



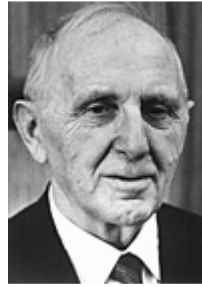
K. Frisch, 1969



Tinbergen, 1969



Samuelson, 1970



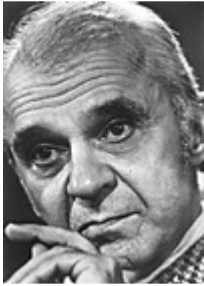
Kuznets, 1971



Hicks, 1972



Arrow, 1972



Leontief, 1973



Myrdal, 1974



*The Bank of Sweden Prize in Economic Sciences in Memory of Alfred Nobel*

Edición 2010



Hayek, 1974



Koopmans, 1975



Kantorovich, 1975



Friedman, 1976



Ohlin, 1977



Meade, 1977



Simon, 1978

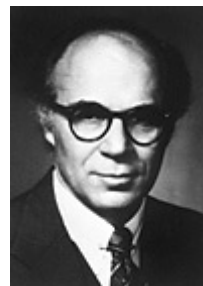


Lewis, 1979



Schultz, 1979

# MANUAL DE ECONOMETRÍA 1º Parte



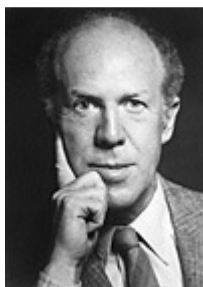
Klein, 1980



Tobin, 1981



Stigler, 1982



Debreu, 1983



Stone, 1984



Modigliani, 1985



Buchanan Jr, 1986



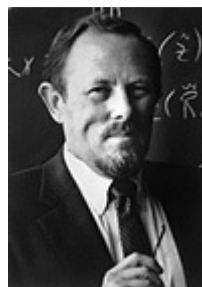
Solow, 1987



Allais, 1988



Haavelmo, 1989



Sharpe, 1990



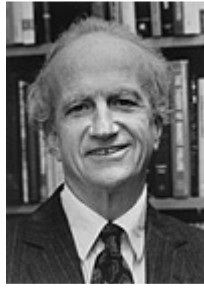
Miller, 1990



Markowitz, 1990



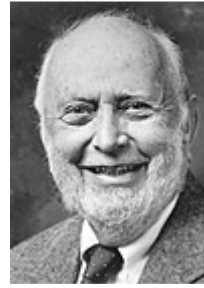
Coase, 1991



Becker, 1992



Fogel, 1993



North, 1993



Harsanyi, 1994



Nash, 1994



Selten, 1994



Lucas Jr., 1995



Mirrlees, 1996



Vickrey, 1996



Merton, 1997



Scholes, 1997



Sen, 1998



Mundell, 1999



Heckman, 2000



McFadden, 2000



Akerlof, 2001



Spence, 2001



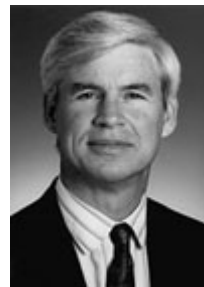
Stiglitz, 2001



Kahneman, 2002



Smith, 2002



Engle III, 2003



Granger, 2003



Kydland, 2004



Prescott, 2004



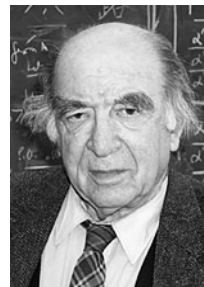
Aumann, 2005



Schelling, 2005



Phelps, 2006



Hurwicz, 2007



Maskin, 2007



Myerson, 2007



Krugman, 2008



Ostrom, 2009



Williamson, 2009



2010



MEDAL

## The Bank of Sweden Prize in Economic Sciences in Memory of Alfred Nobel

- 1969 "for having developed and applied dynamic models for the analysis of economic processes". Ragnar Frisch (1/2 of the prize). Norway. University of Oslo, Oslo, Norway. b. 1895 d. 1973. *Jan Tinbergen* (1/2 of the prize). The Netherlands. The Netherlands School of Economics, Rotterdam, the Netherlands. b. 1903 d. 1994
- 1970 "for the scientific work through which he has developed static and dynamic economic theory and actively contributed to raising the level of analysis in economic science". *Paul A. Samuelson*. (Complete prize) USA. Massachusetts Institute of Technology. (MIT). Cambridge, MA, USA. b. 1915
- 1971 "for his empirically founded interpretation of economic growth which has led to new and deepened insight into the economic and social structure and process of development". *Simon Kuznets*. (Complete prize) USA. Harvard University. Cambridge, MA, USA. b. 1901 (in Russia). d. 1985
- 1972 "for their pioneering contributions to general economic equilibrium theory and welfare theory". *John R. Hicks*. (1/2 of the prize). United Kingdom. USA. All Souls College Oxford, United Kingdom b. 1904 d. 1989. *Kenneth J. Arrow* (1/2 of the prize). Harvard University. Cambridge, MA, USA. b. 1921
- 1973 "for the development of the input-output method and for its application to important economic problems". *Wassily Leontief*. USA. Harvard University. Cambridge, MA, USA. b. 1906 (in St. Petersburg, Russia). d. 1999
- 1974 "for their pioneering work in the theory of money and economic fluctuations and for their penetrating analysis of the interdependence of economic, social and institutional phenomena". *Gunnar Myrdal*. (1/2 of the prize). Sweden. b. 1898 d. 1987. *Friedrich August von Hayek*. (1/2 of the prize). United Kingdom. b. 1899 (in Vienna, Austria) d. 1992
- 1975 "for their contributions to the theory of optimum allocation of resources". *Leonid Vitaliyevich Kantorovich* (1/2 of the prize). USSR. Academy of Sciences. Moscow, USSR. b. 1912 d. 1986. *Tjalling C. Koopmans*. (1/2 of the prize) USA. Yale University. New Haven, CT, USA. b. 1910 (in 's Graveland, the Netherlands) d. 1985
- 1976 "for his achievements in the fields of consumption analysis, monetary history and theory and for his demonstration of the complexity of stabilization policy" *Milton Friedman*. (Complete prize) USA. University of Chicago. Chicago, IL, USA. b. 1912
- 1977 "for their pathbreaking contribution to the theory of international trade and international capital movements" *Bertil Ohlin* (1/2 of the prize). Sweden. Stockholm School of Economics. Stockholm, Sweden. b. 1899 d. 1979. *James E. Meade* (1/2 of the prize). United Kingdom. University of Cambridge Cambridge, U. K. b. 1907 d. 1995
- 1978 "for his pioneering research into the decision-making process within economic organizations". *Herbert A. Simon*. (Complete prize) USA. Carnegie Mellon University. Pittsburgh, PA, USA. b. 1916 d. 2001
- 1979 "for their pioneering research into economic development research with particular consideration of the problems of developing countries". Theodore W. Schultz (1/2 of the prize) USA. University of Chicago. Chicago, IL, USA. b. 1902 d. 1998. Sir Arthur Lewis (1/2 of the prize). United Kingdom. Princeton University. Princeton, NJ, USA. b. 1915 (in Saint Lucia) d. 1991
- 1980 "for the creation of econometric models and the application to the analysis of economic fluctuations and economic policies". *Lawrence R. Klein*. (Complete prize) USA. University of Pennsylvania. Philadelphia, PA, USA. b. 1920
- 1981 "for his analysis of financial markets and their relations to expenditure decisions, employment, production and prices". *James Tobin*. (Complete prize) USA. Yale University. New Haven, CT, USA. b. 1918 d. 2002
- 1982 "for his seminal studies of industrial structures, functioning of markets and causes and effects of public regulation". *George J. Stigler*. (Complete prize) USA. University of Chicago. Chicago, IL, USA. b. 1911 d. 1991
- 1983 "for having incorporated new analytical methods into economic theory and for his rigorous reformulation of the theory of general equilibrium". *Gerard Debreu*. (Complete prize) USA. University of California. Berkeley, CA, USA. b. 1921 (in Calais, France) d. 2004
- 1984 "for having made fundamental contributions to the development of systems of national accounts and hence greatly improved the basis for empirical economic analysis". *Richard Stone*. (Complete prize) United Kingdom. University of Cambridge. Cambridge, U. K. b. 1913 d. 1991
- 1985 "for his pioneering analyses of saving and of financial markets". *Franco Modigliani*. (Complete prize) USA. Massachusetts Institute of Technology (MIT). Cambridge, MA, USA. b. 1918 (in Rome, Italy) d. 2003
- 1986 "for his development of the contractual and constitutional bases for the theory of economic and political decision-making". *James M. Buchanan Jr.* (Complete prize) USA. Center for Study of Public Choice. Fairfax, VA, USA. b. 1919
- 1987 "for his contributions to the theory of economic growth". *Robert M. Solow*. (Complete prize) USA. Massachusetts Institute of Technology (MIT). Cambridge, MA, USA. b. 1924
- 1988 "for his pioneering contributions to the theory of markets and efficient utilization of resources". *Maurice Allais*. (Complete prize) France. École Nationale Supérieure des Mines de Paris. Paris, France. b. 1911
- 1989 "for his clarification of the probability theory foundations of econometrics and his analyses of simultaneous economic structures". *Trygve Haavelmo*. (Complete prize) Norway. University of Oslo. Oslo, Norway. b. 1911 d. 1999
- 1990 "for their pioneering work in the theory of financial economics". *Harry M. Markowitz* (1/3 of the prize). USA. City University of New York. New York, NY, USA. b. 1927. *Merton H. Miller*. (1/3 of the prize). USA. University of Chicago. Chicago, IL, USA. b. 1923 d. 2000. *William F. Sharpe*. (1/3 of the prize). USA. Stanford University. Stanford, CA, USA. b. 1934
- 1991 "for his discovery and clarification of the significance of transaction costs and property rights for the institutional structure and functioning of the economy". *Ronald H. Coase*. (Complete prize) United Kingdom. University of Chicago. Chicago, IL, USA. b. 1910

1992 "for having extended the domain of microeconomic analysis to a wide range of human behaviour and interaction, including nonmarket behaviour". *Gary S. Becker*. (🏆 Complete prize) USA. University of Chicago. Chicago, IL, USA. b. 1930

1993 "for having renewed research in economic history by applying economic theory and quantitative methods in order to explain economic and institutional change". *Robert W. Fogel*. (🏆1/2 of the prize). USA. University of Chicago. Chicago, IL, USA. b. 1926. *Douglass C. North* (🏆1/2 of the prize). USA. Washington University. St. Louis, MO, USA. b. 1920

1994 "for their pioneering analysis of equilibria in the theory of non-cooperative games". *John C. Harsanyi* (🏆1/3 of the prize). USA. University of California. Berkeley, CA, USA. b. 1920 (in Budapest, Hungary). d. 2000. *John F. Nash Jr.* (🏆1/3 of the prize). USA. Princeton University. Princeton, NJ, USA. b. 1928. *Reinhard Selten* (🏆1/3 of the prize). Federal Republic of Germany. Rheinische Friedrich-Wilhelms-Universität. Bonn, Federal Republic of Germany. b. 1930

1995 "for having developed and applied the hypothesis of rational expectations, and thereby having transformed macroeconomic analysis and deepened our understanding of economic policy". *Robert E. Lucas Jr.* (🏆 Complete prize) USA. University of Chicago. Chicago, IL, USA. b. 1937

1996 "for their fundamental contributions to the economic theory of incentives under asymmetric information". *James A. Mirrlees* (🏆1/2 of the prize) United Kingdom. University of Cambridge. Cambridge, U. K. b. 1936. *William Vickrey* (🏆1/2 of the prize). USA. Columbia University. New York, NY, USA. b. 1914 (in Victoria, BC, Canada) d. 1996

1997 "for a new method to determine the value of derivatives". *Robert C. Merton*. (🏆1/2 of the prize). USA. Harvard University. Cambridge, MA, USA. b. 1944. *Myron S. Scholes* (🏆1/2 of the prize). USA. Long Term Capital Management. Greenwich, CT, USA. b. 1941 (in Timmins, ON, Canada)

1998 "for his contributions to welfare economics". *Amartya Sen*. (🏆 Complete prize) India. Trinity College. Cambridge, United Kingdom. b. 1933

1999 "for his analysis of monetary and fiscal policy under different exchange rate regimes and his analysis of optimum currency areas" *Robert A. Mundell*. (🏆 Complete prize) Canada. Columbia University. New York, NY, USA. b. 1932

2000 "for his development of theory and methods for analyzing selective samples" James J. Heckman. (🏆1/2 of the prize). USA. University of Chicago. Chicago, IL, USA. b. 1944. And "for his development of theory and methods for analyzing discrete choice". *Daniel L. McFadden* (🏆1/2 of the prize). USA. University of California. Berkeley, CA, USA. b. 1937

2001 "for their analyses of markets with asymmetric information". *George A. Akerlof* (🏆1/3 of the prize). USA. University of California. Berkeley, CA, USA. b. 1940. *A. Michael Spence* (🏆1/3 of the prize). USA. Stanford University. Stanford, CA, USA. b. 1943. *Joseph E. Stiglitz*. (🏆1/3 of the prize). USA. Columbia University. New York, NY, USA. b. 1943

2002 "for having integrated insights from psychological research into economic science, especially concerning human judgment and decision-making under uncertainty". *Daniel Kahneman* (🏆1/2 of the prize). USA and Israel. Princeton University. Princeton, NJ, USA. b. 1934 (in Tel Aviv, Israel). And "for having established laboratory experiments as a tool in empirical economic analysis, especially in the study of alternative market mechanisms" *Vernon L. Smith* (🏆1/2 of the prize) USA. George Mason University. Fairfax, VA, USA. b. 1927

2003 "for methods of analyzing economic time series with time-varying volatility (ARCH)". *Robert F. Engle III* (🏆1/2 of the prize). USA. New York University. New York, NY, USA. b. 1942. And "for methods of analyzing economic time series with common trends (cointegration)". *Clive W.J. Granger* (🏆1/2 of the prize) United Kingdom University of California. San Diego, CA, USA. b. 1934

2004 "for their contributions to dynamic macroeconomics: the time consistency of economic policy and the driving forces behind business cycles". *Finn E. Kydland* (🏆1/2 of the prize). Norway. Carnegie Mellon University. Pittsburgh, PA, USA; University of California. Santa Barbara, CA, USA. b. 1943. *Edward C. Prescott* (🏆1/2 of the prize). USA. Arizona State University. Tempe, AZ, USA; Federal Reserve Bank of Minneapolis. Minneapolis, MN, USA. b. 1940

2005 "for having enhanced our understanding of conflict and cooperation through game-theory analysis". *Robert J. Aumann* (🏆1/2 of the prize) Israel and USA. Center for Rationality, Hebrew University of Jerusalem. Jerusalem, Israel. b. 1930 (in Frankfurt-on-the-Main, Germany). *Thomas C. Schelling* (🏆1/2 of the prize). USA. Department of Economics and School of Public Policy, University of Maryland. College Park, MD, USA. b. 1921

2006 "for his analysis of intertemporal tradeoffs in macroeconomic policy". *Edmund S. Phelps*, (🏆 Complete prize) USA. Columbia University, New York, NY, USA. b. 1933.

2007 "for having laid the foundations of mechanism design theory". *Leonid Hurwicz*. (🏆1/3 of the prize). USA. University of Minnesota. Minneapolis, MN, USA. b. 1917. (in Moscow, Russia). *Eric S. Maskin*. (🏆1/3 of the prize). USA. Institute for Advanced Study Princeton, NJ, USA. b. 1950. *Roger B. Myerson*. (🏆1/3 of the prize). USA. University of Chicago Chicago, IL, USA. b. 1951

2008 "for his analysis of trade patterns and location of economic activity". *Paul Krugman*. (🏆 Complete prize) Princeton University. Princeton, NJ, USA. b. 1953.

2009 "for her analysis of economic governance, especially the commons" *Elinor Ostrom* (🏆1/2 of the prize) Indiana University Bloomington, IN, USA; Arizona State University Tempe, AZ, USA b. 1933; "for his analysis of economic governance, especially the boundaries of the firm" *Oliver E. Williamson* (🏆1/2 of the prize), University of California Berkeley, CA, USA, b. 1932

*INDICE GENERAL*

Capítulo 1. Conociendo la Econometría	7
Capítulo 2. Investigación econométrica	57
Capítulo 3. Planteamiento de una tabla de datos	87
Capítulo 4. Distribuciones en el muestreo	115
Capítulo 5. Diseño de muestreo	161
Capítulo 6. Organización de los datos	183
Capítulo 7. Teoría de la estimación	247
Capítulo 8. Teoría de la decisión	273
Capítulo 9. Introducción al análisis de información	289
Capítulo 10. Simulación en contextos con y sin información histórica	309
Capítulo 11. Análisis exploratorio de datos	437

*PRIMERA PARTE*

***Materiales y Métodos de la Econometría***

**Capítulo 1. Conociendo la Econometría**

**Capítulo 2. El Proceso de Investigación Econométrica**

<i>Primera Parte</i> .....	5
<i>Materiales y Métodos de la Econometría</i> .....	6
Capítulo 1. CONOCIENDO LA ECONOMETRÍA .....	9
1.1. ¿Qué es la Econometría? .....	9
Instrumentos de la econometría.....	11
Objetivos de la Econometría .....	16
1.2. Epistemología de la Econometría.....	17
Síntesis de la evolución de las ideas sobre el método y la epistemología en economía .....	19
a) La visión de los clásicos .....	19
b) Methodestreit .....	21
c) El apriorismo de la escuela austríaca y la obra de Robbins .....	22
d) La reacción de Hutchison y la introducción de las ideas de Popper en Economía .....	23
e) Las ideas de Lakatos .....	27
f) Dos visiones americanas: el instrumentalismo de Friedman y el operacionalismo de Samuelson.....	27
g) McCloskey.....	29
La econometría como herramienta metodológica de la economía .....	30
a) Los comienzos .....	30
b) El debate en torno al método de Tinbergen .....	31
c) La revolución probabilística de Haavelmo .....	32
d) El debate entre Koopmans y los seguidores de Mitchel .....	33
e) El método econométrico tradicional .....	33
Problemas específicos de algunas metodologías econométricas actuales.....	34
a) Sims y los Vectores Autorregresivos (VAR).....	34
b) Leamer y los métodos bayesianos .....	36
c) El método de Hendry .....	36
Reflexiones finales .....	37
1.3. Análisis cuantitativo de la información .....	41
Relaciones entre el análisis de la información, la investigación econométrica y otras disciplinas .....	42
Expectativas racionales vs. Expectativas adaptativas .....	44
1.4. El Premio Nobel en Economía y la Econometría .....	47

1.5. Econometría, sistemas de información y producción de datos .....	51
CASOS DE ESTUDIO, PREGUNTAS Y PROBLEMAS .....	53
Caso 1.1. Formulación de una investigación regional.....	53



## Capítulo 1. CONOCIENDO LA ECONOMETRÍA

### 1.1. ¿Qué es la Econometría?

Schumpeter (1934), al remontar el origen de la econometría a algunos trabajos realizados por economistas en el siglo XVII, manifiesta que

*“tienen en común... el espíritu del análisis numérico. Todos ellos han sido económetras. Su obra ilustra realmente a la perfección qué es la econometría y qué intentan hacer los económetras”<sup>1</sup>.*

A pesar de la afirmación de Schumpeter y aunque Engel y algunos autores neoclásicos pueden considerarse como precursores, la econometría, tal y como se concibe hoy en día, es una disciplina relativamente joven. Los primeros trabajos de naturaleza verdaderamente econométrica aparecieron en el primer tercio del siglo XX, destacando entre otros los de Moore (1914), Working (1927), Cobb y Douglas (1928), Schultz (1928) y Waugh (1928)<sup>2</sup>.

En el desarrollo histórico de la Econometría juega un papel preponderante la fundación de la *Econometric Society* (1930), que se puede describir como *una sociedad internacional para el progreso de la Teoría económica en su relación con la Estadística y las Matemáticas* y que su objeto esencial es, de acuerdo con los estatutos, *favorecer los estudios de carácter cuantitativo que tienden a relacionar el punto de vista teórico con el empírico en la investigación de los problemas económicos, estando inspirados dichos estudios de un espíritu metódico y riguroso, semejante al que prevalece en las ciencias naturales.*

En diciembre de 1930, Alfred Cowles III, un cuantitativista orientado al análisis de inversiones, lideró la fundación en Cleveland de la “Econometrics Society”, que es una Sociedad Internacional para el progreso de la teoría económica en sus relaciones con la estadística y las matemáticas. Su objeto esencial es favorecer los puntos de vista teóricos y empíricos en la exploración de los problemas económicos y que están inspirados en un estudio metódico y riguroso semejante al que ha prevalecido en las ciencias naturales. Toda actividad susceptible de favorecer mediata o inmediatamente tal unificación en los estudios económicos teóricos y empíricos cae bajo el campo de

---

<sup>1</sup> Schumpeter, *Historia del Análisis Económico: Los económetras y Turgot*. Estudia en ese capítulo, entre otros, la obra de Petty, Cantillon y Quesnay.

<sup>2</sup> Los primeros trabajos econométricos no contaban con el cuerpo teórico del que se dispone hoy en día. De hecho, muchos de los aspectos más sencillos de los modelos econométricos fueron forjándose poco a poco. Una exposición del desarrollo histórico de los modelos formales en econometría puede verse en el libro de Morgan (1990).

acción de la Sociedad. Además, en 1932, Cowles formó la Cowles Commission for Research in Economics<sup>3</sup>.

La citada fundación y la publicación de la revista *Econometrica* (1933) fueron dos grandes hitos en la historia de la econometría. Frisch, primer editor de la revista, expone en el editorial cuál es su principal objetivo y pasa a explicar en qué consiste la econometría:

*... no es lo mismo que la estadística económica. Tampoco es idéntica a lo que llamamos teoría económica general, aunque una parte considerable de esta teoría tiene un carácter cuantitativo. Tampoco, debería ser tomada como sinónimo de la aplicación de las matemáticas a la economía. La experiencia ha demostrado que cada uno de esos tres puntos de vista, el de la estadística, el de la teoría económica y el de las matemáticas, es una condición necesaria pero no suficiente por sí misma para el entendimiento real de las relaciones cuantitativas en la vida económica moderna. Es la unificación de las tres, la que es poderosa. Es esta unificación la que constituye la econometría<sup>4 5</sup>.*

Aunque esta idea persiste, la noción que se tiene de la econometría no es muy clara, como se puede comprobar leyendo las definiciones que proporcionan casi todos los libros de texto. Estas varían desde algunas ciertamente complejas, hasta otras relativamente sencillas, como la de Intriligator et al. (1996): *Econometría es la rama de la economía relacionada con la estimación empírica de relaciones económicas*<sup>6</sup>.

Haciendo uso de estos pioneros argumentos, para nosotros, la Econometría **es la aplicación de métodos matemáticos y estadísticos a tablas de datos, que contienen "unidades" (de observación) por características observables de las mismas ("variables"), con el propósito de dar contenido empírico a las teorías económicas planteadas en modelos, verificándolas a partir del estudio de la semejanza entre unidades y la relación entre variables, en un espacio y tiempo específico.**

Encontramos, en este intento por definir a la econometría, algunos elementos subyacentes de la misma. En primer lugar, de dicha definición surgen **métodos matemáticos y estadísticos**. En este libro, veremos cuáles son los de aplicación directa a la econometría, en particular, los desarrollaremos en su directa aplicación al "conocimiento" de la misma. Estos métodos se aplican a **tablas de datos** (Figura 1.1).

<sup>3</sup> Para un mayor conocimiento sobre la fundación de la Sociedad Econométrica se puede consultar Alfonso G. Barbancho, Capítulo 1, Fundamentos y posibilidades de la Econometría, Ediciones Ariel, Barcelona, 1962.

<sup>4</sup> Frisch recibió en 1969 el primer Premio Nobel de Economía (junto con Jan Tinbergen) por su importante contribución al desarrollo de la econometría.

<sup>5</sup> El primer editorial de *Econometrica* y un buen número de los primeros trabajos econométricos han sido recopilados por A. Darnell (1994) en su obra *The History of Econometrics*.

<sup>6</sup> La falta de consenso sobre la definición de la econometría ha dado lugar a un artículo sobre el tema (Tintner, 1953).

Para “construir” una tabla de datos hace falta tener un problema a investigar que, para el economista, será un problema económico, provendrá de la teoría económica y, según su experiencia, planteará un modelo económico, que tratará de aplicar a un espacio y tiempo determinado.

		<i>variables</i>		
		1	$j$	$k$
<i>Unidades de observación</i>	1	$x_{ij}$		
	⋮			
	$i$			
	⋮			
	$n$			

**Figura 1.1. Tabla de datos**

Para el análisis de esta tabla de datos, necesitará de la “herramienta” econométrica, para estudiar la semejanza entre unidades de observación y la relación entre variables. Las primeras, representarán países, familias, tiempo, personas, etc., en general, en este texto, las denominaremos “individuos” o unidades. Las segundas, seguramente, en su mayoría económicas, definirán “su” modelo económico (a estudiar) a los efectos de darle contenido empírico a la teoría que está desarrollando o estudiando.

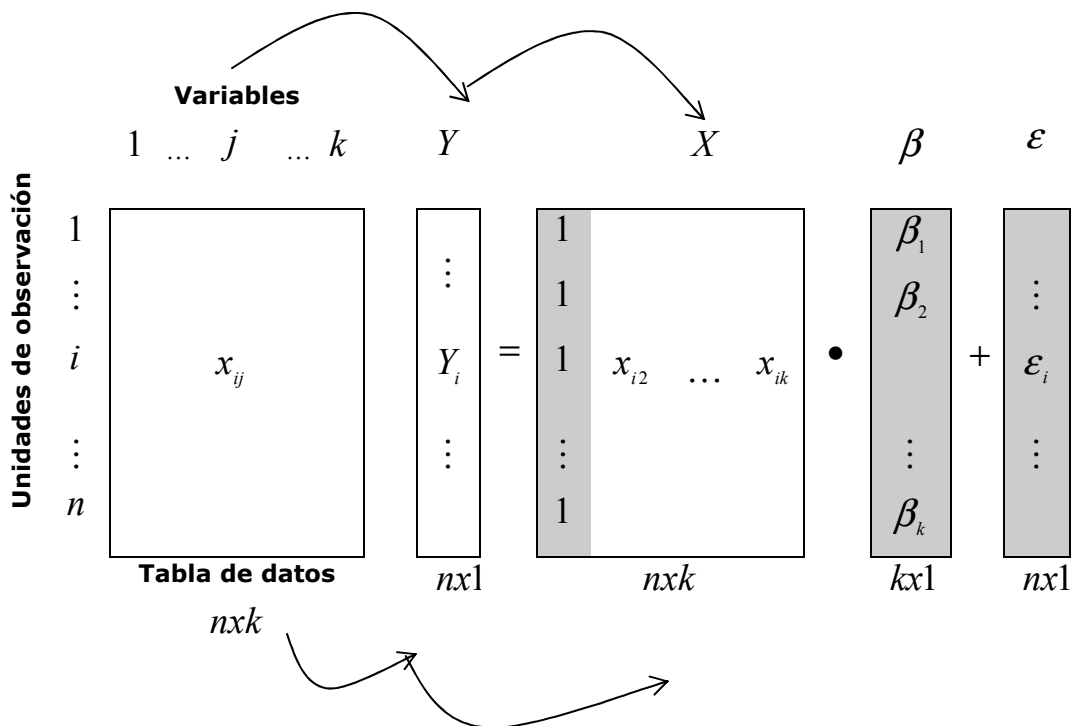
Indudablemente, este trabajo empírico lo hará siguiendo una metodología. Es, justamente aquí, donde el economista se ubica como investigador social y debe aplicar, en su estudio, un proceso de investigación. De esta manera, la econometría se vincula con la metodología de investigación, habida cuenta, que la primera es, en esencia, la aplicación de métodos cuantitativos en una investigación económica.

Pero, además, la econometría ha desarrollado métodos, sobre todo a partir de los últimos años del siglo pasado, que le ha permitido, a un amplio espectro de investigadores, provenientes de las ciencias, tanto sociales, humanas o exactas, como empresariales o biológicas, verificar o refutar modelos representados a través de relaciones funcionales entre variables. Aunque, trabajaremos, fundamentalmente, con las aplicadas a los negocios y economía, no debemos dejar de mencionar el importante aporte de la econometría en otras áreas de investigación.

#### *Instrumentos de la econometría*

Últimamente, se han hecho notables avances para refutar la idea de que la econometría solo utiliza métodos estadísticos que provienen de la inferencia estadística. Por el contrario, los econometristas también

utilizamos los que provienen de la estadística descriptiva. Esto tiene que ver con la idea, cada vez más aceptada, de que la econometría más que una ciencia es un arte en el cual el investigador se apoya para describir de manera, más o menos, ilustrada las características de las unidades que observa y los datos con que trabaja en un espacio y tiempo determinado.



**Figura 1.2 Especificación de un modelo lineal**

Específicamente, el econometrista trabaja para generar un “modelo” que relacione funcionalmente (de forma lineal o no) las variables que generan datos sobre las unidades (personas, hogares, regiones, países o tiempo) que observó. En apretada síntesis, el proceso utilizado descansa en la representación que hacemos en la figura 1.2. y que, dadas  $n$  unidades de observación y  $k$  variables, se puede resumir en la siguiente relación funcional:

$$(1.1) \quad Y = X\beta + \varepsilon$$

Como vemos en la figura, antes de especificar un modelo el investigador deberá, cuanto menos, hacer un estudio exploratorio, descriptivo y correlacional de los datos con que cuenta para su trabajo empírico. En la especificación, hemos agregado tres vectores, distinguidos con un sombreado, que la completan. Ellos son, el *vector unidad*, el *vector de parámetros* -que deberemos estimar-, y un *vector de variables aleatorias* -que se ha especificado con la letra griega  $\varepsilon$ -.

Conviene detenerse por un momento en la misma. El lector atento se habrá dado cuenta que se deduce de la tabla de datos definida en la figura 1.1. Hemos querido destacar con flechas, el hecho de que el vector  $Y$ , que más adelante definiremos como variable endógena, y la matriz de información,  $X$ , que como veremos forma la parte exógena de la especificación, provienen de las  $k$  variables que constituyen nuestra tabla de datos, más un vector de unos. Sobre los otros dos vectores profundizamos más adelante.

Desde esta perspectiva, la econometría no es mas, ni mucho menos, que un método de análisis estadístico multivariado que se complementa con otras técnicas igualmente potentes, como son el análisis factorial y sus derivados.

Por otro lado, utiliza los principios y métodos del análisis matemático y del álgebra lineal para elaborar sus propias técnicas formales. Se apoya, fundamentalmente, en la teoría de funciones y relaciones lineales para abordar temas tan especiales como son los de la especificación de un modelo o el de su confirmación.

Pero también, si pensamos en el contenido de las tablas, matrices y vectores que hemos presentado en la figura 1.2, nos damos cuenta que, salvo por el contenido de los vectores unidad, de parámetros y aleatorio, se trata de datos. Estos datos, –directamente asociados a las unidades de observación, reflejan características de las mismas, denominadas variables –, deben provenir de *información* pertinente, integrada y actualizada. La información es otro instrumento del que se debe ocupar el investigador. Se puede definir, como “la acción y efecto de informar, es decir, un conjunto de informes sobre cosas concretas”. Mientras que el dato, que deriva de esos informes, “es el detalle que sirve de base a un razonamiento o a una investigación, es decir, cada una de las cantidades conocidas que constituyen la base de un problema”<sup>7</sup>.

Al respecto, Barbancho (1962) dice: “es un principio generalmente admitido el que el econométra posee un conocimiento perfecto de las observaciones estadísticas que va a utilizar para la estimación de los parámetros de sus modelos. Así, pues, esta cuestión no suele considerarse explícitamente. Nosotros, sin embargo, queremos detenernos un momento en ella por estimarlo de extraordinaria importancia. En efecto, un mal uso de los datos empíricos no es necesario probar que nos conducirá a resultados pocos fiables, aunque el modelo y el método de estimación sean óptimos”<sup>8</sup>.

Coincidiendo estrechamente con el razonamiento de este autor, Lucas (1976), premio Nobel de economía 1995, establece, entre otras cuestiones muy importantes, como veremos hacia el final de este

<sup>7</sup> *Diccionario de la lengua española*. Ediciones Planeta S. A, México 1994.

<sup>8</sup> Barbancho, op. Cit., p. 150.

capítulo, "...los modelos econométricos están expuestos a la calidad de la información disponible para realizar las estimaciones"<sup>9</sup>.

Para que la información cumpla con las propiedades mencionadas, deberá provenir de fuentes seguras. Al respecto, es importante considerar el breve y sencillo enunciado que un filósofo social hizo en el siglo XIX "*saber para poder y poder para prever*". El mismo, sintetiza, para el mundo contemporáneo, la crucial relación entre la información y la sociedad. Las estadísticas, siendo no solamente una serie de datos cuantificables, sino también, el conjunto de criterios para identificar, recolectar, sistematizar y analizar tales datos, constituyen un elemento básico para la conducción racional de la misma.

La noción "estadística" se derivó originalmente de la palabra "estado" porque ha sido función tradicional de los gobiernos centrales llevar registros de población, nacimientos, defunciones, cosechas, índices de precios y muchas otras clases de cosas y actividades. Contar y medir estos hechos genera un volumen de datos numéricos. Tradicionalmente, estadística se define como: "*compilación, organización, resumen, presentación y análisis de datos numéricos*"<sup>10</sup>.

La cantidad de datos generados a diario por las instituciones públicas son solamente compilados y se encuentran dispersos en oficinas u organismos siendo pocos los esfuerzos que se realizan para organizarlos, resumirlos, presentarlos, analizarlos y en base a ellos tomar decisiones inequívocas generadoras de hechos que sean coherentes con la realidad económica y social, esto lo vemos, de sobre manera en la micro información, referida a regiones, familias o empresas.

Muchas veces, los agentes económicos, públicos y privados, ignoran la importancia de la información estadística y suministran la menor cantidad de información, lo que hace bastante difícil conocer la realidad económica y social que permita al gobierno tomar decisiones positivas para generar las soluciones demandadas.

Es así que, de los tres componentes del instrumental del análisis económico, en la mayoría de los países, la Estadística Económica ha sido la "cenicienta", tanto para los economistas como para los organismos públicos. Es el campo menos prometedor para los agentes económicos, aunque sea, paradójicamente, el que ha precedido buena parte de la tarea de los pensadores y teóricos.

Es oportuno tener en cuenta, por ejemplo, la revolución que significó para la economía y la política económica, en todo el mundo, poseer un buen sistema de cuentas nacionales, o una matriz de insumo producto desagregada. O también, la ventaja de poseer información sobre el nivel de actividad económica desagregado por sectores de actividad, o el desarrollo que ha tenido en los últimos tiempos la investigación de

<sup>9</sup> Lucas, R. "Econometric Policy Evaluation: A Critique", en su *Studies in Business Cycle Theory*.

<sup>10</sup> Libro de ....

mercados. Ésta, conjuntamente con la investigación económica y la política económica, en la actualidad, serían inconcebibles sino existieran las laboriosas tareas de determinar conceptos cuantificables<sup>11</sup>.

Las carencias del sistema de estadística económica obliga a los entes decisores de la política económica y al sector productivo, a tomar decisiones sin un conocimiento acabado de la base empírica. No obstante, en forma incipiente existen "fuerzas" que se encuentran abocadas en la tarea de producir o demandar esa clase de información. Entre las primeras, es decir las productoras de información estadística actualizada, se encuentran instituciones públicas y privadas, que con esfuerzo tratan de compilar datos a través de programas de investigación porque se han dado cuenta de que es la única manera que la comunidad conozca la realidad por la que atraviesa y trate de solucionar sus problemas<sup>12</sup>.

**Ejemplo 1.1.:** En el ámbito regional, el Programa Institucional de Investigación y Extensión de la Facultad de Ciencias Económicas de la Universidad Nacional de Río Cuarto, tuvo a su cargo la medición del INEVE (Índice de Evolución Económica) y, oportunamente, se desarrolló el IDHR (Índice de Desarrollo Humano Regional), el Censo Industrial del Gran Río Cuarto, estudios sobre el Sector Agropecuario del Sur de Córdoba y el Directorio de Comercio Exterior del Sur de Córdoba, entre otros. El objetivo final consistió en permitir la descripción del funcionamiento agregado de Río Cuarto y zona de influencia en función de la observación de los fenómenos que en ella ocurren de lo cual surge la necesidad de contar con información estadística pertinente, integrada y actualizada.

**Ejemplo 1.2 :** En el ámbito internacional la Red de Investigación Urbana en la Unión Europea (NUREC) conjuntamente con la Oficina de Estadística de las Naciones Unidas (UNSTAT); Centro de la Naciones Unidas para los Asentamientos Humanos (HABITAT); Instituto Internacional de Estadística (ISI) y la Unión Internacional de Autoridades Locales (IULA) reunieron información estadística, sobre ciudades de 100000 habitantes o más. Con esta información publican conjuntamente el Internacional Statistical Yearbook of Large Towns. Este Yearbook se ha convertido en un documento sumamente útil para administradores, planificadores urbanos, legisladores, estadísticos, académicos y otros usuarios que desean obtener datos sobre ciudades. En la página [www.uni-duisburg.de/duisburg/lcsp.htm](http://www.uni-duisburg.de/duisburg/lcsp.htm) puedes acceder a la información.

Samuelson (1948), premio Nobel de Economía 1970, expresa que la investigación, la instrucción pública y el pleno empleo pueden acelerar el desarrollo económico de un país. Pero, refiriéndose concretamente a la investigación, observa que esta es la medida que suscita menos controversias entre las destinadas a acelerar el crecimiento. Dice: "...Todo el mundo está de acuerdo en que la ampliación de la investigación -en la ciencia pura y aplicada en la técnica y en la administración- puede rendir unos dividendos sociales muy altos en materia de productividad... las empresas ya realizan buena parte de la investigación y puede alentárselas para que la intensifiquen. De todas formas los directores de las empresas no son tan tontos (sic) que no ven hasta que punto la investigación beneficia a sus respectivas firmas..." y concluye: "...la medida menos discutida de las que pudieran acelerar el desarrollo económico es el fomento y subsidio de la investigación científica..." que para el crecimiento de las regiones

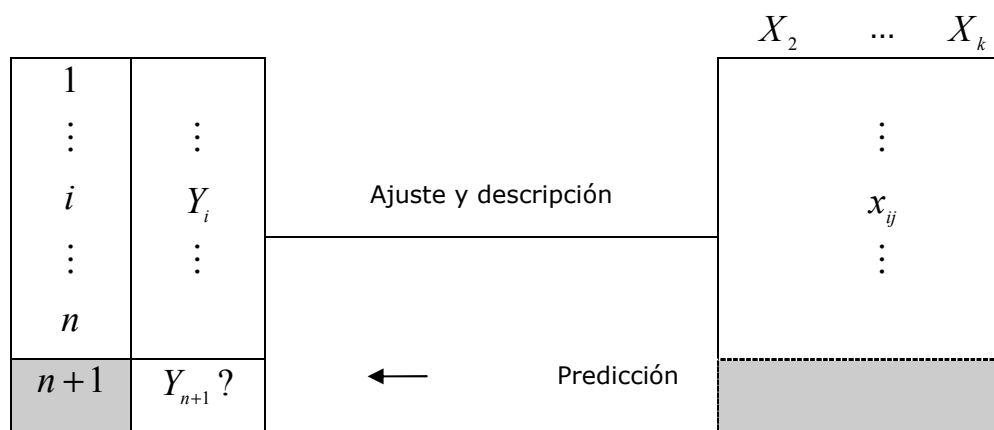
<sup>11</sup> Ferrucci

<sup>12</sup> Resumen de tesis.

debe estar basada, en gran parte, en el apoyo a la generación de estadísticas económicas públicas y a la implementación de investigaciones de mercados y econométricas por parte de los agentes económicos<sup>13</sup>.

### Objetivos de la Econometría

De acuerdo a lo anterior el eje central de la investigación econométrica es el planteamiento de una tabla de datos. Este es nuestro primer objetivo como econométricos. Sin entrar, por ahora, en mayores precisiones respecto a la misma, diremos, sin embargo que, en su planteo inicial, contiene unidades de observación y variables; pero, luego del proceso de observación y recolección, "datos". Las unidades, las variables y los datos dan origen a ecuaciones, funciones o modelos que describen una situación actual que es de interés para el investigador o le permiten realizar predicciones. Estos son los objetivos principales de la investigación del econométrico.



**Figura 1.3 Descripción y Predicción**

Como vemos en la figura 1.3, los métodos econométricos brindan la posibilidad de **describir** una situación económica en particular, para verificar teorías o comprobar empíricamente las mismas. Realizada esa descripción, también pueden ser utilizados para **predecir** o realizar pronósticos sobre dicha situación.

Coincidentemente con Christ (1966), tomamos como objetivo de la econometría la producción de proposiciones económicas cuantitativas que *expliquen o describan* las variaciones de variables ya observadas, o que *pronostiquen o predigan* las variaciones aun no observadas, o que hagan ambas cosas a la vez<sup>14</sup>.

<sup>13</sup> Samuelson, P. *Curso de economía moderna*. Editorial Aguilar 1972. p. 903.

<sup>14</sup> Christ, C. *Modelos y métodos econométricos*. Cowles Foundation for research in Economics at Yale University. 1966. p. 28.



<b>Clasificación</b>	<b>Tipos</b>	<b>Incorporadas al análisis ...</b>
<b>UNIDADES DE OBSERVACIÓN</b>		
De corte transversal (o de espacio)	Personas, Empresas, Instituciones, Regiones, Países, etc.	...por selección Probabilística o casual
De tiempo (o longitudinal)	Anual, Semestral, mensual, semanal, diario, etc.	
De panel	Combinación de corte transversal y de tiempo	
<b>VARIABLES</b>		
Cuantitativas	Aleatorias o determinista	... de manera Estocásticas o no estocásticas
Cualitativas		
<b>DATOS</b>		
Reales	Discretos o continuos	... provenientes de fuentes de información Secundarias o primarias
Catagóricos	Códigos	
Binarios	0 ó 1	
<b>ECUACIONES O FUNCIONES</b>		
Estocásticas o no estocásticas	Exactas o no exactas	... de forma Individual o conjunta
Lineales o no lineales	Simplees o múltiples	
<b>MODELOS</b>		
Teóricos o empíricos	Dinámicos o Estáticos	... para describir o predecir una situación

**Cuadro 1.1. Clasificación, tipos y forma de incorporación al análisis**

Lo que hemos expresado, se puede mostrar en un cuadro resumen (cuadro 1.1.) que es la base de nuestro trabajo. El mismo, refleja cada una de las partes que integra, o que son fuentes de, la investigación econométrica. Siguiendo lo comentado al dar el concepto de tabla de datos, se ilustra, la clasificación en: unidades de observación, variables, datos, ecuaciones o funciones y modelos. También, el cuadro informa, sobre la tipología de las mismas y la forma en que podemos incorporarlas a la investigación.

## 1.2. Epistemología de la Econometría

Navarro (1997) al referirse a las relaciones entre economía, econométrica y epistemología dice<sup>15</sup>:

*El objetivo de este trabajo es analizar algunos de los problemas epistemológicos que enfrenta hoy la economía, en cuanto a lo relacionado con la comprobación empírica de las teorías, para lo cual son la econométrica y la estadística las herramientas más utilizadas. El avance en la utilización de los métodos cuantitativos ha sido muy rápido a partir del nacimiento de la econométrica en la década del 30, y su crecimiento ha resultado exponencial en los últimos años. Cuatro circunstancias se han conjugado para que ello ocurra. Por una parte el desarrollo de la teoría económica, que genera teorías que al expresarse matemáticamente son susceptibles de ser "testeadas", como ocurre con los modelos macroeconómicos de raíz keynesiana. En segundo lugar, el desarrollo de las técnicas estadísticas, que comienza a principios de siglo pasado y llega a su climax con el desarrollo de la econométrica en Inglaterra y los Estados Unidos al comienzo de la década de los años treinta. En tercer*

<sup>15</sup> Navarro, A. *Reflexiones sobre la relación entre economía, econométrica y epistemología*. Trabajo presentado en la conferencia de incorporación a la Academia Nacional de Ciencias Económicas. Buenos Aires. 6/08/1997.

*término, la disponibilidad de información estadística, que mejora su calidad y abarca cada vez más campos, debido en gran parte al desarrollo de las técnicas relacionadas con las cuentas nacionales. Por último un elemento que hace posibles los cálculos complicados: la aparición de los computadores, que si bien se produce durante los años cincuenta se convierte en un elemento de utilización general durante la década siguiente, aunque sólo disponible con restricciones en las universidades y organismos estatales de los países desarrollados. Después el crecimiento es vertiginoso, hasta que hoy todos tenemos acceso a computadores avanzados, a programas que permiten hacer en minutos lo que antes hubiera requerido meses de trabajo y a las posibilidades que ofrece la interconexión, que son ilimitadas<sup>16</sup>.*

*Por otra parte la visión metodológica y epistemológica de la economía es influida por la evolución del pensamiento filosófico. La crisis de la metafísica de fines del siglo pasado había entronizado al positivismo, que fue evolucionando de formas ingenuas a otras más complejas. La idea básica que inspira a esta forma de pensar en su origen está bien definida por Hume, al final de su Enquiry: "...Si tomamos cualquier volumen, sobre teología o metafísica, nos podemos preguntar ¿contiene algún tipo de razonamiento abstracto relativo a la cantidad o el número? No. ¿Contiene algún tipo de razonamiento experimental relativo a la realidad o a la existencia? No. Arrojámosle, pues, al fuego, porque no puede contener sino sofistería e ilusión."*

*En este párrafo, según sostiene Feigl (1966), está representado el espíritu del Círculo de Viena. La negación de la metafísica los llevó a la idea que los enigmas del universo sólo podrían ser develados, en caso de ser posible, por medio de la ciencia. El rol de la filosofía quedaba reducido a la epistemología, que debía pavimentar el camino de la ciencia, mediante la utilización de herramientas tales como la lógica simbólica, siguiendo los desarrollos que hicieron Whitehead y Russell, y mediante el lenguaje, en su función pictórica o simbólica, de acuerdo a Wittgenstein (1995). Mediante estos dos caminos lograrían lo que ellos denominaban la "clarificación" de los conceptos básicos.*

*Estas ideas tuvieron generalizada aceptación hasta principio de los años treinta, cuando comenzaron a ser cuestionadas, entre otros, por Popper, que intenta resolver la llamada "paradoja de Hume"<sup>17</sup>, introduciendo la idea de falsificación de las teorías, en lugar de su verificación. Revive la tradición filosófica del realismo y realiza una demoledora crítica al determinismo y al marxismo. Tiene sin embargo algo en común con los empiristas lógicos del Círculo de Viena: su espíritu crítico y sus standards estrictos en la forma en que debe realizarse el trabajo científico, así como el criterio de demarcación estricto que propone. Debemos ser audaces en la creación de las teorías, nos dice, pero muy estrictos en su crítica.*

*Posteriormente se adopta una visión más descriptiva, como ocurre con los sistemas de Kuhn (1978) y de Lakatos (1971), que tienen presente la*

---

<sup>16</sup> Ver Sosa-Escudero (1997), donde se hace una descripción de las posibilidades que ofrece Internet para los investigadores que usan métodos cuantitativos.

<sup>17</sup> Si sostengo que todos los cisnes son blancos, no puedo aceptar esta proposición como cierta porque no puedo saber que el próximo cisne será también blanco, aunque hubiera controlado el color de muchos de ellos. Por lo tanto lo que puedo hacer es tratar de falsificar esa proposición, encontrando un cisne de otro color. Mientras tanto puedo aceptar esa proposición transitoriamente. Es decir que nada es definitivo y todas las teorías están en el banquillo de los acusados, y solamente las que sobrevivan darwinianamente a los fiscales serán por el momento aceptadas. Ver Popper (1934).

*evolución de la ciencia en la historia y ven a la ciencia como un organismo vivo que evoluciona y se relaciona, en el caso del primero, con las condiciones sociales, mientras que el segundo abandona algunas de las hipótesis más estrictas de Popper, y al distinguir entre "hard core" y "protective belt" permite que las hipótesis correspondientes al primero no sean "testeadas", con lo que permite insertar a las ciencias sociales dentro del mundo de la ciencia, como veremos más adelante.*

### *Síntesis de la evolución de las ideas sobre el método y la epistemología en economía*

#### *a) La visión de los clásicos*

*La preocupación por los problemas metodológicos nace con la ciencia económica. Adam Smith, conocedor de las teorías de Newton, trató de aplicarlas con poco éxito a las ciencias sociales. Pensó que un principio, como el de la gravitación universal, podría ser el elemento básico de su trabajo, y así aplicó el principio del interés personal en "La Riqueza de las Naciones", y el principio de la simpatía hacia el prójimo en su "Teoría de los Sentimientos Morales"<sup>18</sup>.*

*La economía de Ricardo y Malthus llevó, hacia 1830, a enconadas discusiones entre los economistas y los reformadores sociales de la época, lo que hizo que Stuart Mill se preocupara por primera vez en forma sistemática de este tipo de problemas. Su ensayo "On the Definition of Political Economy", escrito en 1836, define su pensamiento acerca del método de la economía, y lo hace de una forma por demás particular y sagaz. Sigue la tradición inglesa de Hume y de Bentham<sup>19</sup>, y recibió alguna influencia de A. Comte<sup>20</sup>, por lo que basa el conocimiento científico en la experiencia, al par que descarta el apriorismo de raíz kantiana, como puede verse en el siguiente párrafo de su Autobiografía: "Pero el System of Logic proporcionaba algo que era muy necesario: un libro de texto de la doctrina opuesta, de aquella que deriva todo conocimiento de la experiencia...La idea de que las verdades externas al espíritu puedan conocerse por intuición o conciencia, independientemente de la observación y la experiencia, estoy persuadido de que es en estos tiempos el gran sostén intelectual de doctrinas falsas."*

*Sin embargo, al referirse a las ciencias sociales, propone una forma diferente, ante la imposibilidad de realizar los experimentos controlados de las ciencias físicas<sup>21</sup>. Parte de un principio "a priori", que establece que el motor de la acción humana es el deseo de riqueza, que está limitado por el deseo de consumir y por la inclinación al ocio. Estos principios*

<sup>18</sup> Ver Ross (1995).

<sup>19</sup> La obra de Bentham tuvo una influencia muy grande en su pensamiento, y también en el de su padre, como cuenta en su "Autobiografía". Este autor trata de hallar la forma de medir la utilidad, al menos así lo decía explícitamente en sus primeros trabajos, y en forma más atenuada más adelante, (Ver Bentham, 1995, pp.69). Es decir que este pensador, uno de los pilares en que se asienta la economía clásica, tiene en mente la posibilidad de aplicar mediciones estadísticas, aún en cuestiones tan poco aptas para ello, como son placer y el dolor. (Ver Mill, 1948).

<sup>20</sup> En la misma obra relata su relación intelectual con Comte: si bien despierta su entusiasmo al principio, pronto se decepciona ante sus excentricidades.

<sup>21</sup> La escasa utilización de las técnicas cuantitativas en la época en que nace la economía clásica está bien ejemplificada por Schumpeter (1971), cuando dice "...un gran paso que contribuyó a esclarecer la importancia del método se dió en 1798 cuando Sir George Evelyn presentó una comunicación a la Sociedad Real en la que, después de presentar excusas por tratar un tema en términos muy por debajo de lo que exigía la dignidad de aquella corporación, utilizó un número índice para medir la devaluación de la moneda...".

vienen de nuestra experiencia o de la de nuestros vecinos y no pueden ser verificados. Se trata de dividir el trabajo de tal manera de estudiar las causas una a una, a diferencia de lo que sucede en la física, y es preciso analizar el efecto de las causas una a una porque de lo contrario el trabajo científico es imposible. De esta manera el principio abstracto, a priori, obtenido por introspección sólo puede observarse en la realidad en algunas ocasiones, dada la presencia de muchas otras causas simultáneamente. Las leyes económicas no pueden ser sino verificadas, pero el hecho de que una circunstancia particular no la verifique no implica que la ley deba descartarse.

Posteriormente Cairnes desarrolló ideas similares, pero despegándose un poco más de la visión empirista de Stuart Mill y de Adam Smith, para acercarse a Ricardo y Senior. Básicamente sostiene que las teorías económicas no pueden ser testeadas, sino que la verificación tiene por objeto establecer si las "disturbing causes" pueden dar razón de la disparidad entre los hechos y las teorías, y en caso de existir ello demostraría que la teoría ha sido inapropiadamente aplicada, ya que no puede ser falsa.

Keynes (1955) realiza un esfuerzo conciliatorio entre posiciones que ya encierran algunos elementos contradictorios que han de estallar después de la Primera Guerra, haciendo una síntesis de la metodología de los economistas clásicos. Marshall (1948), en sus "Principios" si bien no se refiere con mucha profundidad a las cuestiones metodológicas ni a la comprobación de las teorías, sigue una línea conciliadora con la escuela histórica, aunque la economía que propone es una en la cual se parte de algunos principios básicos derivados de la introspección, que son elaborados matemáticamente, aunque se exponen prescindiendo de esa herramienta, y que luego se verifican con ejemplos, y si bien sostiene el principio de la unidad de la ciencia atribuye a las leyes económicas el carácter de tendencias<sup>22</sup> que son mucho más imprecisas que en la física. Se refiere al ejemplo de la ley de la gravedad, que dada la existencia de fricción y de otras fuerzas se transforma en una tendencia, igual que - sostiene- sucede con las mareas, que tienen un componente aleatorio.

Ninguno de los autores mencionados se refiere a las aplicaciones de la estadística porque todavía no ha habido desarrollos en esta dirección. Sin embargo en Oxford ocurre algo un poco diferente, porque las ideas de Edgeworth son distintas en varios aspectos, sobre todo en cuanto a la aplicación de métodos cuantitativos en economía, de lo que es un pionero, y si bien en ambos enfoques se utilizan intensamente las matemáticas, en este caso se comienzan a aplicar los métodos estadísticos.

Pareto (1848-1923), que no puede definirse estrictamente como un continuador de la tradición clásica, realiza tempranamente un análisis del rol de la comprobación empírica que resulta por demás interesante. En su *Manual de Economía Política*, (Pareto, 1945) sostiene que hay tres formas de hacer economía: buscando el bienestar de una persona o de una empresa, el de toda la sociedad o "solamente la búsqueda de uniformidades que presentan los fenómenos sin tener como fin ninguna utilidad práctica directa"<sup>23</sup>, es decir que se persigue la finalidad exclusivamente científica de aumentar el conocimiento por sí. Pareto aboga por esta última posición, y -afirma- se separa así de Smith y de

<sup>22</sup> Ver en Popper (1995) un exhaustivo análisis del concepto de tendencia en los economistas clásicos

<sup>23</sup> Ver Pareto (1945), pág 8.

*Stuart Mill, quienes si bien se refieren en la mayor parte de los casos al tercero de los enfoques enunciados, también adoptan en varios pasajes de sus obras al primero y al segundo. Cuando se refiere más adelante a la comprobación empírica lo hace de esta manera: "Hay que añadir que las teorías no son sino medios para conocer y estudiar los fenómenos. Una teoría puede ser buena para alcanzar cierto fin. Otra puede serlo para alcanzar otro. Pero de todas maneras deben estar de acuerdo con los hechos, porque si no, no tendrían ninguna utilidad. El estudio cualitativo debe ser sustituido por el estudio cuantitativo, y buscar en qué medida la teoría se aparta de la realidad. De dos teorías escogeremos la que se aparte menos. No olvidemos jamás que una teoría no debe ser aceptada sino temporariamente; la que se tiene por verdadera hoy debe ser abandonada mañana. La ciencia es un perpetuo cambio."*

*Este párrafo describe muy bien su pensamiento, que es bastante coincidente con las posiciones más difundidas en la actualidad: el carácter provisional de las teorías, la necesidad de confrontarlas con los hechos, y utilizar esto para la selección de teorías rivales.*

*El común denominador del pensamiento clásico no parece ser muy proclive a los estudios empíricos, más allá de lo necesario para encontrar ejemplos de cómo funcionan las teorías, pero paulatinamente van apareciendo economistas más o menos coincidentes con la posición clásica que van tomando una posición más flexible, al par que se va tomando conciencia de que se está ante un problema de difícil solución.*

#### b) Methodestreit

*El predominio de la escuela clásica inglesa venía siendo cuestionado desde mediados de siglo XIX, entre otros por la escuela histórica alemana, que comienza en 1843 con Roscher, cuya obra es continuada por Hildebrand y Knies . Sin embargo alcanza su máximo esplendor con Schmoller (1838-1917), quien critica el universalismo de las teorías clásicas, así como la insuficiencia y estrechez de su psicología y el abuso del método deductivo. Sugerían una economía basada en la historia, y realizan una cantidad importante de monografías sobre temas puntuales de historia económica. Preocupados por la cuestión social, son casi sin excepción partidarios de la intervención del estado, antiliberales y tienen un cierto tinte hegeliano y nacionalista.*

*En 1883 Menger publica un trabajo sobre el método en la economía, continuando la línea de pensamiento de otro trabajo de 1871, que provoca una violenta reacción de Schmoller, quien le responde airadamente. Es un enfrentamiento entre los que confían en la deducción y quienes prefieren el método inductivo, proponiendo los estudios históricos y estadísticos. Luego de arduas discusiones esta cuestión cae en alguna manera casi en el ridículo, como puede verse en las afirmaciones de Pareto (1945) y en Marshall (1948). La lectura atenta de los escritos de unos y otros nos muestran que ni Menger negaba la importancia de los estudios empíricos, ni los miembros de la escuela histórica consideraban inapropiada la teoría, como bien señala Bucher: "Porque en realidad no hay más método de investigación, para penetrar la complejidad causal de los fenómenos de cambio que la abstracción que aísla y la deducción lógica."*

*Se trató más bien de una discusión motivada en parte por razones políticas y no exclusivamente científicas, que dejó dos consecuencias: por*

*un lado se revalorizaron los estudios empíricos, y si bien prescribieron el método histórico principalmente, abrieron el camino para los estudios cuantitativos, y constituyendo un antecedente para la escuela institucionalista americana, ya que algunos de sus integrantes se formaron en universidades europeas donde estas ideas prevalecían.*

### c) El apriorismo de la escuela austríaca y la obra de Robbins

*El debate que hemos referido antes dió origen al nacimiento de lo que se llamó escuela histórica por una parte, y de la llamada escuela austríaca por el otro, representada principalmente por Menger, de Böhm-Bawerk y de vön Wieser inicialmente. En una segunda etapa esta línea de pensamiento es continuada por vön Mises, que se forma con los anteriores y a su vez tiene como alumnos, entre otros, a Hayek, Haberler, Machlup, Morgestern, Rosenstein-Rodan y Robbins. Vön Mises es en cuanto al pensamiento epistemológico y metodológico, que está expresado en su obra "La Acción Humana", el exponente más conspicuo de este grupo. Su mensaje es que a diferencia de las ciencias físicas, las leyes económicas son descubiertas por deducción lógica a partir de axiomas que son dados "a priori", tal como es la persecución de objetivos por la acción humana. La historia es la resultante de muchos factores y no puede usarse para verificar las teorías, sino más bien, al revés, la teoría de los precios debe ser utilizada para explicar la historia.<sup>24</sup> El mérito de este enfoque es analizar la conducta humana en una dimensión más amplia, siguiendo, aunque desde una perspectiva diferente el camino que había señalado Bentham (1789) en su obra publicada en 1789 y que fue uno de los pilares sobre los que se construyó el modelo clásico<sup>25</sup>. Este camino vuelve a ser retomado recientemente con los trabajos de Aislei (1993), como veremos más adelante.*

*Dada la confianza en la existencia de un orden natural que hace que existan las instituciones sociales, como el lenguaje y la moneda sin que nadie lo haya programado, basan el estudio económico en la conducta individual y consideran que la comprobación empírica tiene un rol muy secundario si es que le cabe alguno. Tampoco las matemáticas son utilizadas en sus formulaciones, que hacen en forma verbal, aún cuando tratan temas como el análisis marginal y la teoría subjetiva del valor, donde realizan destacados aportes.*

*En 1932 Robbins (1951) publica su obra "Naturaleza y Significación de la Ciencia Económica", en la que sigue la línea de pensamiento de la escuela austríaca. Dado el destacado lugar de su autor<sup>26</sup> en la economía inglesa entre 1930 y 1960, su influencia ha sido muy grande. Su idea respecto a la metodología de la economía, influida por el pensamiento de Hayek y Mises principalmente, no es una negación de la comprobación empírica de las teorías, pero mantiene una posición escéptica respecto a los estudios estadísticos: "No hay ningún fundamento para suponer que el estudio de una muestra escogida al azar dé origen a generalizaciones de alguna importancia. Las ciencias no proceden así. Eso, o algo muy parecido, es, empero, el supuesto que yace en la esperanza de que las categorías*

<sup>24</sup> ver Rothbard (1991).

<sup>25</sup> Una explicación del rol de Bentham en el pensamiento clásico se puede encontrar en la Autobiografía de Stuart Mill (1948).

<sup>26</sup> Tuvo gran influencia en la vida pública inglesa. Enseñó en la LSE durante casi toda su larga vida (nació en 1894 y vivió hasta 1983) y además participó en innumerables asuntos no académicos.

formales del análisis económico puedan recibir contenido sustancial del valor constante y permanente”.

A continuación se refiere al ejemplo, ya famoso, sobre la elasticidad de la demanda de arenques. Supongamos, dice, que ha sido fijado su precio en un nivel inferior al de mercado y que la elasticidad de la demanda se estima en 1.3. Podríamos entonces saber cuál es el desplazamiento de la curva que representa a la demanda. Sin embargo, afirma, este valor de elasticidad sólo es válido en un contexto histórico determinado, que no tiene porqué repetirse dada la enorme cantidad de circunstancias que influyen para modificarla, por lo que este tipo de estudio es de una utilidad muy limitada, idea que tuvo un efecto disuasivo importante sobre la investigación empírica. Cabe hacer algunas reflexiones: en primer lugar si se lee con atención el capítulo V de su obra se ve claramente que se está refiriendo a la medición sin teoría de la escuela histórica y de los institucionalistas americanos, ya que la econometría, como la entendemos actualmente, aún no se había desarrollado, y por otra parte el desarrollo de las técnicas estadísticas de su época era todavía bastante rudimentario. Sin embargo, muchos años después, cuando cuenta ya con 81 años reflexiona de una forma más contemporizadora: "Generalizations about physical or economic phenomena are essentially conjectures, capable of deductive development, but, while exposed to refutation by appropriate testing, incapable of final and definite proof. In this respect I am a convinced Popperean...Therefore personally I regard the broad conception of the Logic of Scientific Discovery as generally applicable, whatever particular scientific discipline we practice."<sup>27</sup>

Es decir que las posiciones más extremas van acortando sus distancias a medida que las nuevas técnicas se van generalizando y aceptando y la epistemología evoluciona hacia formas más compatibles con las ciencias sociales.

#### d) La reacción de Hutchison y la introducción de las ideas de Popper en Economía

La obra de Popper "La Lógica de la Investigación Científica" apareció en 1934<sup>28</sup> y en 1938 Hutchinson publica su libro "The Significance and Basic of Postulates of Economic Theory", donde rescata las ideas de aquel y reacciona contra el apriorismo de Vön Mises y de Robbins. Sostiene que existen sólo dos tipos de proposiciones: las tautológicas y las empíricas, y la ciencia económica sólo debe tener presentes las segundas.

Este enfoque es realmente novedoso, y si bien existe el antecedente de la escuela histórica, es a mi juicio, diferente, porque tiene en mente más la comprobación estadística que la histórica, como ocurría con aquellos. De esta manera da paso a una forma diferente de ver las cosas, abriendo el camino a los institucionalistas y a las aplicaciones estadísticas primero y a las econométricas un poco después.

<sup>27</sup> Ver Robbins (1979).

<sup>28</sup> La primera edición inglesa apareció en 1958 y en su prefacio, afirma, "el problema central de la epistemología ha sido siempre, y sigue siéndolo, el aumento del conocimiento. Y el mejor modo de estudiar el aumento del conocimiento es estudiar el del conocimiento científico". Es decir que sostiene una posición diferente del instrumentalismo. (Ver Popper, 1980).

*Si bien la posición de este autor ha sido siempre agresiva y contundente, como pasa frecuentemente, con el transcurso del tiempo evoluciona hacia una posición más flexible, como vimos antes que sucede con Robbins en sentido contrario. Debemos tener presente por una parte el contexto histórico existente en 1938 cuando aparece su primer libro, ya que está por comenzar la guerra, y por otra la edad del autor, que apenas llegaba a los 26 años, lo que explica la energía con que defiende su posición.*

*En su primer enfoque critica a la metodología del enfoque neoclásico y a la escuela austríaca, debido al carácter vacío de la teoría pura, y al método -a su juicio ilegítimo- de la introspección. El test empírico de las teorías es imprescindible, si queremos que nuestra disciplina no quede en el descampado de la pseudociencia, aún en aquellos supuestos básicos como el principio de racionalidad de los agentes económicos, que actúan maximizando, ya sea su utilidad en caso de los consumidores, o sus beneficios en caso de las empresas, sujeto a la restricciones impuestas por el presupuesto y la función de producción respectivamente. Esto debe ser "testado", según Hutchinson y es muy interesante su discusión con Machlup<sup>29</sup>, quien sostiene con razón, en base a la filosofía de la ciencia de su época, que no es posible testear los supuestos básicos, sino que lo que debe ser sujeto a comprobación<sup>30</sup> son las teorías como un todo. Es que ya los positivistas del Círculo de Viena habían elaborado el método hipotético-deductivo, cuya idea central es la imposibilidad de dar sentido empírico a conceptos tales como el átomo o las relaciones de la química del carbono.*

*Posteriormente rechaza el falsificacionismo "naive" de los primeros tiempos y en uno de sus escritos más recientes sostiene que una teoría falsificada no necesita ser abandonada, con lo que deja atrás las posiciones extremas y quita a su enfoque inicial el carácter de destructivo que se le imputa, aceptando las ideas de Lakatos, aunque se lamenta de que este autor no haya definido, debido a su temprana desaparición, la ubicación de la economía dentro de su teoría, elaborada teniendo presente a la física más que a las ciencias sociales. En su trabajo más reciente, mantiene sus ideas anteriores sobre el principio de racionalidad, al que considera un supuesto irreal: "...if the economic world which economists are seeking to explain is one free of uncertainty and ignorance, and made up of omniscient, or approximately omniscient decision-makers, it is a world of overwhelmingly different from, and more simplified than, the actual world in which we live, in very important sectors of which profound ignorance and uncertainty plays such vital rol...In other words, as universal generalization, the vital, full-knowledge component of the rationality postulate is a science-fictional, Utopian fantasy"<sup>31</sup>*

*Considera que un mundo de agentes plenamente racionales e informados sería un mundo de autómatas, distinto del mundo en el que vivimos. Robbins, afirma, nos dice que debemos preguntar por el principio de racionalidad a los filósofos, mientras que de acuerdo a su criterio los economistas son los que deben dar respuesta a esta cuestión<sup>32</sup>.*

<sup>29</sup> Una excelente descripción de este debate se puede ver en Caldwell (1982).

<sup>30</sup> Machlup usa el término "verificación", cuando debiera utilizar "comprobación".

<sup>31</sup> ver Hutchison (1997), pp. 135-136. Es conveniente destacar que a partir de 1997 comienza a aparecer el Journal of Economic Methodology.

<sup>32</sup> Sin embargo flexibiliza su posición cuando diferencia, según dice siguiendo a Marshall, entre *teoría pura*, que sólo tiene que ver con relaciones lógicas o matemáticas, que no pueden ser testeadas empíricamente y *teoría aplicada*, que sí debe ser sometida a tests empíricos.



*¿Cuál es el mérito de Hutchinson? Haber llamado tempranamente la atención sobre la posibilidad de aplicar las ideas de Popper a la economía y sobre todo haber insistido en la necesidad de un mayor uso de la comprobación empírica, cambiando el enfoque de los economistas neoclásicos, pero sobre todo el de Robbins. Es cierto que tal vez se exceda al tomar una posición demasiado rígida, que Machlup califica como "ultraempirista", porque someter a comprobación empírica el principio de racionalidad, por ejemplo, es una tarea plagada de dificultades.*

*Pero éste no es el único embate contra Robbins<sup>33</sup>. En la década de los años sesenta, más precisamente entre 1958 y 1965, otro grupo de disconformes, encabezados por R. Lipsey y C. Archibald, comenzó a sentirse mal con la metodología basada en Robbins y en Samuelson, que analizaban los signos de las derivadas y las condiciones de equilibrio, y hacían caso omiso de las magnitudes de los efectos, recurriendo a los hechos para ejemplificar y no para comprobar o falsificar.*

*Archibald y Lipsey (1958) publican su famoso artículo criticando las conclusiones de Patinkin y quedan disconformes con los resultados. Piensan que tiene que haber algo más que el descubrimiento de la verdad a través de manipulaciones lógicas obtenidas de unos pocos postulados evidentes por sí mismos, sobre todo si estamos pensando en aplicar la economía a la realidad para transformar la sociedad<sup>34</sup>. Las teorías de Popper, que era profesor en la London School of Economics (LSE) en ese tiempo, fueron aceptadas con entusiasmo por estos economistas y un grupo destacado de discípulos, deseosos de una economía más inclinada a testear las teorías y confrontarlas con la realidad empírica.*

*La publicación del libro de Lipsey (1971), "Introducción a la Economía Positiva", del que se vendieron un millón y medio de ejemplares, produjo un cambio en la forma de ver las cosas de la profesión en general, que en gran medida adoptó una posición impregnada por el pensamiento popperiano. La primera página del libro transcribe una cita de Lord Beveridge, que es un llamado hacia una economía con base empírica<sup>35</sup>. Sin embargo pronto estos economistas cambiaron de opinión, ya que vieron las dificultades de la aplicación de las ideas de Popper, para someter a falsificación teorías como el principio de maximización de utilidades por las firmas o las relacionadas con la competencia imperfecta, la inestabilidad de los parámetros y sobre todo el carácter probabilístico de las teorías económicas. Todo esto hizo que Lipsey, en 1965, renegara parcialmente de Popper. De Marchi (1985), sostiene: "Lipsey, too, fairly quickly concluded that falsifiability is a nice ideal but difficult to translate into practice...He concluded that stochastic propositions are not strictly refutable."*

*Este grupo de economistas había sido poco adicto a la econometría y a la estadística, pero en 1964 la presencia de Sargan, que se había formado en la tradición de la Cowles Commission, produjo una revaluación de estas disciplinas, aunque las ideas de Popper fueron poco a poco suplantadas por las de Lakatos, que también enseñó allí por esos años y*

<sup>33</sup> Un prolijo relato de este episodio se puede ver en De Marchi (1985).

<sup>34</sup> Este es precisamente el punto de discusión con Robbins, que no pensaba en ningún tipo de intervención estatal y por lo tanto estaba al margen de la política económica.

<sup>35</sup> Se trata del extracto de un discurso pronunciado por Lord Beveridge en 1937, donde aboga por una posición similar a la de Tycho Brahe y hace una acalorada defensa de los métodos empíricos.

que parecieran adaptarse mejor a las ciencias sociales. Es de notar que hacia ese tiempo Popper excluye a la economía y al principio de racionalidad de ser testeado, dado la inestabilidad de los coeficientes, y se basa para su cambio de opinión precisamente ;en Robbins!<sup>36</sup>. La forma de resolver el problema que utiliza es mediante lo que llama "análisis situacional", tal como lo define en Popper (1965)<sup>37</sup>. La función de las ciencias sociales no es hacer profecías como los historicistas, sino analizar las repercusiones sociales no intencionales de las acciones humanas, actuando los agentes adecuadamente, de acuerdo a la situación<sup>38</sup>. De esta manera cuando testeamos una teoría no estamos testeando el principio de racionalidad sino el modelo que toma como presupuesto.

Pero esta posición y el criterio de falsificación son en alguna manera incompatibles: si el principio de racionalidad es testeable, se puede demostrar que es falso, en cuyo caso el análisis de situación sería no-científico de acuerdo al criterio de falsificación, mientras que mantenerlo implicaría aceptar dos metodologías diferentes, una para las ciencias naturales y otra para las ciencias sociales<sup>39</sup>.

¿Qué queda en pié del pensamiento de Popper en el "mainstream" de la economía? Si bien la aplicación cruda del popperianismo no tiene hoy en día mayor aceptación, queda algo así como una gran simpatía hacia su pensamiento y son muchos los que lamentamos las dificultades para la aplicación de sus prescripciones metodológicas. Así, Hausman (1988), un enconado crítico de Popper y del principio de falsificación, dice: "The most prominent methodologists who have defended part of the Popperian gospel - and these are mayor figures such as Klant, Blaug and Hutchinson - have be unwilling to draw such drastics conclusion...they have simply argued for the importance of critism and empirical testing...in their emphasis on the importance of testing in science, they may have done some significant good... Empirical criticism is important to science and scientific theories must be open to criticism that uses the results of observations and experiments..."

Por otra parte, otro de los críticos de la aplicación de las ideas de Popper a la economía, como es Caldwell (1982), señala: "Methodological falsificasionism in social sciences is an inspiring and noble critical ideal..."

Caldwell (1991) se pregunta acerca del porqué de la gran aceptación de las ideas de Popper entre los economistas, y luego de descartar el contacto personal que significó su presencia en la LSE destaca el hecho de que Popper va al corazón de los problemas que preocupan a los científicos (y a los economistas dentro de ellos), planteando cuestiones como el criterio de demarcación, cómo hacer para descartar teorías que no son científicas, cómo mejorar nuestra objetividad. Si bien dió

<sup>36</sup> Siempre tuvo una excelente relación con los miembros de la escuela austríaca. En su ensayo autobiográfico "Búsqueda sin Término" dice textualmente que Hayek "salvó su vida" en dos ocasiones: la primera cuando intercedió para que fuera publicado su libro "La Sociedad Abierta y sus enemigos", y la segunda cuando le ofreció un cargo en la Universidad de Londres que le permitió volver a Inglaterra. Ver Popper (1974). Por otra parte Hayek adopta gran parte de las ideas de Popper, según Caldwell (1991) después de 1950, aunque no acepta la aplicación del principio de falsificación sobre los principios básicos.

<sup>37</sup> Esta no es la primera incursión de Popper en las ciencias sociales. En 1943, finaliza la redacción de "La Sociedad Abierta y sus Enemigos" (1963) y en ese mismo año envía para su publicación a *Mind* el manuscrito de "La Miseria del historicismo", que dicho sea de paso, fué rechazado.

<sup>38</sup> Ver Padrón (1982)

<sup>39</sup> Los conocimientos de economía de Popper no parecen haber sido muy extensos y posiblemente sus supuestas contradicciones se deban a esa circunstancia.

*respuestas que a veces no se consideran satisfactorias, es evidente que planteó las preguntas adecuadamente y tiene la virtud de darle un marco normativo a quien se dedica a las tareas científicas, una manera de ubicarse ante la realidad que trata de interpretar.*

#### e) Las ideas de Lakatos

*Como vimos el falsificacionismo en su versión ingenua es de difícil aplicación en economía. Existen ciertos principios básicos, ciertos presupuestos de la conducta humana que no pueden ser "testeados", como sucede con el supuesto de conducta maximizadora por parte de consumidores y empresarios, de utilidad y beneficios respectivamente, que difícilmente pueden ser sometidos a comprobación empírica<sup>40</sup>. Por eso el sistema de Lakatos es más posible de conciliar con lo que ha realizado la ciencia económica en sus doscientos años de vida, ya que este autor propone la existencia de diferentes "programas de investigación", que son cúmulos de teorías que evolucionan en el tiempo y tiene un "hard core" con los fundamentos básicos, que son no testeables, y un "protective belt" de hipótesis auxiliares que pueden y deben ser sometidas a tests de comprobación o falsificación. En este caso, por ejemplo, el supuesto de racionalidad pertenecería al "core" del programa de investigación conocido como neoclásico, y no sería necesario someterlo a test de falsificación sin que por eso se cuestione su naturaleza científica. De esta manera el rol de la confirmación o falsificación queda totalmente salvado, aunque sólo permitiría confirmar la aplicación de la teoría a determinado punto de la historia o la geografía, o a determinado sector o unidad económica<sup>41</sup>.*

#### f) Dos visiones americanas: el instrumentalismo de Friedman y el operacionalismo de Samuelson

*El trabajo de Friedman (1953) sobre metodología es uno de los más leídos y aceptados. Tiene la simplicidad y la fuerza común a todos sus trabajos: las teorías, afirma, valen por su poder predictivo, no por la explicación que puedan brindar acerca de los fenómenos observados. Las teorías generan hipótesis capaces de ser "testeadas". El único "test" relevante es la comparación de las predicciones con la experiencia. Si hay varias teorías alternativas que predicen adecuadamente, elegiremos aquella que sea más sencilla. En sus propias palabras:*

*"Viewed as a body of substantive hypothesis, theory is to be judged by its predictive power for the class of phenomena which it is intended to "explain". Only factual evidence can show whether it is right or wrong, or better tentatively accepted as valid or rejected...the only relevant test of the validity of a hypothesis is comparison of its predictions with experience."<sup>42</sup>*

<sup>40</sup> Ver Caldwell (1982), Cap VII, donde se puede encontrar una reseña sobre diversos trabajos realizados para testear el postulado de racionalidad, que ofrecen resultados bastante confusos y no concluyentes.

<sup>41</sup> El enfoque de Lakatos es de tipo descriptivo-prescriptivo, a diferencia del de Kuhn (1962), quien describe el desarrollo de la ciencia como algo discontinuo. Existen períodos de "ciencia normal" seguidos de períodos revolucionarios donde el "paradigma" es cambiado y se adopta uno diferente. Este enfoque tuvo también relativo éxito entre los economistas.

<sup>42</sup> Ver Friedman (1952).

*Inmediatamente define el concepto de "predicción" en los siguientes términos: "To avoid confusion... predictions by which the validity of a hypothesis is tested need not be about phenomena that have occurred..."*

*Es decir que una teoría puede predecir hoy qué hubiera sucedido, dadas ciertas circunstancias, por ejemplo, en 1906 y ello sería una forma de testear la teoría en cuestión. Es bastante amplio en la forma de generar nuevas hipótesis, y en esto se acerca a Popper.<sup>43</sup> La pregunta que surge es si es necesario para pronosticar una teoría subyacente, o si es suficiente con procedimientos de tipo exclusivamente estadísticos, como ocurre con ciertas ramas de la econometría contemporánea que sólo buscan pronosticar, como sucede con el estudio del comportamiento de los mercados financieros<sup>44</sup>.*

*Si nos ubicamos en los debates de su tiempo (hace casi medio siglo) nos encontramos con una economía enfrentada a pocos datos, imposibilidad de hacer experimentos controlados<sup>45</sup> y muy poca oportunidad de hacer experimentos "cruciales", lo que condujo o bien al análisis exclusivamente formal o a un empirismo sin base teórica, es decir al "measurement without theory", que consideraba altamente peligroso. Su metodología obviaba este problema, así como la necesidad de testear los presupuestos básicos. La competencia perfecta o el principio de maximización podían no ser realistas, pero conducirnos a predicciones acertadas, aún cuando, por ejemplo, el principio de la competencia imperfecta fuera más concordante con la realidad.*

*Pero si bien los economistas en su mayoría aceptaron sus puntos de vista no ocurrió lo mismo con los filósofos de la ciencia: sólo los positivistas más recalcitrantes del siglo XIX sostenían que la ciencia no tenía una finalidad explicativa y que su objeto se reducía a hacer correlaciones y hacer predicciones<sup>46</sup>.*

*Nos encontramos con una situación un tanto paradójica: si bien en cierta forma resuelve algunos problemas pendientes en economía, como la "testeabilidad" de los supuestos básicos, la unidad metodológica de las ciencias naturales y sociales y la relación entre la teoría y la comprobación empírica, el supuesto instrumentalista de su metodología sufre demolidores ataques.<sup>47</sup>*

*Lo positivo es que se dió un paso más hacia la idea de que las teorías económicas deben confrontarse con la realidad empírica, y el uso de la estadística y la econometría tuvieron un impulso importante, tanto en sus trabajos como en las de sus discípulos.<sup>48</sup>*

---

<sup>43</sup> Es interesante destacar que en todo el trabajo (Friedman, 1952) no existe ninguna cita de ningún epistemólogo. Su trabajo pareciera la labor aislada de un economista que desea encontrar una metodología apropiada para su disciplina, pero autónomo en su pensamiento.

<sup>44</sup> Ver Campbell, Lo y Kinlay (1997), que analizan una serie de técnicas para el estudio sistemáticos de los mercados financieros. Sin embargo cabe tener presente que modelos que aparentemente son sólo estadísticos, como ocurre con un modelo tan simple como el camino aleatorio, que tiene como fundamento a la teoría de la eficiencia débil de Fama (1970).

<sup>45</sup> Es necesario destacar que Friedman (1952) considera que la diferencia entre un experimento controlado y uno que no lo es, es simplemente una cuestión de grado. Nunca existe un control total de un experimento, aún en la física, como sucede con la caída de los cuerpos en el vacío, el famoso experimento de Galileo.

<sup>46</sup> ver Hempel (1965).

<sup>47</sup> ver Friedman (1992).

<sup>48</sup> Ver sus dos monumentales obras: Friedman (1971) y Friedman (1982), donde se refiere permanentemente a datos estadísticos, aunque con una elaboración econométrica sencilla.

Desde otra perspectiva Samuelson introduce su propio sistema, llamado *operacionalismo*<sup>49</sup>, basado en las ideas de Brigman, quien sostiene que la definición de cualquier concepto en la ciencia no es más que un conjunto de mediciones y solamente atribuye el status de conocimiento a las proposiciones acerca de fenómenos observables. Constituye, conjuntamente con los pragmatistas, la versión americana del Círculo de Viena. Pero los componentes de este grupo, como ocurre con Hempel (1965), aceptan luego la teoría y se diferencian de estas posiciones extremas, que tienen un gran parecido con el positivismo crudo de fines del siglo XIX.

La posición de Samuelson tiene algunas características distintivas: por una parte define su sistema como "operacionalismo", término cuyo sentido no aclara totalmente. En segundo lugar incorpora la idea del test cualitativo, es decir que sostiene que deben considerarse la corrección de los signos de las derivadas en vez de cuantificar los parámetros de los modelos, y además introduce la idea de que los sistemas deben ser estables, por lo que deben satisfacer el principio de correspondencia.

La influencia de este enfoque fué duramente criticada y casi totalmente abandonada, pero dejó sus residuos positivos: la incorporación masiva de las matemáticas en la economía, que es la antesala de la comprobación empírica de los modelos, tuvo un fuerte impulso.

#### g) McCloskey

Este autor representa la heterodoxia, que se rebela contra el orden establecido, laboriosamente, en el "mainstream" de nuestra disciplina. Sostiene<sup>50</sup> que la metodología aceptada por los economistas no es la que se practica en la realidad, y si se aplicara se paralizaría la ciencia, como hubiera paralizado a Galileo si la hubiera aplicado estrictamente. Define el modernismo como la epistemología aceptada, caracterizado por la predicción, el falsificacionismo, la objetividad en lugar de lo subjetivo, el rechazo a la introspección, con criterios muy rígidos de demarcación y desprovisto totalmente de valores y criterios normativos. Considera que el falsificacionismo y la predicción no son posibles en nuestra disciplina y que están en retirada en la epistemología en general: "Modernism promises knowledge free from doubt, metaphysics, morals and personals convictions; what it delivers merely renames as Scientific Method the scientist's and specially the economic scientist's metaphysics, morals and personals convictions...Scientific knowledge is not different from other personal knowledge."

La secuencia Carnap-Popper-Lakatos-Kuhn-Feyerabend representa en la historia de la filosofía de la ciencia un descenso "de los picos frígidos del absolutismo científico a los valles suaves de la anarquía retórica". Esta, de acuerdo a su peculiar definición, "es el arte de probar lo que el hombre cree que se debe creer y no probar la verdad de acuerdo a un método abstracto, y procura "reemplazar la certeza por la sabiduría", volviendo a Aristóteles Cicerón. Considera a la econometría como una manifestación del modernismo y critica su uso con vehemencia.

---

<sup>49</sup> Ver Samuelson (1966).

<sup>50</sup> Ver McCloskey (1983).

*Este enfoque, si tiene algún mérito es cuestionar la ortodoxia, lo que siempre es conveniente, pero como señala Sims (1996), su acción ha sido perjudicial porque incita al charlatanismo. Como sucede generalmente, es mucho más fácil criticar una teoría o un método que proponer algo superior para que sustituya lo existente.*

## *La econometría como herramienta metodológica de la economía*

### *a) Los comienzos*

*La utilización de la estadística y la econometría para el análisis de los fenómenos observables estuvieron subordinados a las concepciones metodológicas de los economistas. Como hemos visto estas ideas han ido cambiando en el tiempo y por otra parte han aparecido nuevos métodos estadísticos, al par que las ciencias sociales, especialmente la economía, fueron reclamando desarrollos de la estadística acordes con sus propias necesidades.*

*El uso de los métodos estadísticos en economía, que constituye el antecedente necesario para los estudios empíricos, se generaliza después del comienzo del siglo pasado<sup>51</sup>. Hasta esa fecha su uso en economía es escaso, ya que las investigaciones empíricas consistían más bien en estudios de historia económica, como pregonaba la escuela histórica alemana. Edgeworth y Marshall, si bien proponen el uso de la deducción y de la inducción como herramientas, hacen poco uso de la estadística, a pesar de que el primero hace un uso más frecuente del instrumental matemático y del cálculo de probabilidades. Algo similar ocurre en el continente con Walras y Cournot, quienes tenían poco tiempo para las estadísticas, como afirma Morgan (1992). Uno de los primeros en utilizar datos y procedimientos estadísticos es Moore, quien es precisamente un disidente de la escuela neoclásica. Es preciso tener presente que quienes comienzan a forjar los métodos estadísticos que luego conformarían la econometría, lejos de renegar de los modelos teóricos y de los modelos matemáticos, pretenden aplicarlos y confirmarlos mediante el análisis de los datos.*

*El período que transcurre entre principios de siglo y la aparición del trabajo de Haavelmo (1944), muestra que los estudios de este tipo se refieren a dos temas muy diferentes entre sí: uno es el análisis de los ciclos económicos y el otro es el estudio de mercados de bienes, en los cuales se procura calcular las elasticidades-precio de la demanda y de la oferta.*

*Ambos campos son un antecedente de investigación empírica con uso de métodos estadísticos sofisticados que no se encuentran en los estudios económicos del siglo pasado, como hemos explicado anteriormente al referirnos a la escuela neoclásica<sup>52</sup>.*

*Respecto al estudio de los ciclos se pueden encontrar dos maneras diferentes de abordar la cuestión. Por una parte estudios de tipo meramente estadísticos, y por otra aquellos que definen un modelo teórico primero y analizan su correspondencia con los datos*

<sup>51</sup> Especialmente con los trabajos de Gauss (1777-1855), quien aplicó sus desarrollos originalmente a la astronomía.

<sup>52</sup> Yule, en 1897 utilizó por primera vez en economía el método de regresión, que había sido empleado exitosamente por los astrónomos desde mediados del siglo pasado para estudiar las trayectorias de los planetas, para las cuales realizaban mediciones que diferían entre sí debido a errores en ellas. Luego fué usado en biometría, disciplina que conjuntamente con la psicometría aparece hacia fines del siglo pasado.

posteriormente, forma ésta que es la aplicada por los primeros econométricos.

## b) El debate en torno al método de Tinbergen

*El primero en construir modelos macroeconómicos de ecuaciones simultáneas fue Jan Tinbergen, quien en 1936 publicó un trabajo que contiene un modelo para la economía de Holanda. Ese mismo año la Liga de las Naciones le encomendó la tarea de analizar estadísticamente la teoría acerca del ciclo económico de Haberler (1953), expuesta en "Prosperidad y Depresión", para lo cual publicó dos informes: uno refiriéndose a los métodos econométricos y otro presentando un modelo para la economía de Estados Unidos para el período 1919-1932, con 71 variables y 48 ecuaciones. En este enfoque se aleja de la formulación meramente estadística de los institucionalistas y presenta algunas ideas metodológicas novedosas: las teorías no pueden ser confirmadas por los estudios empíricos, sino solamente pueden ser declaradas incorrectas o incompletas, pero aún si una teoría concuerda con los hechos es posible que otra teoría sea la "verdadera". El rol del economista que elabora el modelo es descartar lo que no es esencial, para lograr un modelo que sea manejable desde el punto de vista matemático. Las relaciones de causa y efecto, igual que la estructura de rezagos deben surgir de la teoría, y es consciente de que estas relaciones son sumamente complicadas. Si se analiza con atención su trabajo surge claramente que los procedimientos utilizados para la selección de las variables se realizan teniendo en cuenta los resultados estadísticos muy frecuentemente, cosa que no está explícitamente reconocida.*

*Este trabajo despertó un torbellino de discusiones. Su polémica con J.M. Keynes (1939) es una de las más conocidas. La crítica que éste le formula tiene dos soportes: el primero, metodológico, que consiste en afirmar que las teorías no pueden ser modificadas por la investigación empírica. Si el modelo teórico es correcto nada puede añadir la econometría y si los resultados no son concordantes con la teoría debemos culpar al método utilizado. En segundo término realiza una detallada crítica de la econometría como herramienta en función de las violaciones a los supuestos básicos del modelo, que quedan más bien en el campo de la estadística<sup>53</sup>.*

*Tinbergen replica por su parte que si una teoría no es confirmada por los resultados es equivocada o insuficiente, por lo que puede ser preciso incorporar otras variables o formas funcionales que antes no fueron consideradas importantes por la teoría. La evaluación de las teorías rivales puede hacerse a la luz de la econometría. Marshak y Lange (1940) defienden a Tinbergen en este punto y presentan el caso de la relación entre stock de capital y nivel de producto. Según Hayek, nos dicen, suben en la prosperidad y se reducen en la depresión, exactamente lo contrario de lo que sostiene Kaldor. Sugieren que no queda otro camino que, de disponerse de los datos necesarios, dilucidar la cuestión mediante el análisis estadístico.*

---

<sup>53</sup> Hace una descripción detallada de todos los problemas de las estimaciones econométricas: sesgo por variables omitidas, variables cualitativas, colinealidad, correlación espúrea, independencia, causalidad y exogeneidad, linealidad, selección de rezagos, uso del término "trend", estabilidad de los parámetros, etc., todo dentro de su prosa amena y aguda. Algunos de estos problemas han sido tratados por el desarrollo posterior de las técnicas estadísticas y otros subsisten aún como interrogantes.

Posteriormente este enfoque fue adoptado, modificado y elaborado por la Cowles Commission en los Estados Unidos<sup>54</sup>. Es que si bien la econometría tiene un contenido empírico al tratar de testear las teorías, al revés de lo que hacen los institucionalistas, define previamente un modelo teórico, es decir que parte de la teoría y luego recién analiza los datos.

### c) La revolución probabilística de Haavelmo

Haavelmo (1944), formula una idea revolucionaria en un trabajo que comienza a circular en forma mimeografiada en 1941, donde introduce el concepto de probabilidad dentro del modelo. Hasta ese momento se había usado la estadística, que tiene en la base de su formulación el principio estocástico, en alguna manera en forma inconsciente. Haavelmo, en cambio, lo hace en forma explícita y modifica substancialmente el enfoque tradicional. Los modelos económicos requieren de una estadística que contemple la naturaleza especial de los fenómenos sociales, ya que no se pueden realizar experimentos controlados, por lo que debemos aprender a interpretar a los procesos que la Naturaleza nos presenta. Estas ideas alimentan la labor de la Cowles Commission, al introducir el concepto de probabilidad en los modelos econométricos. En sus palabras:

*"The method of econometric research aims, essentially, at a conjunction of economic theory and actual measurements, using the theory and technique of statistical inference as a bridge pier. But the bridge itself was never completely built. So far, the common procedure has been, first to construct an economic theory involving exact functional relationships...Tools of statistical inference have been introduced...The application of such simple "statistics" have been considered legitimate, while at the same time, the adoption of definite probability models has been deemed a crime in economic research...For no tools developed in the theory of statistics has any meaning without being referred to some stochastic scheme".*

Como puede verse el cambio que propone Haavelmo es fundamental: está incorporando los términos estocásticos al modelo en sí, no como un error de medición, sino como un componente más, al lado de las otras variables explicativas, lo que permite que las hipótesis puedan ser sometidas a tests estadísticos. Esta forma de ver la econometría es aceptada rápidamente y marca un hito en la metodología de la econometría y de las aplicaciones estadísticas en economía.

---

<sup>54</sup> Esta institución fué fundada en 1932 en Colorado Springs por Alfred Cowles, un corredor de bolsa que quería estudiar las leyes que rigen el funcionamiento de los mercados de valores. En 1937 se mudó a Yale University por razones fiscales, y en 1939 a Chicago, donde funcionó los 15 años siguientes, donde se realizó el grueso de su tarea científica, con el aporte de economistas americanos, como Klein y de muchos europeos, entre ellos Frisch, Koopmans, Haavelmo y Marshak, que huyeron de la invasión alemana. La labor de esta institución fué el resultado de una serie de condiciones favorables que se dan sólo muy pocas veces en la historia de la ciencia. Nos encontramos con el keynesianismo buscando una herramienta para la programación de la política anticíclica, con un grupo de matemáticos y economistas que se reúnen emigrados de los países escandinavos, de Alemania y de Austria y una institución que los agrupa y cuenta con los recursos económicos necesarios para que puedan desarrollar su proyecto, pero sobre todo con el genio no frecuente de los brillantes científicos precitados. Una detallada reseña sobre el período en que estuvo radicada en Chicago, entre 1939 y 1955, puede verse en Christ (1994). La fundación de la *Econometric Society* en 1933 fué paralela al desenvolvimiento de la Cowles Commission y dió un importante impulso tanto al uso de la estadística como de las matemáticas.



#### d) El debate entre Koopmans y los seguidores de Mitchel

*Durante este período, en el que según Haavelmo todo comenzó<sup>55</sup>, entre las décadas de los veinte y los treinta, aparece el trabajo de Yule (1926) respecto a las correlaciones sin sentido, que constituyó un cambio en la forma de pensar respecto al uso de las correlaciones y regresiones en economía y en las ciencias sociales en general. Si las correlaciones no nos permiten comprobar una teoría y si de ellas no podemos extraer conclusiones respecto a la dirección de la causalidad, resulta muy peligroso "dejar a los datos que hablen por sí mismos", es decir utilizar las herramientas estadísticas sin que estén fundamentadas en la teoría económica, y como señala acertadamente Koopmans (1947) no es posible aplicarlas con la misma liberalidad que en las ciencias físicas.*

*En este contexto es que debe interpretarse la polémica desatada entre Koopmans (1995) y Vinning (1995), donde el primero sostiene que la teoría debe ser siempre el marco dentro del cual se formula investigación empírica<sup>56</sup>*

#### e) El método econométrico tradicional

*Llamamos así al procedimiento econométrico que usamos habitualmente. La teoría económica nos dice que un vector de variables dependientes  $Y$  es función de un vector de variables exógenas  $X$ . Suponemos que las demás variables que podrían afectar a las endógenas no son sistemáticas y que están incluidas en un término de error estocástico que tiene las propiedades del ruido blanco. En ese caso, suponiendo que la relación es lineal, los estimadores minimocuadráticos se supone que tienen lo que se conoce como las propiedades habituales. Realizamos entonces el proceso de estimación y luego comprobamos que los errores tengan las propiedades referidas y que no estén correlados con los regresores. Si éste es el caso podemos verificar la hipótesis bajo estudio, analizando los signos y la significatividad de los parámetros de las distintas variables incluidas. Si el término de error no tuviera esas propiedades los supuestos básicos habrían sido violados y aplicaríamos los correctivos conocidos para las diferentes patologías. Luego volveríamos a considerar el signo y significatividad de los parámetros para ver si la teoría ha sido comprobada.*

*Como vemos este es un procedimiento que difiere del sugerido por Popper en cuanto a perseguir la falsificación y no la verificación o comprobación. En la práctica encierra una peligrosa práctica como es el "data mining". Con este procedimiento se incorporan otras variables y se seleccionan aquellas que resultan estadísticamente significativas. Esto implica que las variables explicativas se seleccionan en función de lo que*

<sup>55</sup> Ver la *Nobel Lecture* de Haavelmo (1992), donde afirma que la econometría comenzó en los veinte, aunque niega la existencia de antecedentes anteriores.

<sup>56</sup> Es conveniente tener en cuenta el razonamiento de Mitchel: si tenemos no menos de veinte teorías explicativas del ciclo económico nos vemos en una situación confusa, ya que es imposible someter a comprobación a todas ellas, y ante esa situación es preciso "dejar hablar a los hechos". No obstante, Mitchel sostiene la necesidad, en principio, de una teoría previa. Esta última posición es abandonada por sus seguidores, especialmente Persons (1995), quien con la utilización de sus "barómetros", desarrolla una metodología totalmente desprovista de base teórica.

*dicen los datos, y estamos en presencia de "medición sin teoría". ¿Cuál es el remedio para este mal tan generalizado, en el que, a fuerza de ser sinceros todos quienes trabajamos en econometría aplicada debemos reconocer haber usado con más frecuencia que la debida? La solución es "testear" con el criterio de falsificar y no de verificar, modificar el modelo en base a consideraciones de teoría económica y volverlo a someter a test estadístico nuevamente. Este es el procedimiento que proponen Darnell y Evans (1990), tratando de rescatar la econometría tradicional.*

*Todos los pasos que se realicen deberían hacerse explícitos y el camino recorrido para llegar a una determinada estimación debiera expresarse claramente para que los resultados puedan interpretarse adecuadamente. Pero a poco que reflexionemos vemos que no estamos ante un problema fácil, y también que la aplicación del falsificacionismo no es posible en forma cruda e inmediata.*

*La naturaleza estadística de nuestra disciplina genera, además, otro tipo de dificultad, que se conoce con la denominación de "pre-test bias". Si estimo el valor de los parámetros de una ecuación determinada y rechazo la hipótesis nula de no autocorrelación, estoy en condiciones de continuar analizando, por ejemplo la significatividad de los parámetros, pero si he cometido un error de tipo II en la primera parte, estoy viciando la segunda.*

*No obstante todo lo anterior, la utilización de la econometría que podemos denominar convencional tiene un alto grado de aceptación y el enfoque a que nos hemos referido de Darnell y Evans (1990) es un intento de rescatar el método tradicional y deja abierta una interesante cuestión de tipo epistemológico.*

### **Problemas específicos de algunas metodologías econométricas actuales**

#### **a) Sims y los Vectores Autorregresivos (VAR)**

*Los fracasos de los modelos econométricos para predecir las fluctuaciones económicas de fines de la década de los años setenta y principios de los ochenta hizo que comenzaran a aparecer críticas a la econometría tradicional y a los modelos econométricos multiecuacionales elaborados sobre los lineamientos de la Cowles Commission. Una fue la famosa y conocida crítica de Lucas, y la otra fue la de Sims, que planteó el carácter "increíble" de las restricciones necesarias para la identificación de las ecuaciones que componen los modelos. Propone un sistema econométrico alternativo<sup>57</sup> en el que todas las variables son endógenas y se regresan todas ellas sobre sus propios rezagos y sobre los rezagos de todas las demás. Como elimina de cada ecuación las variables contemporáneas, los coeficientes carecen del sentido habitual, y no se puede hacer inferencias sobre su signo o significatividad, sino de todos ellos en block, por medio de un test de exclusión tipo Granger (1969). Sin embargo esto permite transformar esas ecuaciones, que son de tipo autorregresivas, en ecuaciones de tipo "moving average", expresando todo en función de los errores pasados, para lo que es necesario lograr una matriz de errores que sea ortogonal. Así es posible analizar el impacto que tiene un shock en cada una de ellas sobre todas las demás y se puede ver cómo es la*

<sup>57</sup> ver Sims (1980)

*relación de prelación temporal entre todas las variables<sup>58</sup> que componen el sistema, permitiendo determinar qué porcentaje de la explicación de la varianza de cada variable corresponde a cada una de las demás.*

*Este sistema recibió críticas<sup>59</sup> por considerárselo una nueva forma de plantear modelos estadísticos sin fundamento teórico, a mi juicio injustamente, ya que cuando se seleccionan las variables que componen el vector se está utilizando la teoría económica como fundamento de esa selección. Sin embargo es necesario tener presente que el concepto de exogeneidad de la Cowles Commission es distinto del de Granger y Sims: mientras el primero se basa en la teoría el segundo mira la muestra sujeta a análisis.*

*Pero es evidente que esta metodología, que es diferente de la convencional, introduce nuevas complicaciones y es y será en futuro próximo fuente de discusiones.*

*Existen dos complementos a la metodología VAR que deben ser tenidos en cuenta porque también tienen implicancias metodológicas: uno es la aplicación de métodos bayesianos que hace Litterman (1984) y otro es la complementación del modelo con vectores de variables cointegradas<sup>60</sup>. Tienen como característica metodológica una forma de ver la econometría como menos predeterminada por la teoría, dejando más margen para que los datos se expresen con mayor libertad que en los modelos convencionales, aunque disintamos, como dijimos más arriba, con quienes consideran a esta línea como ateórica, porque esto implicaría una completa independencia de los datos, que no es el caso.*

*Recientemente Sims (1996) presenta algunas reflexiones metodológicas alternativas. La ciencia, afirma, es una forma de comprimir los datos, de manera de poder interpretarlos mejor. Tycho Brahe acumuló una enorme cantidad de información, Kepler vio que las órbitas eran elípticas, comprimiendo los datos y Newton, al establecer su ley de la gravitación, llega a una compresión aún mayor. Esta visión supera a su criterio la metodología generalmente aceptada, que consiste en enunciar una teoría y confrontarla con los datos, y que había producido un abuso en el uso de la estadística. Coincide en algunas de las críticas que formula McCloskey (1983), pero se opone a la idea de que la economía debe procurar persuadir a los demás, sino que debe trabajar en base a normas muy estrictas:*

*"The only kind of argument that modern science treats as legitimate concerns the match of theory to data generated by experiment and observation...The fact that economics is not physics does not mean that we should not aim to apply the same fundamental standards for what constitutes legitimate argument; we can insist that the ultimate criterion for judging economic ideas is the degree to which they help us order and summarize data, that it is not legitimate to try to protect attractive theories from the data."*

<sup>58</sup> La generalización de esta metodología se debe en parte a la amplia difusión que tuvo el software denominado RATS (Regression Analysis Time Series), que es uno de los programas más usados por los econométricos de todo el mundo.

<sup>59</sup> Ver Cooley y Leroy (1985) y Darnell y Evans (1990), quienes cuestionan la utilización de análisis impulso-respuesta, en razón de que, según afirman no es posible individualizar la variable que sufre el shock. También se oponen al concepto de causalidad de Granger como sinónimo de exogeneidad.

<sup>60</sup> Una descripción del método puede verse en Johansen (1995) y es fácil darse cuenta de que se trata de un procedimiento que se basa en las propiedades estadísticas de las series más que en consideraciones teóricas, como sucede también en los modelos de cointegración más sencillos.

*Es decir que si bien tiene un enfoque que en alguna manera se diferencia del modernismo, como denomina McCloskey al criterio metodológico tradicional, mantiene la defensa de los criterios estrictos de confrontación de las teorías con los datos, que tiene un rol diferente, más relevante que en los criterios anteriores, al par que confirma que su postura no está reñida con la definición previa del modelo teórico.*

## b) Leamer y los métodos bayesianos

*El enfoque de Leamer<sup>61</sup> tiene un cierto componente bayesiano. Critica a la econometría tradicional, proponiendo un sinceramiento de la forma de trabajar. Si vamos a hacer "data mining" hagámoslo explícita y ordenadamente. Las variables explicativas de un modelo dependen de los "priors" del economista que formula el modelo. En un modelo que explique la inflación<sup>62</sup>, un keynesiano incluiría, entre otras variables, la tasa de desempleo, mientras que un monetarista excluiría esa variable e incluiría, por ejemplo la tasa de cambio en la cantidad de dinero. Lo que sugiere es que probemos qué sucede con cada variable en ausencia y en presencia de las excluidas por cada uno. Este tipo de análisis, al que denomina "Extreme Bound Analysis", permite comparar distintos modelos y ver cuán robustos son al cambio de regresores<sup>63</sup>. Es decir, que este método permite incluir las creencias del economista, lo que es necesario, sostiene, porque la naturaleza no experimental de la economía así lo requiere. Es ilusorio, afirma, sostener lo contrario, y la forma tradicional de modelar, si bien es pura e inmaculada desde el punto de vista del teórico, está llena de incumplimientos de los supuestos básicos por parte de los que hacemos econometría aplicada, por lo que procura un sinceramiento y la aceptación de las consecuencias que tiene sobre el modelo la aplicación del "data mining", técnica que considera nefasta y trata de evitar con la aplicación de su método.*

*La diferencia, dicho de otro modo, con el enfoque ortodoxo es que en éste existe una teoría, establecida sobre bases objetivas. En el método que propone Leamer existen tantas teorías como observadores, y los distintos criterios ("priors") de los distintos observadores deben ser tenidos en cuenta al realizarse la estimación del modelo.*

## c) El método de Hendry

*El método de Hendry es una forma de unir en el modelo econométrico la teoría económica con la evidencia que surge de la muestra: "Modeling denotes matching theory and data in a formal quantitative framework, so we will consider the roles of economic theory and empirical evidence and their links<sup>64</sup>."*

<sup>61</sup> Ver en Leamer (1994) los trabajos más importantes que ha publicado en los últimos años. El más conocido, y el más representativo es Leamer (1983).

<sup>62</sup> ver Leamer (1994), pag 155 y ss.

<sup>63</sup> Por un teorema que demuestra en uno de sus trabajos (ver Leamer, 1975), si el valor del estadístico "t" de un determinado coeficiente es mayor que el del coeficiente cuyo valor se restringe igual a cero, la exclusión de este último no producirá cambio en el signo del coeficiente de la primera variable explicativa.

<sup>64</sup> Ver Hendry (1995).

*Por vía de la teoría económica determinamos los regresores que forman el modelo en el equilibrio de largo plazo, y se deja hablar a los datos, pero sólo para establecer el camino hacia el equilibrio en la forma más libre posible, limitando al modelo solamente en cuanto a la linealidad y al número de rezagos. Luego mediante la técnica de reducción se van progresivamente eliminando los regresores superfluos hasta lograr un modelo parsimonioso, consistente con la teoría y admisible para los datos, cuyos regresores deben ser exógenos débilmente, sus parámetros constantes, debe ser capaz de superar a los modelos rivales y los términos de error deben ser ruido blanco. Para eso sugiere testear, testear y testear, sometiendo a los modelos estimados a una batería de tests para comprobar que las propiedades enumeradas están realmente presentes. Este enfoque implica adelantos en el análisis econométrico que tienen importantes efectos epistemológicos: por un lado la introducción de los mecanismos de corrección de errores, y por otro la demostración de que la aplicación de los métodos tradicionales de corrección de la autocorrelación de primer orden implican una restricción que no siempre está presente en las series económicas<sup>65</sup>.*

*Esta metodología trae aparejada una nueva forma de ver la relación entre la teoría y los datos: "Theory driven approaches, in which a model is derived from a priori theory and calibrated from data evidence lie at one extreme. These suffer from theory dependence in that their credibility depends on the credibility of the theory from which they arose. Data driven approaches, lie at the other extreme. They suffer from sample dependence in that accidental and transient data are embodied as tightly in the model as permanent aspects..."<sup>66</sup>*

*Este método, afirma Hendry, no sirve para comprobar teorías, ni brinda soporte para ellas y que una estructura de largo plazo sea consistente con una teoría no quiere decir que sea inconsistente con otra. No es descubrir la verdad el objetivo de los modelos, sino encontrar el mecanismo que subyace en la realidad y genera los datos que observamos: "...empirical evidence is the final arbitrator of the validity of theories. Objectivity of approach, potential falsifiability of claims and progresivity of knowledge and understanding are all key elements. Criticism is thereof the life blood of science..."<sup>67</sup>*

*Es decir que niega la posibilidad del rechazo de las teorías, las cuales más que ser refutadas o falsificadas, caen en desuso.*

## Reflexiones finales

*Para sintetizar la evolución del problema metodológico y epistemológico de la economía es conveniente tener presente la forma como Kuhn (1978) caracteriza las etapas por las que transita la evolución de las ciencias: se parte de una etapa que llama "preciencia" en la cual existe una especie de caos metodológico y poca comunicación entre los científicos. En algún momento se produce un descubrimiento o la enunciación de una teoría que da comienzo a la etapa 2, que llama "el logro". Cuando la idea es aceptada por la comunidad científica que comienza a formarse se cumple la etapa 3, "la conversión", para arribar a*

<sup>65</sup> ver Hendry (1995), pp. 266 y ss.

<sup>66</sup> ver Hendry (1995), pp. 6.

<sup>67</sup> ver Hendry (1995), pp. 9.

la etapa 4, que llama "la ciencia normal", caracterizada por la noción de "paradigma" y por la consolidación de la disciplina como tal. Si analizamos el relato que hemos hecho acerca de la evolución de los problemas epistemológicos de la economía vemos que este camino es similar al recorrido por nuestra ciencia. Como tiene solamente doscientos años de vida estas etapas se han dado mucho después que en las ciencias naturales, aunque antes que en las demás ciencias sociales. Recién a principios de siglo se consolida el paradigma clásico y es partir de entonces que la discusión metodológica se acalla y se hace mucho más moderada, entrando en la cuarta etapa de Kuhn, a mi juicio hace unos veinte años. Se llega entonces a un acuerdo generalizado sobre dos cuestiones: por un lado se conviene en que la economía presenta problemas epistemológicos y metodológicos diferentes de los de las ciencias naturales, lo que dificulta las aplicación de las prescripciones de Popper y por el otro se acepta la idea de que la teoría debe preceder a la investigación empírica, porque se entiende que la comunicación entre la teoría y los datos no es un camino de doble mano, con lo que se define lo que podemos llamar enfoque tradicional, y que cuenta con aceptación generalizada hasta el final de la década de los años setenta. Luego comienza una nueva reacción, con las metodologías de Sims, Hendry y Leamer. Las dos primeras se caracterizan por la preocupación por incorporar al modelo elementos que provienen de la muestra, el que ya no queda totalmente atado a lo prescrito por la teoría previamente, sino que tendrá en cuenta el proceso que genera los datos que se observan y sus desviaciones temporales, mientras que Leamer permite la incorporación de las ideas del investigador dentro de lo que dice la muestra, aplicando un método similar al bayesiano.

Koopmans (1947) en su famoso debate con los institucionalistas define dos formas de trabajo científico: por un lado la metodología de Thyco Brahe, que se basa en la observación detallada de la realidad y por otro la de Newton, que diseña un modelo de mecánica celeste que permite comprender el movimiento de los planetas y pronosticar su comportamiento. En lo que hemos visto la economía y la econometría han oscilado pendularmente entre ambas posiciones. Una etapa inicial, basada en un enfoque más bien teórico, una reacción muy violenta por parte de la escuela histórica y de los institucionalistas, una vuelta a la teoría por parte de la econometría tradicional, pero ya con un importante componente empírico y una vuelta al método de Thyco Brahe, pero mucho más atenuado, con las metodologías recientes referidas, como si paulatinamente se fuera encontrando un punto intermedio.

Las ciencias fácticas pueden dividirse en tres tipos, de acuerdo a su objeto: ciencias de lo inorgánico, de lo orgánico y de lo social. Si los fenómenos estudiados por cada una de ellas son explicados por un modelo del tipo:

$$Y = F(X) + e$$

donde  $Y$  es un vector de variables endógenas,  $X$  un vector de variables exógenas y  $e$  otro vector de términos de error aleatorios, con autocovarianza nula, varianza constante y media cero, tendremos que en principio existen diferencias entre los tres niveles. En el mundo de lo inorgánico, es decir en la física, encontraremos que la parte determinística del modelo explica un altísimo porcentaje de la varianza de las endógenas. Algo menos en la biología y menos aún en los fenómenos

estudiados por las ciencias sociales.<sup>68</sup> Además existen ciertos elementos que permanecen, por lo general, mucho más cerca de la constancia en las ciencias físicas. Por ejemplo, la forma funcional, los valores de los parámetros, el carácter de exogeneidad de las variables y las propiedades homocedásticas del término de error.

La econometría moderna está en condiciones, incipientemente, de atacar estos problemas y de adaptarse a las características propias de las ciencias sociales. Es en alguna medida lógico que sea así: la estadística fué desarrollada para ser usada en otras disciplinas, y los primeros tratamientos econométricos se derivaron de esa misma estadística, por lo que fué necesario un laborioso proceso de adaptación, comenzado en la década de los años treinta. Actualmente, por ejemplo, ya es posible suponer que los coeficientes siguen un comportamiento del tipo del camino aleatorio, o de carácter autorregresivo de primer orden, pero en el futuro seguramente será posible determinar qué variables influyen en los coeficientes, y tal vez qué variables influyen en los coeficientes de estas últimas, llegando a modelos que hoy todavía no soñamos que puedan existir. Baste pensar que la econometría como ciencia tiene solamente sesenta años, y hace apenas treinta que disponemos de computadoras, mientras que hace cinco siglos que la astronomía dispone de telescopios o la biología de microscopios. Los cambios en las técnicas econométricas y de investigación empírica son cada vez más veloces y van a brindar nuevas soluciones así como van a generar problemas metodológicos nuevos.<sup>69</sup>

También es de esperarse cambios en la supuesta imposibilidad de testear las "assumptions". La conducta de los agentes económicos comienza a ser explicada desde la perspectiva de la psicología con más profundidad. Trabajos como el de Ainslei (1992) abren una perspectiva diferente, al interpretar la conducta humana como la resultante de fuerzas contrarias que operan en el interior de la conciencia. Así como la microeconomía es el estudio de las negociaciones entre los individuos, es preciso otra disciplina, sostiene, a la que denomina "picoeconomics", dedicada al estudio de las negociaciones que se producen entre los distintos componentes de la conciencia de los agentes, lo que podría permitir un estudio del principio de racionalidad desde una perspectiva más general, que sigue la tradición que definen Benthan y von Mises, que analizan la conducta económica como una parte de la acción humana.

Por último, una reflexión sobre el pluralismo de Caldwell (1988), quien sostiene que la finalidad de la metodología económica no debe ser prescriptiva, ya que no es posible obtener un algoritmo que permita seleccionar las teorías, por lo que propone que se acepten todos los métodos para no atar a la investigación y hacerle perder su carácter creativo. Esta posición implica el riesgo de la anarquía, y a mi juicio podría ser aplicada sólo transitoriamente, hasta que exista mayor uniformidad en el pensamiento de los epistemólogos y los economistas<sup>70</sup>.

<sup>68</sup> Con otras palabras Sims (1996) desarrolla esta misma idea.

<sup>69</sup> Un ejemplo es el método conocido como *calibración*, que usan los economistas que trabajan con la teoría del ciclo real para sus experimentos computacionales. (Ver Kydland y Prescott, 1996).

<sup>70</sup> El profesor Alfredo Navarro, en su informe, cita las siguientes obras: (1) Ainslei, George. "Picoeconomics". Cambridge: Cambridge University Press. 1992. (2) Archibald, G. y Lipsey. R.G. "Monetary and value theory: a critique of Lange and Patinkin". *Review Of Economic Studies*, 26. 1958, pp. 1-22. (3) Bentham, Jeremy. "An Introduction to the Principles of Morals and Legislation". Oxford: Clarendon Press.[1780].1996. (4) Box, G. y Jenkins, G. "Time Series Analysis: forecasting and control". San Francisco: Holden Day. 1976. (5) Blaug, Mark. "The Methodology of Economics or how economists

- explain". Cambridge: Cambridge University Press.[1980] 1994. (6) Caldwell, Bruce. "Beyond Positivism Economic Methodology in the Twentieth Century". Londres: George Allen and Unwin. 1982. (7) Caldwell, Bruce. "The case for pluralism", en "The Popperian legacy in Economics", editado por Neil de Marchi. Nueva York: Cambridge University Press. 1988. pp. 231-244. (8) Caldwell, Bruce. "Clarifying Popper". *Journal of Economic Literature*, XXIX, 1, 1991, pp 1-33. (9) Campbell, J., Lo A. y Mac Kinley A. "The Econometrics of Financial Markets". Princeton: Princeton University Press. 1997. (10) Christ, Carl F. "The Cowles Commission's Contributions to Econometrics at Chicago, 1939-1955". *Journal of Economic Literature*. XXXII, 1, 1994, pp. 30-59. (11) Cooley, T.F. y Leroy, S.F. "Atheoretical Macroeconomics: a critique. *Journal of Monetary Economics*, 16, 1985, pp. 283-308. (12) Darnell A. y Evans, J. "The Limits of Econometrics". Aldershot: Edward Elgar Pub. 1990. (13) De Marchi Neil. "Popper and the LSE economists" en "The Popperian legacy in Economics", editado por Neil de Marchi. Nueva York: Cambridge University Press. 1988. pp. 139-168. (14) Fama, E. "Efficient Capital Markets: A Review of Theory and Empirical Work", *Journal of Finance*, 25, 1970, pp. 383-417. (15) Feigl Herbert. "El positivismo Lógico". *Revista de Occidente*. Agosto 1966, 41, pp. 137-166. (16) Friedman, Milton. "The methodology of Positive Economics". Chicago: Chicago University Press. [1952] 1953. (17) Friedman Milton. "A Monetary History of the United States 1867-1960". Princeton: Princeton University Press. [1963] 1971. (18) Friedman Milton. "Inflation and Unemployment" en "Nobel Lectures - Economic Sciences", editado por Assar Lindbeck. Londres: Nobel Foundation [1976] 1992. (19) Friedman Milton. "Monetary Trends in the United States and the UK". Chicago: Chicago University Press. 1982. (20) Granger, Clive, W. "Investigating causal relations by econometrics models and cross-spectral methods". *Econometrica*. 37, 1969, pp.424-38. (21) Haavelmo, Trigve. "The Probability Approach in Econometrics". Suplemento de *Econometrica*. 12, 1944, pp. 1-115. (22) Haavelmo, Trigve. "Nobel Lecture" en "Nobel Lectures - Economic Science". Editado por Karl-Goran Maler. Singapur: World Scientific. [1989] 1992. (23) Haberler, Gottfried. "Prosperidad y depresión". México: Fondo de Cultura Económica. [1937]. 1953. (24) Hausman, Daniel M. "An appraisal of Popperian Methodology" en "The Popperian Legacy in Economics", editado por Neil De Marchi. Nueva York: Cambridge University Press. 1988. pp. 65-87. (25) Hempel, Carl G. "La explicación científica". Barcelona: Editorial Paidós. {1965] 1996. (26) Hendry, David F. "Dynamic Econometrics". Oxford: Oxford University Press. 1995. (27) Hutchison, Terence. "On the relations between philosophy and economics". *Journal of Economic Methodology*, 4 (1), 1997, pp.127-154. (28) Johansen, S. "Likelihood-based inference in cointegrated vectors autorregresive models. New York: Oxford University Press. 1995. (29) Keynes, John M. "Professor Tinbergen's Method". *Economic Journal*, 49, 1939, pp. 558-68. (30) Keynes John Neville. "Scope and Method of the Political Economy". Nueva York: Kelley & Milman. [1891] 1955. (31) Koopmans, T.C. "Measurement without theory debate", en *The Foundations of Econometric Analysis*, editado por D. Hendry y M. Morgan. Cambridge: Cambridge University Press. [1947] 1995. pp. 491 y ss. (32) Kuhn, Thomas. S. "La Estructura de las Revoluciones Científicas". Madrid: Fondo de Cultura Económica. [1962] 1978. (33) Kydland, F. y Prescott, E. "The computational experiment: an econometric tool". *Journal of Econometric Perspectives*. 10, 1, 1996. pp. 69-88. (34) Lakatos, Imre. "Historia de la Ciencia y sus reconstrucciones racionales". Madrid: Editorial Tecnos. 1971. (35) Leamer, Edward E. "Sturdy Econometrics". Aldershot: Edward Elgar Pub. 1994. (36) Leamer, Edward E. "Let's take the Cons out of Econometrics". *American Economic Review*. 73(1), 1983, pp. 31-43. (37) Leamer, Edward E. "A Result on the Sign of Restricted Least Squares Estimates". *Journal of Econometrics*. 3, 1975, pp. 387-390. (38) Lipsey, Richard. "Introducción a la Economía Positiva". Barcelona: Editorial Vincens-Vives. [1963] 1971. (39) Litterman, R.B. "Forecasting with Bayesian Vector Autoregressions models". Federal Bank of Minneapolis Review. 1984, pp. 30-41. (40) McCloskey, Donald. "The Rethoric of Economics". *Journal of Economic Literature*. XXI, 2, 1983, pp. 481-517. (41) Marschak, J y Lange O. "Mr Keynes on Statistical Verification of Bussiness Cycle Tehories", en *The Foundation of Econometric Analysis*, editado por David S. Hendry y Mary Morgan. Cambridge: Cambridge University Press, 1995, pp. 390-398. (42) Marshall, Alfred. "Principios de Economía". Madrid: Aguilar. [1890] 1948. (43) Mill, John Stuart. "Autobiografía". Buenos Aires: Editorial Espasa Calpe. [1874] 1948. (44) Morgan, Mary S. "The History of Econometric Ideas". Cambridge: Cambridge University Press. 1990. (45) Padrón, Héctor. "A Propósito de las Conjeturas Parciales o Lógica de Situación de Karl Popper", mim. Instituto de Investigaciones Económicas de la Facultad de Ciencias Económicas de la Universidad Nacional de Mar del Plata. Serie de Estudios No. 1. 1982. (46) Pareto, Wilfredo. "Manual de Economía Política". Buenos Aires: Editorial Atalaya. [1922] 1945. (47) Persons, W. M. "The correlation of Economic Statistics, en *The Foundation of Econometric Analysis*, editado por David S. Hendry y Mary Morgan. Cambridge: Cambridge University Press, [1910] 1995, pp. 130-140. (48) Popper, Karl. "La Lógica de la investigación científica". Madrid: Editorial Tecnos. [1934] 1980. (49) Popper, Karl. "Búsqueda sin Término. Una autobiografía intelectual." Madrid: Editorial Tecnos. [1974] 1985. (50) Popper, Karl. "La Miseria del Historicismo". Madrid: Alianza Taurus. [1945] 1995. (51) Popper, Karl. "La Sociedad Abierta y sus enemigos". Barcelona: Editorial Paidós. [1945] 1981. (52) Rothbard, Murray N. "Ludwig E. von Mises" en "The New Palgrave: a Dictionary of Economics" editado por J. Eatwell, M. Milgate y P. Newman. Londres: The Macmillan Press . 1987. (53) Robbins Lionel. "Naturaleza y Significación de la Ciencia Económica". Mexico: Fondo de Cultura Económica. [1932] 1951. (54) Robbins Lionel. "On Latsis: a Review Essay". *Journal of Economic Literature*. XVII,3, 1979, pp. 996-1004. (55) Ross Ian S. "The Life of Adam Smith". Oxford: Clarendon Press. 1995. (56) Samuelson, Paul A. "Foundations of Economic Analysis", Cambridge: Harvard University Press. [1947]. 1966. (57) Sosa-Escudero, Walter. "La World Wide Web como herramienta de investigación empírica en economía". mim.



### 1.3. Análisis cuantitativo de la información

Presentamos en esta sección algunas de las técnicas de análisis cuantitativo utilizadas en la *investigación econométrica* y que, sobre su importancia, Klein (1958), premio Nobel de economía 1980, sostiene: “lo que comúnmente se denomina *investigación econométrica* es todo lo relativo al problema de estimación de relaciones económicas<sup>71</sup>”.

Las técnicas de análisis de datos se fundamentan en la Estadística. Cualquiera sea el ámbito de actuación del investigador en economía, ya sea que se dedique a las ciencias empresariales o ejerza cualquier otra función que tenga que ver con la investigación económica, en estas líneas, deseamos destacar los elementos necesarios para poder analizar científicamente los problemas que plantea el manejo de información.

Los problemas económicos y sociales, en general, traen cada día más a un primer plano los métodos cuantitativos contenidos en la disciplina estadística, en todo su espectro, y en la matemática aplicada. Por otra parte, el empleo habitual de las computadoras ha difundido rápidamente el análisis estadístico de los fenómenos económicos y sociales. Esto hace que el estudio de las técnicas cuantitativas, para el análisis de la información, sea una necesidad imprescindible. Ningún investigador puede ignorar los conocimientos fundamentales de la estadística. Ésta, en efecto, proporciona las técnicas más apropiadas para organizar e interpretar los datos numéricos que se obtienen en censos, muestreo, encuestas, registros contables y registros de todo tipo.

La expresión análisis cuantitativo de la información no se puede definir de una forma rápida y precisa que evite la tarea que supone asimilar el contenido de estas líneas. Aunque la expresión es un término abstracto, su importancia práctica quedará demostrada por medio de ejemplos. Pero este procedimiento, no es más que una ayuda para que el lector desarrolle su propio concepto filosófico sobre el análisis cuantitativo y llegue a dominar las técnicas de descripción e interpretación de datos.

**Ejemplo 1.3.** La aplicación de estas técnicas a los problemas económicos y sociales nos llevan a responder, entre otras, las siguientes preguntas:  
1) ¿Cómo se mide la tasa del crecimiento del producto bruto de una nación?  
¿Qué relación tiene este crecimiento con la actividad de una empresa?

---

University of Illinois. Julio 1997. (58) Schumpeter, Joseph. “*Historia del Análisis Económico*”. Mexico: Fondo de Cultura Económica. [1954] 1971. (59) Sims, Chistopher. “Macroeconomics and Reality”. *Econometrica*, 48,1, 1980, pp. 1-48. (60) Sims, Chistopher. “Macroeconomics and Methodology”. *Journal of Economics Perspectives*. 10,1, 1996, pp. 105-120. (61) Vinning, Rutledge. “Methodological Issues in Quantitative Economics”, en *The Foundation of Econometric Analysis*, editado por David S. Hendry y Mary Morgan. Cambridge: Cambridge University Press, [1946] 1995, pp. 501-513. (62) Yule G. V. “Why do we sometimes get Nonsense Correlations between Time Series?”. *Journal of The Royal Statistics Society*, vol 89, 1926, pp. 2-41 (63) Wittgesntein, Ludwig. “*Tractatus Logico-Philosophicus*”. Madrid: Alianza Universidad. [1922] 1995.

<sup>71</sup> Econometric análisis for public policy. Iowa State Cpllege Press. Ames. Iowa. P. 24. 1958.

- 2) ¿Porqué y cómo es posible pronosticar los gastos de consumo para el siguiente trimestre con el ingreso disponible del corriente trimestre?.
- 3) ¿Qué métodos cuantitativos se deben aplicar para hacer, con los datos actuales, un pronóstico de oferta y demanda para realizar una adecuada planeación de los beneficios?.
- 4) ¿Debe comercializarse o abandonarse un nuevo producto?. ¿Porqué?.
- 5) Suponga que algunas oportunidades de inversión requieren la misma cantidad de capital, pero producen diferentes ganancias en diferentes condiciones económicas (inflación, prosperidad, recesión o depresión), ¿qué oportunidad debe escogerse?.
- 6) ¿Es importante la diferencia observada durante los últimos veinte años entre las propensiones marginales a consumir a corto y a largo plazo, en una región determinada?.
- 7) Un estudio de mercadotecnia de la demanda potencial de un nuevo producto ¿debe ser realizado en una, dos o aún más etapas del desarrollo del producto?.
- 8) ¿Cuál es el Modelo Econométrico adecuado para predecir el precio óptimo esperado de un producto agrícola?.

*Relaciones entre el análisis de la información, la investigación econométrica y otras disciplinas*

Veamos, ahora, distintas disciplinas que se vinculan con las técnicas de análisis de datos y la investigación econométrica.

- a) *Teoría y técnicas contables.* Un auditor es un experto en contabilidad que lleva a cabo un examen de la información contable con el fin de dar su opinión respecto a la confiabilidad de dicha información. Tanto los auditores internos como externos recurren a pruebas selectivas y a las técnicas de muestreo. Ninguno de ellos puede hacer lo que podría llamarse un examen completo y detallado de cada transacción, debe estimarlas y explicarlas a través de investigaciones econométricas. Por otro lado, para presentar la información correctamente, los Estados Contables deben cumplir con el principio generalmente aceptado de "exposición". Se deberá hacer uso de técnicas no paramétricas de presentación de la información. Cuadros, gráficos, índices e informes deben cumplir con aquel principio y para ello el conocimiento de la estadística no paramétrica será de ayuda especial. En la Contabilidad de Costos, entre otras aplicaciones, tanto en el presupuesto de venta como en el de costos, para el planeamiento de utilidades, se hace uso de los métodos econométricos, aplicando análisis de tendencia y de correlación, estimación de demanda, etc. Pero también, se emplean los costos en la toma de decisiones para la determinación de precios, en estos casos, es importante la contribución de los métodos econométricos. Además, el análisis estadístico proporciona la medida más precisa de variabilidad de costo, a través del diagrama de dispersión que descubre la línea de tendencia. En sindicatura, en el informe general, en la determinación del periodo de sospecha y en la calificación de la conducta, establecida legalmente, el Síndico debe hacer uso del análisis de correlación, regresión, índices, etc., para hacer más objetiva su interpretación.

- b) *Economía*. Los organismos públicos y privados divulgan ingentes cantidades de información numérica. La mayor parte de lo que sabemos sobre las formas de las curvas de la teoría económica procede del cuidadoso análisis estadístico de los datos reales y es de observar que las técnicas de análisis cuantitativo encuentran en el campo de la economía sus aplicaciones más importantes. Por otra parte, es de destacar el carácter acentuadamente analítico y cuantitativo del enfoque de las decisiones de inversión y financiación de una empresa. Esta orientación analítica y cuantitativa se desprende naturalmente de la importancia cada vez mayor que se asigna al análisis matemático y estadístico de los estudios financieros. Destacamos, en este campo fundamentalmente, el Modelo Econométrico de Harry Markowitz para la diversificación eficiente de inversiones.
- c) *Ciencias empresariales*. Es bien sabido el uso que de los modelos estadísticos, econométricos y de análisis multivariante hace la mercadotecnia para cuantificar la información del mercado con respecto a precios, productos, publicidad, necesidades, deseos y demanda de los consumidores.

Como decíamos, estas técnicas basadas, fundamentalmente, en el análisis de la información a través de la investigación econométrica, utilizan elementos de matemática, estadística y teoría económica. Como expresa Pulido (1993), al diferenciar **medición sin teoría y teoría sin medición**, *si la medición sin teoría debe considerarse como un camino erróneo de perfeccionamiento científico, también debe ser radicalmente rechazada – en el extremo opuesto la teoría no sometida a medición, a contrastación*. Ya hace muchos años Joseph Schumpeter (1933) afirmaba categóricamente que *cada economista es un econométra*, lo desee o no y añadía: *porque mientras no seamos capaces de exponer nuestros argumentos en cifras, la voz de nuestra ciencia, aunque ocasionalmente puede ayudar a dispersar errores groseros, nunca será oída por los hombres prácticos. Son, por instinto, econométras todos, en su desconfianza de las cosas no sujetas a una prueba exacta*<sup>72</sup>.

Como ya ha quedado delineado, varias son las disciplinas que dan origen a las técnicas de análisis econométrico. Todas ellas, a su vez, tienen un fundamento matemático, más precisamente en el Análisis Matemático. De esta manera se transforman en disciplinas cuantitativas que nos proveen de un vasto herramental, rico en posibilidades, para complementar con nuestro conocimiento, con nuestra formación, eminentemente fundada en la interpretación de datos de la realidad micro o macroeconómica.

Los investigadores estamos habituados a recibir cuantiosa información que debemos interpretar y sistematizar para contrastar hipótesis que permitan resolver problemas y tomar decisiones. Al definir la investigación no nos detendremos en el estudio de la presentación de

<sup>72</sup> Pulido, Antonio. Modelos econométricos. Editorial Pirámide. Madrid 1993. p. 48

datos sino que profundizaremos a través de ellos, logrando interpretar, sistematizar y descubriendo relaciones empíricas que permitan realizar un análisis cuantitativo de la información a través de las técnicas contenidas en la disciplina econométrica.

Por último, podemos mencionar que nuestras materias se relacionan con la toma de decisiones: la Estadística se podría definir como el conjunto de métodos para la toma de decisiones frente a la incertidumbre; y la Econometría como la aplicación de aquella a datos para tomar decisiones frente a los fenómenos micro o macroeconómicos. Por tanto, todas las técnicas cuantitativas, basadas en esas disciplinas, tienen que ver con el análisis de la información para la toma de decisiones, ya sea a través de la descripción de fenómenos económicos o de predicción de los mismos.

### *Expectativas racionales vs. Expectativas adaptativas*

Para el uso de la econometría en el análisis económico debemos tener presente la *Crítica de Lucas*<sup>73</sup>.

La econometría ha sido un instrumento muy utilizado para verificar la validez de las hipótesis planteadas por las teorías económicas y, con base en ello, estimar modelos econométricos para inferir la evolución futura de algunas variables ante políticas económicas dinámicas.

Estos modelos utilizan información histórica para estimar los parámetros que relacionan a las variables económicas y predicen el comportamiento futuro de las variables de interés, con base en dichos coeficientes estimados. Para Lucas, este procedimiento puede llevar a conclusiones incorrectas, puesto que se asume que los coeficientes estimados permanecen constantes aún ante situaciones de políticas económicas distintas.

Específicamente, Lucas afirma que la econometría considera que los agentes económicos miran hacia atrás para formular sus proyecciones futuras, es decir, tienen expectativas adaptativas, cuando en realidad ellos se comportan de acuerdo con la teoría de las expectativas racionales. Además, hace énfasis en las decisiones microeconómicas de todos los agentes mediante criterios racionales y que al agregarlos permiten sacar conclusiones acerca de la posible evolución de las variables macroeconómicas.

Tal y como lo expone Lucas los modelos econométricos estiman una relación de comportamiento de la forma:

---

<sup>73</sup> Extraído del informe del Banco Central de Costa Rica, División económica, Departamento de investigaciones económicas (1996) *Evaluación del uso de la econometría en el Análisis económico: la crítica de Lucas*. Rigoberto Araya Monge. Norman Orozco Coto (Autorizado por Claudio A. Ureña Ch), Incluye observaciones del Dr. Juan E. Muñoz G. Lucas, R. op. Cit. Precisamente, el importante y revolucionario aporte que dio Lucas con su teoría de las expectativas racionales al análisis económico durante la década de los 70, le valió el Premio Nobel de Economía concedido en 1995.

$$(1.2) \quad Y_{t+1} = f(Y_t, X_t, \varepsilon_t)$$

Donde:

$Y_{t+1}$  variable que se quiere explicar en el periodo  $t+1$  (variable dependiente)

$Y_t$  variable estado en el periodo  $t$

$X_t$  variables exógenas en el periodo  $t$

$\varepsilon_t$  término aleatorio en el periodo  $t$

La estimación de esta ecuación se realiza con los valores pasados de las variables exógenas y se obtiene un vector  $k$  de parámetros estimados que miden la forma y magnitud del efecto de las variables exógenas sobre las dependientes. Al estimarse  $k$  con observaciones históricas, se supone que permanece invariable, por lo que pretender realizar pronósticos de las variables económicas con modelos econométricos, en un contexto en el que los agentes económicos prevén cambios en la política económica, sería un procedimiento incorrecto para *tomar decisiones*.

El punto fundamental de la crítica se refiere a que las estimaciones econométricas a nivel macroeconómicas consideran que los agentes económicos, como un todo, se basan en expectativas adaptativas para pronosticar la evolución de las variables económicas de interés. En este sentido el valor futuro de una variable se basa en los valores pasados de ésta y de sus determinantes.

Este enfoque ignora que las decisiones en el nivel macro provienen de las del nivel micro, donde los agentes pueden diferir en cuanto a sus percepciones acerca de las políticas económicas futuras. Lucas expresa que las personas se comportan racionalmente, lo cual implica que ellos conocen el "modelo" que describe el comportamiento de las variables de su interés y que por tanto son capaces de modificar su actuación presente ante cambios esperados en el entorno macroeconómico. En este caso se supone que las proyecciones futuras de los individuos son, en promedio, correctas.

Por ejemplo, la estimación de un modelo para el consumo privado en el cual éste depende, entre otras cosas, del ingreso permanente, puede dar señales incorrectas acerca de la evolución futura del consumo, en tanto los agentes económicos esperen cambios en sus ingresos permanentes. De ser así, el coeficiente que relaciona el ingreso con el consumo estimado con datos históricos, no reflejaría el nuevo nivel de consumo que estarían dispuestos a realizar los individuos.

Por otra parte, Lucas comenta que además de la deficiencia que tienen los modelos econométricos por no contemplar las expectativas racionales de los agentes económicos, también están expuestos a la *calidad de la información* disponible para realizar las estimaciones. Critica el hecho de que muchas veces el tamaño de la muestra utilizada se circunscribe a periodos de bastante estabilidad de las variables económicas, lo cual no considera correcto debido a que se

ignoran los cambios de política o shocks que se hayan dado y que afectan las decisiones de los individuos y por ende los coeficientes estimados (el vector  $k$ ).

No necesariamente esta crítica invalida por completo la utilización de la econometría en el análisis económico. Lucas considera que este instrumento podría ser útil para pronosticar a corto plazo pero no a largo plazo, debido a que, por lo general, los ajustes de la economía ante posibles *shocks* no se realizan en forma inmediata por la rigidez misma del sistema o debido a que en el corto plazo los agentes se basan en expectativas de tipo adaptativo para tomar sus decisiones. No obstante, en el largo plazo las expectativas racionales de los agentes se cumplen, de aquí que los pronósticos basados en modelos econométricos pueden ser erróneos.

Cabe señalar que se han desarrollado técnicas econométricas relativamente recientes con las que se puede prescindir la crítica de Lucas. Estos enfoques se refieren al *test de exogeneidad* y *simulación contrafactual*, que tienen cierto grado de complejidad.

Sin lugar a dudas *esta crítica* al uso de la econometría en el análisis económico constituye un aporte revolucionario, por cuanto pone en entredicho la validez de este instrumental utilizado por muchos años para el análisis económico.

Esta crítica es especialmente relevante en épocas en las que el sistema económico atraviesa por cambios continuos y profundos, acerca de los cuales los agentes económicos forman sus propias expectativas de una forma racional.

Unido a esta crítica existe, en algunas ocasiones, el problema de la calidad de los datos de las variables involucradas en las distintas estimaciones econométricas y la disponibilidad de los mismos para periodos amplios. Esto, indudablemente hace aún más relevante la conclusión a la que llega Lucas.

En esta línea de razonamiento surge entonces el cuestionamiento sobre la utilidad y la bondad de las investigaciones econométricas.

Al respecto cabe destacar que la crítica de Lucas no invalida el uso de la econometría por completo. Más bien ha conducido al desarrollo y a la aplicación de técnicas más avanzadas que han hecho de la econometría un instrumento cada vez más útil y complementario del análisis económico formal. Prueba de ello lo constituyen la cada vez mayor cantidad de textos dedicados al análisis econométrico y la elaboración de *software* econométrico.

En el contexto de la investigación económica se pueden citar varios determinantes que respaldan el uso de la econometría en el estudio de las series económicas. Entre ellos:

- 1) La estimación de los coeficientes de una regresión lineal incorpora las respuestas de los agentes económicos ante

cambios de política. En otras palabras, las series de tiempo, aun cuando históricas, muestran la reacción racional de los agentes de una economía.

- 2) Es poco probable que en el corto plazo un cambio de política o un *shock* al sistema conduzca a variaciones significativas de las estimaciones de los parámetros de un modelo.
- 3) Si las estimaciones superan las pruebas de estabilidad, es poco probable que, con base en el mismo modelo, los eventuales *shocks* que se puedan presentar en un futuro generen estimaciones significativamente diferentes de las ya obtenidas.
- 4) El análisis contrafactual, o de simulación, puede brindar un complemento importante para evaluar el futuro curso de las variables endógenas, en vista de que no sólo considera el pronóstico con base en el comportamiento de las variables exógenas, sino, también incluye la variación de los estimadores dentro de rangos racionalmente permisibles.

Sin embargo, es necesario tener conciencia de la "*crítica de Lucas*" para ser cauteloso a la hora de utilizar la econometría y no olvidar sus limitaciones, de tal forma que no se "abuse" de este instrumental más allá de lo que éste puede dar. Para llevar a cabo una buena investigación económica, es necesario contar con una buena teoría económica del fenómeno que se pretende explicar y usar la econometría en forma correcta.

#### 1.4. El Premio Nobel en Economía y la Econometría

Además de las contribuciones de Robert Lucas, otros premios Nobel en economía, enriquecieron nuestra disciplina. A continuación, sobre la base de un artículo de Pulido (1993), contamos el avance de la econometría a través del tiempo<sup>74</sup>. Esta no es una historia de la econometría sino que introduce la importancia de la misma en el desarrollo científico de la economía contemporánea<sup>75</sup>.

En 1969, la Academia de Ciencias de Suecia, otorga, por primera vez, un premio Nobel en el campo de la Economía. Ese año, el primer Nobel en Economía fue para dos econométricos, R. Frish y J. Tinbergen. A lo largo de la historia, de entrega de estos premios máximos en la labor de los científicos, en el área de Economía, el trabajo de los econométricos fue galardonado en varias oportunidades.

---

<sup>74</sup> Pulido, A. *25 años de experiencia en econometría aplicada*. Revista Estudios de Econometría Aplicada. 1993 ps. 81-98.

<sup>75</sup> Para un detalle histórico de la Econometría, puede consultarse a Roy Epstein *A history of Econometrics*, North-Holland. Ámsterdam 1987.

Como decíamos, en 1969 se le otorga el premio a Ragnar Frish (1895-1973) **por haber desarrollado y aplicado modelos dinámicos en el análisis de los procesos económicos**. Pero, además, realizó trabajos sobre contabilidad nacional, economía axiomática, teoría econométrica, índices de precios y modelos de decisión en economía; fue creador de las palabras econometría y multicolinealidad. Decía, *la econometría es una herramienta poderosa, pero también peligrosa. Hay muchas ocasiones de abusar de ella y de emplearla más para el mal que para el bien, por lo que sólo debe ponerse en manos de hombres de primera clase*<sup>76</sup>.

Jan Tinbergen, compartió el primer Nobel con Frish, se le otorgó dicho premio por el mismo motivo, en relación con la aplicación del primer modelo macroeconómico (modelo multiecuacional – Holanda, 1937). Realizó trabajos sobre ciclos económicos y planificación del desarrollo. Fue autor del modelo de la Telaraña. Decía: *una vez aceptada con mayor generalidad la necesidad de una activa política gubernamental, en la esfera del desarrollo económico, se han construido modelos orientados en el sentido de las decisiones del gobierno, en otras palabras, destinados al problema político*<sup>77</sup>.

A Wassily Leontief, se le otorga el premio Nobel en el año 1973, **por el desarrollo del método input-output y por su aplicación a importantes problemas económicos**. Su argumento se basaba en la siguiente reflexión, *sin invocar una analogía metodológica fuera de lugar, la tarea de asegurar un flujo masivo de datos económicos primarios puede compararse a la del físico de proporcionar energía suficiente a un gigantesco acelerador atómico. Los científicos tienen sus máquinas mientras que los economistas están todavía esperando sus datos. En nuestro caso, no sólo debe la sociedad estar dispuesta a proporcionar, año tras año, los millones de dólares requeridos para el mantenimiento de la vasta máquina estadística, sino que un amplio número de ciudadanos debe disponerse a jugar un papel pasivo y, ocasionalmente, incluso tomar parte activa en las operaciones actuales de búsqueda de datos. Es como si debiera persuadirse a electrones y protones a cooperar con el físico*<sup>78</sup>.

Tjalling Koopmans, premio Nobel 1975, compartido con L. Kantorovich, **por sus aportaciones a la teoría de la asignación óptima de recursos**. Discípulo de Frisch y Tinbergen, fue el propulsor del enfoque de máxima verosimilitud (debido a R. Fisher) en econometría y de la teoría de la identificación. En sus palabras, *fue capaz (Kepler) de encontrar simples leyes empíricas que estaban de acuerdo con observaciones pasadas y permitían la predicción de observaciones futuras. Este resultado fue un triunfo para el enfoque en que la recogida, aporte y depuración de datos en gran escala precede a la formulación de teorías y a su contraste posterior por los hechos*<sup>79</sup>.

<sup>76</sup> Frish, R. *The responsibility of the econometrician*. Econometrica. 1945

<sup>77</sup> Tinbergen, J *Planificación del desarrollo*. Editorial Guadarrama. 1967.

<sup>78</sup> Leontief, W. *Theoretical assumptions and non observed facts*. The American Economic Review. 1971.

<sup>79</sup> Koopmans, T. *Measurement without theory*. Review of Economic and Statistics. 1974.



Lawrence Klein, premio Nobel en Economía 1980, porque gracias a sus contribuciones, la construcción de modelos económicos ha alcanzado difusión general, por no decir universal. Su sistema es utilizado en todo el mundo no sólo por instituciones científicas, sino también en la practiad de la Administración Pública y las grandes empresas. Considerado el padre de la Econometría moderana, ha realizado trabajos en economía teórica, política económica, métodos econométricos y modelos macroeconómicos aplicados, como el modelo de Wharton y, ha contribuido al desarrollo del Proyecto Link<sup>80</sup>.

Richard Stone, premio Nobel de Economía 1984, **por haber efectuado trabajos de una importancia fundamental para la elaboración de sistemas de Contabilidades Nacionales, los que permiten mejorar radicalmente las bases del análisis económico empírico.** Fue discípulo de Keynes, se destacan sus trabajos en modelos económicos, contabilidad social y demografía. Director del modelo de Cambridge: "A Computable Model of Economic Growth", expresaba que *construir un modelo económico o social es una actividad compleja que implica muchos ingredientes. Si uno quiere modelizar el mundo real debemos disponer ante todo de datos... En segundo lugar tenemos que tener teorías, esto es hipótesis acercad de en qué sentido se relacionan unas variables con otras. Como tercer componente, debemos tener métodos de estimación, ... En cuarta posición habrá de disponerse de métodos de solución para resolver el conjunto de ecuaciones que constituyen al modelo... y finalmente de método de control, que aseguren que nuestra solución satisface ciertas condiciones o restricciones*<sup>81</sup>.

Trygve Haavelmo, premio Nobel en el año 1989, **por la elaboración de los fundamentos probabilísticos de la metodología econométrica y su análisis de las estructuras económicas simultáneas.** También realizó trabajos en métos de estimación, problemas de interdependencia, identificación y modelos aplicados. Es conocido el Multiplicador de Haavelmo del gasto público y su crítica a la estimación aislada de ecuaciones. Nos deja esta reflexión: *En otras ciencias cuantitativas el descubrimiento de leyes ha traspasado el estudio de individuos aislados para pasar a grandes laboratorios científicos donde trabajan multitud de expertos...¿Puede esperarse menos de la investigación económica, cuando sus resultados constituyen la base de la política económica de la que pueden depender miles de millones de dólares de renta nacional y el bienestar general de millones de personas?*<sup>82</sup>.

Harry Markowitz, premio Nobel 1990, compartido con Merton Miller y William Sharpe por su trabajo pionero en la teoría de la economía financiera. Publicó en 1952 el artículo que se considera el origen de la teoría de selección de carteras y la consiguiente teoría de equilibrio en el mercado de capitales. El principal aporte de Markowitz se encuentra en recoger de forma explícita en su modelo los rasgos fundamentales

<sup>80</sup> Klein, R. *Prologo al libro de A. pulido, Modelos Econométricos. 1989.* op. Cit.

<sup>81</sup> Stone, R. *Aspect of Economic and Social Modelling.* Droz. 1982.

<sup>82</sup> Haavelmo, T. *The Probability Approach in Econometrics.* Econometrica 1944.

de lo que en un principio se puede calificar como conducta racional del inversor, consistente en buscar aquella composición de la cartera que haga máxima la rentabilidad para un determinado nivel de riesgo, o bien, un mínimo el riesgo para una rentabilidad dada.

Robert Merton y Myron Scholes, obtuvieron el Premio Nobel de economía en 1997, **por desarrollar un nuevo método para determinar el valor de los derivados**. Se destaca el aporte de estos economistas al estudio de las series de tiempo aplicadas al análisis financiero.

James Heckman y Daniel L. McFadden fueron reconocidos conjuntamente con el Premio Nobel de economía en 2000, **por desarrollar la teoría y los métodos de análisis de datos estadísticos que son actualmente utilizados, ampliamente, para estudiar comportamientos individuales en economía y en otras ciencias sociales**.

Daniel Kahneman a obtenido el Premio Nobel en el año 2002, **por haber integrado los avances de la investigación psicológica en la ciencia económica especialmente en lo que se refiere al juicio humano y a la adopción de decisiones bajo incertidumbre**. Afirma que cuando elegimos no siempre lo hacemos objetivamente". Mediante estudios experimentales ha demostrado que tales faltas de objetividad tienden a seguir patrones regulares que admiten una descripción matemática. Realizó una profunda crítica a la teoría de la utilidad como modelo de la adopción de decisiones bajo riesgo. En general, los individuos subestiman los resultados que son solo probables en comparación con los resultados que son obtenidos con seguridad. Esta tendencia, a la que llama "efecto certidumbre", contribuye a la aversión al riesgo en elecciones que implican ganancias seguras y a la preferencia por el riesgo en elecciones que implican pérdidas seguras. Señala también, el efecto aislamiento: "la gente tiende a ignorar componentes que son compartidos por todas las alternativas por lo que aparecen inconsistencias en las preferencias cuando la misma elección es presentada de forma diferente"<sup>83</sup>.

Robert Engle fue galardonado con el Nobel de Economía 2003, por sus métodos de análisis de series de tiempos económicas. Clive Granger fue reconocido, conjuntamente con Engle, por sus investigaciones que han dado nuevas herramientas para analizar y predecir evoluciones económicas y financieras.

El premio noble de economía 2004 fue otorgado al noruego Finn Kydland y al estadounidense Edward Prescott, **por su análisis de la influencia del modo de la determinación de la política económica en los ciclos económicos**. Según la Real Academia Sueca de Ciencias, su investigación ha transformado la teoría de los ciclos económicos al integrarla con la teoría del crecimiento económico. Las reglas de juego estables serían claves para el crecimiento económico. Un sistema de reglas puede garantizar que el

---

<sup>83</sup>Kahneman, D. *Teoría de la prospección*. Econometrica. 1979

gobierno no sorprenderá a los agentes económicos modificando inesperadamente su política, permite alcanzar resultados superiores a aquellos que se obtienen bajo un régimen discrecional. Sería preferible, según sus conclusiones, *una legislación que introduzca imperfecciones a la economía a una ausencia absoluta de reglas*. En 1977 ambos economistas tomaron como referencia a la **economía argentina** para el análisis de la inconsistencia intertemporal. Demostraron que las autoridades tienden a abandonar las metas a largo plazo para obtener beneficios a corto plazo. Además, con la utilización de un modelo, demuestran que los impactos por el lado de la oferta, como los *shocks* tecnológicos, son determinantes del ciclo económico. Hasta la década del '70 la influencia del keynesianismo había dominado el análisis económico, y las investigaciones habían puesto énfasis sobre el lado de la demanda, como determinantes del ciclo económico. A partir de entonces, y como resultado de los *shocks* de los precios del petróleo, se comenzó a tener en cuenta el lado de la oferta como determinante de los ciclos. Kidland trabaja en proyectos de investigación sobre las políticas monetarias de Irlanda y Argentina. Opinó que *en los países latinoamericanos, la política monetaria no es para nada creíble*. Es coautor de un artículo llamado "La Década perdida de la Argentina y su Recuperación: fracasos y éxitos del modelo neoclásico." Sin reglas, es probable que haya inflación, aunque la inflación no sea un resultado deseado por las autoridades económicas.

Hemos puesto de manifiesto la importancia de la información y los datos, por eso nos ocuparemos, no sólo del análisis econométrico, sino de la generación de información, de sobre manera en la instancia local o regional, donde la orfandad de la misma es más que evidente; por otra parte, se manifiesta de prioridad para la aplicación de "buenos" modelos econométricos, que puedan corroborar empíricamente teorías económicas, en espacio y tiempo determinado. Por lo visto, la mayoría de los nobeles en economía han dedicado parte de su discurso a destacar dicha importancia.

### 1.5. Econometría, sistemas de información y producción de datos

Presentamos, en estas líneas, el fundamento de la aplicación de la econometría al estudio de los espacios sociales (regionales, nacionales o supranacionales). La diferencia sustancial entre la teoría econométrica y la investigación econométrica radica que en la segunda, a veces llamada econometría empírica, estamos interesados en la comprobación empírica de teorías económicas, por lo que debemos ocuparnos de los sistemas de información y de la producción del dato. Más que una diferencia, en este manual, consideramos que los econométricos, debemos ocuparnos de ambas cuestiones. Formalizar la teoría y demostrar la aplicación de la misma a los efectos de colaborar con la interpretación socioeconómica de los espacios, en tiempos determinados.

En los estudios econométricos empíricos, usamos el razonamiento estadístico como una prueba subalterna de verificación de conceptos, por lo que, el trabajo de conceptualización no se resume a contrastar asociaciones y correlaciones entre características medidas. Sobrepasamos los límites del análisis objetivo de la realidad por la inevitable carga de expresiones con las que designamos conceptos y formulamos enunciados. Incorporamos, a la interpretación de las informaciones producidas, la consideración de las condiciones de producción de esa información.

Cualquier proposición inicial de una investigación econométrica se apoya tanto en la teoría estadística como en la teoría económica. La función del econometrista es la generación y el análisis de la información, pero también, es la interpretación de esta información empírica. Es aquí, justamente, donde la teoría se complementa con la investigación, dándole a los métodos econométricos un sentido de aplicación.

En Ciencias Sociales, en general, y en la Economía, en particular, para enunciar y generalizar el resultado de una observación, no se puede conservar el marco de un razonamiento empírico estricto. No es posible olvidar: (1) la definición de la investigación, que permite formular esos enunciados, designar el contexto de medición de las interacciones entre las variables tratadas –lo cual rige la generalización y las comparaciones posibles– y enumerar elecciones metodológicas de la construcción de la información –las formas técnicas de recolección de datos rigen el sentido de la información construida– (2) la selección de características y de unidades de observación en cada caso –tabla de datos– (3) la fuente de información apropiada, (4) la recolección de la misma, para su posterior procesamiento y (5) la estrategia de tratamiento de los datos.

Asimismo, todo enunciado de una proposición general sobre un objeto social extraída de la medición empírica de una realidad social, implica, necesariamente, el paso a la interpretación de resultados y aceptar el riesgo que comporta ese paso obligado. Para formular el enunciado de una proposición general sobre un objeto social, extraída de la medición de una realidad, el investigador econométrico debe pasar por las etapas de una investigación.

Dos lecturas son posibles cuando se procede a interpretar variables a partir de la observación de un fenómeno social: (1) la empírica, que determina la variable y (2) la contextual, que toma en cuenta una información que no figura estrictamente en la variable. La manera de tener en cuenta el contexto de la información es por medio de la construcción de modelos.

Por tanto, el paso al sentido de lo observado, depende de la generación, descripción y análisis de la información social y la interpretación de esa información a través de modelos que resultan del cruce de variables para un mejor conocimiento del contexto socioeconómico.

Trataremos de dar contenido metodológico a la generación y análisis de información y a la construcción de modelos. Es lo que se ha denominado investigación econométrica.

**El objetivo es desarrollar una metodología de investigación aplicada a la econometría para el estudio de variables socioeconómicas que permita la elaboración de modelos econométricos para describir, analizar, comparar, predecir, corroborar e interpretar empíricamente, sobre regiones, más o menos homogéneas, en el contexto regional, nacional o supranacional, teorías económicas.**

Para alcanzar el objetivo se estudia el método de la econometría que permite a cualquier investigador aplicar esta metodología y obtener como resultado indicadores sociales y económicos sobre poblaciones, definidas como regiones, a partir de la evidencia empírica proporcionada por diseños de fuentes de información. Las etapas son: (1) definición de la investigación, (2) planteamiento de una tabla de datos, (3) diseño de fuentes de información, (4) recolección y procesamiento de datos y (5) análisis e interpretación econométrica de la información.

---

---

### *CASOS DE ESTUDIO, PREGUNTAS Y PROBLEMAS*

#### Caso 1.1. Formulación de una investigación regional

En nuestro trabajo sobre metodología estadística para el estudio de regiones, establecemos la correspondencia teórica de definición de investigación con la aplicada al estudio de variables socioeconómicas en un espacio regional. Deseamos especificar una metodología para cuantificar variables sociales y económicas que permitan el estudio y comparación de regiones.

De acuerdo a esto, fijamos objetivos y formulamos hipótesis para el diseño de la investigación. La población a estudiar es un espacio regional homogéneo que contiene información para la cuantificación de variables. Aunque, en general, se dispone de datos económicos y sociales, no siempre precisos y adecuados al espacio de decisión, afirmamos que en los países hay pocos indicadores ajustados a la dimensión humana y regional. Esta situación hace necesario el desarrollo de técnicas que provean de indicadores a los entes de decisión.

Con estos argumentos, tratamos los sistemas para organizar la información. Definamos una tabla de  $n$  unidades de observación que, en este contexto, son regiones, por  $k$  características observables, variables socioeconómicas, que son la base para la elaboración de indicadores.

Luego, presenta una metodología para determinar las macromagnitudes regionales, tanto sociales como económicas, como guía para la compilación de sistemas de estadísticas económica, sociales y demográficas.

Asimismo, desarrolla técnicas de muestreo que permiten relevar la información social y económica y aplica técnicas de muestreo para empresas, sector rural y hogares.

Además, incluye instrumentos de recolección de datos para los diseños presentados y una metodología para el procesamiento de los datos, tomando como ejemplo uno de los indicadores definidos.

Por último, para el análisis de la información, plantea caracterizar las regiones a través de Indicadores considerados básicos para determinar la evolución de las mismas. Presenta, también, los elementos necesarios para la evaluación de regiones y un modelo de análisis factorial que permita determinar las características que deben reunir las mismas para ser consideradas homogéneas y comparables.

De esta manera, el enfoque metodológico propuesto entiende que se debe primero determinar qué variables se estudiarán y dónde se observarán, segundo qué diseño de muestreo aplicar y por último, a través de qué métodos se analizará la información obtenida. Por otra parte, enfatiza en regionalizar la información para ayudar en la toma de decisiones descentralizadas y aliviar la situación social de los espacios emergentes dentro de un país.

## Preguntas

- 1.1 ¿Qué le sugerirías al investigador social respecto a la metodología empleada?
- 1.2 ¿Cuál serían, a tu entender, las variables que este investigador debe incluir en su investigación?
- 1.3 ¿Te parece que el investigador debe usar solamente fuentes primarias de datos?
- 1.4 ¿Qué relación tiene esta investigación con una investigación econométrica tal cual tu parecer?
- 1.5 ¿Puedes sugerir una tabla de datos para este investigador?. En base a ella, ¿cómo formularías un modelo econométrico y cuál o cuáles serían las teorías económicas a corroborar empíricamente?

## Problemas

- 1.1 ¿Qué es la econometría?
- 1.2 Resume en una carilla la epistemología de la econometría, según el estudio presentado por el Dr. Alfredo Navarro
- 1.3 ¿Qué diferencia hay entre teoría econométrica y econometría empírica?
- 1.4 ¿A quién se debe y en qué consiste la Crítica de Lucas?
- 1.5 ¿Cuál es el objetivo de la econometría?
- 1.6 Entra a la página web <http://nobelprize.org/economics/laureates/index.html> elige uno de los economistas que fueron premiados por sus trabajos en econometría y has un resumen de su autobiografía.

## Bibliografía

- Araya Monge, Rigoberto y N.Orozco Coto. (1996) *Evaluación del uso de la econometría en el Análisis económico: la crítica de Lucas*. Extraído del informe del Banco Central de Costa Rica, División económica, Departamento de investigaciones económicas.
- Barbancho, Alfonso G. (1962) *Fundamentos y posibilidades de la Econometría*, Ediciones Ariel, Barcelona.
- Christ, C. (1966) *Modelos y métodos econométricos*. Cowles Foundation for research in Economics at Yale University.
- Frish, R. (1945) *The responsibility of the econometrician*. Econometrica.
- Haavelmo, T. (1944) *The Probability Approach in Econometrics*. Econometrica
- Hernández Sampieri, Roberto; Fernández Collado, Carlos y Baptista Lucio, Pilar. (2000). *Metodología de la Investigación*. Segunda Edición. McGraw Hill. México.
- Kahneman, D. (1979) *Teoría de la prospección*. Econometrica.
- Klein, R. (1989) *Prologo al libro de A. pulido, Modelos Econométricos*.
- Koopmans, T. (1974) *Measurement without theory*. Review of Economic and Statistics.
- Leontief, W. (1971) *Theoretical assumptions and non observed facts*. The American Economic Review.
- Lucas, R. "Econometric Policy Evaluation: A Critique", en *Studies in Business Cycle Theory*.
- Navarro, A. (1997) *Reflexiones sobre la relación entre economía, econometría y epistemología*. Trabajo presentado en la conferencia de incorporación a la Academia Nacional de Ciencias Económicas. Buenos Aires.
- Padua, Jorge. (1996) *Técnicas de Investigación Aplicadas a las Ciencias Sociales*. Fondo de Cultura Económica. México.

- Pulido, A. (1993) *25 años de experiencia en econometría aplicada*. Revista Estudios de Econometría Aplicada.
- Pulido, Antonio. (1993) Modelos econométricos. Editorial Pirámide. Madrid.
- Samuelson, P. (1972) *Curso de economía moderna*. Editorial Aguilar
- Stone, R. (1982) *Aspect of Economic and Social Modelling*. Droz.
- Timbergen, J (1967) *Planificación del desarrollo*. Editorial Guadarrama.



Capítulo 2. INVESTIGACIÓN ECONOMÉTRICA .....	59
2.1. El proceso de investigación.....	59
2.2. El marco teórico .....	61
Distintos tipos de investigación .....	62
El sentido econométrico del análisis en investigación .....	64
2.3. Método de la econometría .....	66
Etapa 1: Definición de la Investigación, Orientación en el campo de la Investigación y Formulación de un sistema de hipótesis ...	66
Etapa 2: Planteo de una tabla de datos y diseño de encuesta ...	67
Etapa 3: Diseño de las fuentes de información.....	67
Etapa 4: Recolección, procesamiento y organización de los datos .....	68
Etapa 5: Análisis y presentación de la información .....	68
2.4. Definición de la investigación econométrica .....	71
Objetivo, Hipótesis y Tesis .....	71
Fuentes de información para desarrollar la hipótesis.....	73
<i>Documentación descriptiva</i> .....	73
<i>Estudio de la situación</i> .....	74
<i>Documentación explicativa</i> .....	75
<i>Teoría</i> .....	75
2.5 Clasificación de modelos .....	78
2.6. Modelos económicos y modelos econométricos.....	78
2.7. Modelos Econométricos y contrastes de Teorías Económicas .	81

CASOS DE ESTUDIO, PREGUNTAS Y PROBLEMAS .....	84
Caso 2.1: ¿Qué investigación econométrica te gustaría realizar? ..	84
Problemas.....	84
Bibliografía .....	85

## Capítulo 2. INVESTIGACIÓN ECONOMÉTRICA

### 2.1. El proceso de investigación

El proceso científico de la economía encuentra en la econometría métodos plausibles capaces de comprobar empíricamente la validez de sus modelos. Es posible justificar el razonamiento que emplearemos en esta obra mencionando lo que Klein (1957) decía: *En la teoría económica de la producción y en la de la empresa, la programación lineal y el análisis de factores y productos – al referirse al modelo de Leontief – contienen ramales econométricos. Cuando estos dos métodos se llevan a un nivel empírico de reunir datos y estimar coeficientes o relaciones completas, nos encontramos con problemas econométricos. El trabajo empírico en estos campos **no ha sido adecuadamente enlazado** con la teoría de la probabilidad y con la inferencia estadística, pero esto no quiere decir que los problemas estén fuera de los límites de estas dos materias. Se debe sencillamente a que no han sido todavía adecuadamente formulados como problemas econométricos y porque los datos empíricos utilizados no han constituido una muestra razonable<sup>1</sup>.*

Aunque Kinnear y Taylor (1993) se refieren a la investigación de mercado, seguimos lo citado por ellos, en el sentido de que elaboramos este capítulo, pensando lograr una visión general de las etapas del proceso de investigación econométrica; satisfacer la necesidad, frecuentemente expresada por los estudiantes, de conocer qué es una investigación econométrica y, finalmente, introducir el método que utiliza la econometría empírica<sup>2</sup>. Asimismo, ilustramos las etapas de la investigación econométrica y los tipos de investigación que pueden utilizarse. Por último, deseamos demostrar que la econometría está inserta en un proceso que, paso a paso, va llevando al investigador a corroborar sus teorías. *En este sentido, este capítulo sirve de referencia unificadora para quien lea los siguientes que tratan aspectos específicos de la econometría, en tanto, aplicación de la estadística y las matemáticas a la economía<sup>3</sup>.*

El proceso formal de investigación econométrica se puede considerar como una serie de etapas que encuentran su origen en la metodología de la investigación. El cuadro 2.1 ilustra las etapas y pasos de este proceso.

Para realizarlo de manera efectiva es necesario prever todas las etapas y reconocer su interdependencia. Sintéticamente, estas etapas ilustran las partes de este manual. Se trata de elaborar los pasos necesarios para que el investigador social lleve adelante el proceso empírico para la corroboración de las teorías.

<sup>1</sup> Klein, L. *The scope and limitations of Econometrics*. Applied statistics. 1957. p.

<sup>2</sup> Kinnear y Taylor. *Investigación de Mercado, un enfoque aplicado*. Mc Graw Hill. 1993. p. 61

<sup>3</sup> Definición de la Real Academia Española.

	ETAPAS	PASOS	Clasificación	Tipos	Incorporadas al análisis ...	CONCEPTOS teóricos o prácticos, aplicados en esta etapa
1	Definición de la Investigación, Orientación en el campo de la Investigación y Formulación de un sistema de hipótesis	<ul style="list-style-type: none"> <li>✓ Documentación Descriptiva</li> <li>✓ Estudio de la Situación</li> <li>✓ Documentación Explicativa</li> <li>✓ Sistema de hipótesis para el estudio</li> </ul>	<ul style="list-style-type: none"> <li>teoría Estática</li> <li>teoría Dinámica</li> </ul>	Exacta o no exacta	... a través de funciones o ecuaciones	Teoría Económica, Modelos Económicos, Metodología de la Investigación, Economía matemática
2	Planteo de una tabla de datos y diseño de encuesta	<ul style="list-style-type: none"> <li>✓ Definición de las unidades de observación</li> <li>✓ Definición de las variables a observar</li> <li>✓ Definición de los Métodos</li> </ul>	<ul style="list-style-type: none"> <li>de corte transversal (o de espacio)</li> <li>de tiempo (o longitudinal)</li> <li>de panel</li> <li>cuantitativas</li> <li>cualitativas</li> <li>de observación</li> <li>de análisis estadístico</li> <li>de utilización estadística de los resultados</li> <li>de precisión que se desea tengan los resultados</li> </ul>	<ul style="list-style-type: none"> <li>Personas, Empresas, Instituciones, Regiones, Países, etc.</li> <li>Anual, Semestral, mensual, semanal, diario, etc.</li> <li>Combinación de corte transversal y de tiempo</li> <li>Aleatorias o determinista</li> </ul>	<ul style="list-style-type: none"> <li>... por selección Probabilística o casual</li> <li>... de manera Estocásticas o no estocásticas</li> </ul>	Teoría Estadística, Teoría Económica
3	Diseño de las fuentes de información	<ul style="list-style-type: none"> <li>✓ Ordenación de las fuentes de información</li> </ul>	<ul style="list-style-type: none"> <li>primarias</li> <li>secundarias</li> </ul>		... por definición del proceso de selección y de estimación	Inferencia Estadística. Muestreo
4	Recolección, procesamiento y organización de los datos	<ul style="list-style-type: none"> <li>✓ Diseño del trabajo de campo</li> <li>✓ Obtención de datos brutos</li> <li>✓ Codificación de la información</li> <li>✓ Organización de los datos</li> </ul>	<ul style="list-style-type: none"> <li>reales</li> <li>categóricos</li> <li>binarios</li> </ul>	<ul style="list-style-type: none"> <li>Discretos o continuos</li> <li>Códigos 0 ó 1</li> </ul>	... provenientes de fuentes de información secundarias o primarias	Diseño de recorrido territorial, Planillas de cálculo, Homogeneización de datos, Software econométrico
5	Análisis y presentación de la información	<ul style="list-style-type: none"> <li>✓ Ecuaciones o funciones</li> <li>✓ Modelos Econométricos</li> </ul>	<ul style="list-style-type: none"> <li>estocásticas o no estocásticas</li> <li>simples o múltiples</li> </ul>	<ul style="list-style-type: none"> <li>Individuales o conjuntas</li> <li>Dinámicos o Estáticos o de estática comparativa</li> </ul>	... para describir o predecir una situación o ambas cosas	Estadísticas Descriptivas Pruebas de hipótesis Regresión, Análisis Estadístico Multivariado. Modelos dinámicos Análisis de Series de Tiempo, Modelos Multiecuacionales, Modelos especiales.

**Cuadro 2.1. Etapas de una Investigación Econométrica**

Si la econometría es, fundamentalmente, operativa tenemos que acercarnos más a la realidad, introducirnos más en los pormenores de ella para conocer cuál es su metodología específica, o sea, cómo opera un investigador con econometría y qué debe conocer un econométrista desde el comienzo hasta el fin de su trabajo. Por *método* se entiende *el modo de decir o hacer con orden una cosa*<sup>4</sup>. Al respecto, dice Barbancho (1962): *aun cuando no es posible fijar, de manera concreta, las etapas por las que ha de discurrir un econométrista, entre otros motivos porque dependen de la clase de trabajo que se vaya a realizar*<sup>5</sup>, debemos mostrar de forma exhaustiva el proceso de investigación econométrica.

Pero, ¿cuál es el fundamento metodológico de este razonamiento?. Es decir, el econométrista actúa como un investigador y debe fundamentarse como tal. Para ello, Sampieri (2000) establece que el proceso de investigación está constituido por una serie de partes íntimamente relacionadas, entre ellas el marco teórico y el análisis desempeñan un papel fundamental<sup>6</sup>.

## 2.2. El marco teórico

El marco teórico inicia el proceso dando lugar a una problemática expresada a priori en la forma de un conjunto de proposiciones. Si éstas se presentan en forma aislada, el estudio adopta el carácter de descriptivo; mientras que, si están interrelacionadas, el estudio será explicativo.

En Ciencias Sociales la **Teoría** es una construcción lógica, es decir, un *sistema deductivo* que sólo debe cumplir con los requisitos lógicos – hipótesis y tesis – y, aunque se construye sobre la base de la experiencia, no debe ser sometida a pruebas de falsificación. En cambio, **Modelo** es una construcción lógica – empírica que debe cumplir con los requisitos lógicos – hipótesis y tesis – y con los empíricos caracterizados por las pruebas de validez.

Dagum y Dagum (1970) expresan que *la combinación de la Teoría con la Experiencia permite la formulación de un modelo cuyo contenido es en parte teórico y en parte empírico. La relación entre Teoría y Experiencia constituye el esquema de referencia en la construcción de los Modelos*<sup>7</sup>.

El marco teórico del econométrista está determinado por la teoría económica, más específicamente, por la teoría del problema que desea estudiar y que puede o no estar expresada a través de un modelo económico.

---

<sup>4</sup> Real Academia Española. Op. Cit.

<sup>5</sup> Barbancho, A. Op. Cit. P. 183

<sup>6</sup> Hernández Sampieri, R., Fernández Collado, C. Baptista Lucio, P.(1998). Metodología de la Investigación. McGraw Hill.

<sup>7</sup> Dagum, C. y Dagum, E.M.de. (1971). "Introducción a la Econometría". Editorial Siglo XXI. México.

La investigación econométrica comienza con la definición de la investigación que se hará a partir del conocimiento de ese marco teórico. La primera tarea del investigador es la decodificación de la realidad, y esa decodificación sólo puede ser realizada según una teoría implícita o explícita.

**Ejemplo 2.1.** Una teoría económica es el conjunto de proposiciones y definiciones extraídas de la realidad social y que explica los fenómenos económicos concretos. El modelo económico es construido a partir de términos generales, definiciones y supuestos de la teoría o de una porción de ella. Para poder realizar la prueba de validez de este modelo, el econometrista, señala cuáles son los problemas más significativos, las maneras cómo se seleccionarán los datos, la selección de diseño más correcto, así como la búsqueda de orden o de patrones entre ellos y la interpretación de los hallazgos de investigación.

El diseño de investigación se refiere al conjunto particular de métodos seleccionados por el investigador, tanto para la búsqueda de nuevos hechos como para la determinación de sus conexiones. Aquí se decide cómo se van a seleccionar los datos, cuáles serán los métodos analíticos, cómo se va a formular el problema, que tipos de instrumentos específicos se van a utilizar, o como se va a realizar el pretest.

En la fase empírica el investigador es guiado por la teoría y el modelo hacia los fenómenos concretos que contrastarán sus hipótesis teóricas. En la fase interpretativa se comparan los hechos con su teoría inicial, examinando las consecuencias que tienen para la teoría la comprobación o refutación de las hipótesis. Como señalamos, el proceso de investigación debe ser sistematizado en etapas.

## Distintos tipos de investigación

Debemos señalar que toda investigación econométrica comienza con algún tipo de interrogante que tratará de ser resuelto. Lo que expresamos a continuación muestra los distintos tipos de investigación que se pueden realizar. Es, para nosotros, una certeza de que, en su investigación, el econometrista debe emplear los estudios exploratorios, descriptivos, correlacionales y, fundamentalmente, explicativos.

La investigación científica tiene como objetivo teórico, más general, dar respuesta inteligible, confiable y válida a preguntas específicas o problemas de investigación. Las respuestas se dan por lo general en términos de qué, cómo, dónde, cuándo y porqué. Sin embargo, no toda la investigación tiene como propósito responder a todos los interrogantes existiendo la posibilidad de que se trate de responder solamente a alguno de ellos.

La formulación del **problema de investigación** es uno de los pasos principales y más difíciles de resolver en cualquier diseño de investigación. El tipo particular de estilo cognoscitivo de una investigación de carácter científico, exige del investigador no solamente claridad en la formulación del problema a investigar, sino también especificidad en

términos del tipo de respuesta que se busca a tipos específicos de preguntas.

Si en un comienzo los intereses del investigador pueden ser de carácter muy amplio, en términos de la investigación econométrica siempre hay que tener bien claro qué es lo que se está buscando, y el tipo de información que dará la respuesta a sus preguntas.

**Ejemplo 2.2.** El econometrista puede tener como área de interés general aspectos vinculados al modelo de mercado, e imponer como objetivo buscar la resolución de las condiciones de equilibrio parcial del mismo, algún interrogante sobre sus características, encontrar las variables que satisfarán esa condición del modelo, etc. Pero para los propósitos de la investigación, *el problema* tiene que ser formulado de manera más específica, buscando las respuestas a nivel de los cinco interrogantes formulados más arriba planteando preguntas tales como:

- ✓ ¿qué tipos de modelos teóricos han sido formulados?
- ✓ ¿qué tipo de variables fueron especificadas?
- ✓ ¿cómo están contruidos?
- ✓ ¿dónde fueron aplicados?
- ✓ ¿cuándo y quién lo formuló?
- ✓ ¿porqué se aplicó al estudio de la estática económica?

La bibliografía especializada acostumbra diferenciar los estudios o diseños de investigación en estudios exploratorios, descriptivos, correlacionales y explicativos<sup>8</sup>.

Los **estudios exploratorios** se efectúan cuando el objetivo es examinar un tema o problema de investigación poco estudiado o que no ha sido abordado antes. Es decir, cuando la revisión de la bibliografía reveló que únicamente hay guías no investigadas e ideas vagamente relacionadas con el problema de estudio.

**Ejemplo 2.3.** Si alguien desea investigar lo que opinan los habitantes de alguna región sobre un nuevo producto y cómo insertarlo en el mercado; revisa la literatura y encuentra que se han hecho muchos estudios similares pero en otros contextos (otras regiones del país o del extranjero). Estos estudios le servirán para ver cómo han abordado la situación de investigación y le sugerirán preguntas que puede hacer, sin embargo, el producto y el mercado son diferentes, la relación entre ambos es única.

Los estudios exploratorios son como realizar un viaje a un lugar que no conocemos, del cual no hemos visto ningún documental ni leído algún libro, sino simplemente alguien nos ha hecho un breve comentario sobre el lugar.

Los **estudios descriptivos** son más específicos y organizados que los estudios exploratorios, el interés está enfocado en las propiedades del objeto o de la situación. Se centran en medir con la mayor precisión posible y dan por resultado diagnósticos. Este tipo de estudios requiere

---

<sup>8</sup> Sampieri y . ob. Cit p.

un considerable conocimiento del área que se investiga para formular las preguntas específicas que busca responder.

**Ejemplo 2.4.** Un investigador econométrico puede pretender describir a los sectores industriales en términos de su producción, tecnología, capital, y trabajo. Entonces, mide esas variables en un número considerable de empresas para describir que tan automatizado está cada sector industrial (tecnología), cuánta es la diferenciación horizontal (producción), vertical (capital) y espacial (número de puestos de trabajo), y en qué medida pueden innovar o realizar cambios en los métodos de trabajo o maquinaria (capacidad de innovación).

Los **estudios correlacionales** tienen como propósito medir el grado de relación que exista entre dos o más conceptos o variables en un contexto particular.

**Ejemplo 2.5.**

- ✓ a mayor variedad de productos en el mercado, ¿corresponde mayor nivel de ventas, en un sector o empresa, en particular?;
- ✓ los productores rurales que adoptan más rápidamente una innovación tecnológica, ¿poseen mayor capital que los productores que la adoptan después?

Los **estudios explicativos** dan respuesta a los porqué. Están dirigidos a responder a las causas de los eventos físicos o sociales y van más allá de la descripción de conceptos o fenómenos o del establecimiento de relaciones entre conceptos. Su interés se centra en explicar porqué ocurre un fenómeno y en qué condiciones se da éste, o porqué dos o más variables están relacionadas.

**Ejemplo 2.6.** Dar a conocer la tecnología incorporada en un sector industrial, el capital empleado, los empleos generados y la producción realizada es un estudio descriptivo. Relacionar dichas variables con la producción realizada, con el nivel de actividad económica y con el nivel de consumo y precios (estudio correlacional) es diferente de señalar porqué la producción realizada por el sector industrial analizado responde con ciertos parámetros a la tecnología incorporada, al trabajo y al capital (estudio explicativo).

El sentido econométrico del análisis en investigación

Decíamos que en la investigación econométrica el marco teórico, conjuntamente con el análisis, desempeñaba un papel fundamental. Como una primera aproximación al método econométrico, podemos decir que para llevar a cabo una investigación económica empírica se requiere, en la mayoría de los casos, la utilización de modelos econométricos. Estos se asocian fundamentalmente con la etapa de análisis de la información del proceso de investigación. Pero para llegar a esta etapa se requiere al menos, a) exponer en forma precisa la información que se necesita en espacio y tiempo determinado; b) resumir las principales "operaciones de campo" que comporta todo proceso de observación, esto es, a partir de la selección de ciertos atributos de un conjunto de unidades de una población, aquellas operaciones consisten en medir esas características a partir de una muestra o de fuentes secundarias y resumirlas en una tabla de datos; c) la tabla de datos es una exposición completa, entonces, de las observaciones, temporales o espaciales, en fila y de las características o atributos observados de las mismas – que pueden ser variables cuantitativas o



cualitativas, cuyos valores se obtendrán de una muestra estadística o de fuentes secundarias aplicando un proceso de recolección y procesamiento de la información-; d) evaluar la asociación o relación entre las características seleccionadas y observadas, temporal o espacialmente, aplicando modelos econométricos.

La primera etapa mencionada en el párrafo anterior recibe, generalmente, el nombre de *Definición de la investigación* y contempla tanto el o los objetivos e hipótesis de investigación. Las fuentes de información para elaborar las hipótesis de investigación son, fundamentalmente, en una investigación económica, la teoría, la investigación exploratoria y la experiencia. En cuanto a la teoría habrá que considerar, principalmente, la teoría económica y la teoría estadística. De esta forma, se define el método de la econometría como aquél que tiene por finalidad corroborar hipótesis y teorías formuladas por la teoría económica. Esta verificación se hace con el auxilio, imprescindible, de las matemáticas y de la estadística.

Según esta concepción de los modelos económicos y de la econometría, los métodos econométricos utilizan las técnicas de la estadística matemática, fundamentada en la teoría de la probabilidad, para describir analíticamente una situación de la economía en espacio y tiempo determinado, para hacer pronósticos o analizar políticas probando o refutando empíricamente la validez de un modelo económico.

Se utiliza el término "modelo" conjuntamente con "econométrico" para distinguir entre modelo económico y modelo econométrico, que se diferencia de aquel en su formulación. Este último tiene en cuenta una parte sistemática – a veces, determinista – y una parte aleatoria, quedando especificado por la siguiente ecuación:

$$(2.1) \quad Y_t = \beta_1 + \beta_2 X_{2t} + \dots + \beta_k X_{kt} + \varepsilon_t; \quad t = 1, \dots, T$$

donde

$t = 1, \dots, T$ , indica que se trata de unidades de observación temporales en contraposición con las espaciales (también llamadas de corte transversal v.g.  $t = 1, \dots, n$ ).

$\beta_1 + \beta_2 X_{2t} + \dots + \beta_k X_{kt}$ ; es la parte sistemática del modelo econométrico (también llamada determinista en el caso de que los coeficientes  $\beta_j$ ,  $j = 1, \dots, k$ , no sean cambiantes o aleatorios).

$\varepsilon_t$ ; es la parte aleatoria del modelo econométrico.

$Y_t, X_{2t}, \dots, X_{kt}$ ; son las variables económicas o atributos de las unidades de observación entre las que se quiere establecer alguna relación proveniente de la teoría económica y que quedaron especificadas en la tabla de datos del proceso de investigación.

De esta manera, consideramos que toda investigación económica que quiera hacer uso de los métodos econométricos debe estar basada en una

teoría económica. No se concibe la "medición sin teoría"<sup>9</sup>. Por lo tanto, no se concibe la econometría fuera de la teoría económica (micro o macroeconómica). Es decir, la econometría, en tanto ciencia de la medida de los fenómenos económicos, provee elementos de análisis que, en su conjunto, hacen plausible la investigación econométrica, que, en palabras de Fossati (1958), *supera, por una parte, a la estadística económica porque ésta, por así decirlo, nos presenta medida sin teoría y, por otra, a la economía pura por cuanto se manifiesta como teoría sin medida*<sup>10</sup>.

### 2.3. Método de la econometría

Antes del desarrollo completo de las etapas de la investigación econométrica, mostradas en el cuadro 2.1, veremos que las mismas contienen (lo mostraremos en forma exhaustiva) fases que determinan el método econométrico. Con ello daremos cuenta cabal de la complejidad de este tipo de investigaciones<sup>11</sup>.

#### *Etapas 1: Definición de la Investigación, Orientación en el campo de la Investigación y Formulación de un sistema de hipótesis*

Fase 1.1) Partiendo del supuesto de que la formulación de una teoría económica cualquiera no corresponde al econométrico sino al economista teórico, lo primero que ha de hacer aquél es conocer dicha teoría. Es en este momento cuando debemos aplicar nuestro conocimiento de cómo formular un marco teórico.

Fase 1.2) Si la teoría no está expresada en lenguaje matemático, la segunda fase, de esta etapa, consistirá en proceder a ello, es decir en obtener el modelo económico.

Fase 1.3) Construcción del modelo, es decir, especificación de las relaciones de comportamiento, institucionales, legales, técnicas y contables, con *mención* expresa de las variables que, según la teoría, entran en cada relación y de su forma funcional.

Fase 1.4) En relación directa con la fase anterior, especificar las ecuaciones que relacionan a las variables con lo que quedará formulado un sistema de hipótesis. En esta misma fase se suele incorporar las tesis a las que se puede arribar de acuerdo a las hipótesis formuladas.

---

<sup>9</sup> Una excelente referencia para ampliar estos conceptos el lector la puede encontrar en Pulido, A. *Modelos econométricos*. Pirámide. 1993. p. 47 a 53

<sup>10</sup> Fossati. *Introduzione di calcolo statistico all' econometria*. 1958 p.160. Cit. Por Barranco. Op. Cit. P.188.

<sup>11</sup> Estructuramos la exposición siguiendo el esquema presentado por Barbancho. Op. Cit. P. 183 y Ss.

## Etapa 2: *Planteo de una tabla de datos y diseño de encuesta*

Fase 2.1) Enumeración de todas las variables relevantes para el fenómeno que pretende explicar la teoría. Detallando si son exógenas o endógenas, y el tipo de variables (cuantitativas o cualitativas), retardadas o no. En el caso de modelos multiecuacionales deberá, a su vez, analizarse el tipo de relación, interdependiente o causal, que liga a las variables endógenas (esto nos servirá para saber si el modelo es, respectivamente, interdependiente o recursivo).

Fase 2.2) Indicación expresa del tipo de observaciones que van a utilizarse, es decir si van a ser unidades temporales o espaciales. El modelo difiere en uno u otro caso.

Fase 2.3) En caso de que las fuentes de información, sobre las unidades de observación, sean de carácter primarias, elaboración del instrumento de recolección de datos.

## Etapa 3: *Diseño de las fuentes de información*

Fase 3.1) Identificación de las variables y los datos que van a ser observados con las fuentes de información.

Fase 3.2) Fijación del espacio geográfico al cual se van a referir las observaciones.

Fase 3.3) En el caso de datos temporales, determinación del tiempo que ha de tomarse como unidad de observación. Esto se hará de acuerdo a la teoría contenida en el modelo. Cabe señalar que el tomar como unidad años, trimestres o meses puede afectar a la calificación de las variables y a la eficacia de los retardos.

Fase 3.4) En el caso de datos temporales, determinación del período total de tiempo al cual se van a referir las observaciones. Por lo general, debe tenderse a elegir ciclos completos porque de no obrar así, los resultados dependerán de la fase del ciclo al cual corresponden los datos. En todo caso, se tendrá en cuenta la teoría contenida en el modelo.

Fase 3.5) En el caso de datos espaciales o de corte transversal, la elección del momento temporal suele venir casi siempre, mas o menos, "forzada". Pero debe tenerse presente, para interpretar correctamente los resultados, la coyuntura del tiempo al cual se refiere la encuesta. Las reacciones de los consumidores y de los empresarios no son las mismas, evidentemente, en todas las fases de un ciclo; esta particularidad es de especial relevancia para interpretar correctamente los resultados y también cuando se pretende aplicar dichos resultados a otras fases del ciclo.

Fase 3.6) En el caso de datos espaciales y de fuente primaria, establecer el tamaño adecuado de la muestra y el tipo de muestreo a utilizar de acuerdo a la distribución de las unidades de observación en la población.

Fase 3.7) En el caso del tratamiento conjunto de datos temporales y espaciales (datos de panel), elegir un criterio adecuado para obtener la suficiente cantidad de datos que haga posible ese tratamiento.

#### *Etapa 4: Recolección, procesamiento y organización de los datos*

Fase 4.1) En el caso de fuentes de información secundaria, organización de las mismas y procesamiento en una planilla de cálculo, con características adecuadas a una tabla de datos, que facilite su exportación a un software econométrico, con las unidades en las filas y las variables en las columnas.

Fase 4.2) En el caso de fuentes de información primaria, organización del relevamiento de las unidades de observación, supervisión de la recolección de los datos, organización y procesamiento en una planilla de cálculo, con las mismas recomendaciones que establecimos en la fase anterior.

Fase 4.3) Elección de un software econométrico adecuado para la importación de los datos procesados en las fases anteriores.

Fase 4.4) Realización de los estudios exploratorios, descriptivos y correlacionales, a partir de los datos procesados. Verificación del posible agrupamiento de unidades de observación, temporales o espaciales, estimación de las características poblacionales de las variables relevadas y, de corresponder, aplicación de la teoría de la decisión estadística.

Fase 4.5) En relación con la fase anterior, consideración de la conveniencia de eliminar de los datos observados, los influjos sistemáticos no recogidos por la teoría o de introducir una variable específica que los represente.

#### *Etapa 5: Análisis y presentación de la información*

Fase 5.1. Especificación del modelo econométrico. Transformar el modelo económico en modelo econométrico. Recordamos que un modelo econométrico es un modelo económico que contiene la especificación necesaria para su aplicación empírica.

Fase 5.2) En el caso de modelos multiecuacionales, identificación de los parámetros estructurales del modelo.

Fase 5.3) Estimación de los parámetros estructurales. Suponemos, en esta fase, que el modelo econométrico que se va a estimar ya es definitivo, esto es, que no se están realizando ensayos para probar, por ejemplo, la conveniencia de incluir o excluir variables en ciertas relaciones; no obstante, si esto aún fuera necesario habrá que realizar, previamente a la estimación definitiva, las pruebas estadísticas necesarias para evitar los errores de especificación. La estimación requiere de ciertas *especificaciones*, las cuales se relacionan a continuación junto a otras particularidades de interés.

- a) Si algunas de las relaciones no son lineales y esto dificulta o no hace posible la estimación, habrá que buscar un método aproximado para convertir en lineales aquellas relaciones.
- b) Elección del método de estimación más apropiado. En el caso de los modelos de ecuaciones múltiples, es una práctica bastante corriente el utilizar más de un método de estimación. El avance de la econometría teórica, también hace plausible esta situación en los modelos de una sola ecuación.
- c) Formulación de las hipótesis estadísticas necesarias en relación con las perturbaciones aleatorias y sobre la parte sistemática del modelo, para poder proceder a la estimación del modelo.
- d) Determinación del error estándar de los parámetros estimados.

Fase 5.4) Verificación de la bondad de ajuste del modelo y la significación individual y conjunta de los parámetros estructurales.

Fase 5.5) Verificación de las hipótesis sobre la parte sistemática del modelo: multicolinealidad, cambio estructural, errores de especificación.

Fase 5.6) Verificación de las hipótesis estadísticas formuladas sobre las perturbaciones aleatorias: media nula, no autocorrelación y homocedasticidad.

Fase 5.7) Verificación de las hipótesis económicas del modelo:

- a) En cuanto a la especificación del modelo, mediante el coeficiente de correlación múltiple para cada relación.
- b) En cuanto a la validez del modelo para explicar el fenómeno, mediante su poder predictivo.
- c) En cuanto a ciertos parámetros estructurales, mediante los contrastes de restricciones lineales adecuadas a la teoría.

Fase 5.8) Interpretación y análisis de los resultados en relación con la teoría económica y el grado de agregación de las variables;

propiedades dinámicas del modelo, cuando es de este tipo y, por último, comparación con otros estudios análogos.

Fase 5.9) Formulación de predicciones.

Fase 5.10) Comunicación al economista teórico de los resultados y posiblemente, de la necesidad de una reparación de la teoría, si ésta no resulta satisfactoria.

Por lo visto, establecer un método significa proporcionar una idea acerca de cómo llevar a cabo una investigación aplicada a la economía, considerando la utilización de modelos econométricos.

Al decir de Barbancho<sup>12</sup>, *“como en la mayor parte de los casos ocurre, también para la Econometría es difícil dar un esquema, que sea generalmente válido, en la investigación. Más, a pesar de ello, existen unos determinados estadios funcionales por los que necesariamente tiene que pasar toda investigación econométrica para que cumpla los requisitos de un proceso cuantitativo. Dichos estadios son de naturaleza teórico-económica, por una parte, y estadística, por otra. El orden por el que hay que pasar por ellos no es fijo, ya que depende de las condiciones particulares de cada caso. Frecuentemente, se pasa varias veces de uno a otro campo, es decir de la Teoría económica a la Estadística, como una especie de juego de “vaivén” o, utilizando el símil de Timbergen, como si se tratara de un partido de tenis entre economistas puros y econométristas”*.

De esta forma, el econométrista prueba la validez de un modelo económico y, si es incorrecto, lo devuelve al teórico para que la reformule. Este movimiento entre los especialistas se produce hasta que se acepta la validez.

El desarrollo del conocimiento de las ciencias, tanto experimentales como no experimentales, dado su origen empírico, en general, se ajusta a un método que sigue las siguientes etapas:

1. definición del problema de la teoría económica
2. planteamiento de una tabla de datos
3. diseño de fuentes de datos
4. recolección y procesamiento de los datos
5. análisis econométrico

El análisis detallado de cada una de estas etapas arrojará luz sobre el proceso empírico de construcción de los modelos econométricos, incluyendo sus características de generalidad y validez y la elaboración de las teorías económicas. Se puede decir que se sigue *un proceso lógico – empírico en la construcción de los modelos econométricos*<sup>13</sup>.

---

<sup>12</sup> Barbancho. Op.cit.

<sup>13</sup> Dagum-Dagum. Op.cit.

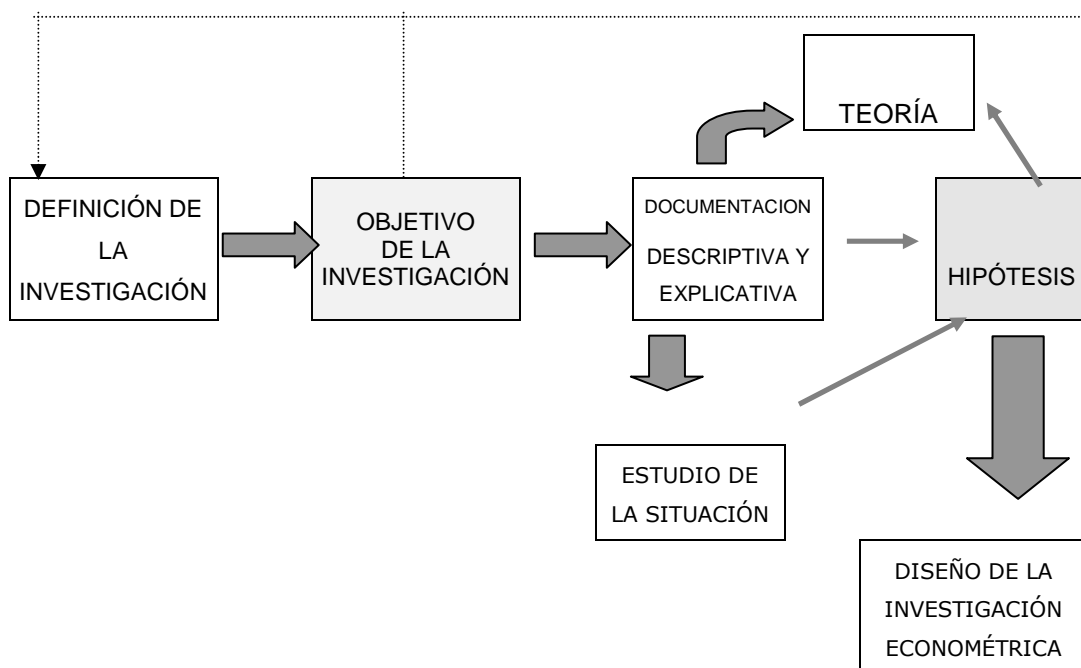
Lo anterior debe tenerse en cuenta siempre. Incorporar esas fases a las etapas del proceso de investigación econométrica es de suma importancia si queremos realizar buenas corroboraciones empíricas. A continuación, en este capítulo, estudiaremos qué se debe tener en cuenta a los efectos de realizar esa incorporación en la etapa de definición de la investigación. En los capítulos siguientes abordaremos las otras etapas e incorporaremos las fases del método econométrico a las mismas.

## 2.4. Definición de la investigación econométrica

Podemos esquematizar, como se muestra en la figura 2.1, el problema de la definición de la investigación, recordando que el primer paso que debe dar el investigador es tener una sólida orientación en el campo que va a investigar. Esta orientación se refiere a las elaboraciones abstractas de teoría, a los resultados de investigaciones y a las particulares circunstancias concretas que constituyen el objeto o situación a investigar. La definición de la investigación es una exposición lo más precisa posible de la información que se necesita.

### Objetivo, Hipótesis y Tesis

El **objetivo** de la investigación pregunta qué información se requiere de acuerdo a la definición de la investigación planteada. Señala los elementos en el modelo que van a ser investigados, incluso puede ser la selección de un modelo. Para ello, habrá que observar la realidad; agrupar las observaciones; realizar un análisis ex-ante; especificar un modelo explicativo; realizar un análisis ex-post; reespecificar el modelo propuesto y, por último, comprobar la utilidad práctica del modelo.



**Figura 2.1. Esquema de definición de la investigación**

Las **hipótesis** son juicios de carácter conjetural y su función es sugerir nuevos experimentos o nuevas observaciones. Indica lo que estamos buscando o tratando de probar y pueden definirse como explicaciones tentativas del fenómeno estudiado. Es una respuesta posible a los objetivos de una investigación econométrica que, para ser desarrollada, utiliza la teoría estadística y económica, la investigación exploratoria sobre el tema de interés particular y la experiencia anterior del investigador con problemas similares.

La *hipótesis* es el axioma o conjunto de axiomas o proposiciones iniciales referente a las relaciones de comportamiento de los sujetos de la actividad económica, teniendo en cuenta las relaciones institucionales y tecnológicas vigentes. Equivale a describir las causas del fenómeno bajo estudio.

Las **Tesis o proposiciones finales** obtenidas son lógicamente consistentes con los postulados o proposiciones iniciales del sistema.

Las Tesis son las proposiciones finales o conclusiones referentes al comportamiento de los sujetos de la actividad económica deducidas a partir de las hipótesis o proposiciones iniciales que posean las propiedades de consistencia e independencia.

Las tesis obtenidas deben ser legítimamente consistentes con el conjunto de las hipótesis, o sea debe haber una afirmación única de verdad o falsedad. Un modelo generalmente tiene más de una tesis o sea más de una conclusión.

Por otro lado, en economía, se puede presentar un orden jerárquico en la construcción de modelos, donde los supuestos iniciales de uno son conclusiones o tesis de otros modelos de orden superior.

La parte empírica de la construcción de los modelos exige su contraste con la experiencia a fin de tener una medida de su realidad, esto es, el grado de representatividad de los mismos y, por lo tanto, del alcance de sus aplicaciones empíricas.

Las hipótesis y tesis de un modelo se contrastan con la experiencia en términos de probabilidad. La probabilidad de que las hipótesis y tesis no sean contradichas por la experiencia aumenta con el número de pruebas que la corroboran y la diversidad entre ellas.

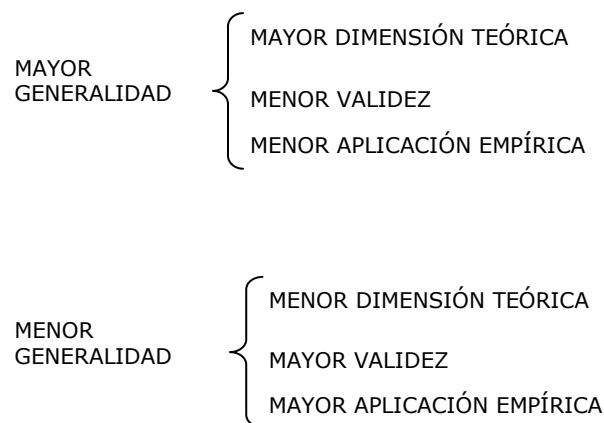
Dos aspectos fundamentales hay que considerar en cuanto al contraste de los modelos con la experiencia: la generalidad y la validez.

La generalidad supone reducir el grado de especificación de las hipótesis con respecto a la conducta "real" de los sujetos de la actividad económica en tiempo y espacio determinado. Cuando más general sea el modelo más probabilidades tiene de aplicación empírica, pero pierde validez en cuanto a sus conclusiones.

Por el contrario, cuanto más especificaciones se agreguen a las hipótesis se pierde generalidad y se gana en validez.



Se entiende por validez el grado en que las conclusiones o tesis de un modelo explican la conducta "real" de los sujetos de la actividad económica en tiempo y espacio determinado.



**Figura 2.2. Relaciones entre generalidad y validez**

En la construcción de los modelos se plantea, como problema crucial, el de la generalización del conocimiento "contra" la especialización del mismo. Es decir, cómo agregar sustancias a las hipótesis que hagan más válido el modelo sin sacrificar en gran medida su dimensión teórica.

Los siguientes criterios son válidos para el avance simultáneo de generalidad y validez

- formación humanista del economista;
- formación matemática y estadística del economista;
- el desarrollo de las matemáticas sociales (esto es, las necesarias para la solución de los problemas planteados por la ciencia económica);
- el desarrollo de los métodos estadísticos aplicables a las ciencias sociales;
- perfeccionamiento de sistemas computados electrónicamente;
- series estadísticas abundantes y precisas.

Una vez establecidas las hipótesis y tesis viene la etapa de diseño de la investigación que contempla cuestiones sobre la población (espacio y tiempo) a estudiar y origina las siguientes etapas de una investigación econométrica.

Fuentes de información para desarrollar la hipótesis

#### *Documentación descriptiva*

La literatura actual y los documentos históricos de información pueden dar luz a los problemas a investigar, sobre todo en relación a los aspectos y peculiaridades concretas. Una revisión de informaciones de prensa,

diarios, radio y televisión pueden resultar de mucha utilidad cuando se trata de hacer un análisis de contenido en investigaciones donde el objetivo es la medición de actitudes u opiniones, por ejemplo.

La consulta de archivos públicos y de documentos oficiales es también útil. Si existe el interés o la necesidad de extraer de ellos algunos datos el trabajo puede ser más arduo porque depende de cómo esté organizado el archivo.

El estudio puede recurrir a las informaciones disponibles en archivos privados, en los cuales hasta las notas escritas a mano pueden ser usados. Debe tenerse especial cuidado con el material estadístico disponible, el cual debe ser estudiado detenidamente, ya que la utilidad de este material depende de la manera en que fue obtenido y calculado.

### *Estudio de la situación*

Si el área de investigación es totalmente desconocida para el investigador es recomendable un primer contacto como observador participante.

**Ejemplo 2.7.** El investigador debe vivir con la gente que quiere estudiar, tomar un trabajo en una fábrica del mismo tipo que la que va a estudiar o actuar como espectador en un determinado ambiente.

Como complemento a los contactos como observador participante, es de gran importancia llevar a cabo entrevistas no estructuradas con "personas claves". En estas entrevistas hay que buscar que el sujeto entrevistado esté en su ambiente natural, de tal modo que pueda contar sus deseos y opiniones con referencia a los problemas en los cuales el investigador esté actuando de la manera más natural posible.

**Ejemplo 2.8.** La persona clave puede ser alguien que ocupe una posición destacada dentro de una industria, el presidente de un club, periodistas especializados, políticos.

La idea es utilizar las entrevistas para hacer un cuadro de situación y de los problemas involucrados a través de lo que diferentes personas piensan o actúan con respecto a estos problemas y de las actitudes que podrían tomar.

Estas entrevistas deben ser realizadas por el mismo investigador. El número de entrevistas varía de acuerdo al tipo de investigación y es conveniente detallar preguntas para fijar algunas áreas que se desee cubrir. A veces de una pregunta se desprende un área desconocida que es conveniente desarrollar.

Con estas entrevistas el aspecto cualitativo entra en la investigación. Luego de la exploración es posible comenzar a formalizar la hipótesis o incluso cambiar todo el carácter de la investigación. El proceso que sigue todo el primer paso de orientación de una materia y la formalización de una hipótesis se concentra en la recolección de datos a través de un cuestionario y el análisis estadístico que da a la investigación su carácter principal de ser cuantitativa.

### *Documentación explicativa*

Se trata de analizar qué han expresado sobre nuestro tema de interés otros autores, cómo han afrontado y formulado el problema, cómo lo han resuelto y a qué conclusiones han llegado, cómo han definido sus conceptos, como han determinado sus observaciones

Las entrevistas cualitativas son el primer intento para integrar la documentación descriptiva y la conceptualización, más o menos generalizada, a la situación concreta de la investigación particular. A este nivel el investigador comienza a definir sus preguntas más específicamente así como a la formulación de sus primeras hipótesis.

Es muy importante consultar la literatura explicativa o teórica porque permite insertar la investigación en un marco de referencia teórico más general.

Puede suceder que el investigador quiera replicar un estudio ya realizado con anterioridad o aplicarlo en otra área de la realidad, esto no debe interpretarse como plagio, pues es la manera en que la ciencia opera con mayor frecuencia con objeto de avanzar y ampliar el campo de la teoría.

### *Teoría*

Básicamente, la formulación de un sistema de hipótesis en la investigación econométrica descansa en la metodología de la investigación, la teoría económica, la teoría estadística y la teoría matemática.

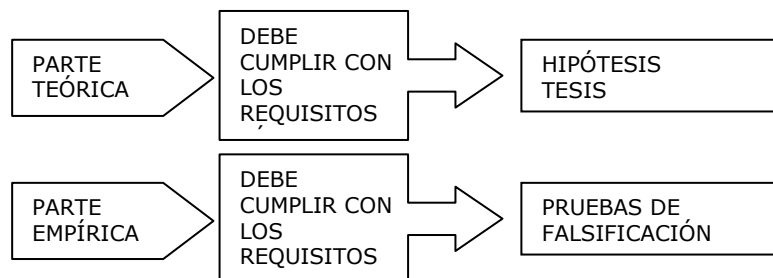
- a) Es interesante tener presente lo que los científicos, en general, dicen respecto de la teoría. En este sentido, una **teoría** es un conjunto de conceptos, definiciones y proposiciones relacionadas entre sí, que presentan un punto de vista sistemático de fenómenos especificando relaciones de variables, con el objeto de explicar y predecir los fenómenos. Los **conceptos** son abstracciones formuladas a partir de generalizaciones de observaciones particulares que son definidas nominalmente y que proporcionan el nivel de significado. Los conceptos describen los fenómenos y pueden ser: *conceptos categóricos*, que son complejos y se miden al nivel de categorías nominales, y las *variables*, que representan dimensiones de los fenómenos admitiendo grados de variación que se miden a niveles ordinales, intervalares o racionales. El concepto es un nombre al que hay que agregarle una **definición**. Cuando se ordenan conceptos y definiciones se obtienen esquemas descriptivos que sirven para clasificar o diagnosticar la realidad. Las definiciones son una forma de explicar, cuando se conectan conceptos tenemos un juicio teórico. Por **proposición** se entiende cualquier generalización que puede probarse como consistente o inconsistente con respecto a otras generalizaciones que forman

parte del cuerpo organizado de conocimiento; las proposiciones científicas deben ser sometidas a verificación empírica.

En ciencias sociales la Teoría es una construcción lógica, es decir, un *sistema deductivo* que sólo debe cumplir con los requisitos lógicos (hipótesis y tesis) y, aunque se construye sobre la base de la experiencia, no debe ser sometida a pruebas de falsificación. En cambio, modelo es una construcción lógica – empírica que debe cumplir con los requisitos lógicos (hipótesis y tesis) y con los empíricos caracterizados por las pruebas de validez.

La *combinación* de la teoría con la experiencia permite la formulación de un modelo cuyo contenido es en parte teórico y en parte empírico. La *relación* entre teoría y experiencia constituye el esquema de referencia en la construcción de los modelos, que se muestra en la figura 2.2.

Los requisitos lógicos hacen a la construcción ordenada y coherente de una teoría o de la parte teórica de un modelo, respondiendo al esquema verdadero o falso de la lógica formal, pero independientemente de la validez empírica que puedan tener.



**Figura 2.3. Esquema de referencia en la construcción de modelos**

La ciencia económica elabora sus teorías y modelos a partir de las *observaciones empíricas y no experimentales* de los sujetos de la actividad económica, que presentan características de permanencia y regularidad.

- b) Si el objetivo de la Econometría es la verificación de las hipótesis económicas no es necesario probar que la investigación que realizamos está en íntima relación con la teoría económica (micro y macroeconómica).

Al respecto, queremos coincidir con Samuelson y Nordhaus (1999), la microeconomía es la *rama de la economía que se ocupa actualmente de la conducta de entidades individuales como los mercados, las empresas y los hogares*; en este sentido los métodos

y modelos econométricos son aplicados en la administración de negocios *ya que los actos interesados de los individuos generan un beneficio económico*. Estos mismos autores definen a la macroeconomía como la rama *que se ocupa del funcionamiento general de la economía – ciencia que se ocupa del estudio de la manera en que las sociedades utilizan los recursos escasos para producir mercancías valiosas y distribuirlas entre los diferentes individuos* -. En este sentido por *funcionamiento general de la economía* entendemos tanto los espacios sociales (*sociedades*) regionales, como nacionales y supranacionales. *Hubo un tiempo en que la frontera entre ambas era muy nítida; últimamente las dos subdisciplinas se han fusionado al aplicar los economistas los instrumentos de la microeconomía a cuestiones como el desempleo y la inflación*<sup>14</sup>

En la investigación econométrica, lo primero que deberá hacer el econometrista es conocer la teoría económica formulada por el economista teórico. Deberá saber si la teoría formulada es dinámica o estática. Por teoría económica dinámica se entiende la que tiene en cuenta o explica los cambios en los valores de las variables endógenas a medida que transcurre el tiempo, aún cuando no haya modificaciones en la estructura económica o en las variables exógenas, excepto el tiempo. Mientras que, por estática económica (o análisis de equilibrio en la economía) entendemos al análisis teórico de problemas en los que el tiempo no interviene explícitamente como variable, ni en la forma de variables desfasadas o tasas de cambio. La estática maneja las situaciones de equilibrio, es decir, las que se mantienen una vez alcanzadas. Veremos más adelante ejemplos de esta clasificación de la teoría.

Conviene, sin embargo, observar que esta relación no es del mismo tipo que la existente entre la Econometría y sus otros dos pedestales, ya que las Estadística y las Matemáticas, no son, en cierto modo más que un instrumento y un lenguaje, en cambio la teoría económica viene a ser la fuente o el manantial que suministra teorías para ser refutadas.

- c) La econometría no puede existir sin la teoría estadística. Esto es así, simplemente, porque la econometría no es más que la estadística especialmente adaptada a la investigación económica. Los orígenes de la econometría tenemos que buscarlos en la denominada Estadística Económica, o sea cuando se buscaba la "medición sin teoría". Recordemos que el análisis de regresión y de correlación fueron las técnicas utilizadas por los precursores de la econometría.

El cálculo de probabilidades, el análisis multivariante, los procesos estocásticos, la teoría de la estimación, la de contrastación de hipótesis (teoría de la decisión estadística), y la de predicción, incluyendo, por supuesto, las técnicas de muestreo, constituyen la base de formación imprescindible de todo econometrista.

<sup>14</sup> Samuelson y Nordhaus. *Economía*. Mc Graw Hill. 1999. ps. 4 y 5.

Sin embargo, la econometría y por lo tanto el economista, debe circunscribirse, desde el punto de vista estadístico a los temas más específicos que, dentro de la investigación econométrica, adquieren una fisonomía particular. Por otro lado, así como la econometría se relaciona con la teoría económica, fundamentalmente porque sirve para corroborar las teorías, se relaciona con la estadística en sentido similar; ya que, a lo largo de su historia ha servido para enriquecerla, como es el caso de regresión ortogonal o la estimación de ecuaciones interdependientes, no incluidos en tratados generales de estadística.

- d) Las matemáticas, en realidad, vienen a ser dentro de la econometría una especie de medio unificador de la teoría económica y de la estadística. La teoría económica puede formularse de una manera literaria. El mejor ejemplo, es la teoría Keynesiana formulada sin el uso de las matemáticas; posteriormente, Hicks y otros teóricos la formularon matemáticamente.

Sin el lenguaje de las matemáticas el economista, y por tanto el economista sin el lenguaje de las matemáticas, no puede llevar a cabo su trabajo; aunque, se deber tener presente, que en todo momento el problema económico debe ser el rector de la investigación econométrica. Esto se analizará en profundidad en el punto siguiente.

## 2.5 Clasificación de modelos

Dagum, C. y Dagum, E.M.de. (1971). "Introducción a la Econometría". Editorial Siglo XXI. México.

CHIANG, ALPHA. (2006). "Métodos Fundamentales de Economía Matemática". Editorial McGraw Hill. México.

## 2.6. Modelos económicos y modelos econométricos

Una gran parte de los esfuerzos de los economistas ha consistido en elaborar modelos genéricos que sean aplicables con validez y generalidad a diversos sistemas concretos; a este tipo de modelos, expuestos en forma matemática, se los denomina modelos económicos. Pero como un modelo económico es demasiado simplificado y excesivamente general como para recoger todos los aspectos de un sistema, en un espacio y tiempo determinado, se han desarrollado los modelos econométricos, que se caracterizan por ser específicos para su aplicación a sistemas reales, concretos y que generalmente se basan en un modelo económico más o menos formalizado.

Sobre la base de los comentarios anteriores se pueden establecer las siguientes diferencias entre un modelo económico y un modelo econométrico:

<i>DIFERENCIAS</i>	<b>MODELO ECONÓMICO</b>	<b>MODELO ECONOMÉTRICO</b>
<b>Espacio y tiempo</b>	Generalidad, sin definición	<i>Definido, concreto a un sistema real</i>
<b>Relaciones funcionales</b>	Exactas o deterministas	<i>Aleatorias (no deterministas)</i>
<b>Funciones</b>	Con pocos o algunos requisitos, generalmente sin explicitar. $Y = f(x, z)$	<i>Definida, <math>Y = X^\alpha + Z^\beta + \mu</math></i>
<b>Otras diferencias</b>	<ul style="list-style-type: none"> <li>● Puede no incluir una variable</li> <li>● Puede incluir una variable implícita</li> <li>● Considera variables teóricas con posible constancia</li> </ul>	<ul style="list-style-type: none"> <li>● <i>Incluye la variable ante situaciones determinadas</i></li> <li>● <i>No lo hace</i></li> <li>● <i>No la puede considerar de esa manera</i></li> </ul>

**Cuadro 2.2. Diferencias entre modelo económico y econométrico**

Por lo tanto, la Econometría se ocupa de estudiar estructuras (o modelos) que permitan explicar el comportamiento o propiedades de una variable económica utilizando como causas explicativas otra u otras variables económicas. Por ejemplo, explicar el comportamiento de la inflación tomando como variables explicativas la oferta monetaria y el nivel de consumo agregado.

En resumen el análisis econométrico considerará:

- a) la especificación de la estructura: MODELO ECONOMETRICO
- b) el análisis de las propiedades estadísticas del modelo
- c) su estimación
- d) su utilización con fines predictivos o para el análisis de cuestiones de política económica de índole micro o macroeconómica.

A veces, es necesario utilizar modelos multiecuacionales en donde una variable a explicar en una ecuación pasa a ser explicativa en otra ecuación del mismo modelo.

Por lo tanto, un modelo econométrico puede asumir diferentes formas según características propias, número de variables, número de ecuaciones, forma funcional, etc.

**Ejemplo 2.11.** Sea la función de consumo

$$C_t = \alpha + \beta Y_t + \mu_t; \quad t = 1, \dots, 100$$

En este modelo se pretende explicar la evolución temporal de los gastos en consumo por medio de una variable que determine el nivel de renta.

De acuerdo a esta especificación, en cada período se debería haber consumido una proporción de la renta, medida por  $\beta Y_t$ ; la diferencia entre ambas cifras se supone constante en el tiempo ( $\alpha$ ).

Este modelo de consumo se puede utilizar a tres niveles distintos:

A nivel agregado, en cuyo caso las variables  $C_t$  e  $Y_t$  serán indicadores del nivel de consumo y la renta agregados. Para este análisis se requieren observaciones numéricas de las variables durante un periodo de tiempo  $t = 1, \dots, 100$ . Por lo tanto, las observaciones correspondientes a cada una de las variables es una serie temporal

T	PERÍODO	$C_t$	$Y_t$
1	1974.I	130	1400
2	.II	133	1420
3	.III	137	1460
⋮	⋮	⋮	⋮
⋮	⋮	⋮	⋮
⋮	⋮	⋮	⋮
100	1998.III	421	2175

A nivel desagregado, por ejemplo relacionando los gastos en un cuatrimestre en consumo y los ingresos de las familias. En este caso:

$$C_i = \alpha + \beta Y_i + \mu_i; \quad i = 1, \dots, 500$$

En donde los subíndices indican que cada observación corresponde a una familia distinta y no a un periodo de tiempo diferente. Por lo tanto, las observaciones correspondientes a cada una de las variables es un dato obtenido de una muestra de un conjunto de familias y se denominan datos de sección cruzada.

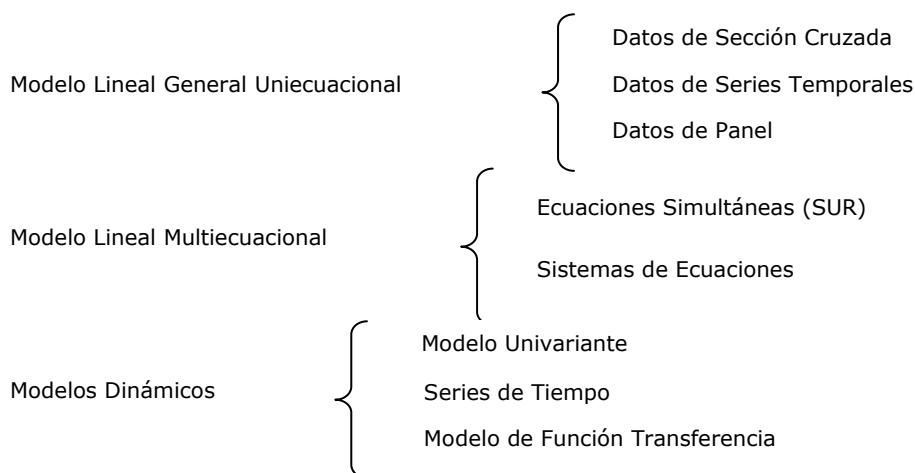
MUESTRA	$C_t$	$Y_t$
Familia 1	12	2500
Familia 2	16,5	2800
Familia 3	18	2000
⋮	⋮	⋮
⋮	⋮	⋮
⋮	⋮	⋮
Familia 500	75	33000

A nivel agregado temporal y desagregado por familia, esto es una combinación de observaciones a través de una muestra de individuos en el tiempo, que se denomina datos de panel.

T	PERÍODO	$C_t$	$Y_t$
1	1974.I	130	1400
	Familia 1	10	20
	Familia 2	35	55
	⋮	⋮	⋮
	Familia 500	9	33
2	.II	133	1420
	Familia 1		
	Familia 2		
	⋮		
	Familia 500		
⋮	⋮	⋮	⋮
⋮	⋮	⋮	⋮
100	1998.III	421	2175
	Familia 1		
	Familia 2		
	⋮		
	Familia 500		



La siguiente clasificación de los modelos econométricos se hace necesaria para su posterior especificación:

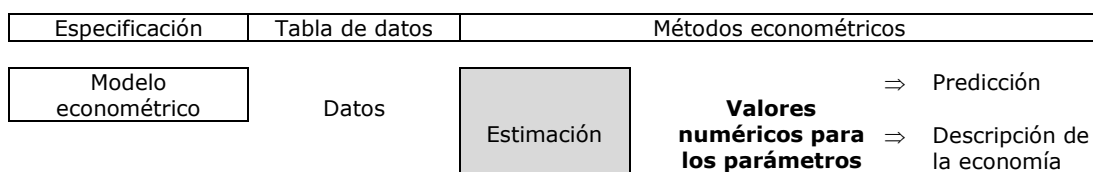


**Figura 2. 4. Tipología de modelos econométricos**

Una vez especificados, los modelos deben ser estimados a partir de datos muestrales o de fuentes secundarias. Este proceso de estimación se complementa con el proceso de inferencia, para posteriormente poder realizar predicciones, esto es valores futuros de las variables del modelo.

Dentro del análisis econométrico de una determinada cuestión económica se responden tres interrogantes, de los cuales los dos primeros ya se han analizado: Qué modelo especificar?; Qué tipo de datos utilizar?; Qué valores se asignan a los parámetros del modelo?

Esto es,



**Cuadro 2.3. Esquema del análisis econométrico**

## 2.7. Modelos Econométricos y contrastes de Teorías Económicas

Puede parecer, en torno a lo que estudiamos más arriba, que siempre resulta sencillo corroborar una teoría económica realizando una investigación econométrica. Al menos inicialmente, parecería ser que con buenos datos y una adecuada especificación econométrica de la teoría resultaría suficiente. Sin embargo, existe razones más que suficientes

para que el proceso no resulte tan automático. Entre ellas, de acuerdo a Pulido (1993), podemos citar<sup>15</sup>:

- a) *Las teorías pueden exigir una adaptación previa a su contraste.* En general, mientras que las Leyes de bajo nivel son fácilmente contrastables, las de alto nivel exigen generalmente una adaptación previa. Por ejemplo, el mecanismo del acelerador en la demanda de inversión puede resultar fácilmente contrastable, mientras que las teorías económicas referidas al proceso dinámico del equilibrio pueden exigir una adaptación previa de los modelos matemáticos en diferencias resultantes.
- b) *Las teorías económicas no son leyes universales.* Mientras que las teorías de las ciencias formales (lógica y matemática) y de las ciencias naturales (física, química, biología, etc.) pueden pretender su validez sin condiciones tempo-espaciales, las ciencias sociales tratan de explicar no solamente un sistema complejo en exceso, sino además en constante mutación. Por tanto, la contrastación de una teoría económica queda limitada al marco de referencia (explícito o no para el que ha sido concebida).
- c) *La confrontación con los datos puede realizarse con modelos econométricos diferentes.* Ya hemos indicado anteriormente que cada investigador puede tener su propio modelo del sistema. De hecho, es posible especificar varios modelos econométricos, diferentes todos ellos, con base en una misma teoría o modelo económico. La elección de las variables consideradas como más relevantes; su introducción en valores absolutos diferentes, porcentajes o índices; los retardos temporales en las influencias, etc., son algunas de las razones de posibles discrepancias entre modelos econométricos referidos al mismo fenómeno y sobre la base de un planteamiento teórico común.
- d) *La confrontación puede realizarse con datos temporal y espacialmente diferentes.* Cuando un modelo econométrico ha sido diseñado, su aplicación se hace en un ámbito dado. Lo habitual es que cada investigador aplique su modelo a su país, región, producto o empresa. Con ello, la contrastación empírica de que se dispone con respecto a una teoría suele ser excesivamente parcial e incompleta como para dar una idea definitiva sobre su validez.
- e) *Los resultados del contraste serán siempre en términos de probabilidad.* Incluso dentro de un mismo país estamos acostumbrados al establecimiento de controversias entre investigadores sobre la base no ya de una teoría común, sino incluso con el mismo modelo econométrico, aunque con diferentes fuentes de datos y métodos de estimación. La

---

<sup>15</sup> Pulido, A. op. Cit. P.53.

inseguridad de las series numéricas de base, la limitación muchas veces arbitraria de la muestra, la dificultad de contrastación de ciertas hipótesis estadísticas básicas, entre otras, hacen que los resultados de un modelo econométrico sólo puedan ser entendidos en términos de una cierta probabilidad de ocurrencia. Como dicen, Dagum y Dagum (1971), *no podemos afirmar de un modelo que sea verdadero o falso en un sentido absoluto, sino más probable o menos probable en el sentido de la lógica probabilística*<sup>16</sup>.

- f) *La evidencia empírica permite refutar pero no verificar.* Las hipótesis y teorías que resultan acordes con los resultados de un contraste observacional, serán confirmadas, pero en ningún caso verificadas lógicamente. En palabras de Papandreu (1961): *si las predicciones incluidas en las hipótesis no son refutadas por la evidencia empírica el teórico puede adoptarlas, pero sólo hipotéticamente, pues siempre son susceptibles de refutación por nueva evidencia empírica. Se ha hecho habitual calificar tales teoremas o hipótesis de "operativamente significativos"*<sup>17</sup>.
- g) *El proceso "elaboración de teoría-contraste de resultados" debe ser iterativo y no terminal.* Consideramos que un diálogo fructífero teoría-medición sólo puede hacerse sobre la base de una contrastación en diferentes fases de elaboración teórica y no una vez que ésta se considere como terminada, aunque sea parcialmente. En la polémica metodológica sobre si una teoría debe contrastarse a partir de sus hipótesis o por la validez de sus resultados, nos inclinamos más hacia la primera aun reconociendo que puede haber ciertas tesis en el proceso de abstracción, que la teoría exige, que no son directamente contrastables; pero siempre que este contraste sea factible, creemos que resulta aconsejable incorporarlo al proceso creativo. Terciando en la polémica de celebres economistas, Leontief (1971) expresa al respecto que: *un verdadero avance puede únicamente conseguirse a través de un proceso iterativo en el que la formulación teórica mejorada plante nuevas cuestiones empíricas y las respuestas a estas cuestiones, a su vez, conduzcan a nuevas ideas teóricas. Los datos de hoy se convierten en incógnitas que deberán explicarse mañana. Esto, incidentalmente, hace insostenible la posición metodológica a veces admitida, de acuerdo con la cual un teórico no necesita verificar directamente las hipótesis factuales elegidas como base de sus argumentos deductivos, con tal de que sus conclusiones empíricas parezcan ser correctas. La prevalencia de tal punto de vista es, en gran parte, responsable del glorioso*

<sup>16</sup> Dagum, C. y Bee de Dagum, E. op. Cit. P.

<sup>17</sup> Pulido. Op.cit. pág.54

*estado de aislamiento en el que nuestra disciplina hoy en día se encuentra*<sup>18</sup>

En síntesis, los modelos econométricos no pueden, por sí solos, ni crear teoría económica ni tan siquiera confirmarla o refutarla definitivamente; su ya importante misión se limita a señalar ciertos caminos de investigación que parecen, en ese momento, más seguros.

## CASOS DE ESTUDIO, PREGUNTAS Y PROBLEMAS

Caso 2.1: ¿Qué investigación econométrica te gustaría realizar?

- a) Te proponemos que pienses un tema y que formules las preguntas a las que quieras encontrarle una respuesta.
- b) Describe la información que necesitas para desarrollar tu investigación, explicita el objetivo perseguido con la misma, menciona las hipótesis con las que vas a trabajar y la fuente de información para desarrollar la hipótesis. Recuerda que puedes recurrir a más de una fuente.

### Problemas

2.1. Comenta, luego de revisar la teoría económica, el siguiente modelo de Jorgenson. Luego propone un modelo econométrico para la verificación del mismo.

El autor ha desarrollado un modelo económico en que la empresa vende toda la cantidad  $Q_t$  que produce el período  $t$  a un precio  $p$ , usando los insumos necesarios, capital y trabajo, y sólo limitada por la tecnología de la empresa.

Es decir, este modelo está basado en la teoría neoclásica en su versión tradicional de competencia perfecta.

Este modelo parte de una función de producción,

$$Q_t = f(L_t; K_t)$$

y demuestra que la optimización de los ingresos netos exige un stock de capital, tal que

<sup>18</sup> Naturalmente la profundidad y relevancia de toda polémica sobre metodología científica exige no caer en posiciones dogmáticas. El tema aludido de la contrastabilidad de las hipótesis ha sido objeto de debate por economistas de primera fila, entre los que se encuentran premios Nobel como Leontief, Samuelson o Friedman. Recomendamos, al efecto, la lectura de la selección de trabajos realizada por J. R. Quintás (1972) sobre *Método científico y teoría económica*. Revista Española de Economía, 1972, ps. 243 ySs.

$$K_t^* = \varphi \left( \frac{s_t}{p_t}, \frac{q_t}{p_t}; \mu_t; v_t; \gamma_t; r_t \right)$$

siendo,

$s$ : precio imput variable

$q$ : precio de los bienes de capital

$\mu$ : tipo impositivo sobre beneficios

$v$ : tipo permitido de deducción fiscal por amortización

$\gamma$ : tasa de depreciación física

$r$ : tipo de descuento

Jorgenson y Stephenson (1967) demuestran que bajo ciertas hipótesis, la anterior expresión puede transformarse en,

$$I_t = f[\Delta(Y/C)_{t-1}, \gamma K_{t-1}]$$

es decir, la inversión queda definida en función del stock de capital del período precedente (ponderado por la tasa de depreciación física) y de los incrementos de la relación entre la renta y una variable de costo de utilización del capital, que esta dada por:

$$C_t = \frac{q_t}{1 - \mu_t} [(1 - \mu_t v_t) \gamma_t + r_t]$$

con base a lo anterior, Pulido (1974) ha estimado el proceso de inversión privada fija no residencial en España durante en período 1958 - 1971, es decir con base en un modelo económico limitado en espacio y tiempo y formulando un modelo econométrico con el siguiente resultado:

$$I = -0,007(Y_t/C_t - Y_{t-1}/C_{t-1}) + 0,004(Y_{t-1}/C_{t-1} - Y_{t-2}/C_{t-2}) + 0,135 K$$

- ¿Cuál es la hipótesis?
- ¿Cuál es la tesis?

### *Bibliografía*

- Barbancho, Alfonso G. (1962) Fundamentos y posibilidades de la Econometría, Ediciones Ariel, Barcelona.
- Chiang, A. (1982) Métodos fundamentales de economía matemática. Amorrortu.
- Dagum, C. y Dagum, E.M.de. (1971). "Introducción a la Econometría". Editorial Siglo XXI. México.
- Hernández Sampieri, R., Fernández Collado, C. Baptista Lucio, P., Metodología de la Investigación. McGraw Hill. 1998
- Kinneer y Taylor. (1993). Investigación de Mercado, un enfoque aplicado. Mc Graw Hill.
- Klein, L. (1957) The scope and limitations of Econometrics. Applied statistics.

- Pulido, Antonio. (1993) Modelos econométricos. Editorial Pirámide. Madrid.
- Quintás, J. R. (1972) Método científico y teoría económica. Revista Española de Economía.
- Samuelson y Nordhaus. Economía. Mc Graw Hill. 1999.

Capítulo 3. PLANTEAMIENTO DE UNA TABLA DE DATOS .....	89
3.1 ¿Cómo obtener los datos? .....	90
Unidades y conceptos de la encuesta .....	91
Valores de la población y estadísticas .....	92
3.2 Características de los datos .....	93
3.3 Diseño del cuestionario .....	96
3.4. Distribuciones de probabilidad discreta y continua .....	99
CASOS DE ESTUDIO, PREGUNTAS Y PROBLEMAS .....	104
Caso 3.1: Tabla de datos para la investigación planteada en el Caso 2.1.....	104
Caso 3.2: El desempleo en Córdoba y Argentina .....	104
Preguntas .....	104
Bibliografía .....	106
Encuesta de Opinión .....	107

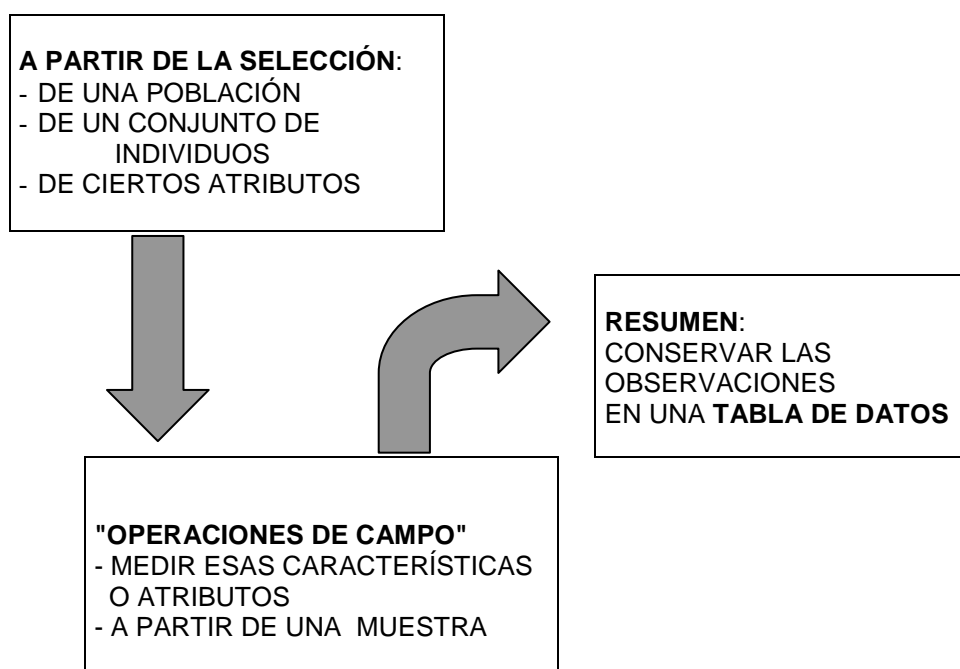




### Capítulo 3. PLANTEAMIENTO DE UNA TABLA DE DATOS

Se puede establecer una interesante comparación entre el trabajo de un artista plástico y el de un estadístico. Ambos necesitan algún recurso básico con el cual trabajar. Para el plástico, ese recurso es algún tipo de material. Para el estadístico, el recurso básico para realizar cualquier investigación son los datos. Así como el plástico utiliza diversos implementos para crear una obra con el material seleccionado, el estadístico, al igual que el artista, también debe seleccionar los instrumentos apropiados para moldear su "material" en un producto terminado: **los datos**. No obstante aquí termina la analogía.

La obra de arte se va a apreciar por su valor intrínseco, el del estadístico se apreciara en un sentido extrínseco de como auxiliará a la solución de diversos problemas de **administración, economía, mercados y otros relacionados con la empresa**. Ahora bien, tanto para el artista como para el estadístico, la calidad del producto final depende de la calidad de la materia prima utilizada. Por ello, la recolección correcta de datos es de extrema importancia para el investigador estadístico.



La Tabla de Datos resume las principales "operaciones de campo" que comporta todo proceso de observación. La definición de la investigación planteada en el capítulo anterior, determina el conjunto de características a estudiar sobre un grupo de individuos que pertenecen a una determinada población. Las características de esa población se miden sobre una muestra y los resultados de esa medición se disponen en una tabla de datos.

La Tabla de Datos es una tabla rectangular de **n individuos** por características observadas de los mismos. Estas características pueden ser **p variables reales** o **p variables categóricas**. Las primeras tendrán propiedades numéricas las segundas no.

¿Para qué se construye una Tabla de Datos?. Para que en la etapa de análisis de la información podamos evaluar:

- la semejanza entre los individuos a través de los atributos seleccionados,
- la asociación entre las características observadas sobre el conjunto de individuos.

	$A_1$	$A_2$	... $A_j$ ...	$A_p$
1				
$\vdots$				
$i$			$x_{ij}$	
$\vdots$				
$n$				

- la  $i$ -ésima línea contiene los valores que resultan de la "medición" de los **p** atributos sobre el  $i$ -ésimo individuo observado.

- la  $j$ -ésima columna contiene los **n** valores que resultan de la "medición" del  $j$ -ésimo atributo sobre los **n** individuos observados.

### 3.1 ¿Cómo obtener los datos?

Existen tres métodos básicos con los cuales el investigador puede obtener los datos deseados:

#### 1) **uso de fuentes de datos ya publicados,**

**Ejemplo.** Datos publicados en nuestro país por el INDEC o por las direcciones de estadísticas provinciales o los informes publicados por otros investigadores.

Direcciones de Internet:

INDEC, [www.indec.gov.ar](http://www.indec.gov.ar)

Ministerio de Economía, [www.mecon.gov.ar](http://www.mecon.gov.ar)

Ministerio del Interior, [www.mininterior.gov.ar](http://www.mininterior.gov.ar)

Gobierno de Córdoba, [www.cba.gov.ar](http://www.cba.gov.ar)

Presidencia de la Nación, [www.presidencia.gov.ar/enlaces](http://www.presidencia.gov.ar/enlaces)

Centro de Estudios para América Latina, [www.cepal.org](http://www.cepal.org)

Consejo Latinoamericano de Ciencias Sociales, [www.clacso.org](http://www.clacso.org)

Fundación de Investigaciones Económicas Latinoamericanas, [www.fiel.org](http://www.fiel.org)

- 2) **diseño de un experimento** o diseño experimental, cuyos conceptos fundamentales se estudiarán en otros capítulos.
- 3) **elaboración de una encuesta**, que es el de mayor aplicación en una investigación estadística.

Supóngase que el Decano de la Facultad de Ciencias Económicas, al enfrentarse a la tarea de proponer políticas que tendrían un efecto marcado en la calidad de vida y educación en el campus, desea llevar a cabo una encuesta para obtener información respecto a la composición del claustro estudiantil y respecto a los atributos físicos (características), actitudes e intereses de los estudiantes

Si la encuesta ha de resultar útil, el cuestionario debe ser válido; es decir, se deben plantear las preguntas correctas en una forma que extraiga respuestas significativas. Por ello a fin de ayudar en la elaboración del cuestionario, el Decano podría solicitar los servicios de un estadístico o investigador. Para poder diseñar el cuestionario se requiere el conocimiento de diferentes tipos de datos.

El estadístico, con toda seguridad, querrá desarrollar un instrumento en que se hagan diversas preguntas y trate una variedad de fenómenos. Por tanto, la materia manejada por el estadístico son **variables aleatorias**, o sea, fenómenos de interés cuyos resultados observados son **datos**. Como se describe en el cuadro existen dos tipos básicos de variables aleatorias que producen dos tipos de datos: cualitativos y cuantitativos. La diferencia entre ellos es que las variables aleatorias cualitativas arrojan respuestas categóricas (sí o no, por ejemplo); mientras que, las variables aleatorias cuantitativas dan respuestas numéricas que pueden, a su vez, clasificarse en continuas o discretas.

Tipos de datos	Tipos de preguntas	Respuestas
Cualitativos	Posee Ud. cuotas partes de algún Fondo de Inversión?	Sí No
Cuantitativos: Continuos Discretos	Cuál es su estatura? A cuántas revistas está suscripto?	_____ cmts _____ (cantidad)

### Unidades y conceptos de la encuesta

Los elementos de una población son las unidades en las que se busca la información; son los individuos, las unidades elementales que forman la población acerca de las cuales se van a hacer inferencias. Son las unidades del análisis, y su naturaleza se determina mediante los objetivos de la encuesta.

La población se define conjuntamente con los elementos: la población es el agregado de los elementos, y los elementos son las unidades básicas que forman y definen la población. Esta debe definirse en términos de:

- el contenido;
- las unidades;
- la extensión, y
- el tiempo.

**Ejemplo.** En el diseño de una encuesta de las expectativas de consumidores se podría especificar:

- a) todas las personas;
- b) en unidades familiares;
- c) en todo un país, y
- d) en 1992.

A menudo debe redefinirse la población deseada para obtener una población de encuesta que sea factible, de modo que lo anterior podría definirse como:

- a) todas las personas mayores de 18 años de edad que vivan en viviendas particulares,
- b) en unidades de gasto,
- c) en determinada región de un país,
- d) el 1º de enero de 1992.

La población de encuesta que realmente se obtenga tal vez difiera un poco de la población objetivo que se buscaba. La diferencia principal suele deberse a la no respuesta y a la cobertura deficiente. En un sentido estricto, solo queda representada en la muestra la población de encuesta. Pero esto puede constituir una descripción difícil, si se quiere hacerlo exactamente y es más fácil escribir acerca de la población objetivo definida.

Una encuesta puede dar información acerca de varias y diversas poblaciones. La misma encuesta puede abarcar contenidos diferentes. Por ejemplo, los datos acerca de consumidores pueden presentarse en términos de personas, unidades de gasto, familias o viviendas.

Las unidades de observación son aquellas de las que se obtienen observaciones. En las encuestas de entrevistas se denominan "unidades de respuesta". Estas unidades son a menudo los elementos, como sucede en las encuestas de actitudes, pero las dos pueden ser diferentes. El jefe del hogar puede dar datos de todas las personas de la familia, o de todos los niños; en el primer caso, las personas que dan los datos pertenecen, con las otras, a la población; en el segundo no sucede lo mismo. En una encuesta de niños de escuela, el que da los datos puede ser el profesor.

### Valores de la población y estadísticas

La investigación empírica puede hacerse de diferentes maneras: por medio de observaciones casuales, de observaciones controladas, de experimentos o de encuestas. La investigación por medio de encuestas tiene por objetivo estimar valores específicos de la población. Un valor de la población es una expresión numérica que sintetiza los valores de una o varias características de los  $N$  elementos de una población completa; es una medida resumida de una cualidad de la distribución de la variable o variables en la población definida. El ejemplo básico en encuestas por muestreo es la media de la población:  $\mu$ . Algunos valores de la población

muy relacionados con la media son las proporciones, medianas, y otros cuantiles, y el total o valor agregado de la población. Otros valores de la población miden relaciones; el más común de estos es la diferencia entre dos medias; otros ejemplos son el coeficiente de correlación. Cualquier valor de la población se determina por cuatro factores que se mencionan en los primeros tres objetivos de una encuesta:

- a) la población definida de encuesta;
- b) la naturaleza de las variables de la encuesta y, en algunos casos, sus distribuciones;
- c) los métodos de observación, y
- d) la expresión matemática para derivar el valor de la población a partir de los valores individuales de los elementos.

El valor de la muestra, o estadística, es una estimación que se calcula a partir de los  $n$  elementos en la muestra. Es una variable aleatoria que depende del diseño de la muestra y de la combinación particular de los elementos que resultaron seleccionados. Por tanto, la estimación que se hace es solamente una de las que pudieron haberse obtenido con el mismo diseño de muestra. Por el contrario, el valor de la población depende de los  $N$  valores en dicha población. Es una constante independiente de las fluctuaciones de la selección, aunque por lo general se desconozca.

### 3.2 Características de los datos

Las características de los datos se estudian a través de sus medidas de posición y dispersión. La medida de posición es la característica más importante que describe o resume un grupo de datos. Todo grupo de datos tiene asociado un valor típico descriptivo denominado *promedio o media aritmética*. Otras medidas de posición: son la **mediana**, la **moda**, el **rango medio** y los **percentiles**, cuartiles y deciles.

La **media aritmética** es el promedio que surge de sumar todos los valores de la muestra y dividirlos por el total de observaciones. Puesto que su cálculo se basa en cada observación, la media aritmética se ve afectada en gran medida por cualquier valor extremo; es decir, actúa como punto de equilibrio de tal forma que las observaciones menores compensan aquellas que son mayores.

$$\bar{X} = \frac{\sum_{i=1}^n X_i}{n} = \frac{X_1 + X_2 + \dots + X_n}{n}$$

donde:  $\bar{X}$ , media aritmética de la muestra  
 $n$ , tamaño de la muestra

$X_i$ ,  $i$ -ésima observación de la variable aleatoria  $X$

$\sum_{i=1}^n$  símbolo griego que significa "suma de todos los valores de 1 a n"

Las propiedades de la media son:

$$\text{Equilibrio: } \sum_{i=1}^n (x - \bar{x}) = 0$$

$$\text{Mínimo: } \sum_{i=1}^n (x - \bar{x})^2 = \text{Total mínimo}$$

$$\text{Total: } \text{Valor total} = N\bar{x} \quad \begin{array}{l} N, \text{ tamaño de la población} \\ \bar{x}, \text{ media aritmética de la muestra} \end{array}$$

La mediana es el valor que aparece en el medio de una secuencia ordenada de datos y no se ve afectada por ninguna observación extrema. Para calcular su valor, primero debemos ordenar los datos y luego usar la fórmula de posicionamiento  $(n+1)/2$ . Si el tamaño de la muestra es un número impar, la mediana se representa mediante el valor numérico correspondiente al punto de posicionamiento; mientras que, si el tamaño de la muestra es par, el punto de posicionamiento cae entre las dos observaciones medias de la clasificación ordenada, el promedio de estas dos observaciones es la mediana.

La moda es el valor más frecuente en la muestra. Se obtiene fácilmente de una clasificación ordenada y no se ve afectada por la ocurrencia de valores extremos.

El rango medio es el promedio de las observaciones menores y mayores de una serie de datos. A menudo es usado como medición de resumen, por ejemplo para datos meteorológicos, puesto que puede proporcionar una medición adecuada, rápida y simple para caracterizar toda una serie de datos. Dado que involucra los menores y mayores de una sucesión, el rango medio se distorsiona como una medición de resumen de tendencia central si está presente una observación extrema.

Los percentiles son medidas que subdividen una distribución de mediciones de acuerdo con la proporción de frecuencias observadas. Así el  $p$ -ésimo percentil es el valor de  $X$  que representa el  $p$  por ciento de observaciones que se encuentran por debajo de  $X$  y el  $(100-p)$  por ciento de observaciones que se encuentran por encima de  $X$ . En general la localización del  $k$ -ésimo percentil está dado por:

$$P_k = \frac{k}{100} n$$

El vigésimo quinto percentil representa el *primer cuartil*, el quincuagésimo percentil (la mediana) se conoce como *segundo cuartil* o

*cuartil medio* y el septuagésimo quinto percentil es el *tercer cuartil*. Para el cálculo del primer y tercer cuartil se utilizan las siguientes fórmulas:

$$Q_1 = \frac{n+1}{4}, \quad Q_3 = \frac{3(n+1)}{4}$$

Las medidas de dispersión se ocupan de describir la variabilidad, diseminación o dispersión de los datos. Las más utilizadas son el recorrido, la varianza, el desvío estándar, y el coeficiente de variación.

El recorrido, rango o amplitud es la diferencia entre el valor más grande y el más pequeño.  $R = x_n - x_1$

La varianza y el desvío estándar tienen en cuenta cómo se distribuyen todas las observaciones en los datos. La varianza mide el promedio del cuadrado de las diferencias entre cada observación y su media

$$s^2 = \frac{\sum_{i=1}^n (x_i - \bar{x})^2}{n-1}, \text{ donde } n-1 \text{ son los grados de libertad.}$$

El desvío estándar mide la raíz cuadrada del promedio del cuadrado de las diferencias alrededor de la media

$$s = \sqrt{s^2} = \sqrt{\frac{\sum_{i=1}^n (x_i - \bar{x})^2}{n-1}}$$

Para las situaciones en las cuales resulta apropiado utilizar una cantidad total, estimada a partir de la media de la muestra, la desviación estándar total es:

$$s_{total} = Ns$$

El coeficiente de variación es una medida relativa, de la desviación estándar respecto de la media de la distribución, que se utiliza para comparar dos o más conjuntos de datos.

$$CV = \frac{s}{\bar{x}} 100$$

**Ejemplo.** Para dos acciones de empresas de la industria electrónica, el precio promedio de cierre en el mercado de valores durante un mes fue, para la acción A, de \$1500.00, con desviación estándar de \$500.00. Para la acción B, el precio promedio fue de \$5000.00, con desviación estándar de \$300.00. haciendo una comparación absoluta, resultó ser superior la variabilidad en el precio de la acción A debido a que muestra una mayor desviación estándar. Pero, con respecto al nivel de precios, deben compararse los respectivos coeficientes de variación. A partir de allí, puede concluirse que el precio de la acción A ha sido casi dos veces más variable de la acción A con respecto al precio promedio para cada una de las dos.

	Acción A	Acción B
Desvío	500	300
Media	15000	5000
Coeficiente de variación	3.33%	6%

### 3.3 Diseño del cuestionario

El objetivo básico del área estadística será brindar al usuario el diseño de Encuestas y de Muestras para que los aplique atendiendo a sus problemas. Para ello se proveerán los elementos fundamentales para:

- a) la definición de las variables de la encuesta
- b) los métodos de observación
- c) los métodos de análisis estadístico
- d) la utilización estadística de los resultados de la encuesta
- e) la precisión que se desea que tengan los resultados de la encuesta.

Los objetivos de la encuesta deben determinar el diseño de la muestra, pero la determinación es, en realidad, un proceso reversible, puesto que los problemas de diseño de la muestra influyen e incluso cambian los objetivos de la encuesta.

En este paso de la investigación se trata de construir un instrumento que sirva para medir los conceptos que se han seleccionado. Los métodos de recolección más utilizados en el tipo de investigación que se están desarrollando son la observación, la entrevista y el cuestionario.

La **observación** se aplica preferentemente en aquellas situaciones en las que se trata de detectar aspectos conductuales, como ocurre en situaciones externas y observables. Esta observación puede ser **participante** o **sistemática**, la primera es utilizada, por ejemplo, por la antropología en la cual el investigador se familiariza con la situación a estudiar; en la segunda se aplica en situaciones de diagnóstico y clasificación en base a tipologías ya establecidas de modo que la observación se convierte en una tarea de registro.

La **entrevista** es una técnica de recolección de datos que implica una pauta de interacción verbal, inmediata y personal, entre un entrevistador y un respondente. Dependiendo del tipo de investigación, las entrevistas se clasifican en:



**No estandarizadas**, se utilizan en etapas exploratorias para detectar las dimensiones más relevantes y generar las hipótesis generales. El rasgo distintivo de este tipo de entrevistas es la flexibilidad en la relación entrevistador-respondente.

**Semiestandarizada**, permiten margen para la reformulación y la profundización de algunas áreas. Por lo general existe una pauta de guía de la entrevista, en donde se respeta el orden y fraseo de las preguntas.

**Estandarizadas** son entrevistas basadas en un cuestionario donde el entrevistador leerá las preguntas a su entrevistado siguiendo el orden preestablecido.

La elaboración del *cuestionario* es un arte que mejora con la experiencia y está compuesto de un espacio para registrar la encuesta como unidad, preguntas y espacio para la respuesta a esas preguntas.

**Ejemplo.** Al final del Módulo se encuentra el cuestionario adoptado en una localidad de la Provincia de Córdoba para el estudio del perfil de la comunidad.

Se debe tomar en cuenta tanto la longitud como el modo de obtener respuestas lo cual da lugar a preguntas abiertas, cerradas o semicerradas. Las preguntas cerradas, con respuestas categóricas, no deben ser siempre de la categoría dicotómica, -sí o no, blanco o negro- sino que pueden dar lugar a la posibilidad de responder en diferentes grados, para ello se emplean escalas de respuestas del tipo:

MUY BUENO  
BUENO  
NI BUENO NI MALO  
MALO  
MUY MALO

o las que utilizan grados de acuerdo, como:

COMPLETAMENTE DE ACUERDO  
DE ACUERDO  
INDIFERENTE  
EN DESACUERDO  
COMPLETAMENTE EN DESACUERDO

Estos ejemplos muestran solo formas de desarrollar las escalas. Algunas, como las de grado de acuerdo, han sido elaboradas con criterio psicosocial por el autor Lickert, a la que deben su nombre, y se centra en el estímulo más que en el sujeto. Otras formas dependerán de la clase de respuestas que necesita el investigador para llevar a cabo sus mediciones.

El formulario no debe ser tan extenso de forma que no desconcentre al entrevistado y lo haga contestar de acuerdo a sus impulsos y no a sus reflexiones.

Las preguntas abiertas del tipo:

*"¿por que eligió asistir a esta facultad en vez de otras facultades?\_\_\_\_\_"*

que requiere que el sujeto responda libremente en forma de composición, suelen ser indeseables, pues el sujeto puede opinar que esas preguntas requieren un largo pensamiento previo y demasiado tiempo para elaborar una respuesta.

Además, como esas preguntas requieren tiempo para contestarlas, también requieren tiempo para valorarlas y procesarlas. En su lugar, si creemos conveniente poseer una respuesta a la elección de una facultad en lugar de otra, deberíamos pensar en codificar una escala para esa pregunta, como sería por ejemplo:

"porque el nivel de enseñanza es muy superior"

"porque el nivel de enseñanza es superior"

"por otros motivos no vinculados al nivel de enseñanza"

Las preguntas semicerradas tienen categorías de respuesta donde la última es una pregunta abierta del tipo

*"Otra, ¿cuál? \_\_\_\_\_"*

a efectos de captar categorías de análisis no previstas con anterioridad.

Hay tres formas de realizar entrevistas:

- personales (a través de encuestadores)
- por correo
- telefónicas (telemarketing)

Indudablemente las más efectivas, por diversas razones, son las entrevistas personales donde existe una definición clara del marco poblacional, seguridad en la información suministrada, definición clara de la unidad muestral y, quien realiza la entrevista, está lo suficientemente preparado en el tema de estudio.

Uno de los objetivos principales de la investigación estadística es la **inferencia estadística**. Esta consiste en utilizar estadísticos, media aritmética, desviación estándar y proporción que se obtienen con los datos de la muestra, para estimar su verdadero valor en la población.

Es necesario darse cuenta de que el investigador estadístico, a través de los datos recolectados en una encuesta se interesa en sacar conclusiones de la población y no de la muestra. Por ejemplo, un encuestador político se interesa en los resultados de la muestra solo como medio para estimar la proporción real de votos que recibirá cada candidato entre la población de votantes.

En la práctica, una muestra individual de tamaño determinado se selecciona aleatoriamente entre la población. Las unidades muestrales se van a obtener mediante, por ejemplo, una tabla de números aleatorios.

En el caso hipotético. a fin de poder utilizar la media de la muestra para estimar la media de la población, se deberían examinar todas las muestras posibles de la población y calcular una media para cada muestra. La distribución de estos resultados se denomina distribución en el muestreo.

Aunque en la práctica solo se selecciona una muestra, se debe examinar el concepto de la distribución en el muestreo, con el propósito de poder utilizar la teoría de la probabilidad para hacer inferencias en cuanto a los valores de la población. Este tema será desarrollado en la Unidad II.

### 3.4. Distribuciones de probabilidad discreta y continua

En la investigación de cualquier fenómeno es necesario observar y recoger algunas características de los objetos en estudio. Debemos obtener datos de alguna forma. Para obtenerlos se realizan observaciones.

El término **observación** se usará para representar cualquier clase de "recogida" numérica de información. Una forma de observación directa en la que se controlan algunos o todos los factores que pueden influir sobre la variable que se estudia se lo conoce con el nombre de **experimento**.

Estos datos, de los cuales vamos a obtener información, por medio de la observación, pueden estar contenidos en cualquier **población** o **muestra** de una población. He aquí dos conceptos fundamentales para cualquiera que desee realizar un análisis cuantitativo de los datos recogidos.

Una **población** puede definirse como el conjunto de todas las observaciones posibles de datos. Relacionado con este concepto, aparece el de **muestra**, que es un conjunto de datos seleccionados a partir de la población.

**Ejemplo.** Los ingresos de los habitantes de un país en un período de tiempo determinado y todos los Estados de Resultados mensuales de los ejercicios económicos de la Empresa "A" obtenidos entre 1980 y 1991. Cualquier par de resultados mensuales de la Empresa "A" seleccionados entre 1980 y 1991 constituye una muestra.

Nos van a interesar las muestras probabilísticas o sea, aquellas obtenidas por medio de un mecanismo casual determinado. Una **muestra**

**aleatoria** es una muestra probabilística en donde cada dato observable de la población tiene la misma probabilidad de ser elegido o seleccionado.

Cuando no es posible obtener la información buscada a través de la observación directa utilizamos la **encuesta**. Este es el proceso por el cual obtenemos los datos solicitando la información a otras personas. En una encuesta los datos pueden obtenerse a través de una entrevista, llamada telefónica o cuestionario escrito.

A partir de lo que se ha dicho hasta aquí, debería quedar claro que se estudiarán los fenómenos que pueden ser medidos o contados. Un fenómeno que se puede medir es una **variable**; esta indica una cantidad homogénea que puede tomar valores distintos en distintas observaciones. Las variables cuyos valores no se pueden determinar antes de ser observados se denominan **variables aleatorias**. Estas se clasifican en **variables aleatorias cualitativas** o **categorías** y **variables aleatorias cuantitativas** o **reales**.

**Variables Categóricas** o **Cualitativas**, son aquellas que toman un número limitado de modalidades y a cada modalidad corresponde una categoría de individuos. Estas categorías forman una partición de la población.

**Ejemplo.** El sexo es una variable categórica o cualitativa que tiene dos modalidades, de modo que la población se particiona entre "hombres" y "mujeres".

A veces el interés se centra en una categoría particular **A**. La proporción de **A** con respecto a la población se simboliza  $\pi$ , y proporción de **A** con respecto a la muestra  $\hat{\pi}$  o  $p$ .

Generalmente las variables categóricas se codifican con números, pero no pueden considerarse como variables reales. Los cálculos de media o de varianza sobre los valores no tienen sentido porque carecen de propiedades numéricas.

Sea **A** una categoría de una variable cualitativa, se denomina **indicadora** de **A** una variable que toma el valor 1 para cualquier individuo que pertenece a **A** y 0 si no pertenece. La media de la indicadora de **A** es la proporción  $\pi$  de **A** en la población y la varianza es igual a  $\pi(1 - \pi)$ .

**Variables Reales** o **Cuantitativas**, son aquellas que toman valores reales para los cuales podemos calcular resúmenes numéricos como la media, varianza y desviación estándar.

**Ejemplo.** Edad, sueldo, PBI, ventas, contenido de un paquete de galletitas, kilómetros recorridos.

Dada una variable real particular,  $Y$ , se tienen la media y la varianza de  $Y$  en la población,  $\mu$  y  $\sigma^2$ ; la media y la varianza de  $Y$  en la muestra,  $\bar{Y}$  y  $s^2$ .

Las variables reales se distinguen entre **variables continuas** y **variables discretas**. Las primeras pueden tomar cualquier valor dentro del campo de los números reales, por ejemplo, renta o gasto. Las segundas sólo pueden tomar valores enteros o fracciones pero sin continuidad, por ejemplo número de hijos.

Una **distribución de frecuencias** representa una ordenación de los datos de tal forma que indique el número de observaciones en cada clase de la variable surgidas de una muestra. El "número de observaciones" en cada valor se denomina frecuencia **absoluta** que se distingue de la frecuencia **relativa** ya que esta indica la "proporción de observaciones" en cada clase y no su número.

**Ejemplo.** Se observa una muestra de 6 estados de resultados de la Empresa "A", considerando en cada uno la incidencia de los gastos financieros sobre el total de la venta. Los resultados aparecen en el Cuadro I-2.

**Cuadro I-2: Muestra de los Estados de Resultados de la Empresa "A"**

% gasto financiero (variable)	Nº de Estados de Resultados (frec.absoluta)	Proporción de Estados de Resultados (frec.relativa) <sup>1</sup>
menos de 5%	1	0.16672
5% a 9,99%	3	0.50000
10% a 19,99%	2	0.33330
20% o más	0	0.00000
TOTALES	6	1.00000

En una distribución de frecuencias los datos pueden agruparse, como en el Cuadro I.2, en diferentes clases. Para cada una de las clases de una distribución de frecuencias, los límites nominales de clase inferior y superior indican los valores incluidos dentro de la clase.

Las **fronteras de clase** o **límites exactos** son los puntos específicos de la escala de medición que sirven para separar clases adyacentes cuando se trata de variables continuas. Los límites exactos de clase pueden determinarse identificando los puntos que están a la mitad entre los límites superior e inferior de las clases adyacentes.

El **intervalo de clase** indica el rango de los valores incluidos dentro de una clase y puede ser determinado restando el límite exacto inferior de clase de su límite exacto superior. Los valores de una clase se representan a menudo por el **punto medio de clase** que puede ser determinado sumando los límites superior e inferior y dividiendo por dos.

En una población, el concepto correspondiente a la distribución de frecuencias de una muestra se conoce con el nombre de **distribución de probabilidades**.

**Ejemplo.** La población de todos los estados de resultados mensuales de la Empresa "A" en los ejercicios económicos entre los años 1980 y 1991, clasificados según la incidencia de los gastos financieros se encuentran en el Cuadro I-3.

**Cuadro I-3: Estados de Resultados de la Empresa "A"**

% gasto financiero (variable)	% de Estados de Resultados (frec.relativa)
menos de 5%	16.67 %
5% a 9,99%	50.00 %
10% a 19,99%	33.33 %
20% o más	0.00 %
TOTAL	100.00 %

Resulta evidente que el hecho de decir que el 16,67% de los Estados de Resultados mensuales tuvieron una incidencia de gastos financieros inferior al 5% equivale a afirmar que la probabilidad de seleccionar aleatoriamente un Estado de Resultado con gastos financieros inferiores al 5% es 0,1667.

Queda claro que el Cuadro I-3 representa una distribución poblacional y el Cuadro I-2 una distribución muestral. No siempre, obviamente, la muestra representa exactamente a la población, como en este caso; pero, con la distribución muestral se puede inferir la poblacional, evaluando estadísticamente el error cometido a partir de la inferencia.

Las distribuciones de frecuencia pueden graficarse a través de barras, o **histograma**, en el cual el eje horizontal representa los intervalos de clase y el eje vertical el número de observaciones.

El **polígono de frecuencias** es la línea que une los puntos medios de una distribución de frecuencias. Por su parte la **curva de frecuencias** es un polígono de frecuencias suavizado.

Esto define la naturaleza de la **inferencia estadística** que se dedica a efectuar generalizaciones respecto a la población a partir de la información proporcionada por la muestra.

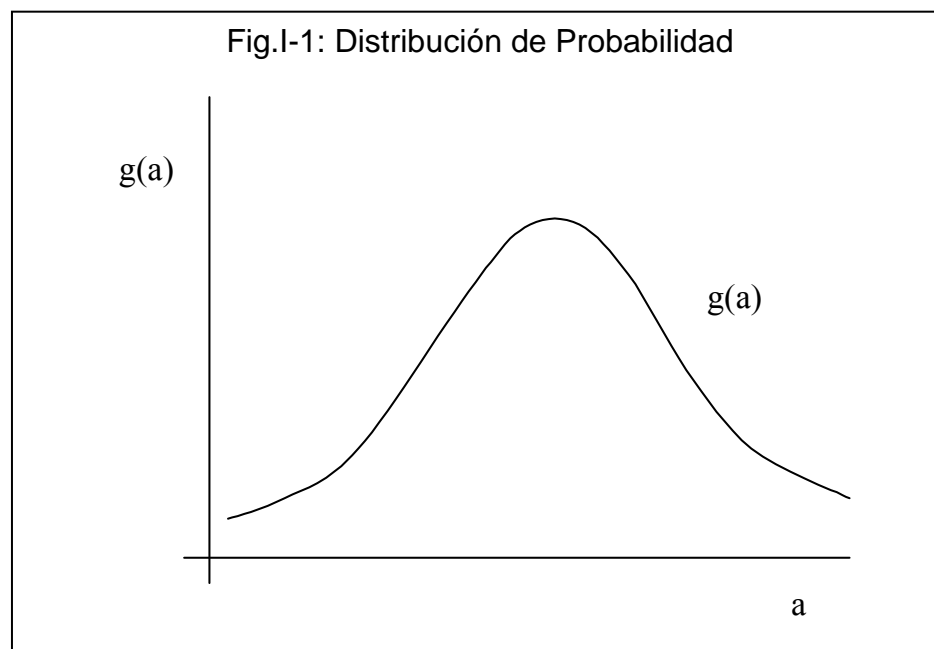
**Ejemplo.** Los ejecutivos de la Empresa "A" estarán acostumbrados a realizar generalizaciones respecto a lo que ocurrirá con los gastos financieros durante el próximo ejercicio económico de acuerdo a la experiencia del pasado. Esto es precisamente lo que se hace en la Inferencia Estadística, salvo que, en este último caso, se realiza en forma científica.

En resumen, se utilizan las muestras para hacer juicios respecto a las poblaciones de las que provienen las muestras. Estos juicios se pueden referir tanto a situaciones del pasado como a estimar lo que ocurrirá en el futuro. En general, sólo interesa alguna característica de la población

que se denomina **parámetro**, sobre él serán nuestros juicios basándonos en el estadístico de la muestra, que recibe el nombre de **estimador**.

**Ejemplo.** Si el parámetro que se desea estimar es el gasto financiero medio, el estimador muestral que se usa es el promedio muestral del gasto financiero.

Ahora bien, para obtener una distribución muestral deben obtenerse todas las muestras posibles de la población. Puede imaginarse que se van obteniendo una detrás de otra todas las muestras posibles y para cada una se calcula el valor en cuestión (por ejemplo, el gasto financiero). La distribución resultante de este número de muestras se denomina distribución muestral. Una **distribución muestral** es la distribución de probabilidad de un estimador.



De esta forma, si se estudia la población de todos los valores posibles de gasto financiero de los últimos 12 ejercicios económicos de la Empresa "A" (variable  $X$ ) e interesa estimar el gasto financiero medio (un parámetro  $\alpha$ ) se usa el estadístico muestral  $a$ , donde  $a$  es un estimador de  $\alpha$ , y un valor concreto de  $a$  (obtenido a partir de una muestra determinada) es una estimación de  $\alpha$ . Como el gasto financiero de una empresa es una variable continua, la distribución muestral de  $\alpha$  tendrá, posiblemente, una forma como la que se muestra en la Figura I-1.

Sin mayor rigurosidad científica se puede decir que el concepto de distribución muestral se puede entender mejor si se lo considera como una distribución de frecuencias relativas calculadas a partir de un número grande de observaciones (en este caso, de muestras) del mismo tamaño. En la Figura I-1, las frecuencias relativas de  $a$  se miden a lo largo del eje  **$g(a)$** .

Con el análisis anterior se quiere dejar en claro la gran importancia que tiene el estudio de las distribuciones muestrales, ya sean experimentales, que son las obtenidas a partir de un número grande o infinito de muestras posibles; o las teóricas, que están basadas en la teoría de la

probabilidad. Estas últimas, las **distribuciones teóricas**, permiten estudiar poblaciones sin necesidad de realizar repeticiones de experimentos muestrales. Se distinguen según se refieran a variables discretas o continuas. Dentro de las continuas, la Distribución Normal es la más utilizada ya que responde a la distribución de muchos fenómenos de las ciencias sociales.

---

---

### CASOS DE ESTUDIO, PREGUNTAS Y PROBLEMAS

Caso 3.1: Tabla de datos para la investigación planteada en el Caso 2.1.

Diseñe la tabla de datos que se corresponda con esa investigación y describa

- a) cómo obtendría los datos
- b) cuáles serían las observaciones
- c) qué observaría sobre ellas
- d) cómo sería el cuestionario que permita reunir la información para la investigación.

Caso 3.2: El desempleo en Córdoba y Argentina

Elabore una tabla de datos con información sobre desempleo en los aglomerados Gran Río Cuarto, Gran Córdoba y Argentina desde 1995. Analice las similitudes y diferencias que presenta la evolución del desempleo en cada área geográfica.

### Preguntas

**3.1** De los siguientes conceptos, ¿cuáles indican muestras? y ¿cuáles población?

- a) grupos de medidas llamados *parámetros*
- b) uso de *inferencia estadística*



- c) hacer un *censo*
- d) juzgar la calidad de un embarque de fruta inspeccionando varias de las cajas
- e) Universo
- f) grupo de medidas denominadas *estadísticas*
- g) aplicación de conceptos de probabilidad
- h) inspección de todos los artículos que se fabrican

**3.2** De los siguientes procedimientos de recopilación de datos, ¿cuáles indican experimentos? y ¿cuáles encuestas?

- a) una investigación política sobre la intención de votos en las próximas elecciones
- b) se entrevista a los clientes de un centro comercial sobre las razones que tienen para hacer sus compras en ese lugar
- c) se comparan dos métodos para comercializar una póliza de seguros, utilizando cada uno de los métodos en áreas geográficas diferentes
- d) i'987r2dx comparar los resultados de un nuevo método, para entrenar agentes de boletos de avión, con el método tradicional
- e) evaluar dos conjuntos diferentes de instrucciones para el armado de un juguete, haciendo que dos grupos comparables de niños armen el juguete utilizando diferentes instrucciones
- f) hacer que una revista envíe a sus suscriptores un cuestionario pidiéndoles que evalúen un producto recientemente adquirido

**3.3** Clasifique las siguientes variables en cuantitativas, discretas y continuas, o cualitativas.

- a) el peso del contenido de un paquete de cereales
- b) el diámetro de un rodamiento
- c) el número de artículos defectuosos
- d) el número de personas pensionadas en un área geográfica
- e) el número promedio de posibles clientes que han sido visitados por un vendedor el mes pasado
- f) el total de ventas en pesos
- g) número de unidades de un artículo en existencia
- h) relación de activos circulantes con pasivos circulantes
- i) tonelaje total embarcado
- j) cantidad embarcada, en unidades
- k) número de vehículos en una ruta provincial
- l) asistencia a la reunión anual de una compañía
- m) Duración útil de un foco eléctrico de 100 Watts.
- n) Partido político al cual están afiliados los empleados públicos.
- o) Número de quiebras de empresas por mes.
- p) Pagos de dividendos de diversas empresas de Servicios Públicos.
- q) Número de llegadas a tiempo, por hora, en un aeropuerto grande.

---

---

### *Bibliografía*

- Berenson, Mark y Levine, David. *Estadística Básica en Administración*. Prentice Hall. México, 1996.
- Hernández Sampieri, Roberto; Fernández Collado, Carlos y Baptista Lucio, Pilar. *Metodología de la Investigación*. Segunda Edición. McGraw Hill. México, 2000.
- Kazmier, Leonard y Diaz Mata, Alfredo. *Estadística Aplicada a la Administración y a la Economía*. Mc.Graw Hill. México, 1993.
- Padua, Jorge. *Técnicas de Investigación Aplicadas a las Ciencias Sociales*. Fondo de Cultura Económica. México, 1996.

*Encuesta de Opinión***PERFIL DEL  
HABITANTE DE  
HHH**

Encuestador \_\_\_\_\_ Circuito \_\_\_\_\_ Sector \_\_\_\_\_ Manzana \_\_\_\_\_ Cuestionario N°: \_\_\_\_\_  
 Dirección: \_\_\_\_\_ Barrio: \_\_\_\_\_  
 Fecha: \_\_\_\_\_ Hora: \_\_\_\_\_ Tiempo de duración de la entrevista: \_\_\_\_\_

Supervisor: \_\_\_\_\_ Fecha: \_\_\_\_\_  
 Observación del Supervisor:

---



---



---



---

BUENOS DIAS/BUENAS TARDES: Estamos haciendo una encuesta sobre temas de actualidad y quisiéramos hablar con una persona que habite esta vivienda.  
 (Decir que son de XXConsultora . Si es necesario mostrar la credencial. Aclarar el carácter ANONIMO de la entrevista y pasar a seleccionar al entrevistado).

**CONFIDENCIAL**

La encuesta se realiza a personas mayores de 18 años que habiten la vivienda seleccionada. El encuestado deberá seleccionarse utilizando el cuadro siguiente. Es necesario detallar en orden decreciente de edad todos los integrantes de la vivienda consignando edad y sexo. Luego cruzar la fila del último integrante del hogar, en condiciones de responder, con la columna correspondiente al último número del cuestionario. El número resultante del cruce indicará el orden de la persona que deben encuestar. Si la persona seleccionada no se encuentra podrán reemplazar en orden decreciente de edad por otra persona de igual sexo. Si esto no es posible, deberán buscar el reemplazo por una persona de igual sexo y similar edad en la vivienda que corresponda de acuerdo al conteo establecido dentro de la manzana.

**CUADRO DE SELECCION**

Número de Orden	Edad	SEXO		Ultimo dígito de número cuestionario										Selección	
		Hombre	Mujer	1	2	3	4	5	6	7	8	9	0	SI	NO
1				1	1	1	1	1	1	1	1	1	1		
2				2	1	2	1	2	1	2	1	2	1		
3				3	1	3	2	2	1	2	1	2	3		
4				2	3	4	1	3	2	1	1	3	4		
5				5	4	3	4	2	1	2	1	3	5		
6				5	5	4	4	2	2	1	3	3	6		
7				1	6	5	7	2	6	4	1	3	7		
8				7	4	3	7	1	2	6	3	5	8		
9				1	2	3	9	5	6	7	8	4	9		
10				9	3	6	5	2	10	7	8	1	4		

OBSERVACIONES: (indicar cuando la persona encuestada corresponda a un reemplazo, consignar domicilio de entrevista)

---



---



---

1. En términos generales, cómo es el habitante de HHH? (L)
1. Conservador
  2. Progresista
  3. Ni uno ni otro
  4. Otro. Cuál? \_\_\_\_\_
  5. Ns/Nc

2. Cómo se entera **primero** Ud. y su familia sobre lo que pasa en HHH y en el país? (L)

HHH		Argentina
1	Por Radio	1
2	Por TV	2
3	Por Diarios Locales	3
4	Por Diarios Provinciales	4
5	Por Diarios Nacionales	5
6	Alternativamente por unos y otros	6
7	Por comentario en el ámbito de trabajo	7
8	Por comentario en otro ámbito (bar, etc)	8
9	Ns/Nc	9

3. Cuánto le gusta.....?. (Mostrar T1)

	No me gusta	Me gusta poco	Me gusta	Me gusta mucho	Me gusta muchísimo	Ns/Nc
Tomar fotografías						
Ir a la Iglesia						
Ir a ver Deportes						
Practicar Deportes						
Hacer gimnasia						
Jugar al naípe						
Escuchar radio						
Conversar de política						
Tocar algún instrumento						
Estar en familia						
Reunirse con amigos						
Leer el diario						
Bailar						
Ir al cine						
Leer libros						
Trabajar en el jardín						
Ir a Exposiciones de arte						
Leer revistas						
Pescar o Cazar						
Ver Televisión						
Participar en Asociaciones						
Mirar vidrieras						
Comer en restaurantes						
Viajar						
Navegar por Internet						

4. Cómo es su interés con respecto a los hechos.....? (L)

	Muy Interesado	Interesado	No muy Interesado
Políticos			
Sociales			
Económicos			
Culturales			
Deportivos			

5. Lee Ud. diarios? (E)

1. SI
2. NO. Porqué? \_\_\_\_\_ (Ir a 9)
3. Ns/Nc

6. Cuál es el diario que **lee más**? (E)

1. Local
2. Ambito Financiero
3. La Voz del Interior
4. Puntal
5. Clarín
6. Otro. Cuál? \_\_\_\_\_
7. Ns/Nc

7. El diario que Ud. más lee

1. Lo compra?
2. Lo ve por Internet?
3. Se lo presta un vecino?
4. Lo lee en el trabajo?
5. Otra forma. Cuál? \_\_\_\_\_
6. Ns/Nc

8. Ud. lee también otro diario? (E)

1. SI. Cuál? \_\_\_\_\_
2. NO
3. Ns/Nc

9. Dónde escucha Ud. habitualmente radio (Marcar solo el lugar donde lo hace más frecuentemente) (E)

1. No escucha (IR A 11)
2. Hogar
3. Trabajo
4. Auto
5. Ns/Nc

10. Qué radio escucha Ud. y en que momento del día? (L)

	Mañana	Tarde	Noche
1. Radio X1			
2. Radio X2			
3. Radio X3			
4. Radio X4			
5. Radio X5			
6. Radio X6			
7. Otra. Cuál? _____			
8. Otra. Cuál? _____			
9. Ns/Nc			

11. Ud. mira televisión? (E)

1. SI
2. NO (IR A 16)
3. Ns/Nc

12. Qué canales de TV mira Ud. y en que momento, del día? (E)

	Mañana	Tarde	Noche
1) América 2			
2) Canal 9 Libertad			
3) Canal 13 Bs.As.			
4) Canal 11 Telefe			
5) Canal 7 ATC			
6) Canal Local			
7) Crónica TV			
8) TN Noticias			
9) TyC Sports			
10) Utilísima			
11) Otro. _____			
12) Otro: _____			

13. Está abonado al servicio de televisión por cable?

- 1.SI      2.NO (Seguir en 15)    3.Ns/Nc

14. A qué servicio de televisión por cable está abonado?

- 1.SIT SRL                      2. Multicanal  
3.SKY                            4.Cablevisión  
5.Direc TV

15.Cuál es su programa preferido por televisión?  
(A)

\_\_\_\_\_

16. De los medios de comunicación existentes en HHH, cuál representa mejor el sentir de la gente?  
(A)

\_\_\_\_\_

17.Qué publicidad recuerda Ud. más frecuentemente? (L)

1. La de radio                      2. La de televisión  
3. La Gráfica                      4. Todas por igual  
5. Ninguna                          6.Ns/Nc

18. Con qué Banco opera Ud. habitualmente? (E)

1. No opero con Bancos (Pasar a 21)  
2. Banco Provincia de Córdoba  
3. Banco Nación Argentina  
4. Otro. Cuál? \_\_\_\_\_  
5. Ns/Nc

19.Qué servicios tiene con su banco? (L)

- 1.Caja de ahorro  
2.Cuenta corriente  
3.Inversiones, Plazo fijo, Fondo común  
4.Crédito hipotecario  
5.Crédito personal  
6.Tarjeta de crédito  
7.Otro. Cuál? \_\_\_\_\_  
8.Ns/Nc

20.Se siente satisfecho con el servicio que le brinda su banco? (E)

- 1.SI  
2.NO Porqué? \_\_\_\_\_  
3.NS/Nc

21.Cuál o cuáles tarjetas de crédito posee Ud.?  
(E)

- 1.No tengo tarjeta (IR a 36)    7.Naranja  
2.Argencard/Mastercard      8.Provencred  
3.Visa/Visa Internacional      9.American Express  
4.Diners                            10.Otra. Cuál? \_\_\_\_\_  
5.Bisel                                11.Ns/Nc  
6.Cabal

22. Habitualmente para que tipo de compras utiliza la tarjeta (E)

- 1.Supermercado  
2.Indumentaria  
3.Libros/material de estudio/trabajo  
4.Viajes  
5.Combustible  
6.Todo  
7.Otro. Cuál? \_\_\_\_\_  
8.Ns/Nc

23. Cuántas tarjetas adicionales tiene? \_\_\_\_\_

24.Cuál es su gasto promedio mensual? \$ \_\_\_\_\_  
(Es el gasto por resumen, incluye lo consumido por adicionales. Si tiene más de 1 tarjeta consignar la suma de todos los resúmenes)  
(Si NO responde espontáneamente Mostrar T2)

1	Menos de 50
2	De 50 a 100
3	De 101 a 150
4	De 151 a 200
5	De 201 a 300
6	De 301 a 400
7	De 401 a 500
8	De 501 a 700
9	De 701 a 900
10	De 901 a 1100
11	De 1101 a 1300
12	De 1301 a 1500
13	De 1501 a 1750
14	De 1751 a 2000
15	De 2001 a 2500
16	De 2501 a 3000
17	De 3001 a 4000
18	Más de 4000
19	Ns/Nc

25. Cuánto paga de costo de mantenimiento? (A)  
\$ \_\_\_\_\_

(Si tiene más de una tarjeta consignar el promedio)

26. Considera que es caro? (E)

- 1.SI      2.No      3Ns/Nc

27. Cómo calificaría a su tarjeta de crédito. Ud. diría que el servicio es... (L) (Mostrar T3)

- 1.Muy Bueno      2.Bueno      3.Regular  
4.Malo              5.Muy malo      6.Ns/Nc

28. Qué es lo que hace atractiva a una tarjeta de crédito? (L)

- 1.El emisor  
2.El color que la identifica  
3.El slogan  
4.El costo de mantenimiento y renovación  
5.Los premios o promociones que otorga  
6.Otro, cuál? \_\_\_\_\_  
7.Ns/Nc

29. Qué es lo que NO debe tener una tarjeta de crédito? (A)

30. Le gustaría tener otra tarjeta de.....? (E)

Débito	Crédito
1.SI	1.SI
2.NO	2.NO
3.NS/NC	3.NS/NC

31. Si Ud. tuviera la posibilidad de diseñarla a su gusto: qué color debería tener, y qué servicio debería prestarle, para que Ud. accediera a tenerla? (A)

Color: \_\_\_\_\_ Porqué? \_\_\_\_\_

Servicio: \_\_\_\_\_

32. Actualmente su tarjeta cumple con esa condición? (E)

1.SI 2.No 3.Ns/Nc

33. Acostumbra Ud. a comprar en cuotas? (E)

1.Si. Hasta en cuántas cuotas? \_\_\_\_\_  
2.NO  
3.Ns/Nc

34. Su compra en cuotas la realiza con...(L)

1. Tarjeta de crédito  
2. Financiación directa del comercio  
3. Otra. Cuál? \_\_\_\_\_  
4. Ns/Nc

35. Le resulta relativamente fácil acceder a planes de cuotas

1.SI 2.NO 3.NS/Nc

Porqué? \_\_\_\_\_

**(PASAR A 39)**

36. (dijo no tener tarjeta) Porqué no tiene? (E)

1.No tengo ingreso mínimo necesario  
2.No quiero tarjeta  
3.No me hace falta, trabajo con cheque  
4.Solicitan demasiados requisitos y no alcanzo a ellos  
5.Otro. Cuál? \_\_\_\_\_  
6.Ns/Nc

37. Le gustaría tener tarjeta de débito o crédito? (E)

Débito	Crédito
1.SI	1.SI
2.NO	2.NO
3.NS/NC	3.NS/NC

38. Qué debería tener, o qué servicio debería prestarle, para que Ud. accediera a tenerla? (A)

39. Supongamos que se lanza a nivel local una tarjeta de compra a través de la cual Ud. realiza todas las compras necesarias para su consumo, incluso podría comprar en cuotas y pagar un resumen a fin de mes. Es decir, en lugar de visitar a cada comercio donde tiene un crédito, unificaría todo el gasto en una tarjeta. Le gustaría tenerla? (Mostrar T4)

1.Me gustaría mucho 2.Me gustaría  
3.Me es indistinto 4.No me gustaría  
5.No me gustaría para nada 6.Ns/Nc

40. Si Ud. tuviera que decidir que nombre ponerle a la tarjeta. Cuál le pondría? (A)

41. Dónde recurre Ud. para su atención médica? (E)

1.Hospital Público  
2.Consult.de Médico Privado  
3.Clínica Privada  
4.Otro. Cuál? \_\_\_\_\_  
5.Ns/Nc

42. Cómo paga Ud. su atención médica privada? (E)

1) Pago en efectivo  
2) Paga la obra social  
3) Tiene un plan de medicina prepaga  
4) Tiene plan prepago y obra social  
5) Otros \_\_\_\_\_  
6) Ns/Nc

43. Qué obra social tiene? (A)

44. Que plan de medicina prepaga tiene? (A)

45. Realiza Ud. aportes jubilatorios? (E)

1. Si 2.No 3.Ns/Nc

46. A qué AFJP realiza sus aportes? (E)

1.Consolidar 2.Máxima  
3.Orígenes 4.Nación  
5.Prewinter  
6.Caja de Profesionales  
7.Otra. Cuál? \_\_\_\_\_  
8. Ns/Nc

47. Del total de sus compras, que participación tiene...

1. HHH \_\_\_\_\_%
2. Localidades vecinas \_\_\_\_\_%
3. Ciudades cercanas \_\_\_\_\_%
4. Otra. \_\_\_\_\_% Cuál? \_\_\_\_\_
5. Ns/Nc

48. Qué es lo que compra en HHH? (A) \_\_\_\_\_

49. Por qué compra en los comercios locales? (L)

1. Por buena atención
2. Por higiene
3. Por variedad de productos
4. Por precios convenientes
5. Por trayectoria
6. Por disponibilidad de productos
7. Por servicio post-venta
8. Otros. Cuál? \_\_\_\_\_
9. Ns/Nc

50. Qué opina de la atención en el comercio local (Mostrar T3)

- |              |             |            |
|--------------|-------------|------------|
| 1. Muy Bueno | 2. Bueno    | 3. Regular |
| 4. Malo      | 5. Muy malo | 6. Ns/Nc   |

51. Qué opina de la higiene en el comercio local? (Mostrar T3)

- |              |             |            |
|--------------|-------------|------------|
| 1. Muy Bueno | 2. Bueno    | 3. Regular |
| 4. Malo      | 5. Muy malo | 6. Ns/Nc   |

52. La vivienda en la cual vive es... (L)

- a) Propia (Seguir en 57)
- b) Alquilada (Seguir en 53)
- c) De familiares (Seguir en 55)
- d) Otros (Seguir en 55)
- e) NS/NC

53. Cuánto paga de alquiler por mes? (A)

\$ \_\_\_\_\_

(Si NO responde espontáneamente Mostrar Tabla 5)

1	Menos de 150
2	De 150 a 250
2	De 251 a 350
3	De 351 a 500
4	De 501 a 700
5	De 701 a 1000
6	Más de 1000
7	Ns/Nc

54. Estaría de acuerdo en pagar el mismo importe por una cuota de su propia casa? (Mostrar T6)

- 1) Muy de acuerdo
- 2) De acuerdo
- 3) Ni de acuerdo ni en desacuerdo
- 4) En desacuerdo
- 5) Muy en desacuerdo
- 6) Ns/Nc

**(PASAR A 56)**

55. Estaría de acuerdo en pagar el equivalente a una cuota de alquiler por el plan de una casa o departamento? (Mostrar T6)

- 1) Muy de acuerdo
- 2) De acuerdo
- 3) Ni de acuerdo ni en desacuerdo
- 4) En desacuerdo
- 5) Muy en desacuerdo
- 6) Ns/Nc

56. Que tipo de vivienda prefiere Ud.

(Marcar lo que corresponda)	De cuántos dormitorios?
a) Casa	
b) Departamento	
c) Country	

Las preguntas que siguen se realizan a efectos de poder conocer el nivel económico social de HHH.

57. Cuál es su ocupación principal y la del Jefe de Familia?. (Cuando el entrevistado sea Jefe consignar los datos en Jefe. Cuando el entrevistado sea Inactivo, consignar la actividad que desarrollaba)

Entrevistado	Jefe
<b>Inactivo</b>	
1	Desocupado 1
2	Estudiante 2
3	Ama de Casa 3
4	Jubilado/pensionado 4
<b>Autónomos</b>	
5	Changarín 5
6	Trabajos no especializados 6
7	Comerciante sin personal 7
8	Técnico/artesanos/trabajador especializado 8
9	Profesionales independientes 9
10	Otros autónomos 10
<b>Empleadores</b>	
11	1-5 empleados 11
12	6-20 empleados 12
13	más de 20 empleados 13
<b>Relación de dependencia</b>	
14	Empleada doméstica 14
15	Trabajo familiar 15
16	Obrero no calificado 16
17	Obrero calificado 17
18	Técnico/capataz 18
19	Empleado s/jerarquía del Estado 19
20	Empleado s/jerarquía del Privado 20
21	Jefe intermedio del Estado 21
22	Jefe intermedio Privado 22
23	Gerencia del Estado 23
24	Gerencia Privada 24
25	Alta Dirección del Estado 25
26	Alta Dirección Privada 26
27	Ns/Nc 27
28	Otro. Cuál? _____ 28

58. Qué estudios cursó Usted? (Y el Jefe de familia?)

Entrevistado		Jefe
1	Sin instrucción	1
2	Primario incompleto	2
3	Primario completo	3
4	Secundario incompleto	4
5	Secundario completo	5
6	Terciario incompleto	6
7	Terciario completo	7
8	Universitario incompleto	8
9	Universitario completo	9
10	Estudios de posgrado	10
11	Ns/Nc	11

59. Cuáles son los ingresos mensuales de su grupo familiar? (Mostrar T7)

1	Menos de 150
2	De 150 a 250
3	De 251 a 350
4	De 351 a 450
5	De 451 a 550
6	De 551 a 700
7	De 701 a 850
8	De 851 a 1000
9	De 1001 a 1250
10	De 1251 a 1500
11	De 1501 a 2000
12	De 2001 a 2500
13	De 2501 a 3000
14	De 3001 a 4000
15	Más de 4000
16	Ns/Nc

60. Usted tiene.... (L)

	S	N		S	N
	I	O		I	O
Computadora			Gas por redes		
Computadora portátil			Agua potable		
Agenda electrónica			Energía eléctrica		
Conexión a Internet			Cloacas		
Dirección de correo electrónico			Teléfono inalámbrico		
Televisión			Teléfono fijo		
TV por cable			Celular		
Contestador automático			Alarma en vivienda		
Fax			Teléfono en el auto		
Video grabador			Aire acondicionado		
Satélite receptor			Horno microondas		
Lectora de CD room			Radio		
Tarjeta de crédito o débito			Filmadora		
Auto 1 Marca _____ Modelo _____			Auto 2 Marca _____ Modelo _____		

61. Es Ud. jefe de familia?

1. Sí
2. No
3. Ns/Nc

62. Perciben una parte de los ingresos en ticket?

1. SI.
2. NO (Seguir en 67)
3. NS/NC

63. Cuánto representa aproximadamente en dinero? \$ \_\_\_\_\_ (Si no responde espontáneamente mostrar T8)

1	Menos de 25
2	De 25 a 50
3	De 51 a 100
4	De 101 a 150
5	De 151 a 200
6	De 201 a 300
7	De 301 a 400
8	Más de 400
9	Ns/Nc

64. Los tickets que reciben en su hogar a qué emisor pertenecen?

1. Ticket total
2. Luncheon Check
3. Ticket Proms
4. Ticket Super compras
5. Ticket canasta
6. Ticket Premium
7. Ticket Restaurant
8. Ticket Total Restaurant
9. Ticket Proms Restaurant
10. Otro. Cuál? \_\_\_\_\_
11. Ns/Nc

65. Qué opinión le merece el sistema de tickets? (Mostrar T3)

1. Muy Bueno
2. Bueno
3. Regular
4. Malo
5. Muy Malo
6. Ns/Nc

66. Ha tenido en alguna oportunidad problemas para comprar con tickets?

1. Sí. Por qué motivo? \_\_\_\_\_
2. No
3. Ns/Nc

67. Una última preguntita y terminamos, cuántas personas viven en el hogar? \_\_\_\_\_

*MUCHAS GRACIAS POR COLABORAR  
CON NOSOTROS*

*Antes de iniciar la próxima entrevista el  
encuestador deberá observar el nivel de  
vivienda*



### Características para identificar el nivel de vivienda

1. **Lujo.** viviendas de personas de alto poder adquisitivo; casas, chalets, departamentos de piso o semipiso de categoría. Petit hotel. En general son casas (de una o dos plantas) o departamentos en barrios residenciales con garaje, cochera o jardín. El frente de los terrenos suele ser de 15 mts.
2. **Muy buena calidad:** vivienda de clase media alta, casa o departamento de muy buen aspecto exterior. Casa aislada o chalet con garaje o jardín
3. **Buena calidad:** casa o departamento de buen aspecto exterior. Semipisos o departamentos con ascensor, entrada de servicio interna. Chalet con jardín al frente, al fondo o circundando la vivienda. Puede poseer garaje o cochera, en el caso de los departamentos –esta característica no es excluyente-. En general, es una construcción moderna pero no lujosa.
4. **Regular calidad:** Vivienda urbana de menor valor; casas o departamentos deteriorados, en mal estado de mantenimiento. Viviendas de planes habitacionales tipo. Casas de corredor abierto (departamento en fondo) de una planta.
5. **Mala calidad:** viviendas humildes; casa antigua deteriorada, departamentos en edificios sin ascensor. Viviendas en planes habitacionales de mala calidad.
6. **Extremadamente de mala calidad:** villa miseria, conventillos, viviendas espontáneas (construidas con materiales de descarte), ranchos rurales con pisos de tierra, sin servicios en el interior de los mismos.

OBSERVACIONES:

---

---

---



Capítulo 4. DISTRIBUCIONES EN EL MUESTREO.....	117
4.1 Distribución en el muestreo de la media .....	118
4.2 Función generatriz de momentos de la distribución de la media y de la población.....	121
¿Cómo se distribuye $\bar{X}$ si $X \sim N(0,1)$ ?.....	124
¿Cómo se distribuye $\bar{X}$ si $X \sim N(\mu_x, \sigma_x)$ ? .....	126
¿Cómo se distribuye $\bar{X}$ si X tiene otra distribución? .....	130
4.3 Factor de corrección para poblaciones finitas.....	138
4.4 Distribución Gamma .....	142
4.5 Distribución chi-cuadrado ( $\chi^2$ ) de Pearson .....	144
Propiedad reproductiva de $\chi^2$ .....	146
4.6 Distribución F de Snedecor o Fischer .....	147
4.7 Relaciones entre las distribuciones Normal y $\chi^2$ .....	150
4.8 Distribución T de Student .....	154
4.9 Los grados de libertad .....	156
 CASOS DE ESTUDIO, PREGUNTAS Y PROBLEMAS .....	 157
Caso 4.1: Edades y estaturas del curso Inferencia Estadística. ..	157
Caso 4.2: Ingreso medio en los hogares de Río Cuarto .....	158

Preguntas .....	159
Problemas .....	159
Bibliografía .....	160

### Capítulo 4. DISTRIBUCIONES EN EL MUESTREO

Hasta aquí se ha desarrollado el concepto de probabilidad y sus distribuciones asociadas, ahora utilizando aquella base se comenzará con el enfoque de la forma en que los estadísticos de la muestra se pueden utilizar para hacer inferencias sobre los parámetros de la población.

Es necesario darse cuenta de que el investigador estadístico, a través de los datos recolectados en una encuesta, se interesa en sacar conclusiones de la población y no de la muestra. Por ejemplo, un encuestador político se interesa en los resultados de la muestra solo como medio para estimar la proporción real de votos que recibirá cada candidato entre la población de votantes.

En la práctica, una muestra individual de tamaño determinado se selecciona aleatoriamente entre la población. Las unidades muestrales se van a obtener mediante, por ejemplo, una tabla de números aleatorios. A fin de poder utilizar la media de la muestra para estimar la media de la población se deberían examinar todas las muestras posibles de la población y calcular una media para cada muestra. La distribución de estos resultados se denomina distribución en el muestreo, en la práctica con la selección de una muestra, y utilizando la teoría de la probabilidad, se hacen inferencias a la población.

La inferencia estadística es uno de los objetivos de la investigación estadística. La inferencia se hace utilizando estadísticos de la muestra, que son "mediciones" que se realizan en la muestra para estimar las mismas "mediciones" de la población que se denominan parámetros.

El investigador estadístico, a través de las medidas que obtiene de una muestra, busca sacar conclusiones respecto de la población y no respecto de la muestra.

**Ejemplo.** Un encuestador político se interesa en los resultados de la muestra sólo como medio para estimar la proporción de votos que recibirá cada candidato entre la población de votantes.

En la siguiente tabla se observa la relación entre estadísticos y parámetros

Característica	Parámetros	Estimador
Media	$\mu_x$	$\bar{X}$
Diferencia entre las medias de dos poblaciones	$\mu_{x_1} - \mu_{x_2}$	$\bar{X}_1 - \bar{X}_2$
Proporción	$\pi$	$p$
Varianza	$\sigma_X^2$	$s_X^2$
Desviación estándar	$\sigma_X$	$s_X$

**Ejemplos.**

El estadístico  $\bar{X}$ , que surge de una muestra, es estimador del parámetro  $\mu_X$ , en la población.

El estadístico  $\bar{X}_1 - \bar{X}_2$ , diferencia de las medias provenientes de dos muestras, es estimador del parámetro  $\mu_{X_1} - \mu_{X_2}$ , que indica la diferencia en las medias proveniente de dos poblaciones.

En la práctica, una muestra individual de tamaño determinado se selecciona aleatoriamente entre la población. Las unidades muestrales se obtienen mediante, por ejemplo, una tabla de números aleatorios. En el caso hipotético, a fin de poder utilizar la media de la muestra para estimar la media de la población se deben examinar todas las muestras posibles de la población y calcular una media para cada muestra. La distribución de estos resultados se denomina distribución en el muestreo.

Aunque en la práctica se ha seleccionado solo una muestra, se debe examinar el concepto de la distribución en el muestreo, con el propósito de poder utilizar la teoría de la probabilidad para hacer inferencias en cuanto a los valores de la población.

En este módulo se estudiarán las distribuciones en el muestreo sin estudiar específicamente como se obtiene una muestra aleatoria, tema que será abordado más adelante.

#### 4.1 Distribución en el muestreo de la media

Existen diversas medidas con que se pueden caracterizar los datos de la muestra. Dentro de las medidas de posición, tanto el modo como la mediana y la media, calculadas de los datos muestrales, podrían ser considerados estadísticos útiles para estimar la media poblacional.

No obstante, el mejor estimador para inferir la media poblacional, es la media aritmética obtenida de los datos muestrales. Esto es así pues, este estadístico, tiene tres propiedades deseables y muy importantes en todo proceso de estimación: es insesgado, eficiente y consistente.

La media aritmética, obtenida de los datos de una muestra aleatoria, es una variable aleatoria, ya que se ha calculado a partir de un proceso de estimación aleatorio. Por lo tanto, al ser una variable aleatoria tiene una distribución de probabilidad, una esperanza matemática y una varianza. Esta propiedad es propia de todo estadístico calculado con datos provenientes de una muestra aleatoria.

Para el caso particular de la media aritmética de los datos de una muestra aleatoria, se demostrará empíricamente que

$$E(\bar{X}) = \mu_X$$

Esto es, la esperanza matemática de la media aritmética de los datos obtenidos a partir de una muestra aleatoria, es igual al parámetro poblacional que se está estimando (la media de la población). Al cumplir

con esta propiedad, se dice que la media aritmética es un estimador insesgado de la media poblacional.

Por lo tanto, si se toma una cantidad determinada de muestras aleatorias de una población,  $n$  muestras, y de cada una se calcula, a partir de los datos observados de una variable, una media aritmética, se obtendrán  $n$  medias aritméticas muestrales. La esperanza matemática de todas estas medias aritméticas muestrales es la media de la población,

$$\sum_{i=1}^n \frac{\bar{X}_i}{n} = \mu_X$$

**Experimento.** Bajo el supuesto de que la variable  $X$  representa el número de visitas al dentista en un año determinado por individuos de una cierta población, y que es una variable aleatoria discreta con distribución uniforme. Esto es,  $X$  puede asumir solo valores enteros y cada uno de estos valores se observa con la misma probabilidad.

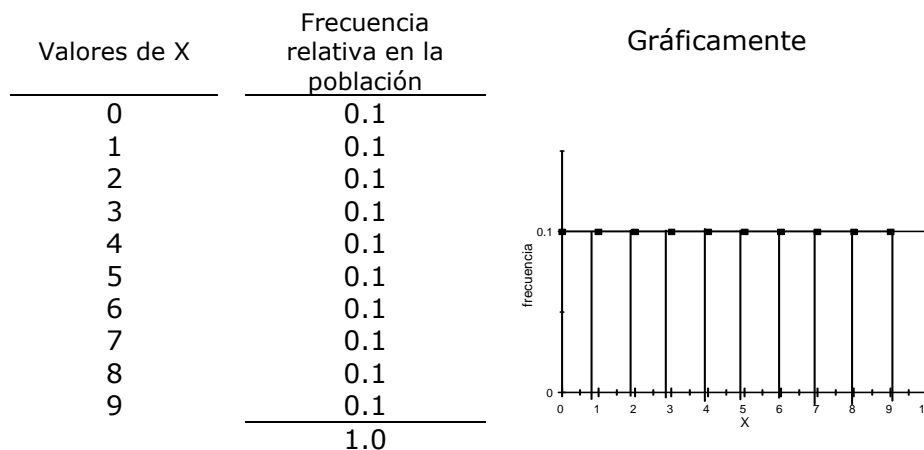
Identificando con  $a$  el valor mínimo y con  $b$  el valor máximo de  $X$ , la media de una distribución uniforme discreta es

$$E(X) = \mu_X = \frac{a+b}{2} = \frac{0+9}{2} = 4.5$$

y la desviación estándar:

$$\sigma_X = \sqrt{\frac{[(b-a)+1]^2 - 1}{12}} = \sqrt{\frac{[(9-0)+1]^2 - 1}{12}} = 2.87$$

De modo que la media y la desviación estándar de la población bajo estudio asumen los valores 4.5 y 2.87.



De esta población se extraen 100 muestras posibles de tamaño 5. Es decir, se seleccionan 5 personas al azar de la población y se le pregunta cuántas veces visitó al dentista en el año. En una de las muestras las respuestas de las personas fueron:

PERSONA	NUMERO DE VISITAS
Primera	2
Segunda	0
Tercera	4
Cuarta	1
Quinta	1

Sobre esta muestra se calcula la media aritmética:

$$\frac{(2 + 0 + 4 + 1 + 1)}{5} = 1.6$$

Este experimento se repitió 100 veces y se calcularon 100 medias aritméticas. Los resultados del experimento se muestran en el Cuadro II-2:

Cuadro II.2			
Valores de la Media de la muestra X		Frecuencia	
Intervalo	Punto Medio	Absoluta	Relativa
0.5 a 1.499	1	1	0.01
1.5 a 2.499	2	5	0.05
2.5 a 3.499	3	12	0.12
3.5 a 4.499	4	31	0.31
4.5 a 5.499	5	28	0.28
5.5 a 6.499	6	15	0.15
6.5 a 7.499	7	5	0.05
7.5 a 8.499	8	3	0.03
8.5 a 9.499	9	0	0.00
		100	1.00

Aunque la variable ( $y$ , por lo tanto, los datos individuales) solo pueden tomar valores enteros entre 0 y 9, las medias de las muestras no serán, generalmente, enteros y la distribución muestral será una distribución de frecuencias cuyas clases estarán definidas por intervalos y no por puntos. Se sabe que para representar cada clase se puede elegir un valor, como, por ejemplo, el centro de cada intervalo.

Las principales características de esta distribución son

$$\begin{aligned} \text{Media} &= \sum_{i=1}^9 \bar{X}_i f_i = 1(0.01) + 2(0.05) + 3(0.12) + 4(0.31) + 5(0.28) \\ &\quad + 6(0.15) + 7(0.05) + 8(0.03) + 9(0.00) = 4.60 \end{aligned}$$

$$\text{Desviación estándar} = \sqrt{\sum_{i=1}^9 (\bar{X}_i - 4.60)^2 f_i} = 1.3638$$

Los resultados indican que un 59% de los valores estimados (0.31+0.28) están comprendidos en el intervalo  $\pm 1$  alrededor del valor real 4.5 y que el 86% de las estimaciones están comprendidas en el intervalo  $\pm 2$  alrededor de dicho valor.

Realizado el mismo experimento con la variante de tomar 100 muestras de tamaño 10 se obtuvo la siguiente característica

$$\text{Media} = \sum_{i=1}^9 \bar{X}_i f_i = 4.57,$$

$$\text{Desviación estándar} = \sqrt{\sum_{i=1}^9 (\bar{X}_i - 4.57)^2 f_i} = 1.0416$$



Los resultados del experimento permiten realizar las siguientes generalizaciones

- a) la media de la muestra,  $\bar{X}$ , es un estimador insesgado de la media de la población,  $\mu_X$
- b) a medida que aumenta el tamaño de la muestra, la distribución muestral de  $\bar{X}$  se va concentrando cada vez más en torno a la media de la población y, por tanto,  $\bar{X}$  es un estimador consistente de  $\mu_X$
- c) el estimador  $\bar{X}$  es más estable de una muestra a otra que la mediana o el modo, esto lo convierte en un estimador eficiente.

Queda claro que se ha trabajado con distribuciones muestrales experimentales. Si se lo hubiese hecho con las distribuciones teóricas se hubieran obtenido idénticos resultados pero con mucho menos trabajo ya que sólo se tomaría una muestra.

#### 4.2 Función generatriz de momentos de la distribución de la media y de la población

¿A que es igual la función generatriz de momentos de la suma de variables aleatorias independientes  $X_1, X_2, \dots, X_n$ ?

*Variable aleatoria independiente:*

*Cuando la probabilidad de ocurrencia de una variable no depende de la probabilidad de ocurrencia de la otra.*

La función generatriz de momentos de una variable aleatoria continua es:

$$M_x = E(e^{\theta x}) = \int_{-\infty}^{\infty} e^{\theta x} f(x) dx \quad (1)$$

Si la variable aleatoria es discreta, se expresa

$$M_x = E(e^{\theta x}) = \sum_x e^{\theta x} f(x)$$

( $M_x$ , o  $M_x(\theta)$  donde  $\theta$  es el parámetro, simboliza la función generatriz de momentos de una variable aleatoria)

La función generatriz de momentos de la sumatoria de variables aleatorias independientes en el parámetro  $\theta$

$$M_{X_1+X_2+\dots+X_n}(\theta) = E\left\{ e^{\theta(x_1+x_2+\dots+x_n)} \right\} \quad (2)$$

Aplicando la propiedad distributiva

$$E\left\{ e^{\theta x_1} e^{\theta x_2} e^{\theta x_3} \dots e^{\theta x_n} \right\} \quad (3)$$

Dado que las variables son independientes lo anterior puede expresarse

$$E(e^{\theta x_1}) E(e^{\theta x_2}) \dots E(e^{\theta x_n}) \quad (4)$$

Por lo expresado en (1) se sabe que

$$E(e^{\theta x}) = M_x$$

Si tenemos en cuenta (1) en (4) tendremos

$$M_{X_1}(\theta) M_{X_2}(\theta) \dots M_{X_n}(\theta) \quad (5)$$

Pero en muestreo aleatorio las características de la población se obtienen a partir de las características de la muestra

$$M_{X_i}(\theta) = M_x^n(\theta) \quad (6)$$

Por ende  $M_x^n(\theta)$  es la función generatriz de momento elevado a la potencia n.

En síntesis, la función generatriz de momentos de una suma de variables independientes

$$M_{X_1+X_2+\dots+X_n}(\theta)$$

es el producto de las n funciones generatrices de momentos del parámetro  $\theta$  para las n variables

$$M_x(\theta) M_x(\theta) \dots M_x(\theta)$$

y esto es igual a la función generatriz de momentos del parámetro  $\theta$  elevado a la  $n$

$$M_X^n(\theta)$$

$$M_{X_1+X_2+\dots+X_n}(\theta) = M_X(\theta) M_X(\theta) \dots M_X(\theta) = M_X^n(\theta) \quad (7)$$

La función generatriz de momentos del producto de una constante por una variable para el parámetro  $\theta$

$$M_{cX}(\theta)$$

Es la función generatriz de momentos del producto del parámetro ( $\theta$ ) por la constante ( $c$ ) para la variable  $X$

$$M_{cX}(\theta) = M_X(c\theta) \quad (8)$$

Con (7) y (8) se puede hallar la función generatriz de momentos de la media de la muestra en el parámetro  $\theta$  ( $M_{\bar{X}}(\theta)$ )

$$M_{\bar{X}}(\theta) = M_{\frac{1}{n}(X_1+X_2+\dots+X_n)}(\theta) = M_{(X_1+X_2+\dots+X_n)}\left(\frac{\theta}{n}\right)$$

Con este ultimo resultado y teniendo en cuenta (7)

$$M_{\bar{X}}(\theta) = M_X^n\left(\frac{\theta}{n}\right) \quad (9)$$

La función generatriz de momentos de la  $\bar{X}$  en el parámetro ( $\theta$ ) es la función generatriz de momentos de la variable  $X$  elevado a la  $n$  en el parámetro  $\theta/n$

La expresión (9) relaciona la función generatriz de momentos de la media muestral y de la población.

### 4.3 Distribuciones en el muestreo de la media

- 1) Una población de valores de  $X$
- 2) Muestras aleatorias de tamaño  $n$  de donde se obtiene  $\bar{X}$  (de cada muestra)

¿Cómo se distribuye  $\bar{X}$ ?

Veremos que pasa si

$$X \sim N(0,1)$$

$$X \sim N(\mu_x, \sigma_x)$$

$X \sim (\mu_x, \sigma_x)$  con cualquier tipo de distribución de probabilidad

¿Cómo se distribuye  $\bar{X}$  si  $X \sim N(0,1)$ ?

Si la variable aleatoria  $X \sim N(\mu_x, \sigma_x)$  su función de densidad es

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{\left[-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2\right]}; \quad -\infty < X < \infty$$

Si definimos  $Z = \frac{X - \mu}{\sigma}$  donde  $Z \sim N(0,1)$  entonces

$$f(Z) = \frac{1}{\sqrt{2\pi}} e^{\left(-\frac{1}{2}Z^2\right)}$$

La función generatriz de momentos de la variable  $Z$  es (usando (1))

$$M_z(\theta) = E\{e^{\theta \cdot Z}\} = \int_{-\infty}^{\infty} e^{\theta \cdot Z} f(z) dz \quad (10)$$

Reemplazando  $f(z)$  por su igual

$$M_z(\theta) = \int_{-\infty}^{\infty} e^{\theta \cdot z} \frac{1}{\sqrt{2\pi}} e^{\left(-\frac{1}{2}z^2\right)} dz = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{\theta \cdot z} e^{\left(-\frac{z^2}{2}\right)} dz \quad (11)$$

Observemos  $e^{\theta Z} \cdot e^{-\frac{Z^2}{2}}$

Se tiene un producto de exponente de igual base, el resultado de este producto es conservar la base y sumar los exponentes

$$e^{\theta Z - \frac{Z^2}{2}} \quad (12)$$

A fin de encontrar una expresión menos compleja, se suma y resta  $\frac{1}{2}\theta^2$  en el exponente de e

$$\theta Z - \frac{Z^2}{2} + \frac{1}{2}\theta^2 - \frac{1}{2}\theta^2$$

Al ordenar convenientemente y sacar factor común  $-\frac{1}{2}$

$$-\frac{1}{2}(\theta^2 - 2\theta Z + Z^2) + \frac{1}{2}\theta^2$$

$$-\frac{1}{2}(-2\theta Z + Z^2 + \theta^2) + \frac{1}{2}\theta^2$$

Se tiene el desarrollo del cuadrado

$$-\frac{1}{2}(Z - \theta)^2 + \frac{1}{2}\theta^2$$

En síntesis:

$$\theta Z - \frac{Z^2}{2} = -\frac{1}{2}(Z - \theta)^2 + \frac{1}{2}\theta^2 \quad (13)$$

Reemplazando (13) en (12)

$$e^{\theta \cdot Z} e^{-\frac{Z^2}{2}} = e^{\left[-\frac{1}{2}(Z - \theta)^2 + \frac{1}{2}\theta^2\right]} = e^{-\frac{1}{2}(Z - \theta)^2} e^{\frac{1}{2}\theta^2} \quad (14)$$

El resultado de (14) en (11) da lugar a una nueva expresión de la función generatriz de momentos de la variable Z en el parámetro  $\theta$

$$M_Z(\theta) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{-\frac{1}{2}(Z - \theta)^2} e^{\frac{1}{2}\theta^2} dz = \frac{e^{\left(\frac{1}{2}\theta^2\right)}}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{-\frac{1}{2}(Z - \theta)^2} dz \quad (15)$$

La diferencia entre (15) y (11) radica en la expresión a integrar, (15) es más sencilla y se resuelve sustituyendo variables. Para lo cual se define  $Z - \theta = t$

Entonces si  $t = z - \theta$   
 $dt = dz$  (porque  $\theta$  es constante)

Remplazando en (15)

$$M_Z(\theta) = \frac{e^{\left(\frac{1}{2}\theta^2\right)}}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{-\frac{1}{2}t^2} dt$$

Esta última expresión tiene la ventaja de contener una integral notable cuyo resultado es  $\sqrt{2\pi}$

$$\int_{-\infty}^{\infty} e^{-\frac{1}{2}t^2} dt = \sqrt{2\pi} \quad (16)$$

Entonces

$$M_Z(\theta) = \frac{e^{\left(\frac{1}{2}\theta^2\right)}}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{-\frac{1}{2}t^2} dt = \frac{e^{\left(\frac{1}{2}\theta^2\right)}}{\sqrt{2\pi}} \sqrt{2\pi} = e^{\left(\frac{1}{2}\theta^2\right)}$$

En síntesis, la función generatriz de momentos de la variable Z en el parámetro  $\theta$  es una exponencial con exponente  $\theta^2/2$

$$M_Z(\theta) = e^{\left(\frac{\theta^2}{2}\right)} \quad (17)$$

¿Cómo se distribuye  $\bar{X}$  si  $X \sim N(\mu_x, \sigma_x)$ ?

La variable normal estándar Z es una variable aleatoria normal centrada y reducida<sup>1</sup>.

$$Z = \frac{x - \mu_x}{\sigma_x}$$

Si se multiplica en ambos miembros por  $\sigma_x$ , la diferencia entre la observación de la variable y la media es la medida estándar por el desvío:

$$Z \cdot \sigma_x = x - \mu_x \quad (18)$$

<sup>1</sup> Una variable centrada es aquella a la que se le ha restado, a cada valor, el de su media; y es reducida cuando se la divide por su desvío.

La función generatriz de la variable aleatoria normal centrada  $x - \mu_x$ , en el parámetro  $\theta$ , es la función generatriz de momentos de la medida estándar por el desvío en el parámetro  $\theta$

$$M_{x-\mu_x}(\theta) = M_{z\sigma_x}(\theta) \quad (\text{donde } \sigma_x \text{ es constante})$$

Si se tiene en cuenta lo visto en (8), donde  $M_{cX}(\theta) = M_X(c\theta)$ , la función generatriz de momentos de la variable centrada  $x - \mu_x$  en el parámetro  $\theta$ , es la función generatriz de momentos de la medida estándar por el desvío en el parámetro  $\theta$ , que es igual a la función generatriz de momentos de la variable estándar Z en el parámetro  $\sigma_x\theta$

$$M_{x-\mu_x}(\theta) = M_{z\sigma_x}(\theta) = M_Z(\sigma_x\theta)$$

Por (17) se sabe que  $M_Z(\theta) = e^{\frac{\theta^2}{2}}$  la función generatriz de momentos de la variable Z en el parámetro  $\theta$  es  $e^{\theta^2/2}$

Entonces

$$M_Z(\theta\sigma_x) = e^{\left(\frac{(\theta\sigma_x)^2}{2}\right)} = e^{\left[\frac{1}{2}(\theta\sigma_x)^2\right]}$$

Es decir, la función generatriz de momentos de la variable normal aleatoria centrada  $x - \mu_x$  en el parámetro  $\theta$  es

$$M_{x-\mu_x}(\theta) = e^{\left[\frac{1}{2}(\theta\sigma_x)^2\right]} \quad (19)$$

La variable aleatoria Z fue reexpresada de acuerdo a (18), si se suma  $\mu_x$  en ambos miembros se tiene la expresión de la variable aleatoria X

$$X = Z\sigma_x + \mu_x \quad (20)$$

La función generatriz de momentos de una variable aleatoria x en el parámetro  $\theta$ , teniendo en cuenta (20), es la función generatriz de momentos de la suma de un producto (la variable Z ponderada por el desvío) y una constante en el parámetro  $\theta$

$$M_X(\theta) = M_{z\sigma_x + \mu_x}(\theta) \quad (21)$$

Pero, teniendo en cuenta las propiedades de función generatriz de momentos

$$M_{c+x}(\theta) = e^{c\theta} M_x(\theta) \quad (22)$$

$$M_{cX}(\theta) = M_X(c\theta)$$

Se observa que esta última expresión es la indicada en (8)

La expresión (21) es igual a

$$M_X(\theta) = M_{Z\sigma_X + \mu_X}(\theta) = e^{\mu_X\theta} M_Z(\sigma_X\theta) \quad (23)$$

La función generatriz de momentos de la variable X en el parámetro  $\theta$  es el producto de una constante  $e^{\mu_X\theta}$  y la función generatriz de momentos de una variable Z en el parámetro  $\sigma_X\theta$

Siguiendo a (17)

$$M_Z(\sigma_X\theta) = e^{\frac{(\sigma_X\theta)^2}{2}}$$

Reemplazando en (23) se tiene

$$\Rightarrow M_X(\theta) = e^{\mu_X\theta} e^{\frac{(\sigma_X\theta)^2}{2}} = e^{\left(\mu_X\theta + \frac{1}{2}\sigma_X^2\theta^2\right)} \quad (24)$$

Esto es la función generatriz de momentos de la variable aleatoria X normal de media  $\mu_X$  y desviación estándar  $\sigma_X$

Con  $X \sim N(\mu_X, \sigma_X)$  y la expresión hallada en (9)

$$M_{\bar{X}}(\theta) = M_X^n\left(\frac{\theta}{n}\right)$$

y sustituyendo por (24)

$$M_{\bar{X}}\left(\frac{\theta}{n}\right) = \left[ e^{\mu_X \frac{\theta}{n} + \frac{1}{2}\sigma_X^2 \frac{\theta^2}{n^2}} \right]^n$$

Por lo tanto, la función generatriz de momentos de  $\bar{X}$  cuando X es normal

$$M_{\bar{X}}(\theta) = \left[ e^{\mu_X \frac{\theta}{n} + \frac{1}{2}\sigma_X^2 \frac{\theta^2}{n^2}} \right]^n = \left[ e^{\mu_X \frac{\theta}{n}} e^{\frac{1}{2}\sigma_X^2 \frac{\theta^2}{n^2}} \right]^n = \left( e^{\mu_X \frac{\theta}{n}} \right)^n \left( e^{\frac{1}{2}\sigma_X^2 \frac{\theta^2}{n^2}} \right)^n = e^{\mu_X \frac{\theta}{n} n} e^{\frac{1}{2}\sigma_X^2 \frac{\theta^2}{n^2} n}$$

$$M_{\bar{X}}(\theta) = e^{\mu_X\theta} e^{\frac{1}{2}\sigma_X^2 \frac{\theta^2}{n}} = e^{\mu_X\theta + \frac{1}{2}\sigma_X^2 \frac{\theta^2}{n}} \quad (25)$$



a la función generatriz de momentos de  $\bar{X}$  cuando  $X$  es normal.

Al comparar las funciones generatrices de momentos de las variables  $X$  y  $\bar{X}$

$$M_X(\theta) = e^{\theta\mu_x + \frac{1}{2}\theta^2\sigma_x^2} \quad (24)$$

$$M_{\bar{X}}(\theta) = e^{\theta\mu_x + \frac{1}{2}\theta^2\frac{\sigma_x^2}{n}} \quad (25)$$

Se acepta el teorema que dice: "Si la variable aleatoria  $X$  asociada a una población tiene distribución normal de media  $\mu_x$  y desviación estándar  $\sigma_x$ , entonces la variable aleatoria  $\bar{X}$  asociada a la distribución de medias de las muestras también es normal, de media:

$$E(\bar{X}) = \mu_{\bar{X}} = \mu_X$$

Y desvío:

$$\sigma_{\bar{X}} = \frac{\sigma_X}{\sqrt{n}}$$

Para demostrar el teorema calculemos el momento de orden 1 y el momento de orden 2. Los momentos se calculan derivando la función generatriz de momentos y evaluando  $\theta$  en 0. Entonces,

$$M_1 = \frac{\partial M_{\bar{X}}(\theta)}{\partial \theta} = e^{\mu_x \theta + \frac{1}{2}\theta^2\frac{\sigma_x^2}{n}} \left( \mu_x + \frac{\sigma_x^2 \theta}{n} \right) \quad (26)$$

Evaluando en  $\theta = 0$

$$M_1 = e^{\mu_x \cdot 0 + \frac{1}{2}\sigma_x^2\frac{0^2}{n}} \left( \mu_x + \frac{\sigma_x^2}{n} \cdot 0 \right) = \mu_x$$

De modo que el momento de orden 1 es

$$M_1 = E(\bar{X}) = \mu_X$$

$$M_2 = \frac{\partial^2 M_{\bar{X}}(\theta)}{\partial \theta^2} = e^{\mu_X \theta + \frac{1}{2} \theta^2 \frac{\sigma_X^2}{n}} \left( \mu_X + \frac{\sigma_X^2}{n} \theta \right)^2 + e^{\mu_X \theta + \frac{1}{2} \theta^2 \frac{\sigma_X^2}{n}} \frac{\sigma_X^2}{n} \quad (27)$$

Evaluando en  $\theta = 0$

$$M_2 = e^0 \left( \mu_X + \frac{\sigma_X^2}{n} 0 \right)^2 + e^0 \frac{\sigma_X^2}{n}$$

Entonces el momento centrado de orden 2 es

$$M_2 = \mu_X^2 + \frac{\sigma_X^2}{n}$$

La varianza de  $\bar{X}$  es,

$$V(\bar{X}) = \sigma_{\bar{X}}^2 = M_2 - M_1^2 = \mu_X^2 + \frac{\sigma_X^2}{n} - \mu_X^2 = \frac{\sigma_X^2}{n}$$

Y el desvío de  $\bar{X}$  será,

$$\sigma_{\bar{X}} = \sqrt{\frac{\sigma_X^2}{n}} = \frac{\sigma_X}{\sqrt{n}}$$

¿Cómo se distribuye  $\bar{X}$  si X tiene otra distribución?

Es decir, que pasa si X tiene alguna distribución distinta a la normal pero tiene una  $\mu_X$  y un  $\sigma_X$ .

Denominemos

$$t = \frac{\bar{X} - \mu_X}{\sigma_{\bar{X}}} = \frac{\bar{X} - \mu_X}{\sigma_X / \sqrt{n}}$$

A una variable aleatoria estandarizada correspondiente a la distribución de medias de las muestras.

La función generatriz de momentos de  $t$

$$M_t(\theta) = M_{\frac{\bar{X} - \mu_x}{\sigma_x / \sqrt{n}}}(\theta) \quad \text{donde } \sigma_x / \sqrt{n} \text{ es constante}$$

Aplicando lo visto en (8) que  $M_{CX}(\theta) = M_X(C\theta)$

$$M_t(\theta) = M_{\bar{X} - \mu_x} \left( \frac{\theta}{\sigma_x / \sqrt{n}} \right) \quad \text{donde } \mu_x \text{ es constante}$$

En (22) vimos que  $M_{C+X}(\theta) = e^{C\theta} M_X(\theta)$

Con lo cual

$$M_t(\theta) = e^{\frac{-\mu_x \theta}{\sigma_x / \sqrt{n}}} M_{\bar{X}} \left( \frac{\theta}{\sigma_x / \sqrt{n}} \right)$$

Pero en (9) vimos que  $M_{\bar{X}}(\theta) = M_X^n \left( \frac{\theta}{n} \right)$

Entonces,

$$M_t(\theta) = e^{\frac{-\mu_x \theta \sqrt{n}}{\sigma_x}} M_X^n \left( \frac{\theta / n}{\sigma_x / \sqrt{n}} \right) = e^{\frac{-\mu_x \theta \sqrt{n}}{\sigma_x}} M_X^n \left( \frac{\theta}{\sigma_x \sqrt{n}} \right) \quad (28)$$

Esto es así porque

$$\frac{\theta/n}{\sigma_X/\sqrt{n}} = \frac{\theta\sqrt{n}}{\sigma_X n} = \frac{\theta n^{1/2}}{\sigma_X n} = \frac{\theta n^{1/2-1}}{\sigma_X} = \frac{\theta n^{-1/2}}{\sigma_X} = \frac{\theta}{\sigma_X \sqrt{n}}$$

Tomemos logaritmo natural en (28)

$$\ln[M_t(\theta)] = \frac{-\mu_X \theta \sqrt{n}}{\sigma_X} + n \cdot \ln \left[ M_X \left( \frac{\theta}{\sigma_X \sqrt{n}} \right) \right] \quad (29)$$

Tengamos en cuenta que

$$M_X(\theta) = E\{e^{\theta X}\}$$

Y que  $e^{\theta X}$  podemos desarrollarlo en series de potencias

$$e^w = 1 + w + \frac{w^2}{2!} + \frac{w^3}{3!} + \dots$$

Entonces

$$M_X(\theta) = E\left\{1 + \theta X + \frac{(\theta X)^2}{2!} + \frac{(\theta X)^3}{3!} + \dots\right\}$$

La esperanza es distributiva respecto de la suma

$$M_X(\theta) = E\{1\} + E\{\theta X\} + E\left\{\frac{(\theta X)^2}{2!}\right\} + E\left\{\frac{(\theta X)^3}{3!}\right\} + \dots$$

Pero luego  $E\{CX\} = C \cdot E\{X\}$

$$M_X(\theta) = E\{1\} + \theta E\{X\} + \frac{\theta^2}{2!} E\{X^2\} + \frac{\theta^3}{3!} E\{X^3\} + \dots$$

De modo que la función generatriz de la variable X es

$$M_X(\theta) = 1 + \theta \mu'_1 + \frac{\theta^2}{2!} \mu'_2 + \frac{\theta^3}{3!} \mu'_3 + \dots \quad (30)$$

Donde se ha tenido en cuenta que

$$E\{1\} = E\{C\} = C = 1$$

Y que  $E\{X\} = \mu'_1$  es el momento con respecto al origen de la variable aleatoria  $X$  de órdenes crecientes.

Derivando sucesivamente la expresión (30)

$$\begin{aligned} \frac{d}{d\theta} M_X(\theta) &= \frac{d}{d\theta} \left[ 1 + \theta\mu'_1 + \frac{\theta^2}{2!} \mu'_2 + \frac{\theta^3}{3!} \mu'_3 + \frac{\theta^4}{4!} \mu'_4 + \dots \right] \\ &= 0 + \mu'_1 + \frac{2}{2} \theta\mu'_2 + \frac{3}{3 \cdot 2} \theta^2 \mu'_3 + \frac{4}{4 \cdot 3 \cdot 2} \theta^3 \mu'_4 + \dots \end{aligned}$$

Si  $\theta = 0$  entonces  $\frac{d}{d\theta} M_X(\theta) = \mu'_1$

Busquemos ahora

$$\begin{aligned} \frac{d^2}{d\theta^2} M_X(\theta) &= \frac{d}{d\theta} \left[ \mu'_1 + \frac{2}{2} \theta\mu'_2 + \frac{3}{3 \cdot 2} \theta^2 \mu'_3 + \frac{4}{4 \cdot 3 \cdot 2} \theta^3 \mu'_4 + \dots \right] \\ &= 0 + \mu'_2 + \frac{2}{2} \theta\mu'_3 + \frac{3}{3 \cdot 2} \theta^2 \mu'_4 + \dots \end{aligned}$$

Si  $\theta = 0$  entonces  $\frac{d^2}{d\theta^2} M_X(\theta) = \mu'_2$

$$\begin{aligned} \frac{d^3}{d\theta^3} M_X(\theta) &= \frac{d}{d\theta} \left[ \mu'_2 + \frac{2}{2} \theta\mu'_3 + \frac{3}{3 \cdot 2} \theta^2 \mu'_4 + \dots \right] \\ &= 0 + \mu'_3 + \theta\mu'_4 + \dots \end{aligned}$$

Si  $\theta = 0$  entonces  $\frac{d^3}{d\theta^3} M_X(\theta) = \mu'_3$

Generalizando

$$\left. \frac{d^k}{d\theta^k} M_X(\theta) \right|_{\theta=0} = \mu_k'$$

La derivada k-ésima respecto al parámetro  $\theta$  k veces, evaluada en  $\theta = 0$ , es el k-ésimo momento de la variable aleatoria con respecto al origen.

Volviendo a (29) teníamos que

$$\ln[M_t(\theta)] = \frac{-\mu_x \theta \sqrt{n}}{\sigma_x} + n \cdot \ln \left[ M_X \left( \frac{\theta}{\sigma_x \cdot \sqrt{n}} \right) \right]$$

En (30) encontramos una expresión para  $M_X \left( \frac{\theta}{\sigma_x \cdot \sqrt{n}} \right)$  reemplazando

$$\begin{aligned} \ln[M_t(\theta)] = \frac{-\mu_x \theta \sqrt{n}}{\sigma_x} + n \cdot \left[ 1 + \frac{\theta}{\sqrt{n} \sigma_x} \mu_1' + \frac{1}{2!} \frac{\theta^2}{n \sigma_x^2} \mu_2' + \right. \\ \left. + \frac{1}{3!} \frac{\theta^3}{n^{3/2} \sigma_x^3} \mu_3' + \frac{1}{4!} \frac{\theta^4}{n^{4/2} \sigma_x^4} \mu_4' + \dots \right] \end{aligned} \quad (31)$$

Si  $n > 30$  se lo considera lo suficientemente grande como para que lo contenido en el corchete sea una suma de términos que tiende a 1+ algo cercano a 0. Esto se puede desarrollar usando una expansión en serie de potencia donde

$$\ln(1+Z) = Z - \frac{Z^2}{2} + \frac{Z^3}{3} - \frac{Z^4}{4} + \dots \quad |Z| < 1$$

Llamemos Z a

$$Z = \frac{\theta}{\sqrt{n} \sigma_x} \mu_1' + \frac{1}{2!} \frac{\theta^2}{n \sigma_x^2} \mu_2' + \frac{1}{3!} \frac{\theta^3}{n^{3/2} \sigma_x^3} \mu_3' + \frac{1}{4!} \frac{\theta^4}{n^{4/2} \sigma_x^4} \mu_4' + \dots$$

Si desarrollamos la serie en potencia

$$\begin{aligned} \ln(1+Z) = & \left[ \frac{\theta}{\sqrt{n}\sigma_X} \mu_1' + \frac{1}{2!} \frac{\theta^2}{n\sigma_X^2} \mu_2' + \frac{1}{3!} \frac{\theta^3}{n^{3/2}\sigma_X^3} \mu_3' + \frac{1}{4!} \frac{\theta^4}{n^{4/2}\sigma_X^4} \mu_4' + \dots \right] - \\ & - \frac{1}{2} \left[ \frac{\theta}{\sqrt{n}\sigma_X} \mu_1' + \frac{1}{2!} \frac{\theta^2}{n\sigma_X^2} \mu_2' + \frac{1}{3!} \frac{\theta^3}{n^{3/2}\sigma_X^3} \mu_3' + \frac{1}{4!} \frac{\theta^4}{n^{4/2}\sigma_X^4} \mu_4' + \dots \right]^2 + \\ & + \frac{1}{3} \left[ \frac{\theta}{\sqrt{n}\sigma_X} \mu_1' + \frac{1}{2!} \frac{\theta^2}{n\sigma_X^2} \mu_2' + \frac{1}{3!} \frac{\theta^3}{n^{3/2}\sigma_X^3} \mu_3' + \frac{1}{4!} \frac{\theta^4}{n^{4/2}\sigma_X^4} \mu_4' + \dots \right]^3 - \\ & - \frac{1}{4} \left[ \frac{\theta}{\sqrt{n}\sigma_X} \mu_1' + \frac{1}{2!} \frac{\theta^2}{n\sigma_X^2} \mu_2' + \frac{1}{3!} \frac{\theta^3}{n^{3/2}\sigma_X^3} \mu_3' + \frac{1}{4!} \frac{\theta^4}{n^{4/2}\sigma_X^4} \mu_4' + \dots \right]^4 + \dots \end{aligned}$$

Luego tenemos que desarrollar los cuadrados, los cubos, las cuartas potencias... para poder agrupar

$$\underbrace{\left[ \frac{\theta}{\sqrt{n}\sigma_X} \mu_1' + \frac{1}{2!} \frac{\theta^2}{n\sigma_X^2} \mu_2' + \frac{1}{3!} \frac{\theta^3}{n^{3/2}\sigma_X^3} \mu_3' + \frac{1}{4!} \frac{\theta^4}{n^{4/2}\sigma_X^4} \mu_4' + \dots \right]^2}_{\text{A} \quad \text{B}}$$

Desarrollamos el cuadrado

$$\begin{aligned} & \frac{\theta^2}{n\sigma_X^2} (\mu_1')^2 + 2 \frac{\theta}{\sqrt{n}\sigma_X} \mu_1' \left[ \frac{1}{2!} \frac{\theta^2}{n\sigma_X^2} \mu_2' + \frac{1}{3!} \frac{\theta^3}{n^{3/2}\sigma_X^3} \mu_3' + \frac{1}{4!} \frac{\theta^4}{n^{4/2}\sigma_X^4} \mu_4' + \dots \right] + \\ & + \left[ \frac{1}{2!} \frac{\theta^2}{n\sigma_X^2} \mu_2' + \frac{1}{3!} \frac{\theta^3}{n^{3/2}\sigma_X^3} \mu_3' + \frac{1}{4!} \frac{\theta^4}{n^{4/2}\sigma_X^4} \mu_4' + \dots \right]^2 \end{aligned}$$

Tenemos

$$\begin{aligned} \ln(1+Z) &= \frac{\theta}{\sqrt{n}\sigma_X} \mu'_1 + \frac{1}{2!} \frac{\theta^2}{n\sigma_X^2} \mu'_2 + \frac{1}{3!} \frac{\theta^3}{n^{3/2}\sigma_X^3} \mu'_3 + \frac{1}{4!} \frac{\theta^4}{n^{4/2}\sigma_X^4} \mu'_4 + \dots - \\ &- \frac{\theta^2}{2\sigma_X^2 n} (\mu'_1)^2 - \frac{\theta^3}{\sigma_X^3 n^{3/2}} \mu'_1 \mu'_2 - \frac{\theta^4}{3!\sigma_X^4 n^{4/2}} \mu'_1 \mu'_3 - \frac{\theta^5}{4!\sigma_X^5 n^{5/2}} \mu'_1 \mu'_4 + \dots + \\ &+ \left[ \frac{\theta^2}{2!\sigma_X^2 n} \dots \right]^2 + \dots \end{aligned}$$

$$\ln(1+Z) = \frac{\theta}{\sqrt{n}\sigma_X} \mu'_1 + \frac{1}{2!} \frac{\theta^2}{n\sigma_X^2} \left[ \mu'_2 - (\mu'_1)^2 \right] + \frac{1}{3!} \frac{\theta^3}{n^{3/2}\sigma_X^3} (\mu'_3 - \mu'_1 \mu'_2) + \dots$$

Volviendo a (29) y teniendo en cuenta lo anterior

$$\begin{aligned} \ln[M_t(\theta)] &= \frac{-\mu_X \theta \sqrt{n}}{\sigma_X} + \frac{n\theta}{\sqrt{n}\sigma_X} \mu'_1 + \frac{n\theta^2}{2!n\sigma_X^2} \left[ \mu'_2 - (\mu'_1)^2 \right] + \\ &+ \frac{n\theta^3}{3!n^{3/2}\sigma_X^3} (\mu'_3 - \mu'_1 \mu'_2) + \dots \end{aligned}$$

$$\ln[M_t(\theta)] = \frac{-\mu_X \sqrt{n}}{\sigma_X} \theta + \frac{\sqrt{n}}{\sigma_X} \theta \mu'_1 + \frac{\mu'_2 - (\mu'_1)^2}{2!\sigma_X^2} \theta^2 + \frac{\mu'_3 - \mu'_1 \mu'_2}{3!n^{1/2}\sigma_X^3} \theta^3 + \dots \text{términos en}$$

$\theta^k$  en  $k > 3$

Tengamos en cuenta que  $\mu_X = \mu'_1$  y que  $\sigma_X^2 = \mu'_2 - (\mu'_1)^2$



Entonces

$$\ln[M_t(\theta)] = \underbrace{\left[ \frac{-\mu_X \sqrt{n}}{\sigma_X} + \frac{\mu_X \sqrt{n}}{\sigma_X} \right]}_{\ln[M_t(\theta)]} \theta + \frac{\sigma_X^2}{2! \sigma_X^2} \theta^2 + \dots$$

$$\ln[M_t(\theta)] = \frac{\theta^2}{2}$$

Los términos  $\theta^k$  tienden a 0 cuando  $n \rightarrow \infty$

$$\text{Entonces } \lim_{n \rightarrow \infty} \ln[M_t(\theta)] = \frac{\theta^2}{2} \quad (32)$$

$$\text{Es decir } \lim_{n \rightarrow \infty} M_t(\theta) = e^{\frac{\theta^2}{2}} \text{ porque } \ln \cdot e^{\frac{\theta^2}{2}} = \frac{\theta^2}{2}$$

Si comparamos (32) con (22)

$$\lim_{n \rightarrow \infty} \ln[M_t(\theta)] = \frac{\theta^2}{2} \quad (32)$$

$$M_t(\theta) = e^{\frac{\theta^2}{2}} \quad (22)$$

Se observa que la función generatriz de momentos de la variable aleatoria estándar  $t$  tiende a la función generatriz de momentos de la distribución normal estándar a medida que el tamaño de la muestra aumenta.

Este se puede expresar en el siguiente teorema: "Si  $X$  es una variable aleatoria con cualquier distribución de probabilidad de media  $\mu_X$  y desviación estándar  $\sigma_X$ , para la cual existe su función generatriz de momentos, entonces la variable aleatoria  $\bar{X}$  tiende asintóticamente a la distribución normal de media  $\mu_X$  y desviación estándar  $\sigma_X/\sqrt{n}$ ".

En el Teorema Central del Límite radica la importancia de la distribución normal, ya que basta con que el tamaño de la muestra sea

suficientemente grande, mayor a 30, para poder asegurar la buena aproximación de la distribución de medias de la muestra a la distribución normal, independientemente de la propia distribución que tenga la población de donde se extraigan las muestras.

Con los subincisos anteriores queda resuelto el problema de la distribución de probabilidad de las medias de las muestras, cuando estas provienen de una población normal o de una población con cualquier distribución de probabilidad y el tamaño de las muestras es grande. En ambos casos se recurre a la distribución normal.

#### 4.3 Factor de corrección para poblaciones finitas

Las poblaciones pueden ser finitas o infinitas. Si de una población finita se obtienen muestras con reemplazo, puede considerarse teóricamente que la población es infinita, ya que cualquier número de muestras pueden obtenerse de la población sin agotarla. Además, para casi cualquier propósito práctico, el muestreo de poblaciones finitas de tamaño muy grande puede considerarse como muestreo de poblaciones infinitas. De acuerdo al concepto de muestreo aleatorio, el muestreo sin reemplazo de poblaciones finitas no conduce a muestras aleatorias, debido a que al dejar permanentemente fuera el elemento ya extraído, influye en la probabilidad de extracción del siguiente; o sea, que las pruebas sucesivas para obtener una muestra no son independientes. Esto no ocurre si la población es infinita. Resumiendo lo anterior puede decirse que, para obtener muestras aleatorias, el muestreo debe ser de cualquier tipo en poblaciones infinitas y sólo con reemplazo en poblaciones finitas, a excepción del caso de poder considerar a la población infinita por ser de tamaño muy grande.

Con respecto al teorema, cabe indicar que cuando la población es finita y el muestreo se realiza sin reemplazo, hay que incluir en la determinación de  $\sigma_{\bar{X}}$  el valor del tamaño de la población, es decir,

$$\sigma_{\bar{X}} = \frac{\sigma_X}{\sqrt{n}} \cdot \sqrt{\frac{N-n}{N-1}}$$

donde  $\frac{N-n}{N-1}$  recibe el nombre de factor de corrección para poblaciones finitas en muestreo sin reemplazo. Obsérvese que este factor de corrección tiende a cero cuando  $n \rightarrow N$ .

**Ejemplo.** La vida útil de los focos fabricados por una empresa tiene distribución normal de media 200 horas y desviación estándar 25 horas. Calcular la probabilidad de que la duración media de 25 focos escogidos al azar sea superior a 208 horas.

De acuerdo al enunciado del problema, la población de las duraciones de los focos de la empresa tiene distribución normal de parámetros.

$$\mu_x = 200 \text{ horas y } \sigma_x = 25 \text{ horas}$$

y puede considerarse que es de tamaño infinito. Entonces, la distribución de las medias de las muestras también tendrá distribución normal de parámetros:

$$n = 25 \text{ focos}$$

$$\mu_{\bar{x}} = \mu_x = 200 \text{ horas}$$

$$\sigma_{\bar{x}} = \frac{\sigma_x}{\sqrt{n}} = \frac{25}{\sqrt{25}} = 5 \text{ horas}$$

Estandarizando la variable aleatoria  $\bar{X}$ , se tiene

$$Z = \frac{\bar{X} - \mu_{\bar{x}}}{\sigma_{\bar{x}}} = \frac{\bar{X} - 200}{5}$$

$$\text{Si } \bar{X} = 208, \quad z = \frac{208 - 200}{5} = 1.6$$

Por lo tanto, la probabilidad de que la media de los 25 focos de la muestra tenga un valor mayor a 208 horas valdrá:

$$P(\bar{X} > 208) = P(z > 1.6)$$

$$= 1 - P(z \leq 1.6)$$

$$= 1 - 0.9452$$

$$= 0.0548$$

Sólo el 5.48% de los focos de la muestra tendrán una vida superior a las 208 horas, es decir, un solo foco.

**Ejemplo.** Los pesos de 600 ejes producidos en un torno tienen distribución normal de media 53 kg y desviación estándar de 2.5 kg. Estos se empaquetan en cajas de 10, las que soportan hasta 540 kg de peso. Si se envían 35 de estas cajas, calcular cuántas cajas cabe esperar se rompan por exceso de peso.

La población de pesos de los ejes tiene distribución normal de acuerdo al enunciado explícito del problema. Sus parámetros son:

$$N = 600, \quad \mu_x = 53\text{Kg}, \quad \sigma_x = 2.5\text{Kg}$$

Como la población es normal, la distribución de medias de las muestras también es normal, y sus parámetros valen;

$$n = 10 \quad \mu_{\bar{x}} = \mu_x = 53\text{Kg}$$

$$\sigma_{\bar{x}} = \frac{\sigma_x}{\sqrt{n}} \sqrt{\frac{N-n}{N-1}} = \frac{2.5}{\sqrt{10}} \sqrt{\frac{600-10}{600-1}} = 0.785 \text{Kg}$$

Una caja de ejes se romperá por exceso de peso, si el peso conjunto de los 10 ejes es mayor de 540 kg; y el peso conjunto de los 10 ejes es mayor de 540 kg si el peso medio de los 10 ejes es superior a 54kg (540/10).

Entonces, se trata de calcular cuántas de las 35 muestras tienen media mayor de 54 kg, lo que se hará por medio del cálculo de la probabilidad  $P(\bar{X} > 54)$ , en donde  $\bar{X}$  tiene distribución normal de media 53 y desviación estándar 0.785.

Estandarizando  $\bar{X} = 54$ , se tiene

$$z = \frac{\bar{X} - \mu_{\bar{x}}}{\sigma_{\bar{x}}} = \frac{54 - 53}{0.785} = 1.274$$

Por lo tanto,

$$\begin{aligned} P(\bar{X} > 54) &= P(z > 1.274) \\ &= 1 - P(z \leq 1.274) \\ &= 1 - 0.8987 \\ &= 0.1013 \end{aligned}$$

El 10.13% de las 35 cajas se romperán, o sea, de 3 a 4 cajas.

**Ejemplo.** Los tabiques comprimidos que se usan en una construcción tienen un peso medio de 5.50 kg y una desviación estándar de 0.85 kg. Estos se elevan en lotes al lugar en donde se emplean por medio de una grúa cuyo límite de seguridad es de 200 kg. Calcular el tamaño máximo de los lotes de manera de que la probabilidad de exceder el límite de seguridad de la grúa sea menor de 5%.

Como en el ejemplo anterior, se vuelve a tener una población de pesos cuyos parámetros son

$$\mu_x = 5.50 \text{Kg} \text{ y } \sigma_x = 0.85 \text{Kg}$$

con distribución de probabilidad desconocida.

La simple comparación entre el peso de un tabique y el límite de seguridad de la grúa, hace que los lotes de tabiques que se cargan den muestras de tamaño grande. Esto permite asegurar que la distribución de medias de las muestras de pesos de los tabiques es aproximadamente normal de parámetros

$$\mu_{\bar{x}} = \mu_x = 5.50 \text{Kg} \text{ y } \sigma_{\bar{x}} = \frac{\sigma_x}{\sqrt{n}} = \frac{0.85}{\sqrt{n}}$$

en donde sólo se sabe que n es grande (mayor de 30).

El peso de la muestra de n tabiques excede el límite de seguridad de la grúa de 200 kg, si la media de la muestra es superior a 200/n. Se busca saber cuál es el tamaño de la muestra del lote (cantidad de tabiques) que hace que el límite de seguridad sea superado. El problema se plantea como:

$$P\left(\bar{X} > \frac{200}{n}\right) < 0.05$$

Estandarizando  $x$  resulta 
$$z = \frac{\bar{X} - \mu_x}{\sigma_x} = \frac{\bar{X} - 5.50}{0.85/\sqrt{n}}$$

Si  $\bar{X} = \frac{200}{n}$ ;

Entonces 
$$z = \frac{\frac{200}{n} - 5.50}{\frac{0.85}{\sqrt{n}}} = \frac{200 - 5.50n}{0.85\sqrt{n}}$$

De modo que

$$P\left(\bar{X} > \frac{200}{n}\right) = P\left(z > \frac{200 - 5.50n}{0.85\sqrt{n}}\right) < 0.05$$

lo que es lo mismo 
$$P\left(z \leq \frac{200 - 5.50n}{0.85\sqrt{n}}\right) \geq 0.95$$

De la tabla de probabilidades de la normal se obtiene que esta es efectivamente mayor de 0.95 si el valor crítico de  $z$  es mayor o igual a 1.654.

Luego se obtiene

$$\frac{200 - 5.50n}{0.85\sqrt{n}} \geq 1.645$$

$$200 - 5.50n \geq 1.398\sqrt{n}$$

Elevando ambos miembros al cuadrado

$$(200 - 5.50n)^2 \geq (1.398\sqrt{n})^2$$

$$40000 - 2200n + 30.25n^2 \geq 1.955n$$

$$30.25n^2 - 2201.955n + 40000 \geq 0$$

Resolviendo la ecuación se obtienen las raíces 34.86 y 37.93. Cada una de estas raíces representa dos tamaños de muestras diferentes.

Si  $z \leq \frac{200 - 5.50n}{0.85\sqrt{n}}$ , cuando  $n = 34.86 \Rightarrow z \leq -1.645$   
 $n = 37.93 \Rightarrow z \leq 1.645$

Luego

$$P(z < -1.645) = 0.05$$

$$P(z < 1.645) = 0.95$$

Se busca que el tamaño del lote no supere el límite de seguridad de la grúa fijado en el 5%. Por esta razón, el tamaño de muestra requerido es

$$n \leq 34.86$$

#### 4.4 Distribución Gamma

La **distribución gamma** servirá de punto de partida para obtener las funciones de densidad de las distribuciones exactas, que comprende a:

distribución chi cuadrado de Pearson,  $\chi^2$

distribución t de Student,  $t$

distribución F de Snedecor o Fischer,  $F$

La función gamma del análisis matemático es:

$$\Gamma(\lambda) = \int_0^{\infty} e^{-x} x^{\lambda-1} dx; \quad \lambda > 0 \quad (1)$$

esta integral se resuelve fácilmente y se obtiene

$$\Gamma(\lambda) = (\lambda - 1)! \quad (2)$$

Se debe tener presente, en los desarrollos posteriores, el valor notable:

$$\Gamma\left(\frac{1}{2}\right) = \sqrt{\pi} \quad (3)$$

Hay dos integrales de las que resulta útil conocer su resolución por la aplicación que de ellas se hace más adelante:

1. En la integral:

$$\int_0^{\infty} e^{-\alpha x} x^{\lambda-1} dx; \quad \alpha > 0, \quad \lambda > 0 \quad (4)$$

se realiza la siguiente sustitución:  $x\alpha = \mu$

de donde se deduce:  $x = \frac{\mu}{\alpha}; \quad dx = \frac{d\mu}{\alpha}$

sustituyendo en (10):  $\int_0^{\infty} e^{-\mu} \frac{\mu^{\lambda-1}}{\alpha^{\lambda-1}} \frac{d\mu}{\alpha} = \int_0^{\infty} e^{-\mu} \frac{\mu^{\lambda-1}}{\alpha^{\lambda}} d\mu$

si se saca  $\frac{1}{\alpha^{\lambda}}$  fuera de la integral y se compara con (7) queda:

$$\frac{1}{\alpha^{\lambda}} \int_0^{\infty} e^{-\mu} \mu^{\lambda-1} d\mu = \frac{1}{\alpha^{\lambda}} \Gamma(\lambda)$$

o sea que:

$$\int_0^{\infty} e^{-\alpha x} x^{\lambda-1} dx = \frac{\Gamma(\lambda)}{\alpha^{\lambda}} \quad (5)$$

2. En la integral

$$\int_0^{\infty} e^{-\alpha x^2} dx, \quad \alpha > 0$$

sustituyendo  $x^2 = \mu$ , se tiene

$$x = \mu^{1/2}; \quad dx = \frac{1}{2} \mu^{-\frac{1}{2}} du$$

al sustituir se obtiene:

$$\frac{1}{2} \int_0^{\infty} e^{-\alpha \mu} \mu^{-\frac{1}{2}} du = \frac{\Gamma\left(\frac{1}{2}\right)}{2\alpha^{\frac{1}{2}}} = \frac{\sqrt{\pi}}{2\alpha^{\frac{1}{2}}} \quad (6)$$

La **distribución gamma** de la variable aleatoria  $\mathbf{X}$ , con parámetros  $\alpha$  y  $\lambda$ , se define mediante la siguiente función de densidad:

$$f(x) = \begin{cases} \frac{\alpha^\lambda}{\Gamma(\lambda)} e^{-\alpha x} x^{\lambda-1}, & \alpha > 0; \lambda > 0; x > 0 \\ 0, & x \leq 0 \end{cases} \quad (7)$$

observando (5) se verifica que:  $\int_0^{\infty} f(x) dx = 1$

La función generatriz de momentos de esta distribución con función de densidad  $f(x)$ , es

$$m_x(\theta) = E(e^{\theta x}) = \int_0^{\infty} e^{\theta x} f(x) dx$$

que, haciendo uso de (13), queda:

$$\frac{\alpha^\lambda}{\Gamma(\lambda)} \int_0^{\infty} e^{-\alpha x} e^{\theta x} x^{\lambda-1} dx = \frac{\alpha^\lambda}{\Gamma(\lambda)} \int_0^{\infty} e^{-x(\alpha-\theta)} x^{\lambda-1} dx$$

La integral es análoga a (5) pero con  $\alpha - \theta$  en lugar de  $\alpha$ , luego,

$$m_x(\theta) = \frac{\alpha^\lambda \Gamma(\lambda)}{\Gamma(\lambda)(\alpha - \theta)^\lambda} = \frac{\alpha^\lambda}{(\alpha - \theta)^\lambda} = \frac{(\alpha - \theta)^{-\lambda}}{\alpha^{-\lambda}} = \left(\frac{\alpha - \theta}{\alpha}\right)^{-\lambda} = \left(\frac{\alpha}{\alpha} - \frac{\theta}{\alpha}\right)^{-\lambda}$$

$$m_x(\theta) = \left(1 - \frac{\theta}{\alpha}\right)^{-\lambda} \quad (8)$$

Derivando sucesivamente la función generatriz se obtienen los momentos de orden 1 y orden 2 ( $m_1$  y  $m_2$ ), que posibilitan el cálculo de la media y la varianza de la distribución.

$$E(x) = m_1 = \frac{\lambda}{\alpha}$$

$$V(x) = m_2 - m_1^2 = \frac{\lambda}{\alpha^2}$$

#### 4.5 Distribución chi-cuadrado ( $\chi^2$ ) de Pearson

Si  $X_1, X_2, \dots, X_n$ , son  $n$  observaciones muestrales en una población normal, con parámetros  $(\mu, \sigma)$ , el estadístico (variable aleatoria):

$$Y = \frac{\sum_{i=1}^n (X_i - \mu)^2}{\sigma^2} \sim \chi_n^2 \quad (9)$$

(9) representa una Chi-Cuadrado con  $n$  grados de libertad, donde la variable  $X$  se distribuye normal con media,  $\mu$ , y desvío,  $\sigma$ .

Una forma alternativa de escribir (9) es la siguiente:

$$Y = \sum_{i=1}^n z_i^2 \sim \chi_n^2 \quad \text{donde } z_i \sim N(0, 1)$$

La función generatriz de momentos de la variable  $Y \sim \chi_n^2$  es:

$$m_Y(\theta) = m_{\sum z_i^2}(\theta) = m_{z_1^2}(\theta) m_{z_2^2}(\theta) \dots m_{z_n^2}(\theta) \Rightarrow m_Y(\theta) = (m_{z^2}(\theta))^n$$

Pero como  $z \sim N(0, 1)$ , según lo analizado en (10)

$$m_{z^2}(\theta) = E(e^{\theta z^2}) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{\theta z^2} e^{-\frac{z^2}{2}} dz$$



Como se trata de una función par, integramos entre 0 e  $\infty$  y multiplicamos por 2

$$\frac{2}{\sqrt{2\pi}} \int_0^{\infty} e^{-z^2 \left(\frac{1}{2}-\theta\right)} dz \quad (10)$$

(10) es análoga a (6) para  $\alpha=(1/2)-\theta$ , por lo tanto

$$m_{z^2}(\theta) = \frac{2}{\sqrt{2\pi}} \frac{\sqrt{\pi}}{2\sqrt{\frac{1}{2}-\theta}} = \frac{1}{\sqrt{2}\sqrt{\frac{1}{2}-\theta}} = \frac{1}{\sqrt{1-2\theta}}$$

o sea que  $m_Y(\theta) = [m_{z^2}(\theta)]^n = \sqrt{\left(\frac{1}{1-2\theta}\right)^n}$  (11)

(11) es la función generatriz de momentos de una variable Y con distribución chi cuadrado de n grados de libertad.

Comparando (11) con (8) se observa que la función generatriz de momentos tiene una estructura similar a una distribución gamma de:

*variable:* **Y**

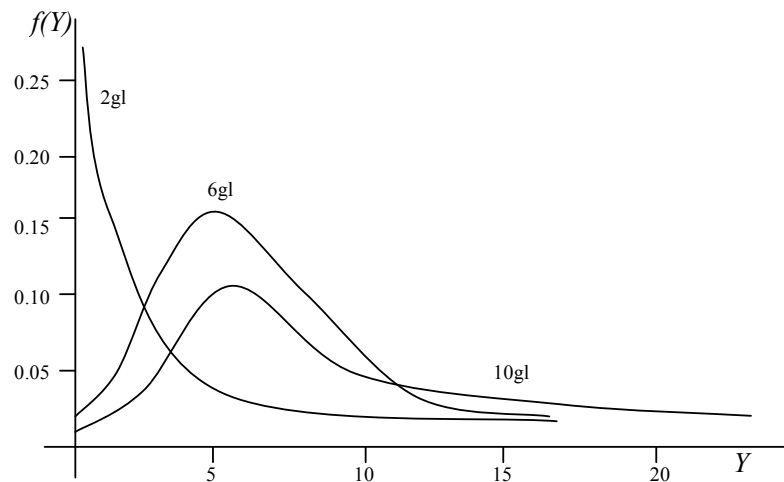
*parámetros:*  **$\alpha = 1/2, \lambda = n/2$**

En consecuencia la función de densidad de Y, teniendo en cuenta (7) es:

$$f(Y) = \frac{e^{-Y/2} Y^{(n/2)-1}}{\sqrt{2^n} \Gamma(n/2)} \quad (12)$$

Y se distribuye chi-cuadrado con n grados de libertad.

El diagrama presenta funciones de distribución de chi cuadrado con diferentes grados de libertad



La media y la desviación típica de la distribución  $\chi^2$  se obtienen de la función generatriz de momentos, hallada en (11)

$$E(\chi^2) = \left. \frac{\partial m_Y(\theta)}{\partial(\theta)} \right|_{\theta=0} = n$$

$$V(\chi^2) = \left. \frac{\partial^2 m_Y(\theta)}{\partial^2(\theta)} \right|_{\theta=0} - [E(\chi^2)]^2 = 2n$$

Propiedad reproductiva de  $\chi^2$

Dadas dos variables independientes  $Y_1 \sim \chi_{n_1}^2$  e  $Y_2 \sim \chi_{n_2}^2$

la suma de ellas es otra variable  $\chi^2$  con  $n_1 + n_2$  grados de libertad.

En efecto,

$$m_{Y_1+Y_2}(\theta) = m_{Y_1}(\theta) m_{Y_2}(\theta) = \sqrt{\frac{1}{(1-2\theta)^{n_1}}} \sqrt{\frac{1}{(1-2\theta)^{n_2}}} = \sqrt{\frac{1}{(1-2\theta)^{(n_1+n_2)}}$$

que según (11) es la función generatriz de momentos de  $\chi_{n_1+n_2}^2$

**Conclusión.** Dada  $Z \sim N(0,1)$ , variable normal tipificada, si se seleccionan  $n$  valores aleatorios  $Z_1, Z_2, \dots, Z_n$  de esta distribución, se elevan al

cuadrado y se suman, el estadístico resultante tiene una distribución  $\chi^2$  con  $n$  grados de libertad,

$$Z_1^2 + Z_2^2 + \dots + Z_n^2 \sim \chi_n^2$$

Cuando los grados de libertad tienden a  $\infty$ ,  $\chi_n^2 \rightarrow Z \sim N(0,1)$

**Ejemplo.** La variable  $X$  tiene distribución chi-cuadrado con 20 grados de libertad:

$$X \sim \chi_{20}^2$$

La probabilidad de que la variable  $X$  asuma el valor 34.17, o menos, es 0,75. La notación es

$$P(X \geq 34.17) = 0.025$$

Otra forma de notación es  $P(\chi^2 \geq 34.17) = 0.025$

Para hallar esta probabilidad es necesario utilizar la **Tabla E.4 Valores críticos de  $\chi^2$**  publicada por *Mark Berenson y David Levine*. La tabla tiene los grados de libertad a la izquierda y arriba la probabilidad en el extremo superior, es decir, la probabilidad acumulada desde el valor crítico a infinito.

La variable tiene 20 grados de libertad, por ende es necesario ubicar la fila correspondiente a 20 grados. Luego debe buscarse sobre la fila el punto 34.17. Una vez hallado debe leerse arriba (en Areas de extremo superior) la probabilidad, en este caso 0.025.

La probabilidad de los menores a 34.17 se halla de la siguiente manera:

$$P(X \leq 34.17) = 1 - P(X \geq 34.17) = 1 - 0.025 = 0.975$$

#### 4.6 Distribución F de Snedecor o Fischer

Sea  $\chi_m^2$  una variable chi-cuadrado con  $m$  grados de libertad y  $\chi_n^2$  otra variable chi-cuadrado con  $n$  grados de libertad, supongamos que  $\chi_m^2$  y  $\chi_n^2$  son independientes, entonces la variable aleatoria

$$F = \frac{\chi_m^2}{\chi_n^2} \sim F_{m,n}$$

se distribuye F de Snedecor con  $m$  y  $n$  grados de libertad en el numerador y denominador respectivamente.

Si las dos variables independientes son  $X$  e  $Y$ , con  $m$  y  $n$  grados de libertad se define el estadístico F

$$F = \frac{X/m}{Y/n} = \frac{nX}{mY} \sim F(m, n) \quad (13)$$

La función de densidad conjunta de las variables  $X$  e  $Y$  es:

$$g(X, Y) dx dy = k_m(X) dx k_n(Y) dy = \frac{e^{-\left(\frac{x}{2}\right)} x^{\frac{m}{2}-1} e^{-\left(\frac{y}{2}\right)} y^{\frac{n}{2}-1}}{2^{\frac{m}{2}} \Gamma(m/2) 2^{\frac{n}{2}} \Gamma(n/2)} dx dy \quad (14)$$

donde  $k_m(X) \sim \chi_m^2$  y  $k_n(Y) \sim \chi_n^2$

$$\iint g(X, Y) dx dy = \iint \frac{x^{\frac{m}{2}-1} y^{\frac{n}{2}-1} e^{-\left(\frac{x+y}{2}\right)}}{2^{\frac{m+n}{2}} \Gamma(m/2) \Gamma(n/2)} dx dy \quad (15)$$

Se sustituye  $X$  en función de  $F$  e  $Y$  ya que según (19)

$$X = \frac{mYF}{n}; \quad dX = \frac{mY}{n} dF \Rightarrow \frac{dX}{dF} = \frac{mY}{n}$$

queda

$$\begin{aligned} \iint g(F, Y) dF dY &= \iint \frac{m^{\frac{m}{2}-1} Y^{\frac{m}{2}-1} F^{\frac{m}{2}-1} Y^{\frac{n}{2}-1} e^{-\left(\frac{mYF}{2n}\right) - \left(\frac{Y}{2}\right)} mY}{2^{\frac{m+n}{2}} \Gamma(m/2) \Gamma(n/2) n^{\frac{m}{2}-1} n} dF dY \\ &= \iint \frac{m^{\frac{m}{2}-1} Y^{\frac{m}{2}-1} F^{\frac{m}{2}-1} Y^{\frac{n}{2}-1} e^{-\left(\frac{Y^{n+mF}}{2n}\right)} mY}{n^{\frac{m}{2}-1} 2^{\frac{m+n}{2}} \Gamma(m/2) \Gamma(n/2) n} dF dY \end{aligned} \quad (16)$$

al integrar  $g(F, Y) dF dY$  respecto de  $Y$  se obtiene la función de densidad marginal de  $F$ , esto es

$$\int_0^{\infty} g(F, Y) dY = h(F) dF$$

Por lo tanto,

$$\int h(F) dF = \int_0^{\infty} \frac{m^{\frac{m}{2}} F^{\frac{m}{2}-1}}{n^{\frac{m}{2}} 2^{\frac{m+n}{2}} \Gamma(m/2) \Gamma(n/2)} e^{-Y \frac{n+mF}{2n}} Y^{\frac{n+m}{2}-1} dY \quad (17)$$

Se observa que:

$$\begin{aligned}
 1) \quad & m^{\frac{m}{2}-1} m = m^{\frac{m}{2}-1+1} = m^{\frac{m}{2}} \\
 2) \quad & n^{\frac{n}{2}-1} n = n^{\frac{n}{2}} \\
 3) \quad & Y^{\frac{m}{2}-1} Y^{\frac{n}{2}-1} Y = Y^{\frac{m}{2}-1+\frac{n}{2}-1+1} = Y^{\frac{m+n}{2}-1}
 \end{aligned}$$

de donde:

$$h(F)dF = \frac{m^{\frac{m}{2}} F^{\frac{m}{2}-1}}{n^{\frac{m}{2}} 2^{\frac{m+n}{2}} \Gamma(m/2) \Gamma(n/2)} \int_0^{\infty} e^{-Y \frac{n+mF}{2n}} Y^{\frac{n+m}{2}-1} dY$$

que es análoga a la integral de la Distribución Gamma hallada en (5)

$$\int_0^{\infty} e^{-\alpha X} X^{\lambda-1} dX = \frac{\Gamma(\lambda)}{\alpha^{\lambda}}; \quad \text{donde: } \alpha = \frac{n+mF}{2n} \text{ y } \lambda = \frac{n+m}{2}$$

Entonces:

$$\begin{aligned}
 h(F) &= \frac{m^{\frac{m}{2}} F^{\frac{m}{2}-1}}{n^{\frac{m}{2}} 2^{\frac{m+n}{2}} \Gamma(m/2) \Gamma(n/2)} \frac{\Gamma\left(\frac{n+m}{2}\right)}{\left(\frac{n+mF}{2n}\right)^{\frac{n+m}{2}}} \\
 &= \frac{m^{\frac{m}{2}} F^{\frac{m}{2}-1}}{n^{\frac{m}{2}} 2^{\frac{m+n}{2}} \Gamma(m/2) \Gamma(n/2)} \frac{\Gamma\left(\frac{n+m}{2}\right) 2^{\frac{m+n}{2}} n^{\frac{n+m}{2}}}{(n+mF)^{\frac{n+m}{2}}} \\
 &= \frac{\Gamma\left(\frac{n+m}{2}\right) m^{\frac{m}{2}} F^{\frac{m}{2}-1} n^{\frac{m}{2}+\frac{n+m}{2}}}{\Gamma(m/2) \Gamma(n/2) (n+mF)^{\frac{n+m}{2}}}
 \end{aligned}$$

$$\Rightarrow h(F) = \frac{\Gamma\left(\frac{n+m}{2}\right) m^{\frac{m}{2}} F^{\frac{m-2}{2}} n^{\frac{n}{2}} (n+mF)^{-\left(\frac{m+n}{2}\right)}}{\Gamma(m/2) \Gamma(n/2)}$$

es la función de densidad de la distribución  $F$  de Snedecor

Las características de esta distribución son

$$E(F) = \frac{n}{n-2}; \quad n > 2$$

$$V(F) = \frac{n^2(2n+2m-4)}{m(n-2)^2(n-4)}; \quad n > 4$$

**Ejemplo.** Una variable  $F$  se distribuye  $F$  de Snedecor con 4 grados de libertad en el numerador y 6 grados de libertad en el denominador. La probabilidad asociada a los puntos mayores al valor crítico 9.15 es 0.01. Su notación es:

$$P(F_{4,6} \geq 9.15) = 0.01$$

Para hallar esta probabilidad es necesario utilizar la **Tabla E.5 Valores críticos de  $F$**  publicada por *Mark Berenson y David Levine*. La tabla tiene los grados de libertad del denominador a la izquierda y arriba los grados de libertad en el numerador. Hay una tabla para cada probabilidad en el extremo superior, es decir, la probabilidad acumulada desde el valor crítico a infinito.

La variable tiene 4 grados de libertad en el numerador y 6 grados de libertad en el denominador, por ende es necesario ubicar la celda en la cual, la columna y la fila correspondientes a estos grados de libertad, se crucen. En la Tabla donde se encuentre el punto crítico, allí se hallará la probabilidad.

#### 4.7 Relaciones entre las distribuciones Normal y $\chi^2$

**Teorema:** Suponiendo que  $X_1, X_2, \dots, X_n$  son datos provenientes de una muestra aleatoria de una variable  $X$  con esperanza  $\mu$  y varianza  $\sigma^2$ . Sea

$$S^2 = \frac{\sum_{i=1}^n (X_i - \bar{X})^2}{n-1}$$

un estimador de la varianza, en donde  $\bar{X}$  es el promedio muestral, se tiene que:

(a)  $E(S^2) = \sigma^2$

(b) Si  $X$  está distribuida normalmente,  $\frac{n-1}{\sigma^2} S^2$  tiene una distribución  $\chi^2$  con  $(n-1)$  grados de libertad

Para demostrar la parte **(a)** del teorema, a la suma de cuadrados de los desvíos respecto de la media, se le resta y suma  $\mu$

$$\sum_{i=1}^n (X_i - \bar{X})^2 = \sum_{i=1}^n (X_i - \mu + \mu - \bar{X})^2$$

Agrupando convenientemente y desarrollando el cuadrado

$$= \sum_{i=1}^n [(X_i - \mu)^2 + 2(X_i - \mu)(\mu - \bar{X}) + (\mu - \bar{X})^2]$$

Distribuyendo el sumatorio

$$\begin{aligned} &= \sum_{i=1}^n (X_i - \mu)^2 + 2(\mu - \bar{X}) \sum_{i=1}^n (X_i - \mu) + n(\mu - \bar{X})^2 \\ &= \sum_{i=1}^n (X_i - \mu)^2 + 2(\mu - \bar{X}) \left( \sum_{i=1}^n X_i - n\mu \right) + n(\mu - \bar{X})^2 \end{aligned}$$

Multiplicando y dividiendo el segundo término por  $(-n)$

$$= \sum_{i=1}^n (X_i - \mu)^2 + 2(\mu - \bar{X}) \left( \sum_{i=1}^n X_i - n\mu \right) \frac{-n}{-n} + n(\mu - \bar{X})^2$$

$$= \sum_{i=1}^n (X_i - \mu)^2 - 2n(\mu - \bar{X})(\mu - \bar{X}) + n(\mu - \bar{X})^2$$

$$= \sum_{i=1}^n (X_i - \mu)^2 - 2n(\mu - \bar{X})^2 + n(\mu - \bar{X})^2$$

De modo que,

$$\sum_{i=1}^n (X_i - \bar{X})^2 = \sum_{i=1}^n (X_i - \mu)^2 - n(\mu - \bar{X})^2$$

Luego, se toma esperanza

$$E(S^2) = E\left[\frac{\sum_{i=1}^n (X_i - \bar{X})^2}{n-1}\right] = E\left[\frac{\sum_{i=1}^n (X_i - \mu)^2 - n(\bar{X} - \mu)^2}{n-1}\right]$$

Aplicando las propiedades de esperanza matemática

$$= \frac{1}{n-1} E\left[\sum_{i=1}^n (X_i - \mu)^2 - n(\bar{X} - \mu)^2\right] = \frac{1}{n-1} \left[\sum_{i=1}^n E(X_i - \mu)^2 - nE(\bar{X} - \mu)^2\right]$$

De modo que

$$E(S^2) = \frac{n(\sigma^2 - \sigma_{\bar{X}}^2)}{n-1}$$

$$\text{Entonces, } E(S^2) = \frac{n\sigma^2 - n\frac{\sigma^2}{n}}{n-1} = \frac{n\sigma^2 - \sigma^2}{n-1} = \frac{\sigma^2(n-1)}{n-1} = \sigma^2$$

con lo que se demuestra la primera parte del Teorema.

El estimador

$$\hat{\sigma}^2 = \frac{\sum_{i=1}^n (X_i - \bar{X})^2}{n}$$

es un estimador sesgado de la varianza poblacional; siendo el sesgo igual a  $\frac{n-1}{n}$ .



Para demostrarlo se parte consignando

$$E(\hat{\sigma}^2) = \frac{n-1}{n} \sigma^2$$

de donde  $\frac{n}{n-1} E(\hat{\sigma}^2) = \sigma^2$

Esto puede escribirse reemplazando el estimador por su igual

$$\frac{n}{n-1} E \left[ \frac{\sum (X_i - \bar{X})^2}{n} \right] = \sigma^2$$

esto es,

$$E \left( \frac{\sum (X_i - \bar{X})^2}{n-1} \right) = \sigma^2$$

Pero  $\frac{\sum (X_i - \bar{X})^2}{n-1} = S^2$

es decir,  $E(S^2) = \sigma^2$ , con lo que queda demostrado que el sesgo es  $(n-1)/n$ . De este modo  $S^2$  es un estimador insesgado de la varianza poblacional y el factor de corrección que permite evitar el sesgo es  $n/(n-1)$ .

Se demostrará ahora la parte **(b)** del teorema. El enunciado dice que si

$$X \sim N(\mu, \sigma^2)$$

Entonces

$$\frac{(n-1)s^2}{\sigma^2} \sim \chi_{n-1}^2$$

$$\text{Si } \chi_{n-1}^2 = \frac{(n-1)s^2}{\sigma^2} = \frac{(n-1) \frac{\sum_{i=1}^n (X_i - \bar{X})^2}{n-1}}{\sigma^2} = \frac{\sum_{i=1}^n (X_i - \bar{X})^2}{\sigma^2}$$

En general, una suma de variables aleatorias independientes con distribución normal se distribuye  $\chi^2$  con grados de libertad determinados por la cantidad de variables menos 1.

**Ejemplo.** Considerando el caso especial de  $n=2$

$$\begin{aligned} \sum_{i=1}^2 (X_i - \bar{X})^2 &= (X_1 - \bar{X})^2 + (X_2 - \bar{X})^2 \\ &= \left( X_1 - \frac{X_1 + X_2}{2} \right)^2 + \left( X_2 - \frac{X_1 + X_2}{2} \right)^2 \\ &= \frac{(2X_1 - X_1 - X_2)^2}{4} + \frac{(2X_2 - X_1 - X_2)^2}{4} \\ &= \frac{(X_1 - X_2)^2 + (X_2 - X_1)^2}{4} = \frac{(X_1 - X_2)^2}{2} \end{aligned}$$

Puesto que  $X_1$  y  $X_2$  están independientemente distribuidos  $N(\mu, \sigma^2)$  se encuentra que  $(X_1 - X_2)$  tiene una distribución  $N(0, 2\sigma^2)$ , luego

$$\frac{\sum_{i=1}^2 (X_i - \bar{X})^2}{\sigma^2} = \left( \frac{(X_1 - X_2) - 0}{\sqrt{2} \sigma} \right)^2 = \frac{1}{\sigma^2} S^2$$

tiene la distribución  $\chi^2$  con 1 grado de libertad.

#### 4.8 Distribución T de Student

Dadas las variables  $X \sim N(\mu, \sigma)$  e  $Y \sim \chi_n^2$

se define

$$Z = \frac{X - \mu_x}{\sigma_x}$$

$$Y = \sum \left( \frac{x - \mu_x}{\sigma_x} \right)^2 = \frac{\sum (x - \mu_x)^2}{\sigma_x^2}$$

Multiplicando y dividiendo esta última expresión por  $(n-1)$

$$Y = \frac{(n-1) \sum (x - \mu_x)^2}{\sigma_x^2 (n-1)} = \frac{(n-1)}{\sigma_x^2} s^2$$

Se dice que el estadístico  $t$  se distribuye como una T de Student con  $(n-1)$  grados de libertad

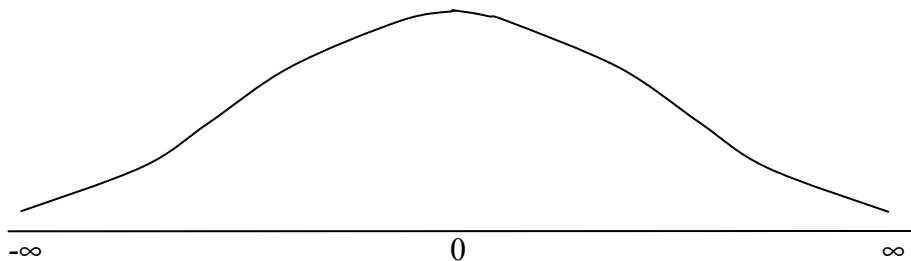
$$t = \frac{\frac{X - \mu_x}{\sigma_x}}{\sqrt{\frac{(n-1) s^2}{\sigma^2 (n-1)}}} = \frac{X - \mu}{s} \sim t_{n-1}$$

siendo sus características

$$E(t) = 0; \quad n > 0$$

$$V(t) = \frac{n}{n-2}; \quad n > 2$$

La distribución  $t$  es simétrica con respecto a 0 y asintóticamente tiende a una distribución normal tipificada.



Si se eleva al cuadrado la expresión de  $t$ , el resultado se puede escribir como  $t^2 = \frac{Z^2/1}{V/v}$  en donde  $Z^2$ , al ser el cuadrado de una variable normal tipificada, tiene una distribución  $\chi_1^2$ . Así pues,

$$t^2 = F(1, v)$$

es decir, el cuadrado de una variable  $t$  con  $v$  grados de libertad es una  $F$  con  $(1, v)$  grados de libertad.

**Ejemplo.** En una variable  $X$ , con distribución  $t$  de Student con 10 grados de libertad, la probabilidad de asumir valores superiores a 1.3722 es 0.10. Su notación es:

$$P(t_{10} \geq 1.3722) = 0.10$$

Para hallar esta probabilidad es necesario utilizar la **Tabla E.3 Valores críticos de  $t$**  publicada por *Mark Berenson y David Levine*. La tabla tiene los grados de libertad a la izquierda y arriba el área en el extremo superior a partir del valor crítico, es decir, la probabilidad acumulada desde el punto crítico a infinito.

La variable tiene 10 grados de libertad, debe ubicarse la fila correspondiente a estos grados de libertad y en ellos ubicar el punto crítico 1.3722. Luego leer en la fila correspondiente a Areas de extremo superior la probabilidad desde 1.3722 hasta infinito.

La probabilidad de que la variable asuma valores inferiores a 1.3722 es:

$$P(t_{10} \leq 1.3722) = 1 - P(t_{10} \geq 1.3722) = 1 - 0.10 = 0.90$$

#### 4.9 Los grados de libertad

En el numerador del estadístico chi-cuadrado,

$$\sum_{i=1}^n (X_i - \bar{X})^2$$

el número total de cuadrados en esta expresión es  $n$ , pero sólo hay  $(n-1)$  cuadrados independientes porque una vez calculados los  $n-1$  primeros, el valor del último queda determinado automáticamente. La razón de ello es la presencia de  $\bar{X}$  en la expresión y el hecho de que una de las características bien conocidas de  $\bar{X}$  es que

$$\sum_i (X_i - \bar{X}) = 0$$

Esto representa una restricción que debe cumplirse. Por ejemplo, supongamos que tenemos tres cuadrados y que los valores de los dos primeros son

$$(x_1 - \bar{X})^2 = 2^2$$

$$(x_2 - \bar{X})^2 = 4^2$$

Si el tercer cuadrado fuese independiente de los otros dos podría tener cualquier valor, por ejemplo  $5^2$ . Pero esto no puede suceder, porque en este caso tendríamos

$$\sum_i (X_i - \bar{X}) = 2 + 4 + 5$$

cuya suma es 11 y no cero como tendría que ser. En realidad si los dos primeros cuadrados son  $2^2$  y  $4^2$  el tercero tiene que ser necesariamente  $(-6)^2 = 36$ , porque sólo en este caso obtenemos  $2+4-6=0$ .

En síntesis, los grados de libertad hacen referencia al número de cuadrados independientes.

**Ejemplo.** Tres personas entran a un comercio a comprar chocolate, pero en el lugar solo hay tres chocolates de distinto tipo. La primer persona que entra tiene la posibilidad de elegir entre los tres chocolates. La segunda persona tiene posibilidades de elegir entre dos. La tercer persona no tiene posibilidades de elegir, si quiere chocolate debe comprar el último que quedó. El no poder elegir es similar a la pérdida de grados de libertad.

### **CASOS DE ESTUDIO, PREGUNTAS Y PROBLEMAS**

Caso 4.1: Edades y estaturas del curso Inferencia Estadística.

El tema de estudio es la distribución de las edades y las estaturas en el curso de Inferencia Estadística del presente ciclo lectivo.

El objetivo general es conocer las medidas antropométricas de la población; específicamente se busca hallar la media de estatura y la media de edad, adicionalmente utilizar la teoría de distribuciones en el muestreo para inferir datos a la población.

Se parte de las siguientes premisas:

- la existencia de datos extremos en la muestra aumenta la dispersión de la media
- a medida que aumenta el número de muestras, la media de las muestras tiende a la media de la población

Para realizar este estudio se debe:

- a) Generar una tabla de datos de n observaciones por 2 variables cuantitativas, donde la unidad de observación sean los alumnos de Inferencia Estadística y las variables edad y estatura.
- b) Cada alumno debe
  1. seleccionar cinco muestras de tamaño cinco.
  2. calcular la media de cada muestra
  3. calcular la media de las cinco muestras

- c) Generar una tabla con las medias de cada una de las cinco muestras generadas por cada alumno, el número de observaciones de esta tabla será  $n \times 5$ , y calcular la media de las medias.
- d) Generar una tabla con la media de cada alumno, el número de observaciones de esta tabla será  $n$ , y calcular la media de las medias.
- e) En la tabla generada en a), calcular la media de la población.
- f) Comparar las medias obtenidas en c), d) y e).

#### Caso 4.2: Ingreso medio en los hogares de Río Cuarto

La tabla clasifica a los hogares de Río Cuarto de acuerdo al nivel de ingresos declarados a la Encuesta Permanente de Hogares en Mayo de 2003.

Con esta información, y teniendo en cuenta que la cantidad de hogares relevados fue de 636, se debe calcular:

1. el ingreso medio de los hogares
2. el desvío del ingreso
3. la probabilidad de que un hogar, extraído al azar de la población, tenga
  - ingreso superior a \$3000.00
  - ingreso inferior a \$750.00
  - ingreso entre \$800.00 y \$1300.00
  - ingreso inferior a \$125.00
  - ingreso inferior a \$2400.00

Nivel de ingreso mensual del hogar	Total de hogares (en %)
Hasta 150	4.93
Entre 151 y 300	13.21
Entre 301 y 450	13.41
Entre 451 y 600	13.41
Entre 601 y 750	11.64
Entre 751 y 1000	13.21
Entre 1001 y 1250	7.49
Entre 1251 y 1500	4.14
Entre 1501 y 2000	4.93
Entre 2001 y 3000	3.16
Más de 3000	2.37
Ingreso desconocido	8.10
Total de hogares	100.00

FUENTE: Encuesta Permanente de Hogares. Mayo de 2003

## Preguntas

4.1. ¿Porqué es válida la siguiente igualdad?

$$\frac{\sum_{i=1}^n (x_i - \bar{x})^2}{n-1} = \frac{\sum_{i=1}^n x_i^2 - \frac{\left(\sum_{i=1}^n x_i\right)^2}{n}}{n-1},$$

siendo:  $\sum_{i=1}^n x_i^2$  la suma del cuadrado de cada observación

$\left(\sum_{i=1}^n x_i\right)^2$  el cuadrado de la suma total

Demostrar

## Problemas

4.1. Se encontró que un grupo de calificaciones de examen final de un curso de introducción a la estadística, tenía distribución normal con una media de 73 y una desviación estándar de 8.

1. ¿cuál es la probabilidad de obtener cuando mucho una calificación de 91, en este examen?
2. ¿qué porcentaje de estudiantes calificó entre 65 y 89?
3. ¿qué porcentaje de estudiantes calificó entre 81 y 89?
4. ¿cuál debe ser la calificación final del examen, si solo el 5% de los estudiantes examinados tuvieron calificación más alta?

El profesor decide aprobar al 10% de los alumnos que tengan la mayor nota, sin que importe su calificación.

5. El alumno que haya obtenido una calificación de 81 en este examen, ¿aprueba o no?.

Con una calificación de 68 en un examen en que la media es 62 y la desviación es 3,

6. ¿el alumno aprueba o no?.
- 7.Cuál de las dos situaciones es más conveniente para el alumno? Múestrello estadísticamente y explíquello.

4.2. Calcular las siguientes probabilidades

$$P(\chi_{12}^2 \leq 26.2) = \quad P(\chi_{12}^2 > 6.3) =$$

$$P(5.23 \leq \chi_{12}^2 \leq 14.8) = \quad P(\chi_{12}^2 < X_0) = 0.90$$

4.3. Dada una variable G que se distribuye como una t de Student con 15 grados de libertad, calcular

$$P(g < 1.341) = \quad P(-2.131 < g < 2.131) =$$

$$P(g > -1.753) = \quad P(g > 0.691) =$$

4.4. Dada una variable aleatoria F se desea saber

$$P\left(\frac{1}{3.85} < F_{10,8} < 4.30\right) = \quad P(F_{10,8} > F_0) = 0.95$$

$$P(3.35 < F_{10,8} < 4.30) = \quad P(F_{10,8} > 4.30) =$$

### **Bibliografía**

- Berenson, Mark y Levine, David. (1996) *Estadística Básica en Administración*. Prentice Hall. México.
- Daniel, Wayne W. (1999) *Bioestadística, base para el análisis de las ciencias de la salud*. Tercera Edición. Editorial Limussa. México.
- Dixon W.J. y Massey F.J. (1957) *Introduction to Statistical Analysis*. Nueva York, McGraw-Hill.
- Hernández Sampieri, Roberto; Fernández Collado, Carlos y Baptista Lucio, Pilar. (2000) *Metodología de la Investigación*. Segunda Edición. McGraw Hill. México.
- Kazmier, Leonard y Diaz Mata, Alfredo. (1993) *Estadística Aplicada a la Administración y a la Economía*. Mc.Graw Hill. México.
- Kinnear, T. Taylor, J. (1993) *Investigación de Mercados. Un enfoque aplicado*. Mc.Graw Hill.
- Mao, James C.T. (1980) *Análisis Financiero*. El Ateneo. Buenos Aires.
- Mendenhall, W. Wackerly, D. Scheaffer, R. (1990) *Estadística Matemática con Aplicaciones*. Grupo Editorial Iberoamerica. 2ª Edición.
- Meyer, P.L. (1973) *Probabilidad y Aplicaciones Estadísticas*. Fondo Educativo Interamericano. México.
- Padua, Jorge. (1996) *Técnicas de Investigación Aplicadas a las Ciencias Sociales*. Fondo de Cultura Económica. México.
- Tramutola, C.D. *Modelos Probabilísticos y Decisiones Financieras*. E.C.Moderan. Lectura de Administración de Empresas. Capital Federal.



Capítulo 5. DISEÑO DE MUESTREO .....	163
5.1 Teoría estadística .....	164
5.2 Tamaño de la muestra .....	167
Error, riesgo y tamaño muestral.....	168
Relaciones entre error, riesgo y n.....	169
Desviación estándar desconocida .....	169
5.3. Selección de Muestras.....	171
Criterios del diseño de la muestra .....	171
5.4 Muestras informales y casuales .....	172
5.5 Muestras Probabilística.....	172
Muestreo Aleatorio Simple.....	174
Muestreo Sistemático .....	174
Muestreo Estratificado. ....	175
Muestreo De Conglomerados .....	177
Diseños De Etapas Múltiples .....	178
CASOS DE ESTUDIO, PREGUNTAS Y PROBLEMAS .....	178
Caso 5.1: Cooperativa de Alimentos.....	178
Problemas.....	182
Bibliografía .....	182



## Capítulo 5. DISEÑO DE MUESTREO

Muestreo aparece en el tercer paso de una investigación estadística y es aquella teoría que establece los procedimientos que permiten generalizar sobre la población a través del estudio de una parte de la misma, es decir, a través del estudio de la muestra.

Cada observación o elemento tomado de la población contiene cierta cantidad de información acerca del parámetro o parámetros de interés. Ya que la información cuesta dinero, el investigador debe determinar cuánta información debe comprar: demasiado poca impedirá realizar buenas estimaciones; mientras que mucha información ocasiona un despilfarro de dinero. La cantidad de información depende del número de elementos muestreados y de la cantidad de variación de los datos, ambos fenómenos pueden ser controlados a través del diseño de la encuesta y el tamaño de la muestra.

Esta introducción a los métodos estadísticos para la selección de muestras aleatorias en marcos poblacionales determinados, no se dedica al estudio de métodos específicos de muestreo; sino que trata de elaborar cuáles serán los objetivos de la aplicación de metodologías estadísticas a problemas de investigación. Constituye las respuestas al porqué, cómo y dónde, algunas razones y motivaciones para usar buenos métodos de muestreo, una visión panorámica de los problemas básicos y de los métodos para resolverlos, y una indicación de la manera que el concepto de muestreo de poblaciones cabe dentro de los métodos de muestra y dentro de la búsqueda general de conocimiento científico.

El diseño de muestra tiene dos aspectos: un proceso de selección, que consiste en la regla y operaciones mediante las cuales se incluyen en la muestra algunos miembros de la población, y un proceso de estimación para calcular los estadísticos de la muestra, que son estimaciones muestrales de los parámetros (valores) de la población.

En la teoría del muestreo son importantes las siguientes definiciones:

*Elemento*: objeto sobre el cual se toman las mediciones.

*Población*: conjunto de elementos acerca de los cuales se desea hacer alguna inferencia, es el universo de referencia.

*Unidades de muestreo*: conjunto no superpuestos de elementos de la población que cubren la población completa. Si bien en el estudio necesito encuestar individuos, es cierto que un hogar significa un conjunto de individuos (es decir un conjunto de elementos) y el proceso de seleccionar hogares, y dentro de los hogares seleccionar el elemento, puede resultar más eficiente siempre y cuando la persona no sea encuestada dos veces. Es posible que el número de elementos y el número de unidades de muestreo coincidan, esto es así cuando se muestrean individuos en lugar de hogares.

*Marco*: es una lista de unidades de muestreo. La selección del elemento (individuo) puede hacerse directamente del marco, esto es si se poseen listas de individuos. También pueden darse marcos

múltiples: primero seleccionando viviendas y dentro de las viviendas individuos.

**Muestra:** conjunto de unidades seleccionadas de un marco o de varios marcos. De la muestra se obtendrán los datos objeto de la investigación que se utilizará para describir la población y realizar estimaciones sobre ella.

**Ejemplo.** Una importante firma industrial alimenticia quiere conocer la proporción de habitantes de cierta población que han consumido una nueva línea de polvos para helados de reciente aparición en el mercado.

Elemento será la persona que habita en esa población y que fue seleccionada por algún método de muestreo para que respondiera acerca de su conocimiento o no del producto.

Población son todos los habitantes de la población que superen un cierto límite de edad a definir por el investigador.

## 5.1 Teoría estadística

Una pregunta práctica en gran parte de la investigación tiene que ver con la determinación del tamaño de la muestra. La decisión del tamaño de la muestra está directamente relacionada con el costo de la investigación y por tanto debe ser justificada.

Las características poblacionales de interés son la *media*, la *varianza* y la *desviación estándar* de la población.

Normalmente estas medidas son desconocidas, y nuestra meta consiste en determinar su valor lo más aproximadamente posible, tomando una muestra de la población.

**Ejemplo.** El interés está centrado en conocer el ingreso promedio de los 57 clientes de un banco que poseen la tarjeta de crédito AA. La población son los 57 clientes y los datos se ordenan en 8 intervalos de clase de acuerdo al ingreso que posea cada cliente.

Intervalo de clase	Punto medio de la clase $m_j$	Frecuen de clase Absoluta $f_j$	$m_j f_j$	$(m_j - \mu)$	$(m_j - \mu)f$	$(m_j - \mu)^2 f_j$
De \$150.00 a \$300.99	225.00	6	1350.00	-606.58	-3639.48	2207635.78
De \$301.00 a \$500.99	400.00	13	5200.00	-431.58	-5610.54	2421396.85
De \$501.00 a \$700.99	600.00	13	7800.00	-231.58	-3010.54	697180.85
De \$701.00 a \$1000.99	850.00	8	6800.00	18.42	147.36	2714.37
De \$1001.00 a \$1250.99	1125.00	5	5625.00	293.42	1467.10	430476.48
De \$1251.00 a \$1500.99	1375.00	3	4125.00	543.42	1630.26	885915.89
De \$1501.00 a \$2000.99	1750.00	6	10500.00	918.42	5510.52	5060971.78
Más de \$2000	2000.00	3	6000.00	1168.42	3505.26	4095615.89
Sumas		57	47400		0.00	15801907.89

La medida promedio son los ingresos de los 57 clientes del banco que poseen la

tarjeta AA, lo cual se denomina media de la población.  $\mu = \frac{\sum_{i=1}^n m_i f_i}{\sum_{i=1}^n f_i} = 831.58$

La varianza es el grado en el que el ingreso de un cliente difiere del ingreso promedio de la población,  $\mu$ , y se la denomina varianza de la población

$$\sigma^2 = \frac{\sum_{i=1}^n (m_i - \mu)^2 f_i}{\sum_{i=1}^n f_i} = 277226,45$$

El desvío estándar de la población es la raíz cuadrada de la varianza

$$\sigma = \sqrt{\sigma^2} = \sqrt{277226,45} \cong 526.52$$

Del mismo modo que la *población* tiene un conjunto de características, cada *muestra* también tiene un conjunto de características. Una de ellas es el promedio o media muestral  $\bar{X}$  que, por ser una característica muestral, cambiará si se obtuviera una nueva muestra.

**Ejemplo.** Se toma una muestra aleatoria, de tamaño 10, de los clientes del banco que poseen la tarjeta de crédito AA. Las diez personas seleccionadas y el ingreso respectivo se muestran en la tabla:

	$x_j$	$x_j - \mu$	$(x - \mu)^2$
1	1700.00	907.00	822649.00
2	900.00	107.00	11449.00
3	550.00	-243.00	59049.00
4	1300.00	507.00	257049.00
5	300.00	-493.00	243049.00
6	600.00	-193.00	37249.00
7	350.00	-443.00	196249.00
8	1100.00	307.00	94249.00
9	450.00	-343.00	117649.00
10	680.00	-113.00	12769.00
$\Sigma$	7930.00	0.00	1851410.00

La media muestral es

$$\bar{x} = \frac{\sum_{i=1}^n x_i}{n} = \frac{7930.00}{10} = 793.00$$

La varianza de la muestra

$$s^2 = \frac{\sum_{i=1}^n (x_i - \mu)^2}{n-1} = \frac{1851410.00}{9} = 205712.22$$

El desvío de la muestra

$$s = \sqrt{s^2} \cong 453.56$$

$S^2$  es una estimación sin sesgo de la varianza de la población porque incluye el término  $n-1$ .

Otra característica es la varianza de la muestra,  $S^2$ . Esta será pequeña si las respuestas de la muestra son similares, y grande si se encuentran dispersas.  $S^2$  es la varianza de la población estimada a partir de los datos de la muestra.

Si el tamaño de la población, denominado  $N$ , es pequeño con relación al tamaño de la muestra  $n$ , debe ser añadido un "factor de corrección de población finita"

$$\frac{N-n}{N-1}$$

De modo que la varianza es

$$S^2 = \frac{N-n}{N-1} \left[ \frac{\sum (X_i - \bar{X})^2}{n-1} \right]$$

Si  $N$  es más de 10 veces el tamaño de la muestra, el factor de corrección tiene rara vez significado.

Desde luego, todas las muestras no generarán el mismo valor de  $\bar{X}$  (ó s). El punto de importancia, es que  $\bar{X}$  variará de muestra a muestra.

La varianza de  $\bar{X}$  es medida por su error estándar:

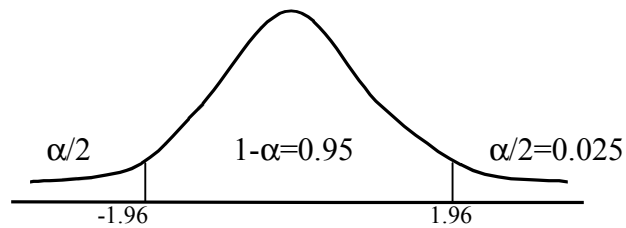
$$\sigma_{\bar{X}} = \text{error estándar de } \bar{x} = \sigma / \sqrt{n}$$

La varianza en  $\bar{X}$  será más grande a medida que la varianza de la población,  $\sigma^2$ , sea más grande. A medida que aumente el tamaño de la muestra, la variación en  $\bar{X}$  disminuirá.

La media de la muestra,  $\bar{X}$ , tiene una distribución de probabilidad normal (con media igual a  $\mu$  y varianza igual a  $\sigma_{\bar{X}}^2$ ), cualquiera sea la distribución de probabilidad de la población, aunque si esta no es normal se requerirá de que la muestra sea lo suficientemente grande de acuerdo al Teorema del Límite Central.

Por consiguiente,  $\bar{X} \sim N(\mu, \sigma / \sqrt{n})$

Existe un nivel de confianza de 0.95 (95%) de que  $\bar{x}$  caiga dentro de  $\pm 1.96$  errores estándar de la media de la población  $\mu$ . 1,96 es el valor de la distribución normal estándar para un 95% de probabilidad.



**Ejemplo.** En el estudio de los ingresos de los clientes del banco que poseen la tarjeta de crédito AA, si se han obtenido 100 muestras diferentes de 10 individuos, aproximadamente el 95% de las medias muestrales resultantes ( $\bar{X}$ ) estarán dentro de  $\pm 1.96$  errores estándar ( $\sigma_{\bar{x}} = 166.50$ ) de la media de la población ( $\mu = 831.58$ )

## 5.2 Tamaño de la muestra

Para determinar el tamaño de la muestra debe especificarse

- 1 El tamaño del error de la muestra que se desea.
- 2 El nivel de confianza, por ejemplo, del 95%.

El error de la muestra es

$$e = z \frac{\sigma}{\sqrt{n}}$$

de donde se deduce el tamaño de la muestra

$$n = \frac{z^2 \sigma^2}{e^2}$$

**Ejemplo.** Existe la necesidad de estar un 95% seguros de que el error muestral, al estimar la media de la población, no exceda de 100 pesos. En este caso, el error muestral ( $e$ ) es igual a 100, y puesto que el nivel de confianza es del 95%,  $z$  asume el valor 1.96. La desviación estándar de la población, es de 531.20, por tanto el tamaño de la muestra debe ser:

$$n = \frac{(1.96)^2 (531.20)^2}{100^2} \cong 108$$

El tamaño de la muestra es independiente del tamaño de la población y dependiente de:

Nivel de Confianza	$> z \Rightarrow > n$
Desviación estándar de la población	$> \sigma \Rightarrow > n$

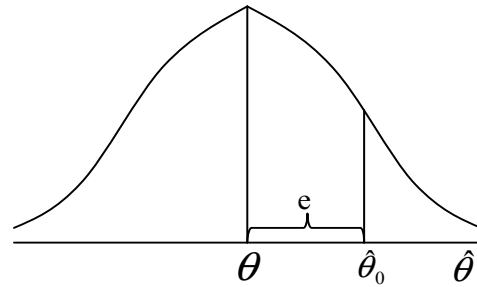
## Error, riesgo y tamaño muestral

Dado el estimador  $\hat{\theta}$ , que se distribuye normal con parámetros  $(\theta, \sigma_{\hat{\theta}})$  y es estimador insesgado del parámetro  $\theta$

$$\hat{\theta} \sim N(\theta, \sigma_{\hat{\theta}}) \quad E(\hat{\theta}) = \theta.$$

El error de estimación es la diferencia entre el valor obtenido en la estimación puntual,  $\hat{\theta}_0$ , y el valor del parámetro,  $\theta$

$$e = \hat{\theta}_0 - \theta$$



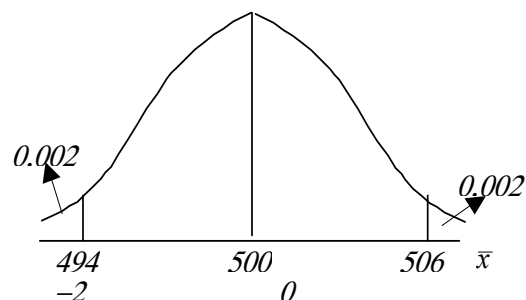
Si el error máximo aceptable por el investigador es  $e$ , habrá un riesgo de cometer un error superior a él. Es decir, un riesgo de que  $|\hat{\theta} - \theta| > e$ . El riesgo es la probabilidad de que la diferencia entre el parámetro y su estimador supere la cuantía de  $e$ .

$$P(|\hat{\theta} - \theta| > e) = \text{riesgo de cometer un error superior a } e.$$

**Ejemplo.** Suponga que  $X \sim N(500, 30)$ ,  $n = 100$ ,  $|e| > 6$  y  $\hat{\mu} = \bar{X}$ . Se quiere saber cuál es el riesgo de cometer un error de estimación de  $\mu$  superior a 6.

$$z = \frac{\bar{x} - \mu}{\sigma_{\bar{x}}} = \frac{e}{\sigma/\sqrt{n}} = \frac{e\sqrt{n}}{\sigma},$$

$$\text{por lo tanto } z = \frac{6\sqrt{100}}{30} = \pm 2$$



La probabilidad (o sea, el riesgo) de que al efectuar la estimación puntual se cometa un error en valor absoluto mayor a 6 es

$$P(|z| > 2) = 1 - P(-2 \leq z \leq 2) = 1 - (0.9772 - 0.0228) = 0.0456 = \text{riesgo}$$



## Relaciones entre error, riesgo y n

a) Riesgo

- $$z = \frac{e\sqrt{n}}{\sigma}$$
- el riesgo es función inversa de  $z$ . Cuando  $z$  crece el riesgo disminuye.
  - $z$  depende del error, de  $n$  y de la varianza poblacional, por lo tanto el riesgo, que es función de  $z$ , también depende de ellos.
  - si aumenta el error  $e$ , disminuye el riesgo, aumenta  $z$
  - si aumenta  $n$ , disminuye el riesgo, aumenta  $z$
  - si aumenta  $\sigma^2$ , aumenta el riesgo y disminuye  $z$

b) Error

- $$e = \frac{z\sigma}{\sqrt{n}}$$
- si aumenta  $n$ , disminuye el error.
  - si aumenta  $\sigma^2$ , aumenta el error

c) Tamaño muestral

- $$n = \frac{z^2\sigma^2}{e^2}$$
- si aumenta  $\sigma^2$ , aumenta  $n$  si se desea mantener el mismo error y el mismo riesgo. Podemos calcular  $n$  para estimar  $\mu$  si se conocen error, riesgo y  $\sigma^2$  poblacional.

## Desviación estándar desconocida

El procedimiento que se acaba de mostrar *supone* que la desviación estándar de la población es conocida. En la mayor parte de las situaciones prácticas no es conocida y debe ser estimada a partir de

- Usar una desviación estándar de la muestra obtenida de una encuesta anterior comparable o de una encuesta piloto.
- Estimar  $\sigma$  subjetivamente
- Usar proporciones.

Los dos primeros casos se resuelven de la manera desarrollada. Para el tercer caso, el procedimiento es usar una proporción de la muestra para *estimar* una proporción de la población desconocida,  $\pi$ .

El error muestral usando proporciones es de

$$e = z\sigma_p$$

donde  $\sigma_p$  es el desvío de la proporción en la población.

En la distribución binomial el desvío es

$$\sigma = \sqrt{n\pi(1-\pi)}$$

Su estimación a partir de datos muestrales es

$$\sigma_p = \frac{\sqrt{n\pi(1-\pi)}}{n} = \frac{\sqrt{\pi(1-\pi)}}{\sqrt{n}}$$

donde,  $\pi$  = la proporción de la población

$p$  = la proporción en la muestra, usada para estimar la proporción en la población desconocida

$\sigma_p^2$  = la varianza poblacional de  $p$ .

Reemplazando  $\sigma_p$  se obtiene

$$e = z \frac{\sqrt{\pi(1-\pi)}}{\sqrt{n}}$$

De modo que la fórmula para el cálculo del tamaño de muestra utilizando proporciones es

$$n = \frac{z^2 \pi(1-\pi)}{e^2}$$

En el peor de los casos, en donde la varianza de la población es máxima, ocurre cuando la proporción poblacional es igual a 0.50:

$$\pi(1 - \pi) = 0.25$$

$$\pi = 0.50$$

Debido a que la proporción poblacional es desconocida, un procedimiento común consiste en suponer el peor de los casos. La fórmula para el tamaño de la muestra se simplifica entonces a:

$$n = \frac{z^2 (0.25)}{e^2}$$

De este modo, si la proporción de la población debe ser estimada dentro de un error, de 0.05 (ó 5 %) a un nivel de confianza del 95%, el tamaño necesario de la muestra es:

$$n = \frac{(1.96)^2 (0.25)}{(0.05)^2} \cong 384$$

Generalmente, un instrumento de encuesta o un experimento no se basará sólo en una pregunta. Algunas veces, cientos de ellas pueden estar involucradas, pero no será necesario pasar por todo este proceso en todas las preguntas. Generalmente se toman unas cuantas preguntas representativas y se determina el tamaño de la muestra a partir de ellas.

En el análisis deben incluirse las preguntas más importantes para el estudio y que tengan la varianza esperada más alta.

### 5.3. Selección de Muestras

El muestreo estudia los métodos para seleccionar y observar una parte de la población con el fin de hacer inferencias acerca de toda la población. Una muestra puede tener varias ventajas sobre un censo completo:

- a) economía,
- b) rapidez y oportunidad;
- c) posibilidad de hacerse (si la observación es destructiva, el empleo de un censo no es práctico);
- d) calidad y precisión (en algunas situaciones no hay dinero suficiente para pagar el personal adiestrado y los supervisores necesarios para realizar un buen censo, o aun para obtener una muestra grande).

Los censos completos poseen ventajas especiales en algunas situaciones:

- a) se pueden obtener datos para unidades pequeñas;
- b) la aceptación pública es más fácil de alcanzar para datos completos;
- c) la colaboración y la respuesta del público se pueden obtener más fácilmente.

#### Criterios del diseño de la muestra

1. Orientación hacia la meta. El diseño completo, tanto en la selección como en la estimación, debe orientarse a los objetivos de la investigación, hechos a la medida del diseño de la encuesta y ajustados a las condiciones de la encuesta.
2. La medibilidad es una característica de los diseños que permite calcular, a partir de la propia muestra, estimaciones válidas o aproximaciones de su variabilidad de muestreo. Esto se suele expresar en las encuestas con los errores estándares, pero a veces, pueden utilizarse otras expresiones de la función de verosimilitud o de la distribución de muestreo. Esta es la base necesaria para la inferencia estadística, y sirve como puente científico y objetivo, entre el resultado de la muestra y el valor desconocido de la población.
3. La practicidad se refiere a los problemas que deben resolverse para llevar a cabo el diseño esencialmente como se propuso. Una muestra probabilística no puede crearse por suposición, ni estará dada, como sucede en los problemas teóricos. El método de los muestreadores de cuota a sus entrevistadores: "vayan y obtengan una muestra aleatoria", es sumamente impráctico; ni el entrevistador ni el que lo

envía pueden hacerlo. Se requiere de cuidado para traducir el modelo de selección teórico a un conjunto de instrucciones de oficina y campo. Estas instrucciones deben ser simples, claras, prácticas y completas. Por ejemplo, para identificar un segmento de muestra, no se le debe pedir al entrevistador que localice una línea marcada arbitrariamente en un mapa; sus deberes deben confinarse a localizar calles y direcciones.

4. La economía se refiere a cumplir los objetivos de la encuesta con un costo mínimo y al grado en que se alcanza este objetivo.

En líneas generales hay dos formas de obtener una muestra: de manera informal y casual, o bien, de manera probabilística.

#### 5.4 Muestras informales y casuales

Las muestras no probabilísticas constituyen un problema en la inferencia pues no hay una forma probabilística de estimar qué tan representativas son esas muestras seleccionadas. Los procedimientos usados en este tipo de muestreo son:

- muestras de juicio: el entrevistador selecciona a cualquier sujeto que desee.
- muestras de cuotas: son muestras de juicios pero con previa asignación de cuotas por sexo, edad, clase social, raza, entre otros, que tratan de simular características conocidas de la población.
- trozo de pastel: en este caso el entrevistador no interviene en el proceso de selección pues consiste en una autoselección, personas que responden a un cupón, que concurren a un centro de exhibición o el público de un teatro en particular, por ejemplo.

Los peatones pueden ser interrogados en cuanto a sus opiniones de un nuevo producto. Si la respuesta de todos en la población es uniforme, todos ellos lo odian o lo aman, tal enfoque puede ser satisfactorio.

#### 5.5 Muestras Probabilística

Todos los miembros de la población tienen una probabilidad conocida de estar en la muestra.

Una muestra probabilística tiene las ventajas de:

- permitir al investigador demostrar la representatividad de la muestra.

- permitir un planteamiento explícito en cuanto a la cantidad de variación que será introducida, porque se usa una muestra en lugar de un censo de la población.
- hacer posible la identificación más explícita de las probables desviaciones.

El objetivo de un muestreo probabilístico es hacer una inferencia acerca de la población con base en la información contenida en la muestra.

Existen dos factores que pueden afectar la información contenida en la muestra:

- ⇒ El tamaño de la muestra
- ⇒ La cantidad de variación en los datos (que puede ser controlada por el *método* de selección de una muestra)

Las unidades de muestreo contienen los elementos, y se usan para seleccionarlos en la muestra. En el muestreo de elementos, cada unidad de muestreo contiene solamente un elemento; pero en el muestreo de conglomerados cualquier unidad de muestreo, llamada conglomerado, puede contener varios elementos.

**Ejemplo.** Una muestra de estudiantes se puede obtener de una muestra de aulas; o una muestra de viviendas, de una muestra de manzanas de la ciudad.

Una misma encuesta puede usar diferentes clases de unidades de muestreo, y en muestreo polietápico se usa una jerarquía de unidades de muestreo o conglomerados, de manera que el elemento pertenezca únicamente a una unidad de muestreo en cada etapa.

**Ejemplo.** Una muestra de los habitantes de una Región puede tomarse al seleccionar sucesivamente los municipios, las manzanas, las viviendas y, finalmente, las personas.

Las unidades de listado se usan para identificar y seleccionar unidades de muestreo a partir de listas. A veces se necesitan procedimientos detallados para convertir listados en unidades de muestreo, como por ejemplo, para convertir un listado de direcciones en viviendas y hogares. Los problemas pueden ser serios si los elementos no se identifican unívocamente con los listados. Por ejemplo, una muestra de familias tomada de los listados de teléfonos puede involucrar serias dificultades.

Entre los métodos de selección de una muestra probabilística se encuentran:

1. Muestreo aleatorio simple
2. Muestreo sistemático
3. Muestreo estratificado
4. Muestreo por conglomerados
5. Muestreo por etapas múltiples

## Muestreo Aleatorio Simple

El muestreo aleatorio simple es un enfoque en el cual cada miembro de la población, y por tanto cada muestra posible, tiene una probabilidad igual de ser seleccionado.

Las muestras aleatorias simples pueden ser seleccionadas mediante el uso de una tabla de números aleatorios. La manera de hacer inferencias es estimar ciertos parámetros de la población utilizando la información de la muestra.

Frecuentemente, el objetivo es estimar una media poblacional (parámetro  $\mu$ ) o un total poblacional (parámetro  $\tau$ ). Otras veces, se requiere estimar la proporción poblacional (parámetro  $\pi$ ).

Para lograr la estimación se usa el promedio muestral (estadístico  $\hat{\mu}$ ), el estimador del total poblacional (estadístico  $\hat{\tau}$ ) y la proporción muestral (estadístico  $\hat{\pi}$  o  $p$ ), donde:

$$\hat{\mu} = \bar{x} = \frac{\sum_{i=1}^n x_i}{n} ; \quad \hat{\tau} = N\hat{\mu} ;$$

siendo,

N, tamaño de la población

n, tamaño de la muestra

$x_i$ , datos muestrales,  $i = 1, \dots, n$

En el muestreo aleatorio simple, el tamaño de la muestra, es decir, la cantidad de observaciones necesarias para estimar un parámetro poblacional con un límite para el error de estimación (asumir un riesgo determinado de cometer dicho error), viene dado, por

$$n = \frac{z^2 \sigma^2}{e^2}$$

## Muestreo Sistemático

Consiste en esparcir sistemáticamente la muestra a lo largo de la lista de miembros de la población. Si la población tiene 10.000 individuos y se desea un tamaño de muestra de 1.000, cada décima persona es seleccionada para la muestra.

Aunque en casi todos los ejemplos prácticos tal procedimiento generaría una muestra equivalente a una muestra aleatoria simple, el investigador debe estar consciente de las regularidades dentro de la lista.

Se debe hallar primero la frecuencia de extracción de elementos, es decir, cada cuántos elementos se extrae uno. Esto se logra haciendo

$$\frac{N}{n} = K$$

Para determinar cuál es el primer elemento de la muestra, se selecciona de una tabla de números aleatorios un valor inferior a  $K$ . A este elemento se lo denomina arranque aleatorio ( $a$ ). El segundo individuo será el  $a+K$ , el tercero  $a+K+K$  y así sucesivamente.

La ventaja de este método de selección es su practicidad; la desventaja surge a partir de la determinación de los elementos  $a$  y  $K$ , donde algunas unidades de observación pasan a tener probabilidad cierta de ser seleccionadas y otras probabilidad nula.

### Muestreo Estratificado.

En el muestreo aleatorio simple, una muestra aleatoria se selecciona de una lista, o de un marco muestral, que representa a la población.

Al desarrollar un plan de muestreo, es aconsejable buscar subgrupos naturales que sean más homogéneos que la población total. Tales subgrupos se denominan "estratos".

Este tipo de muestreo es conveniente cuando la población o universo puede ser dividido en categorías estratos o grupos que reúnen cierto interés analítico y que por razones teóricas y empíricas presentan diferencias entre ellos.

1. La estratificación puede producir un límite más pequeño para el error de estimación que el que se generaría por una muestra aleatoria del mismo tamaño. Este resultado es particularmente cierto si las mediciones dentro de los estratos son homogéneas.
2. El costo por observación en la encuesta puede ser reducido mediante la estratificación de los elementos de la población en grupos convenientes.
3. Se pueden obtener estimaciones de los parámetros poblacionales para subgrupos de la población. Los subgrupos deben ser entonces estratos identificables.

**Ejemplo.** Se necesita información sobre las actitudes de los estudiantes hacia una nueva instalación atlética dentro de la Universidad. Se conoce que existen tres grupos de estudiantes con características diferenciadas:

A) **los que viven en residencias estudiantiles** tienen actitudes muy homogéneas hacia la instalación propuesta, la variación o la varianza en sus actitudes es muy pequeña.

B) **residentes en la ciudad**, son menos homogéneos.

C) **residentes fuera de la ciudad**, varían ampliamente en sus opiniones.

En tal situación, en lugar de permitir que la muestra provenga de la totalidad de los tres grupos aleatoriamente, será más prudente tomar un menor número de miembros del grupo de residentes y extraer más del grupo ajeno al campo. Se particiona la lista de los estudiantes en los tres grupos y se extrae una muestra aleatoria simple de cada uno de los tres grupos.

Una muestra estratificada puede ser

1. Proporcional, donde la fracción de muestreo es igual en cada estrato de la muestra que la existente en la población
2. No proporcional

La estimación de la media de la población, en el muestreo estratificado, es un promedio ponderado de las medias de las muestras encontradas en cada estrato:

$$\mu_X = \sum_i \pi_i \bar{x}_i$$

donde,  $\bar{x}_i$  = la media de la muestra para el estrato i

$\pi_i$  = la proporción de la población en el estrato i

**Ejemplo.** El propósito de la investigación es estudiar el rendimiento escolar según sea su extracción de clase social. Para ello, el investigador se sitúa en una escuela a la cual concurren 500 alumnos y de los cuales le informan la siguiente composición de clase social

Población		Muestra	
Clases sociales	Elementos	Proporcional	No proporcional
Alta	50	5	25
Media	300	30	25
Baja	150	15	25
Total	500	50	75

Se quiere que el tamaño de la muestra sea del 10% de la población.

Para realizar un muestreo proporcional se aplica esa fracción de muestreo a cada estrato. Ahora bien en el estrato Clase Alta solo se tienen 5 casos y, puede ocurrir, que no alcancen para realizar ciertas estimaciones. Puede establecerse que se necesita 25 casos en cada clase para poder realizar el estudio. En este caso el tamaño muestral es de 75, y los resultados a obtener dentro de cada estrato se deben ponderar por el peso del estrato en la población a efectos de obtener el resultado. En el caso del muestreo proporcional esto no ocurría porque era autoponderado.

*¿Cómo se determina la mejor aplicación del presupuesto de muestreo a diversos estratos?*

Este problema clásico de muestreo fue solucionado en 1935 por Jerzy Neyman, a partir de la siguiente expresión:

$$n_i = \frac{\pi_i \sigma_i / \sqrt{c_i}}{\sum_i \pi_i \sigma_i / \sqrt{c_i}} n$$

donde,

$n_i$  = el tamaño de la muestra para el estrato i

$\pi_i$  = la proporción de la población en el estrato i

$\sigma_i$  = la desviación estándar de la población en el estrato i

$c_i$  = el costo de una entrevista en el estrato i

$\Sigma_i$  = la suma a lo largo de todos los estratos

$n$  = tamaño de la muestra



**Ejemplo.** La tabla presenta información de la encuesta sobre el uso mensual de los cajeros automáticos. La población se encuentra estratificada por ingreso. El segmento de ingresos *altos* tiene la variación más alta y el costo de entrevista más alto. Los estratos de ingresos medios y bajos tienen el mismo costo de entrevista, pero difieren con respecto a la desviación estándar del uso de los procesadores bancarios.

Estrato (i)	Proporción ( $\pi_i$ )	Desviación estándar ( $\sigma_i$ )	Costo ( $c_i$ ) entrevista	$\pi_i \sigma_i / \sqrt{c_i}$	$n_i$
Bajo	0.3	1	25	0.06	177
Medio	0.5	2	25	0.20	588
Alto	0.2	4	100	0.08	235
				$\Sigma_i = 0.34$	$n=1000$

La última columna presenta la división de la muestra de 1.000 personas en los tres estratos. Al estrato de ingresos altos le corresponden 235 personas que representan el 23.5% de la muestra; sin embargo, la proporción de personas de altos ingresos en la población es de 20.0%. Si se hubiera seleccionado una muestra aleatoria simple de tamaño 1000, 200 personas pertenecerían al estrato de altos ingresos.

La disponibilidad presupuestaria determina, en alguna medida, el tamaño de la muestra. Este es ajustado hacia arriba hasta que alcanza el límite presupuestal, de modo que el presupuesto debe ser calculado como:

$$\text{Presupuesto} = \sum_i c_i n_i$$

Conocido el tamaño de la muestra, se determina el error muestral y se decide si es excesivo o no. La fórmula del error muestral es:

$$\text{Error muestral} = z \sigma_{\bar{x}}$$

donde,

$$\sigma_{\bar{x}} = \sqrt{\sum_i \frac{\pi_i^2 \sigma_i^2}{n_i}}$$

Si el error muestral es excesivo se debe ampliar el presupuesto, lo que permite tomar un tamaño de muestra mayor; de no ser posible el proyecto debe desecharse.

## Muestreo De Conglomerados

En el muestreo de conglomerados, la población se divide nuevamente en subgrupos. En esta técnica se selecciona una muestra aleatoria de subgrupos y todos los miembros de los subgrupos forman parte de la muestra. Este método es útil cuando se pueden identificar aquellos subgrupos que sean representativos de la totalidad de la población.

**Ejemplo.** Se tomó una muestra de estudiantes universitarios de segundo año, que cursaban estadística en todas las facultades de Argentina, para estudiar la prevención del SIDA. Se contaba en ese momento con 200 comisiones de estadística, cada una de las cuales tenía, en promedio, 30 estudiantes.

El muestreo utilizado fue el de conglomerados: se seleccionaron 15 comisiones y, dentro en ellas, la totalidad de los alumnos. De modo que el tamaño de muestra fue de 450 alumnos.

Si se hubiera decidido tomar una muestra aleatoria simple reuniendo 450 alumnos en el total de 200 cursos de estadística, el costo sería significativamente mayor.

La gran pregunta, desde luego, es si los cursos son representativos de la población. Si los cursos de las áreas de ingresos superiores tienen diferentes opiniones acerca de la prevención del SIDA, que los cursos con estudiantes de ingresos más bajos, el supuesto que fundamenta al enfoque, no se mantendría.

La gran ventaja del muestreo de conglomerados es que su costo es más bajo. Los subgrupos o conglomerados son seleccionados de modo que, el costo para obtener la información deseada dentro del conglomerado, sea mucho más pequeño que si se obtuviera una muestra aleatoria simple.

### Diseños De Etapas Múltiples

Si se usan otros diseños muestrales, la lógica para generar el tamaño óptimo de la muestra aún se mantendrá; sin embargo, la fórmula puede complicarse.

**Ejemplo.** En un diseño por áreas, el primer paso puede ser el de seleccionar comunidades al azar. De este modo, el procedimiento puede ser escoger porciones de radios censales, luego manzanas y finalmente familias.

En tal diseño, la expresión para determinar el error estándar de  $\bar{X}$  se vuelve sumamente compleja. La situación consiste en repetir la totalidad del plan de muestreo y obtener dos, tres o cuatro estimaciones independientes de  $\bar{X}$ . Estas diferentes estimaciones pueden ser usadas para estimar el error estándar de  $\bar{X}$ .

---



---

## **CASOS DE ESTUDIO, PREGUNTAS Y PROBLEMAS**

### Caso 5.1: Cooperativa de Alimentos

J.L, Gerente General de una Cooperativa de Alimentos, se dio cuenta recientemente que había perdido contacto con los patrones de compra de los miembros de la cooperativa. Simplemente la cooperativa parecía mucho más grande ahora en comparación con los primeros años. El Gerente se preguntaba si podría hacer uso de algún tipo de dato que estuviese a su alcance con el propósito de ampliar su comprensión de los hábitos de compra de los miembros.

Esperaba poder utilizar este conocimiento para planificar mejor la mezcla y el volumen de artículos que la cooperativa ofrecía.

J.L. y un pequeño grupo de voluntarios fundaron la cooperativa en 1974. Esta cooperativa había aumentado de diez miembros iniciales en enero de 1974 a 500 miembros en setiembre de 1990. La empresa estaba localizada en una vieja bodega del noroeste de Milán, Michigan. Milán era una comunidad de 7500 personas localizada en el sureste de Michigan, aproximadamente a 40 millas al suroeste de Detroit. La cooperativa obtuvo sus miembros de una gran cantidad de comunidades alrededor del Milán, incluyendo Ann Arbor y Monroe.

Antecedentes. El objetivo de la cooperativa era proveer productos alimenticios de alta calidad a un precio por debajo de los que se ofrecían en los supermercados locales. Para este objetivo, la cooperativa utilizaba las cajas de empaque como estanterías, solicitaba la colaboración de los mismos compradores para que marcaran sus precios en los artículos, vendía sólo las mejores marcas del mercado y, por lo general, no ofrecía los "lujos" relacionados con los supermercados tradicionales. Para poder comprar en la cooperativa, las personas tenían que ser socios. La cuota de afiliación era de \$25 anuales. Cualquier ganancia obtenida por la cooperativa durante un año, se devolvía a los socios en forma de crédito para sus compras. J.L. estaba convencida de que los socios compraban la mayor parte de sus alimentos en la cooperativa.

Preocupaciones de J.L. durante los primeros años de la cooperativa. J.L. se sentía orgullosa de conocer a todos sus miembros. Había gastado una cantidad considerable de tiempo en la tienda y sentía que conocía lo que las personas compraban y cuánto dinero estaban gastando. A medida que creció el número de socios, sus deberes administrativos lo mantenían más tiempo en la oficina. Por tanto, ya no podía decir que conocía a todos los socios, ni que tenía idea de sus patrones de gastos. Deseaba conocer mejor estos aspectos de su negocio y pensó que tal vez parte de los datos que se habían recolectado con anterioridad sobre los miembros podrían proporcionarle respuestas.

### **Datos disponibles**

En junio de 1990 se utilizó un cuestionario para recolectar datos sobre los socios. Durante ese mes todos los miembros vinieron a la cooperativa por lo menos una vez. Por tanto existían datos sobre cada uno de los socios. Los datos consistían en las características demográficas de los miembros y en sus gastos semanales en alimentos.

Los datos se encontraban en las tarjetas que los socios habían llenado en el momento de la entrevista. J.L. tenía estas tarjetas en un archivador en su oficina. A continuación se encuentra una descripción del contenido de las tarjetas. Los valores reales de las tarjetas se hallan tabulados en las páginas siguientes.

Con el propósito de poder conocer mejor a los socios, J.L. quiere saber inicialmente el promedio semanal de gastos en alimentos, puesto que no dispone de demasiado tiempo, quiere hacer esto sin tener que mirar las quinientas tarjetas. Sin embargo, también desea asegurarse de que el promedio que calcule fuera exacto. Para esto contrata a un consultor y este le recomienda realizar una muestra probabilística sobre las 500 tarjetas que se encuentran tabuladas en la tabla. Las tarjetas tienen la siguiente información:

**Descripción de las variables**

**A** = número de identificación de la unidad familiar; 1 - 500

**B** = gastos reales semanales en alimentos

**C** = número de personas en la unidad familiar; 1 - 9

- 1 = una persona
- 2 = dos personas
- 3 = tres personas
- 4 = cuatro personas
- 5 = cinco personas
- 6 = seis personas
- 7 = siete personas
- 8 = ocho personas
- 9 = nueve personas

**D** = ingreso familiar anual real

**E** = educación del jefe del hogar; 1 - 5

- 1 = menos del 8 grado
- 2 = entre los grados 9 y 11
- 3 = con título en secundaria
- 4 = algunos años de educación secundaria
- 5 = con título universitario

**F** = edad actual del jefe del hogar

**G** = gasto semanal en alimentos, codificados en 7 categorías; 1 - 7

- 1 = menos de \$ 15
- 2 = \$ 15 a \$ 29.99
- 3 = \$ 30 a \$ 44.99
- 4 = \$ 45 a \$ 59.99
- 5 = \$ 60 a \$ 74.99
- 6 = \$ 75 a \$ 89.99
- 7 = \$90 ó más

**H** = algún niño menor de 6 años en el hogar; 1 - 2

- 1 = no
- 2 = sí

**I** = algún hijo entre 6 y 18 años en el hogar; 1 - 2

- 1 = no
- 2 = sí

**J** = ingreso anual familiar, codificado en seis categorías; 1 - 6

- 1 = menos de \$ 3000
- 2 = \$ 3000 - \$ 5999
- 3 = \$ 6000 - \$ 9999
- 4 = \$ 10000 - \$ 14999
- 5 = \$ 15000 - \$ 24999
- 6 = \$ 25000 ó más

**K** = edad del jefe del hogar, codificado en siete categorías; 1 - 7

- 1 = menos de 25
- 2 = 25 - 34
- 3 = 35 - 44
- 4 = 45 - 54
- 5 = 55 - 64
- 6 = 65 - 74
- 7 = 75 ó más

Se le pide:

1. Seleccione una *muestra aleatoria simple* de tamaño 50. Construya la tabla con las observaciones seleccionadas. Calcule los estadísticos que permitan describir la información. Realice una estimación por intervalo de la media con un nivel de confianza del 95%.
2. Seleccione una *muestra sistemática* de tamaño 50. Construya la tabla con las observaciones seleccionadas. Calcule los estadísticos que permitan describir la información. Realice una estimación por intervalo de la media con un nivel de confianza del 95%.
3. Seleccione una *muestra estratificada proporcional* según la presencia de niños menores de 6 años en el hogar de tamaño 25. Construya la tabla con las observaciones seleccionadas. Calcule los estadísticos que permitan describir la información. Realice una estimación por intervalo de la media con un nivel de confianza del 95%.
4. Seleccione una *muestra estratificada no proporcional* según la presencia de niños menores de 6 años en el hogar donde cada estrato sea de tamaño 25. Construya la tabla con las observaciones seleccionadas. Calcule los estadísticos que permitan describir la información. Realice una estimación por intervalo de la media con un nivel de confianza del 95%.

A	B	C	D	E	F	G	H	I	J	K
1	12	1	2500	1	56	1	1	1	1	5
2	16.5	1	2800	1	70	2	1	1	1	6
3	18	1	2000	1	20	2	1	1	1	1
4	17	1	4500	1	60	2	1	1	2	5
5	46.5	1	8000	1	40	4	1	1	3	3
6	45	1	7000	1	51	4	1	1	3	4
7	15	1	3500	1	76	2	1	1	2	7
8	60	2	2800	1	20	5	1	1	1	1
9	15	2	2500	1	51	2	1	1	1	4
10	18	2	4000	1	32	2	1	1	2	2
11	22.5	2	5000	1	47	2	1	1	2	4
12	20	2	8000	1	35	2	1	1	3	3
13	97	2	5500	1	58	7	1	1	2	5
14	57	2	6000	1	27	4	1	1	3	2
15	39	2	3000	1	38	3	1	1	2	3
16	30	2	4000	1	40	3	1	1	2	3
17	42	2	3000	1	19	3	1	1	2	1
.	.	.	.	.	.	.	.	.	.	.
.	.	.	.	.	.	.	.	.	.	.
.	.	.	.	.	.	.	.	.	.	.
.	.	.	.	.	.	.	.	.	.	.
.	.	.	.	.	.	.	.	.	.	.
497	115	6	24000	5	36	7	2	2	5	3
498	75	7	28000	5	37	6	2	2	6	3
499	105	7	20000	5	39	7	2	2	5	3
500	75	8	33000	5	42	6	2	2	6	3

Tabla extraída de Kinnear, T. Taylor, J. *Investigación de Mercados. Un enfoque aplicado*. Mc.Graw Hill. 1993. Puedes bajar de la web el archivo Caso 51.xls

## Problemas

Problema 1. En una compañía, se quiere determinar los gastos médicos familiares anuales promedio de los empleados. La gerencia de la compañía quiere tener 95% de confianza de que el promedio de la muestra está correcto con aproximación  $\pm \$50$  de los gastos familiares reales. Un estudio piloto indica que la desviación estándar se puede estimar en \$1400. ¿Qué tan grande es el tamaño necesario de la muestra?

Problema 2. Se desea estimar la suma promedio de ventas con aproximación de  $\pm \$100$  con 99% de confianza y se supone que la desviación estándar es de \$200, ¿qué tamaño de muestra se necesita?

Problema 3. Un grupo de estudio quiere estimar la facturación mensual promedio por luz eléctrica en el mes de julio en una ciudad grande. En base en estudios efectuados en otras ciudades se supone que la desviación estándar es de \$20. El grupo querría estimar la facturación promedio de julio con aproximación  $\pm \$5$  del promedio real con 99% de confianza ¿Qué tamaño de muestra se necesita?

---

---

## **Bibliografía**

- Berenson, Mark y Levine, David. (1996) *Estadística Básica en Administración*. Prentice Hall. México.
- Daniel, Wayne W. (1999). *Bioestadística, base para el análisis de las ciencias de la salud*. Tercera Edición. Editorial Limussa. México.
- Dixon W.J. y Massey F.J. (1957) *Introduction to Statistical Analysis*. Nueva York, McGraw-Hill.
- Hernández Sampieri, Roberto; Fernández Collado, Carlos y Baptista Lucio, Pilar. (2000) *Metodología de la Investigación*. Segunda Edición. McGraw Hill. México.
- Kazmier, Leonard y Diaz Mata, Alfredo. (1993) *Estadística Aplicada a la Administración y a la Economía*. Mc.Graw Hill. México.
- Kinnear, T. Taylor, J. (1993). *Investigación de Mercados. Un enfoque aplicado*. Mc.Graw Hill.
- Padua, Jorge. (1996). *Técnicas de Investigación Aplicadas a las Ciencias Sociales*. Fondo de Cultura Económica. México.

Capítulo 6. ORGANIZACIÓN DE LOS DATOS .....	185
6.1 Diseño del trabajo de campo.....	185
Elección del punto muestral.....	186
6.2 Recolección de información.....	191
6.3 Codificación de información. ....	193
6.4 Diseño de base de datos.....	194
6.5 Procesamiento de la información. ....	194
Organización y homogeneización de los datos.....	195
6.6 Marco legal de las estadísticas oficiales .....	198
Ley Nacional Nº 17.622 .....	198
Ley Provincial Nº 5454.....	201
CASOS DE ESTUDIO, PREGUNTAS Y PROBLEMAS .....	203
Caso 6.1: Diario regional .....	203
Caso 6.2: Razas de perros .....	203
Bibliografía .....	204
Anexo A: Diario Regional.....	205
Anexo B: Razas de perros .....	245





## Capítulo 6. ORGANIZACIÓN DE LOS DATOS

### 6.1 Diseño del trabajo de campo

La recolección de información de fuentes primarias requiere el diseño de una estrategia que permita llegar al elemento integrante de la muestra, a efectos de realizarle la entrevista que permita completar la tabla de datos.

Hay situaciones en las que el diseño se realiza en su totalidad en laboratorio y otras en las que es un trabajo compartido entre el laboratorio y el terreno. La primera situación se observa cuando la selección del elemento muestral se realiza a través de un padrón o listado exhaustivo de la población; por ejemplo, los alumnos integrantes de la muestra para responder la Encuesta Alumnos en la Facultad. En esta encuesta la población la constituyen, dentro de cada materia, todos los alumnos que la regularizaron y promocionaron. La metodología consiste en extraer una muestra por materia e independiente del resto; solo debe responder quien ha sido seleccionado y su eventual reemplazo lo realiza el investigador (en laboratorio) que lo seleccionó.

Hay situaciones en las que no se dispone un listado de individuos de la población, lo que da lugar a diseñar un muestreo en etapas. En esta metodología están contemplados muchos de los relevamientos a hogares o empresas. En laboratorio se decide el ámbito geográfico donde se realizará el relevamiento, el cual puede estratificarse, luego se decide aleatoriamente las manzanas a relevar, cómo será el recorrido y quién tendrá la responsabilidad del relevamiento en ese sector. El encuestador deberá completar la selección del elemento muestral, siguiendo las indicaciones impartidas por el laboratorio, las que a su vez tienen estrecha relación con el objetivo de la investigación. Por ejemplo, un relevamiento en la ciudad para conocer la intención de voto en las próximas elecciones contempla la selección aleatoria de la persona de 18 años o más que debe responder en el hogar; mientras que, una encuesta de opinión sobre la calidad de atención en los supermercados de la ciudad va a requerir que responda la persona que vive en ese hogar y habitualmente concurre al supermercado.

El relevamiento censal también requiere de la participación activa del censista en la selección de las unidades censales, con las instrucciones recibidas es quien decide si corresponde censar. Por ejemplo, en un censo de población deberá decidir si corresponde censar a aquellas personas que han dormido en ese domicilio; en un censo de actividad industrial deberá decidir si está frente a una industria.

En el siguiente punto se describe una metodología para seleccionar el punto muestral en la que el encuestador tienen participación en la selección aleatoria del elemento muestra.

### Elección del punto muestral

El objetivo es seleccionar  $n$  hogares en una ciudad, de un total de  $N$ , para que la muestra sea representativa del total de hogares de la población.

En primer lugar se determina la fracción de muestreo. Esta es la proporción de unidades totales, en cualquier segmento dado del universo, que va a incluirse en la obtención de los datos. Si el estudio consiste en encuestar a personas de una vivienda, entonces la fracción de muestreo será aquella que permita visitar tantas viviendas como personas se deben encuestar. Por lo tanto  $n$ , tamaño de la muestra, debe coincidir con  $v$ , cantidad de viviendas a visitar. De modo que se deben seleccionar  $v$  viviendas de un total de  $V$  en el estrato.

La *fracción de muestreo básica* es

$$\frac{1}{k} = \frac{v}{V}$$

esto significa que se deben encuestar 1 cada  $k$  viviendas. El valor de  $k$  se obtiene tomando la inversa de la fracción de muestreo:

$$k = \frac{V}{v}$$

El paso siguiente es convertir esta fracción de muestreo básica, en fracciones de muestreo prácticas, las cuales permitan elegir sólo algunas manzanas del estrato y unidades de vivienda dentro de esas manzanas.

La fracción de muestreo básica es igual al producto de la fracción de la manzana ( $1/m$ ) y la fracción dentro de la manzana ( $1/V_m$ ), es decir

$$\frac{1}{k} = \frac{1}{m} * \frac{1}{V_m}$$

De modo que, decidiendo un término del segundo miembro, se resuelve el problema.

Para determinar la fracción de muestreo dentro de la manzana ( $1/V_m$ ), hay que tener en cuenta la cantidad de viviendas en promedio existentes por manzana,  $\bar{v}_m$ , y la cantidad de encuestas recomendables por manzana. Esto último va a depender de que tan probable sea encontrar respuestas en los hogares a las preguntas del cuestionario, una ayuda para esto es tener en cuenta la siguiente relación:

Si la entrevista puede hacerse:		encuestas a realizar		frecuencia de recorrido
-en una alta proporción de hogares	⇒	de 4 a 6	⇒	Mayor
-en una pequeña proporción de hogares	⇒	de 10 a 12	⇒	Menor

Entonces se tendrá que la fracción de muestreo dentro de la manzana será

$$\frac{1}{V_m} = \frac{n_m}{\bar{v}_m}$$

donde  $n_m$  es el número de encuestas recomendables y  $\bar{v}_m$  es el número medio de viviendas por manzana. Se realizarán 1 cada  $V_m$  viviendas dentro de la manzana seleccionada.

Determinado  $(1/V_m)$  y sabiendo  $(1/k)$  se puede calcular la fracción de muestreo de la manzana  $(1/m)$  de la siguiente manera:

$$\frac{1}{m} * \frac{1}{V_m} = \frac{1}{k} \Rightarrow \frac{1}{m V_m} = \frac{1}{k}$$

De esta última expresión se deduce que

$$m V_m = k \Rightarrow m = \frac{k}{V_m}$$

Por lo tanto, se debe seleccionar 1 cada  $m$  manzanas y 1 cada  $V_m$  viviendas dentro de la manzana.

**Ejemplo. Elección de punto muestra en la ciudad de Río Cuarto.**

Se requiere replicar la investigación acerca del Perfil del Habitante de HHH en la ciudad de Río Cuarto. Para esto es necesario la realización de 600 encuestas a hogares en un muestreo estratificado proporcional en la ciudad.

La localización física del hogar es la vivienda, de modo que se debe encontrar la manera de relevar 600 viviendas, en una población que tiene 60484 viviendas, a efectos de tener información de 600 hogares.

Se decide dividir a la ciudad en estratos geográficos:

Alberdi  
Banda Norte  
Centro

De acuerdo al peso que cada estrato tiene, en términos de viviendas en el total de la ciudad, se asignan la cantidad de encuestas a realizar en cada uno. De esta manera los 600 hogares se reúnen al visitar 106 viviendas en Alberdi, 125 viviendas en Banda Norte y 369 viviendas en el sector Centro.

Según la nomenclatura catastral, la ciudad se divide en Circunscripción y Sectores. Cada uno de ellos tiene un número definido de manzanas a las cuales se asocian viviendas y, por ende, guardan una relación de población.

Estrato	Frecuencia Absoluta		Frecuencia Relativa	
	Manzanas	Viviendas	Manzanas	Viviendas
A.Alberdi	304	10709	18.06	17.71
B.Banda Norte	403	12604	23.95	20.84
C.Centro	976	37171	57.99	61.46
Total	1683	60484	100.00	100.00

Alberdi		Frecuencia Absoluta		Frecuencia Relativa	
Circ.	Sector	Manzanas	Viviendas	Manzanas	Viviendas
4	1	218	7786	71.71	72.71
4	2	86	2923	28.29	27.29
Total		304	10709	100.00	100.00

Banda Norte		Frecuencia Absoluta		Frecuencia Relativa	
Circ.	Sector	Manzanas	Viviendas	Manzanas	Viviendas
6	1	177	4615	43.92	39.51
6	2	177	4930	43.92	42.21
6	3	16	802	3.97	6.87
7	3	33	1334	8.19	11.42
Total		403	11681	100.00	100.00

C.Centro		Frecuencia Absoluta		Frecuencia Relativa	
Circ.	Sector	Manzanas	Viviendas	Manzanas	Viviendas
1	1	100	5787	10.25	15.57
1	2	115	7951	11.78	21.39
2	1	128	4066	13.11	10.94
2	2	92	3361	9.43	9.04
3	1	148	5117	15.16	13.77
3	2	103	3014	10.55	8.11
5	1	92	1835	9.43	4.94
5	2	154	4168	15.78	11.21
7	1	24	1163	2.46	3.13
7	2	20	709	2.05	1.91
Total		976	37171	100.00	100.00

es decir, deben encuestarse una cada 101 viviendas

Se ejemplificará la metodología con el estrato Centro. En primer término, debe definirse la fracción de muestreo como la proporción de unidades totales de la población que van a incluirse en la muestra. El estudio consiste en encuestar a personas, pero las personas viven en viviendas, entonces la fracción de muestreo básica será la que permita visitar tantas viviendas como personas deben encuestarse dentro de cada estrato.

$$\frac{1}{k} = \frac{v}{V} = \frac{369}{37171} = \frac{1}{100.73} \cong \frac{1}{101}$$

Esto es útil hasta cierto punto, se necesita alguna expresión más práctica en términos de números de manzanas y número de viviendas por manzanas, a efectos de disminuir costos de relevamiento. Cuanto más distancia debe recorrer el encuestador para hacer una encuesta adicional, mayor es el tiempo empleado y, paralelamente, mayor el costo asociado. Por esto, se debe hallar alguna relación que permita, sin perder aleatoriedad, concentrar en, relativamente, pocas manzanas el mayor número de encuestas.

Esto se logra teniendo en cuenta que, la fracción de muestreo básica es igual a la fracción de la manzana por la fracción dentro de la manzana.

$$\frac{1}{101} = \frac{1}{m} * \frac{1}{V_m}$$

Debe decidirse ahora  $1/V_m$ , es decir cada cuántas viviendas en la manzana se debe encuestar. Se conoce que existen 37171 viviendas en 976 manzanas, por lo tanto, el número medio de viviendas es de aproximadamente 38. Se considera que para el trabajo de investigación que se está realizando pueden recolectarse 5 encuestas por manzana. Si hay 38 viviendas por manzana, en promedio, y van a encuestarse 5, se tendrá que

$$\frac{1}{V_m} = \frac{n_m}{\bar{v}_m} = \frac{5}{38} \cong \frac{1}{7} \Rightarrow \frac{1}{V_m} = \frac{1}{7}$$

Conociendo la fracción de muestreo básica ( $1/k$ ) y la fracción de muestreo dentro de la manzana ( $1/V_m$ ), solo resta saber cuántas manzanas se necesitan seleccionar para hacer el trabajo. Si

$$\frac{1}{m} * \frac{1}{V_m} = \frac{1}{k} \Rightarrow \frac{1}{m} * \frac{1}{7} = \frac{1}{101}$$

puede reexpresarse como

$$m * 7 = 101 \Rightarrow m = \frac{101}{7} \cong 14$$

Es decir, se deben seleccionar una cada 14 manzanas. y una cada 7 viviendas dentro de la manzana, para que la fracción de muestreo sea de 1 cada 101 viviendas. Esto permite relevar 369 viviendas.

¿Que quiere decir esto? Que recorriendo una cada 14 manzanas, y visitando 1 cada 7 viviendas de cada manzana, se tendrá la fracción de muestreo de 1 cada 101 viviendas que van a permitir relevar 369 opiniones en el estrato Centro. El problema a resolver siguiente es como se seleccionan las 70 manzanas.

En las tablas se observa que cada estrato se subdivide en Circunscripciones y Sectores, cada uno de ellos tiene un peso determinado en el total de manzanas del estrato. De acuerdo a ese peso se asigna la fracción de muestreo de manzanas que le corresponde. Es necesario numerar en el plano, dentro de cada circunscripción sector, las manzanas de a grupos de cuatro:

3	2	7	6	11	10
4	1	8	5	12	9

donde cada grupo representa un punto muestra. Se seleccionan las manzanas aplicando muestreo sistemático: utilizando la fracción de manzanas,  $m$ , se determina al azar el valor de arranque y a partir de allí se aplica la frecuencia hasta seleccionar las  $m$  manzanas.

### Ejemplo. Muestreo por Diseño de etapas múltiples

Un país de 22.200.000 habitantes está distribuido en 4000 municipios de la manera siguiente

Municipios de menos De 2000 habitantes	Municipios de 2000 a 10000 habitantes	Municipios de Mas de 10000 habitantes
3000 municipios 3.000.000 habitantes (media: 1.000 hab.)	700 municipios 4.200.000 habitantes (media: 6.000 hab.)	300 municipios 15.000.000 habitantes (media: 50.000 hab.)

Se decide constituir una muestra de 2.220 habitantes, con una tasa de muestreo de  $1/10.000$ , utilizando, en cada categoría de municipios, un muestreo bietápico (muestra de municipios luego muestra de habitantes en los municipios extraídos) de tal manera que cada habitante tenga la misma probabilidad igual a  $1/10.000$  de pertenecer a la muestra.

Se puede procesar según dos métodos

#### Primer Método

Se realiza un muestreo aleatorio estratificado proporcional de municipios con una tasa uniforme de 1/100, luego, en los municipios extraídos, un muestreo aleatorio simple de un habitante para cada 100 habitantes.

Se obtiene entonces para cada muestra de municipios:

$$m_1 = 30 \qquad m_2 = 7 \qquad m_3 = 3 \text{ municipios}$$

y, en promedio, para cada municipio de la muestra de cada categoría:

$$n_1 = 10 \qquad n_2 = 60 \qquad n_3 = 500 \text{ personas}$$

lo que da, como previsto, para cada categoría, un número medio de habitantes en la muestra de:

$$300, \quad 420 \text{ y } 1500 \text{ respectivamente}$$

Con este método, los tamaños de las muestras de habitantes son aleatorios y el manejo del trabajo de los encuestadores se vuelve difícil. Además el número de "puntos de encuesta", que corresponden aquí al número de municipios de la muestra, es débil: solamente 40.

#### Segundo Método

Se decide encuestar a 10 habitantes por municipio de la muestra, lo que, cuando se toma en cuenta la duración del cuestionario, corresponde a un día de trabajo de un encuestador.

Hay pues que constituir una muestra de 222 municipios. Se constituye esta muestra de municipios con probabilidades desiguales, proporcionales al tamaño del municipio.

Se obtiene entonces, para la muestra de municipios:

$$m_1 = 30 \qquad m_2 = 42 \qquad m_3 = 150$$

y, como tenemos que encuestar 10 habitantes por municipio de la muestra, al final tenemos, tal como lo previsto:

$$300, \quad 420 \text{ y } 1500 \text{ habitantes encuestados en cada categoría}$$

Con este método, se facilita la gestión del trabajo de los encuestadores y el número de puntos de encuestas crece (222).

Para un municipio, su probabilidad de pertenecer a la muestra de municipios es la misma dentro de una categoría pero distinta de una categoría a otra.

Tenemos  $\pi_i = 1/100$  para la primera categoría,  $\pi_i = 6/100$  para la segunda y  $\pi_i = 50/100$  para la tercera.

Se trata ahora de estimar  $\mu$  utilizando el estimador de estratificación.

Supongamos que los resultados sobre los tres estratos sean los siguientes.

	Primer estrato	Segundo estrato	Tercer estrato	Conjunto ( $N$ )
$N_k$	3.000.000	4.200.000	15.000.000	22.200.000
$N_k / N$	3/22,2	4,2/22,2	15/22,2	1
$\hat{\mu}_k$	624	700	750	744
$\hat{V}(\hat{\mu}_k)$	1700	1600	500	317

El resultado expuesto en Conjunto se obtiene de hacer:

$$\hat{\mu} = \sum \frac{N_k}{N} \hat{\mu}_k \quad \hat{V}(\hat{\mu}) = \sum \left( \frac{N_k}{N} \right)^2 \hat{V}(\hat{\mu}_k)$$

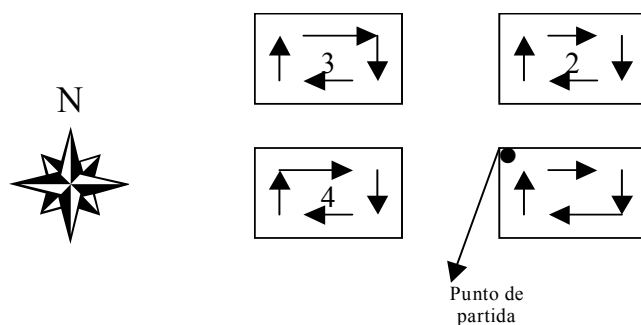
Luego, la estimación de  $\mu$  por intervalo de confianza al 95%:

$$\hat{\mu} \pm 1.96 \sqrt{\hat{V}(\hat{\mu})} \quad 744 \pm 35 \quad [709; 779]$$

Se realizó aquí un muestreo de 3 etapas. Como se trata de una estratificación a la primera etapa, al final tenemos, en cada estrato, un muestreo bietápico: una muestra de municipios, luego, en cada municipio de la muestra, una muestra de habitantes.

## 6.2 Recolección de información.

Las manzanas seleccionadas son origen de un punto muestra. El recorrido comienza en la esquina nor-oeste y se gira con la mano derecha sobre la pared, es decir, se gira la manzana en el sentido de las agujas del reloj. Debe encuestarse la vivienda ubicada en la esquina nor-oeste y luego sistemáticamente una cada dos. Si al terminar la manzana seleccionada no se ha completado la totalidad de encuestas, se debe continuar en la manzana 2 como indican las flechas en el esquema, y así sucesivamente; si al terminar la manzana 4 no se completaron la totalidad de formularios requeridos en el punto muestra, se deberá buscar un reemplazo.



Si la investigación indica que se deben seleccionar personas de manera aleatoria, el formulario tendrá un cuadro de selección del entrevistado que asegura la aleatoriedad de la elección. En el ejemplo se describe una metodología a seguir luego de describir, en el cuadro de selección del formulario, todas las personas que vivan en la casa, consignando el sexo y la edad.

**Ejemplo.** La encuesta se realiza a las personas mayores de 18 años. Una vez consignado el grupo familiar, deberá cruzarse la fila correspondiente a la última persona detallada y en condiciones de responder, con la columna correspondiente al último número del cuestionario. La intersección da por resultado el número de orden de la persona a encuestar. En la tabla, el hogar tiene 5 integrantes pero sólo 3 en condiciones de responder. La selección aleatoria indica que la mujer de 43 años debe responder la encuesta.

CUADRO DE SELECCIÓN		Encuesta N°15														
Número de Orden	Edad	SEXO		Ultimo dígito de número cuestionario										Selección		
		H	M	1	2	3	4	5	6	7	8	9	0	SI	NO	
1	45	X		1	1	1	1	1	1	1	1	1	1	1		
2	43		X	2	1	2	1	2	1	2	1	2	1	2	X	
3	20		X	3	1	3	2	2	1	2	1	2	3			
4	15	X		2	3	4	1	3	2	1	1	3	4			
5	13	X		5	4	3	4	2	1	2	1	3	5			
6				5	5	4	4	2	2	1	3	3	6			
7				1	6	5	7	2	6	4	1	3	7			
8				7	4	3	7	1	2	6	3	5	8			
9				1	2	3	9	5	6	7	8	4	9			
10				9	3	6	5	2	10	7	8	1	4			

En el recorrido de la manzana puede ocurrir que algunas viviendas rechacen la encuesta en forma total y, en otros casos, que la persona no se encuentre o no quiera contestar. Todos estos casos se recuperan al final del recorrido de la manzana. Suponga que deben hacerse 10 encuestas en la manzana, luego de encuestar a la persona correspondiente a la vivienda 10, se selecciona una nueva vivienda y se realiza la encuesta que había quedado pendiente anteriormente. Si el rechazo era de hogar, con lo cual no se había realizado selección de persona para encuestar, entonces deberá realizarse el cuadro de selección. Si el rechazo proviene de una persona que había sido seleccionada y no quiso responder, y no pudo ser reemplazada por otra de igual sexo y edad decreciente dentro del hogar, entonces no se hace cuadro de selección y se reemplaza por otra de igual sexo en la vivienda que corresponda.

Es necesario prestar atención a la cuota de sexo. Al final del relevamiento de la manzana deben existir respuestas de hombres y de mujeres de acuerdo a la cuota de sexo asignada.

En la Hoja de ruta deberá consignarse el domicilio seleccionado, la aceptación o rechazo a la encuesta, el sexo y la edad de la persona seleccionada y el sexo y la edad de la persona que responde. Este formulario es útil en el momento de supervisar un trabajo de encuestas.



### 6.3 Codificación de información.

La información relevada a través de encuestas puede requerir de algún tratamiento adicional antes de incorporarla en la base de datos. Esto se asocia directamente al tipo de pregunta: cerrada, abierta o semicerrada.

Las preguntas cerradas no requieren codificación adicional porque las respuestas se asocian a un código preestablecido.

**Ejemplo.** En la pregunta

11. Ud. mira televisión? (E)

1. SI
2. NO (IR A 16)
3. Ns/Nc

las respuestas se ingresan, en la base de datos, a través de los códigos asignados. Así, quienes contesten que sí adoptarán el código 1 y quienes contesten que no adoptarán el código 2.

Las preguntas abiertas y las semicerradas requieren de un tratamiento adicional denominado "codificación". Esto consiste en asignarles códigos a cada respuesta distinta que tenga esa pregunta. Para las preguntas abiertas, se procede de igual manera.

**Ejemplo.** En la pregunta

6. Cuál es el diario que **lee más**? (E)

1. Local
2. Ambito Financiero
3. La Voz del Interior
4. Puntal
5. Clarín
6. Otro. Cuál? \_\_\_\_\_
7. Ns/Nc

es necesario codificar la respuesta 6. Otro. Cuál? \_\_\_\_\_.

Entre las respuestas posibles se encuentran: La Nación, Ambito Financiero, Cronista Comercial, El Economista, El Independiente de La Rioja, La Gaceta de Tucumán y El Mundo de España.

La tabla muestra el proceso de codificación para la pregunta ¿Cuál? en las 10 primeras encuestas de un relevamiento.

Encuestas	Respuestas a la pregunta ¿Cuál?	Código asignado
1	El Economista	1
2		
3	La Nación	2
4	El Independiente LR	3
5	La Nación	2
6	Infobae	4
7		
8		
9	Infobae	4
10	La Nación	2

Cuando no se menciona un diario, no se asigna código

Cada vez que aparece el mismo diario, se le asigna igual código

#### 6.4 Diseño de base de datos.

El diseño de la base para generar la tabla de datos brutos puede generarse en una planilla de cálculo. En general, se diseña reservando una fila para cada encuesta y tantas columnas como variables tenga la investigación; de este modo, las filas contienen a los individuos y las columnas a las características medidas sobre ellos. La primera fila de la base tiene los rótulos de las variables.

En la tarea de diseño se deben identificar las características de la variable bajo análisis. Toda pregunta da lugar a una o más variables, de acuerdo a que sus respuestas sean mutuamente excluyentes o no.

**Ejemplo.** En la pregunta

22. Habitualmente para que tipo de compras utiliza la tarjeta de crédito (E)

1. Supermercado
2. Indumentaria
3. Libros/material de estudio/trabajo
4. Viajes
5. Combustible
6. Todo
7. Otro. Cuál? \_\_\_\_\_
8. Ns/Nc

las respuestas pueden ser múltiples; es decir, se puede responder por más de una opción. En la base cada respuesta debe considerarse como una variable en sí misma, lo que lugar a reservar tantas columnas (1 para cada variable) como respuestas tenga la pregunta.

En la pregunta:

27. Cómo calificaría a su tarjeta de crédito. Ud. diría que el servicio es.... (L)

- (Mostrar T3)
1. Muy Bueno
  2. Bueno
  3. Regular
  4. Malo
  5. Muy malo
  6. Ns/Nc

Es válida una sola respuesta porque son mutuamente excluyentes. En la base de datos se reserva sólo una columna para ingresar la opinión de cada encuestado.

#### 6.5 Procesamiento de la información.

La tarea de procesamiento de la información consiste en ingresar toda la información contenida en una encuesta en la base de datos previamente diseñada. Terminada esta tarea se cuenta con la tabla de datos brutos. Es necesario realizar controles que permitan asegurarnos que el ingreso de la información a la tabla estuvo bien realizado. La carga de la información se controla a través de la selección aleatoria de encuestas y de conocer el valor máximo y el valor mínimo de cada variable. Con lo primero se pretende controlar que el procesador no haya ingresado códigos en columnas que no corresponden y con lo segundo verificar la no existencia de duplicación de códigos o tipeo accidental de algún número.

Esta etapa también contempla la construcción de nuevas variables o indicadores que surgen de la combinación de otras variables. Por ejemplo, la variable Ingreso per cápita del hogar se construye dividiendo el ingreso del hogar por la cantidad de personas que integran ese hogar.

Con las variables cualitativas es necesario realizar un conteo de la frecuencia de cada respuesta a efectos de reunir bajo un mismo código respuestas similares.

### Organización y homogeneización de los datos.

Es habitual en Econometría, al trabajar con datos reales provenientes del contexto económico, tener variables medidas con diferente periodicidad. Por ejemplo, el PBI medido en forma trimestral y el Índice de Precios medido mensualmente.

Al incorporarlas en un modelo econométrico donde la unidad de observación es el tiempo, las variables deben referir a la misma unidad de tiempo. Por ejemplo, hay situaciones en las que es necesario contar con el dato del PBI medido en forma mensual. Para esto es necesario construir la serie mensual a partir de los datos trimestrales.

Un criterio que permite construir la serie mensual supone que las variaciones dentro del trimestre se producirán de manera constante. Los pasos a seguir consisten en obtener el coeficiente de variación trimestral, y a este convertirlo en su equivalente mensual. Luego se asigna el valor de un trimestre al último mes del trimestre, y se le aplica a este el coeficiente de actualización mensual para obtener el mes siguiente.

**Ejemplo.** La tabla muestra el PBI medido en millones de pesos corrientes entre el primer trimestre de 1998 y el cuarto trimestre de 1999.

PBI a precios de mercado en millones de pesos corrientes		
PBI		
1998	I	282764
	II	312129
	III	305475
	IV	295425
1999	I	270746
	II	288830
	III	285087
	IV	289429

FUENTE: Ministerio de Economía de la Nación. [www.mecon.gov.ar](http://www.mecon.gov.ar)

Para convertir la serie en datos mensuales se hará lo siguiente:

1. Se le asigna al último mes del trimestre el valor del trimestre

PBI a precios de mercado en millones de pesos corrientes		
		PBI
1998	Marzo	282764
	Junio	312129
	Setiembre	305475
	Diciembre	295425
1999	Marzo	270746
	Junio	288830
	Setiembre	285087
	Diciembre	289429

FUENTE: Ministerio de Economía de la Nación.  
www.mecon.gov.ar

2. Se calcula la variación entre dos trimestres

$$\frac{\text{junio}'98}{\text{marzo}'98} = \frac{312129}{282764} = 1.10384985$$

3. Se calcula la variación mensual a partir de la trimestral

$$(1.10384985)^{1/3} = 1.03348299$$

PBI a precios de mercado en millones de pesos corrientes				
		PBI	Variación trimestral	Variación mensual
1998	Marzo	282764		
	Junio	312129	1.10384985	1.03348299
	Setiembre	305475	0.97868189	0.99284286
	Diciembre	295425	0.96710042	0.98891096
1999	Marzo	270746	0.91646272	0.97134073
	Junio	288830	1.06679323	1.02178632
	Setiembre	285087	0.98704082	0.99566148
	Diciembre	289429	1.01523044	1.00505125

FUENTE: Elaboración propia en base a datos de PBI del Ministerio de Economía de la Nación. www.mecon.gov.ar

4. Se calculan los valores del PBI mensual

$$\text{abril}'98 = \text{marzo}'98 * 1.03348299 = 292232$$

$$\text{mayo}'98 = \text{abril}'98 * 1.03348299 = 302017$$

$$\text{junio}'98 = \text{mayo}'98 * 1.03348299 = 312129$$

PBI a precios de mercado en millones de pesos corrientes					
		PBI trimestral	Variación trimestral	Variación mensual	PBI mensual
1998	Marzo	282,764			282764
	Abril				292232
	Mayo				302017
	Junio	312,129	1.10384985	1.03348299	312129
	Julio				309895
	Agosto				307677
	Setiembre	305,475	0.97868189	0.99284286	305475
	Octubre				302088
	Noviembre				298738
	Diciembre	295,425	0.96710042	0.98891096	295425
	Enero				286958
	Febrero				278734
1999	Marzo	270,746	0.91646272	0.97134073	270746
	Abril				276645
	Mayo				282672
	Junio	288,830	1.06679323	1.02178632	288830
	Julio				287577
	Agosto				286329
	Setiembre	285,087	0.98704082	0.99566148	285087
	Octubre				286527
	Noviembre				287974
	Diciembre	289,429	1.01523044	1.00505125	289429

FUENTE: Elaboración propia en base a datos trimestrales suministrados por el Ministerio de Economía de la Nación. [www.mecon.gov.ar](http://www.mecon.gov.ar)

Para la conversión a dato trimestral de los índices de precios mensuales elaborados por el INDEC se puede calcular el promedio del índice de tres meses:

Índice de Precios al consumidor del Gran Buenos Aires. Nivel general serie base 1999=100						
Año	Mes	IPC mensual		Año	Trimestre	IPC trimestral
1998	1	101.6677		1998	I	101.861333
1998	2	102.0222				
1998	3	101.8941				
1998	4	101.9056			II	101.921333
1998	5	101.8320				
1998	6	102.0264				
1998	7	102.3453			III	102.3485
1998	8	102.3659				
1998	9	102.3343				
1998	10	101.9599			IV	101.7939
1998	11	101.7178				
1998	12	101.7040				

FUENTE: Elaboración propia en base a datos del INDEC. [www.indec.gov.ar](http://www.indec.gov.ar)

## 6.6 Marco legal de las estadísticas oficiales

### Ley Nacional Nº 17.622

Buenos Aires, 25 de enero de 1968

En uso de las atribuciones conferidas por el artículo 5º del Estatuto de la Revolución Argentina

EL PRESIDENTE DE LA NACION ARGENTINA  
SANCIONA Y PROMULGA CON FUERZA DE LEY

ARTICULO 1º.- Las actividades estadísticas oficiales y la realización de los censos que se efectúen en el territorio de la Nación, se regirán por las disposiciones de la presente ley.

ARTICULO 2º.- Créase el Instituto Nacional de Estadística y Censos, que dependerá de la Secretaría del Consejo Nacional de Desarrollo y estará a cargo de un Director que será designado por el Poder Ejecutivo.

ARTICULO 3º.- Son objetivos del Instituto Nacional de Estadística y Censos:

- a) Unificar la orientación y ejercer la dirección superior de todas las actividades estadísticas oficiales que se realicen en el territorio de la Nación;
- b) Estructurar, mediante la articulación y coordinación de los servicios estadísticos nacionales, provinciales y municipales, el Sistema Estadístico Nacional, y ponerlo en funcionamiento de acuerdo con el principio de centralización normativa y descentralización ejecutiva.

ARTICULO 4º.- El Sistema Estadístico Nacional estará integrado por:

- a) El Instituto Nacional de Estadística y Censos.
- b) Los organismos centrales de estadística, que son:
  - I) Los servicios estadísticos de los Ministerios y Secretarías de Estado
  - II) Los servicios estadísticos de los Comandos en Jefe de las Fuerzas Armadas
  - III) Los servicios estadísticos de organismos descentralizados de la Administración Nacional
  - IV) Los servicios estadísticos de las Empresas del Estado
- c) Los organismos periféricos de estadística, que son:
  - I) Los servicios estadísticos de los gobiernos provinciales
  - II) Los servicios estadísticos de los gobiernos municipales
  - III) Los servicios estadísticos de las reparticiones autárquicas y descentralizadas, provinciales y municipales
  - IV) Los servicios estadísticos de las empresas provinciales y municipales
  - V) Los servicios estadísticos de los entes interprovinciales

ARTICULO 5º.- Son funciones del Instituto Nacional de Estadística y Censos:

- a) Planificar, promover y coordinar las tareas de los organismos que integran el Sistema Estadístico Nacional;
- b) Confeccionar el programa anual de las estadísticas y censos nacionales, con su correspondiente presupuesto por programa, basándose especialmente en las necesidades de información formuladas por las Secretarías del Consejo Nacional de Desarrollo (CONADE) y del Consejo Nacional de Seguridad (CONASE), sin perjuicio de tener en cuenta los requerimientos que puedan plantear otras entidades públicas y privadas;
- c) Establecer las normas metodológicas y los programas de ejecución de las estadísticas que se incluyan en el programa anual;
- d) Distribuir, entre los organismos que integran el Sistema Estadístico Nacional las tareas detalladas en el programa anual de estadística y censos nacionales, así como los fondos necesarios para su ejecución, cuando correspondiere;
- e) Promover la creación de nuevos servicios estadísticos en el territorio nacional;
- f) Promover la adecuada difusión de toda la información estadística en los Ministerios, Comandos en Jefe, Secretarías de Estado, Gobierno provinciales y municipales, organizaciones públicas y privadas y población en general;

- g) Concretar investigaciones de carácter metodológico y estadístico, tendientes a elevar el nivel técnico y científico del Sistema Estadístico Nacional;
- h) Celebrar acuerdos o convenios de carácter estadístico, con entidades públicas y privadas y promoverlos con organismos extranjeros e internacionales;
- i) Realizar cursos de capacitación técnica estadística, con la colaboración de organismos internacionales, nacionales y privados, y otorgar becas para capacitar personal, con el objeto de perfeccionar el nivel técnico y científico del Sistema Estadístico Nacional;
- j) Enviar delegados a los congresos, conferencias y reuniones nacionales e internacionales, que tengan por objeto el tratamiento de cuestiones estadísticas;
- k) Organizar un centro de intercambio e interpretación de informaciones estadísticas nacionales e internacionales;
- l) Realizar conferencias, congresos y reuniones estadísticas nacionales;
- m) Elaborar las estadísticas que considere conveniente, sin afectar el principio de descentralización ejecutiva establecido en el inc. d);
- n) Toda otra función que contribuya al cumplimiento de los objetivos fijados en el artículo 3º de la presente ley.

ARTICULO 6º.- El presupuesto de recursos del Instituto Nacional de Estadística y Censos, estará integrado por:

- a) Los recursos que determinen la Ley General de Presupuesto de la Nación.
- b) Los ingresos provenientes de la venta de publicaciones, certificaciones, registros y trabajos para terceros.
- c) Las multas aplicadas por infracciones a la presente ley.
- d) Contribuciones, aportes, y subsidios de provincias, municipalidades, dependencias o reparticiones oficiales, organismos mixtos, privados e internacionales.
- e) Los legados y donaciones.

ARTICULO 7º.- El presupuesto de gastos del Instituto Nacional de Estadística y Censos preverá las sumas destinadas a:

- a) Las erogaciones necesarias para el cumplimiento del programa anual de estadísticas, investigaciones y censos nacionales;
- b) El pago de los servicios que, eventualmente, acuerden con las reparticiones periféricas del Sistema Estadístico Nacional y que no estuvieren previstos en los presupuestos propios de ellas;
- c) La ampliación o perfeccionamiento de los servicios de las reparticiones periféricas del Sistema Estadístico Nacional;
- d) El mejoramiento de los métodos de trabajo de los organismos integrantes del Sistema Estadístico Nacional;
- e) La organización de misiones científicas o técnicas, relacionadas con los programas estadísticos;
- f) La contratación de trabajos técnicos o científicos estadísticos especializados;
- g) El pago de becas de perfeccionamiento que forman parte de los programas de capacitación del Instituto;
- h) Todas las otras erogaciones que estén vinculadas con el funcionamiento del Instituto.

ARTICULO 8º.- Todas las reparticiones que integran el Sistema Estadístico Nacional elevarán anualmente al Instituto Nacional de Estadísticas y Censos, los presupuestos por programas de todas las tareas estadísticas a ejecutar, para su integración en el programa nacional, de acuerdo con las normas que establezca el Instituto. Las reparticiones centrales atenderán con sus asignaciones presupuestarias la realización de los programas estadísticos que formen parte del programa nacional, a cuyo efecto los respectivos Ministerios y Secretarías de Estado, Comandos en Jefe, organismos descentralizados y Empresas del Estado o mixtas deberán proveer los recursos pertinentes. Las reparticiones periféricas podrán solicitar al Instituto financiación complementaria para atender los gastos correspondientes a:

- a) La ejecución de estadísticas y censos nacionales;
- b) Programas de asistencia técnica que haya formulado el Instituto para esas reparticiones;
- c) Inversiones que el Instituto considere necesarias para elevar el nivel de eficiencia de esas reparticiones.

- d) La provisión de estos recursos tendrá lugar cuando, a juicio del Instituto, las asignaciones presupuestarias con que cuentan las reparticiones resultaren insuficientes.

ARTICULO 9º.- A los efectos de la realización de las tareas que integran el programa anual o los planes estadísticos que formule el Instituto Nacional de Estadística y Censos, las reparticiones centrales y periféricas dependerán normativamente de éste y utilizarán los métodos, definiciones, formularios, cartografía, clasificaciones, fórmulas y toda otra disposición o norma técnica que el Instituto establezca para la reunión, elaboración, análisis y publicación de las estadísticas y censos.

ARTICULO 10º.- Las informaciones que se suministren a los organismos que integran el Sistema Estadístico Nacional, en cumplimiento de la presente ley, serán estrictamente secretos y sólo se utilizarán con fines estadísticos. Los datos deberán ser suministrados y publicados, exclusivamente en compilaciones de conjunto, de modo que no pueda ser violado el secreto comercial o patrimonial, ni individualizarse las personas o entidades a quienes se refieran. Quedan exceptuados del secreto estadístico los siguientes datos de registro: nombre y apellido, o razón social, domicilio y rama de actividad.

ARTICULO 11º.- Todos los organismos y reparticiones nacionales, provinciales y municipales, las personas de existencia visible o ideal, públicas o privadas con asiento en el país, están obligadas a suministrar a los organismos que integran el Sistema Estadístico Nacional los datos e informaciones de interés estadístico que éstos le soliciten.

ARTICULO 12º.- Facultase al Instituto Nacional de Estadísticas y Censos para exigir, cuando lo considere necesario, la exhibición de libros y documentos de contabilidad de las personas o entidades que estén obligadas a suministrar informaciones de carácter estadístico a los efectos exclusivos de la verificación de dichas informaciones. Cuando los datos consignados en las declaraciones presentadas, no se encuentren registradas en libros de contabilidad, deberán exhibirse los documentos originales y los antecedentes que sirvieron de base a las informaciones suministradas.

ARTICULO 13º.- Todas las personas que por razón de sus cargos o funciones tomen conocimiento de datos estadísticos o censales, están obligados a guardar sobre ellos absoluta reserva.

ARTICULO 14º.- Las personas que deban realizar tareas estadísticas o censales, con carácter de carga pública, estarán obligadas a cumplir estas funciones. Si no lo hicieran, se harán pasibles de las penalidades preceptuadas en el artículo 239 del Código Penal, salvo que aquellas estuviesen comprendidas en las excepciones que establezca reglamentariamente el Poder Ejecutivo.

ARTICULO 15º.- Incurrirán en infracción y serán pasibles de multa de diez mil (10.000) a quinientos mil (500.000) pesos moneda nacional, conforme al procedimiento que se establezca en la reglamentación de la presente ley, quienes no suministren en término, falseen o produzcan con omisión maliciosa las informaciones necesarias para la estadística y los censos a cargo del Sistema Estadístico Nacional.

ARTICULO 16º.- Cuando se trate de entidades civiles o comerciales, con personalidad jurídica o sin ella, serán personalmente responsables de las infracciones a la presente Ley los directores, administradores, gerentes o miembros de la razón social que hayan intervenido en los actos considerados punibles. Para las multas se establece que existirá responsabilidad subsidiaria de las entidades sancionadas. En caso de reincidencia dentro del período de un (1) año, contando desde la fecha de la sanción impuesta conforme al artículo 13º serán pasibles de la pena establecida por el artículo 239 del Código Penal, sin perjuicio de la nueva multa que correspondiera.

ARTICULO 17º.- Los funcionarios o empleados que releven a terceros o utilicen en provecho propio cualquier información individual de carácter estadístico censal, de la cual tengan conocimiento por sus funciones, o que incurran dolosamente en tergiversación, omisión o adulteración de datos de los censos o estadísticas, serán pasibles de exoneración y sufrirán además las sanciones que correspondan conforme con lo previsto por el Código Penal (Libro II, Título V, Capítulo III).

#### DISPOSICIONES TRANSITORIAS

ARTICULO 18º.- Dentro de los noventa (90) días de la fecha de promulgación de la presente Ley, los organismos que integran el Sistema Estadístico Nacional deberán suministrar al Instituto Nacional de Estadísticas y Censos, las informaciones que éste les requiera sobre las tareas estadísticas que realizan, el personal y los equipos afectados a ello, así como los recursos presupuestarios que demanda su realización.

ARTICULO 19º.- Dentro de los ciento ochenta (180) días de promulgada la presente Ley, el Instituto Nacional de Estadística y Censos, propondrá al Poder Ejecutivo su propia estructura orgánico-funcional y la estructura completa del Sistema Estadístico Nacional, estableciendo las áreas de competencia de cada uno de los organismos que lo integran.

ARTICULO 20º.- La Dirección Nacional de Estadística y Censos, dependiente de la Secretaría de Estado de Hacienda, pasará a integrar el Instituto Nacional de Estadística y Censos con su presupuesto, personal, inmuebles, muebles, útiles y antecedentes.

ARTICULO 21º.- Derogase la Ley 14.046 y toda otra norma que se oponga a la presente Ley.



ARTICULO 22º.- Comuníquese, publíquese, dese a la Dirección Nacional del Registro Oficial y archívese.

Ley Nº 17.622

FDO.: ONGANIA

GUILLERMO A. BORDA

## Ley Provincial Nº 5454

VISTO: La autorización del Gobierno Nacional concedida por Decreto Nro. 717/71 Artículo 1ro., Punto 1.1.4. - y la Política Nacional Nro. 127, en ejercicio de las facultades legislativas que le confiere el artículo noveno del Estatuto de la Revolución Argentina, EL GOBERNADOR DE LA PROVINCIA DE CÓRDOBA, SANCIONA Y PROMULGA CON FUERZA DE

### LEY

**Art. 1ro.** - La Dirección Área Estadística de la Secretaría de Desarrollo es el organismo encargado de coordinar y satisfacer los requerimientos que en materia estadística le formule el Gobierno Provincial, el Instituto Nacional de Estadística y Censos a través de su Programa Anual y otras entidades.

**Art. 2do.** - A los efectos del cumplimiento de la labor estadística, créase el Servicio Estadístico Provincial, que estará integrado por:

- a. La Dirección Área Estadística de la Secretaría de Desarrollo, como organismo rector de toda actividad estadística en la Provincia y nexos con el Sistema Estadístico Nacional (SEN), creado por Ley Nacional Nro. 17622,
- b. Los organismos de estadística compuestos por:
  - I. Los servicios estadísticos de los Ministerios y Secretaría Ministerios.
  - II. Los servicios estadísticos de los organismos descentralizados y empresas del estado de la Administración Pública Provincial;
  - III. Los servicios estadísticos de los gobiernos municipales y entes descentralizados dependientes.

A los fines de este artículo, todas las oficinas, secciones o departamentos de estadística que funcionan actualmente o que se creen en el futuro, en las distintas reparticiones de la Provincia o sus entes, actuarán como delegaciones de la Dirección Área Estadística, ajustando su acción a las normas y directivas de carácter técnico que imparta dicha Dirección.

**Art. 3ro.** - Son funciones de la Dirección Área Estadística:

- a. Planificar, promover y coordinar las tareas de los organismos que integran el Servicio Estadístico Provincial;
- b. Confeccionar el programa anual de estadísticas y censos, compatibilizando los requerimientos provenientes del Gobierno de la Provincia y del Instituto Nacional de Estadística y Censos; sin perjuicio de tener en cuenta otros requerimientos que puedan plantear entidades públicas y privadas;
- c. Distribuir entre los organismos que integran el Servicio Estadístico Provincial, las tareas detalladas en el Programa anual de estadística y censos,
- d. Promover la adecuada difusión de toda información estadística entre los distintos sectores de la comunidad;
- e. Concretar investigaciones de carácter metodológico y estadístico, tendientes a elevar el nivel técnico y científico del sistema estadístico provincial;
- f. Celebrar acuerdos o convenios de carácter estadístico con entidades públicas o privadas, promoverlos con organismos extranjeros e internacionales;
- g. Realizar cursos de capacitación en técnicas estadísticas, con la colaboración de organismos internacionales, nacionales y privados y gestionar el otorgamiento de becas para capacitar en el país o exterior a su personal, con el objeto de perfeccionar el nivel técnico y científico del servicio estadístico provincial;
- h. Enviar delegados a los congresos, conferencias y reuniones nacionales e internacionales, que tengan por objeto el tratamiento de cuestiones estadísticas;
- i. Toda otra función que contribuya al cumplimiento de los objetivos fijados en el artículo 1ro. de la presente ley.

**Art. 4to.** - Todas las reparticiones e instituciones dependientes de los poderes públicos de la provincia y municipalidades y las personas de existencia visible o ideal, públicas o privadas están obligadas a proporcionar a la Dirección Área Estadística u a los organismos integrantes del Servicio estadístico provincial, todo dato informe de interés estadístico que le soliciten y dentro de los plazos fijados a tales fines.

**Art. 5to.** - Las informaciones que en virtud del artículo precedente se suministren en cumplimiento de la presente ley, serán estrictamente secretas y sólo se utilizarán con fines estadísticos.

Los datos así receptados, deberán ser proporcionados y publicados, exclusivamente en compilaciones de conjunto, de modo que no pueda ser violado el secreto comercial, patrimonial o profesional, ni individualizarse las personas o entidades a quienes se refieran. Quedan exceptuados del secreto estadístico los siguientes datos de registro: nombre y apellido y razón social, domicilio y rama de actividad.

**Art. 6to.** - Facúltase a la Dirección Área Estadística de la Secretaría de Desarrollo, para exigir cuando lo considere necesario, la exhibición de libros y documentos de contabilidad de las personas o entidades informantes, a los efectos exclusivos de la verificación de los datos suministrados.

Cuando estos datos consignados en las declaraciones presentadas, no se encuentren registrados en libros de contabilidad, deberán exhibirse los documentos originales y los antecedentes que sirvieron de base a las informaciones suministradas.

**Art. 7mo.** - Todas las personas que por razón de sus cargos o funciones, tomen conocimiento de datos estadísticos o censales están obligadas a guardar sobre ellos, absoluta reserva.

**Art. 8vo.** - Las personas que deban realizar tareas estadísticas o censales, con carácter de carga pública, estarán obligadas a cumplir estas funciones. Si no lo hicieran, se harán pasibles de las penalidades preceptuadas en el artículo 239 del Código Penal, salvo los casos de excepción que se establezcan reglamentariamente.

**Art. 9no.** - Incurrirán en infracción y serán posibles de multas de Cien (100) a Cinco Mil (5.000) Pesos, conforme al procedimiento establecido por la Ley Nacional No 17622 y su Decreto Reglamentario Nro. 3110; quienes no suministren en término, falseen o produzcan con omisión maliciosa las informaciones necesarias para las estadísticas y los censos a cargo del Servicio Estadístico Provincial y/o Nacional.

**Art. 10mo.** - A los fines previstos en el artículo anterior, actuará como jefe del servicio estadístico provincial, el Director del Arca Estadística de la Secretaría de Desarrollo y el pago de las multas aplicadas serán satisfechas mediante depósito en la Casa Central, sucursales o agencias del Banco de la Provincia de Córdoba, en la cuenta titulada: "Rentas Generales". Tratándose de sanciones aplicadas a funcionarios o empleados del sector público el importe de las multas les será descontado directamente de sus haberes, debiendo ingresar en la forma establecida en el párrafo anterior.

**Art. 11ro.** - Cuando se trate de entidades civiles o comerciales, con personería jurídica o sin ella, serán personalmente responsables de las infracciones a la presente ley,, los directores, administradores, gerentes o miembros de la razón social que hayan intervenido en los actos considerados punibles. Para las multas se establece que existirá responsabilidad subsidiaria de las entidades sancionadas.

**Art. 12do.** - En caso de reincidencia dentro del período de un año contado desde la fecha de la sanción impuesta, los infractores serán pasibles de la pena establecida por el artículo 239 del Código Penal, sin perjuicio de la nueva multa que correspondiera.

**Art. 13ro.** - En ningún caso el pago de la multa eximirá al infractor de la obligación de suministrar los datos solicitados.

**Art. 14to.** - Los funcionarios o empleados que revelen a terceros o utilicen en provecho propio cualquier información individual de carácter estadístico o censal, de la cual tengan conocimiento por sus funciones o que incurran dolosamente en tergiversación, omisión o adulteración de datos de los censos o estadísticas, serán pasibles de exoneración y sufrirán además las sanciones que correspondan conforme con lo previsto por el Código Penal (Libro II - Título V -Capítulo III)

**Art. 15to.** - El presupuesto de recursos de la Dirección Arca Estadística de la Secretaría de Desarrollo, estará integrado por:

- a. Los recursos que determine la Ley de Presupuesto de la Provincia;
- b. Las contribuciones provenientes de reparticiones oficiales, organismos mixtos, privados o internacionales.

**Art. 16to.** - El Director y Subdirector del Arca Estadística deberán ser argentinos nativos, poseer título universitario habilitante y antecedentes de una probada preparación en la materia.

**Art. 17mo.** - El personal técnico de los departamentos, divisiones o secciones en que se organice la Dirección Arca Estadística, deberá poseer título habilitante para el eficaz desempeño de las funciones que se le confíen y será designado de acuerdo con los resultados de un concurso de títulos y/o antecedentes en las condiciones establecidas en el estatuto del Empleado Público (Ley 5172)

**Art. 18vo.** - Derógase la Ley, 4355 y toda otra disposición que se oponga a la presente ley.

**Art. 19no.** - Téngase por Ley de la Provincia, cúmplase, comuníquese, publíquese, dése al Registro y Boletín Oficial y archívese.

GUOZDEN - LÓPEZ CARUSILLO  
SUAREZ BENGOCHEA - RAFFO

## **CASOS DE ESTUDIO, PREGUNTAS Y PROBLEMAS**

### Caso 6.1: Diario regional

Un diario regional de alta circulación en el Norte del país estaba interesado en conocer el perfil del habitante de la región de influencia del diario y el perfil de sus lectores. A tales efectos realizó una encuesta de opinión que alcanzó a 505 hogares de los cuales 322 respondieron que leían el diario. De este grupo de 322 hogares, se han seleccionado aleatoriamente 20 encuestas de las cuales se han extraído algunas preguntas. En el Anexo A se encuentran los formularios.

Se solicita que:

1. Identifique las unidades de observación
2. Identifique las variables del problema y sus modalidades cuando corresponda
3. Codifique las preguntas del formulario
4. Construya en Excel una tabla de datos brutos
5. Elija dos variables y construya una tabla de contingencia
6. Calcule las estadísticas descriptivas para cada variable.

### Caso 6.2: Razas de perros

Una importante veterinaria del sur de Córdoba disponía de información acerca de las características de 27 razas de perros. La intención del grupo empresario es conocer si existen razas que presenten características semejantes y si hay relación entre pertenecer a un grupo de razas y cumplir una determinada función. En el Anexo B hay información de diferentes razas de perros para las cuales se solicita:

1. Identifique las unidades de observación
2. Identifique las variables del problema y sus modalidades cuando corresponda
3. Codifique las preguntas del formulario
4. Construya en Excel una tabla de datos brutos
5. Elija dos variables y construya una tabla de contingencia
6. Calcule las estadísticas descriptivas para cada variable.

---

---

### **Bibliografía**

- Berenson, Mark y Levine, David. (1996). *Estadística Básica en Administración*. Prentice Hall. México.
- Daniel, Wayne W. (1999). *Bioestadística, base para el análisis de las ciencias de la salud*. Tercera Edición. Editorial Limussa. México.
- Hernández Sampieri, Roberto; Fernández Collado, Carlos y Baptista Lucio, Pilar. (2000). *Metodología de la Investigación*. Segunda Edición. McGraw Hill. México.
- Kazmier, Leonard y Diaz Mata, Alfredo. (1993). *Estadística Aplicada a la Administración y a la Economía*. Mc.Graw Hill. México.
- Meyer, P.L. (1973). *Probabilidad y Aplicaciones Estadísticas*. Fondo Educativo Interamericano. México.
- Padua, Jorge. (1996). *Técnicas de Investigación Aplicadas a las Ciencias Sociales*. Fondo de Cultura Económica. México.
- Tramutola, C.D. *Modelos Probabilísticos y Decisiones Financieras*. E.C.Moderan. Lectura de Administración de Empresas. Capital Federal.

**Anexo A: Diario Regional**

Extracto de la encuesta realizada por  
un diario regional de alta circulación

ENCUESTA DE OPINIÓN  
**QUIEN LEE QUÉ?**

Domicilio: \_\_\_\_\_ N° de encuesta: 1

1.Sexo:  H  M

2.Edad: 46

49.Es Ud. jefe de familia?  SI  NO

15. De las secciones que tiene el diario cuál lee, cuál no lee y cuál no conoce? (L)

Tipo de noticia	LEE	NO LEE	Ns/Nc
Noticias sobre la realidad política, económica y general del ámbito internacional		<b>X</b>	
Clasificados		<b>X</b>	
Noticias sobre sexo, violencia, droga,		<b>X</b>	
Noticias policiales		<b>X</b>	
Noticias nacionales económicas		<b>X</b>	
Noticias nacionales políticas		<b>X</b>	
Información sobre la corrupción política y social		<b>X</b>	
Deportes	<b>X</b>		
Comentarios Científicos	<b>X</b>		
Información Gremial	<b>X</b>		
Notas Editoriales	<b>X</b>		
Comentario firmados por analistas económicos y políticos		<b>X</b>	
Comentarios firmados por analistas locales sobre aspectos de la realidad regional	<b>X</b>		
Noticias y comentarios Agropecuarios		<b>X</b>	
Noticias regionales		<b>X</b>	
Investigaciones periodísticas		<b>X</b>	
Información sobre Espectáculos		<b>X</b>	
Sociales		<b>X</b>	
Chistes, Horóscopos y entretenimiento en general	<b>X</b>		
Suplementos y revista dominical	<b>X</b>		

50. Cuál es su ocupación principal?.

1	Desocupado
2	Estudiante
3	Ama de Casa
4	Jubilado/pensionado
<input checked="" type="radio"/> 5	Empleadores
6	Trabajos no especializados o changarín
7	Comerciante sin personal
8	Técnico
9	Profesionales independientes y otros autónomos
10	Relación de dependencia
11	Obrero, cal. y no cal., capataz
12	Empleado s/jerarquía
13	Jefe intermedio
14	Ns/Nc

9	Universitario completo
10	Ns/Nc

52. Cuáles son los ingresos mensuales de su grupo familiar?

1	Menos de 300
<input checked="" type="radio"/> 2	De 300 a 500
3	De 501 a 700
4	De 701 a 1000
5	De 1001 a 1500
6	De 1501 a 2500
7	De 2501 a 4000
8	Más de 4000
9	Ns/Nc

54. Ud tiene...?

	SI	NO
Computadora		<b>X</b>
Televisor	<b>X</b>	
Horno a microondas		<b>X</b>
Radio	<b>X</b>	
Auto		<b>X</b>
Video		<b>X</b>
Vivienda		

51. Cuál es el máximo nivel de estudios cursados por Ud.?

1	Sin instrucción
2	Primario incompleto
<input checked="" type="radio"/> 3	Primario completo
4	Secundario incompleto
5	Secundario completo
6	Terciario incompleto
7	Terciario completo
8	Universitario incompleto

53. Cuántas personas en el hogar comparten este presupuesto?

1. Una persona
2. Dos personas
3. De 3 a 4 personas
4. De 5 a 6 personas
5. 7 o más personas
6. Ns/Nc

55. Antes de iniciar la próxima entrevista el encuestador deberá observar el nivel de vivienda

1	Media Superior
2	Media media
3	Media baja
<input checked="" type="radio"/> 4	Muy modesta o Marginal

Extracto de la encuesta realizada por  
un diario regional de alta circulación

ENCUESTA DE OPINIÓN  
QUIEN LEE QUÉ?

Domicilio: \_\_\_\_\_ N° de encuesta: 5

1. Sexo:  H  M

2. Edad: 39

49. Es Ud. jefe de familia?  SI  NO

15. De las secciones que tiene el diario cuál lee, cuál no lee y cuál no conoce? (L)

Tipo de noticia	LEE	NO LEE	Ns/Nc
Noticias sobre la realidad política, económica y general del ámbito internacional		<input checked="" type="checkbox"/>	
Clasificados		<input checked="" type="checkbox"/>	
Noticias sobre sexo, violencia, droga,		<input checked="" type="checkbox"/>	
Noticias policiales		<input checked="" type="checkbox"/>	
Noticias nacionales económicas		<input checked="" type="checkbox"/>	
Noticias nacionales políticas		<input checked="" type="checkbox"/>	
Información sobre la corrupción política y social		<input checked="" type="checkbox"/>	
Deportes		<input checked="" type="checkbox"/>	
Comentarios Científicos		<input checked="" type="checkbox"/>	
Información Gremial		<input checked="" type="checkbox"/>	
Notas Editoriales		<input checked="" type="checkbox"/>	
Comentario firmados por analistas económicos y políticos		<input checked="" type="checkbox"/>	
Comentarios firmados por analistas locales sobre aspectos de la realidad regional		<input checked="" type="checkbox"/>	
Noticias y comentarios Agropecuarios	<input checked="" type="checkbox"/>		
Noticias regionales		<input checked="" type="checkbox"/>	
Investigaciones periodísticas	<input checked="" type="checkbox"/>		
Información sobre Espectáculos	<input checked="" type="checkbox"/>		
Sociales	<input checked="" type="checkbox"/>		
Chistes, Horóscopos y entretenimiento en general	<input checked="" type="checkbox"/>		
Suplementos y revista dominical	<input checked="" type="checkbox"/>		

50. Cuál es su ocupación principal?

1	Desocupado
2	Estudiante
3	Ama de Casa
4	Jubilado/pensionado
5	Empleadores
6	Trabajos no especializados o changarín
7	Comerciante sin personal
8	Técnico
9	Profesionales independientes y otros autónomos
10	Relación de dependencia
<input checked="" type="radio"/> 11	Obrero, cal. y no cal., capataz
12	Empleado s/jerarquía
13	Jefe intermedio
14	Ns/Nc

51. Cuál es el máximo nivel de estudios cursados por Ud.?

1	Sin instrucción
2	Primario incompleto
3	Primario completo
4	Secundario incompleto
<input checked="" type="radio"/> 5	Secundario completo
6	Terciario incompleto
7	Terciario completo
8	Universitario incompleto
9	Universitario completo
10	Ns/Nc

52. Cuáles son los ingresos mensuales de su grupo familiar?

1	Menos de 300
2	De 300 a 500
<input checked="" type="radio"/> 3	De 501 a 700
4	De 701 a 1000
5	De 1001 a 1500
6	De 1501 a 2500
7	De 2501 a 4000
8	Más de 4000
9	Ns/Nc

53. Cuántas personas en el hogar comparten este presupuesto?

1. Una persona
2. Dos personas
3. De 3 a 4 personas
4. De 5 a 6 personas
5. 7 o más personas
6. Ns/Nc

54. Ud tiene...?

	SI	NO
Computadora	<input checked="" type="checkbox"/>	
Televisor	<input checked="" type="checkbox"/>	
Horno a microondas		<input checked="" type="checkbox"/>
Radio	<input checked="" type="checkbox"/>	
Auto		<input checked="" type="checkbox"/>
Video		<input checked="" type="checkbox"/>
Vivienda	<input checked="" type="checkbox"/>	

55. Antes de iniciar la próxima entrevista el encuestador deberá observar el nivel de vivienda

1	Media Superior
2	Media media
3	Media baja
<input checked="" type="radio"/> 4	Muy modesta o Marginal

Extracto de la encuesta realizada por  
un diario regional de alta circulación

ENCUESTA DE OPINIÓN  
QUIEN LEE QUÉ?

Domicilio: \_\_\_\_\_ N° de encuesta: **26**

1. Sexo: H **(M)**

2. Edad: **44**

49. Es Ud. jefe de familia? SI **(NO)**

15. De las secciones que tiene el diario cuál lee, cuál no lee y cuál no conoce? (L)

Tipo de noticia	LEE	NO LEE	Ns/Nc
Noticias sobre la realidad política, económica y general del ámbito internacional	<b>X</b>		
Clasificados	<b>X</b>		
Noticias sobre sexo, violencia, droga,		<b>X</b>	
Noticias policiales		<b>X</b>	
Noticias nacionales económicas	<b>X</b>		
Noticias nacionales políticas	<b>X</b>		
Información sobre la corrupción política y social	<b>X</b>		
Deportes	<b>X</b>		
Comentarios Científicos	<b>X</b>		
Información Gremial	<b>X</b>		
Notas Editoriales	<b>X</b>		
Comentario firmados por analistas económicos y políticos	<b>X</b>		
Comentarios firmados por analistas locales sobre aspectos de la realidad regional		<b>X</b>	
Noticias y comentarios Agropecuarios	<b>X</b>		
Noticias regionales		<b>X</b>	
Investigaciones periódicas	<b>X</b>		
Información sobre Espectáculos	<b>X</b>		
Sociales	<b>X</b>		
Chistes, Horóscopos y entretenimiento en general	<b>X</b>		
Suplementos y revista dominical	<b>X</b>		

50. Cuál es su ocupación principal?.

1	Desocupado
<b>2</b>	Estudiante
<b>3</b>	Ama de Casa
4	Jubilado/pensionado
5	Empleadores
6	Trabajos no especializados o changarín
7	Comerciante sin personal
8	Técnico
9	Profesionales independientes y otros autónomos
10	Relación de dependencia
11	Obrero, cal. y no cal., capataz
12	Empleado s/jerarquía
13	Jefe intermedio
14	Ns/Nc

51. Cuál es el máximo nivel de estudios cursados por Ud.?

1	Sin instrucción
2	Primario incompleto
<b>3</b>	Primario completo
4	Secundario incompleto
5	Secundario completo
6	Terciario incompleto
7	Terciario completo
8	Universitario incompleto
9	Universitario completo
10	Ns/Nc

52. Cuáles son los ingresos mensuales de su grupo familiar?

1	Menos de 300
<b>2</b>	De 300 a 500
3	De 501 a 700
4	De 701 a 1000
5	De 1001 a 1500
6	De 1501 a 2500
7	De 2501 a 4000
8	Más de 4000
9	Ns/Nc

53. Cuántas personas en el hogar comparten este presupuesto?

1. Una persona
2. Dos personas
3. De 3 a 4 personas
4. De 5 a 6 personas
- 5.** 7 o más personas
6. Ns/Nc

54. Ud tiene...?

	SI	NO
Computadora		<b>X</b>
Televisor	<b>X</b>	
Horno a microondas		<b>X</b>
Radio	<b>X</b>	
Auto	<b>X</b>	
Video		<b>X</b>
Vivienda	<b>X</b>	

55. Antes de iniciar la próxima entrevista el encuestador deberá observar el nivel de vivienda

1	Media Superior
2	Media media
<b>3</b>	Media baja
4	Muy modesta o Marginal

Extracto de la encuesta realizada por  
un diario regional de alta circulación

ENCUESTA DE OPINIÓN  
**QUIEN LEE QUÉ?**

Domicilio: \_\_\_\_\_ N° de encuesta: 77

1. Sexo: H  M

2. Edad: 55

49. Es Ud. jefe de familia? SI  NO

15. De las secciones que tiene el diario cuál lee, cuál no lee y cuál no conoce? (L)

Tipo de noticia	LEE	NO LEE	Ns/Nc
Noticias sobre la realidad política, económica y general del ámbito internacional		<b>x</b>	
Clasificados		<b>x</b>	
Noticias sobre sexo, violencia, droga,			<b>x</b>
Noticias policiales		<b>x</b>	
Noticias nacionales económicas		<b>x</b>	
Noticias nacionales políticas		<b>x</b>	
Información sobre la corrupción política y social			<b>x</b>
Deportes	<b>x</b>		
Comentarios Científicos			<b>x</b>
Información Gremial			<b>x</b>
Notas Editoriales			<b>x</b>
Comentarios firmados por analistas económicos y políticos			<b>x</b>
Comentarios firmados por analistas locales sobre aspectos de la realidad regional			<b>x</b>
Noticias y comentarios Agropecuarios			<b>x</b>
Noticias regionales			<b>x</b>
Investigaciones periodísticas			<b>x</b>
Información sobre Espectáculos			<b>x</b>
Sociales			<b>x</b>
Chistes, Horóscopos y entretenimiento en general		<b>x</b>	
Suplementos y revista dominical	<b>x</b>		

50.Cuál es su ocupación principal?.

1	Desocupado
2	Estudiante
<input checked="" type="radio"/> 3	Ama de Casa
4	Jubilado/pensionado
5	Empleadores
6	Trabajos no especializados o changarín
7	Comerciante sin personal
8	Técnico
9	Profesionales independientes y otros autónomos
10	Relación de dependencia
11	Obrero, cal. y no cal., capataz
12	Empleado s/jerarquía
13	Jefe intermedio
14	Ns/Nc

51.Cuál es el máximo nivel de estudios cursados por Ud.?

1	Sin instrucción
2	Primario incompleto
<input checked="" type="radio"/> 3	Primario completo
4	Secundario incompleto
5	Secundario completo
6	Terciario incompleto
7	Terciario completo
8	Universitario incompleto
9	Universitario completo
10	Ns/Nc

52. Cuáles son los ingresos mensuales de su grupo familiar?

<input checked="" type="radio"/> 1	Menos de 300
2	De 300 a 500
3	De 501 a 700
4	De 701 a 1000
5	De 1001 a 1500
6	De 1501 a 2500
7	De 2501 a 4000
8	Más de 4000
9	Ns/Nc

53. Cuántas personas en el hogar comparten este presupuesto?

1. Una persona
2. Dos personas
3. De 3 a 4 personas
4. De 5 a 6 personas
5. 7 o más personas
6. Ns/Nc

54. Ud tiene...?

	SI	NO
Computadora		<b>x</b>
Televisor	<b>x</b>	
Horno a microondas		<b>x</b>
Radio	<b>x</b>	
Auto		<b>x</b>
Video		<b>x</b>
Vivienda		

55. Antes de iniciar la próxima entrevista el encuestador deberá observar el nivel de vivienda

1	Media Superior
2	Media media
<input checked="" type="radio"/> 3	Media baja
4	Muy modesta o Marginal



Extracto de la encuesta realizada por  
un diario regional de alta circulación

ENCUESTA DE OPINIÓN  
QUIEN LEE QUÉ?

Domicilio: \_\_\_\_\_ N° de encuesta: **126**

1. Sexo **H** M

2. Edad: **33**

49. Es Ud. jefe de familia? **SI** NO

15. De las secciones que tiene el diario cuál lee, cuál no lee y cuál no conoce? (L)

Tipo de noticia	LEE	NO LEE	Ns/Nc
Noticias sobre la realidad política, económica y general del ámbito internacional		<b>x</b>	
Clasificados		<b>x</b>	
Noticias sobre sexo, violencia, droga,			<b>x</b>
Noticias policiales			<b>x</b>
Noticias nacionales económicas		<b>x</b>	
Noticias nacionales políticas		<b>x</b>	
Información sobre la corrupción política y social		<b>x</b>	
Deportes		<b>x</b>	
Comentarios Científicos			<b>x</b>
Información Gremial			<b>x</b>
Notas Editoriales			<b>x</b>
Comentario firmados por analistas económicos y políticos			<b>x</b>
Comentarios firmados por analistas locales sobre aspectos de la realidad regional			<b>x</b>
Noticias y comentarios Agropecuarios		<b>x</b>	
Noticias regionales			<b>x</b>
Investigaciones periódicas			<b>x</b>
Información sobre Espectáculos		<b>x</b>	
Sociales			<b>x</b>
Chistes, Horóscopos y entretenimiento en general			<b>x</b>
Suplementos y revista dominical			<b>x</b>

50. Cuál es su ocupación principal?

1	Desocupado
2	Estudiante
3	Ama de Casa
4	Jubilado/pensionado
5	Empleadores
6	Trabajos no especializados o changarín
7	Comerciante sin personal
<b>8</b>	Técnico
9	Profesionales independientes y otros autónomos
10	Relación de dependencia
11	Obrero, cal. y no cal., capataz
12	Empleado s/jerarquía
13	Jefe intermedio
14	Ns/Nc

51. Cuál es el máximo nivel de estudios cursados por Ud.?

1	Sin instrucción
2	Primario incompleto
3	Primario completo
4	Secundario incompleto
<b>5</b>	Secundario completo
6	Terciario incompleto
7	Terciario completo
8	Universitario incompleto
9	Universitario completo
10	Ns/Nc

52. Cuáles son los ingresos mensuales de su grupo familiar?

1	Menos de 300
2	De 300 a 500
3	De 501 a 700
<b>4</b>	De 701 a 1000
5	De 1001 a 1500
6	De 1501 a 2500
7	De 2501 a 4000
8	Más de 4000
9	Ns/Nc

53. Cuántas personas en el hogar comparten este presupuesto?

1. Una persona
2. Dos personas
- 3**. De 3 a 4 personas
4. De 5 a 6 personas
5. 7 o más personas
6. Ns/Nc

54. Ud tiene...?

	SI	NO
Computadora		<b>x</b>
Televisor	<b>x</b>	
Horno a microondas		<b>x</b>
Radio	<b>x</b>	
Auto	<b>x</b>	
Video		<b>x</b>
Vivienda		<b>x</b>

55. Antes de iniciar la próxima entrevista el encuestador deberá observar el nivel de vivienda

1	Media Superior
2	Media media
<b>3</b>	Media baja
4	Muy modesta o Marginal

Extracto de la encuesta realizada por  
un diario regional de alta circulación

ENCUESTA DE OPINIÓN  
**QUIEN LEE QUÉ?**

Domicilio: \_\_\_\_\_ N° de encuesta: **195**

1. Sexo:  F M

2. Edad: 70

49. Es Ud. jefe de familia?  SI NO

15. De las secciones que tiene el diario cuál lee, cuál no lee y cuál no conoce? (L)

Tipo de noticia	LEE	NO LEE	Ns/Nc
Noticias sobre la realidad política, económica y general del ámbito internacional		<b>X</b>	
Clasificados		<b>X</b>	
Noticias sobre sexo, violencia, droga,		<b>X</b>	
Noticias policiales		<b>X</b>	
Noticias nacionales económicas		<b>X</b>	
Noticias nacionales políticas		<b>X</b>	
Información sobre la corrupción política y social		<b>X</b>	
Deportes	<b>X</b>		
Comentarios Científicos			<b>X</b>
Información Gremial			<b>X</b>
Notas Editoriales			<b>X</b>
Comentario firmados por analistas económicos y políticos			<b>X</b>
Comentarios firmados por analistas locales sobre aspectos de la realidad regional			<b>X</b>
Noticias y comentarios Agropecuarios		<b>X</b>	
Noticias regionales			<b>X</b>
Investigaciones periódicas			<b>X</b>
Información sobre Espectáculos	<b>X</b>		
Sociales			<b>X</b>
Chistes, Horóscopos y entretenimiento en general	<b>X</b>		
Suplementos y revista dominical	<b>X</b>		

50. Cuál es su ocupación principal?.

1	Desocupado
2	Estudiante
3	Ama de Casa
4	Jubilado/pensionado
5	Empleadores
6	Trabajos no especializados o changarín
7	Comerciante sin personal
8	Técnico
<input checked="" type="radio"/> 9	Profesionales independientes y otros autónomos
10	Relación de dependencia
11	Obrero, cal. y no cal., capataz
12	Empleado s/jerarquía
13	Jefe intermedio
14	Ns/Nc

51. Cuál es el máximo nivel de estudios cursados por Ud.?

1	Sin instrucción
2	Primario incompleto
3	Primario completo
4	Secundario incompleto
5	Secundario completo
6	Terciario incompleto
<input checked="" type="radio"/> 7	Terciario completo
8	Universitario incompleto
9	Universitario completo
10	Ns/Nc

52. Cuáles son los ingresos mensuales de su grupo familiar?

1	Menos de 300
2	De 300 a 500
3	De 501 a 700
4	De 701 a 1000
5	De 1001 a 1500
<input checked="" type="radio"/> 6	De 1501 a 2500
7	De 2501 a 4000
8	Más de 4000
9	Ns/Nc

53. Cuántas personas en el hogar comparten este presupuesto?

1. Una persona
2. Dos personas
3. De 3 a 4 personas
4. De 5 a 6 personas
5. 7 o más personas
6. Ns/Nc

54. Ud tiene...?

	SI	NO
Computadora		<b>X</b>
Televisor	<b>X</b>	
Horno a microondas	<b>X</b>	
Radio	<b>X</b>	
Auto	<b>X</b>	
Video		<b>X</b>
Vivienda	<b>X</b>	

55. Antes de iniciar la próxima entrevista el encuestador deberá observar el nivel de vivienda

<input checked="" type="radio"/> 1	Media Superior
2	Media media
3	Media baja
4	Muy modesta o Marginal

Extracto de la encuesta realizada por  
un diario regional de alta circulación

ENCUESTA DE OPINIÓN  
QUIEN LEE QUÉ?

Domicilio: \_\_\_\_\_ N° de encuesta: **203**

1. Sexo:  H  M

2. Edad: 16

49. Es Ud. jefe de familia? SI  NO

15. De las secciones que tiene el diario cuál lee, cuál no lee y cuál no conoce? (L)

Tipo de noticia	LEE	NO LEE	Ns/Nc
Noticias sobre la realidad política, económica y general del ámbito internacional		<b>X</b>	
Clasificados	<b>X</b>		
Noticias sobre sexo, violencia, droga,		<b>X</b>	
Noticias policiales		<b>X</b>	
Noticias nacionales económicas		<b>X</b>	
Noticias nacionales políticas		<b>X</b>	
Información sobre la corrupción política y social			<b>X</b>
Deportes		<b>X</b>	
Comentarios Científicos		<b>X</b>	
Información Gremial		<b>X</b>	
Notas Editoriales		<b>X</b>	
Comentario firmados por analistas económicos y políticos		<b>X</b>	
Comentarios firmados por analistas locales sobre aspectos de la realidad regional		<b>X</b>	
Noticias y comentarios Agropecuarios		<b>X</b>	
Noticias regionales		<b>X</b>	
Investigaciones periódicas		<b>X</b>	
Información sobre Espectáculos		<b>X</b>	
Sociales		<b>X</b>	
Chistes, Horóscopos y entretenimiento en general		<b>X</b>	
Suplementos y revista dominical		<b>X</b>	

50. Cuál es su ocupación principal?.

1	Desocupado
<input checked="" type="radio"/> 2	Estudiante
3	Ama de Casa
4	Jubilado/pensionado
5	Empleadores
6	Trabajos no especializados o changarín
7	Comerciante sin personal
8	Técnico
9	Profesionales independientes y otros autónomos
10	Relación de dependencia
11	Obrero, cal. y no cal., capataz
12	Empleado s/jerarquía
13	Jefe intermedio
14	Ns/Nc

51. Cuál es el máximo nivel de estudios cursados por Ud.?

1	Sin instrucción
2	Primario incompleto
3	Primario completo
<input checked="" type="radio"/> 4	Secundario incompleto
5	Secundario completo
6	Terciario incompleto
7	Terciario completo
8	Universitario incompleto
9	Universitario completo
10	Ns/Nc

52. Cuáles son los ingresos mensuales de su grupo familiar?

1	Menos de 300
<input checked="" type="radio"/> 2	De 300 a 500
3	De 501 a 700
4	De 701 a 1000
5	De 1001 a 1500
6	De 1501 a 2500
7	De 2501 a 4000
8	Más de 4000
9	Ns/Nc

53. Cuántas personas en el hogar comparten este presupuesto?

1. Una persona
2. Dos personas
3. De 3 a 4 personas
4. De 5 a 6 personas
5. 7 o más personas
6. Ns/Nc

54. Ud tiene...?

	SI	NO
Computadora	<b>X</b>	
Televisor	<b>X</b>	
Horno a microondas	<b>X</b>	
Radio	<b>X</b>	
Auto	<b>X</b>	
Video	<b>X</b>	
Vivienda	<b>X</b>	

55. Antes de iniciar la próxima entrevista el encuestador deberá observar el nivel de vivienda

<input checked="" type="radio"/> 1	Media Superior
2	Media media
3	Media baja
4	Muy modesta o Marginal

Extracto de la encuesta realizada por  
un diario regional de alta circulación

ENCUESTA DE OPINIÓN  
QUIEN LEE QUÉ?

Domicilio: \_\_\_\_\_ N° de encuesta: **210**

1. Sexo: H  M

2. Edad: **26**

49. Es Ud. jefe de familia? SI  NO

15. De las secciones que tiene el diario cuál lee, cuál no lee y cuál no conoce? (L)

Tipo de noticia	LEE	NO LEE	Ns/Nc
Noticias sobre la realidad política, económica y general del ámbito internacional		<b>x</b>	
Clasificados			<b>x</b>
Noticias sobre sexo, violencia, droga,			<b>x</b>
Noticias policiales		<b>x</b>	
Noticias nacionales económicas		<b>x</b>	
Noticias nacionales políticas		<b>x</b>	
Información sobre la corrupción política y social	<b>x</b>		
Deportes	<b>x</b>		
Comentarios Científicos			<b>x</b>
Información Gremial			<b>x</b>
Notas Editoriales			<b>x</b>
Comentario firmados por analistas económicos y políticos	<b>x</b>		
Comentarios firmados por analistas locales sobre aspectos de la realidad regional			<b>x</b>
Noticias y comentarios Agropecuarios			<b>x</b>
Noticias regionales		<b>x</b>	
Investigaciones periódicas		<b>x</b>	
Información sobre Espectáculos			<b>x</b>
Sociales			<b>x</b>
Chistes, Horóscopos y entretenimiento en general		<b>x</b>	
Suplementos y revista dominical			<b>x</b>

50. Cuál es su ocupación principal?.

1	Desocupado
2	Estudiante
<b>3</b>	Ama de Casa
4	Jubilado/pensionado
5	Empleadores
6	Trabajos no especializados o changarín
7	Comerciante sin personal
8	Técnico
9	Profesionales independientes y otros autónomos
10	Relación de dependencia
11	Obrero, cal. y no cal., capataz
12	Empleado s/jerarquía
13	Jefe intermedio
14	Ns/Nc

51. Cuál es el máximo nivel de estudios cursados por Ud.?

1	Sin instrucción
2	Primario incompleto
3	Primario completo
<b>4</b>	Secundario incompleto
5	Secundario completo
6	Terciario incompleto
7	Terciario completo
8	Universitario incompleto
9	Universitario completo
10	Ns/Nc

52. Cuáles son los ingresos mensuales de su grupo familiar?

1	Menos de 300
<b>2</b>	De 300 a 500
3	De 501 a 700
4	De 701 a 1000
5	De 1001 a 1500
6	De 1501 a 2500
7	De 2501 a 4000
8	Más de 4000
9	Ns/Nc

53. Cuántas personas en el hogar comparten este presupuesto?

1. Una persona
2. Dos personas
3. De 3 a 4 personas
- 4.** De 5 a 6 personas
5. 7 o más personas
6. Ns/Nc

54. Ud tiene...?

	SI	NO
Computadora		<b>x</b>
Televisor	<b>x</b>	
Horno a microondas		<b>x</b>
Radio	<b>x</b>	
Auto		<b>x</b>
Video		<b>x</b>
Vivienda		<b>x</b>

55. Antes de iniciar la próxima entrevista el encuestador deberá observar el nivel de vivienda

1	Media Superior
2	Media media
3	Media baja
<b>4</b>	Muy modesta o Marginal

Extracto de la encuesta realizada por  
un diario regional de alta circulación

ENCUESTA DE OPINIÓN  
QUIEN LEE QUÉ?

Domicilio: \_\_\_\_\_ N° de encuesta: 274

1. Sexo:  H  M

2. Edad: 16

49. Es Ud. jefe de familia? SI  NO

15. De las secciones que tiene el diario cuál lee, cuál no lee y cuál no conoce? (L)

Tipo de noticia	LEE	NO LEE	Ns/Nc
Noticias sobre la realidad política, económica y general del ámbito internacional	<input checked="" type="checkbox"/>		
Clasificados	<input checked="" type="checkbox"/>		
Noticias sobre sexo, violencia, droga,		<input checked="" type="checkbox"/>	
Noticias policiales		<input checked="" type="checkbox"/>	
Noticias nacionales económicas		<input checked="" type="checkbox"/>	
Noticias nacionales políticas	<input checked="" type="checkbox"/>		
Información sobre la corrupción política y social		<input checked="" type="checkbox"/>	
Deportes		<input checked="" type="checkbox"/>	
Comentarios Científicos	<input checked="" type="checkbox"/>		
Información Gremial	<input checked="" type="checkbox"/>		
Notas Editoriales	<input checked="" type="checkbox"/>		
Comentarios firmados por analistas económicos y políticos	<input checked="" type="checkbox"/>		
Comentarios firmados por analistas locales sobre aspectos de la realidad regional		<input checked="" type="checkbox"/>	
Noticias y comentarios Agropecuarios	<input checked="" type="checkbox"/>		
Noticias regionales		<input checked="" type="checkbox"/>	
Investigaciones periodísticas	<input checked="" type="checkbox"/>		
Información sobre Espectáculos	<input checked="" type="checkbox"/>		
Sociales	<input checked="" type="checkbox"/>		
Chistes, Horóscopos y entretenimiento en general		<input checked="" type="checkbox"/>	
Suplementos y revista dominical	<input checked="" type="checkbox"/>		

50. Cuál es su ocupación principal?

<input checked="" type="radio"/> 1	Desocupado
<input type="radio"/> 2	Estudiante
<input type="radio"/> 3	Ama de Casa
<input type="radio"/> 4	Jubilado/pensionado
<input type="radio"/> 5	Empleadores
<input type="radio"/> 6	Trabajos no especializados o changarín
<input type="radio"/> 7	Comerciante sin personal
<input type="radio"/> 8	Técnico
<input type="radio"/> 9	Profesionales independientes y otros autónomos
<input type="radio"/> 10	Relación de dependencia
<input type="radio"/> 11	Obrero, cal. y no cal., capataz
<input type="radio"/> 12	Empleado s/jerarquía
<input type="radio"/> 13	Jefe intermedio
<input type="radio"/> 14	Ns/Nc

51. Cuál es el máximo nivel de estudios cursados por Ud.?

<input type="radio"/> 1	Sin instrucción
<input type="radio"/> 2	Primario incompleto
<input type="radio"/> 3	Primario completo
<input checked="" type="radio"/> 4	Secundario incompleto
<input type="radio"/> 5	Secundario completo
<input type="radio"/> 6	Terciario incompleto
<input type="radio"/> 7	Terciario completo
<input type="radio"/> 8	Universitario incompleto
<input type="radio"/> 9	Universitario completo
<input type="radio"/> 10	Ns/Nc

52. Cuáles son los ingresos mensuales de su grupo familiar?

<input type="radio"/> 1	Menos de 300
<input checked="" type="radio"/> 2	De 300 a 500
<input type="radio"/> 3	De 501 a 700
<input type="radio"/> 4	De 701 a 1000
<input type="radio"/> 5	De 1001 a 1500
<input type="radio"/> 6	De 1501 a 2500
<input type="radio"/> 7	De 2501 a 4000
<input type="radio"/> 8	Más de 4000
<input type="radio"/> 9	Ns/Nc

53. Cuántas personas en el hogar comparten este presupuesto?

1. Una persona
2. Dos personas
3. De 3 a 4 personas
4. De 5 a 6 personas
5. 7 o más personas
6. Ns/Nc

54. Ud tiene...?

	SI	NO
Computadora		<input checked="" type="checkbox"/>
Televisor	<input checked="" type="checkbox"/>	
Horno a microondas		<input checked="" type="checkbox"/>
Radio	<input checked="" type="checkbox"/>	
Auto	<input checked="" type="checkbox"/>	
Video		<input checked="" type="checkbox"/>
Vivienda		

55. Antes de iniciar la próxima entrevista el encuestador deberá observar el nivel de vivienda

<input type="radio"/> 1	Media Superior
<input type="radio"/> 2	Media media
<input checked="" type="radio"/> 3	Media baja
<input type="radio"/> 4	Muy modesta o Marginal

Extracto de la encuesta realizada por  
un diario regional de alta circulación

ENCUESTA DE OPINIÓN  
**QUIEN LEE QUÉ?**

Domicilio: \_\_\_\_\_ N° de encuesta: 300

1. Sexo:  H  M

2. Edad: 59

49. Es Ud. jefe de familia?  SI  NO

15. De las secciones que tiene el diario cuál lee, cuál no lee y cuál no conoce? (L)

Tipo de noticia	LEE	NO LEE	Ns/Nc
Noticias sobre la realidad política, económica y general del ámbito internacional		<b>X</b>	
Clasificados	<b>X</b>		
Noticias sobre sexo, violencia, droga,		<b>X</b>	
Noticias policiales		<b>X</b>	
Noticias nacionales económicas	<b>X</b>		
Noticias nacionales políticas	<b>X</b>		
Información sobre la corrupción política y social		<b>X</b>	
Deportes		<b>X</b>	
Comentarios Científicos		<b>X</b>	
Información Gremial		<b>X</b>	
Notas Editoriales		<b>X</b>	
Comentarios firmados por analistas económicos y políticos	<b>X</b>		
Comentarios firmados por analistas locales sobre aspectos de la realidad regional	<b>X</b>		
Noticias y comentarios Agropecuarios	<b>X</b>		
Noticias regionales		<b>X</b>	
Investigaciones periodísticas		<b>X</b>	
Información sobre Espectáculos	<b>X</b>		
Sociales		<b>X</b>	
Chistes, Horóscopos y entretenimiento en general		<b>X</b>	
Suplementos y revista dominical	<b>X</b>		

50.Cuál es su ocupación principal?

1	Desocupado
2	Estudiante
3	Ama de Casa
<input checked="" type="radio"/> 4	Jubilado/pensionado
5	Empleadores
6	Trabajos no especializados o changarín
7	Comerciante sin personal
8	Técnico
9	Profesionales independientes y otros autónomos
10	Relación de dependencia
11	Obrero, cal. y no cal., capataz
12	Empleado s/jerarquía
13	Jefe intermedio
14	Ns/Nc

51.Cuál es el máximo nivel de estudios cursados por Ud.?

1	Sin instrucción
2	Primario incompleto
<input checked="" type="radio"/> 3	Primario completo
4	Secundario incompleto
5	Secundario completo
6	Terciario incompleto
7	Terciario completo
8	Universitario incompleto
9	Universitario completo
10	Ns/Nc

52.Cuáles son los ingresos mensuales de su grupo familiar?

1	Menos de 300
2	De 300 a 500
<input checked="" type="radio"/> 3	De 501 a 700
4	De 701 a 1000
5	De 1001 a 1500
6	De 1501 a 2500
7	De 2501 a 4000
8	Más de 4000
9	Ns/Nc

53. Cuántas personas en el hogar comparten este presupuesto?

1. Una persona
2. Dos personas
3. De 3 a 4 personas
4. De 5 a 6 personas
5. 7 o más personas
6. Ns/Nc

54. Ud tiene...?

	SI	NO
Computadora		<b>X</b>
Televisor	<b>X</b>	
Horno a microondas		<b>X</b>
Radio	<b>X</b>	
Auto	<b>X</b>	
Video		<b>X</b>
Vivienda	<b>X</b>	

55. Antes de iniciar la próxima entrevista el encuestador deberá observar el nivel de vivienda

1	Media Superior
<input checked="" type="radio"/> 2	Media media
3	Media baja
4	Muy modesta o Marginal

Extracto de la encuesta realizada por  
un diario regional de alta circulación

ENCUESTA DE OPINIÓN  
**QUIEN LEE QUÉ?**

Domicilio: \_\_\_\_\_ N° de encuesta: **310**

1. Sexo: **(H)** M

2. Edad: **49**

49. Es Ud. jefe de familia? SI **(NO)**

15. De las secciones que tiene el diario cuál lee, cuál no lee y cuál no conoce? (L)

Tipo de noticia	LEE	NO LEE	Ns/Nc
Noticias sobre la realidad política, económica y general del ámbito internacional	<b>X</b>		
Clasificados	<b>X</b>		
Noticias sobre sexo, violencia, droga,	<b>X</b>		
Noticias policiales	<b>X</b>		
Noticias nacionales económicas		<b>X</b>	
Noticias nacionales políticas		<b>X</b>	
Información sobre la corrupción política y social		<b>X</b>	
Deportes		<b>X</b>	
Comentarios Científicos	<b>X</b>		
Información Gremial	<b>X</b>		
Notas Editoriales	<b>X</b>		
Comentario firmados por analistas económicos y políticos	<b>X</b>		
Comentarios firmados por analistas locales sobre aspectos de la realidad regional	<b>X</b>		
Noticias y comentarios Agropecuarios	<b>X</b>		
Noticias regionales		<b>X</b>	
Investigaciones periódicas		<b>X</b>	
Información sobre Espectáculos		<b>X</b>	
Sociales		<b>X</b>	
Chistes, Horóscopos y entretenimiento en general	<b>X</b>		
Suplementos y revista dominical	<b>X</b>		

50. Cuál es su ocupación principal?.

1	Desocupado
2	Estudiante
3	Ama de Casa
4	Jubilado/pensionado
5	Empleadores
6	Trabajos no especializados o changarín
<b>7</b>	Comerciante sin personal
8	Técnico
9	Profesionales independientes y otros autónomos
10	Relación de dependencia
11	Obrero, cal. y no cal., capataz
12	Empleado s/jerarquía
13	Jefe intermedio
14	Ns/Nc

51. Cuál es el máximo nivel de estudios cursados por Ud.?

1	Sin instrucción
2	Primario incompleto
<b>3</b>	Primario completo
4	Secundario incompleto
5	Secundario completo
6	Terciario incompleto
7	Terciario completo
8	Universitario incompleto
9	Universitario completo
10	Ns/Nc

52. Cuáles son los ingresos mensuales de su grupo familiar?

1	Menos de 300
<b>2</b>	De 300 a 500
3	De 501 a 700
4	De 701 a 1000
5	De 1001 a 1500
6	De 1501 a 2500
7	De 2501 a 4000
8	Más de 4000
9	Ns/Nc

53. Cuántas personas en el hogar comparten este presupuesto?

1. Una persona
2. Dos personas
3. De 3 a 4 personas
- 4.** De 5 a 6 personas
5. 7 o más personas
6. Ns/Nc

54. Ud tiene...?

	SI	NO
Computadora		<b>X</b>
Televisor	<b>X</b>	
Horno a microondas		<b>X</b>
Radio	<b>X</b>	
Auto		<b>X</b>
Video	<b>X</b>	
Vivienda	<b>X</b>	

55. Antes de iniciar la próxima entrevista el encuestador deberá observar el nivel de vivienda

1	Media Superior
2	Media media
3	Media baja
<b>4</b>	Muy modesta o Marginal

Extracto de la encuesta realizada por  
un diario regional de alta circulación

ENCUESTA DE OPINIÓN  
QUIEN LEE QUÉ?

Domicilio: \_\_\_\_\_ N° de encuesta: **317**

1. Sexo: H **(M)** 2. Edad: **45** 49. Es Ud. jefe de familia? **(SI)** NO

15. De las secciones que tiene el diario cuál lee, cuál no lee y cuál no conoce? (L)

Tipo de noticia	LEE	NO LEE	Ns/Nc
Noticias sobre la realidad política, económica y general del ámbito internacional		<b>X</b>	
Clasificados		<b>X</b>	
Noticias sobre sexo, violencia, droga,	<b>X</b>		
Noticias policiales		<b>X</b>	
Noticias nacionales económicas		<b>X</b>	
Noticias nacionales políticas		<b>X</b>	
Información sobre la corrupción política y social		<b>X</b>	
Deportes	<b>X</b>		
Comentarios Científicos		<b>X</b>	
Información Gremial	<b>X</b>		
Notas Editoriales	<b>X</b>		
Comentario firmados por analistas económicos y políticos		<b>X</b>	
Comentarios firmados por analistas locales sobre aspectos de la realidad regional		<b>X</b>	
Noticias y comentarios Agropecuarios	<b>X</b>		
Noticias regionales		<b>X</b>	
Investigaciones periódicas		<b>X</b>	
Información sobre Espectáculos		<b>X</b>	
Sociales		<b>X</b>	
Chistes, Horóscopos y entretenimiento en general		<b>X</b>	
Suplementos y revista dominical		<b>X</b>	

50. Cuál es su ocupación principal?.

1	Desocupado
<b>2</b>	Estudiante
<b>3</b>	Ama de Casa
4	Jubilado/pensionado
5	Empleadores
6	Trabajos no especializados o changarín
7	Comerciante sin personal
8	Técnico
9	Profesionales independientes y otros autónomos
10	Relación de dependencia
11	Obrero, cal. y no cal., capataz
12	Empleado s/jerarquía
13	Jefe intermedio
14	Ns/Nc

51. Cuál es el máximo nivel de estudios cursados por Ud.?

1	Sin instrucción
<b>2</b>	Primario incompleto
<b>3</b>	Primario completo
4	Secundario incompleto
5	Secundario completo
6	Terciario incompleto
7	Terciario completo
8	Universitario incompleto
9	Universitario completo
10	Ns/Nc

52. Cuáles son los ingresos mensuales de su grupo familiar?

<b>1</b>	Menos de 300
2	De 300 a 500
3	De 501 a 700
4	De 701 a 1000
5	De 1001 a 1500
6	De 1501 a 2500
7	De 2501 a 4000
8	Más de 4000
9	Ns/Nc

53. Cuántas personas en el hogar comparten este presupuesto?

1. Una persona
2. Dos personas
- 3**. De 3 a 4 personas
4. De 5 a 6 personas
5. 7 o más personas
6. Ns/Nc

54. Ud tiene...?

	SI	NO
Computadora		<b>X</b>
Televisor	<b>X</b>	
Horno a microondas		<b>X</b>
Radio	<b>X</b>	
Auto		<b>X</b>
Video		<b>X</b>
Vivienda		

55. Antes de iniciar la próxima entrevista el encuestador deberá observar el nivel de vivienda

1	Media Superior
2	Media media
3	Media baja
<b>4</b>	Muy modesta o Marginal



Extracto de la encuesta realizada por  
un diario regional de alta circulación

ENCUESTA DE OPINIÓN  
QUIEN LEE QUÉ?

Domicilio: \_\_\_\_\_ N° de encuesta: **318**

1. Sexo: H  M

2. Edad: **41**

49. Es Ud. jefe de familia? SI  NO

15. De las secciones que tiene el diario cuál lee, cuál no lee y cuál no conoce? (L)

Tipo de noticia	LEE	NO LEE	Ns/Nc
Noticias sobre la realidad política, económica y general del ámbito internacional	<input checked="" type="checkbox"/>		
Clasificados	<input checked="" type="checkbox"/>		
Noticias sobre sexo, violencia, droga,	<input checked="" type="checkbox"/>		
Noticias policiales	<input checked="" type="checkbox"/>		
Noticias nacionales económicas	<input checked="" type="checkbox"/>		
Noticias nacionales políticas	<input checked="" type="checkbox"/>		
Información sobre la corrupción política y social	<input checked="" type="checkbox"/>		
Deportes	<input checked="" type="checkbox"/>		
Comentarios Científicos	<input checked="" type="checkbox"/>		
Información Gremial	<input checked="" type="checkbox"/>		
Notas Editoriales	<input checked="" type="checkbox"/>		
Comentario firmados por analistas económicos y políticos	<input checked="" type="checkbox"/>		
Comentarios firmados por analistas locales sobre aspectos de la realidad regional	<input checked="" type="checkbox"/>		
Noticias y comentarios Agropecuarios	<input checked="" type="checkbox"/>		
Noticias regionales		<input checked="" type="checkbox"/>	
Investigaciones periódicas	<input checked="" type="checkbox"/>		
Información sobre Espectáculos	<input checked="" type="checkbox"/>		
Sociales		<input checked="" type="checkbox"/>	
Chistes, Horóscopos y entretenimiento en general		<input checked="" type="checkbox"/>	
Suplementos y revista dominical	<input checked="" type="checkbox"/>		

50. Cuál es su ocupación principal?.

1	Desocupado
2	Estudiante
3	Ama de Casa
4	Jubilado/pensionado
5	Empleadores
6	Trabajos no especializados o changarín
7	Comerciante sin personal
8	Técnico
9	Profesionales independientes y otros autónomos
<input checked="" type="radio"/> 10	Relación de dependencia
11	Obrero, cal. y no cal., capataz
12	Empleado s/jerarquía
13	Jefe intermedio
14	Ns/Nc

51. Cuál es el máximo nivel de estudios cursados por Ud.?

1	Sin instrucción
2	Primario incompleto
3	Primario completo
<input checked="" type="radio"/> 4	Secundario incompleto
5	Secundario completo
6	Terciario incompleto
7	Terciario completo
8	Universitario incompleto
9	Universitario completo
10	Ns/Nc

52. Cuáles son los ingresos mensuales de su grupo familiar?

1	Menos de 300
<input checked="" type="radio"/> 2	De 300 a 500
3	De 501 a 700
4	De 701 a 1000
5	De 1001 a 1500
6	De 1501 a 2500
7	De 2501 a 4000
8	Más de 4000
9	Ns/Nc

53. Cuántas personas en el hogar comparten este presupuesto?

1. Una persona
2. Dos personas
3. De 3 a 4 personas
4. De 5 a 6 personas
5. 7 o más personas
6. Ns/Nc

54. Ud tiene...?

	SI	NO
Computadora		<input checked="" type="checkbox"/>
Televisor	<input checked="" type="checkbox"/>	
Horno a microondas		<input checked="" type="checkbox"/>
Radio	<input checked="" type="checkbox"/>	
Auto		<input checked="" type="checkbox"/>
Video		<input checked="" type="checkbox"/>
Vivienda		

55. Antes de iniciar la próxima entrevista el encuestador deberá observar el nivel de vivienda

1	Media Superior
2	Media media
3	Media baja
<input checked="" type="radio"/> 4	Muy modesta o Marginal

Extracto de la encuesta realizada por  
un diario regional de alta circulación

ENCUESTA DE OPINIÓN  
QUIEN LEE QUÉ?

Domicilio: \_\_\_\_\_

N° de encuesta: **334**

1. Sexo: **(H)** M

2. Edad: **51**

49. Es Ud. jefe de familia? **(SI)** NO

15. De las secciones que tiene el diario cuál lee, cuál no lee y cuál no conoce? (L)

Tipo de noticia	LEE	NO LEE	Ns/Nc
Noticias sobre la realidad política, económica y general del ámbito internacional			<b>x</b>
Clasificados			<b>x</b>
Noticias sobre sexo, violencia, droga,			<b>x</b>
Noticias policiales			<b>x</b>
Noticias nacionales económicas			<b>x</b>
Noticias nacionales políticas		<b>x</b>	
Información sobre la corrupción política y social		<b>x</b>	
Deportes			<b>x</b>
Comentarios Científicos			<b>x</b>
Información Gremial			<b>x</b>
Notas Editoriales		<b>x</b>	
Comentario firmados por analistas económicos y políticos		<b>x</b>	
Comentarios firmados por analistas locales sobre aspectos de la realidad regional		<b>x</b>	
Noticias y comentarios Agropecuarios			<b>x</b>
Noticias regionales			<b>x</b>
Investigaciones periódicas			<b>x</b>
Información sobre Espectáculos			<b>x</b>
Sociales			<b>x</b>
Chistes, Horóscopos y entretenimiento en general		<b>x</b>	
Suplementos y revista dominical		<b>x</b>	

50. Cuál es su ocupación principal?.

1	Desocupado
2	Estudiante
3	Ama de Casa
4	Jubilado/pensionado
5	Empleadores
6	Trabajos no especializados o changarín
<b>7</b>	Comerciante sin personal
8	Técnico
9	Profesionales independientes y otros autónomos
10	Relación de dependencia
11	Obrero, cal. y no cal., capataz
12	Empleado s/jerarquía
13	Jefe intermedio
14	Ns/Nc

51. Cuál es el máximo nivel de estudios cursados por Ud.?

1	Sin instrucción
2	Primario incompleto
<b>3</b>	Primario completo
4	Secundario incompleto
5	Secundario completo
6	Terciario incompleto
7	Terciario completo
8	Universitario incompleto
9	Universitario completo
10	Ns/Nc

52. Cuáles son los ingresos mensuales de su grupo familiar?

1	Menos de 300
2	De 300 a 500
<b>3</b>	De 501 a 700
4	De 701 a 1000
5	De 1001 a 1500
6	De 1501 a 2500
7	De 2501 a 4000
8	Más de 4000
9	Ns/Nc

53. Cuántas personas en el hogar comparten este presupuesto?

1. Una persona
2. Dos personas
- 3.** De 3 a 4 personas
4. De 5 a 6 personas
5. 7 o más personas
6. Ns/Nc

54. Ud tiene...?

	SI	NO
Computadora		<b>x</b>
Televisor	<b>x</b>	
Horno a microondas		<b>x</b>
Radio	<b>x</b>	
Auto		<b>x</b>
Video		<b>x</b>
Vivienda		

55. Antes de iniciar la próxima entrevista el encuestador deberá observar el nivel de vivienda

1	Media Superior
2	Media media
<b>3</b>	Media baja
4	Muy modesta o Marginal

Extracto de la encuesta realizada por  
un diario regional de alta circulación

ENCUESTA DE OPINIÓN  
QUIEN LEE QUÉ?

Domicilio: \_\_\_\_\_ N° de encuesta: **384**

1. Sexo: **(H)** M

2. Edad: **32**

49. Es Ud. jefe de familia? **(SI)** NO

15. De las secciones que tiene el diario cuál lee, cuál no lee y cuál no conoce? (L)

Tipo de noticia	LEE	NO LEE	Ns/Nc
Noticias sobre la realidad política, económica y general del ámbito internacional		<b>X</b>	
Clasificados		<b>X</b>	
Noticias sobre sexo, violencia, droga,		<b>X</b>	
Noticias policiales		<b>X</b>	
Noticias nacionales económicas		<b>X</b>	
Noticias nacionales políticas		<b>X</b>	
Información sobre la corrupción política y social		<b>X</b>	
Deportes		<b>X</b>	
Comentarios Científicos		<b>X</b>	
Información Gremial	<b>X</b>		
Notas Editoriales	<b>X</b>		
Comentario firmados por analistas económicos y políticos	<b>X</b>		
Comentarios firmados por analistas locales sobre aspectos de la realidad regional		<b>X</b>	
Noticias y comentarios Agropecuarios		<b>X</b>	
Noticias regionales		<b>X</b>	
Investigaciones periódicas		<b>X</b>	
Información sobre Espectáculos		<b>X</b>	
Sociales		<b>X</b>	
Chistes, Horóscopos y entretenimiento en general		<b>X</b>	
Suplementos y revista dominical		<b>X</b>	

50. Cuál es su ocupación principal?.

1	Desocupado
2	Estudiante
3	Ama de Casa
4	Jubilado/pensionado
5	Empleadores
6	Trabajos no especializados o changarín
7	Comerciante sin personal
8	Técnico
9	Profesionales independientes y otros autónomos
<b>(10)</b>	Relación de dependencia
11	Obrero, cal. y no cal., capataz
12	Empleado s/jerarquía
13	Jefe intermedio
14	Ns/Nc

52. Cuáles son los ingresos mensuales de su grupo familiar?

1	Menos de 300
<b>(2)</b>	De 300 a 500
3	De 501 a 700
4	De 701 a 1000
5	De 1001 a 1500
6	De 1501 a 2500
7	De 2501 a 4000
8	Más de 4000
9	Ns/Nc

54. Ud tiene...?

	SI	NO
Computadora		<b>X</b>
Televisor	<b>X</b>	
Horno a microondas		<b>X</b>
Radio	<b>X</b>	
Auto		<b>X</b>
Video		<b>X</b>
Vivienda		

51. Cuál es el máximo nivel de estudios cursados por Ud.?

1	Sin instrucción
<b>(2)</b>	Primario incompleto
<b>(3)</b>	Primario completo
4	Secundario incompleto
5	Secundario completo
6	Terciario incompleto
7	Terciario completo
8	Universitario incompleto
9	Universitario completo
10	Ns/Nc

53. Cuántas personas en el hogar comparten este presupuesto?

1. Una persona
2. Dos personas
3. De 3 a 4 personas
- (4)** De 5 a 6 personas
5. 7 o más personas
6. Ns/Nc

55. Antes de iniciar la próxima entrevista el encuestador deberá observar el nivel de vivienda

1	Media Superior
<b>(2)</b>	Media media
3	Media baja
4	Muy modesta o Marginal

Extracto de la encuesta realizada por  
un diario regional de alta circulación

ENCUESTA DE OPINIÓN  
**QUIEN LEE QUÉ?**

Domicilio: \_\_\_\_\_ N° de encuesta: **418**

1. Sexo: **(H)** M

2. Edad: **52**

49. Es Ud. jefe de familia? **(S)** NO

15. De las secciones que tiene el diario cuál lee, cuál no lee y cuál no conoce? (L)

Tipo de noticia	LEE	NO LEE	Ns/Nc
Noticias sobre la realidad política, económica y general del ámbito internacional	<b>X</b>		
Clasificados	<b>X</b>		
Noticias sobre sexo, violencia, droga,		<b>X</b>	
Noticias policiales		<b>X</b>	
Noticias nacionales económicas		<b>X</b>	
Noticias nacionales políticas		<b>X</b>	
Información sobre la corrupción política y social		<b>X</b>	
Deportes		<b>X</b>	
Comentarios Científicos	<b>X</b>		
Información Gremial	<b>X</b>		
Notas Editoriales	<b>X</b>		
Comentario firmados por analistas económicos y políticos	<b>X</b>		
Comentarios firmados por analistas locales sobre aspectos de la realidad regional	<b>X</b>		
Noticias y comentarios Agropecuarios		<b>X</b>	
Noticias regionales		<b>X</b>	
Investigaciones periódicas		<b>X</b>	
Información sobre Espectáculos	<b>X</b>		
Sociales	<b>X</b>		
Chistes, Horóscopos y entretenimiento en general	<b>X</b>		
Suplementos y revista dominical	<b>X</b>		

50. Cuál es su ocupación principal?.

1	Desocupado
2	Estudiante
3	Ama de Casa
4	Jubilado/pensionado
5	Empleadores
6	Trabajos no especializados o changarín
7	Comerciante sin personal
<b>8</b>	Técnico
9	Profesionales independientes y otros autónomos
10	Relación de dependencia
11	Obrero, cal. y no cal., capataz
12	Empleado s/jerarquía
13	Jefe intermedio
14	Ns/Nc

51. Cuál es el máximo nivel de estudios cursados por Ud.?

1	Sin instrucción
<b>2</b>	Primario incompleto
3	Primario completo
4	Secundario incompleto
5	Secundario completo
6	Terciario incompleto
7	Terciario completo
8	Universitario incompleto
9	Universitario completo
10	Ns/Nc

52. Cuáles son los ingresos mensuales de su grupo familiar?

1	Menos de 300
2	De 300 a 500
<b>3</b>	De 501 a 700
4	De 701 a 1000
5	De 1001 a 1500
6	De 1501 a 2500
7	De 2501 a 4000
8	Más de 4000
9	Ns/Nc

53. Cuántas personas en el hogar comparten este presupuesto?

1. Una persona
2. Dos personas
3. De 3 a 4 personas
- 4.** De 5 a 6 personas
5. 7 o más personas
6. Ns/Nc

54. Ud tiene...?

	SI	NO
Computadora		<b>X</b>
Televisor	<b>X</b>	
Horno a microondas		<b>X</b>
Radio	<b>X</b>	
Auto	<b>X</b>	
Video	<b>X</b>	
Vivienda		<b>X</b>

55. Antes de iniciar la próxima entrevista el encuestador deberá observar el nivel de vivienda

1	Media Superior
<b>2</b>	Media media
3	Media baja
4	Muy modesta o Marginal

Extracto de la encuesta realizada por  
un diario regional de alta circulación

ENCUESTA DE OPINIÓN  
QUIEN LEE QUÉ?

Domicilio: \_\_\_\_\_ N° de encuesta: **448**

1. Sexo: H  M

2. Edad: **45**

49. Es Ud. jefe de familia? SI  NO

15. De las secciones que tiene el diario cuál lee, cuál no lee y cuál no conoce? (L)

Tipo de noticia	LEE	NO LEE	Ns/Nc
Noticias sobre la realidad política, económica y general del ámbito internacional		<b>X</b>	
Clasificados	<b>X</b>		
Noticias sobre sexo, violencia, droga,		<b>X</b>	
Noticias policiales	<b>X</b>		
Noticias nacionales económicas		<b>X</b>	
Noticias nacionales políticas		<b>X</b>	
Información sobre la corrupción política y social		<b>X</b>	
Deportes	<b>X</b>		
Comentarios Científicos		<b>X</b>	
Información Gremial		<b>X</b>	
Notas Editoriales		<b>X</b>	
Comentario firmados por analistas económicos y políticos	<b>X</b>		
Comentarios firmados por analistas locales sobre aspectos de la realidad regional	<b>X</b>		
Noticias y comentarios Agropecuarios	<b>X</b>		
Noticias regionales		<b>X</b>	
Investigaciones periódicas		<b>X</b>	
Información sobre Espectáculos		<b>X</b>	
Sociales		<b>X</b>	
Chistes, Horóscopos y entretenimiento en general		<b>X</b>	
Suplementos y revista dominical		<b>X</b>	

50. Cuál es su ocupación principal?.

1	Desocupado
2	Estudiante
3	Ama de Casa
4	Jubilado/pensionado
5	Empleadores
6	Trabajos no especializados o changarín
7	Comerciante sin personal
8	Técnico
9	Profesionales independientes y otros autónomos
10	Relación de dependencia
11	Obrero, cal. y no cal., capataz
12	Empleado s/jerarquía
<b>13</b>	Jefe intermedio
14	Ns/Nc

51. Cuál es el máximo nivel de estudios cursados por Ud.?

1	Sin instrucción
2	Primario incompleto
3	Primario completo
4	Secundario incompleto
<b>5</b>	Secundario completo
6	Terciario incompleto
7	Terciario completo
8	Universitario incompleto
9	Universitario completo
10	Ns/Nc

52. Cuáles son los ingresos mensuales de su grupo familiar?

1	Menos de 300
2	De 300 a 500
<b>3</b>	De 501 a 700
4	De 701 a 1000
5	De 1001 a 1500
6	De 1501 a 2500
7	De 2501 a 4000
8	Más de 4000
9	Ns/Nc

53. Cuántas personas en el hogar comparten este presupuesto?

1. Una persona
2. Dos personas
- 3. De 3 a 4 personas**
4. De 5 a 6 personas
5. 7 o más personas
6. Ns/Nc

54. Ud tiene...?

	SI	NO
Computadora		<b>X</b>
Televisor	<b>X</b>	
Horno a microondas		<b>X</b>
Radio	<b>X</b>	
Auto	<b>X</b>	
Video		<b>X</b>
Vivienda		

55. Antes de iniciar la próxima entrevista el encuestador deberá observar el nivel de vivienda

1	Media Superior
<b>2</b>	Media media
3	Media baja
4	Muy modesta o Marginal

Extracto de la encuesta realizada por  
un diario regional de alta circulación

ENCUESTA DE OPINIÓN  
QUIEN LEE QUÉ?

Domicilio: \_\_\_\_\_ N° de encuesta: **456**

1. Sexo: H  M

2. Edad: 22

49. Es Ud. jefe de familia? S  NO

15. De las secciones que tiene el diario cuál lee, cuál no lee y cuál no conoce? (L)

Tipo de noticia	LEE	NO LEE	Ns/Nc
Noticias sobre la realidad política, económica y general del ámbito internacional		<input checked="" type="checkbox"/>	
Clasificados		<input checked="" type="checkbox"/>	
Noticias sobre sexo, violencia, droga,		<input checked="" type="checkbox"/>	
Noticias policiales		<input checked="" type="checkbox"/>	
Noticias nacionales económicas		<input checked="" type="checkbox"/>	
Noticias nacionales políticas		<input checked="" type="checkbox"/>	
Información sobre la corrupción política y social	<input checked="" type="checkbox"/>		
Deportes		<input checked="" type="checkbox"/>	
Comentarios Científicos		<input checked="" type="checkbox"/>	
Información Gremial		<input checked="" type="checkbox"/>	
Notas Editoriales	<input checked="" type="checkbox"/>		
Comentario firmados por analistas económicos y políticos	<input checked="" type="checkbox"/>		
Comentarios firmados por analistas locales sobre aspectos de la realidad regional		<input checked="" type="checkbox"/>	
Noticias y comentarios Agropecuarios	<input checked="" type="checkbox"/>		
Noticias regionales		<input checked="" type="checkbox"/>	
Investigaciones periódicas	<input checked="" type="checkbox"/>		
Información sobre Espectáculos		<input checked="" type="checkbox"/>	
Sociales		<input checked="" type="checkbox"/>	
Chistes, Horóscopos y entretenimiento en general		<input checked="" type="checkbox"/>	
Suplementos y revista dominical		<input checked="" type="checkbox"/>	

50. Cuál es su ocupación principal?.

1	Desocupado
2	Estudiante
<input checked="" type="radio"/> 3	Ama de Casa
4	Jubilado/pensionado
5	Empleadores
6	Trabajos no especializados o changarín
7	Comerciante sin personal
8	Técnico
9	Profesionales independientes y otros autónomos
10	Relación de dependencia
11	Obrero, cal. y no cal., capataz
12	Empleado s/jerarquía
13	Jefe intermedio
14	Ns/Nc

51. Cuál es el máximo nivel de estudios cursados por Ud.?

1	Sin instrucción
2	Primario incompleto
<input checked="" type="radio"/> 3	Primario completo
4	Secundario incompleto
5	Secundario completo
6	Terciario incompleto
7	Terciario completo
8	Universitario incompleto
9	Universitario completo
10	Ns/Nc

52. Cuáles son los ingresos mensuales de su grupo familiar?

<input checked="" type="radio"/> 1	Menos de 300
2	De 300 a 500
3	De 501 a 700
4	De 701 a 1000
5	De 1001 a 1500
6	De 1501 a 2500
7	De 2501 a 4000
8	Más de 4000
9	Ns/Nc

53. Cuántas personas en el hogar comparten este presupuesto?

1. Una persona
2. Dos personas
3. De 3 a 4 personas
4. De 5 a 6 personas
5. 7 o más personas
6. Ns/Nc

54. Ud tiene...?

	SI	NO
Computadora		<input checked="" type="checkbox"/>
Televisor	<input checked="" type="checkbox"/>	
Horno a microondas		<input checked="" type="checkbox"/>
Radio	<input checked="" type="checkbox"/>	
Auto		<input checked="" type="checkbox"/>
Video		<input checked="" type="checkbox"/>
Vivienda		

55. Antes de iniciar la próxima entrevista el encuestador deberá observar el nivel de vivienda

1	Media Superior
2	Media media
<input checked="" type="radio"/> 3	Media baja
4	Muy modesta o Marginal

Extracto de la encuesta realizada por  
un diario regional de alta circulación

ENCUESTA DE OPINIÓN  
QUIEN LEE QUÉ?

Domicilio: \_\_\_\_\_ N° de encuesta: **464**

1. Sexo: H (M) 2. Edad: **47** 49. Es Ud. jefe de familia? SI (NO)

15. De las secciones que tiene el diario cuál lee, cuál no lee y cuál no conoce? (L)

Tipo de noticia	LEE	NO LEE	Ns/Nc
Noticias sobre la realidad política, económica y general del ámbito internacional		<b>X</b>	
Clasificados	<b>X</b>		
Noticias sobre sexo, violencia, droga,	<b>X</b>		
Noticias policiales	<b>X</b>		
Noticias nacionales económicas		<b>X</b>	
Noticias nacionales políticas		<b>X</b>	
Información sobre la corrupción política y social	<b>X</b>		
Deportes	<b>X</b>		
Comentarios Científicos		<b>X</b>	
Información Gremial	<b>X</b>		
Notas Editoriales	<b>X</b>		
Comentario firmados por analistas económicos y políticos		<b>X</b>	
Comentarios firmados por analistas locales sobre aspectos de la realidad regional		<b>X</b>	
Noticias y comentarios Agropecuarios		<b>X</b>	
Noticias regionales		<b>X</b>	
Investigaciones periodísticas		<b>X</b>	
Información sobre Espectáculos	<b>X</b>		
Sociales	<b>X</b>		
Chistes, Horóscopos y entretenimiento en general	<b>X</b>		
Suplementos y revista dominical	<b>X</b>		

50. Cuál es su ocupación principal?.

1	Desocupado
2	Estudiante
3	Ama de Casa
4	Jubilado/pensionado
5	Empleadores
6	Trabajos no especializados o changarín
7	Comerciante sin personal
8	Técnico
9	Profesionales independientes y otros autónomos
10	Relación de dependencia
11	Obrero, cal. y no cal., capataz
<b>12</b>	Empleado s/jerarquía
<b>13</b>	Jefe intermedio
14	Ns/Nc

51. Cuál es el máximo nivel de estudios cursados por Ud.?

1	Sin instrucción
2	Primario incompleto
3	Primario completo
4	Secundario incompleto
5	Secundario completo
6	Terciario incompleto
<b>7</b>	Terciario completo
8	Universitario incompleto
9	Universitario completo
10	Ns/Nc

52. Cuáles son los ingresos mensuales de su grupo familiar?

1	Menos de 300
2	De 300 a 500
3	De 501 a 700
4	De 701 a 1000
<b>5</b>	De 1001 a 1500
6	De 1501 a 2500
7	De 2501 a 4000
8	Más de 4000
9	Ns/Nc

53. Cuántas personas en el hogar comparten este presupuesto?

1. Una persona
2. Dos personas
3. De 3 a 4 personas
4. De 5 a 6 personas
- 5** 7 o más personas
6. Ns/Nc

54. Ud tiene...?

	SI	NO
Computadora	<b>X</b>	
Televisor	<b>X</b>	
Horno a microondas		<b>X</b>
Radio	<b>X</b>	
Auto	<b>X</b>	
Video	<b>X</b>	
Vivienda	<b>X</b>	

55. Antes de iniciar la próxima entrevista el encuestador deberá observar el nivel de vivienda

<b>1</b>	Media Superior
2	Media media
3	Media baja
4	Muy modesta o Marginal

Extracto de la encuesta realizada por  
un diario regional de alta circulación

ENCUESTA DE OPINIÓN  
QUIEN LEE QUÉ?

Domicilio: \_\_\_\_\_ N° de encuesta: 502

1. Sexo:  H  M      2. Edad: 35      49. Es Ud. jefe de familia?  SI  NO

15. De las secciones que tiene el diario cuál lee, cuál no lee y cuál no conoce? (L)

Tipo de noticia	LEE	NO LEE	Ns/Nc
Noticias sobre la realidad política, económica y general del ámbito internacional	<input checked="" type="checkbox"/>		
Clasificados	<input checked="" type="checkbox"/>		
Noticias sobre sexo, violencia, droga,	<input checked="" type="checkbox"/>		
Noticias policiales		<input checked="" type="checkbox"/>	
Noticias nacionales económicas		<input checked="" type="checkbox"/>	
Noticias nacionales políticas	<input checked="" type="checkbox"/>		
Información sobre la corrupción política y social	<input checked="" type="checkbox"/>		
Deportes	<input checked="" type="checkbox"/>		
Comentarios Científicos		<input checked="" type="checkbox"/>	
Información Gremial	<input checked="" type="checkbox"/>		
Notas Editoriales	<input checked="" type="checkbox"/>		
Comentario firmados por analistas económicos y políticos	<input checked="" type="checkbox"/>		
Comentarios firmados por analistas locales sobre aspectos de la realidad regional	<input checked="" type="checkbox"/>		
Noticias y comentarios Agropecuarios	<input checked="" type="checkbox"/>		
Noticias regionales	<input checked="" type="checkbox"/>		
Investigaciones periodísticas	<input checked="" type="checkbox"/>		
Información sobre Espectáculos	<input checked="" type="checkbox"/>		
Sociales	<input checked="" type="checkbox"/>		
Chistes, Horóscopos y entretenimiento en general	<input checked="" type="checkbox"/>		
Suplementos y revista dominical	<input checked="" type="checkbox"/>		

50.Cuál es su ocupación principal?.

1	Desocupado
2	Estudiante
3	Ama de Casa
4	Jubilado/pensionado
5	Empleadores
6	Trabajos no especializados o changarín
7	Comerciante sin personal
8	Técnico
9	Profesionales independientes y otros autónomos
10	Relación de dependencia
11	Obrero, cal. y no cal., capataz
12	Empleado s/jerarquía
<input checked="" type="radio"/> 13	Jefe intermedio
14	Ns/Nc

51.Cuál es el máximo nivel de estudios cursados por Ud.?

1	Sin instrucción
2	Primario incompleto
3	Primario completo
4	Secundario incompleto
<input checked="" type="radio"/> 5	Secundario completo
6	Terciario incompleto
7	Terciario completo
8	Universitario incompleto
9	Universitario completo
10	Ns/Nc

52. Cuáles son los ingresos mensuales de su grupo familiar?

1	Menos de 300
<input checked="" type="radio"/> 2	De 300 a 500
3	De 501 a 700
4	De 701 a 1000
5	De 1001 a 1500
6	De 1501 a 2500
7	De 2501 a 4000
8	Más de 4000
9	Ns/Nc

53. Cuántas personas en el hogar comparten este presupuesto?

1. Una persona
2. Dos personas
3. De 3 a 4 personas
4. De 5 a 6 personas
5. 7 o más personas
6. Ns/Nc

54. Ud tiene...?

	SI	NO
Computadora		<input checked="" type="checkbox"/>
Televisor	<input checked="" type="checkbox"/>	
Horno a microondas		<input checked="" type="checkbox"/>
Radio	<input checked="" type="checkbox"/>	
Auto	<input checked="" type="checkbox"/>	
Video	<input checked="" type="checkbox"/>	
Vivienda	<input checked="" type="checkbox"/>	

55. Antes de iniciar la próxima entrevista el encuestador deberá observar el nivel de vivienda

1	Media Superior
<input checked="" type="radio"/> 2	Media media
3	Media baja
4	Muy modesta o Marginal



### **Anexo B: Razas de perros**

Raza Basset: tamaño pequeño, peso bajo; velocidad, inteligencia, afectividad y agresividad leves. Se los utiliza para caza
Raza Beauceron: tamaño grande, peso mediano, gran velocidad; inteligencia, afectividad y agresividad mediana. Su función es la utilidad
Raza Boxer: tamaño, peso mediano, velocidad e inteligencia mediana; afectividad y agresividad grande. Son útiles para compañía
Raza Bul-dog: tamaño y peso pequeño; velocidad leve; inteligencia mediana y gran afectividad; poca agresividad. Son útiles para compañía
Raza Bull-massif: tamaño y peso grande; velocidad leve; gran inteligencia; poca afectividad; muy agresivos. Su función es la utilidad
Raza Caniche: tamaño y peso pequeño; velocidad mediana; gran inteligencia y afectividad; leve agresividad. Son útiles para compañía
Raza Chiguagua: tamaño y peso pequeño; velocidad e inteligencia leve; gran afectividad y leve agresividad. Son útiles para compañía
Raza Cocker: tamaño mediano; peso pequeño; velocidad leve; inteligencia mediana, gran afectividad; muy agresivos. Son útiles para compañía
Raza Colley: gran tamaño; mediano peso; gran velocidad; inteligencia mediana; afectividad y agresividad leves. Son útiles para compañía
Raza Dálmata: tamaño, peso, velocidad e inteligencia mediana; gran afectividad y poca agresividad. Son útiles para compañía
Raza Doberman: gran tamaño; peso mediano; gran velocidad e inteligencia; leve afectividad; gran agresividad. Su función es la utilidad
Raza Dogo alemán: gran tamaño, peso y velocidad; inteligencia y afectividad leve; gran agresividad. Su función es la utilidad
Raza Fox-hound: tamaño grande; peso mediano; velocidad grande; inteligencia y afectividad leve; gran agresividad. Son útiles para caza
Raza Fox-terrier: tamaño y peso pequeño; velocidad e inteligencia mediana; gran afectividad y agresividad. Son útiles para compañía
Raza Galgo: tamaño grande; peso mediano; velocidad grande; inteligencia, afectividad y agresividad leve. Son útiles para caza
Raza Gascoño: tamaño grande; peso y velocidad mediana; inteligencia y afectividad leve; gran agresividad. Son útiles para caza
Raza Labrador: tamaño, peso, velocidad e inteligencia mediana; gran afectividad; poca agresividad. Son útiles para caza
Raza Mastín alemán: gran tamaño; peso mediano; velocidad e inteligencia grande; gran afectividad y agresividad. Su función es la utilidad
Raza Mastiff: tamaño y peso grande; velocidad, inteligencia y afectividad leve; gran agresividad. Su función es la utilidad
Raza Pekinés: tamaño y peso pequeño; velocidad e inteligencia leve; gran afectividad; poca agresividad. Son útiles para compañía
Raza Podenco Bretón: tamaño, peso y velocidad mediana; gran inteligencia y afectividad; poca agresividad. Son útiles para caza
Raza Podenco Francés: tamaño grande; peso, velocidad e inteligencia mediana; afectividad y agresividad leve. Son útiles para caza
Raza Pointer: tamaño gran; peso mediano; gran velocidad e inteligencia; poca afectividad y agresividad. Son útiles para caza
Raza Setter: gran tamaño; peso mediano; gran velocidad; inteligencia mediana; leve afectividad y agresividad. Son útiles para caza
Raza San Bernardo: tamaño y peso grande; velocidad leve; inteligencia mediana; leve afectividad; poca agresividad. Su función es la utilidad
Raza Teckel: tamaño y peso pequeño; velocidad leve; inteligencia mediana y gran afectividad; poca agresividad. Son útiles para compañía
Raza Terranova: tamaño y peso grande; velocidad leve; inteligencia mediana; leve afectividad y agresividad. Su función es la utilidad

Capítulo 7. TEORIA DE LA ESTIMACION	249
7.1 Estimación puntual	249
7.2 Método de Máxima verosimilitud	251
7.3 Estimación por Intervalo	253
Intervalo de confianza cuando $\hat{\theta}$ se distribuye normal o aproximadamente normal	254
Intervalo de confianza para estimar $\mu$	255
Intervalo de confianza para la varianza de una población normal	258
Intervalo de confianza para la diferencia de medias	261
Caso 1. $\sigma_1^2$ y $\sigma_2^2$ conocidas,	261
Caso 2. $\sigma_1^2$ y $\sigma_2^2$ desconocidas y muestras grandes.	262
Caso 3. $\sigma_1^2$ y $\sigma_2^2$ desconocidas pero iguales y muestras pequeñas.	263
Caso 4. $\sigma_1^2$ y $\sigma_2^2$ desconocidas y distintas en muestras pequeñas.	264
Intervalo de confianza para el cociente de varianzas poblacionales	266
Intervalo de confianza para estimar la proporción de ocurrencia de un evento en la población	268
Intervalo de confianza para la igualdad de proporciones poblacionales	269
CASOS DE ESTUDIO, PREGUNTAS Y PROBLEMAS	270
Caso 7.1: Edades y Estaturas	270

	248
Caso 7.2: Actividad Económica en Río Cuarto	270
Caso 7.3: Demanda de diario regional	271
Problemas	271
Bibliografía	272

## Capítulo 7. TEORIA DE LA ESTIMACION

### 7.1 Estimación puntual

En la discusión que sigue el tema de interés es la estimación de parámetros (generalmente un solo parámetro) de distribuciones de probabilidad de forma conocida, es decir casos en que la forma de la función de densidad  $f(x, \theta)$  es conocida.

Todas las estimaciones se hacen con base en la información contenida en la muestra. Esta muestra tendrá siempre un tamaño fijo, es decir no se tendrá en cuenta en el análisis de los métodos de estimación en que el tamaño de la muestra es también una variable aleatoria, que depende de los mismos datos de los parámetros a medida que van apareciendo.

Los estimadores puntuales son requeridos, en particular, cuando la misma estimación debe insertarse como un número en una estructura matemática más compleja, como es común en la economía moderna.

En la estimación puntual se construye una sola función de los datos de la muestra para estimar el valor de un parámetro poblacional.

Dada  $f(\bar{x}, \theta)$  conocida, existe una muestra ( $n$ ) de una población con distribución conocida y el tamaño de la muestra es fijo.

Los estimadores puntuales tienen las propiedades de: insesgamiento, invarianza, consistencia y eficiencia o suficiencia.

Insesgamiento. Esta primera propiedad fue analizada cuando se desarrolló la distribución muestral de la media aritmética de la muestra, donde, empíricamente, se demostró que

$$E(\bar{X}) \cong \mu$$

siendo  $\mu$  (parámetro poblacional) la media de la población.

Generalizando, hay una población descrita por la función de densidad  $f(\bar{x}, \theta)$ , donde  $f$  es conocida y el valor del parámetro  $\theta$  es desconocido.

Sea  $x_1, x_2, \dots, x_n$  una muestra aleatoria obtenida de esta población, el estadístico  $t(x_1, x_2, \dots, x_n)$  es un estimador insesgado del parámetro  $\theta$  si  $E(t) = \theta$  para todo  $n$  y para cualquier valor de  $\theta$ . En cualquier otro caso decimos que  $\theta$  es sesgado.

**Ejemplo.** Estimadores insesgados

a)  $E(X_1) = \mu, E(\bar{X}) = \mu$

Si  $\mu$  existe, tanto una sola observación de la variable  $X$ ,  $X_1$ , como la media,  $\bar{X}$ , de una variable aleatoria de  $n$  observaciones de una población de media  $\mu$  son estimadores insesgados de  $\mu$ . Sus varianzas respectivas son  $\sigma^2$  y  $\sigma^2/n$ . Por lo tanto es preferible  $\bar{X}$  a  $X_1$ , siempre y cuando  $n > 1$ , ya que  $(\sigma^2/n) < \sigma^2$ .

$X_1$  es una observación muestral que puede asumir distintos valores ( $n$  valores) de la población. De donde

$$E(X_1) = X_1P(x_1 = X_1) + X_2P(x_1 = X_2) + \dots + X_nP(x_1 = X_n)$$

b)  $E(\bar{x} - \bar{y}) = E(x) - E(y)$

Esto es, la diferencia de las medias de dos muestras aleatorias independientes es un estimador insesgado de la diferencia de las medias de la población.

c)  $E(S^2) = \sigma^2$

Dado  $\hat{\sigma}^2 = \frac{\sum (x_i - \bar{x})^2}{n}$

$$\begin{aligned} E(\hat{\sigma}^2) &= \frac{1}{n} E\left\{ \sum [x_i - E(x) + E(x) - \bar{x}]^2 \right\} \\ &= \frac{1}{n} \sum \left\{ E[x_i - E(x)]^2 - E[\bar{x} - E(x)]^2 \right\} \end{aligned}$$

donde,  $E[x_i - E(x)]^2 = \sigma^2$  y  $E[\bar{x} - E(x)]^2 = \sigma_{\bar{x}}^2 = \frac{\sigma^2}{n}$

Reemplazando

$$E(\hat{\sigma}^2) = \frac{1}{n} \left\{ n\sigma^2 - n \frac{\sigma^2}{n} \right\} = \sigma^2 - \frac{\sigma^2}{n} = \frac{n\sigma^2 - \sigma^2}{n} = \frac{n-1}{n} \sigma^2$$

Por lo tanto,  $\hat{\sigma}^2$  es un estimador sesgado de  $\sigma^2$ . Pero  $S^2$  es insesgado ya que

$$s^2 = \frac{\sum (x_i - \bar{x})^2}{n-1} = \frac{n}{n-1} \frac{\sum (x_i - \bar{x})^2}{n} = \frac{n}{n-1} \hat{\sigma}^2$$

$$E(s^2) = \frac{n}{n-1} E(\hat{\sigma}^2) = \frac{n}{n-1} \frac{n-1}{n} \sigma^2$$

Por lo tanto  $E(s^2) = \sigma^2$

Invarianza Si  $\hat{\theta}$  es un estimador de  $\theta$ , podemos pensar que  $\hat{\theta}^2$  es un estimador de  $\theta^2$ . Por lo tanto un estimador es invariante cuando  $g(\hat{\theta})$  es un estimador de  $g(\theta)$ . La  $\bar{X}$  no es un estimador invariante de  $\mu$  ya que

$$E(\bar{X}^2) = \mu^2 + \frac{\sigma^2}{n}$$

**Observación.** Sabemos que

$$\begin{aligned} V(X) &= m_2 - m_1^2 \\ &= E(X^2) - [E(X)]^2 \\ V(\bar{X}) &= E(\bar{X}^2) - [E(\bar{X})]^2 \\ \frac{\sigma^2}{n} &= E(\bar{X}^2) - \mu^2 \\ E(\bar{X}^2) &= \mu^2 + \frac{\sigma^2}{n} \end{aligned}$$

Consistencia.  $W$  es un estimador consistente de  $\theta$  si para números pequeños arbitrarios  $a$  y  $b$ , existe un valor  $n_1$  de  $n$  tal que si  $n > n_1$ ,

$$p\{|W - \theta| < a\} > 1 - b$$

En breve, cuando  $n \rightarrow \infty$ ,  $W \rightarrow \theta$ .

**Ejemplo.**  $\bar{X}$  es un estimador consistente de  $\mu$  ya que dado  $E(\bar{X}) = \mu$ ,

$$\sigma_{\bar{X}}^2 = \frac{\sigma^2}{n}. \text{ Si } n \rightarrow \infty, \frac{\sigma^2}{n} \rightarrow 0 \text{ por lo tanto } \sigma_{\bar{X}}^2 \rightarrow \frac{\sigma^2}{n} \text{ y } \bar{X} \rightarrow \mu.$$

Suficiencia. Es un estimador que utiliza toda la información contenida en la muestra, respecto del parámetro a estimar.

## 7.2 Método de Máxima verosimilitud

Uno de los procedimientos para obtener buenos estimadores es el denominado método de máxima verosimilitud. La función de verosimilitud es la función de densidad conjunta de las  $n$  observaciones muestrales, condicionada por el, o por los, parámetros a estimar.

$$f(x_1, x_2, x_3, \dots, x_n / \theta) = \prod_{i=1}^n f(x_i / \theta)$$

Se puede notar por  $L(x_1, x_2, x_3, \dots, x_n / \theta)$  o  $L(\theta)$

Para obtener los estimadores maximoverosímiles se debe:

1°. Definir la función de verosimilitud  $L(\theta_1, \theta_2, \dots, \theta_k)$

2°. Tomar  $\ln$  de  $L$ :  $\ln[L(\theta_1, \theta_2, \dots, \theta_k)]$

3°. Los estimadores de  $\theta_1, \theta_2, \dots, \theta_k$  serán aquellas expresiones para las cuales la función  $L$  sea máxima. Para esto es necesario

a) derivar  $\ln(L)$  respecto a cada  $\theta_i$

b) igualar a cero

c) reemplazar los  $\theta_i$  por los estimadores  $\hat{\theta}_i$

d) resolver el sistema de ecuaciones

**Ejemplo.** Dada una población normal de parámetro  $(\mu, \sigma)$ , la función de verosimilitud para obtener buenos estimadores para la media  $\mu$  y para la varianza  $\sigma^2$ , será

$$f(\mu/\sigma^2) = \prod_{i=1}^n f(x_i) = \prod_{i=1}^n \frac{e^{-\frac{(x_i-\mu)^2}{2\sigma^2}}}{\sigma\sqrt{2\pi}}$$

1° Se define la función de verosimilitud

$$\begin{aligned} L(\mu, \sigma^2) &= \prod_{i=1}^n \frac{1}{\sigma\sqrt{2\pi}} \prod_{i=1}^n e^{-\frac{(x_i-\mu)^2}{2\sigma^2}} = \left(\frac{1}{\sigma\sqrt{2\pi}}\right)^n e^{-\frac{\sum(x_i-\mu)^2}{2\sigma^2}} \\ &= \left(\frac{1}{\sigma^2 2\pi}\right)^{\frac{n}{2}} e^{-\frac{\sum(x_i-\mu)^2}{2\sigma^2}} \end{aligned}$$

2° Se toma  $\ln$

$$\ln L(\mu, \sigma^2) = -\frac{n}{2} \ln(\sigma^2 2\pi) - \frac{\sum(x_i - \mu)^2}{2\sigma^2} = -\frac{n}{2} \ln(\sigma^2) - \frac{n}{2} \ln(2\pi) - \frac{\sum(x_i - \mu)^2}{2\sigma^2}$$

3° Se deriva el  $\ln L(\mu, \sigma^2)$  con respecto a  $\mu$  y  $\sigma^2$ , y se iguala a cero

$$\frac{\partial \ln(L)}{\partial \mu} = -2 \frac{\sum(x_i - \mu)}{2\sigma^2} (-1) = \frac{\sum(x_i - \mu)}{2\sigma^2} = 0$$

$$\frac{\partial \ln(L)}{\partial \sigma^2} = -\frac{n}{2} \frac{1}{\sigma^2} + \frac{\sum(x_i - \mu)^2}{2(\sigma^2)^2} = 0$$

Se reemplaza  $\mu$  por  $\bar{x}$ , y  $\sigma^2$  por  $\hat{\sigma}^2$ , y se forma y resuelve el sistema:

$$\frac{\sum(x_i - \bar{x})^2}{\hat{\sigma}^2} = 0 \quad (1)$$

$$-\frac{n}{2} \frac{1}{\hat{\sigma}^2} + \frac{\sum(x_i - \bar{x})^2}{2(\hat{\sigma}^2)^2} = 0 \quad (2)$$

En (1) se multiplica ambos miembros por  $\hat{\sigma}^2$ , de modo que

$$\sum x_i - n\bar{x} = 0 \Rightarrow \bar{x} = \frac{\sum x_i}{n}$$

En (2) se despeja  $\hat{\sigma}^2$  y se obtiene:

$$\frac{-n\hat{\sigma}^2 + \sum(x_i - \bar{x})^2}{2(\hat{\sigma}^2)^2} = 0 \Rightarrow \sum(x_i - \bar{x})^2 = n\hat{\sigma}^2 \therefore \hat{\sigma}^2 = \frac{\sum(x_i - \bar{x})^2}{n}$$

La media muestral es un Estimador Máximo Verosímil (EMV) de la media poblacional.

El EMV de  $\sigma^2$  es  $\hat{\sigma}^2$  que, por lo demostrado anteriormente, es un estimador sesgado.

### 7.3 Estimación por Intervalo

En la estimación por intervalo, generalmente, se construyen dos funciones basadas en los mismos datos muestrales, dentro de los cuales, en un sentido probabilístico, el valor desconocido del parámetro ha de encontrarse. Esto consiste en obtener un cierto intervalo aleatorio que contiene el parámetro a estimar. Este intervalo aleatorio se denomina intervalo de confianza y se establece su magnitud de la siguiente manera

$$P\{L_i \leq \theta \leq L_s\} = 1 - \alpha$$

Donde:

$\theta$  es el parámetro a estimar

$L_i$  límite inferior del intervalo

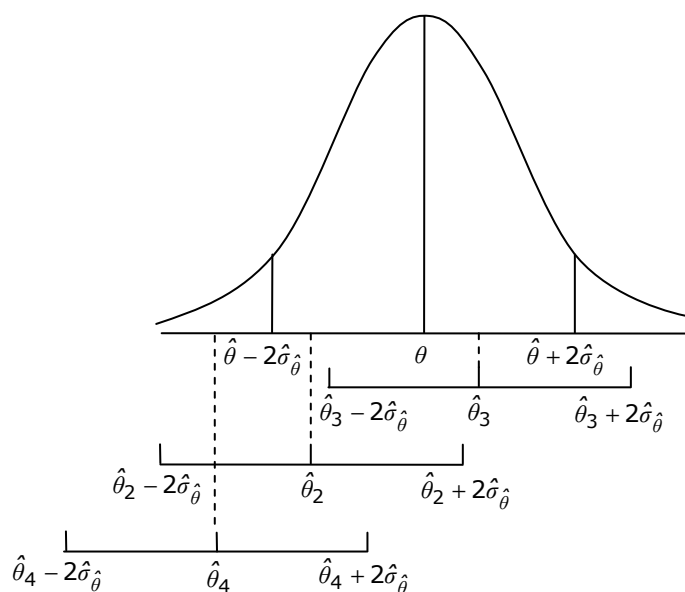
$L_s$  límite superior del intervalo

$1 - \alpha$  nivel de confianza, esto es la probabilidad que de cada 100 veces que se realice la estimación  $(1 - \alpha)100$  veces, el parámetro a estimar va a caer dentro de este intervalo. Es una probabilidad conocida.

Los límites de confianza  $L_i$  y  $L_s$  son variables aleatorias que dependen de las observaciones muestrales y del nivel de confianza  $1 - \alpha$ .

En la estimación puntual se estima  $\theta$ , puntualmente, a través de  $\hat{\theta}$  con un riesgo  $\alpha$  de cometer un error de estimación.

En la estimación por intervalo se estima  $\theta$  a través de un intervalo aleatorio que tiene una probabilidad conocida de contener el parámetro.





Para entender aún más el concepto, suponga que el estimador  $\hat{\theta} \sim N$ , o asintóticamente normal, y que la amplitud del intervalo sea  $\pm 2\sigma_{\hat{\theta}}$ .

Suponga, además, que se hacen varias estimaciones puntuales  $\hat{\theta}_1, \hat{\theta}_2, \dots, \hat{\theta}_n$ . La probabilidad de que este intervalo contenga al parámetro  $\theta$  se obtiene de la tabla normal y es 0,9544. Se observa que los intervalos de confianza contenidos a partir de  $\hat{\theta}_2$  y  $\hat{\theta}_3$  contiene al parámetro, no así el construido a partir de  $\hat{\theta}_4$ .

Por lo tanto el 95,44% de las veces el intervalo de amplitud  $\pm 2\sigma_{\hat{\theta}}$  contendrá al parámetro y en el 4,56% no lo contendrá. Si la amplitud del intervalo fuera  $\pm 2,4$  la probabilidad sería 0,9836.

Elementos de la estimación por intervalos

- 1)  $\theta$ ; parámetro a estimar
- 2)  $x_1, x_2, \dots, x_n$ ;  $n$  observaciones muestrales de una población cuya función de densidad es  $g(x/\theta)$ .
- 3)  $\hat{\theta}$ ; un estimador de  $\theta$  y como tal variable aleatoria,  $\hat{\theta} = \hat{\theta}(x_1, x_2, \dots, x_n)$ .
- 4)  $K = K(\hat{\theta}, \theta)$ ; es un estadístico que por ser función de  $\hat{\theta}$  es una variable aleatoria.  $K(\hat{\theta}, \theta)$  se elige de manera que tenga una función de probabilidad conocida y tabulada.
- 5)  $g(K)$ ; es la función de densidad de la variable  $K(\hat{\theta}, \theta)$ .
- 6)  $1 - \alpha = P\{K_1 \leq K(\hat{\theta}, \theta) \leq K_2\}$ ; es el nivel de confianza.
- 7)  $K_1, K_2$ ; son los coeficientes de confianza y se obtienen de la tabla de la distribución  $g(K)$ . Su valor depende del nivel de confianza  $(1 - \alpha)$ . Los coeficientes de confianza son funciones de  $(1 - \alpha)$ , esto es, al fijar  $(1 - \alpha)$  quedan determinados  $K_1$  y  $K_2$ .
- 8)  $L_i, L_s$ ; son límites del intervalo de confianza para estimar  $\theta$ . Se obtienen al despejar  $\theta$  en la desigualdad  $K_1 \leq K(\hat{\theta}, \theta) \leq K_2$ , de tal manera que  $L_i \leq \theta \leq L_s$ .  $L_i, L_s$  son variables aleatorias que dependen de las observaciones muestrales y de los coeficientes de confianza y son función del nivel de confianza y del tamaño de la muestra.

*Intervalo de confianza cuando  $\hat{\theta}$  se distribuye normal o aproximadamente normal*

Suponga  $\hat{\theta} \sim N(\theta, \sigma_{\hat{\theta}})$ . El estadístico  $K(\hat{\theta}, \theta)$  que se utilizará es  $Z \sim N(0,1)$

$$K(\hat{\theta}, \theta) = \frac{\hat{\theta} - \theta}{\sigma_{\hat{\theta}}} = Z \sim N(0,1)$$

$$P\left\{K_1 \leq \frac{\hat{\theta} - \theta}{\sigma_{\hat{\theta}}} \leq K_2\right\} = 1 - \alpha$$

Multiplicando por  $(-\sigma_{\hat{\theta}})$  y sumando  $\hat{\theta}$ , en los tres miembros, nos queda:

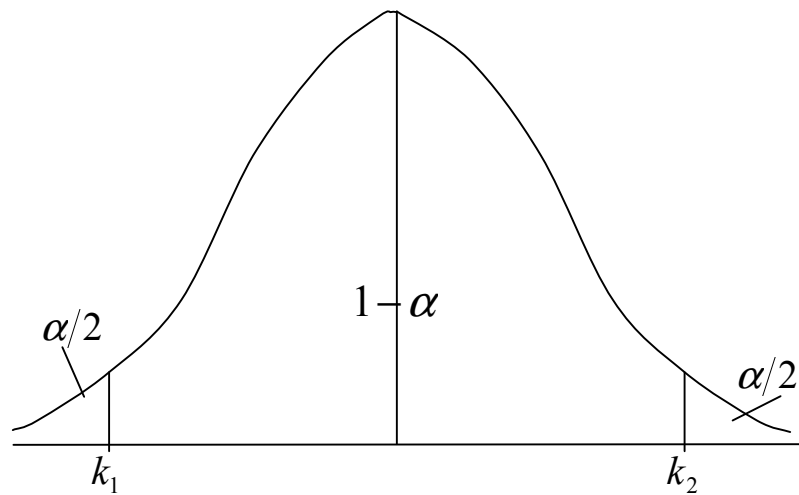
$$-K_1\sigma_{\hat{\theta}} \geq \theta - \hat{\theta} \geq -K_2\sigma_{\hat{\theta}} = \hat{\theta} - K_2\sigma_{\hat{\theta}} \leq \theta \leq \hat{\theta} - K_1\sigma_{\hat{\theta}}$$

los límites de confianza son

$$L_i = \hat{\theta} - K_2\sigma_{\hat{\theta}}; \quad K_2 = z_{1-\alpha/2}$$

$$L_s = \hat{\theta} - K_1\sigma_{\hat{\theta}}; \quad K_1 = z_{\alpha/2}$$

Dada la simetría de la curva normal, en todos los casos se verifica que  $z_{\alpha/2} = -z_{1-\alpha/2}$



o sea que:  $L_i = \hat{\theta} - z_{1-\alpha/2}\sigma_{\hat{\theta}}; \quad L_s = \hat{\theta} + z_{1-\alpha/2}\sigma_{\hat{\theta}}$

*Intervalo de confianza para estimar  $\mu$*

Si la población es normal de parámetro  $(\mu, \sigma)$ ,  $\bar{X}$  se distribuye normal con parámetro  $(\mu, \sigma/\sqrt{n})$

$$\bar{X} \sim N(\mu, \sigma/\sqrt{n})$$

El estadístico es  $K(\bar{X}/\mu) = \frac{\bar{X} - \mu}{\sigma_{\bar{X}}} = \frac{\bar{X} - \mu}{\sigma/\sqrt{n}}$

a) si  $\sigma$  es conocida,

$$K(\bar{X}/\mu) \sim N(0,1)$$

$$L_i = \bar{X} - z_{1-\alpha/2} \sigma / \sqrt{n};$$

$$L_s = \bar{X} + z_{1-\alpha/2} \sigma / \sqrt{n}$$

- b) Si  $\sigma$  es desconocida y la muestra es grande  $S \rightarrow \sigma$  y  $K(\bar{X}/\mu) \rightarrow Z(0,1)$ , pero al ser el desvío desconocido debe estimarse con  $s$

$$s = \sqrt{\frac{\sum (X_i - \bar{X})^2}{n-1}}$$

$$L_i = \bar{X} - z_{1-\alpha/2} s / \sqrt{n};$$

$$L_s = \bar{X} + z_{1-\alpha/2} s / \sqrt{n}$$

- c) Si  $\sigma$  es desconocida y la muestra es pequeña

$$K(\bar{X}, \mu) \sim t$$

$$\text{Si } Y = \frac{\sum_{i=1}^n (X_i - \bar{X})^2}{\sigma^2} \sim \chi_{n-1}^2$$

Multiplicando y dividiendo por  $n-1$ :

$$Y = \frac{\sum_{i=1}^n (X_i - \bar{X})^2}{\sigma^2} \frac{n-1}{n-1} = (n-1) \frac{s^2}{\sigma^2} \sim \chi_{n-1}^2$$

En consecuencia, dada una muestra de tamaño  $n$ , y siendo

$$Z = \frac{\bar{X} - \mu}{\sigma_{\bar{X}}} = \frac{(\bar{X} - \mu)\sqrt{n}}{\sigma} \sim N(0,1)$$

se obtiene

$$t_{n-1} = \frac{Z}{\sqrt{\frac{Y}{n-1}}} = \frac{\frac{(\bar{X} - \mu)\sqrt{n}}{\sigma}}{\sqrt{\frac{(n-1) \frac{s^2}{\sigma^2}}{(n-1)}}} = \frac{(\bar{X} - \mu)\sqrt{n}}{\frac{\sigma}{s}}$$

Esto significa que

$$t_{n-1} = \frac{(\bar{X} - \mu)\sqrt{n}}{s}$$

El nivel de confianza es:  $P\left\{K_1 \leq \frac{(\bar{X} - \mu)\sqrt{n}}{s} \leq K_2\right\} = 1 - \alpha$

donde:

$$K_1 = t_{n-1; \alpha/2}$$

$$K_2 = t_{n-1; 1-\alpha/2}$$

por lo tanto

$$L_i = \bar{X} - t_{n-1; 1-\alpha/2} s / \sqrt{n}$$

$$L_s = \bar{X} - t_{n-1; \alpha/2} s / \sqrt{n}$$

En resumen, la media de la muestra ( $\bar{X}$ ) es usada para estimar la media de la población desconocida ( $\mu$ ). Debido a que  $\bar{X}$  varía de muestra en muestra, no es igual a la media de la población ( $\mu$ ); es decir, existe un error de la muestra. Para reflejar el alcance de este error muestral es útil una estimación de intervalo en torno de  $\bar{X}$ .

El tamaño del intervalo dependerá de la confianza que se desee, de que contenga a la media de la población verdadera y desconocida.

Para tener una confianza del 95% de que la estimación del intervalo contenga a la media de la población verdadera, la estimación de intervalo es:

$$\bar{x} \pm \text{error de la muestra (e)} = \bar{x} \pm z \frac{\sigma}{\sqrt{n}} = \text{estimación de intervalo de } \mu$$

**Ejemplo.** En el estudio del ingreso promedio de los clientes de un banco que poseen la tarjeta AA, se obtuvo

$$\bar{X} = 793.00$$

$$\sigma = 531.20$$

El intervalo es

$$\bar{x} \pm 1.96 \frac{\sigma}{\sqrt{n}} = 793.00 \pm 1.96 \frac{531.20}{\sqrt{10}} = 793.00 \pm 329.24$$

e incluye la media poblacional verdadera ( $\mu = 831.58$ ).

Aproximadamente el 95% de las muestras generarán una estimación de intervalo que incluirá a la verdadera media de la población.

El tamaño de la estimación del intervalo dependerá de cuatro factores.

- ⇒ nivel de confianza.
- ⇒ desviación estándar de la población.
- ⇒ tamaño de la muestra.
- ⇒ conocimiento de la desviación estándar de la población  
 $\sigma_X = \sigma$ .

Si la desviación estándar de la población no es conocida, es necesario estimarla con la desviación estándar de la muestra,  $s$ . Por consiguiente, la estimación del intervalo de 95%, sería:

$$\bar{x} \pm t \frac{s}{\sqrt{n}}$$

Donde, debido a que el tamaño de la muestra es pequeño, se reemplaza la distribución normal con la distribución t.

**Ejemplo.** En el estudio del ingreso promedio de los clientes de un banco que poseen la tarjeta AA, se obtuvo

$$\bar{X} = 793.00$$

$$s = 453.56$$

$$n = 10$$

Entonces el intervalo es:

$$\bar{x} \pm t \frac{s}{\sqrt{n}} = 793.00 \pm 2.26 \frac{453.56}{\sqrt{10}} = 793.00 \pm 324.15$$

A medida que n se vuelva más grande, la distribución de t se aproxima a la distribución normal. Por ejemplo, si  $n=20 \Rightarrow t=2.09$ ;  $n=60 \Rightarrow t=2.00$

### *Intervalo de confianza para la varianza de una población normal*

El parámetro a estimar es ahora  $\sigma^2$  y su estimador es

$$S^2 = \frac{\sum_{i=1}^n (X_i - \bar{X})^2}{n-1}$$

El estadístico que se utiliza es

$$K(S^2/\sigma^2) = n-1 \frac{S^2}{\sigma^2} \sim \chi_{n-1}^2$$

Nivel de confianza:

$$P\left\{K_1 \leq (n-1) \frac{S^2}{\sigma^2} \leq K_2\right\} = 1 - \alpha$$

para despejar  $\sigma^2$  se deben seguir los siguientes pasos

1. se toman recíprocas y se cambia el sentido de la desigualdad entre llaves, esto es:

$$\frac{1}{K_1} \geq \frac{\sigma^2}{(n-1)S^2} \geq \frac{1}{K_2}$$

2. se multiplican los miembros de la desigualdad por  $(n-1)S^2$

$$\frac{(n-1)S^2}{K_2} \leq \sigma^2 \leq \frac{(n-1)S^2}{K_1}$$

$K_1$  y  $K_2$  son valores particulares de  $\chi^2$  que se encuentran en la tabla

$$K_1 = \chi_{n-1; \alpha/2}^2$$

$$K_2 = \chi_{n-1; 1-\alpha/2}^2$$

Los límites del intervalo se definen

$$L_i = \frac{(n-1)S^2}{\chi_{n-1; 1-\alpha/2}^2}$$

$$L_s = \frac{(n-1)S^2}{\chi_{n-1; \alpha/2}^2}$$

De modo que la probabilidad será

$$P\left(\frac{(n-1)S^2}{\chi_{n-1; 1-\alpha/2}^2} \leq \sigma^2 \leq \frac{(n-1)S^2}{\chi_{n-1; \alpha/2}^2}\right) = 1 - \alpha$$

**Ejemplo.** Por requerimiento del Departamento Técnico de la empresa INTI SA se desea estimar un intervalo de confianza para la varianza del llenado de las botellas de una bebida gaseosa a un nivel del 95%. Se conoce que el llenado de las botellas se distribuye aproximadamente normal. Se toma una muestra de 19 botellas al azar con los siguientes resultados.

$$\sum X_i = 185.2dc^3 \quad \sum X_i^2 = 3545.8dc^3$$

El intervalo de confianza para la varianza,  $\sigma^2$ , es:

$$P\left(\frac{(n-1)S^2}{\chi_{n-1;1-\alpha/2}^2} \leq \sigma^2 \leq \frac{(n-1)S^2}{\chi_{n-1;\alpha/2}^2}\right) = 1 - \alpha$$

El enunciado del problema no indica el valor de  $S^2$ , pero ofrece algunos resultados para la variable bajo estudio que pueden ser de utilidad.

$$S^2 = \frac{\sum_{i=1}^n (X_i - \bar{X})^2}{n-1} = \frac{\sum_{i=1}^n X_i^2 - n\bar{X}^2}{n-1}$$

Reemplazando

$$S^2 = \frac{3545.8 - 19\left(\frac{185.2}{19}\right)^2}{18} = 96.70$$

La distribución chi cuadrado con 18 grados de libertad, para un nivel de confianza de 0.95, arroja los siguientes estadísticos:

$$\chi_{n-1;\alpha/2}^2 = \chi_{18;0.025}^2 = 8.23$$

$$\chi_{n-1;1-\alpha/2}^2 = \chi_{18;0.975}^2 = 31.53$$

Reemplazando en el intervalo

$$P\left(\frac{18 \cdot 96.70}{31.53} \leq \sigma^2 \leq \frac{18 \cdot 96.70}{8.23}\right) = 0.95$$

$$P(55.26 \leq \sigma^2 \leq 211.49) = 0.95$$

El resultado indica que la variabilidad en el llenado de las botellas de una bebida gaseosa se encuentra entre  $55.26 dc^3$  y  $211.49 dc^3$

*Intervalo de confianza para la diferencia de medias*

El parámetro a estimar es la diferencia en las medias de dos poblaciones,  $\mu_1 - \mu_2$ . Para realizar la estimación se cuenta con muestras provenientes de ambas poblaciones de las cuales se obtienen las respectivas medias,  $\bar{X}_1$  y  $\bar{X}_2$ .

Si  $\bar{X}_1$  y  $\bar{X}_2$  son independientes, entonces

$$1) E(\bar{X}_1 - \bar{X}_2) = E(\bar{X}_1) - E(\bar{X}_2) = \mu_1 - \mu_2$$

$$2) V(\bar{X}_1 - \bar{X}_2) = V(\bar{X}_1) + V(\bar{X}_2) = \frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}$$

$$3) \text{ Si } \bar{X}_1 \sim N\left(\mu_1, \frac{\sigma_1^2}{n_1}\right) \text{ y } \bar{X}_2 \sim N\left(\mu_2, \frac{\sigma_2^2}{n_2}\right), \text{ entonces}$$

$$\bar{X}_1 - \bar{X}_2 \sim N\left(\mu_1 - \mu_2; \sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}\right)$$

Al igual que en el caso de intervalos para la media,  $\mu$ , se deben considerar tres casos:

*Caso 1.  $\sigma_1^2$  y  $\sigma_2^2$  conocidas,*

Se utiliza un estadístico  $Z \sim N(0,1)$

$$Z = \frac{\bar{X}_1 - \bar{X}_2 - (\mu_1 - \mu_2)}{\sigma_{\bar{X}_1 - \bar{X}_2}} = \frac{\bar{X}_1 - \bar{X}_2 - (\mu_1 - \mu_2)}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}}$$

De acuerdo a lo desarrollado oportunamente, los límites para estimadores que se distribuyen normal son:

$$L_i = \hat{\theta} - Z_{1-\frac{\alpha}{2}} \sigma_{\hat{\theta}}$$

$$L_s = \hat{\theta} + Z_{1-\frac{\alpha}{2}} \sigma_{\hat{\theta}}$$

En consecuencia, para la diferencia de dos medias de poblaciones con varianzas conocidas,

$$P(Z_1 \leq Z \leq Z_2) = 1 - \alpha$$



$$P\left(Z_1 \leq \frac{\bar{X}_1 - \bar{X}_2 - (\mu_1 - \mu_2)}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}} \leq Z_2\right) = 1 - \alpha$$

$$P\left((\bar{X}_1 - \bar{X}_2) - Z_{1-\frac{\alpha}{2}} \sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}} \leq \mu_1 - \mu_2 \leq (\bar{X}_1 - \bar{X}_2) + Z_{1-\frac{\alpha}{2}} \sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}\right) = 1 - \alpha$$

De modo que los límites serán:

$$L_i = \bar{X}_1 - \bar{X}_2 - Z_{1-\frac{\alpha}{2}} \sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}$$

$$L_s = \bar{X}_1 - \bar{X}_2 + Z_{1-\frac{\alpha}{2}} \sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}$$

Caso 2.  $\sigma_1^2$  y  $\sigma_2^2$  desconocidas y muestras grandes.

Los estimadores de las varianzas poblacionales,  $s_1^2$  y  $s_2^2$ , tienden a  $\sigma_1^2$  y  $\sigma_2^2$ , respectivamente, por lo que se utiliza la distribución de probabilidad Normal tipificada. Entonces:

$$Z = \frac{\bar{X}_1 - \bar{X}_2 - (\mu_1 - \mu_2)}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}}$$

El intervalo de confianza se plantea

$$P\left((\bar{X}_1 - \bar{X}_2) - Z_{1-\frac{\alpha}{2}} \sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}} \leq \mu_1 - \mu_2 \leq (\bar{X}_1 - \bar{X}_2) + Z_{1-\frac{\alpha}{2}} \sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}\right) = 1 - \alpha$$

De modo que los límites serán

$$L_i = \bar{X}_1 - \bar{X}_2 - Z_{1-\frac{\alpha}{2}} \sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}$$

$$L_S = \bar{X}_1 - \bar{X}_2 + Z_{1-\frac{\alpha}{2}} \sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}$$

Caso 3.  $\sigma_1^2$  y  $\sigma_2^2$  desconocidas pero iguales y muestras pequeñas.  
En este caso se debe utilizar la distribución de probabilidad t.

$$t_{n_1+n_2-2} = \frac{Z(0,1)}{\sqrt{\frac{\chi_{n_1+n_2-2}^2}{n_1+n_2-2}}}$$

Hay que tener en cuenta que la suma de variables chi cuadrado dan por resultado otra chi cuadrado.

$$\chi_{n_1+n_2-2}^2 = \chi_{n_1-1}^2 + \chi_{n_2-1}^2$$

Utilizando el estadístico definido en el cálculo del intervalo de confianza para la varianza, se tiene:

$$\chi_{n_1+n_2-2}^2 = (n_1-1) \frac{s_1^2}{\sigma^2} + (n_2-1) \frac{s_2^2}{\sigma^2} = \frac{(n_1-1)s_1^2 + (n_2-1)s_2^2}{\sigma^2} \sim \chi_{n_1+n_2-2}^2$$

Esto significa que

$$t_{n_1+n_2-2} = \frac{\frac{\bar{X}_1 - \bar{X}_2 - (\mu_1 - \mu_2)}{\sqrt{\frac{\sigma^2}{n_1} + \frac{\sigma^2}{n_2}}}}{\sqrt{\frac{(n_1-1)s_1^2 + (n_2-1)s_2^2}{\sigma^2}}}$$

multiplicándolo por  $\sqrt{\sigma^2}$  queda

$$t_{n_1+n_2-2} = \frac{\frac{\bar{X}_1 - \bar{X}_2 - (\mu_1 - \mu_2)}{\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}}}{\sqrt{\frac{(n_1-1)s_1^2 + (n_2-1)s_2^2}{n_1+n_2-2}}}$$

Multiplicando extremos por medios

$$t_{n_1+n_2-2} = \frac{\bar{X}_1 - \bar{X}_2 - (\mu_1 - \mu_2)}{\sqrt{\frac{n_1+n_2}{n_1 n_2} \frac{(n_1-1)s_1^2 + (n_2-1)s_2^2}{n_1+n_2-2}}}$$

de donde

$$L_i = \bar{X}_1 - \bar{X}_2 - t_{n_1+n_2-2; 1-\frac{\alpha}{2}} \sqrt{\frac{n_1+n_2}{n_1 n_2} \frac{(n_1-1)s_1^2 + (n_2-1)s_2^2}{n_1+n_2-2}}$$

$$L_s = \bar{X}_1 - \bar{X}_2 + t_{n_1+n_2-2; 1-\frac{\alpha}{2}} \sqrt{\frac{n_1+n_2}{n_1 n_2} \frac{(n_1-1)s_1^2 + (n_2-1)s_2^2}{n_1+n_2-2}}$$

De modo que la probabilidad será

$$P\left( (\bar{X}_1 - \bar{X}_2) - t_{n_1+n_2-2; 1-\frac{\alpha}{2}} \sqrt{\frac{n_1+n_2}{n_1 n_2} \frac{(n_1-1)s_1^2 + (n_2-1)s_2^2}{n_1+n_2-2}} \leq \mu_1 - \mu_2 \leq \right. \\ \left. (\bar{X}_1 - \bar{X}_2) + t_{n_1+n_2-2; 1-\frac{\alpha}{2}} \sqrt{\frac{n_1+n_2}{n_1 n_2} \frac{(n_1-1)s_1^2 + (n_2-1)s_2^2}{n_1+n_2-2}} \right) = 1 - \alpha$$

Para cualquiera de los casos, si el intervalo de confianza contiene el cero, las medias son iguales; y, cuando no lo contiene, son distintas.

Caso 4.  $\sigma_1^2$  y  $\sigma_2^2$  desconocidas y distintas en muestras pequeñas.

Si  $\sigma_1^2$  y  $\sigma_2^2$  son desconocidas y distintas, la distribución de

$$K = \frac{\bar{X}_1 - \bar{X}_2 - (\mu_1 - \mu_2)}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}}$$

Depende de  $\sigma_1^2/\sigma_2^2$  y, en principio, K no puede utilizarse para el cálculo del intervalo. En su lugar existen diferentes soluciones al problema, una de ellas debida a Hsu consiste en aproximar la distribución por una t con grados de libertad fijadas en el mínimo  $(n_1, n_2) - 1$  obteniéndose el intervalo

$$P\left(\bar{X}_1 - \bar{X}_2 - k_{2, (n_1, n_2) - 1} \sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}} \leq \mu_1 - \mu_2 \leq \bar{X}_1 - \bar{X}_2 + k_{2, (n_1, n_2) - 1} \sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}\right) = 1 - \alpha$$

**Ejemplo.** Para un determinado producto de consumo popular, el promedio de ventas en una muestra de 10 comercios fue de \$342500 con un desvío estimado de \$20000. Para un segundo producto el promedio de ventas en una muestra de 12 comercios fue de \$325000 con un desvío estimado de \$17500.

Se supone que los montos de las ventas por tienda tienen una distribución normal para ambos productos. La Cámara Empresaria quiere saber, a un nivel de confianza de 0.95, si hay diferencias significativas en el nivel promedio de ventas por comercio.

La información brindada en el enunciado puede ordenarse de la siguiente manera

		Producto 1	Producto 2
Ventas promedio	$\bar{X}$	342500	325000
Variabilidad	$S$	20000	17500
Tamaño de la muestra	$n$	10	12

Dado que no se conoce el desvío en la población, y la muestra es pequeña, debe aplicarse el tercer caso:

$$P\left((\bar{X}_1 - \bar{X}_2) - t_{n_1+n_2-2; 1-\frac{\alpha}{2}} \sqrt{\frac{n_1+n_2}{n_1 n_2} \frac{(n_1-1)s_1^2 + (n_2-1)s_2^2}{n_1+n_2-2}} \leq \mu_1 - \mu_2 \leq (\bar{X}_1 - \bar{X}_2) + t_{n_1+n_2-2; 1-\frac{\alpha}{2}} \sqrt{\frac{n_1+n_2}{n_1 n_2} \frac{(n_1-1)s_1^2 + (n_2-1)s_2^2}{n_1+n_2-2}}\right) = 1 - \alpha$$

La distribución t, al nivel de confianza de 0.95 y con 20 grados de libertad, tiene los siguientes estadísticos:

$$t_{n_1+n_2-2; 1-\alpha/2} = t_{10+12-2; 0.975} = t_{20; 0.975} = \pm 2.086$$

El error es igual a:

$$\begin{aligned} e &= t_{n_1+n_2-2; 1-\frac{\alpha}{2}} \sqrt{\frac{n_1+n_2}{n_1 n_2} \frac{(n_1-1)s_1^2 + (n_2-1)s_2^2}{n_1+n_2-2}} \\ &= \pm 2.086 \sqrt{\frac{10+12}{10*12} \frac{9*20000^2 + 11*17500^2}{10+12-2}} \\ &= \pm 2.0860 \sqrt{63880208.3} \\ &= \pm 16672.38 \end{aligned}$$

Reemplazando en el intervalo de confianza

$$P((342500 - 325000) - 16672.38 \leq \mu_1 - \mu_2 \leq (342500 - 325000) + 16672.38) = 1 - \alpha$$

$$P(17500 - 16672.38 \leq \mu_1 - \mu_2 \leq 17500 + 16672.38) = 0.95$$

$$P(827.63 \leq \mu_1 - \mu_2 \leq 34172.38) = 0.95$$

El intervalo de confianza alcanzado indica que el producto 1 tiene un promedio de ventas más elevado que el producto 2.

### *Intervalo de confianza para el cociente de varianzas poblacionales*

Al igual que con la diferencia de medias, el cociente de varianzas se utiliza para comparar dos poblaciones. Si el intervalo de confianza contiene el 1 se considera que las dos poblaciones son iguales y si no lo contiene, son distintas.

El parámetro a estimar es  $\sigma_2^2/\sigma_1^2$  y las poblaciones bajo estudio deben ser normales.

En cada una de las poblaciones se toman muestras, de tamaño  $n_1$  y  $n_2$ , con las cuales se estimará  $\sigma_2^2/\sigma_1^2$  a partir del cálculo de  $s_1^2$  y  $s_2^2$ . El estadístico a utilizar es una F que se obtiene mediante el cociente de dos chi cuadrado con  $n-1$  grados de libertad ( $\chi_{n-1}^2$ ).

En el intervalo de confianza de la varianza el estadístico a utilizar es

$$K = (n-1) \frac{S^2}{\sigma^2} \sim \chi_{n-1}^2$$

En el intervalo de confianza del cociente de varianzas, el estadístico a utilizar es una F

$$F_{n_1-1; n_2-1} = \frac{(n_1-1) \frac{S_1^2}{\sigma_1^2}}{(n_2-1) \frac{S_2^2}{\sigma_2^2}} = \frac{S_1^2 \sigma_2^2}{S_2^2 \sigma_1^2}$$

El intervalo de confianza será

$$P\left(K_1 \leq \frac{S_1^2 \sigma_2^2}{S_2^2 \sigma_1^2} \leq K_2\right) = 1 - \alpha$$

de donde resulta

$$P\left(K_1 \frac{S_2^2}{S_1^2} \leq \frac{\sigma_2^2}{\sigma_1^2} \leq K_2 \frac{S_2^2}{S_1^2}\right) = 1 - \alpha$$

$K_1$  y  $K_2$  se encuentran en la tabla F

$$K_1 = F_{n_1-1; n_2-1; \alpha/2}$$

$$K_2 = F_{n_1-1; n_2-1; 1-\alpha/2}$$

Siendo los límites

$$L_i = F_{n_1-1; n_2-1; \alpha/2} \frac{S_2^2}{S_1^2}$$

$$L_s = F_{n_1-1; n_2-1; 1-\alpha/2} \frac{S_2^2}{S_1^2}$$

De modo que la probabilidad

$$P\left(F_{n_1-1; n_2-1; \alpha/2} \frac{S_2^2}{S_1^2} \leq \frac{\sigma_2^2}{\sigma_1^2} \leq F_{n_1-1; n_2-1; 1-\alpha/2} \frac{S_2^2}{S_1^2}\right) = 1 - \alpha$$

**Ejemplo.** Se utilizarán los datos del ejemplo de intervalo de confianza para las diferencias de medias.

		Producto 1	Producto 2
Ventas promedio	$\bar{X}$	342500	325000
Variabilidad	$S$	20000	17500
Tamaño de la muestra	$n$	10	12

Para calcular el intervalo de confianza de la diferencia de varianzas en las ventas promedios de dos productos se utiliza el intervalo

$$P\left(F_{n_1-1; n_2-1; \alpha/2} \frac{S_2^2}{S_1^2} \leq \frac{\sigma_2^2}{\sigma_1^2} \leq F_{n_1-1; n_2-1; 1-\alpha/2} \frac{S_2^2}{S_1^2}\right) = 1 - \alpha$$

Los estadísticos de la distribución  $F$  para 9 grados de libertad en el numerador y 11 grados de libertad en el denominador, para un nivel de confianza de 0.95, son:

$$F_{n_1-1; n_2-1; \alpha/2} = F_{9, 11; 0.025} = \frac{1}{F_{11, 9; 0.975}} = \frac{1}{3.59}$$

$$F_{n_1-1; n_2-1; 1-\alpha/2} = F_{9, 11; 0.975} = 3.59$$

Reemplazando en el intervalo de confianza

$$P\left(\frac{1}{3.96} \frac{17500^2}{20000^2} \leq \frac{\sigma_2^2}{\sigma_1^2} \leq 3.78 \frac{17500^2}{20000^2}\right) = 0.95$$

$$P\left(0.1933 \leq \frac{\sigma_2^2}{\sigma_1^2} \leq 2.8941\right) = 0.95$$

Esto significa que las varianzas son iguales

*Intervalo de confianza para estimar la proporción de ocurrencia de un evento en la población*

Se utiliza para conocer con qué frecuencia se presenta un evento en la población. El parámetro  $\theta = \pi$  y el estimador  $\hat{\theta} = p$ .

El estadístico  $k = k(p/\pi)$  tiene asociado una función de densidad  $g(k)$  conocida de igual manera que con la estimación de la media poblacional

$$k = \frac{p - \pi}{\sqrt{\frac{p(1-p)}{n}}}$$

Generalmente no se conoce el desvío en la población por lo que se estima a partir de  $\sqrt{p(1-p)/n}$

$$P\left(k_1 \leq \frac{p - \pi}{\sqrt{\frac{p(1-p)}{n}}} \leq k_2\right) = 1 - \alpha$$

$$P\left(k_1 \sqrt{\frac{p(1-p)}{n}} \leq p - \pi \leq k_2 \sqrt{\frac{p(1-p)}{n}}\right) = 1 - \alpha$$

$$P\left(p - k_2 \sqrt{\frac{p(1-p)}{n}} \leq \pi \leq p + k_1 \sqrt{\frac{p(1-p)}{n}}\right) = 1 - \alpha$$

Si  $n$  es grande ( $n > 30$ )  $k \sim N(0,1)$

Si  $n$  es pequeña ( $n \leq 30$ )  $k \sim t_{n-1}$

*Intervalo de confianza para la igualdad de proporciones poblacionales*

El parámetro  $\theta = \pi_1 = \pi_2 \rightarrow \pi_1 - \pi_2 = 0$

Estimador  $\hat{\theta} = p_1 = p_2$

Para cada una de las muestras tendremos

$$E(p_1) = \pi_1 \quad V(p_1) = \frac{p_1(1-p_1)}{n_1}$$

$$E(p_2) = \pi_2 \quad V(p_2) = \frac{p_2(1-p_2)}{n_2}$$

Esto significa que

$$E(p_1 - p_2) = E(p_1) - E(p_2) = \pi_1 - \pi_2$$

$$V(p_1 - p_2) = V(p_1) + V(p_2) = \frac{p_1(1-p_1)}{n_1} + \frac{p_2(1-p_2)}{n_2}$$

Si  $n_1$  y  $n_2$  son mayores que 30  $(p_1 - p_2) \sim N(\pi_1 - \pi_2; \sigma_{\pi_1 - \pi_2})$

La variable estandarizada será

$$k = \frac{p_1 - p_2 - (\pi_1 - \pi_2)}{\sqrt{\frac{p_1(1-p_1)}{n_1} + \frac{p_2(1-p_2)}{n_2}}}$$

$$P\left( (p_1 - p_2) - k_2 \sqrt{\frac{p_1(1-p_1)}{n_1} + \frac{p_2(1-p_2)}{n_2}} \leq \pi_1 - \pi_2 \leq (p_1 - p_2) + k_1 \sqrt{\frac{p_1(1-p_1)}{n_1} + \frac{p_2(1-p_2)}{n_2}} \right) = 1 - \alpha$$



## **CASOS DE ESTUDIO, PREGUNTAS Y PROBLEMAS**

### Caso 7.1: Edades y Estaturas

Mediciones realizadas en cursos de Inferencia Estadística informaron las edades y estaturas de los integrantes. La tabla reúne la información del tamaño de muestra y las medidas promedio

Inferencia Estadística	Curso año 1998 n=12		Curso año 2007 n=12	
	Edad	Estatura	Edad	Estatura
Media	20.3750	1.73125	21.8330	1.70500
Desvío	2.2500	0.09258	4.8586	0.09719

Obtiene para cada curso, con un nivel de significación del 0.05,

- el intervalo de confianza para la media
- el intervalo de confianza para el desvío

¿Hay similitudes entre las dos poblaciones? ¿Porqué sí? O ¿porqué no? Indique si es necesario obtener información adicional

### Caso 7.2: Actividad Económica en Río Cuarto

La actividad del comercio minorista, excluido los sectores de alimentos y bebidas y textil, se desarrollaba en el 18, 27% de las empresas de la ciudad. Una encuesta realizada por muestreo estratificado, entre 1998 y 2004, permitió monitorear la evolución de las ventas del sector; esta información se reúne en la tabla.

Los estratos se definieron en función de la venta mensual que realizara la empresa

- Hasta \$100000
- Más de \$100000

Evolución de ventas del sector comercio minorista								
Fecha	Empresas con ventas hasta \$100.000				Empresas con ventas mayores a \$100.000			
	Media	Desvío	n	N	Media	Desvío	n	N
Julio 98	10290,33	7316,15	11	1014	459803,80	428747,81	20	40
Julio 99	9235,69	7752,72	11	1023	467755,06	400134,52	20	40
Julio 00	10470,89	7211,74	11	1059	437535,36	395982,72	20	40
Julio 01	6833,18	5696,52	11	1078	244791,27	230112,77	20	40
Julio 02	16613,58	14523,07	12	1054	251042,89	234032,71	20	40
Julio 03	18865,16	18848,22	13	1128	354017,00	309265,69	22	40
Julio 04	32928,64	30721,44	13	1230	526558,14	426254,49	22	40

(\*) Excluye los sectores alimentos, bebidas y textiles.  
FUENTE: Índice de Evolución Económica. Programa Institucional de Investigación y Extensión. Facultad de Ciencias Económicas. Universidad Nacional de Río Cuarto.

Con un nivel de confianza del 90%, calcula para el mes de julio de 2004

- Intervalo de confianza para la media poblacional
- Intervalo de confianza para el desvío poblacional

Estima las ventas totales del sector comercio minorista para el mes de julio de 2004.

¿Hay diferencias significativas en los niveles y la variabilidad de las ventas de julio de 2004 respecto de las registradas en julio de 1998?

### Caso 7.3: Demanda de diario regional

Una editorial del interior del país realiza un estudio con el objetivo de estimar la demanda semanal de un nuevo diario regional. Los resultados que arroja el estudio, sobre una muestra de 400 casos, se resumen en las tablas que se adjuntan.

Los directivos de la editorial necesitan que realices una estimación por intervalo de confianza, con un nivel de significación del 0,05, del número promedio de diarios a vender por semana.

Región de cobertura del nuevo diario	
Total de habitantes	101289
Total de hogares	27980
FUENTE: INDEC.	

Le gustaría un diario regional?	
SI	87,67
NO	5,02
Ns/Nc	7,31
Total de hogares que leen	100,00
FUENTE: Encuesta de opinión. Estudios & Mercados Consultora. Diciembre 2003.	

#### Hábitos de los hogares hacia la lectura de diarios

Cuándo lee el diario? (en % sobre el total de hogares)		Para leer el diario... (en % sobre el total de hogares que leen)	
Todos los días (5 a 7)	8,5	Lo compra	55,25
Algunos días (2 a 4)	21,0	Se lo prestan	25,11
Fin de semana	7,5	Lo lee en su trabajo	10,96
Algunas veces	17,25	De otra forma	8,22
De otra forma	0,25	Ns/Nc	0,46
Ns/Nc	1,75	Total de hogares que leen	100,00
No lee diarios	43,75		
Total de hogares	100,00		

FUENTE: Encuesta de opinión. Estudios & Mercados Consultora. Diciembre 2003.

FUENTE: Encuesta de opinión. Estudios & Mercados Consultora. Diciembre 2003.

### Problemas

7.1. El gerente de control de calidad de una fábrica de lámparas eléctricas desea estimar la duración promedio de un embarque de lámparas (focos). Se selecciona una muestra aleatoria de 64 focos. Los resultados indican una duración promedio de la muestra de 540 horas con una desviación estándar de 120 horas.

Realiza una estimación de la duración promedio real de los focos de este embarque:

- a) con intervalo de confianza de 90%,
- b) con intervalo de confianza de 95%,
- c) con intervalo de confianza de 99%.

Observa cómo cambia el tamaño del intervalo.

7.2. El administrador de una sucursal de un banco de ahorro local desea estimar la cantidad promedio que se tiene en las cuentas de ahorro de los clientes del banco. Seleccionó una muestra aleatoria de 30 depositantes y los resultados indicaron un promedio de muestra de \$4750.00 y un desvío estándar de \$1200.00

- a. Establezca una estimación por intervalo de confianza del 95%, de la cantidad promedio que se tiene en todas las cuentas de ahorro
- b. Si un cliente tiene \$4000.00 ¿puede considerárselo fuera de lo normal? ¿Porqué?.

7.3. Una línea de colectivos piensa establecer una ruta desde un barrio hasta el centro de la ciudad. Se selecciona una muestra aleatoria de 50 posibles usuarios y 18 indicaron que utilizarían esta ruta de colectivos. Establece una estimación por intervalo con 95% de confianza de la proporción real de usuarios para esta nueva ruta de autobuses.

7.4. De dos grupos de pacientes A y B, compuestos de 50 y 100 individuos respectivamente, al primero le fue dado un nuevo tipo de píldoras para dormir y al segundo le fue dado un tipo convencional.

Para los pacientes del primer grupo el número medio de horas de dormir fue de 8.32 con una desviación típica estimada de 0.24 horas

Para los pacientes del grupo B el número medio de horas de dormir fue de 6.75 con una desviación típica estimada de 0.30 horas

Encuentra los límites de confianza para la diferencia en las varianzas del número medio de horas de sueño inducidas por los dos tipos de píldoras, para un nivel de significación del 0,10 y el 0,02.

### **Bibliografía**

- Berenson, Mark y Levine, David. (1996). *Estadística Básica en Administración*. Prentice Hall. México.
- Daniel, Wayne W. (1999). *Bioestadística, base para el análisis de las ciencias de la salud*. Tercera Edición. Editorial Limussa. México.
- Dixon W.J. y Massey F.J. (1957) *Introduction to Statistical Analysis*. Nueva York, McGraw-Hill.
- Gomez Villegas, Miguel Angel (2005) *Inferencia Estadística*. Ediciones Díaz de Santos.
- Graybill, Mood. *Introducción a la Teoría de la Estadística*.
- Kazmier, Leonard y Diaz Mata, Alfredo. (1993). *Estadística Aplicada a la Administración y a la Economía*. Mc.Graw Hill. México.
- Meyer, P.L. (1973). *Probabilidad y Aplicaciones Estadísticas*. Fondo Educativo Interamericano. México.
- Sheldon M. Ross (2007) *Introducción a la Estadística*. Editorial Reverté.

Capítulo 8. TEORIA DE LA DECISIÓN	275
8.1 Pruebas de hipótesis	275
Tipos de d́cima	275
1. <i>D́cima bilateral</i>	276
2. <i>D́cima lateral izquierda</i>	276
3. <i>D́cima lateral derecha</i>	277
8.2 Tabla de contingencia	282
El estadístico chi-cuadrado	283
La prueba de la hipótesis	285
CASOS DE ESTUDIO, PREGUNTAS Y PROBLEMAS	285
Caso 8.1: Edades y Estaturas.	285
Caso 8.2: Demanda de diario regional	285
Caso 8.3: Razas de perros	285
Problemas	285
Bibliografía	288



## Capítulo 8. TEORIA DE LA DECISIÓN

### 8.1 Pruebas de hipótesis

La prueba de hipótesis, también conocida como dócima o contraste de hipótesis, proporciona una excelente oportunidad para analizar la información. Cuando un hallazgo interesante y empírico emerge del análisis de información basado en una muestra, una hipótesis sencilla pero relevante debe ocurrírsele a todo investigador:

⇒ *Representa el hallazgo empírico solo un accidente de muestreo?*

Una hipótesis estadística es una conjetura con respecto a uno o más parámetros poblacionales y el método de contraste es el que permite tomar decisiones en base a datos muestrales y riesgos conocidos.

La prueba de hipótesis, o contraste de hipótesis, consiste en:

- 1- Formular hipótesis nula,  $H_0$
- 2- Formular hipótesis alternativa,  $H_1$
- 3- Establecer nivel de confianza para  $H_0$
- 4- Definir estadístico de contraste
- 5- Aplicar regla de decisión

La hipótesis nula indica que en la población no se han producido cambios y la hipótesis alternativa es una afirmación que enfrenta lo establecido por la hipótesis nula. Se fija un nivel de confianza  $1-\alpha$  de que la hipótesis nula sea cierta. Este nivel de confianza fija los puntos críticos dentro de los cuales se acepta o se rechaza la hipótesis nula.

Tipos de dócima

Pueden plantearse tres situaciones:

Dócima bilateral

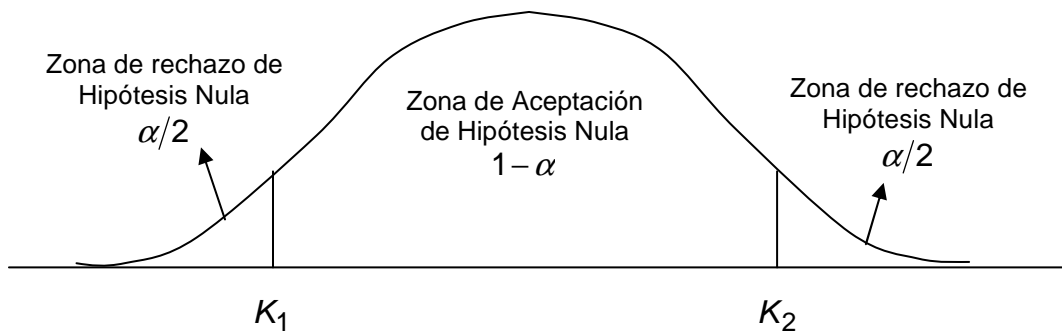
Dócima lateral izquierda

Dócima lateral derecha

### 1. Dócima bilateral

$$H_0 : \theta = \theta_0$$

$$H_1 : \theta \neq \theta_0$$



$\alpha$  es la probabilidad de rechazar la  $H_0$  siendo cierta. Esto se lo conoce como Error tipo I. Si se acepta la  $H_0$  siendo falsa se comete el Error tipo II y se lo conoce como  $\beta$ , esto es la Potencia de la prueba. Los estadísticos de contraste,  $K_1$  y  $K_2$ , son los puntos críticos que definen la región de aceptación y de rechazo y se obtienen de la distribución de probabilidad asociada a la prueba al nivel de confianza  $1 - \alpha$  fijado.

### 2. Dócima lateral izquierda

$$H_0 : \theta \geq \theta_0$$

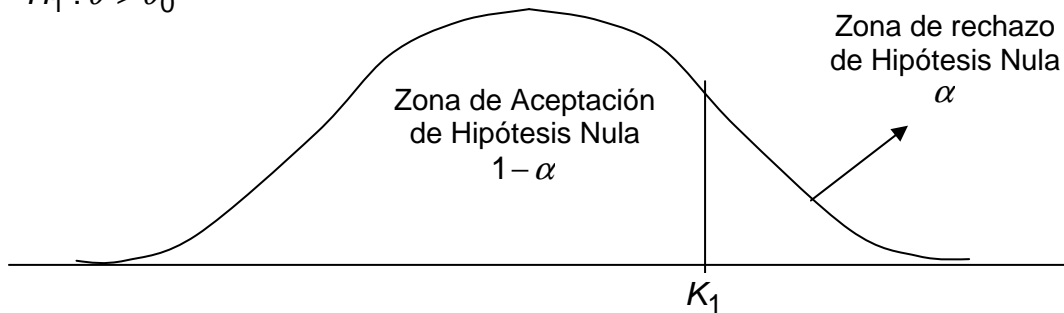
$$H_1 : \theta < \theta_0$$



### 3. Dócima lateral derecha

$$H_0 : \theta \leq \theta_0$$

$$H_1 : \theta > \theta_0$$



En la siguiente tabla se reúnen las diferentes situaciones de pruebas de hipótesis, para variables cuantitativas, con el estadístico a utilizar.

Dócima	Estadístico
Media con $\sigma^2$ poblacional conocida	$Z = \frac{\bar{X} - \mu}{\sigma/\sqrt{n}} \sim N(0,1)$
Media con $\sigma^2$ desconocida, n grande	$Z = \frac{\bar{X} - \mu}{s/\sqrt{n}} \sim N(0,1)$
Media con $\sigma^2$ desconocida, n pequeña, probabilidad Normal	$t = \frac{\bar{X} - \mu}{s/\sqrt{n}} \sim t_{n-1}$
Proporción	$Z = \frac{p - \pi}{\sqrt{\frac{\pi(1-\pi)}{n}}}$
Diferencias de medias con $\sigma^2$ conocida y n de cualquier tamaño	$Z = \frac{\bar{X}_1 - \bar{X}_2 - (\mu_1 - \mu_2)}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}} \sim N(0,1)$
Diferencia de medias con $\sigma^2$ desconocida muestras grandes	$Z = \frac{\bar{X}_1 - \bar{X}_2 - (\mu_1 - \mu_2)}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}} \sim N(0,1)$
Diferencia de medias, $\sigma^2$ desconocida pero iguales y n pequeño	$t = \frac{\bar{X}_1 - \bar{X}_2 - (\mu_1 - \mu_2)}{\sqrt{\frac{n_1 + n_2}{n_1 n_2} \frac{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2}{n_1 + n_2 - 2}}} \sim t_{n_1 + n_2 - 2}$
Diferencia de medias, $\sigma^2$ desconocida y distintas, n pequeño	$t = \frac{\bar{X}_1 - \bar{X}_2 - (\mu_1 - \mu_2)}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}} \sim t_{\min(n_1, n_2) - 1}$



Igualdad de Proporciones

$$k = \frac{p_1 - p_2 - (\pi_1 - \pi_2)}{\sqrt{\frac{p_1(1-p_1)}{n_1} + \frac{p_2(1-p_2)}{n_2}}} \sim N(0,1)$$

 $\sigma^2$  de población normal

$$W = (n-1) \frac{s^2}{\sigma^2} \sim \chi_{n-1}^2$$

Igualdad de varianza

$$\frac{s_1^2}{s_2^2} \frac{\sigma_2^2}{\sigma_1^2} \sim F_{n_1-1; n_2-1}$$

**Ejemplo.** Un grupo de investigadores está interesado en conocer la concentración de una enzima en cierta población de seres humanos. Toman una muestra de 20 individuos de la población bajo estudio y calculan que la media muestral es de 22 y la varianza muestral  $s^2 = 45$ . Se supone que la muestra proviene de una población que sigue una distribución normal con varianza desconocida. Los investigadores saben que habitualmente los niveles de concentración de esta enzima en las poblaciones es de 25. Una teoría a ser investigada era que los niveles de concentración de la enzima en esta población es distinto de 25.

1. Suponga, la **Hipótesis Nula** de que la concentración de enzima es igual a 25 y que si todas las personas de la población fueran contactadas, encontraríamos que el nivel de enzima es 25.

$$H_0 : \mu = 25$$

2. Otro argumento, que será denominado **Hipótesis Alternativa**, es que la concentración media de la enzima de la población es diferente de 25. Es importante destacar que se está identificando la hipótesis alternativa con la conclusión a la que se quiere llegar, de manera que si los datos permiten rechazar la hipótesis nula, la conclusión de los investigadores tendrá mayor peso, dado que la probabilidad complementaria de rechazar una hipótesis nula verdadera es pequeña.

$$H_1 : \mu \neq 25$$

Pero, qué tan convincente es la evidencia? Después de todo, solo se conocen los resultados de una muestra de 20 personas. La diferencia entre 25 (el nivel de concentración de enzima conocido), y 22 (el promedio de la muestra), podría ser mas bien un caso de error de muestreo que diferencias reales en la existencia de enzimas en la población.

3. Para evaluar la evidencia estadísticamente, se establece una probabilidad, denominada **nivel de confianza**  $(1 - \alpha)$ , de 0.95 (ó 95%) de que la hipótesis nula es verdadera, es decir que el resultado es sólo una opinión parcial o sea un accidente de muestreo.

4. A continuación, debemos definir un estadístico que nos permita contrastar la hipótesis nula a un nivel de confianza del 95%. Ese estadístico es:

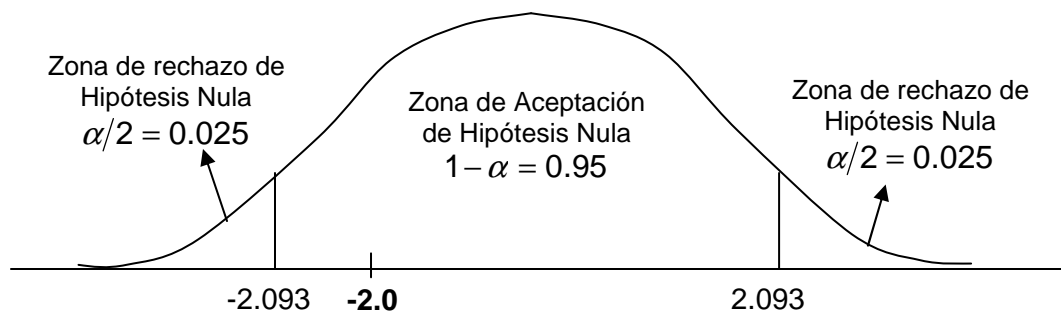
$$t = \frac{\bar{x} - \mu}{s_{\bar{x}}} = \frac{22 - 25}{\sqrt{45} / \sqrt{20}} = -2.00$$

donde,  $t$  es un estadístico que cuando la muestra es pequeña ( $n < 30$ ), como en este caso, tiene una distribución  $t$  de Student y refleja, en unidades tipificadas, en cuánto difiere la media de la muestra  $\bar{X}$  de la media de la población  $\mu$ .

Luego, comparamos el valor empírico de este estadístico con el valor teórico obtenido de una tabla de distribución  $t$  de Student para 19 ( $n - 1$ ) grados de libertad y para un valor  $\alpha$  de 0.05,

$$\text{Valor teórico de } t = 2.093$$

5. La regla de decisión es, acepte la hipótesis nula, al nivel de confianza establecido, si el valor empírico de  $t$  es, en valor absoluto, menor al valor teórico. Caso contrario se debe rechazar.



En nuestro caso, debemos aceptar la hipótesis nula, es decir que no basta la evidencia de la muestra para decir que la población tiene una concentración de enzima diferente a 25.

**Ejemplo.** El fabricante de un medicamento afirma que el mismo es efectivo en el 90% de los casos o más en que se aplicó para aliviar una alergia. En una muestra de 200 pacientes se encontró que produjo alivio en 160 casos. ¿Se justifica la afirmación del fabricante a un nivel del 1%?

Este caso se corresponde con una dócima lateral derecha donde:

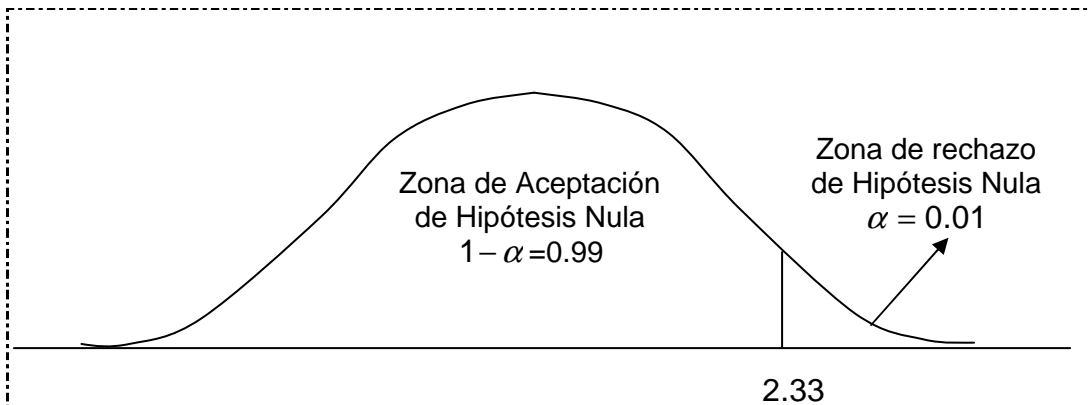
$$H_0 : \Pi \leq 90$$

$$H_1 : \Pi > 90$$

En la muestra el estimador es:

$$p = \frac{160}{200} = 0,80$$

Dado que la muestra es grande,  $n=200$ , el estadístico a utilizar es una  $Z \sim N(0,1)$ . Al nivel de confianza de 0.99, el valor crítico de la distribución normal es 2.33. Este divide la zona de aceptación de la zona de rechazo de la hipótesis nula.



Para realizar el contraste se debe calcular el valor empírico del estadístico Z.

$$Z^* = \frac{p - \pi}{\sqrt{\frac{\pi(1-\pi)}{n}}} = \frac{0,80 - 0,90}{\sqrt{\frac{0,90 * 0,1}{200}}} = \frac{-0,1}{0,02121} = -4.7148$$

El valor empírico cae en la zona de aceptación de la hipótesis nula. Esto significa que la efectividad es menor a 0.90, por lo tanto el fabricante no tiene razón.

También es válido calcular el intervalo de confianza para la hipótesis nula y ver a qué región pertenece el valor muestral. El conjunto de hipótesis se mantiene:

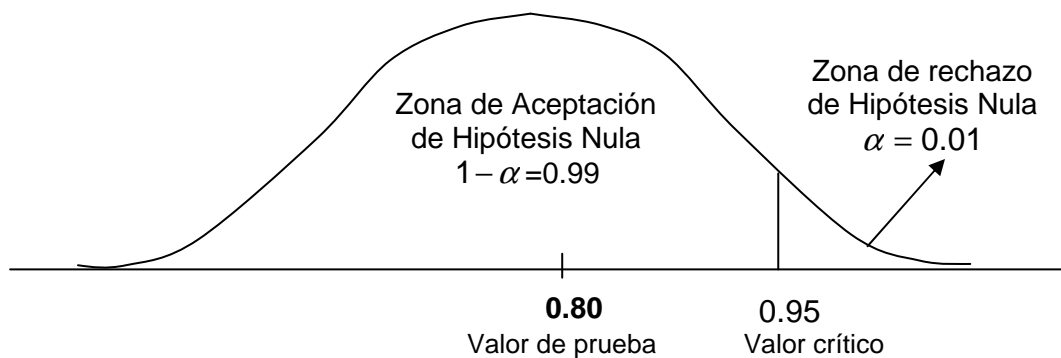
$$H_0 : \Pi \leq 90$$

$$H_1 : \Pi > 90$$

Debe calcularse ahora el límite superior del intervalo

$$L_s = \Pi + Z_1 \sqrt{\frac{\Pi(1-\Pi)}{n}} = 0,90 + 2,33 * 0,0212 = 0,949$$

El valor de 0.949, para el límite superior del intervalo, separa la zona de aceptación de la zona de rechazo de la hipótesis nula.



El valor empírico, o valor de prueba:

$$p = \frac{160}{200} = 0,80$$

cae en la zona de aceptación de la hipótesis nula.

**Ejemplo.** Al asignar subsidios federales entre distintas localidades, una determinada ciudad se clasificó como área de renta elevada. Las autoridades locales se mostraron disconformes y presentaron pruebas, basada en una muestra aleatoria de 25 familias, de que la renta familiar media de la ciudad era de \$7145 y se encontraba en niveles similares a la media nacional. Un estudio a escala nacional indicó que la renta familiar media era de \$6500 y la desviación estándar \$920.

¿Son consistentes los resultados muestrales presentados por las autoridades locales con su afirmación de que la ciudad en cuestión no es más próspera que las demás ciudades similares del país?

1. el parámetro a docimar es  $\mu$

2. la población es normal  $\mu = 6500$

$$\sigma = 920$$

3.  $H_0 : \mu = 6500$

$H_1 : \mu \neq 6500$

4.  $K(\bar{X}, \mu) = Z = \frac{\bar{X} - \mu}{\sigma_{\bar{X}}}$

$$\sigma_{\bar{X}} = \frac{\sigma}{\sqrt{n}} = \frac{920}{\sqrt{25}} = 184$$

5. El nivel de confianza  $1 - \alpha = 0,95$ . Con este nivel de confianza y la distribución normal, la zona de aceptación de la hipótesis nula se encuentra entre los valores críticos  $-1.96$  y  $1.96$



6. El valor empírico del estadístico Z es

$$Z^* = \frac{\bar{X}^* - 6500}{184} = \frac{7145 - 6500}{184} = 3.51$$

7.  $Z^*$  es mayor a  $1.96$ , se rechaza la  $H_0$  y se concluye que los ingresos medios de esta ciudad son diferentes al de otras ciudades similares.

También es válido en el punto 6 calcular los límites del intervalo de confianza

$$\bar{X}^* = \pm Z \frac{\sigma}{\sqrt{n}} + 6500 = \pm 1.96 * 184 + 6500$$

$$\bar{X}_1^* = 6139.36, \quad \bar{X}_2^* = 6860.64$$

El valor muestral  $\bar{X} = 7145$  cae en la zona de rechazo de la hipótesis nula por lo que la afirmación de las autoridades locales no es cierta.

## 8.2 Tabla de contingencia

Para estudiar la asociación de variables cuantitativas se utiliza el coeficiente de correlación. Cuando se trabaja con variables cualitativas, en lugar de la correlación, se analiza la independencia de las variables.

El concepto de **Independencia** es central, no solo para el estadístico de chi-cuadrado, sino para todas las medidas de asociación. Dos variables son estadísticamente independientes, si el conocimiento de una no ofrece ninguna información en cuanto a la identidad de la otra.

La tabla muestra los resultados de una encuesta de 200 asistentes a las funciones de ópera que fueron interrogados acerca de la frecuencia con la que asistían a los conciertos de la orquesta sinfónica en una ciudad vecina. La frecuencia de asistencia, fue dividida en las categorías de frecuente, ocasional y nunca; y se les preguntó si consideraban la ubicación de la sinfónica como conveniente o inconveniente.

ASISTENTES A LA OPERA				
Tabla de frecuencias absolutas				
Asistencia a los conciertos de la sinfónica (A)	Ubicación (U)		Total	Perfil fila $p_A$
	Conveniente	No conveniente		
Frecuente	22	18	40	0.20
Ocasional	48	52	100	0.50
Nunca	10	50	60	0.30
Total	80	120	200	1.00
Perfil columna $p_U$	0.40	0.60	1.00	

La tabulación cruzada resultante, muestra la clasificación porcentual de asistencia en cada categoría de localización. Los totales de las filas, los totales de las columnas y las proporciones ( $P_A$  y  $P_U$ ), son tabulados al margen.  $P_A$  y  $P_U$  son las distribuciones de frecuencias marginales para las variables respectivas. Por ejemplo, la fila total indica que 80 entrevistados (0.40 de todos los entrevistados), sentían que la localización era conveniente y 120 (0.60 de todos los entrevistados), sentían que era inconveniente.

Si el experimento se repitiera, un 20% de los resultados mostrarían que la Asistencia a los conciertos de la sinfónica caerían en la categoría de "frecuentes". La **frecuencia absoluta esperada**, bajo la condición de independencia estadística, sería de  $0.20n$ , donde  $n$  es el número de nuevos experimentos realizados. También se espera, bajo las mismas condiciones que el 40% de los casos tengan una ubicación "conveniente" y el 60% una ubicación "no conveniente".

De igual manera, el número de experimentos que darían como resultado una asistencia frecuente en una ubicación conveniente "se esperaría" que fueran

$$(0.40) (0.20n)$$

Tabla de frecuencias absolutas esperadas				
Asistencia a los conciertos de la sinfónica (A)	Ubicación (U)		Total	Perfil fila $P_A$
	Conveniente	No conveniente		
Frecuente	16	24	40	0.20
Ocasional	40	60	100	0.50
Nunca	24	36	60	0.30
Total	80	120	200	1.00
Perfil columna $P_U$	0.40	0.60	1.00	

### El estadístico chi-cuadrado

El estadístico chi-cuadrado, es una medida de la diferencia entre las frecuencias observadas en la celda  $i$ , con la frecuencia esperada bajo el supuesto de independencia estadística. El estadístico de chi-cuadrado, se define como:

$$\chi^2_{(r-1)(c-1)} = \sum \frac{(f_a - f_a^*)^2}{f_a^*}$$

donde,

$r$  es el número de filas o renglones

$c$  el número de columnas.

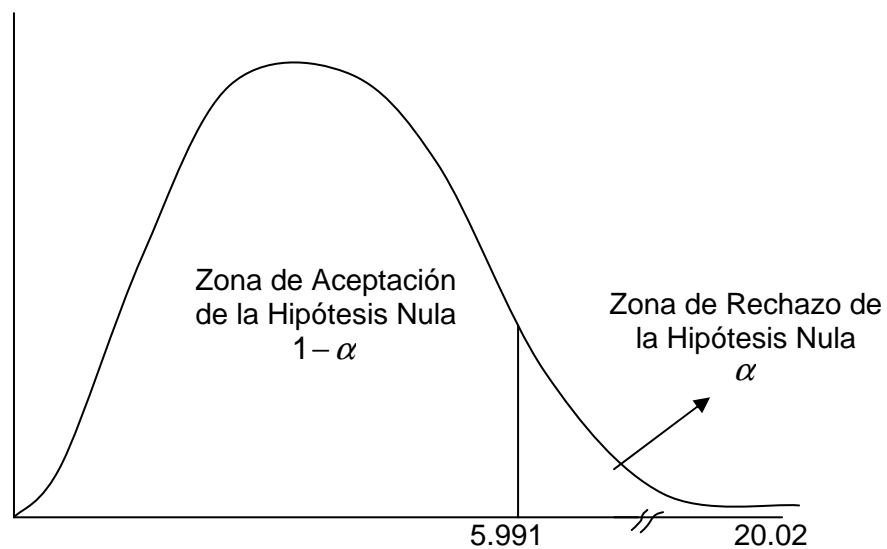
$(r-1)(c-1)$  son los grados de libertad.

Si las variables son estadísticamente independientes, el valor  $\chi^2$  debería ser relativamente pequeño. Sin embargo, si las variables no son independientes -si están asociadas o relacionadas-, entonces el valor de  $\chi^2$  debería ser relativamente grande. El valor empírico para el estadístico chi cuadrado es:

$$\chi^2_{(r-1)(c-1)} = \sum \frac{(f_a - f_a^*)^2}{f_a^*}$$

$$= \frac{(22-16)^2}{16} + \frac{(18-24)^2}{24} + \frac{(48-40)^2}{40} + \frac{(52-60)^2}{60} + \frac{(10-24)^2}{24} + \frac{(50-36)^2}{36} = 20.02$$

El estadístico de chi-cuadrado teórico, para 2 grados de libertad y nivel de confianza de 0.95, es de 5.991. Este valor crítico divide la zona de aceptación de la hipótesis nula de la zona de rechazo. El valor empírico de 20.02 es mayor al teórico. Esto indica que las dos variables pueden no ser estadísticamente independientes; al menos en este caso particular, se encontró una **asociación muestral** entre las dos variables.



## La prueba de la hipótesis

La **hipótesis nula** asociada con el estadístico muestral chi-cuadrado, es que las dos variables (categóricas), son estadísticamente independientes.

La **hipótesis alternativa**, es que la dos variables no son independientes.

La prueba de la hipótesis se basa en el hecho de que el estadístico chi-cuadrado está distribuido como la distribución chi-cuadrado con  $(r-1)(c-1)$  grados de libertad, suponiendo que la hipótesis nula es verdadera. O, en otras palabras, dado un nivel de significación  $(1-\alpha)$  se puede determinar, de acuerdo a los grados de libertad, el valor teórico de Chi-Cuadrado, se lo compara con el valor empírico y si el segundo es superior, se rechaza la hipótesis de independencia.

## **CASOS DE ESTUDIO, PREGUNTAS Y PROBLEMAS**

### Caso 8.1: Edades y Estaturas.

Con la información muestral suministrada en el Caso 6.1 y con un nivel de significación de 0.05, pruebe la hipótesis de que la estatura promedio de los cursantes 2007 de Inferencia Estadística supera 1.71cm.

### Caso 8.2: Demanda de diario regional

El responsable del diario regional del Caso 6.3 afirma que editará un nuevo diario regional si tiene evidencias empíricas suficientes que le garanticen, con un nivel de significación del 0.10, que el 75% de los hogares compran habitualmente algún diario.

### Caso 8.3: Razas de perros

A partir de la tabla de datos generada en el Caso 7.2, seleccione dos variables y pruebe la hipótesis de independencia.

## Problemas

8.1. El saldo que arroja la cuenta deudores varios en el balance general de una empresa es de \$120000. Se conoce además que el total de deudores es de 500. Tomada una muestra al azar de 50 de esos deudores se encontró que debían en total \$11900. Es confiable a un nivel del 1% la cifra que reflejan los registros contables si suponemos población aproximadamente normal y desvío poblacional de 100?



8.2. En una fábrica de alambres se sabe por registros estadísticos anteriores que la resistencia media de un tipo de alambre era de 12.46kg con una desviación típica de 1.8 kg. Se toma una muestra de 25 pedazos de alambre y se encontró que la media era de 31.01 y la varianza estimada de 4.5. Como se planea efectuar una docima de la media necesitan saber si la varianza no se ha alterado. Docime a un nivel de significación del 5% si tal hecho ocurre. Suponga que la población es normal.

8.3. En una muestra de 400 lámparas producidas por una empresa dedicada al ramo se encontró que la vida media de cada una es de 1585 horas. Se conoce además que la desviación típica de la población es de 120 horas. El Departamento de Ingeniería asegura que la vida media de las lámparas producidas debe ser de 1600 horas. ¿Aceptaría Ud. esta aseveración a la luz de los resultados muestrales a un nivel de significación del 1%?

8.4. Un constructor está considerando dos lugares alternativos para un centro comercial regional. Como los ingresos de los hogares de la comunidad son una consideración importante en esa selección, desea probar la hipótesis nula de que no existe diferencia entre el ingreso promedio por hogar en las dos comunidades. Consistente con esta hipótesis supone que la desviación estándar del ingreso por hogar es también igual en las dos comunidades. Para una muestra de 30 hogares de la primera comunidad encuentra que el ingreso promedio es de \$35500 con desviación estándar muestral de \$1800. Para una muestra de 40 hogares de la segunda comunidad encuentra que el ingreso promedio es de \$34600 con desviación estándar muestral de \$2400. Probar la hipótesis nula en el nivel de significación del 5%.

8.5. Aceros S.A. fabrica barras de acero. El proceso de producción hace barras con una longitud promedio de, cuanto menos, 2.8 pies cuando el proceso funciona correctamente. Se selecciona una muestra de 25 barras en la línea de producción. La muestra indica una longitud promedio de 2.43 pies y una desviación estándar de 0.20 pies. La compañía desea determinar si la máquina necesita algún ajuste. A un nivel del 5%, que decisión tomaría?

8.6. De 100 graduados en contaduría, una muestra aleatoria de 12 estudiantes tiene un promedio de calificación de 2.70 (el máximo es 4) con una desviación estándar muestral de 0.40. Para 50 egresados de sistemas de información una muestra aleatoria de 10 estudiantes tiene un promedio de calificación de 2.90 con una desviación estándar muestral de 0.30. Se supone que las calificaciones tienen distribución normal. Pruebe la hipótesis nula de que la calificación promedio para las dos categorías de estudiantes es distinta utilizando el nivel de significación del 5%.

8.7. Se plantea la hipótesis de que la desviación estándar del salario por hora de los trabajadores a destajo en una determinada industria es de, cuanto menos, \$3000. Para una muestra de 15 trabajadores elegidos al azar se encuentra que la desviación estándar es de 2000. Se supone que las cifras de ingresos de los trabajadores de la población tienen una distribución normal. Con base en este resultado muestral ¿puede rechazarse la hipótesis nula utilizando el nivel de significación del 5%?.

8.8. Una fábrica de galletitas incorporó un nuevo proceso de elaboración de bizcochos dulces que les trae aparejado ciertas ventajas. Pero un detalle muy importante es el peso de las galletas. El gerente de fabricación desea saber el grado de variación del nuevo proceso respecto del anterior en cuanto a peso de cada galleta. Tomada una muestra de cada proceso se obtuvo

$$n = 16$$

$$\sum_{i=1}^{16} X_i = 319.20$$

$$\sum_{i=1}^{16} X_i^2 = 6373.11$$

$$n = 13$$

$$\sum_{i=1}^{13} Y_i = 261.30$$

$$\sum_{i=1}^{13} Y_i^2 = 5255.54$$

Sabiendo que ambas poblaciones son aproximadamente normales. ¿Hay diferencias significativas en cuanto a las varianzas de los pesos de los bizcochos entre ambos procesos a un nivel del 99%?

8.9. El gerente de marketing de una cadena de tiendas de flores y plantas cree que habrá una gran demanda de bulbos de gladiolos en las tiendas de los suburbios durante el mes de mayo.

Si las ventas de bulbos exceden de \$100 por semana se venderán en todas las tiendas suburbanas de la cadena.

Se selecciona una muestra aleatoria de 16 tiendas y los resultados de la prueba en las tiendas indicaron ventas promedio de \$120 con una desviación estándar de la muestra de \$25

Con un nivel de significación del 1% ¿se deberían vender los bulbos de gladiolos en todas las tiendas suburbanas de la cadena?

8.10. En la tabla de contingencia se muestran las 200 personas que entraron en una tienda de equipos de sonido de acuerdo con el sexo y la edad. Pruebe si las dos variables categóricas son independientes.

Clientes de la tienda de aparatos de sonido			
Edad	Sexo		Total
	Hombre	Mujer	
Menor de 30	60	50	110
30 y más	80	10	90
Total	140	60	200

8.11. En la tabla se muestran las reacciones de los votantes ante un nuevo plan de impuestos sobre bienes raíces, de acuerdo con su afiliación política partidaria. Con estos datos, construya una tabla de frecuencias esperadas suponiendo que no existe relación ante el plan fiscal.

Reacciones de los votantes ante un nuevo plan de impuestos				
Afiliación partidista	Reacción			Total
	A favor	Neutral	Se opone	
Partido A	120	20	20	160
Partido B	50	30	60	140
Otro	50	10	40	100
Total	220	60	120	400

8.12. En la tabla siguiente se presentan los datos relacionados con la reacción de los estudiantes ante la ampliación de un programa deportivo colegial de acuerdo con la clase a la que pertenecen, en donde "división menor" indica que se trata de un alumno de primer o segundo año, y la "división superior" señala que los alumnos se encuentran en el tercer o cuarto año. Pruebe la hipótesis nula de que la posición de clase y la reacción ante

el programa deportivo son variables independientes, utilizando el nivel de significación del 5%.

Reacción de los estudiantes ante el plan deportivo de acuerdo a su generación			
Reacción	Generación		Total
	División inferior	División superior	
A favor	20	19	39
En contra	10	16	26
Total	30	35	65

---



---

### **Bibliografía**

- Berenson, Mark y Levine, David. (1996). *Estadística Básica en Administración*. Prentice Hall. México.
- Dagum, C. y Dagum, E.M.de. (1971). *Introducción a la Econometría*. Editorial Siglo XXI. México.
- Daniel, Wayne W. (1999). *Bioestadística, base para el análisis de las ciencias de la salud*. Tercera Edición. Editorial Limussa. México.
- Dixon W.J. y Massey F.J. (1957). *Introduction to Statistical Analysis*. Nueva York, McGraw-Hill.
- Graybill, Mood. *Introducción a la Teoría de la Estadística*.
- Gujarati, D. (2004). *Econometría*. 4ª Edición. Mc.Graw Hill. México.
- Hildebrand, D. y Lyman Ott R. (1997). *Estadística Aplicada a la administración y a la economía*. Addison Wesley. Wilmington USA.
- Johnston, J. y Dinardo, J. (2001). *Métodos de Econometría*. Editorial Vicens Vives. Barcelona.
- Kazmier, Leonard y Diaz Mata, Alfredo. (1993). *Estadística Aplicada a la Administración y a la Economía*. Mc.Graw Hill. México.
- Meyer, P.L. (1973). *Probabilidad y Aplicaciones Estadísticas*. Fondo Educativo Interamericano. México.
- Padua, Jorge. (1996). *Técnicas de Investigación Aplicadas a las Ciencias Sociales*. Fondo de Cultura Económica. México.
- Tramutola, C.D. *Modelos Probabilísticos y Decisiones Financieras*. E.C.Moderan. Lectura de Administración de Empresas. Capital Federal.

Capítulo 9. INTRODUCCION AL ANALISIS DE INFORMACION	291
9.1 Análisis de varianza	291
9.2. Correlación de muestras	293
9.3. Regresión lineal simple	294
Especificación del modelo	295
Estimación de los parámetros del modelo	296
La interpretación de los parámetros estimados	298
Predicción	299
Coeficiente $r^2$	301
CASOS DE ESTUDIO, PREGUNTAS Y PROBLEMAS	305
Caso 9.1: Función consumo	305
Caso 9.2: Evolución del Consumo	305
Problemas	305
Bibliografía	308



## Capítulo 9. INTRODUCCION AL ANALISIS DE INFORMACION

### 9.1 Análisis de varianza

Este método se utiliza para comparar las diferencias entre las medias de diferentes grupos y la hipótesis nula es que hay igualdad entre las medias de los distintos grupos:

$$H_0 : \mu_1 = \mu_2 = \dots = \mu_k$$

La hipótesis alternativa es que las medias no son todas iguales

$$H_1 : \text{no todas las } \mu_j \text{ son iguales } j = 1, 2, \dots, k$$

Para llevar a cabo una prueba ANOVA, se subdivide la variación total (SST) en aquellas que pueden atribuirse a la variación entre grupos (SSA) y la que se debe a variaciones dentro de los grupos (SSW).

El estadístico de contraste es el cociente de la varianza entre grupos,  $MSA = (SSA / k - 1)$ ; y la varianza dentro de los grupos  $MSW = (SSW / n - k)$

$$F = \frac{MSA}{MSW} \text{ que se distribuye } F_{k-1; n-k}$$

**Ejemplo.** Se asignan en forma aleatoria tres tipos de métodos de instrucción a 15 participantes de un programa técnico. Los métodos desarrollan un nivel determinado de habilidad en diseño auxiliado por computadora. En la tabla se presentan las calificaciones alcanzadas al término de la unidad de instrucción y se presentan las calificaciones promedio correspondientes.

Instrucción	Calificaciones de la prueba por participante					Calificaciones Totales $\sum_{i=1}^5 X_i$	Calificaciones Promedio $\sum_{i=1}^5 X_i / n$
	1	2	3	4	5		
$A_1$	86	79	81	70	84	400	80
$A_2$	90	76	88	82	89	425	85
$A_3$	82	68	73	71	81	375	75

Se calcula primero la media total

$$\bar{X} = \frac{\sum_{j=1}^k \sum_{i=1}^n X_{ij}}{p * n} = \frac{\sum_{j=1}^3 \sum_{i=1}^5 X_{ij}}{3 * 5} = \frac{400 + 425 + 375}{15} = 80$$

Luego, se calculan la variación total (SST), la variación entre grupos (SSA) y la variación dentro del grupo (SSW).

$$\begin{aligned} SST &= \sum_{j=1}^k \sum_{i=1}^{n_k} (X_{ij} - \bar{X})^2 = \sum_{j=1}^3 \sum_{i=1}^5 (X_{ij} - 80)^2 \\ &= (86 - 80)^2 + (79 - 80)^2 + (81 - 80)^2 + (70 - 80)^2 + (84 - 80)^2 \\ &\quad + (90 - 80)^2 + (76 - 80)^2 + (88 - 80)^2 + (82 - 80)^2 + (89 - 80)^2 \\ &\quad + (82 - 80)^2 + (68 - 80)^2 + (73 - 80)^2 + (71 - 80)^2 + (81 - 80)^2 \\ &= 698 \end{aligned}$$

$$\begin{aligned} SSA &= \sum_{j=1}^k n_j (\bar{X}_j - \bar{X})^2 = \sum_{j=1}^3 5 \left( (80 - 80)^2 + (85 - 80)^2 + (75 - 80)^2 \right) \\ &= 5(0 + 25 + 25) = 50 * 5 = 250 \end{aligned}$$

$$\begin{aligned} SSW &= \sum_{j=1}^k \sum_{i=1}^{n_k} (X_{ij} - \bar{X}_j)^2 = \sum_{i=1}^5 (X_{i1} - 80)^2 + \sum_{i=1}^5 (X_{i2} - 75)^2 + \sum_{i=1}^5 (X_{i3} - 85)^2 \\ &= (86 - 80)^2 + (79 - 80)^2 + (81 - 80)^2 + (70 - 80)^2 + (84 - 80)^2 \\ &\quad + (90 - 85)^2 + (76 - 85)^2 + (88 - 85)^2 + (82 - 85)^2 + (89 - 85)^2 \\ &\quad + (82 - 75)^2 + (68 - 75)^2 + (73 - 75)^2 + (71 - 75)^2 + (81 - 75)^2 \\ &= 448 \end{aligned}$$

Fuente de Variación		Grados de libertad
Entre fila variación explicada	$SSA = \sum_{j=1}^k n_j (\bar{X}_j - \bar{X})^2 = 250$	$k - 1 =$ $3 - 1 = 2$
Dentro de la fila variación no explicada	$SSW = \sum_{j=1}^k \sum_{i=1}^{n_k} (X_{ij} - \bar{X}_j)^2 = 448$	$n - k =$ $15 - 3 = 12$
Total	$SST = \sum_{j=1}^k \sum_{i=1}^{n_k} (X_{ij} - \bar{X})^2 = 698$	$n - 1 =$ $15 - 1 = 14$

Fuente de Variación	Estimación de Varianza	Estadístico
Entre fila variación explicada	$MSA = \frac{SSA}{k-1} = \frac{250}{2} = 125$	$F = \frac{MSA}{MSW}$ $= \frac{125}{37.33}$ $= 3.35$
Dentro de la fila variación no explicada	$MSW = \frac{SSW}{n-k} = \frac{448}{12} = 37.333$	
Total	$MST = \frac{SST}{n-1} = \frac{698}{14} =$	

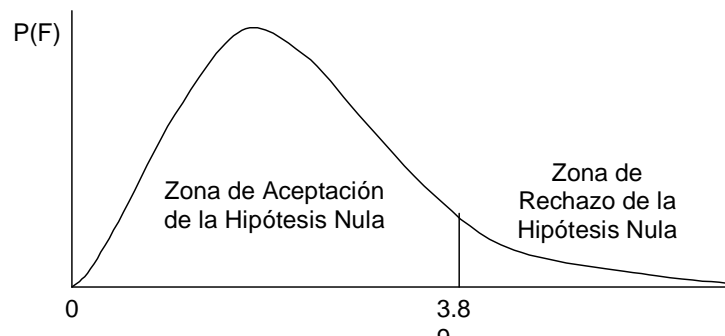
La hipótesis nula a probar es

$$H_0 : \mu_1 = \mu_2 = \mu_3$$

Al nivel de confianza de 0.95, el estadístico crítico es

$$F_{k-1; n-k} = F_{3-1; 15-3} = F_{2; 12} = 3.89$$

El valor empírico  $F = 3.35$  cae en la zona de aceptación de la hipótesis nula, por ende se concluye que las medias son iguales.



## 9.2. Correlación de muestras

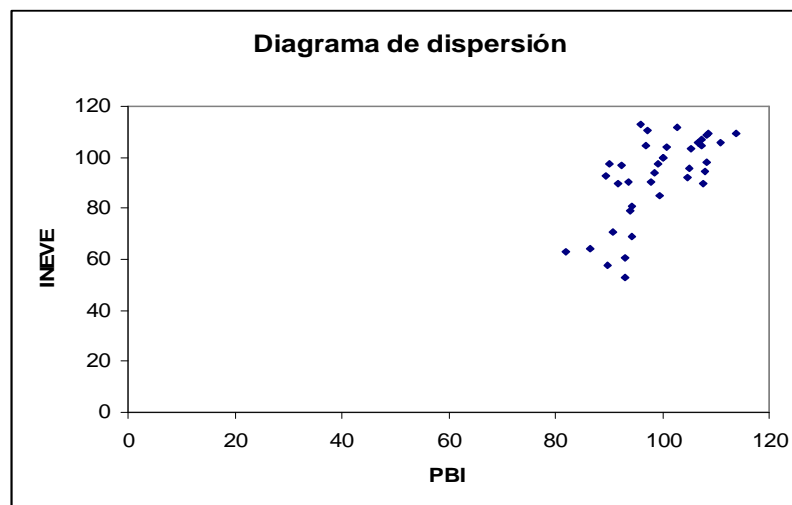
La correlación mide el grado en el cual existe una asociación entre dos preguntas o variables y su aplicación es válida cuando las variables son cuantitativas.

- ⇒ Una correlación positiva, reflejará una tendencia de alto valor para una primera variable que está asociada con un alto valor de una segunda.
- ⇒ Una correlación negativa, refleja una asociación entre un valor alto de una primera variable y un valor bajo de una segunda variable.

La correlación muestral se observa gráficamente al representar dos variables sobre una gráfica bidimensional, denominada Diagrama de Dispersión. La correlación muestral refleja la tendencia para que los puntos se agrupen sistemáticamente alrededor de una línea recta que crece o decrece de izquierda a derecha. El gráfico muestra la dispersión



de las variables PBI a nivel nacional e Índice de Evolución Económica de Río Cuarto (INEVE), las mismas presentan una asociación estadística positiva



El coeficiente de correlación muestral se define como,

$$r = \frac{\sum (x - \bar{x})(y - \bar{y})}{(n-1) s_x s_y}; \quad -1 \leq r \leq 1$$

Una  $r$  de +1, indica una **asociación positiva perfecta** ente las dos variables, mientras que si  $r$  es -1, existe una **asociación negativa perfecta**. Una  $r$  de 0, refleja la ausencia de cualquier asociación lineal.

**Ejemplo.** La varianza observada en el indicador local de actividad económica (INEVE) es de 274.55 y la varianza en el índice de PBI 56.01. La covarianza entre ambas variables es de 71.56. ¿Cuál es la correlación entre ambas variables?

$$r = \frac{\sum (x - \bar{x})(y - \bar{y})}{(n-1) s_x s_y} = \frac{\text{Cov}(X, Y)}{\sqrt{S_X^2} \sqrt{S_Y^2}} = \frac{71.56}{\sqrt{274.55} \sqrt{56.01}} = 0.5927$$

### 9.3. Regresión lineal simple

En el análisis de la información existen, frecuentemente, un número de variables claves que se convierten en el centro de atención del estudio.

El análisis de regresión proporciona una herramienta que puede cuantificar tales relaciones y proporcionar un control estadístico. Es un

análisis de dependencia puesto que involucra una **variable dependiente**, como el punto de atención del análisis, la cual es explicada por las **variables independientes** o variables explicativas.

El análisis de regresión está orientado hacia

- la descripción, obtención de las relaciones entre variables independientes y una variable dependiente para un conjunto de observaciones.
- la predicción, niveles a alcanzar por la variable dependiente bajo el supuesto de comportamiento de la variable explicativa.

### Especificación del modelo

La construcción de un modelo de regresión, generalmente, empieza con la especificación de la variable dependiente y de la variable, o variables, independientes.

$$Y = f(X)$$

$Y$ : variable dependiente o endógena

$X$ : variable independiente o exógena.

Como no se puede explicar en forma perfecta el comportamiento de  $Y$  sólo por  $X$ , se adiciona un término de error,  $\mu$ :

$$Y = f(X) + \mu$$

A ésta se la denomina *ecuación de regresión* de  $Y$  sobre  $X$ .  $\mu$  es el residuo, o error, y es una variable aleatoria. Este término hace referencia a error de medición de las variables, error de especificación del modelo y omisión de otras variables explicativas que influyen sobre el comportamiento de  $Y$ . Al ser  $\mu$  variable aleatoria, también lo es  $Y$ .

Se explicita una relación lineal para  $f(X)$

$$f(X) = \alpha + \beta X$$

Si se dispone de  $n$  observaciones de  $Y$  y de  $X$ , puede especificarse el siguiente modelo de regresión.

$$Y_i = \alpha + \beta X_i + \mu_i \quad i = 1, \dots, n$$

donde,

$Y$  = variable dependiente

$X$  = variable explicativa

$\mu$  = perturbación aleatoria o término de error

$\alpha, \beta$  = parámetros del modelo.

El modelo está basado en las siguientes consideraciones:

La relación especificada es lineal. Esto conduce a tener que linealizar aquellas relaciones funcionales no lineales a efectos de poder aplicar el análisis.

El término de error es una variable aleatoria donde:

- El promedio es cero.

$$E(\mu_i) = 0 \quad \forall i$$

- La varianza es constante. Esto significa que el error no es más grande para valores grandes de X que para valores pequeños de X. Esto se lo conoce como "homocedasticidad".

$$V(\mu_i) = E(\mu_i \mu_j) = \sigma^2 \quad \forall i = j$$

- Los errores son independientes entre sí. Esto significa que los errores de una observación no dependen de la observación anterior. Esto se lo conoce como "no autocorrelación".

$$\text{Cov}(\mu_i \mu_j) = E(\mu_i \mu_j) = 0 \quad \forall i \neq j$$

- El error tiene distribución normal de media cero y varianza constante.

$$\mu_j \sim N(0, \sigma^2)$$

La variable explicativa  $X$  es fija, no estocástica e independiente de los errores

$$\text{Cov}(\mu_i X_i) = E(\mu_i X_i) = 0$$

Bajo el cumplimiento de este conjunto de supuestos, los estimadores minimocuadráticos son lineales, insesgados y óptimos (ELIO) y es posible construir intervalos de confianza y realizar test para  $\hat{\alpha}$  y  $\hat{\beta}$ .

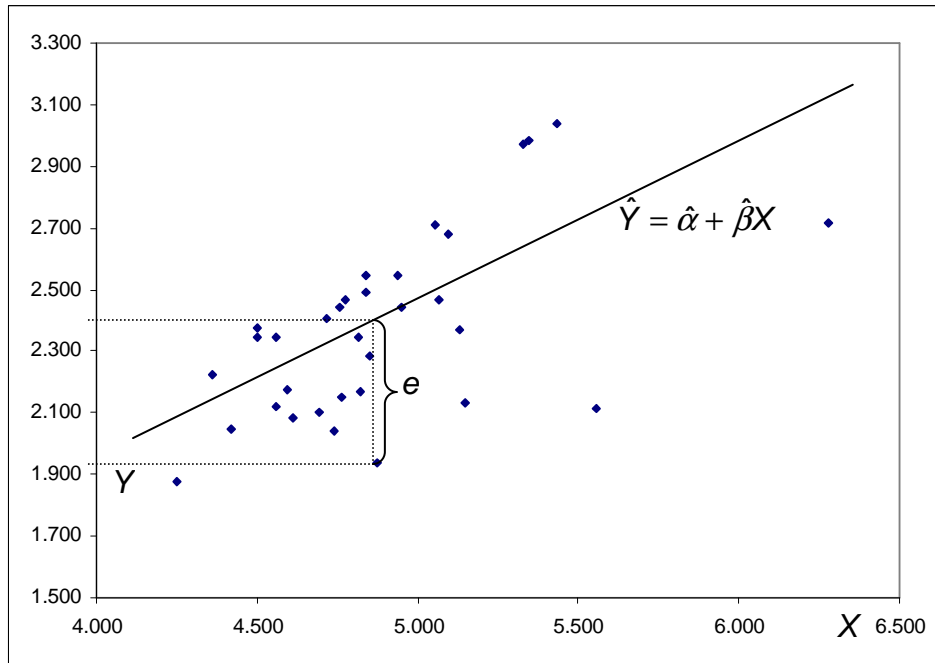
### Estimación de los parámetros del modelo

Los parámetros  $\alpha$  y  $\beta$  se estiman por el método de "mínimos cuadrados". Esto es elegir  $\hat{\alpha}$  y  $\hat{\beta}$  como estimadores de  $\alpha$  y  $\beta$ , respectivamente, de tal manera que la suma de las diferencias al cuadrado entre el valor observado y el valor estimado de la variable dependiente sea mínima. Es decir, que tenga la siguiente propiedad

$$Q = \sum_{i=1}^n (Y_i - \hat{Y}_i)^2 = \text{mínimo}$$

La gráfica muestra la dispersión de la información proveniente de la observación de dos variables X e Y. La línea recta simula el modelo estimado. La diferencia entre el valor observado y la línea estimada es el

error. Lo que se "minimiza" es la suma de cuadrados entre los puntos y la línea, en forma vertical.



Los parámetros  $\alpha$  y  $\beta$  son las características de la relación entre X e Y. Reemplazando en Q,  $\hat{Y}$  por su igual:

$$Q = \sum_{i=1}^n (Y_i - \hat{\alpha} - \hat{\beta}X_i)^2$$

Para que el valor de Q sea mínimo se debe derivar con respecto a  $\hat{\alpha}$  y  $\hat{\beta}$  e igualar a cero.

$$\frac{\partial Q}{\partial \hat{\alpha}} = 2 \sum (Y_i - \hat{\alpha} - \hat{\beta}X_i)(-1) = 0$$

$$\sum Y_i = n\hat{\alpha} + \hat{\beta} \sum X_i$$

$$\frac{\partial Q}{\partial \hat{\beta}} = 2 \sum (Y_i - \hat{\alpha} - \hat{\beta}X_i)(-X_i) = 0$$

$$\sum X_i Y_i = \hat{\alpha} \sum X_i + \hat{\beta} \sum X_i^2$$

De esta manera tenemos un sistema de dos ecuaciones y dos incógnitas denominadas ecuaciones normales. De la primera se obtiene el valor de  $\hat{\alpha}$

$$\sum Y_i = n\hat{\alpha} + \hat{\beta} \sum X_i$$

$$\hat{\alpha} = \bar{Y} - \hat{\beta} \bar{X}$$

Reemplazando ésta en la segunda se obtiene  $\hat{\beta}$

$$\begin{aligned}\sum x_i y_i &= (\bar{y} - \beta \bar{x}) \sum x_i + \hat{\beta} \sum x_i^2 \\ \sum x_i y_i &= (\bar{y} - \hat{\beta} \bar{x}) n \bar{x} + \hat{\beta} \sum x_i^2 \\ \sum x_i y_i - n \bar{x} \bar{y} &= \hat{\beta} \sum x_i^2 - \hat{\beta} n \bar{x}^2 \\ \hat{\beta} &= \frac{\sum x_i y_i - n \bar{x} \bar{y}}{\sum x_i^2 - n \bar{x}^2}\end{aligned}$$

$\hat{\alpha}$  y  $\hat{\beta}$ , se denominan **coeficientes de regresión** y se basan en la muestra aleatoria. Cada muestra aleatoria tendrá asociada un  $\hat{\alpha}$  y un  $\hat{\beta}$ , diferentes. Una medida de esta variación de las estimaciones de los parámetros está dada por sus *errores estándar*. De este modo, así como la media muestral,  $\bar{X}$ , tiene un error estándar que es estimado por  $\sigma_{\bar{X}}$ ; también  $\hat{\alpha}$  y  $\hat{\beta}$  tienen un error estándar asociado con cada uno de ellos,  $S_{\alpha}$  y  $S_{\beta}$ .

$$s_{\alpha} = \sqrt{\sigma_e^2 \left[ \frac{1}{n} + \frac{\bar{x}^2}{\sum (x - \bar{x})^2} \right]} \qquad s_{\beta} = \sqrt{\frac{\sigma_e^2}{\sum (x - \bar{x})^2}}$$

La varianza de los errores de la regresión,  $\sigma_e^2$ , se estima a partir de  $s_e^2$

$$s_e^2 = \frac{\sum (y_i - \hat{y}_i)^2}{n-2} = \frac{\sum e^2}{n-2}$$

La interpretación de los parámetros estimados

Los parámetros tienen un significado muy preciso.

- El parámetro  $\beta$  indica que, si la variable X cambia en una unidad, la variable Y cambiará en  $\beta$  unidades.
- El parámetro  $\alpha$  refleja el nivel que alcanza la variable dependiente cuando la variable explicativa asume el valor nulo.

Tal como se indicó anteriormente, la estimación de  $\hat{\beta}$ , tiene una variación asociada con ella (medida por  $S_{\hat{\beta}}$ ), porque se basa en una

muestra de individuos. Una forma de evaluar la magnitud de  $\hat{\beta}$ , tomando en cuenta su variación, consiste en usar una **prueba estadística de hipótesis**. Para esto se calcula el estadístico t,

$$t = \frac{\hat{\beta} - \beta}{s_{\hat{\beta}}}$$

Bajo la **hipótesis nula**, que será planteada como  $\beta = 0$ , el estadístico t se reduce a,

$$t = \frac{\hat{\beta}}{s_{\hat{\beta}}} \sim t_{n-2}$$

Nuevamente, como se hizo en otras pruebas de hipótesis, este valor empírico de t se compara con el valor teórico. Si el valor empírico es superior, en valor absoluto, al valor teórico se rechaza la hipótesis nula y se decide que  $\beta \neq 0$  y por lo tanto, la relación entre X e Y es significativa, a un valor de probabilidad  $\alpha$  determinado. Para construir intervalos de confianza para los parámetros de la regresión, se utiliza la distribución t, teniendo en cuenta que

$$\frac{\hat{\alpha} - \alpha}{s_{\hat{\alpha}}} \sim t_{n-2} \qquad \frac{\hat{\beta} - \beta}{s_{\hat{\beta}}} \sim t_{n-2}$$

### Predicción

El modelo de regresión, desde luego, puede ser usado como una herramienta predictiva. Dada la ecuación de regresión

$$Y = \hat{\alpha} + \hat{\beta}X$$

se pueden predecir valores futuros de Y ( $Y_F$ ) dado un valor futuro de X ( $X_F$ ) usando

$$\hat{Y}_F = \hat{\alpha} + \hat{\beta}X_F$$

El valor de  $Y_F$  viene dado por

$$Y_F = \alpha + \beta X_F + \mu_F$$

De aquí el **error de predicción** es

$$\hat{Y}_F - Y_F = (\hat{\alpha} - \alpha) + (\hat{\beta} - \beta)X_F - \mu_F$$

Como

$$\left\{ \begin{array}{l} E(\hat{\alpha} - \alpha) = 0 \\ E(\hat{\beta} - \beta) = 0 \\ E(\mu_F) = 0 \end{array} \right\} \Rightarrow E(\hat{Y}_F - Y_F) = 0$$

De donde:  $E(\hat{Y}_F) = Y_F$

Esto significa que el predictor  $\hat{Y}_F$  es insesgado.

La varianza del error de predicción es:

$$\begin{aligned} V(\hat{Y}_F - Y_F) &= V(\hat{\alpha} - \alpha) + X_F^2 V(\hat{\beta} - \beta) - 2X_F \text{cov}(\hat{\alpha} - \alpha, \hat{\beta} - \beta) + V(\mu_F) \\ &= \sigma_e^2 \left( \frac{1}{n} + \frac{\bar{X}^2}{\sum (X - \bar{X})^2} \right) + X_F^2 \frac{\sigma_e^2}{\sum (X - \bar{X})^2} \\ &\quad - 2X_F \sigma_e^2 \frac{\bar{X}}{\sum (X - \bar{X})^2} + \sigma_e^2 \\ &= \sigma_e^2 \left[ 1 + \frac{1}{n} + \frac{(X_F - \bar{X})^2}{\sum (X - \bar{X})^2} \right] \end{aligned}$$

Por lo tanto,  $V(\hat{Y}_F - Y_F)$  se incrementa a medida que  $X_F$  se aleja de  $\bar{X}$ . Como  $\sigma^2$  es desconocida se estima por:

$$s_e^2 = \frac{SCR}{n-2} = \frac{\sum e^2}{n-2}$$

De aquí, el error estándar de la predicción es

$$s_{\hat{Y}_F} = \sqrt{\frac{\sum e^2}{n-2} \left[ 1 + \frac{1}{n} + \frac{(X_F - \bar{X})^2}{\sum (X - \bar{X})^2} \right]}$$

Esta última expresión se puede utilizar para construir intervalos de confianza para  $Y_F$ .

Para tener en cuenta:

- La predicción usando valores extremos de la variable independiente puede ser arriesgada. Recuerde que el supuesto lineal puede ser apropiado solo para un rango limitado de las variables independientes. Además, la muestra aleatoria no proporcionó información acerca de los valores extremos de peso.
- Si el contexto cambia, como el hecho de comunidades diferentes, entonces, los parámetros del modelo se verán afectados. Los datos provenientes de la muestra aleatoria fueron obtenidos bajo un conjunto de condiciones de la población. Si cambian, entonces el modelo puede verse afectado.

## Coeficiente $r^2$

Para evaluar la capacidad predictiva del modelo se usa el coeficiente de determinación o  $r^2$ . Este coeficiente es la razón de la variación explicada sobre la variación total:

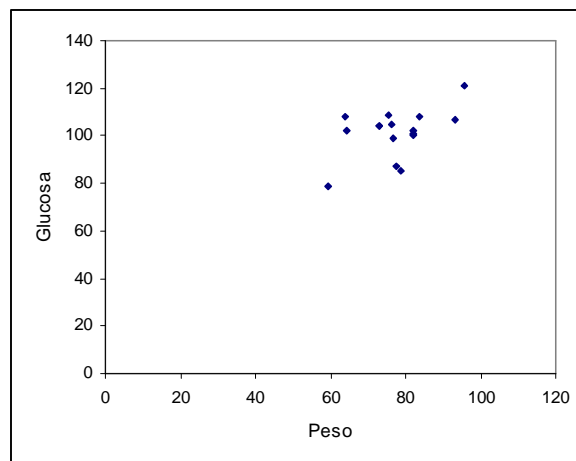
$$r^2 = \frac{\text{var iación total} - \text{var iación no explicada}}{\text{var iación total}} = \frac{\text{var iación explicada}}{\text{var iación total}}$$

$$r^2 = \frac{\sum (Y_i - \bar{Y})^2 - \sum (\hat{Y}_i - Y_i)^2}{\sum (Y_i - \bar{Y})^2} = \frac{\sum (\hat{Y}_i - \bar{Y})^2}{\sum (Y_i - \bar{Y})^2}$$

La variación total en Y se divide en la variación que explica el modelo de regresión, variación explicada, y la variación que no explica el modelo de regresión, variación no explicada.

El término  $r^2$  **es el cuadrado de la correlación entre X e Y**. Por consiguiente, se encuentra entre cero y uno.

**Ejemplo.** En un centro médico quieren estudiar, de la población de hombres adultos aparentemente sanos, la relación existente entre la concentración de glucosa en la sangre con el peso. Específicamente, están interesados en conocer si el peso de las personas determina el nivel de glucosa existente en la sangre.



Las observaciones fueron graficadas en el diagrama de dispersión, donde, cada punto de la gráfica, representa un individuo para el cual se observó el peso y el nivel de glucosa. La dispersión de la nube de puntos presenta una asociación positiva.

El siguiente paso, consiste en obtener una línea que tenga el mejor "ajuste" para estos puntos. La línea se denomina **línea de mínimos cuadrados** y se obtiene a partir de la especificación del modelo de regresión:



$$Y = \alpha + \beta X + \mu$$

donde:

$Y$ , es la variable dependiente "nivel de glucosa"

$X$ , es la variable explicativa "peso"

$\alpha$  y  $\beta$ , parámetros a estimar

$\mu$ , término de error

Se deben calcular los valores de los estimadores de los parámetros  $\alpha$  y  $\beta$  a partir de los datos consignados en la tabla. De allí se obtienen los valores promedio de las variables Peso ( $X$ ) y Concentración de glucosa ( $Y$ ).

$$\bar{X} = \frac{\sum X}{n} = \frac{1237.8}{16} = 77.3625$$

$$\bar{Y} = \frac{\sum Y}{n} = \frac{1621}{16} = 101.3125$$

Observación	Peso (en kgrs.) $X$	Glucosa (mg/100ml) $Y$	$XY$	$X^2$	$\hat{Y}$	$Y - \hat{Y}$	$(Y - \hat{Y})^2$
1	64,0	108	6912,0	4096,00	94,520	13,480	181,710
2	75,3	109	8207,7	5670,09	100,283	8,717	75,986
3	73,0	104	7592,0	5329,00	99,110	4,890	23,912
4	82,1	102	8374,2	6740,41	103,751	-1,751	3,066
5	76,2	105	8001,0	5806,44	100,742	4,258	18,131
6	95,7	121	11579,7	9158,49	110,687	10,313	106,358
7	59,4	79	4692,6	3528,36	92,174	-13,174	173,554
8	93,4	107	9993,8	8723,56	109,514	-2,514	6,320
9	82,1	101	8292,1	6740,41	103,751	-2,751	7,568
10	78,9	85	6706,5	6225,21	102,119	-17,119	293,060
11	76,7	99	7593,3	5882,89	100,997	-1,997	3,988
12	82,1	100	8210,0	6740,41	103,751	-3,751	14,070
13	83,9	108	9061,2	7039,21	104,669	3,331	11,096
14	73,0	104	7592,0	5329,00	99,110	4,890	23,912
15	64,4	102	6568,8	4147,36	94,724	7,276	52,940
16	77,6	87	6751,2	6021,76	101,456	-14,456	208,976
Sumas	1237,8	1621	126128,1	97178,60			1204,648

Se calculan los coeficientes de la regresión

$$\hat{\beta} = \frac{\sum X_i Y_i - n \bar{X} \bar{Y}}{\sum X_i^2 - n \bar{X}^2} = \frac{126128.1 - 16 * 77.3625 * 101.3125}{97178.6 - 16 * 77.3625^2} = 0.50975$$

$$\hat{\alpha} = \bar{Y} - \hat{\beta} \bar{X} = 101.3125 - 0.50975 * 77.3625 = 61.8769$$

Con estos resultados la línea de regresión estimada es

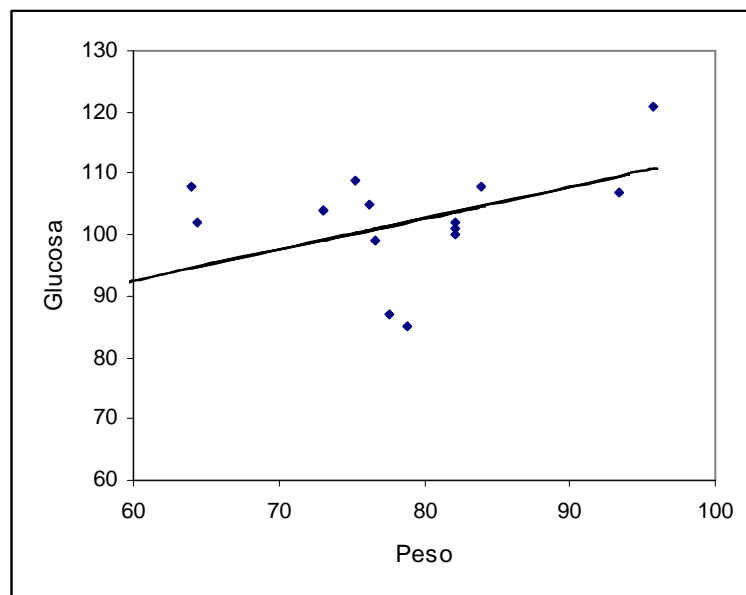
$$\hat{Y} = \hat{\alpha} + \hat{\beta}X = 61.88 + 0.51 X$$

El valor 61.88 es una estimación del parámetro  $\alpha$ , e indica el nivel de glucosa independiente del peso alcanzado por la persona. El valor 0.51, es una estimación del parámetro  $\beta$  que indica en cuánto cambia el nivel de glucosa cuando cambia el peso en una unidad.

El término  $\hat{Y}$  es una estimación del nivel de glucosa basada en el modelo de regresión, por ejemplo, cuando X es 64:

$$\hat{Y} = 61.88 + 0.51 * 64 = 94.520$$

De este modo, para un paciente de 64 Kg., la **estimación** del nivel de glucosa es de 94.52.



Si se obtuviera *otra muestra* aleatoria de 16 varones adultos, contendría individuos diferentes. Como resultado, el gráfico de dispersión XY diferiría de la figura anterior y los coeficientes de regresión,  $\hat{\alpha}$  y  $\hat{\beta}$ , serían diferentes.

La varianza de los errores y de los coeficientes se obtienen de la siguiente manera:

$$s_e^2 = \frac{\sum (Y_i - \hat{Y}_i)^2}{n-2} = \frac{\sum e^2}{n-2} = \frac{1204.648}{14} = 86.0463$$

$$\begin{aligned}
 s_{\hat{\alpha}}^2 &= \sigma_e^2 \left[ \frac{1}{n} + \frac{\bar{X}^2}{\sum (X - \bar{X})^2} \right] = \sigma_e^2 \left[ \frac{1}{n} + \frac{\bar{X}^2}{\sum X^2 - n\bar{X}^2} \right] \\
 &= 86.0463 \left[ \frac{1}{16} + \frac{77.3625^2}{97178.60 - 16 * 77.3625^2} \right] \\
 &= 368.2217
 \end{aligned}$$

$$\begin{aligned}
 s_{\hat{\beta}}^2 &= \frac{\sigma_e^2}{\sum (X - \bar{X})^2} = \frac{\sigma_e^2}{\sum X^2 - n\bar{X}^2} \\
 &= \frac{86.0463}{97178.60 - 16 * 77.3625^2} = 0.0606
 \end{aligned}$$

Con estos valores los desvíos de la regresión y los coeficientes alcanzan los siguientes valores:

$$s_e = \sqrt{86.0463} = 9.2761$$

$$s_{\hat{\alpha}} = \sqrt{368.2217} = 19.1891$$

$$s_{\hat{\beta}} = \sqrt{0.0606} = 0.2462$$

Hasta aquí se ha realizado una estimación puntual del valor de los coeficientes de la regresión. Pero es posible, como ocurre con otros parámetros, realizar una estimación por intervalo asignándole una probabilidad de ocurrencia.

$$P\left(\hat{\alpha} - t_{n-2} s_{\hat{\alpha}} \leq \alpha \leq \hat{\alpha} + t_{n-2} s_{\hat{\alpha}}\right) = 1 - 0.05$$

$$P(61.8769 - 2.1448 * 19.1891 \leq \alpha \leq 61.8769 + 2.1448 * 19.1891) = 0.95$$

$$P(20.7201 \leq \alpha \leq 103.0337) = 0.95$$

$$P\left(\hat{\beta} - t_{n-2} s_{\hat{\beta}} \leq \beta \leq \hat{\beta} + t_{n-2} s_{\hat{\beta}}\right) = 1 - 0.05$$

$$P(0.50975 - 2.1448 * 0.2462 \leq \beta \leq 0.50975 + 2.1448 * 0.2462) = 0.95$$

$$P(-0.01829 \leq \beta \leq 1.03779) = 0.95$$

Cuando se plantea el modelo de regresión, el primer supuesto es que el comportamiento de la variable X afecta el comportamiento de la variable Y. Para probar esto se contrasta la hipótesis de que el parámetro  $\beta$  es nulo. Si esto fuera así; entonces, no habría efecto del peso sobre los niveles de glucosa y el modelo no debería ser usado para ningún propósito.

$$H_0 : \beta = 0$$

El estadístico de prueba es

$$t = \frac{\hat{\beta} - \beta_{H_0}}{s_{\hat{\beta}}} = \frac{0.50975 - 0}{0.2462} = 2.07$$

Al nivel de significación de 0.05,  $\alpha = 0.05$ , el valor de la distribución t con 14 grados de libertad (valor teórico de t) es de  $\pm 2.1448$ . Esto significa que el valor de prueba cae en la zona de aceptación de la  $H_0$ , por lo que el peso alcanzado por los hombres adultos aparentemente sanos no afecta el nivel de glucosa observado en ellos.

Para predecir el valor del nivel de glucosa a partir del modelo de regresión, se supone un nivel de peso y se reemplaza en el mismo. Si se propone un peso de 80 Kg, un nivel de glucosa estimado en base al modelo, sería:

$$\hat{Y} = 61.88 + 0.51 * 80 = 102.68$$

El  $r^2$  es igual a 0.23; por consiguiente, el 23% de la variación total de Y, está explicada por X

## **CASOS DE ESTUDIO, PREGUNTAS Y PROBLEMAS**

### Caso 9.1: Función consumo

Con información suministrada por el Ministerio de Economía de la Nación, especifica un modelo de regresión lineal para estimar la función consumo en Argentina.

### Caso 9.2: Evolución del Consumo

Analiza la serie Consumo utilizada en el Caso 9.1; construye la línea de tendencia y las relativas cíclicas irregulares.

### Problemas

9.1. En el periodo Marzo 1993 y Diciembre 2003, la varianza observada en el consumo de energía eléctrica alcanzó 815151.19 Gwh y la varianza en el PBI 374.07 miles de millones de pesos a valores de 1993. La covarianza entre ambas variables es de 6204.11. ¿Cuál es la correlación entre ambas variables?

9.2. El consumo de gas observado en la Argentina en el periodo Marzo de 1993 y Diciembre de 2003 registró una variación de 510030.89 millones de metros cúbicos y la covarianza con el consumo de energía eléctrica alcanzó, en el mismo periodo 566380.53. Con el dato de varianza en el consumo de energía eléctrica de 815151.19 Gwh, calcule el coeficiente de correlación entre el consumo de Gas y el consumo de energía eléctrica.

9.3. El gerente de marketing de una cadena de autoservicio quiere determinar el efecto del espacio en las estanterías, sobre las ventas de alimentos para animales domésticos. Se seleccionó una muestra aleatoria de 12 autoservicios de igual tamaño cuyos resultados se muestran en la tabla.

- Grafique el diagrama de dispersión
- Especifique el modelo y utilice el método de mínimos cuadrados para estimar los coeficientes de la regresión
- Interprete el significado de la pendiente en este problema
- Prediga las ventas semanales promedio de alimentos para animales domésticos para un autoservicio con 8 metros de estanterías para esos alimentos.
- Calcule el error estándar de la estimación
- Calcule el coeficiente de determinación e interprete su significado
- Calcule el coeficiente de correlación
- Encuentre una estimación de intervalo con el 90% de confianza en las ventas semanales promedio de un autoservicio que tiene 8 metros de estantería para alimentos de animales domésticos
- Con un nivel de significación del 0.10, ¿hay relación lineal entre el espacio en la estantería y las ventas?

Relación entre ventas y espacio en estantería		
Autoservicio	Espacio X (metros)	Ventas semanales Y (miles de pesos)
1	5	1.6
2	5	2.2
3	5	1.4
4	10	1.9
5	10	2.4
6	10	2.6
7	15	2.3
8	15	2.7
9	15	2.8
10	20	2.6
11	20	2.9
12	20	3.1

9.4. Un estadístico de una fábrica de automóviles quiere desarrollar un modelo estadístico para predecir el tiempo de entrega (número de días entre la fecha del pedido y la fecha de entrega) de automóviles nuevos pedidos con mucho equipo opcional. El estadístico cree que hay una relación lineal entre el número de opciones pedidas y el tiempo de entrega. Se selecciona una muestra aleatoria de 16 automóviles que arroja los resultados de la tabla

- Grafique el diagrama de dispersión
- Especifique el modelo y utilice el método de mínimos cuadrados para estimar los coeficientes de regresión
- Interprete el significado de la pendiente en este problema
- Si se ordena un automóvil que tenga 16 opciones, cuántos días cree usted que tardarían para la entrega?
- Calcule el error estándar de la estimación
- Calcule el coeficiente de determinación e interprete su significado
- Calcule el coeficiente de correlación

- h) Encuentre una estimación de intervalo con el 95% de confianza para el tiempo medio de entrega de un automóvil equipado con 16 opciones
- i) Con un nivel de significación del 0.05, ¿hay relación lineal entre el número de opciones y el tiempo de entrega?

Relación entre tiempo de entrega y opciones ordenadas		
Automóvil	Número de opciones X	Tiempo de entrega Y
1	3	25
2	4	32
3	4	26
4	7	38
5	7	34
6	8	41
7	9	39
8	11	46
9	12	44
10	12	51
11	14	53
12	16	58
13	17	61
14	20	64
15	23	66
16	25	70

9.5. Un ingeniero agrónomo quiere determinar el efecto de un fertilizante orgánico natural sobre el rendimiento de tomates. Se van a utilizar cinco cantidades diferentes de fertilizantes en 10 lotes equivalentes. Los niveles de fertilizante se asignan en forma aleatoria a los lotes arrojando los resultados de la tabla.

Relación entre rendimiento y nivel de fertilizante aplicado		
Lote	Cantidad de fertilizante X (kilos/hectárea)	Rendimiento Y (kilos)
1	0	6
2	0	8
3	10	11
4	10	14
5	20	18
6	20	23
7	30	25
8	30	28
9	40	30
10	40	34

- a) Grafique el diagrama de dispersión
- b) Especifique el modelo y utilice el método de mínimos cuadrados para estimar los coeficientes de regresión
- c) Interprete el significado de la pendiente en este problema
- d) Prediga el rendimiento de tomates para un lote al cual se le han aplicado 15 kilos por hectárea de fertilizante orgánico natural.
- e) Calcule el error estándar de la estimación
- f) Calcule el coeficiente de determinación e interprete su significado
- g) Calcule el coeficiente de correlación
- h) Encuentre una estimación de intervalo con el 90% de confianza para el rendimiento promedio de tomates fertilizados con 15 kilos por hectárea de fertilizante orgánico natural
- i) Con un nivel de significación del 0.05, ¿hay relación lineal entre la cantidad de fertilizante utilizado y el rendimiento de tomates?

---

---

### **Bibliografía**

- Berenson, Mark y Levine, David. (1996). *Estadística Básica en Administración*. Prentice Hall. México.
- Dagum, C. y Dagum, E.M.de. (1971). *Introducción a la Econometría*. Editorial Siglo XXI. México.
- Daniel, Wayne W. (1999). *Bioestadística, base para el análisis de las ciencias de la salud*. Tercera Edición. Editorial Limussa. México.
- Dixon W.J. y Massey F.J. (1957). *Introduction to Statistical Analysis*. Nueva York, McGraw-Hill.
- Graybill, Mood. *Introducción a la Teoría de la Estadística*.
- Gujarati, D. (2004). *Econometría*. 4ª Edición. Mc.Graw Hill. México.
- Hildebrand, D. y Lyman Ott R. (1997). *Estadística Aplicada a la administración y a la economía*. Addison Wesley. Wilmington USA.
- Johnston, J. y Dinardo, J. (2001). *Métodos de Econometría*. Editorial Vicens Vives. Barcelona.
- Kazmier, Leonard y Diaz Mata, Alfredo. (1993). *Estadística Aplicada a la Administración y a la Economía*. Mc.Graw Hill. México.
- Meyer, P.L. (1973). *Probabilidad y Aplicaciones Estadísticas*. Fondo Educativo Interamericano. México.
- Padua, Jorge. (1996). *Técnicas de Investigación Aplicadas a las Ciencias Sociales*. Fondo de Cultura Económica. México.
- Tramutola, C.D. *Modelos Probabilísticos y Decisiones Financieras*. E.C.Moderan. Lectura de Administración de Empresas. Capital Federal.

Capítulo 10. SIMULACIÓN EN CONTEXTOS CON Y SIN INFORMACIÓN HISTORICA .....	311
A. Contexto general de la predicción y simulación en economía .....	311
A.1. Predicción, futuro y toma de decisiones .....	311
A.2. Predicción económica y empresarial: campos de aplicación .	312
A.3. Centros mundiales y fuentes de predicción económica y empresarial .....	315
A.4.- Clasificación de las técnicas de predicción .....	322
A.5. Tipología: horizonte de predicción, medios e información ....	326
A.6. Cinco reglas sobre predicción .....	329
A.7. Fundamentación científica de la predicción: antecedentes históricos.....	333
A.8. Etapas genéricas del proceso de predicción/simulación .....	334
B. Técnicas Elementales de Predicción.....	338
B.1. Predicción elemental sin información histórica .....	338
B.2. Información histórica: componentes de una serie económica .....	340
B.3. Medias móviles .....	355
B.4. Alisado exponencial sin tendencia: el alisado simple .....	361
B.5. Alisados con tendencia .....	368
B.6. Ajuste con funciones matemáticas .....	379
B.7. Predicción en series con componente estacional.....	384
B.8. ¿Qué técnica utilizar? .....	398



C. La simulación en ausencia de información histórica .....	411
C.1. Simulación mediante fórmulas recursivas.....	411
C.2. El método Delfos.....	420
CASOS DE ESTUDIO, PREGUNTAS Y PROBLEMAS .....	428
Bibliografía .....	433

## Capítulo 10. SIMULACIÓN EN CONTEXTOS CON Y SIN INFORMACIÓN HISTORICA

### A. Contexto general de la predicción y simulación en economía

■ En este acápite se pretende describir una mínima panorámica del marco en el que se desenvuelve la predicción y simulación económica y empresarial. Describiremos la forma en que la predicción apoya el proceso de toma de decisiones en una amplia variedad de campos y, en especial, en las áreas de la economía y la gestión de empresas. Señalaremos algunos de los centros que realizan predicción, y ejemplificaremos el funcionamiento de uno de estos centros, con detalles sobre la práctica profesional de la predicción. Nos detendremos especialmente en cuáles son las posibles técnicas que podemos utilizar en la predicción, muchas de las cuales serán objeto de atención en el curso, y apuntaremos detalles sobre cómo y cuándo utilizarlas óptimamente. También mostraremos algunas referencias históricas fundamentales.

■ Este acápite se divide en 8 apartados, cada uno de ellos explicado a través de su correspondiente ficha de resumen y lecturas adicionales. Se propone la realización de 2 actividades. El estudio de esta unidad puede exigir de unas 10 horas: 5 para los conceptos teóricos y otras 5 para la realización de actividades, lecturas adicionales y la realización del test de autoevaluación.

#### A.1. Predicción, futuro y toma de decisiones

##### Resumen:

- ✚ El proceso de toma de decisiones lleva aparejada una cierta valoración del futuro.
- ✚ Tomar decisiones basándose en datos actuales es una simplificación inadmisibile.
- ✚ Conceptualmente el proceso de predicción puede sistematizarse.
- ✚ Las predicciones implicadas en el proceso no tienen que ser únicas ni permanentes.
- ✚ En el proceso formal de decisión deben considerarse predicciones alternativas.

#### **Predicción, futuro y toma de decisiones**

La toma de decisiones en la gestión de las empresas y de la política económica conlleva inexcusablemente la necesidad de *valorar el futuro*. Ese proceso de toma de decisiones puede sistematizarse, por ejemplo, en las siguientes etapas:

---

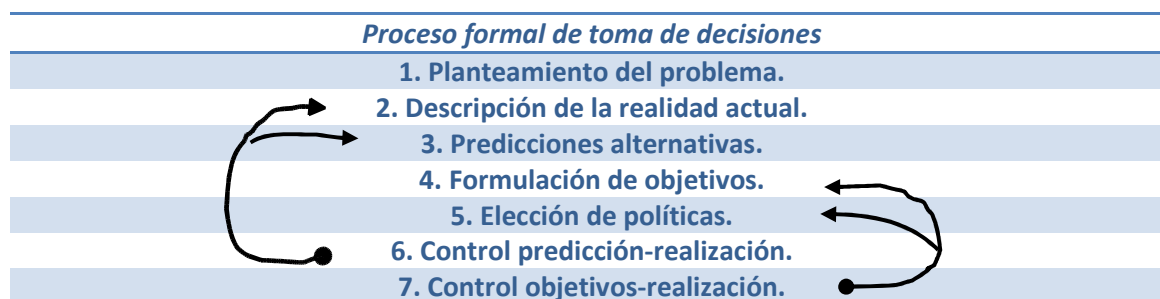
**Etapas conceptuales del proceso de toma de decisiones**


---

1. Planteamiento del problema.
  2. Valoración de la situación actual.
  3. Consideración de las alternativas de futuro.
  4. Formulación de objetivos.
  5. Elección de políticas.
  6. Control de realización del proceso.
- 

Obviamente, en la gestión económica y empresarial se hace imprescindible considerar la *evolución futura* de los elementos que condiciona tal decisión. Por ello, se considera que tomar decisiones basándose en datos actuales es una simplificación inadmisibile del problema. Ahora bien, no es menos cierto que esta consideración del futuro no debe entenderse como una *cuantificación específica* de aspectos concretos de ese futuro, sino que en la mayor parte de las decisiones será suficiente con una valoración *aproximada*. También es cierto que esta sistematización del proceso de toma de decisiones no debe conllevar la existencia de una única predicción, sino que el proceso se enriquece, por propia definición, con la existencia de *alternativas*, y de un proceso de revisión permanente de las predicciones, resultado de cambios en los objetivos, o en los elementos que condicionan la decisión.

Por tanto, se puede considerar que el proceso formal de toma de decisiones en la gestión económica y empresarial podría representarse de la siguiente forma:



### A.2. Predicción económica y empresarial: campos de aplicación

#### Resumen:

- ✚ La necesidad de predecir se da en las áreas más diversas, desde la economía, la meteorología, la política o la defensa militar.
  - ✚ Aunque las técnicas pueden ser comunes a esas áreas cada campo genera problemas específicos.
  - ✚ Sin embargo, las cuestiones relacionadas con la predicción económica y empresarial deben tratarse de forma conjunta.
  - ✚ En la actualidad la predicción se aplica a multitud de temas dentro de la economía y la gestión de empresas.
-

## **Campos de aplicación de la predicción y ejemplos en economía y gestión de empresas**

Como no podría ser de otra manera, la necesidad de realizar predicciones se da en las áreas más diversas; desde las más divulgadas de la economía y la meteorología a las más reservadas de la política, la ecología o la defensa militar. En el cuadro siguiente se recogen algunos de los campos en los que se utilizan técnicas de predicción.

<i>Campos de aplicación de las técnicas de predicción</i>
<b>Salud</b>
<ul style="list-style-type: none"> <li>- Sida.</li> <li>- Utilización de fármacos.</li> <li>- Efectos de nuevos productos farmacéuticos.</li> <li>- Expectativas de vida.</li> </ul>
<b>Ecología</b>
<ul style="list-style-type: none"> <li>- Evolución del ecosistema europeo.</li> <li>- Flujo de agua de un río.</li> <li>- Perturbaciones en bosques.</li> <li>- Hidrología</li> <li>- Cambios en la vegetación natural.</li> <li>- Niveles de contaminación atmosférica.</li> <li>- Establecimiento de un plan estratégico del medio ambiente.</li> <li>- Capturas, tasas de mortalidad y bancos de peces.</li> </ul>
<b>Política</b>
<ul style="list-style-type: none"> <li>- Conflicto Israel-OLP.</li> <li>- Popularidad de la función de presidente de Gobierno.</li> <li>- Riesgo político por países.</li> </ul>
<b>Tecnología</b>
<ul style="list-style-type: none"> <li>- <i>Sustitución de tecnologías.</i></li> <li>- Nuevas tecnologías en energía e información.</li> <li>- Superconductividad.</li> <li>- Evolución tecnológica en países en vías de desarrollo.</li> <li>- Impacto social de la tecnología.</li> <li>- Control de calidad.</li> </ul>
<b>Sociedad</b>
<ul style="list-style-type: none"> <li>- Número de alumnos de universidades.</li> <li>- Costos de construcción.</li> <li>- Política de viviendas.</li> <li>- Funcionamiento de edificios.</li> </ul>
<b>Meteorología</b>
<ul style="list-style-type: none"> <li>- Relación entre lluvia y radiación solar.</li> <li>- Temperaturas.</li> <li>- Intervalos de tiempo seco.</li> </ul>
<b>Demografía</b>
<ul style="list-style-type: none"> <li>- Población mundial.</li> <li>- Nivel de fertilidad.</li> <li>- Movimientos de población.</li> <li>- Formación y disolución de familias</li> </ul>

---

**Campos de aplicación de las técnicas de predicción (Continuación)**
**Temas militares**

- Ciclo de vida del armamento.
- Necesidades de personal militar.
- Sistema de gestión de las fuerzas aéreas.
- Alternativas para valoración y riesgo de estrategias político-militares.
- Juegos de guerra.
- Cambios en el papel de las fuerzas armadas.

Fuente; Elaboración propia a partir de las comunicaciones presentadas en The Eighth International Symposium on Forecasting, Ámsterdam, junio de 1988.

---

Con frecuencia, las técnicas de predicción utilizadas, aun en áreas tan dispares, son las mismas o comparten una base metodológica común. Sin embargo, cada campo de aplicación genera a su vez problemas, metodologías y enfoques específicos. Aun con todo, resulta conveniente tratar en forma conjunta las cuestiones de predicción en economía general y las de carácter específicamente empresarial. Entre las razones que avalan tal proceder están:

- 1) La relevancia de las predicciones macroeconómicas para el mundo de la empresa.
- 2) La posibilidad de desagregar sectorialmente las predicciones macroeconómicas.
- 3) La existencia de campos de predicción comunes a la actuación de empresas y Administraciones Públicas (por ejemplo, tipos de interés, tipos de cambio, cotizaciones de acciones, etc.)

A continuación recogemos un repertorio de temas de predicción económica y empresarial clasificados según el ámbito de aplicación.

---

**Ejemplos de aplicaciones de la predicción en economía y gestión de empresas**
**1.- Economía internacional**

- Tasas de crecimiento comparativas.
- Cuentas nacionales agregadas.
- Comercio internacional.
- Relación entre las cotizaciones de las acciones en diferentes países.
- Conflictos, mediación internacional y desarrollo económico.
- Incidencias de nuevas tecnologías de informática, biotécnica y nuevos materiales.
- Financiación internacional y riesgo-país

**2.- Economía nacional**

- Renta y gasto.
- Relación entre ondas largas y otros ciclos macroeconómicos.
- Desempleo de larga duración.
- Ciclos de negocios y ciclos políticos.
- Confianza empresarial e ilusión monetaria.
- Pensiones por jubilación.

**3.- Economía regional**

- Empleo y salarios de una región.
  - Consumo regional de electricidad.
  - Comercio interregional.
  - Indicadores económicos adelantados por regiones.
  - Turismo regional.
-

<b>4.- Economía sectorial</b>
<ul style="list-style-type: none"> <li>- Costos de la construcción.</li> <li>- Desarrollo sectorial comparado.</li> <li>- Impacto de nuevas tecnologías sobre el tráfico telefónico.</li> <li>- Consumo de energía.</li> </ul>
<b>5.- Mercados financieros</b>
<ul style="list-style-type: none"> <li>- Tipos de cambio y premio de riesgo.</li> <li>- Cotizaciones de acciones y obligaciones.</li> <li>- Riesgo bancario.</li> </ul>
<b>6.- Economía inter-empresarial</b>
<ul style="list-style-type: none"> <li>- Transferencia de información entre empresas.</li> <li>- Difusión de nuevas tecnologías en un mercado competitivo.</li> <li>- Competencia y <i>cash-flow</i>.</li> </ul>
<b>7.- Gestión de empresas</b>
<ul style="list-style-type: none"> <li>- Management information systems.</li> <li>- Planificación estratégica corporativa.</li> <li>- Resultados económicos.</li> </ul>
<b>8.- Comercialización</b>
<ul style="list-style-type: none"> <li>- Ventas por productos.</li> <li>- Mercado potencial de nuevos productos.</li> <li>- Control de información alternativa sobre evolución de las ventas.</li> </ul>
<b>9.- Financiación</b>
<ul style="list-style-type: none"> <li>- Planificación financiera</li> <li>- Evolución de estados financieros.</li> <li>- Rentabilidad bancaria.</li> </ul>
<b>10.- Producción</b>
<ul style="list-style-type: none"> <li>- Planificación de la producción.</li> <li>- Control de inventarios.</li> <li>- Sustitución de tecnologías</li> </ul>
<b>Fuentes bibliográficas utilizadas y más información en; Pulido, A. y A. López, (1999), <i>Predicción y simulación aplicada a la economía y gestión de empresas</i>, Pirámide, Madrid.</b>

### A.3. Centros mundiales y fuentes de predicción económica y empresarial

#### Resumen:

- ✚ La amplia variedad de centros que se dedican a realizar predicciones puede ser clasificado de la siguiente manera:

#### Tipología de centros de predicción

Tipos	Ejemplos
<b>1.- Organismos internacionales</b>	<b>OCDE</b>
<b>2.- Proyectos internacionales</b>	<b>LINK (Toronto, Filadelfia, Nueva York)</b>
<b>3.- Empresas multinacionales</b>	<b>Wefa Group (Filadelfia y diversos centros en Estados Unidos y otros países).</b>
<b>4.- Organismos públicos</b>	<b>State Economic Information Center (Beijing, China).</b>
<b>5.- Bancos centrales</b>	<b>Bank of Russia (Moscú)</b>
<b>6.- Universidades y sus centros asociados</b>	<b>Cambridge Econometrics (Reino Unido)</b>
<b>7.- Instituciones financieras privadas</b>	<b>Standard &amp; Poor's (14 centros internacionales en Europa, América y Asia-Pacífico)</b>

- ✚ A continuación se muestran un conjunto más amplio de instituciones que realizan predicciones sobre España y EE.UU.
- ✚ La predicción empresarial exige la utilización de predicciones, normalmente externas, sobre economía internacional, nacional, sectorial o sobre mercados concretos.
- ✚ En la Actividad 1 se pide al lector que realice un pequeño informe a partir de las predicciones reales de varias fuentes de predicción.

### Centros mundiales y fuentes de predicción económica y empresarial

*Instituciones que predicen en España y EE.UU.*

España	
Analistas Financieros Internacionales (AFI)	Banco Bilbao Vizcaya Argentaria (BBVA)
Banco Santander Central Hispano (BSCH)	Centro de Predicción Económica (CEPREDE)
Comisión Europea	Consensus Economic, Inc.
FMI	FUNCAS
Goldman Sachs	Instituto de Crédito Oficial (ICO)
Instituto de Estudios Económicos (IEE)	JP Morgan
La Caixa	LINK
Merrill Lynch	Ministerio de Economía
OCDE	Santander Investment
The Economist	

Estados Unidos	
Amoco Corporation	Bank America Corp
Brown Brothers	Comisión Europea
Consensus Economic, Inc.	Crédit Suisse First Boston
Daimler Chrysler	DuPont
Eaton Corporation	Fannie Mae
First Union Corp.	FMI
Ford Motor Corp.	Georgia State University
Goldman Sachs	Griggs & Santow
Harriman	Inforum-Univ. of Maryland
JP Morgan	LINK
Merrill Lynch	Nat. Assn. of Home
Builders	OCDE
Oxford Economics	Prudential Insurance
Regional Financial Assocs.	Standard & Poor's
The Conference Board	The Economist
Unites States Trust	Univ. of Michigan-RSQE
WEFA Group	Wells Fargo Bank

*Algunas referencias de informes de predicción*FONDO MONETARIO INTERNACIONAL, *World Economic Outlook*<http://www.imf.org/external/pubind.htm>

Informes semestrales, normalmente en abril y noviembre. Predicción con gran desagregación de variables, países y grandes áreas geográficas. Incluye además análisis de temas de actualidad.

COMISIÓN EUROPEA, *European Economy*[http://europa.eu.int/comm/economy\\_finance/index\\_en.htm](http://europa.eu.int/comm/economy_finance/index_en.htm)

Informes semestrales en primavera y otoño. Predicciones para los países de la Unión, con amplia desagregación. Incluye también una buena base histórica.

MINISTERIO DE ECONOMÍA, *Previsiones Macroeconómicas; The Spanish Economy; y Actualización del Programa de Estabilidad*<http://www.mineco.es/sgpc/405SGPCM.htm>

Predicciones oficiales, pero sin actualización periódica determinada. Interesante porque muestra la visión gubernamental de la marcha de la economía.

ORGANIZACIÓN PARA LA COOPERACIÓN Y EL DESARROLLO ECONÓMICO, *OECD Economic Outlook*<http://www.oecd.org/publications/>

Amplia desagregación por países e indicadores. Con análisis de actualidad

INSTITUTO FLORES DE LEMUS DE ESTUDIOS AVANZADOS DE ECONOMÍA (UNIVERSIDAD CARLOS III),

*Boletín Inflación y Análisis Macroeconómico*<http://www.uc3m.es/uc3m/inst/FL/IFL.HTM><http://www.uc3m.es/uc3m/inst/FL/boletin/index.html>

Predicción de un número reducido de variables. Predicciones limitadas a suscritos. También se puede acceder al informe de la red European Forecasting Network sobre la economía europea, en el que también participa el grupo AQR de la Universidad de Barcelona.

BBVA, *Situación España*<http://www.bbva.es>

(Servicio de Estudios/situación)

Informe mensual de cierta tradición. Incluye informe de coyuntura, predicciones de un cuadro básico y notas de actualidad.

COMMERZBANK, *Economic Research*<https://www.commerzbank.com/>

(Research/Economic Research)

Predicciones sobre los principales indicadores económicos de varios países, incluido España

FUNCAS, *Previsiones FUNCAS; Panel de Previsiones*<http://www.funcas.ceca.es/>

(Indicadores/Coyuntura Nacional e internacional)

Se incluye predicción sobre un cuadro macroeconómico de actualización bimensual básico. El panel incluye una síntesis de las predicciones de varias instituciones

***¿Cómo funciona un centro de predicción? La experiencia CEPREDE– Instituto L.R. Klein (Centro Stone).*****CENTRO DE PREDICCIÓN ECONÓMICA (CEPREDE)**

En 1977 un grupo de profesores universitarios del entonces Departamento de Econometría de la Empresa e Informática de la Universidad Autónoma de Madrid toma contacto con la Wharton Econometric Forecasting Associates, Inc, (WEFA), institución perteneciente a la Universidad de Pennsylvania,



(fundada por el premio Nobel de Economía 1980 Lawrence R. Klein), para la realización de investigaciones conjuntas, fundamentalmente la elaboración de un modelo econométrico de la economía española, en línea con las experiencias que comenzaban a darse en otros países del entorno. Nació el *Proyecto de Predicción Económica (PPE)*.

Los contactos entre ambas instituciones, WEFA y UAM, espontáneos en sus inicios, se formalizan en 1980, cuando el Comité Conjunto Hispano-Norteamericano para Asuntos Educativos y Culturales aprueba el proyecto de cooperación entre ambas instituciones, teniendo como directores a Lawrence Klein y Antonio Pulido, por cada parte. Como resultado de esta colaboración y tras varias etapas preparativas se crea la primera versión del modelo econométrico Wharton-UAM, que habría de constituirse en el eje nuclear de la actividad de predicción de CEPREDE.

Las actividades de CEPREDE se inician formalmente en junio de 1981 con la celebración de su I Junta Semestral de Predicción, si bien entonces sin entidad jurídica propia, sino como un proyecto más de la Fundación Universidad-Empresa y con el apoyo inicial de la Cámara de Comercio e Industria de Madrid. No sería hasta 1985 cuando se constituye el Centro de Predicción Económica (CEPREDE) como asociación sin fines de lucro promovida por 5 instituciones que actuaban como fundadores (Cámara de Comercio, Fundación Universidad-Empresa, Fundesco, Feuga y Fundación Banco Exterior). A ese grupo de miembros fundadores se fueron añadiendo muy diversas empresas, organismos de la Administración y agrupaciones empresariales, en calidad de utilizadores de sus servicios de predicción ("asociados" en la terminología CEPREDE).

Desde entonces, y anualmente, se ha venido firmando un convenio de colaboración entre CEPREDE y la Universidad Autónoma de Madrid, por el que se encarga la dirección técnica del proyecto al Instituto de Predicción Económica L.R. Klein, Instituto Universitario de la Universidad Autónoma de Madrid, bajo la responsabilidad inmediata de su director, Antonio Pulido, por el cual la Universidad Autónoma de Madrid cede instalaciones en la Facultad de Ciencias Económicas a CEPREDE, quien financia parte de las investigaciones que se desarrollan en el Instituto L.R.Klein.

En la actualidad, CEPREDE se ha convertido en la institución española de más prestigio en materia de predicción económica, acumulando una experiencia de más de 20 años.

CEPREDE lanza, periódicamente, predicciones sobre la economía española, utilizando para ello las estimaciones del modelo Wharton-UAM, modelo econométrico que en el momento actual cuenta con más de 1.200 ecuaciones que describen el funcionamiento de la economía española.

Varios son los canales a través de los cuales CEPREDE da a conocer su actividad. El más importante es la *Junta Semestral de Predicción*, reuniones organizadas de forma semestral (mayo y noviembre), en las que participan las empresas e instituciones asociadas, y en la que se dan a conocer varios informes de predicción. Las Juntas Semestrales de Predicción son el principal punto de encuentro entre los distintos colectivos integrados en CEPREDE, y tienen una duración de un día y medio, estructuradas en torno a distintas sesiones, tanto de exposición de los análisis realizados por el

equipo de CEPREDE u otras Instituciones nacionales e internacionales, como de debate abierto de los resultados presentados, donde la visión técnica se contrasta con la realidad experimentada por los distintos representantes de las empresas e instituciones. Adicionalmente, estas Juntas sirven como punto de encuentro de especialistas en el análisis económico de diversas empresas y ramas de actividad, permitiendo el intercambio de experiencias y puntos de vista, en muchas ocasiones comunes, y en otros incluso contrapuestos.

En la actualidad CEPREDE cuenta con una treintena de investigadores distribuidos en 5 áreas organizativas; predicción, internacional, regional, sectorial, y nueva economía, además de un área de administración.

Las empresas e instituciones de CEPREDE reciben periódicamente varios informes de predicción y de análisis económico. Quincenalmente, se publican las *Notas de Alerta*, breves comentarios sobre algún tema de actualidad que necesite de rápido análisis. Mensualmente se edita el *Informe Mensual*, con análisis de coyuntura, últimos datos publicados, comentarios de actualidad económica, y predicción de algunos de las principales variables macroeconómicas de la economía española y, en menor medida, internacionales. Los informes principales de predicción se publican semestralmente con ocasión de las Juntas Semestrales. El informe principal de predicción es el *Informe Semestral. Predicciones, alternativas y riesgos*. En este informe se analizan con gran detalle la salida del modelo econométrico Wharton-UAM, con predicciones de cerca de 300 variables de la economía española en un horizonte de 5 años. El informe *Perspectivas económicas y empresariales* recoge de manera más simplificada las predicciones elementales de CEPREDE, junto con análisis de coyuntura y, sobre todo, el informe semestral del Director General del Centro, Antonio Pulido.

De cara a la difusión internacional de la actividad de CEPREDE, se elabora el Spanish Economic Outlook. Anualmente, CEPREDE publica el *Panorama Sectorial*, con predicciones y análisis en desagregación sectorial, y el *Panorama Regional*. Además, también con frecuencia anual se publican un informe monográfico, Informe especial CEPREDE, sobre algún tema que requiera de un análisis exhaustivo, por ejemplo el más reciente sobre el impacto de la ampliación de la U.E.

La actividad de difusión de los trabajos de predicción, se completa con asesoramiento *on-line* en materia de predicción, así como la elaboración de informes *ad hoc* en función de las necesidades de sus asociados, y la posibilidad de adquirir el Modelo MICROUAM, una "maqueta" a escala del modelo Wharton-UAM, que permite a las empresas la realización de simulaciones. Toda la información de CEPREDE puede adquirirse por los no asociados a través del programa Forest-Forecast & Strategy Club.

Las investigaciones de CEPREDE participan, además en los proyectos de predicción Link de las Naciones Unidas (mundial), EUREN (europeo) y, a través del Instituto L.R.Klein-Centro Stone, en HISPALINK (regional de España).

## **INSTITUTO DE PREDICCIÓN ECONÓMICA LAWRENCE R. KLEIN-CENTRO STONE**

El Instituto de Predicción Económica *Lawrence R. Klein*, es un Instituto Universitario aprobado provisionalmente por el Consejo Social y la Junta de Gobierno de la Universidad Autónoma de Madrid en 1992 y definitivamente por la Comunidad de Madrid en 1999. Hasta 1992, existía el Centro Lawrence R. Klein como una unidad investigadora aprobada por la UAM, a efectos de dar continuidad a las investigaciones iniciadas mediante convenio entre la UAM y la Asociación Centro de Predicción Económica CEPREDE. Desde entonces son múltiples los convenios de colaboración firmados con gran cantidad de Empresas e instituciones públicas y privadas, nacionales e internacionales.

El Instituto *Lawrence R. Klein* está dedicado a la figura científica y humana del profesor Klein, Premio Nobel de Economía 1980, considerado mundialmente como el padre de la econometría moderna, Doctor *honoris causa* por la UAM e impulsor de las actividades investigadoras de un grupo de profesores de la Universidad desde 1977.

Posteriormente, en 2001, se acordó incorporar al Instituto la figura de Sir Richard Stone, y añadir su nombre a uno de los dos centros que lo componen: "Instituto L.R. Klein – Centro Stone".

El Instituto está especializado en modelización económica aplicada y sus áreas actuales de actividad son: Área de predicción, Área Regional, Área Sectorial, Área de Economía Digital e innovación y Área de nuevos proyectos. Además, cuenta con el *Fondo de Investigación e Innovación Richard Stone* constituido para potenciar la actividad investigadora básica y aplicada y la difusión académica de sus resultados.

En la actualidad, el Instituto cuenta con un equipo de cerca de 40 personas entre Profesores, Investigadores, becarios de investigación, colaboradores, y personal de administración. Además, varios alumnos de Licenciatura y doctorado de la Universidad realizan prácticas en el mismo. Al Instituto Klein-Centro Stone acuden gran número de instituciones públicas y compañías privadas, para la realización de proyectos de análisis económico. Varios de ellos se enmarcan en el campo de la predicción económica.

Además, como Instituto Universitario, el Instituto organiza un Programa de Doctorado en *Modelización Económica Aplicada*, centrado en los temas de investigación objeto de su actividad. Los alumnos de Doctorado se incorporan habitualmente a los proyectos en marcha, bajo la supervisión de tutores efectivos y contando con la utilización de bancos de datos, programas de ordenador, desarrollos metodológicos y antecedentes de aplicaciones del Instituto.

Vamos a referirnos ahora a la **estrategia de modelización** que se desarrolla en el Instituto a efectos de predicción. El Instituto L.R. Klein-Centro Stone, sigue un proceso continuo que va desde el entorno internacional al nacional y regional, con diversos niveles de desagregación sectorial. En todos los casos, y como es habitual en la experiencia de otros centros de predicción mundiales se utilizan los resultados tanto de modelos como otras técnicas de análisis de datos, (en particular, de seguimiento de la coyuntura), opiniones de expertos y la referencia de las valoraciones de

situación actual y perspectivas de futuro elaboradas por otros centros. En el cuadro siguiente se sintetiza esta estrategia:

Estrategia de predicción utilizada por el Instituto L.R.Klein–Centro Stone, UAM		
Ámbito	Modelos en funcionamiento	Otras fuentes de predicción
Internacional	- Proyecto LINK (unos 85 modelos de países y uno de congruencia).	- Diversos centros de predicción mundiales
U.E.	- Red EUREN	- Bureau Federal du Plan. Belgica. - Centro Studi Confindustria. - London Business School–Oxford Economic Forecasting. - Centro de Predicción Económica. - Rheinisch-Westfälischens Institut für Wirtschaftsforschung. - Centre d’Observation Economique de la Chambre de Commerce et d’Industrie de Paris
Nacional	- Modelo Anual WHARTON-UAM. - Modelo trimestral. - Modelo MICRO-UAM. - Modelo Sectorial Hermes	- Seguimiento de la coyuntura a partir de indicadores. - Panel de expertos sectoriales. - Diversos centros de predicción
Regional	- Proyecto HISPALINK (un modelo por cada Comunidad Autónoma y uno de congruencia)	- Seguimiento de la coyuntura a partir de indicadores. - Equipos regionales del proyecto Hispalink

Centrándonos en los modelos econométricos en funcionamiento, cada país perteneciente al proyecto LINK mantiene el suyo propio, excepto en el caso de las áreas menos desarrolladas en que, en muchas ocasiones, lo realiza el propio equipo central. El correspondiente a España, el modelo *Wharton-UAM* de CEPREDE, fue desarrollado en colaboración con el equipo del profesor Klein de la Universidad de Pennsylvania. Cuenta con 1200 ecuaciones repartidas entre siete bloques que permiten predecir desde tipos de cambio hasta valores añadidos y empleo. Como un modelo de estas dimensiones resulta de una indudable complejidad, tanto para manejar como para interpretar sus resultados, se dispone de una versión más reducida con sólo 74 ecuaciones, una especie de maqueta a pequeña escala del modelo original. Con ambos modelos (*Wharton-UAM* y *MICRO-UAM*) se realizan predicciones para los próximos años, actualizadas al menos semestralmente, y que cubren un horizonte de cinco años a futuro.

Como ocurre en todo modelo, sus resultados dependen fundamentalmente de los supuestos sobre el entorno internacional y sobre otros aspectos *exógenos* tales como ciertos acontecimientos sociopolíticos que afectan a actuaciones económicas concretas (por ejemplo, a la elaboración y aprobación de los Presupuestos).

A efectos de valorar con detalle la evolución a más corto plazo, el Instituto Klein–Centro Stone dispone de un modelo trimestral en el que las predicciones se actualizan para los restantes trimestres del año en curso y las correspondientes al año siguiente.

Por último, desde sus comienzos en 1988, el Instituto Klein–Centro Stone participa en el proyecto HISPALINK, de modelización regional integrada, y en el que participan 17 universidades españolas. Cada equipo elabora y mantiene su propio modelo, que se alimenta con predicciones del entorno internacional y nacional para el conjunto de España. Cada seis meses se aplica un modelo de congruencia para integrar las predicciones de las 17 regiones, con detalle para nueve grandes sectores económicos, tanto para el año en curso como para el siguiente.

#### A.4.- Clasificación de las técnicas de predicción

##### Resumen:

- ✚ Obviamente, existen multitud de posibles técnicas para llevar a cabo la predicción.
- ✚ Un criterio clasificador de las técnicas, podría ser en función del tipo de información, distinguiendo 3 enfoques.

#### **Enfoques y técnicas de predicción**

Enfoque A: Predicción según información subjetiva	
<b>Datos suministrados por individuos aisladamente o como componentes de un grupo basándose en sus experiencias, opiniones y expectativas de futuro</b>	
<ul style="list-style-type: none"> <li>- Encuestas de intenciones o expectativas</li> <li>- Encuestas de actitudes o sentimientos</li> <li>- Agregación de predicciones individuales</li> </ul>	<ul style="list-style-type: none"> <li>- Consenso de grupo en interacción abierta</li> <li>- Panel de expertos con retroacción controlada de la información</li> <li>- Otras técnicas exploratorias y normativas</li> </ul>
Enfoque B: Predicción según el análisis aislado de series	
<b>La característica clave de este enfoque es el estudio de un fenómeno en sí mismo, a través de su evolución temporal (series temporales o históricas).</b>	
<ul style="list-style-type: none"> <li>- Fórmulas ad hoc de predicción</li> <li>- Análisis de descomposición temporal de series</li> <li>- Análisis frecuencial de procesos estocásticos (análisis espectral)</li> </ul>	<ul style="list-style-type: none"> <li>- Análisis temporal de procesos estocásticos (Box-Jenkins/modelos ARIMA univariantes)</li> <li>- Modelos probabilísticos.</li> </ul>
Enfoque C: Predicción según las relaciones entre variables	
<b>La dependencia entre el comportamiento de diferentes fenómenos posibilita la introducción de predicciones condicionadas con diversos grados de formalización y un análisis de mera relación o de causalidad.</b>	
<ul style="list-style-type: none"> <li>- Establecimiento de analogías</li> <li>- Análisis de regularidades estadísticas</li> </ul>	<ul style="list-style-type: none"> <li>- Modelos de simulación deterministas</li> <li>- Modelos econométricos</li> </ul>

- ✚ Clasificaciones como esta facilitan la elección de la técnica, pero para ello es necesario conocer en qué situación concreta nos encontramos.
- 

### **Enfoques clasificadores de las técnicas de predicción**

Como no podría ser de otra manera, existe una amplia variedad de técnicas que permiten llevar a cabo un proceso de predicción en economía y gestión de empresas. El conocimiento de las técnicas disponibles se presenta como algo necesario, tanto para juzgar las predicciones elaboradas externamente como, con más razón, las elaboradas por el propio analista.

Para alcanzar una visión de las posibilidades es adecuado realizar una clasificación ordenada de todas ellas. Nosotros, realizamos esta clasificación en función *del tipo de información* que utilizan estas técnicas. De esta forma distinguimos entre;

#### **- Enfoque A: Predicción según información subjetiva.**

Este primer enfoque se basa en la propia opinión que tienen ciertas personas sobre el futuro de la cuestión en estudio. Se trata datos suministrados por individuos aisladamente o como componentes de un grupo basándose en sus experiencias, opiniones y expectativas de futuro. Se trataría, por ejemplo, de encuestas de opinión de empresarios, expectativas de compra de ciertos consumidores, perspectivas de ventas de delegados comerciales o valoraciones de expertos sobre el futuro de la presión fiscal o de la productividad. Encontramos varias posibilidades según este enfoque:

##### *A.1. Encuestas de intenciones o expectativas.*

La predicción se obtiene directamente basada en las proporciones, medias y totales elevados al conjunto de la población, de una encuesta por muestreo. Ejemplos: Intenciones de compra de bienes de consumo duradero, expectativas de inversión en equipo de los empresarios.

##### *A.2. Encuestas de actitudes o sentimientos.*

Conocida a priori la conexión entre actitudes y comportamiento, se obtiene de una encuesta por muestreo un indicador de «sentimiento» según una escala preestablecida que se transforma en previsiones basadas en la historia del fenómeno en estudio.

Ejemplos: Índice de sentimiento del consumidor como condicionante de la demanda, índice de sentimiento del empresario como condicionante de la inversión.

### *A.3. Agregación de predicciones individuales.*

Las opiniones de determinados «expertos» se combinan para obtener una previsión global, con posibles ponderaciones diferentes y actuando al nivel de muestra o de población. Ejemplos: Previsiones de ventas a partir de vendedores y delegados, previsiones de precios de materias primas según los suministradores.

### *A.4. Consenso de grupo en interacción abierta.*

A partir de una discusión en grupo sobre un tema en concreto, se establecen predicciones con cierto grado de aceptación de sus componentes. Ejemplos: Brainstorming para predicción tecnológica, comités de empresa.

### *A.5. Panel de expertos con retroacción controlada de la información.*

Los componentes del panel suministran sus previsiones, que son agregadas y tratadas estadísticamente, recibiendo en sucesivas etapas información sobre la opinión global del grupo, pero con anonimato individual. Los expertos pueden proceder a una reconsideración de sus previsiones con tendencia a homogeneizar las previsiones del grupo. Técnicas: Delfos, impactos cruzados.

### *A.6. Otras técnicas exploratorias y normativas.*

El investigador examina los posibles condicionantes de la evolución de un fenómeno y, cuantificando o no sus probabilidades subjetivas de ocurrencia, establece la predicción. Ejemplos: Análisis morfológico y árboles de pertinencia en predicción tecnológica.

## **- Enfoque B: Predicción según el enfoque aislado de series.**

Atendiendo a este enfoque se analiza la propia evolución del fenómeno objeto de estudio en períodos anteriores. Se trata de casos tales como predecir la inflación del próximo trimestre basándonos en los datos mensuales de los últimos años o predecir las ventas de una empresa para el siguiente año tomando como referencia la evolución observada durante, por ejemplo, los últimos 15 años. Las técnicas más utilizadas según este criterio podrían ser:

### *B.1. Fórmulas «ad hoc» de predicción.*

Las predicciones se generan mediante un mecanismo automático establecido a priori y de cálculo recursivo. Técnicas: Medias móviles, alisado exponencial, modelos *naïve*, etc.

### *B.2. Análisis de descomposición temporal de series.*

El tratamiento clásico de las series ha sido la diferenciación y tratamiento aislado de las componentes de tendencia, estacionalidad y movimientos cíclicos. Ejemplos: Curvas de vida de productos, estacionalidad de las ventas, etc. Técnicas: Extrapolación de tendencia, curvas en S, desestacionalización, X-11, X-11 ARIMA, etc.

### *B.3. Análisis frecuencial de procesos estocásticos (análisis espectral).*

Las oscilaciones de una serie histórica de datos pueden descomponerse según la frecuencia o periodicidad de estos movimientos, definiendo el espectro de la misma. Ejemplos: Análisis espectral de las fluctuaciones de precios o de las cotizaciones en bolsa.

### *B.4. Análisis temporal de procesos estocásticos (Box-Jenkins/modelos ARIMA univariantes).*

Un análisis del comportamiento estocástico de una serie permite establecer un modelo estadístico de doble componente autorregresiva (valores anteriores de la variable) y de medias móviles (componente aleatorio), posible incorporación de estacionalidad y eliminación de tendencia. A efectos operativos, Box y Jenkins han propuesto el esquema resolutivo más ampliamente difundido. Ejemplos: Análisis de series temporales de variación de stocks, precios de materias primas, etc.

### *B.5. Modelos probabilísticos.*

Utilizando las distribuciones teóricas de probabilidad, se establecen modelos de predicción basados en los comportamientos reales observados. Ejemplos: Modelos de reposición de bienes de consumo duradero, modelos de difusión de un producto, cadenas de Markov para el establecimiento de las probabilidades de transición entre estados.

### **- Enfoque C: Predicción según las relaciones entre variables.**

El tercer enfoque se basa en la propia *conexión causal* entre el tema a predecir (efecto), y variables que la condicionan (causas). En términos más formales diríamos que la dependencia entre el comportamiento de diferentes fenómenos posibilita la introducción de predicciones condicionadas con diversos grados de formalización y un análisis de mera relación o de causalidad. Es frecuente que las leyes de comportamiento que se supone van a regir en el futuro se deduzcan de la experiencia sobre su funcionamiento en el pasado. Este sería el caso cuando el analista trata de predecir la evolución de los tipos de cambio en función de la posible marcha de otros aspectos que la condicionan, tales como el saldo de la balanza por cuenta corriente, los diferenciales de tipos de interés o los diferenciales de inflación por países. Vamos a distinguir 4 técnicas:



### *C.1. Establecimiento de analogías.*

La forma más simple de relación es la que se establece entre fenómenos similares en contextos distintos (temporal y espacialmente). En este sentido la predicción se realiza por pura analogía de situaciones. Ejemplos: Predicción de compra de un producto en función de la elasticidad de la demanda en otro país, modelos de adopción de nuevos productos a partir de las relaciones: conocimiento, intencionalidad de compra, adopción o prueba, en productos similares ya establecidos

### *C.2. Análisis de regularidades estadísticas.*

La relación existente entre fenómenos puede establecerse por un mero análisis estadístico de regularidad en su comportamiento conjunto, aun ante la inexistencia de una auténtica relación de causalidad. Con tal base pueden realizarse predicciones condicionadas a los valores precedentes de ciertos indicadores o barómetros. Ejemplos: Análisis de correlación y regresión entre variables de producción e indicadores macroeconómicos de actividad, indicadores adelantados.

### *C.3. Modelos de simulación deterministas.*

Las relaciones históricamente observadas entre la actuación de un conjunto de factores y el comportamiento de un fenómeno sirven de base para el establecimiento de predicciones sobre el mismo ante hipótesis de mantenimiento de la relación técnica preexistente. Técnicas: Análisis *input-output*, dinámica de sistemas (enfoque de Forrester).

### *C.4. Modelos econométricos.*

Se parte de un análisis estructural de las relaciones causa-efecto entre fenómenos con incorporación de factores aleatorios. Puede detectar efectos dinámicos (relaciones entre variables referidas a distintos momentos del tiempo), estructuras cambiantes (coeficientes variables en el tiempo) y realizar un análisis conjunto de múltiples relaciones simultáneas. Ejemplos: Modelos de demanda, inversión, costos, integrados de empresa, etc. Técnicas: Modelos econométricos uni y multiecuacionales, funciones de transferencia, SEMTSA, simulación y escenarios, etc.

### *A.5. Tipología: horizonte de predicción, medios e información*

#### Resumen:

- ✚ Las clasificaciones de técnicas facilitan la elección, pero es necesario conocer en qué situación concreta nos encontramos.
- ✚ La situación de predicción se caracteriza por 3 factores: horizonte de predicción, medios disponibles, e información disponible.
- ✚ Según el horizonte de predicción distinguimos 5 períodos.

- ✚ Los medios disponibles se refiere tanto a los humanos y financieros como acceso a programas, bases de datos, predicciones externas, y el tiempo disponible.
- ✚ El tipo de información se refiere a situaciones nuevas o situaciones con historia previa.
- ✚ Con todo ello, podemos realizar una tipología completa de técnicas en función de los factores descritos. Algunas de estas técnicas son objeto de estudio en el resto del curso.

### ¿A qué se refiere la información disponible?

La información de la que se disponga es otro de los factores que caracterizan la *situación de predicción*. En función de la información disponible, podemos encontrarnos con dos tipos de situaciones diferentes;

1) *Situaciones nuevas*, en las que tenemos que tomar decisiones sobre determinados temas y no se dispone de antecedentes. Ejemplos encontraríamos en el lanzamiento de nuevos productos, implantación de tecnologías innovadoras, fusión de empresas o integración comercial de áreas o países.

2) *Situaciones con historia*, en los que se dispone de información previa sobre el tema. Sin embargo, esta información (estos datos), puede estar limitada en el tiempo (series estadísticas cortas), o puede que se limite a la variable en cuestión, desconociéndose la información de otras variables del entorno.

Tipificación de la situación de predicción, según el horizonte de predicción			
Horizonte	Duración aproximada	Tipo de decisiones	Ejemplos
<b>Inmediato</b>	0-3 meses	Decisiones operativas, control presupuestario.	Distribución física, por ejemplo, gestión de inventarios en establecimientos minoristas.
<b>Corto plazo</b>	3-12 meses	Decisiones operativas, control presupuestario, decisiones de compra.	Compra de bienes de temporada para venta al por menor.
<b>Medio plazo</b>	1-3 años	Decisiones estratégicas.	<i>Leasing</i> de plantas y equipos; formación de empleados para nuevos procesos.
<b>Largo plazo</b>	3-10 años	Decisiones estratégicas, incluyendo las de expansión.	Investigación y desarrollo; compra de activos, adquisiciones y absorciones.
<b>Muy largo plazo</b>	> 10 años	Decisiones estratégicas, incluyendo aquellas de cambio de las características principales de la empresa.	Predicción tecnológica en apoyo de investigación y desarrollo y planificación estratégica.

Tipificación de la situación de predicción, según los medios empleados		
Medios	Características	Ejemplos de aplicaciones
<b>Escasos</b>	Personas individuales no especializadas. Aplicaciones a realizar en pocos días o incluso en horas.	La mayor parte de las realizadas en PYMES, predicciones en empresas o áreas tradicionales o donde la precisión requerida es relativamente baja.
<b>Normales</b>	Equipo con 2-3 miembros con experiencia, aplicaciones a realizar en semanas e incluso meses, y organización informática no necesariamente especializada en predicción.	Planes operativos y estratégicos de empresas, presupuestos de organismos oficiales.
<b>Altamente profesionalizados</b>	Equipos especializados, aplicaciones realizadas en forma continuada, acceso a bancos de información y programas de ordenador especializados.	Predicciones macroeconómicas a medio plazo, sistemas integrados de predicción empresarial, planes estratégicos de multinacionales y empresas líderes, predicción de tipos de cambio y tipos de interés.

Tipología de las técnicas de predicción, según las situaciones a las que se aplican			
<b>A) Situaciones sin historia</b>			
Horizonte Medios	Corto	Medio	Largo
<b>Escasos</b>	<ul style="list-style-type: none"> <li>- Agregación predicciones individuales.</li> <li>- Reuniones de directivos (plan operativo).</li> <li>- Indicadores.</li> </ul>	<ul style="list-style-type: none"> <li>- Reuniones de directivos (plan estratégico).</li> <li>- Fórmulas elementales.</li> <li>- Analogías.</li> </ul>	<ul style="list-style-type: none"> <li>- <i>Brainstorming</i>.</li> <li>- Analogías.</li> </ul>
<b>Normales</b>	<ul style="list-style-type: none"> <li>- Encuestas opinión, intenciones, actitudes o sentimiento.</li> </ul>	<ul style="list-style-type: none"> <li>- Modelos de difusión y respuesta.</li> <li>- Diseño de experimentos.</li> <li>- Simulación mediante fórmulas recursivas.</li> </ul>	<ul style="list-style-type: none"> <li>- Delfos.</li> <li>- Impactos cruzados.</li> <li>- Análisis morfológico y árboles de pertinencia.</li> </ul>
<b>B) Situaciones con historia</b>			
<b>Escasos</b>	<ul style="list-style-type: none"> <li>- Alisado exponencial.</li> <li>- Medias móviles.</li> <li>- Modelos <i>naïve</i>.</li> <li>- Desestacionalización.</li> </ul>	<ul style="list-style-type: none"> <li>- Ajuste tendencia.</li> </ul>	<ul style="list-style-type: none"> <li>- Curvas en S.</li> </ul>
<b>Normales</b>	<ul style="list-style-type: none"> <li>- ARIMA univariante.</li> <li>- Análisis intervención.</li> <li>- X-11 ARIMA.</li> <li>- Funciones de transferencia.</li> </ul>	<ul style="list-style-type: none"> <li>- Modelos econométricos uniecuacionales.</li> <li>- Modelos econométricos de pequeño tamaño.</li> <li>- Modelos VAR.</li> </ul>	<ul style="list-style-type: none"> <li>- Modelos con cambio estructural</li> </ul>
<b>Altamente profesionalizados</b>	<ul style="list-style-type: none"> <li>- Análisis espectral.</li> <li>- Modelos dinámicos.</li> <li>- Filtros.</li> </ul>	<ul style="list-style-type: none"> <li>- Modelos econométricos multiecuacionales de gran tamaño.</li> <li>- Simulación.</li> <li>- <i>Input/output</i>.</li> </ul>	<ul style="list-style-type: none"> <li>- Modelos con cambio estructural.</li> <li>- Escenarios.</li> </ul>

## A.6. Cinco reglas sobre predicción

### Resumen:

1. Escepticismo medurado sobre la exactitud de todo dato económico.
  2. Utilización de las predicciones a futuro con cautela y valorar, sobre todo, el cambio de expectativas que supone respecto a predicciones anteriores.
  3. Las predicciones no son neutrales y deben ser valoradas según la fuente de origen.
  4. La discrepancia entre predicciones alternativas, es un indicador importante de los riesgos que hay detrás de una predicción. La coincidencia entre predicciones, puede traslucir ausencia de criterios propios.
  5. En las predicciones alternativas hay que valorar: la independencia de la institución, los errores cometidos, la actualidad de la información utilizada, la disponibilidad de escenarios alternativos y los supuestos de la predicción.
- 

### **Cinco reglas sobre predicción**

Estas reglas surgen de la *carta abierta* de Antonio Pulido a Peter Pauly, responsable operativo del Proyecto LINK, publicada en la prensa económica española.

1ª) El punto de partida para el usuario de cualquier predicción es un *escepticismo medurado sobre la exactitud de todo dato económico*, especialmente sobre los grandes agregados. Aunque no se puede dudar de la seriedad metodológica de los institutos de estadística de todos los países, es evidente que las grandes macromagnitudes de la contabilidad nacional son producto de una compleja labor de encaje de datos estadísticos diversos. Como, además la información económica de base fluye con cierto retraso, las valoraciones de *avance* que se publican pocos meses después de la fecha de referencia se van rectificando año tras año hasta alcanzar la categoría de dato oficial *definitivo*. La experiencia nos indica que entre el dato de avance y el definitivo, por ejemplo, de crecimiento del PIB, se ha dado en los últimos años, para el conjunto de países europeos, una corrección media de medio punto de porcentaje. Así, un crecimiento para 2002 valorado hoy día, para un país, en el 2 por 100 puede verse convertido, dentro de algunos meses o años, en cualquier tasa comprendida entre el 1,5 y el 2,5 por 100. Obviamente, las implicaciones de una u otra son bien distintas, y desde luego que condicionan cualquier predicción.

2ª) Dada la inexactitud de los datos macroeconómicos, es necesario *utilizar las predicciones a futuro con máxima cautela y valorar, sobre todo, el cambio de expectativas* que supone respecto a predicciones anteriores. Por ejemplo, no podemos estar muy seguros de que el PIB de Alemania crezca en 2003 un 0,4 % como afirma la Comisión Europea en su informe de abril de 2003. Pero si en el informe de abril de 2002 apostaban por un crecimiento para 2003 de un 2,7 %, entonces sí podemos estar seguros de

que ahora son más pesimistas que hace un año. Ese, quizás sea el aspecto relevante.

3ª) *Las predicciones no son neutrales y deben ser valoradas según la fuente de origen.* Así, las valoraciones de futuro que cualquier gobierno incorpora a sus presupuestos para el año siguiente o a sus programas a medio plazo tiene casi la obligación de pecar de optimismo. Igualmente, es conveniente saber que las predicciones de las instituciones internacionales, aunque están elaboradas de forma independiente, se comentan y *negocian* previamente con los gobiernos. En este caso, las revisiones periódicas en las predicciones son muy indicativas de cambios de percepción sobre la marcha de la economía.

4ª) La coincidencia entre predicciones no siempre corresponde a acuerdos de valoración y puede traslucir ausencia de criterios propios; *la discrepancia es un indicador importante de los riesgos* que hay detrás de una predicción. Así, la alta inestabilidad de la economía japonesa, con variaciones trimestrales de crecimiento del PIB en 1998 que habían oscilado entre el -3,5 por 100 y el -2,8 por 100 en tasas interanuales, provoca que las predicciones para el siguiente año se repartieran en un amplio intervalo que va de poco más del -0,1 por 100 hasta el -1,8 por 100, mostrando el alto riesgo asociado a las mismas.

5ª) A la hora de elegir entre predicciones internacionales alternativas es importante valorar aspectos tan diversos como: la *independencia* de la institución que las elabora; la experiencia previa sobre *errores* comparativos; la *actualidad* de la fecha efectiva de cierre de la información utilizada para la predicción; la disponibilidad de *escenarios alternativos* y, en todo caso, el que se conozcan los *supuestos* de base que están detrás de cada predicción.

### ***Discrepancias entre instituciones a la hora de predecir.***

Los economistas no iniciados en el seguimiento de la coyuntura y la predicción económica suelen contemplar con cierta animadversión las escasas coincidencias que se suelen dar sobre las perspectivas que tienen las distintas instituciones. Los economistas no sólo discrepan sobre las medidas de política económica más adecuada para afrontar las dificultades que se presentan, o para mantener unos ritmos sostenidos de crecimiento de la producción, sino que ni siquiera coinciden en su valoración sobre la evolución que puede presentar una economía a partir del conocimiento pasado y actual. Cuando un gestor de política económica, o un gestor de una empresa, desean hacerse una idea de la evolución prevista del entorno económico, es absolutamente necesario que consulte varias *fuentes de predicción*. Debe revisar un buen número de informes de centros de predicción. ¿Qué puede encontrarse? Sin duda una alta discrepancia. Comprobémoslo analizando las predicciones que para el crecimiento del PIB español del año 2002, realizaron algunas de las instituciones (españolas y extranjeras, públicas y privadas) en octubre de 2001.

Obsérvese que en el cuadro siguiente se presentan algunos de los rasgos que hemos descrito en la Unidad Didáctica; en primer lugar, la *cuarta regla de la predicción* señalaba que la elevada discrepancia entre instituciones no era más que reflejo de los riesgos que hay detrás de la predicción. En el ejemplo se trata de más de un punto porcentual entre el más y el menos optimista. Pues bien esa gran diferencia, con implicaciones tan distintas en la consideración de la marcha de la economía, no es más que el reflejo de la dificultad de valorar situaciones tan complejas como los efectos del 11-S en la economía española (recuérdese que las predicciones recogidas se efectuaron apenas unas semanas después de los atentados). Queremos destacar un segundo aspecto, que sirve de ejemplo para nuestra tercera regla de predicción: *Las predicciones no son neutrales y deben ser valoradas según la fuente de origen*. Obsérvese que entre las 5 instituciones más optimistas en aquel momento sobre la marcha esperada de la economía, 4 (si incluimos al ICO) son instituciones oficiales. Ya nos hemos acostumbrado a que estas predicciones tiendan al optimismo, y de hecho, se acoge entre los analistas con gran temor las visiones pesimistas de este tipo de fuentes de predicción.

Predicciones dadas en octubre de 2001 por distintas instituciones sobre el crecimiento del PIB español para 2002	
Institución	% variación
Ministerio de Economía	2,9
OCDE (junio)	2,9
Caixa Catalunya	2,8
FMI	2,8
Instituto de Crédito Oficial	2,8
Banesto	2,7
Analistas Financieros Internacionales	2,5
BBVA	2,5
Instituto de Estudios Económicos	2,5
La Caixa	2,4
Caja Madrid	2,3
Consensus Economics Inc.	2,3
FUNCAS	2,3
Instituto Complutense de Análisis Económico	2,3
JP Morgan	2,3
The Economist Poll	2,2
CEPREDE	2,1
Banco Santander Central Hispano	2,0
Goldman-Sachs	2,0
Merrill Lynch	2,0
UBS Warburg	2,0
IFL-Univ. Carlos III	1,9
Morgan Stanley	1,8
Media	2,4

Fuente; CEPREDE, *Perspectivas económicas y empresariales*, noviembre 2001, elaborado a partir de los informes de las instituciones señaladas.

### **Revisiones en la predicción; la experiencia CEPREDE**

El Centro de Predicción Económica (CEPREDE) acumula una experiencia de más de 20 años, con unos orígenes ligados a la referencia mundial en predicción económica, como es la Warthon School de la University of Pennsylvania del Profesor Lawrence R. Klein, Premio Nobel de Economía en 1980. El proceso de realización de las predicciones sigue un patrón de reconocido prestigio internacionalmente comparable.

Sin embargo, las predicciones de CEPREDE, y de cualquier Institución, sobre un período determinado varían a lo largo del tiempo, con una gran inestabilidad cuanto más alejado se encuentran del período de referencia. Pero ello, no es motivo de inexperiencia, o de falta de rigor. Todo lo contrario. La predicción, sobre todo cuando el horizonte es tan amplio (CEPREDE predice más de 300 variables a 5 años), no puede ser estática, sino que debe adaptarse, para recoger aquellas situaciones y fenómenos que se presentan progresivamente en una economía.

Por ejemplo, sería una insensatez que una previsión de crecimiento de los ingresos por turismo en España para 2002 realizada en el año 1999 no se revisase tras los atentados del 11-S. Esa es la garantía de la utilidad de las predicciones, y tales variaciones no son sino un fiel reflejo de "cambios de clima" económico.

En el cuadro siguiente se recoge cómo han variado las predicciones que CEPREDE ha ido realizando sobre la tasa de crecimiento anual del PIB español para el ejercicio 2002.

Además, se compara con la tasa que finalmente se produjo, según el Instituto Nacional de Estadística, organismo encargado de realizar la estadística oficial.

Variación en las predicciones de CEPREDE sobre el crecimiento del PIB español para 2002	
	Tasa de variación anual
<b>Avance INE (marzo 2003)</b>	<b>2.0</b>
<i>Predicción CEPREDE realizada en...</i>	
Noviembre de 2002	1.9
Mayo de 2002	2.0
Noviembre de 2001	2.1
Mayo de 2001	3.3
Noviembre de 2000	3.6
Mayo de 2000	3.8
Noviembre de 1999	3.6
Mayo de 1999	3.4
Noviembre de 1998	3.0
Mayo de 1998	2.9
Fuente; CEPREDE, <i>Perspectivas Económicas y Empresariales</i> , varios años.	

¿Qué lección podemos extraer de este cuadro?

En primer lugar, vemos que, como es natural, cuando la fecha de la predicción se aleja del período de referencia, existe una alta volatilidad de la predicción, que no hace sino reflejar los riesgos e incertidumbres asociados a una predicción a largo plazo, ya que el conjunto de factores que pueden incidir sobre la economía en un período tan largo es de tal amplitud, que cada revisión se convierte en un nuevo ajuste y reconsideración de factores. Observemos que la media de la predicción realizada con una antelación superior a 1 año (años 1998, 1999, y 2000) es de 3,4 %, con un rango de casi un punto de porcentaje (0,9 puntos). Obviamente, es muy distinta la consideración que se puede sacar de un crecimiento del PIB del 2,9 % o del 3,8 %.

Obsérvese que la volatilidad en la predicción, en general, disminuye a medida que la fecha de predicción se aproxima a la fecha de referencia. En el Cuadro se comprueba para las tres predicciones más recientes.

Pero este cuadro nos enseña con claridad una lección importante. **Las revisiones en las predicciones son necesarias, y ello garantiza su utilidad.** Fíjese en el gran salto producido entre mayo de 2001 y noviembre de 2001. Las predicciones ya venían cayendo desde un año antes, reflejo de un descenso en el optimismo sobre la continuidad del ciclo alcista, pero en ese momento la tasa prevista de crecimiento pasa de un 3,3 a un 2,1 por ciento en sólo 6 meses. Un observador crítico en exceso puede pensar que eso es señal de falta de seriedad y rigor de la técnica de predicción, o de la Institución que la lleva a cabo. Pero en absoluto; nadie podía en mayo de 2001 prever un suceso de tal calibre como los atentados en el World Trade Center del 11 de septiembre de 2001. Evidentemente, ese suceso (por supuesto que junto con otros), condiciona la predicción, e implican necesariamente una rebaja en las expectativas de crecimiento de una economía. Cualquier centro de predicción debe revisar sus predicciones a la luz de ese fenómeno. Lo contrario supone una actitud estática alejada de la realidad que la predicción pretende describir. Ahora cobra sentido nuestra *segunda regla de predicción*; utilizar las predicciones a futuro con máxima cautela y valorar, sobre todo, el *cambio de expectativas* que supone respecto a predicciones anteriores.

#### A.7. Fundamentación científica de la predicción: antecedentes históricos

##### Resumen:

- ✚ La predicción económica y empresarial no tiene una base científica unitaria, sino que conjuga diversos enfoques metodológicos provenientes de la estadística, la economía, la psicología y otras ciencias.
- ✚ Las técnicas de información subjetiva están próximas a los desarrollos de la psicología y la ciencia de la dirección.
- ✚ Las técnicas de información histórica se vinculan a desarrollos de la estadística y, en particular, al análisis de procesos estocásticos y series temporales.



- ✚ Las técnicas que aprovechan información relacional se entroncan con los desarrollos de la modelización en ciencias sociales y, en nuestro caso, con la modelización econométrica.
- ✚ En el documento se recogen algunos de los hitos históricos en predicción económica y empresarial.

#### ***Algunos hitos históricos en predicción económica y empresarial***

- 1758** François Quesnay *Le tableau économique*
- 1845** Verhulst *Recherche mathématique sur la loi de l'accroissement de la population*
- 1847** C.D.H. Buys-Ballot "Les changements périodiques de température"
- 1900** Arthur Schuster "The periodogram of magnetic declination", *Trans. Cambridge Phil. Soc.*, 18.
- 1904** J. Spencer "On the graduation of the rates of sickness and normality", *Journal of the Institute of Actuaries*, 38
- 1911** Henry Moore *Laws of wages: An essay in statistical economics*, Mac Millan
- 1919** M. Pearson "Index of business condition", *Review of Economics and Statistics*, 1.
- 1921** William Beveridge, "Weather and harvest cycles", *Economic Journal*, 31.
- 1927** G.U. Yule "On a method of investigating periodicities in disturbed series", *Trans. Roy. Soc.*, 226
- 1928** P.H. Douglas y C.W. Cobb "A theory of production", *American Economic Review*, 18.
- 1934** J. Newman "On the two different aspects of the representative methods", *Journal of the Royal Statistical Society*, 97
- 1937** Jean Tinbergen "An econometric approach to business cycle problems", Herman
- 1937** E. Slutsky "The summation of random causes as the source of cycle processes" *Econometrica*, 5.
- 1938** Herman Wold *Stationary time series*, Almqvist and Wiksells
- 1938** Henry Schultz *The theory and measurement of demand*, University of Chicago Press
- 1988** S. Johansen "Statistical Analysis of Cointegrating Vectors", *Journal of Economic Dynamics and Control*, 12.
- 1995** S. Johansen *Likelihood-based inference in cointegrated vector autorregressive models*, Oxford University Press.
- 2002** D.F. Hendry y M.P. Clements "Economic Forecasting: some lesson from recent research", ECB Workshop on forecasting techniques.

#### ***A.8. Etapas genéricas del proceso de predicción/simulación***

##### **Resumen:**

- ✚ Aunque cada técnica conlleva peculiaridades propias, sí se pueden señalar aspectos comunes.
  - 1)** Partiendo de un tema de carácter general de economía o gestión de empresas es preciso un *planteamiento del problema en términos de predicción*.
  - 2)** *Tipificación de la situación* de predicción que se nos plantee.
  - 3)** *Elección de la técnica*.
  - 4)** *Búsqueda de la información*.
  - 5)** *Aplicación* de la técnica de predicción.
  - 6)** *Evaluación* de los resultados.
  - 7)** *Revisión* del proceso con base en la experiencia.
- ✚ A continuación se recoge una hipotética aplicación: inversión en infraestructura ambiental y reestructuración del sector siderúrgico.

### **Etapas genéricas del proceso de predicción**

- 1) Planteamiento del problema en términos de predicción.
- 2) Tipificación de la situación.
- 3) Elección de técnica.
- 4) Búsqueda de la información.
- 5) Aplicación.
- 6) Evaluación.
- 7) Revisión del proceso.

<b>Etapas</b>	<b>Ejemplo 1: Inversión en infraestructura para reducir la contaminación en el Mediterráneo</b>	<b>Ejemplo 2: Reestructuración del sector siderúrgico</b>
1	Contaminación estimada en fechas claves como años 2000, 2010 y 2020.	Tendencias de producción y consumo para los próximos cinco años.
2	Largo plazo, información subjetiva, medios normales.	Largo plazo, información histórica, medios escasos.
3	Delfos.	Análisis de tendencia (diversas variantes).
4	Envío de cuestionarios a 10 expertos internacionales seleccionados, con retroalimentación de resultados.	Serie histórica de producción, consumo, exportación e importación de productos siderúrgicos.
5	Cálculo de la predicción mediana, primer y tercer cuartil para los años seleccionados.	Cálculo de las ecuaciones de tendencia y predicciones alternativas (según variantes) para los próximos cinco años.
6	Autoevaluación de expertos. Posible comparación con otros estudios internacionales o aplicando otras técnicas.	Análisis de errores históricos. Comparación de predicciones. Diferencias predicción realización hacia futuro.
7	Reunión internacional de todos los equipos nacionales implicados.	Establecimiento de un sistema continuo de rectificación de predicciones en función de nuevos datos.

### ***Programas de ordenador disponibles para la predicción y la simulación.***

Para llevar a cabo ejercicios de predicción y simulación necesitamos inexcusablemente la ayuda de equipos informáticos y software adecuado. Las operaciones necesarias para dar respuesta a la sofisticación de las técnicas disponibles, junto con la propia complejidad de la realidad a predecir, convierten a los instrumentos informáticos en imprescindibles. De hecho, una de las razones para la proliferación de aplicaciones de predicción a cualquier campo de la economía y la gestión de empresas se encuentra en el espectacular desarrollo de los equipos y software informático. Algunos análisis impensables hace ahora unos años, se han convertido en rutinarios con la aparición de equipos informáticos personales muy potentes y de paquetes estadístico-econométricos o lenguajes informáticos muy versátiles.

Como ejemplo del avance desarrollado por la predicción aplicada podemos recoger la anécdota mencionada por Berndt: como vimos, uno de los instrumentos de predicción y simulación más habituales son los modelos econométricos multiecuacionales. Pues bien, este autor apunta que el

tiempo estimado de solución de un modelo econométrico de 10 ecuaciones y 50 parámetros desconocidos en la Cowles Commission de 1946 era de 70 días continuados de trabajo. Los modelos econométricos actuales, como el Warthon-UAM del Instituto Klein y CEPREDE estiman los parámetros de más de 1000 ecuaciones en instantes. Precisamente es la creación de programas informáticos específicos para la predicción, junto con la mejora y popularidad de los sistemas informáticos, lo que ha dotado de dinamismo a la predicción en economía y gestión de empresas y ha provocado la proliferación de todo tipo de trabajos y estudios. Por ello, vamos a señalar, sin ánimo de ser exhaustivos, una breve relación de algunos de los más utilizados programas de ordenador para la predicción y simulación. Todos ellos pueden ser útiles en determinados momentos, si bien el más versátil y utilizado en las Universidades, o incluso en la práctica profesional, es el programa EVIEWS de QMS.

#### I.- Programas estadísticos de uso general

SPSS, SPSS Inc. *Statistical Package for Social Sciences*. El más utilizado en Economía y gestión de empresas. Incluye ya funciones avanzadas más habituales en programas estrictamente econométricos.

STATA, Stata Corp. Análisis general de datos y estadístico.

SAS, SAS Institute Statistical Analysis System

#### II.- Programas globales de predicción

FOCA, Dataplan GmbH. *Forecasting Software for Economical Time Series*. Incluye diversas variantes de técnicas de alisado, ajuste de tendencia, tratamiento de la estacionalidad, análisis espectral, ARIMA y regresión múltiple.

FORMAN, QM Software *Forecasting for management*. Incluye variantes de técnicas de medias móviles, tendencia, alisado exponencial, descomposición de series, filtros adaptativos, ARIMA y regresión múltiple.

FORECAST MASTER, Scientific Systems Electric power Research institute. Incluye variantes de técnicas de alisado exponencial, tendencia, descomposición de series, regresión múltiple, ARIMA, espacio de los estados y vectores autorregresivos bayesianos.

#### III.- Programas de especialización en información subjetiva

FORECAST, Decision Analysis Group. Psychology Dept. City of London Polytechnic. Limitado a predicción basado en información subjetiva

#### IV.- Programas de especialización en series temporales (no ARIMA)

XTRAPOLATOR, Stratix Diversas variantes de alisado, ajuste de tendencia, autorregresión y medias móviles

VISITREND, Visicorp Alisado exponencial, medias móviles y regresión múltiple.

#### V.- Programas de especialización en modelos ARIMA

SCA STATISTICAL SYSTEM, Scientific Computing Ass Análisis univariante y multivariante, análisis de intervención y función de transferencia, Análisis de descomposición de series

MICROBJ, Stratix ARIMA univariante

AUTOBJ/AUTOBOX, Automatic Forecasting INC. Opción automática de modelización univariante

TRAMO-SEATS *Time series Regression with ARIMA noise, Missing values and Outliers-Signal Extraction in ARIMA Time Series Maragall y Gómez (Banco de España)* Detección automática. Extracción de señal.

#### VI.- Programas de especialización en modelos econométricos

EIEWS, Quantitative Micro Software Amplias funciones en modelización econométrica. Desde la versión 4 incluye TRAMO-SEATS.

RATS, VAR Econometrics Inc. *Regression Analysis of Time Series*. Modelos VAR, análisis espectral y filtro de Kalman

LIMDEP, Econometric Software Inc. Modelos lineales y no lineales, incluido panel data. Especializado en modelos de variables dependientes limitadas.

GAUSS Programa matemático-estadístico de cierta utilización y con amplia variedad de modelos econométricos

#### VII.- Programas especializados por temas

SYSPRE, Metrapack Especializado en previsión de ventas

FORSYS/MAVIS, Marketing Systems GmbH. Especializado en información, control, análisis y predicción a corto plazo.

#### VIII.- Programas de cálculo de previsiones alternativas

LOTUS 1-2-3, Lotus Development Corp. Hoja de cálculo general

EXCEL, Microsoft Corp. Hoja de cálculo general

SIMPC, Don Econometrics Especializado en simulación de modelos econométricos de tamaño mediano y grande, inicialmente estimados con otros programas

MICRODYNAMO, Pugh Roberts Ass. Simulación en dinámica de sistemas

## B. Técnicas Elementales de Predicción.

- En este acápite iniciamos el estudio de las técnicas de predicción. Comenzaremos con las más sencillas. Se trata de técnicas caracterizadas por su escasa necesidad de medios y que, sin grandes pretensiones metodológicas, pueden resolver razonablemente bien necesidades sencillas de predicción, con una buena relación eficacia/costo.
- Abordamos el estudio de las distintas alternativas, dispongamos o no de información histórica. Para cada técnica especificaremos sus fundamentos teóricos y realizaremos una aplicación práctica. Sintetizaremos, además, con una breve guía, las distintas posibilidades que se nos presentan.
- Algunos conceptos previos son necesarios para aplicar estas técnicas. Por ello, le dedicaremos epígrafes a los componentes de una serie, a la localización y tratamiento de información estadística, tanto con el programa EViews como en Excel. Por último, y a modo de repaso final, aplicaremos algunas de las técnicas mencionadas a casos reales. Una vez estudiado el acápite, el lector debe estar en condiciones de reproducir, con otras series estadísticas, estos casos.
- Este segundo acápite está articulado en 8 apartados. Cada uno parte de una ficha resumen donde se explican los aspectos teóricos fundamentales. Cuando es necesario se complementa con breves explicaciones adicionales. Cada técnica cuenta con su respectivo ejercicio práctico. Cerramos con una actividad a realizar, dos casos de aplicación (uno para resolver en Excel y otro en EViews - **se supone que el lector es usuario de ambos programas y tiene conocimientos básicos de econometría – de no ser así tomar contacto con el profesor**) y el correspondiente test de autoevaluación. Con todo ello, estimamos que para el aprovechamiento del acápite el lector deberá dedicar unas 25 horas, distribuidas de la siguiente manera: unas 7 para el estudio de los conceptos teóricos; 8 horas para la realización guiada de los ejercicios; unas 4 horas para la resolución de los 2 casos de aplicación práctica; y unas 6 horas para el resto de actividades (lecturas, manejo del programa EViews y test de autoevaluación).

### B.1. Predicción elemental sin información histórica

#### Resumen:

- ✚ Se presenta no sólo ante temas nuevos, sino porque el analista renuncia a la historia por considerarla poco orientadora, o porque prefiere información subjetiva.
- ✚ Con frecuencia estas técnicas se basan en la experiencia y en la intuición, y son de gran utilización en la práctica profesional.
- ✚ Entre estas técnicas señalamos: Reuniones de directivos, agregación de predicciones, *brainstorming*, analogías, indicadores adelantados, fórmulas matemáticas elementales.

- ✚ En general, se trata de técnicas que precisan de medios sencillos, pero que pueden lograr buenas predicciones incluso a medio y largo plazo.

### **Técnicas elementales en situaciones sin información histórica**

Entre estas técnicas señalamos algunas de las más utilizadas en la práctica profesional, como son:

- *Reuniones de directivos.*

Bien sea de las empresas o de los Organismos Públicos. En ellas se trata de adoptar *planes* en diferentes plazos. Por ejemplo, se puede tratar de *planes operativos anuales* y de su desarrollo en *presupuestos*, o de *planes estratégicos*, frecuentemente a tres o cinco años con revisiones anuales "deslizantes", es decir, añadiendo sucesivamente un año para mantener el horizonte de predicción. Estas reuniones de directivos permiten abordar problemas puntuales de predicción, en los conocidos como *jurado de opiniones de directivos* o *consenso de grupo de expertos en interacción abierta*.

- *Agregación de predicciones.*

Surgen normalmente de los consensos de grupo de expertos, y en su opción más difundida se trata de realizar agregaciones de predicciones individuales (por ejemplo, predicciones de vendedores, delegados de ventas, controladores de costos,...). Una versión más sofisticada utiliza tres predicciones por experto (pesimista, optimista y más probable) y las pondera en línea con lo propuesto en el método PERT (*Program Evaluation and Review Technique*, método inicialmente diseñado para el seguimiento de proyectos).

- *Brainstorming o tormenta de ideas.*

Versión cualitativa de las reuniones de expertos o directivos, utilizada fundamentalmente en temas de predicción tecnológica y nuevos productos, en la que un grupo de expertos improvisa sobre un tema con un mínimo de formalidades organizativas.

- *Predicción por analogías.*

La línea más difundida es la que se refiere analogías internacionales. Por ejemplo, el porcentaje de gasto en sanidad sobre el PIB presenta una tendencia al crecimiento en relación con el nivel de desarrollo económico del país, medido por la renta *per cápita*. Si suponemos que en cinco años el país A puede alcanzar el nivel de renta *per cápita* del país B, una predicción para los próximos 5 años sobre el gasto sanitario en el país A puede realizarse a partir de la diferencia entre la renta *per cápita* actual entre ambos países y el porcentaje de gasto sanitario del país B sobre su PIB.

- *Indicadores adelantados de actividad.*

Se trata de una línea de investigación iniciada por Mitchell y Burns en el *National Bureau of Economic Research* (NBER). Un indicador adelantado (*leading indicator*) es una variable económica de actividad cuyas oscilaciones anticipan las caídas y recuperaciones en algunos sectores de la actividad económica, o en el conjunto de la economía de un país. Además, suelen ser variables de actualización periódica y fácil obtención, por lo que su seguimiento y análisis puede mostrar indicios de la trayectoria futura de la actividad económica y, por tanto, se convierten en una forma indirecta de predicción. Uno de los ejemplos más paradigmáticos lo constituye el *consumo aparente de cemento*, que es un claro indicador adelantado del sector de la construcción, que a su vez representa una parte muy considerable de la producción de cualquier país. Sus oscilaciones a lo largo del ciclo anticipan el comportamiento del sector en algunos meses. Por ello, una caída observada en un momento determinado en el mismo, nos permite predecir una contracción en los indicadores de producción del sector de la construcción en un futuro próximo.

- *Fórmulas matemáticas elementales.*

No son producto de una técnica estadística de tratamiento de datos históricos, sino de un supuesto (por ejemplo, continuidad de la tasa de crecimiento del último año) o de una experiencia (por ejemplo, una regla genérica de respuesta de los precios de consumo ante una variación en los precios de la gasolina).

*B.2. Información histórica: componentes de una serie económica*

Resumen:

- En buen número de circunstancias, a la hora de predecir un fenómeno económico, contamos con información histórica. Normalmente esa información queda cuantificada en series estadísticas.
  - Tradicionalmente se considera que una serie económica se puede descomponer, atendiendo a dos esquemas, en cuatro componentes: tendencia (T), ciclo (C), estacionalidad (S), y componente irregular (I).
  - En general, la elección de una u otra técnica de predicción depende del supuesto establecido sobre el esquema de descomposición, y la identificación de los componentes mencionados, tal y como se desarrolla en los epígrafes siguientes.
-

### **Esquemas de descomposición**

Se suele considerar que la descomposición de una serie económica en sus cuatro componentes considerados (tendencia, ciclo, estacionalidad y componente irregular), puede llevarse a cabo atendiendo a dos esquemas de descomposición. Según el **esquema multiplicativo**, la serie económica es el resultado del producto de sus componentes, esto es:

$$Y = T * C * S * I$$

Por el contrario, según el **esquema aditivo**, la serie es el resultado de la suma de sus componentes:

$$Y = T + C + S + I$$

En la práctica, se suele considerar que los cuatro componentes teóricos quedan reducidos realmente a dos: 1) que recogería movimientos de la serie a largo plazo (tendencia, o tendencia más ciclo) y 2) que reflejaría variaciones de la serie por razones de estacionalidad. Los movimientos cíclicos raramente tienen un tratamiento independiente y los de carácter irregular forman el residuo o término de error que servirá para analizar la propia bondad de la aplicación realizada. Además, se suele considerar un esquema multiplicativo. En definitiva, la serie se considera formada por:

$$Y = T * S$$

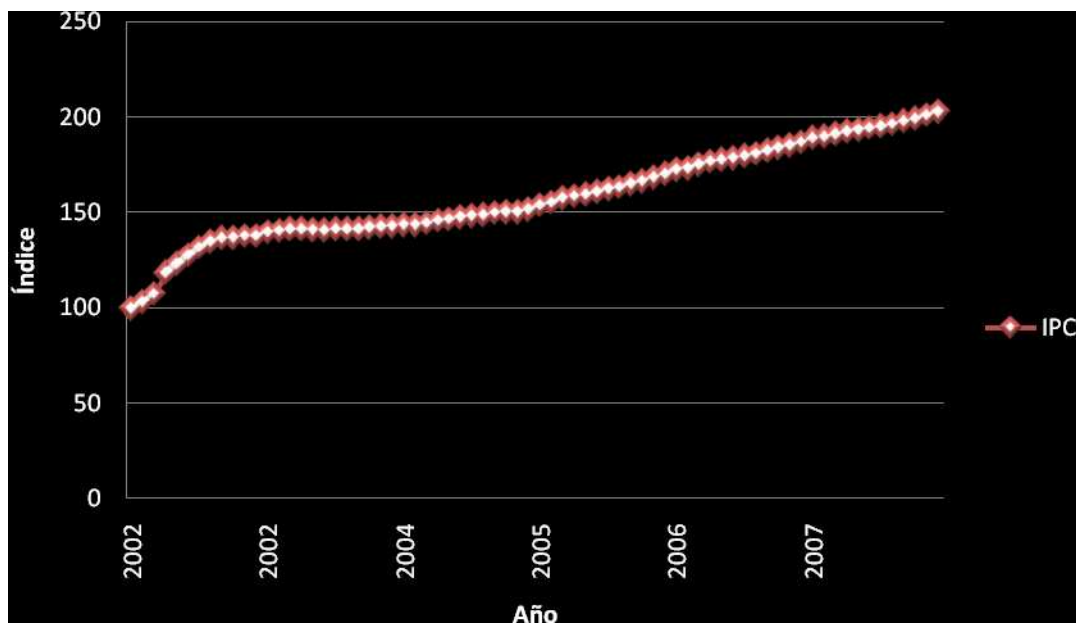
Significa que el valor original de una serie es el resultado de multiplicar un valor de tendencia por un factor de estacionalidad (de media 1 o 100, en índice). No obstante, los tratamientos que analizaremos son generalizables al esquema aditivo, o incluso a esquemas mixtos.

### **Tendencia**

La tendencia es aquel componente de una serie económica vinculada al movimiento a largo plazo de la serie. Se trata del patrón regular de comportamiento a largo plazo, sea éste creciente o decreciente. Por ejemplo, la inflación es un proceso inherente a casi todas las economías en mayor o menor grado. Por ello, cabe esperar que el patrón de comportamiento regular de los índices de precios sea creciente. Debemos esperar a largo plazo que los precios suban. Por ello, la serie estadística que recoge la evolución de los precios debe presentar un importante componente tendencial. Así aparece reflejado en el gráfico inferior, donde se recoge la evolución del IPC argentino (nivel general) desde enero de 2002 hasta diciembre de 2007, base 1999=100.

La clave es comprender que, más allá de subidas y bajadas parciales (de un mes a otro) existe una clara tendencia al crecimiento de esta variable.





En general puede decirse que la tendencia puede ser aislada o eliminada de muy diferentes formas como son, por ejemplo:

- Cálculo de las diferencias entre valores de la serie original (diferenciación sucesiva).
- Ajuste de funciones matemáticas.
- Media móvil simple, ponderada o alisado exponencial.

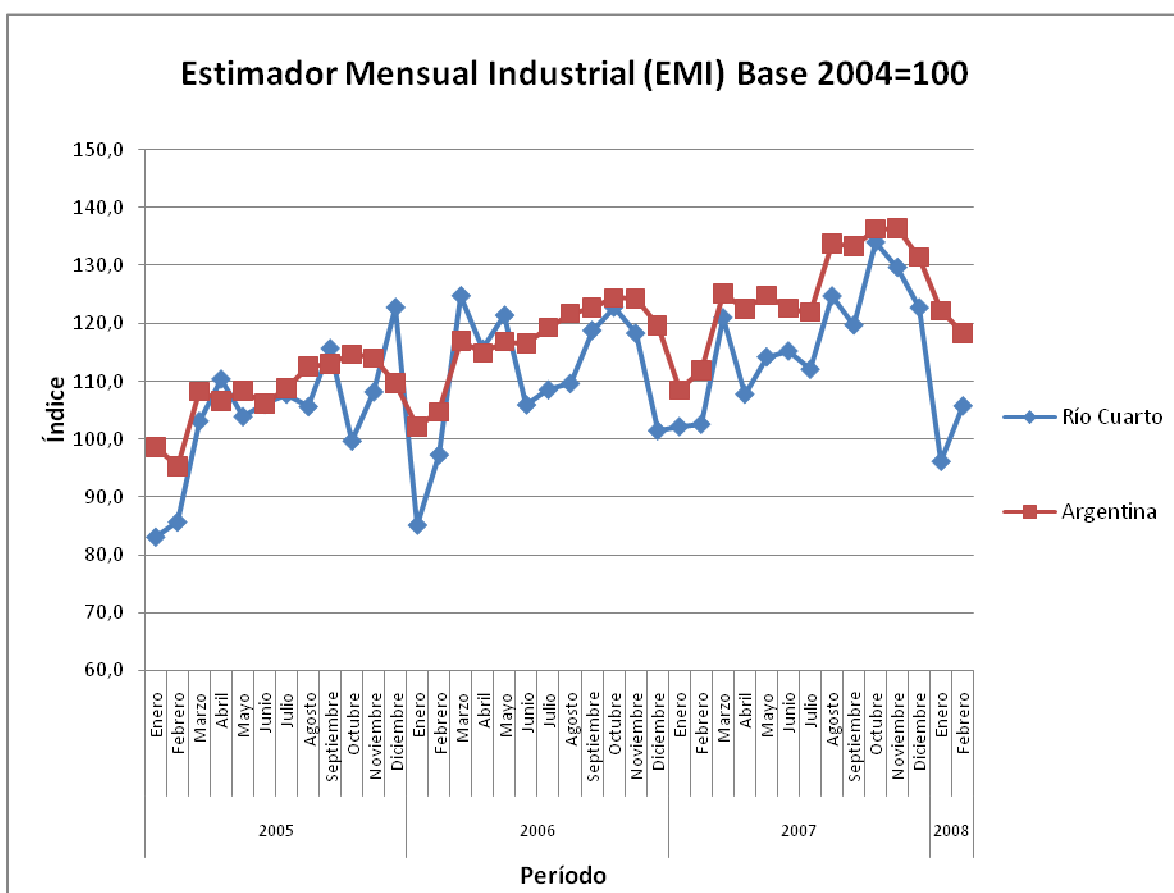
De todos ellos nos ocuparemos en los epígrafes siguientes.

### **Ciclo**

El ciclo es aquel componente de la serie vinculado a oscilaciones a medio plazo. Se suele considerar que no presenta un movimiento con patrón uniforme único, como la tendencia, o la estacionalidad, por lo que hablamos de oscilaciones en períodos superiores al año, aspecto que sirve, entre otros, para distinguirlo de la estacionalidad. En la práctica, se supone que en una serie económica se solapan distintos ciclos, por lo que es habitual considerar conjuntamente este componente con la tendencia, en un único que refleje movimientos de la serie a largo plazo, frente a movimientos causados por la estacionalidad.

## Estacionalidad

La estacionalidad es una componente que se presenta en series de frecuencia inferior a la anual (mensual, trimestral,...), y supone oscilaciones a corto plazo de período regular, inferior al año y amplitud regular. Se trata de la componente que introduce los matices más interesantes de cara a la predicción. En general, todas las series de frecuencia inferior a la anual presentan en mayor o menor medida estacionalidad. Un ejemplo puede ser esclarecedor:



En el gráfico se recoge la evolución del Índice de Producción industrial de Río Cuarto y Argentina, desde enero de 2005 hasta Febrero de 2008. Son datos mensuales, en base 2004=100, obtenidos de la Estimador Mensual Industrial que elabora el INDEC para Argentina y la Municipalidad de Río Cuarto para la ciudad. Podríamos discutir si la serie presenta nítidamente tendencia o ciclo (realmente es un período muy pequeño (para detectarlo gráficamente), pero este es un ejemplo de serie con estacionalidad. Obsérvese que año tras año, la serie presenta un comportamiento en los distintos meses de cada año muy similar. Todos los años la serie parece crecer en los primeros meses, decae ligeramente en abril, y, sobre todo, sufre una drástica caída de forma sistemática en el mes de enero. Y así año tras año. Se detecta pues un patrón regular. Efectivamente, esa evolución

de la serie responde a hechos sencillos. En los meses de enero y de abril (si en este mes cae la Semana Santa) las empresas notan el bajón en su producción debido a las vacaciones. Eso es estacionalidad.

Los procedimientos para obtener el componente estacional de la serie se analizarán con detalle en epígrafes posteriores.

### ***Componente irregular***

También llamado residual o aleatorio. El aspecto importante de este componente es que suponemos que el mismo no sigue ningún patrón sistemático de comportamiento que se pueda modelizar. En todo caso, los residuos se utilizan para comprobar la bondad de la aplicación realizada.

### **Una digresión: Relativas Cíclicas – Irregulares**

El factor componente de una serie de tiempo que se estudia con más frecuencia, es la tendencia. El estudio primario de la tendencia, sirve para fines predictivos

Para obtener alguna impresión visual de los movimientos generales a largo plazo en una serie de tiempo, se construye una gráfica en la cual se trazan los datos observados (variables dependientes) y los períodos de tiempo (variables independientes).

Si fuera posible ajustar en forma adecuada una tendencia lineal a los datos, los dos métodos en uso más extenso para ajustes de tendencias, son el de los mínimos cuadrados y el de la doble suavización exponencial.

Sin embargo, si los datos de la serie de tiempo indican algún movimiento curvilíneo descendente o ascendente a largo plazo, los dos métodos en mayor uso para el ajuste de tendencias son el de los mínimos cuadrados y el de la triple suavización exponencial.

1) **Método de los mínimos cuadrados:** El método de los mínimos cuadrados permite ajustar una línea recta de la forma:

$$\hat{Y}_i = \hat{\alpha} + \hat{\beta}X_i$$

De modo que los valores que se calculen para los dos coeficientes cumplan con el siguiente requisito: minimizar la suma de las diferencias al cuadrado

entre cada valor observado en los datos y cada valor estimado a lo largo de la línea de tendencia, es decir:

$$\sum_{i=1}^n \left( Y_i - \hat{Y}_i \right)^2 = \text{mínimo}$$

Para obtener la recta estimada se calcula la pendiente y la intercepción con:

$$\hat{\beta} = \frac{\sum_{i=1}^n X_i Y_i - \frac{\left( \sum_{i=1}^n X_i \right) \left( \sum_{i=1}^n Y_i \right)}{n}}{\sum_{i=1}^n X_i^2 - \frac{\left( \sum_{i=1}^n X_i \right)^2}{n}} = \frac{\Sigma(X - \bar{X})(Y - \bar{Y})}{\Sigma(X - \bar{X})^2}$$

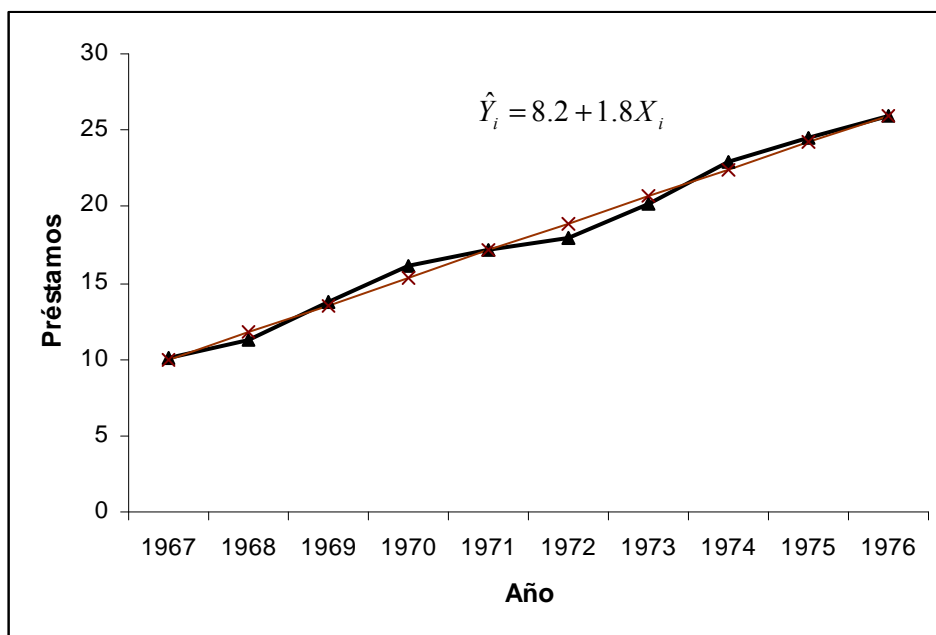
$$\hat{\alpha} = \bar{Y} - \hat{\beta} \bar{X}$$

PRÉSTAMOS ANUALES SOBRE PÓLIZAS Y PRIMAS FRACCIONADAS DE COMPAÑÍAS DE SEGUROS SOBRE LA VIDA (1967 - 1976)						
Año	X	Y	$(X_i - \bar{X})$	$(Y_i - \bar{Y})$	$(X_i - \bar{X})(Y_i - \bar{Y})$	$(X_i - \bar{X})^2$
1967	1	10,1	-4,5	-7,9	35,5	20,3
1968	2	11,3	-3,5	-6,7	23,4	12,3
1969	3	13,8	-2,5	-4,2	10,5	6,3
1970	4	16,1	-1,5	-1,9	2,8	2,3
1971	5	17,1	-0,5	-0,9	0,4	0,3
1972	6	18	0,5	0,0	0,0	0,3
1973	7	20,2	1,5	2,2	3,3	2,3
1974	8	22,9	2,5	4,9	12,3	6,3
1975	9	24,5	3,5	6,5	22,8	12,3
1976	10	25,9	4,5	7,9	35,6	20,3
n = 10	55	179,9	0	0,0	146,7	82,5

Una vez obtenida la línea  $\hat{Y}_i = \hat{\alpha} + \hat{\beta}X_i$ , se pueden substituir los valores de  $X_i$ , en la ecuación para predecir diversos valores para  $Y_i$ .

Al trabajar con datos de una serie de tiempo deben codificarse los valores de  $X_i$  asignándole códigos crecientes de enteros: 1, 2, 3, 4, ..., n.

Por ejemplo, la serie de tiempo presentada en la tabla y trazada en la figura, representa los pagos anuales (en miles de millones de dólares) a las compañías de seguros de vida, tanto por intereses sobre préstamos con garantía de la póliza como por primas fraccionadas en el período de 10 años de 1967 a 1976.



Se determina que:

$$\hat{\beta} = \frac{\sum(X - \bar{X})(Y - \bar{Y})}{\sum(X - \bar{X})^2} = \frac{146.7}{82.5} = 1.8$$

y dado que:

$$\bar{Y} = \frac{\sum_{i=1}^n Y_i}{n} = \frac{179.9}{10} = 17.99 \quad y \quad \bar{X} = \frac{\sum_{i=1}^n X_i}{n} = \frac{55}{10} = 5.5$$

$$\hat{\alpha} = \bar{Y} - b_1 \bar{X} = 17.99 - (1.8)(5.5) \cong 8.2$$

Se tiene:

$$\hat{Y}_i = 8.2 + 1.8X_i$$

La intercepción  $\hat{\alpha} = 8.2$ , es el valor ajustado de la tendencia que refleja la cantidad de dinero (en miles de millones de dólares) pagado a las compañías de seguros de vida por intereses por préstamos con garantía de la póliza y por primas fraccionadas durante el año 1966. La pendiente  $\hat{\beta} = 1.8$ , indica que tales pagos van aumentando a razón de 1.800 millones de dólares por año.

Para ajustar la línea de tendencia a los años observados de la serie, sólo se substituyen los valores codificados correspondientes a  $X_i$ , en la ecuación.

Por ejemplo, en el año 1975, en donde  $X_{1975} = 9$ , el valor predicho (ajustado) de la tendencia, se expresa con:

$$\hat{Y}_{1975} = 8.2 + (1.8)(9) = 24.2 \text{ miles de millones de dolares}$$

Para usar la recta de tendencia en predicción, se puede proyectar la línea ajustada hacia el futuro mediante extrapolación matemática. Por ejemplo, para predecir la tendencia en los pagos para el año 1977, se substituye  $X_{1977} = 11$ , el código para el año 1977, en la ecuación y se pronostica que la tendencia será:

$$\hat{Y}_{1977} = 8.2 + (1.8)(11) = 28.0 \text{ miles de millones de dolares}$$

2) **Aislamiento y eliminación de la tendencia.** Se estudió la tendencia como una ayuda para el pronóstico a corto plazo. Pero, los economistas o los pronosticadores de negocios también pueden desear el estudio de la tendencia de modo que los factores que influyen en ella, se puedan eliminar del modelo multiplicativo de series de tiempo clásico y, por tanto, proveer la estructura para el pronóstico a corto plazo de la actividad general de los negocios. El procedimiento de aislar y eliminar un factor componente de los datos, se llama descomposición de las series de tiempo.

Como el método de los mínimos cuadrados provee valores  $\hat{Y}_i$  de tendencia "ajustados" para cada año en la serie, se puede eliminar con facilidad la componente de tendencia del modelo multiplicativo de series de tiempo clásico (porque en cualquier año dado la componente de la tendencia, se estima con  $\hat{Y}_i$ ).

Por tanto, la componente de la tendencia, se puede eliminar mediante una división como sigue:

$$Y_i = T_i \cdot C_i \cdot I_i$$

De modo que:

$$\frac{Y_i}{\hat{Y}_i} = \frac{T_i \cdot C_i \cdot I_i}{\hat{Y}_i}$$

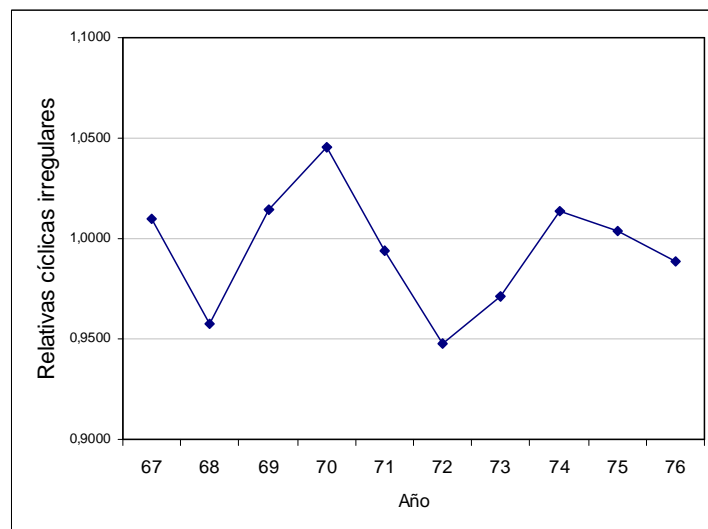
OBTENCIÓN DE LAS RELATIVAS CÍCLICAS - IRREGULARES				
Año	$X_i$	$Y_i$	$\hat{Y}_i = 8.2 + 1.8X_i$	$Y_i/\hat{Y}_i$
(1)	(2)	(3)	(4)	(5)
1967	1	10,1	10	1,010
1968	2	11,3	11,8	0,958
1969	3	13,8	13,6	1,015
1970	4	16,1	15,4	1,045
1971	5	17,1	17,2	0,994
1972	6	18	19	0,947
1973	7	20,2	20,8	0,971
1974	8	22,9	22,6	1,013
1975	9	24,5	24,4	1,004
1976	10	25,9	26,2	0,989

Pero como  $\hat{Y}_i = T_i$  se tiene:

$$\frac{Y_i}{\hat{Y}_i} = \frac{T_i \cdot C_i \cdot I_i}{T_i} = C_i \cdot I_i$$

Las proporciones entre los valores observados y los valores de las tendencias ajustadas,  $Y_i/\hat{Y}_i$  que se calculan cada año en la serie, se llaman **relativas cíclicas-irregulares**. Estos valores, que fluctúan alrededor de una base de 1.0, muestran la actividad tanto cíclica como irregular en la serie.

Volviendo al ejemplo anterior, los cálculos de las relativas cíclicas-irregulares, se muestran en la siguiente tabla para los datos de las series de tiempo de 10 años que reflejan los pagos anuales a las compañías de seguros por intereses de préstamos con garantía de póliza y por pago de primas fraccionadas. Los valores de tendencia ajustada [columna (4)], se determinan con solo substituir los valores apropiados con el código x [columna (2)], al modelo de tendencia lineal obtenido con el método de los mínimos cuadrados. Para cada año en la serie, se ve que el valor observado [columna (3)], se divide después entre el valor de tendencia ajustada [columna (4)] para producir la relativa cíclica-irregular [columna (5)].



Muchas veces para tener una aproximación del componente ciclo y del componente irregular, se utilizan las relativas cíclicas - irregulares, usando la tendencia de la serie.



De esta forma, por ejemplo con los datos del EMI de la ciudad de Río Cuarto se podría calcular primero la tendencia y luego el componente cíclico e irregular.

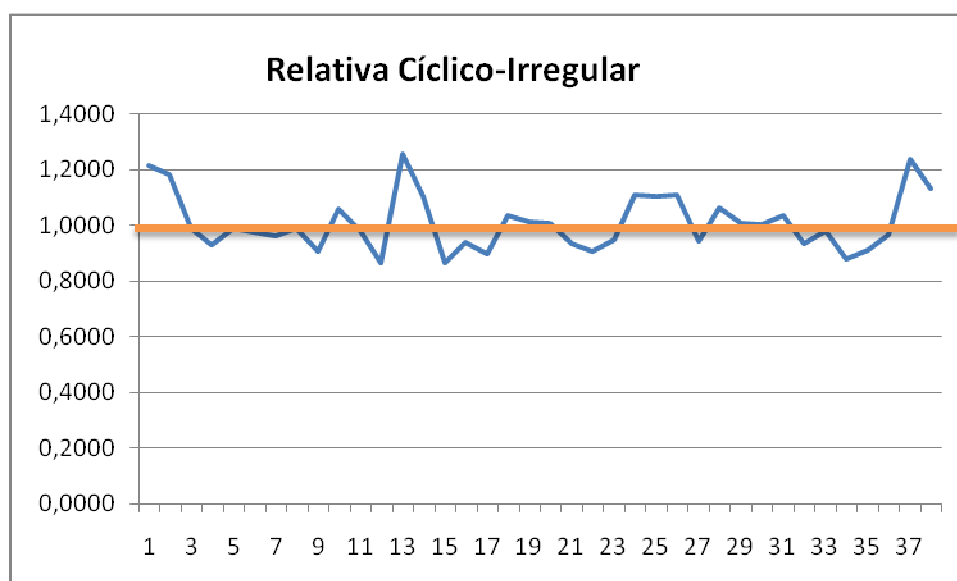
Mes	EMI	Tendencia	Relativas C-I
1	83,1	100,9	1,2146
2	85,7	101,4	1,1835
3	103,1	101,9	0,9880
4	110,4	102,4	0,9273
5	103,9	102,9	0,9903
6	106,1	103,4	0,9751
7	107,8	103,9	0,9640
8	105,6	104,4	0,9885
9	115,7	104,9	0,9067
10	99,7	105,5	1,0578
11	108,2	106,0	0,9792
12	122,8	106,5	0,8667
13	85,1	107,0	1,2568
14	97,3	107,5	1,1047
15	124,9	108,0	0,8649
16	115,6	108,5	0,9384
17	121,5	109,0	0,8974
18	106,0	109,5	1,0336
19	108,6	110,0	1,0135
20	109,6	110,5	1,0082
21	118,8	111,1	0,9344
22	122,8	111,6	0,9081
23	118,4	112,1	0,9468
24	101,5	112,6	1,1092
25	102,2	113,1	1,1066
26	102,6	113,6	1,1071
27	121,1	114,1	0,9424
28	107,8	114,6	1,0631
29	114,2	115,1	1,0076
30	115,3	115,6	1,0030
31	112,1	116,1	1,0358
32	124,8	116,6	0,9345
33	119,8	117,2	0,9781
34	134,1	117,7	0,8771
35	129,8	118,2	0,9105
36	122,8	118,7	0,9665
37	96,2	119,2	1,2395
38	105,8	119,7	1,1315

Para el cálculo de la tendencia se utiliza la fórmula:

$$EMI = \alpha + \beta T$$

Donde T representa el tiempo medido en meses. Tanto el valor mensual estimado del EMI como de la tendencia se muestran en la siguiente tabla de datos.

El siguiente gráfico muestra las relativas cíclicas e irregulares. Por debajo de 1,0000 se observan los ciclos de contracción de la actividad industrial y por arriba los de expansión de la misma.



### **La necesidad de información: los datos, su localización y su tratamiento elemental.**

La mayor parte de las técnicas de predicción hacen uso de información histórica. La información histórica queda cuantificada en series estadísticas. Las técnicas de predicción elemental que hemos visto precisan datos de varios años sobre la variable a predecir. A continuación vamos a analizar tres aspectos; 1) ¿Qué tipo de datos podemos encontrar en el análisis económico?; 2) ¿Dónde podemos encontrar tales datos?; y 3) ¿Qué tratamientos especiales debemos aplicar a esos datos.

#### **1) Tipos de datos.**

Las técnicas de predicción, habitualmente, exigen disponer de datos suficientes sobre el fenómeno a predecir. En general, podemos distinguir

tres tipos de datos utilizables en economía y gestión de empresas. Estos datos con lo que nos encontramos en el análisis pueden ser:

**Series temporales** (serie cronológica o histórica). Son valores de una variable a lo largo del tiempo, es decir, una serie temporal son un conjunto de observaciones, en general de frecuencia regular, sobre una variable determinada, para distintos momentos de tiempo. Ejemplos encontramos en los datos anuales de la Producción de Cereales, los trimestrales del PIB, los indicadores mensuales de coyuntura económica, la evolución de ventas de empresa, etc.

Un concepto muy importante relacionado con los datos de series temporales es el de frecuencia. Por tal entendemos el período de tiempo que separa dos de sus datos, bien entendido que cuanto menor es el período transcurrido, mayor es la frecuencia de los datos. En general, los datos que manejaremos presentan frecuencia anual, mensual o trimestral.

**Datos de corte transversal** (*cross section*): Son valores para diferentes sujetos en un momento dado, esto es, son un conjunto de observaciones de distintos individuos o elementos relativos a un mismo momento de tiempo. El ejemplo clásico de datos de corte transversal lo constituye la Encuesta Permanente de Hogares (EPH) realizada por el INDEC. En esta encuesta se entrevista a un número determinado de familias para determinar distintas variables sobre trabajo de los habitantes del país.

**Datos de Panel** (*Panel data*): Los datos de panel son una combinación de datos de serie temporal y de datos de corte transversal. En los datos de panel se obtienen observaciones sobre distintas unidades en diversos momentos de tiempo, bien entendido que al menos en una parte de las unidades a las que se recaba información no varían en los distintos períodos de tiempo. La Encuesta Mensual Industrial (EMI) que elabora el INDEC constituye un ejemplo de datos de panel. Se entrevista mensualmente a un número considerable de empresas y se realiza un análisis sectorial. Uno podría realizar el análisis solo de la serie agregada o de un determinado sector. Cuando se analizan las dos cosas en forma conjunta estamos en presencia de un análisis de datos de panel. Como veremos en unidades posteriores este caso merece un estudio econométrico particular.

En predicción elemental utilizaremos datos de series temporales, si bien, la simulación y algunas técnicas avanzadas de predicción también hacen uso de los otros tipos. En todo caso, la necesidad de datos no debe entenderse

en el sentido de que no es posible incorporar en el análisis información cualitativa. Como hemos visto, también es posible realizar predicción a partir de información cualitativa. Además, parte de esa información puede, fácilmente, cuantificarse. Algunos casos son: 1) Cuantificaciones inmediatas de aspectos inicialmente cualitativos; 2) Mediciones indirectas de variables cualitativas, como un índice de expectativas de los consumidores; 3) Relaciones de orden y no cuantificaciones estrictas de las variables, como en el caso de establecimiento de semejanzas y preferencias; y 4) Variables definidas por su pertenencia o no a un conjunto, por ejemplo renta alta, baja o media.

## **2) Localización de la información.**

La información contenida en las series estadísticas puede provenir de diversas fuentes alternativas que podemos sistematizar de la siguiente manera:

**Información directa** (Fuentes primarias): Se genera la información específica que precisa el análisis a realizar. A su vez, podemos bien recurrir a los registros propios de la empresa o institución a la que atañe la predicción, o bien recurrir a mediciones directas realizadas específicamente para el desarrollo de la aplicación, por ejemplo a través de encuestas.

**Utilización de estadísticas previamente generadas por otros organismos** (Fuentes secundarias): Esta es la situación más común, y puede consistir tanto en publicaciones estadísticas o bancos de datos. Las posibilidades que ofrecen las tecnologías de la información y las comunicaciones facilitan extraordinariamente la obtención de datos.

Hoy día cualquiera de los proveedores de información estadística (institutos regionales de estadística, bancos, centros de investigación, etc.) proporcionan innumerables datos vía electrónica.

## **3) Tratamiento de los datos.**

Una vez localizadas las fuentes estadísticas, y como paso previo, es frecuente, y en multitud de casos imprescindible, la realización de algunos tratamientos que garanticen la bondad y significatividad de los resultados obtenidos en las distintas aplicaciones.

Para ello precisamos de la disponibilidad de series estadísticas homogéneas, que no presenten discontinuidades, que estén medidas en las mismas unidades, y lo más actualizadas posibles. Por todo ello se hace necesario disponer de una serie de procedimientos de actuación para tratar de solventar los problemas más frecuentes que pueden presentarse en los procesos de obtención de información.

**Deflación de series estadísticas:** El primero de los procedimientos, que merece una especial atención por su singularidad, es el de la deflación de series estadísticas. Uno de los principales problemas que presenta el manejo de las series estadísticas que utilizamos en el análisis económico aplicado es que mayoritariamente son magnitudes económicas expresadas en precios. Esto, que en última instancia es lo que permite la realización de comparaciones entre magnitudes que encierran conceptos heterogéneos, tiene un serio problema, ya que habitualmente se presentan en términos corrientes. Como sabemos, una magnitud monetaria se dice que está expresada en términos corrientes si dicha magnitud se encuentra valorada a los precios del periodo en que se mide. Por el contrario una magnitud se dice que está expresada en precios constantes base  $t$  si dicha magnitud se ha valorado a los precios de un determinado período  $t$ .

El problema, queda planteado en los siguientes términos: si disponemos de una serie estadística expresada en términos corrientes (*en pesos corrientes*) y pretendemos realizar una comparación entre dos períodos distintos, puesto que la serie no es homogénea, tendremos que expresar la serie en términos constantes (*en pesos constantes*) para poder hacer la comparación. Para pasar de una serie en pesos corrientes a una en pesos constantes hay que dividir la serie en términos corrientes por el cociente entre el índice de precios (que corresponda) del período corriente sobre el índice de precio del período base, eliminando, de esta manera, la influencia de los precios. De esta forma, las conclusiones a las que llegásemos en nuestro estudio, como por ejemplo el cálculo de crecimiento de los distintos agregados económicos, serían en valores reales, frente a la posibilidad alternativa, en cuyo caso hablaríamos de crecimientos en valores nominales.

Como sabemos, el paso de una serie en unidades monetarias corrientes a una en unidades constantes se conoce con el nombre de deflación de la serie y al índice elegido para efectuar dicha transformación se le llama deflactor, que técnicamente es definido como el índice de precios que se obtiene por comparación entre una magnitud en términos corrientes y la misma magnitud en pesos constantes. La base de ese índice de precios será, evidentemente, la base en que venga expresada la magnitud en pesos constantes.

**Cambios de base:** Con cierta frecuencia el analista suele encontrarse con que las series con las que trabaja vayan cambiando los periodos de referencia, y ello sin necesidad de que haya ocurrido un cambio metodológico. Pero a veces resulta interesante en sí mismo cambiar esta base de referencia de la magnitud medida en términos constantes, para conseguir que tal magnitud sea representativa de la dinámica del fenómeno que se está explicando.

Para realizar esta operación es suficiente con cambiar la base de referencia del deflactor y después aplicarse sobre la magnitud en corrientes. Para realizar el cambio de base basta con disponer del valor del deflactor inicial en el nuevo año base para, a continuación, realizar una sencilla regla de tres. Debe entenderse que, obviamente, el crecimiento de una magnitud en constantes es el mismo independientemente de la base en la que se exprese el deflactor.

**Cambios metodológicos:** Un problema algo más complejo resulta cuando nos enfrentamos a cambios metodológicos en los procesos de elaboración de series estadísticas, como los que habitualmente se realizan en los sistemas de cuentas nacionales, o en la medición del desempleo, y que en principio nos obligaría a disponer de toda la información original para poder reconstruir las series de acuerdo con la nueva metodología. Sin embargo, en los casos en los que se presente este problema, se suele acudir a procedimientos pragmáticos para la obtención de una serie homogénea.

### *B.3. Medias móviles*

#### Resumen:

- Se trata de una técnica que utiliza información estadística pasada, y es considerada como técnica naïve (en sentido amplio) y de alisado.
- La técnica de medias móviles construye una nueva serie a partir de la media de un número determinado de datos, en la que se va añadiendo sucesivamente un dato nuevo y quitando, al mismo tiempo, el más antiguo de los datos incluidos en la media anterior.
- La expresión general de una media móvil de orden  $S$  consistiría en calcular una serie que para cada momento  $t$  toma el siguiente valor:

$$M_t = \frac{y_t + y_{t-1} + \dots + y_{t-s+1}}{s}$$

- La técnica se aplica óptimamente en series sin tendencia ni estacionalidad. En ese caso, la predicción se realiza con la última media móvil calculada, es decir:

$$\hat{y}_{t+1} = M_t$$

- Cuando hay tendencia, debe corregirse el efecto del incremento medio anual, es decir, para predecir valores mensuales:

$$\hat{y}_{t+1} = M_t + \frac{y_t - y_{t-12}}{12}$$

- Este resultado es similar a calcular

$$\hat{y}_{t+1} = y_t + \Delta\hat{y}_{t+1} = y_t + M_t = y_t + \frac{y_t - y_{t-12}}{12}$$

- Cuando la serie presenta estacionalidad, la media móvil debe ser corregida por coeficientes de estacionalidad, tal y como se desarrollará posteriormente.
- Versiones alternativas consistirían en atribuir ponderaciones distintas a cada valor incluido en la media móvil (normalmente más ponderación con el dato más próximo), es decir, calcular *medias móviles ponderadas*, y *dobles medias*, tal y como se desarrolla en el ejercicio práctico 1.
- No existen criterios generales sobre el orden adecuado de la media móvil.

### ***¿Qué es un método naïve?***

En general, y aplicado a múltiples contextos, es frecuente denominar como *naïve* (ingenuo) a aquel procedimiento de predicción que repite de forma mecánica un comportamiento pasado. De hecho, gran parte de las técnicas que examinamos en este acápite se pueden calificar de *naïve*.

Por ejemplo, el conocido como **modelo naïve I** realiza predicción asumiendo que el valor futuro (en el período  $t + 1$ ) de una variable (es decir, el valor predicho,  $\hat{y}_{t+1}$ ) coincide con el valor actual (el valor de la variable en el momento  $t$ ), es decir:

$$\hat{y}_{t+1} = y_t$$

Una versión alternativa, conocido como **modelo naïve II**, asumiría no la igualdad del valor, sino del incremento, es decir:

$$\hat{Y}_{t+1} - y_t = y_t - y_{t-1}$$

Obviamente, se trata de procedimientos muy simples, por lo que se puede utilizar como predicción el valor medio de un período, en lugar de sólo el último:

$$\hat{Y}_{t+1} = \bar{y}$$

Sin embargo esto sólo sería aconsejable cuando la variable o serie no tuviese tendencia, sino que oscilase aleatoriamente entorno a la media. En ese caso, la media es el valor más probable de predicción.

Como la mayor parte de las series económicas presentan tendencia (piénsese en una serie como el PIB, que más allá de caídas puntuales, muestra tendencia ascendente a lo largo del tiempo), se pueden elaborar alternativas igualmente "ingenuas". Una primera consiste añadir en el modelo *naïve* I un término de tendencia, dada una tasa de crecimiento constante, es decir:

$$\hat{Y}_{t+1} = (1 + c) \cdot y_t$$

Una segunda opción consiste en eliminar la tendencia de la serie, predecir la nueva serie y posteriormente añadir la tendencia. Para ello suele ser suficiente con calcular la primera diferencia de la serie, es decir, si  $\Delta y_t$  representa la serie en diferencias, se trata de calcular para cada momento de tiempo la expresión:

$$\Delta y_t = y_t - y_{t-1}$$

En todo caso, si trabajamos con medias aritméticas como predictor aplicadas a la serie en incrementos, estaríamos utilizando la media de los incrementos, que es igual al recorrido de la serie en niveles (primer valor menos último), dividido por el número de períodos. Es decir:

$$\bar{\Delta y}_t = \frac{1}{n \cdot [(y_t - y_{t-1}) + (y_{t-1} - y_{t-2}) + \dots + (y_1 - y_0)]} = \frac{1}{n \cdot (y_t - y_0)}$$

Con lo que para predecir el incremento en la serie se actuaría:

$$\Delta \hat{y}_{t+1} = \bar{\Delta y}_t = \frac{1}{n \cdot (y_t - y_0)}$$

De forma que la predicción de la serie en niveles adoptaría la expresión:

$$\hat{y}_{t+1} = y_t + \Delta \hat{y}_{t+1} = y_t + \frac{1}{n \cdot (y_t - y_0)}$$

En todo caso, la práctica operativa de este procedimiento exigiría calcular, período a período, la media de los incrementos en los valores de la serie para todo el período muestral para actualizar la predicción. Ello introduce, además de falta de economicidad, dos problemas:

- (1) por una parte cuando el período muestral es muy extenso se corre el riesgo de que cambios estructurales (es decir, cambios sustanciales en el comportamiento del fenómeno) acontecidos en un momento determinado no queden reflejados correctamente, puesto que el valor de predicción se forma con todo el período muestral y no sólo con los más recientes.
- (2) Un segundo problema, más importante con períodos muestrales cortos, hace referencia a la estacionalidad: algunas series económicas presentan valores atípicos (por exceso o por defecto) en determinados períodos del año. Por ejemplo, la producción industrial cae de forma muy relevante en enero. A la hora de predecir el valor

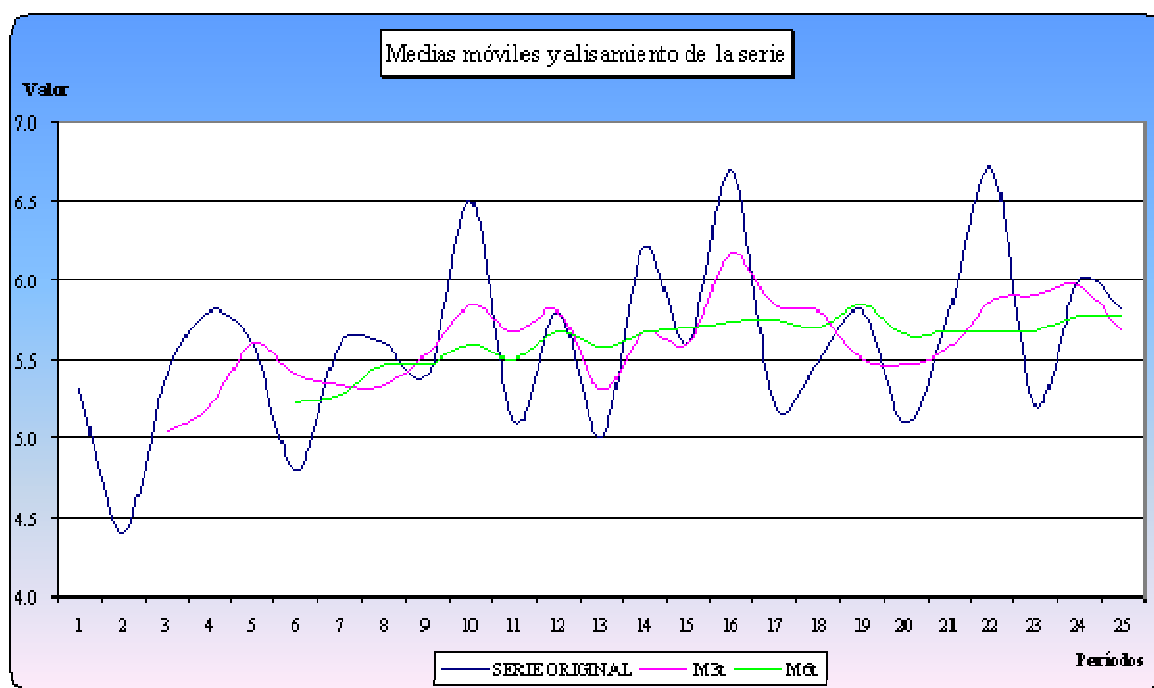


de tal serie en un mes como febrero, la inclusión de un mes como enero en el cálculo de la media, puede infravalorar el valor de predicción.

Por todo ello, es más frecuente utilizar la técnica conocida como *medias móviles*.

### ¿Qué es un método de alisado?

Reciben este nombre aquellas técnicas, como las medias móviles o el alisado exponencial, que "alisan", en el sentido de moderar, las variaciones que una serie económica pueda presentar, sean estas estacionales (sólo en determinados momentos del año), cíclicas (recurrentes cada ciertos años, es decir, debidas al momento del ciclo económico) o erráticas. Con frecuencia resulta de interés eliminar estos comportamientos, puesto que introducen ruido y, con frecuencia, no ayudan en la predicción. En última instancia, estas técnicas provocan que la serie presente un comportamiento más estable.



La técnica de medias móviles, en sus múltiples variantes, provoca esta circunstancia en las series económica, tanto más cuanto mayor sea el número de términos en la media móvil. Esto se observa con claridad con un ejemplo como el recogido en el gráfico inferior. Se ha representado la serie original utilizada en el Ejercicio 1 (a continuación), junto con la media móvil simple de tres (M3t) y de seis términos (M6t). Aprecie cómo, efectivamente, la técnica modera las variaciones de la serie.

### EJERCICIO 1: Cálculo de Medias móviles

Un gran almacén decide aplicar la técnica de medias móviles como método rápido y sencillo de predicción de ventas, a efectos de determinar las necesidades de personal temporal de ventas en un período normal (sin oscilaciones estacionales ni tendencia, al menos especialmente significativas). A partir de la serie y (ventas semanales en millones de pesos), se pide calcular en excel:

- Medias móviles asimétricas de tres términos (M3t).
- Medias móviles asimétricas de seis términos (M6t).
- Medias móviles asimétricas ponderadas de tres términos y ponderaciones 3, 2, 1 (M3wt).
- Doble media móvil de tres y tres términos (M3.3t).

Además realice predicción para dos períodos. Utilice para todo ello la siguiente tabla. Para trabajar copie esta tabla en un libro Excel (fíjese que el primer dato esté ubicado en la **CELDA B17 Y SE REFIERA AL PERÍODO 1** y siga la solución para cada apartado que se detalla a continuación.

	A	B	C	D	E	F	G
16		Periodo	VENTAS	M3t	M6t	M3wt	M3.3t
17		1	5,3	--	--	--	--
18		2	4,4	--	--	--	--
19		3	5,4		--		--
20		4	5,8		--		--
21		5	5,6		--		
22		6	4,8				
23		7	5,6				
24		8	5,6				
25		9	5,4				
26		10	6,5				
27		11	5,1				
28		12	5,8				
29		13	5,0				
30		14	6,2				
31		15	5,6				
32		16	6,7				
33		17	5,2				
34		18	5,5				
35		19	5,8				
36		20	5,1				
37		21	5,8				
38		22	6,7				
39		23	5,2				
40		24	6,0				
41		25	5,8				
42	Predicción	26	--				
43		27	--				

**Solución:****a) MEDIAS MÓVILES DE TRES TÉRMINOS.**

La media móvil de tres términos se calcularía asignando a cada momento del tiempo ( $t$ ) el resultado de sumar los valores de serie y en el momento  $t$ ,  $t-1$ , y  $t-2$ , y dividir por 3. Por ejemplo, la media móvil de orden 3 para el período 10 se calcularía como  $(6,5+5,4+5,6)/3 = 5,83$ .

Para el cálculo en Excel sitúese en la celda D19 y escriba en lenguaje Excel la fórmula de la media móvil de ese período de tiempo, es decir,  $=(C19+C18+C17)/3$  y pulse INTRO. Vuelva a marcar la celda D19 y sitúe el cursor en la esquina inferior derecha, hasta que el cursor se transforme en forma de aspa. Haga CLICK y arrastre el cursor hasta D41. Excel generará automáticamente la media móvil para toda la serie.

Para realizar la predicción para el período 26, asumimos como predicción el último valor de la serie de medias móviles, es decir, la predicción de la serie y en 26 es igual a  $M3t$  en el período 25. La predicción en el período 27 exigiría calcular la media móvil **de** orden 3. Para ello utilizaremos el valor de predicción en el período 26, y los valores de la serie y en los períodos 24 y 25.

En lenguaje Excel, escribiríamos en la celda D42, la fórmula  $=D41$ , mientras que en la celda D43 escribiríamos  $=(D42+C41+C40)/3$ .

**b) MEDIAS MÓVILES DE SEIS TÉRMINOS.**

La media móvil de seis términos se calcula en forma análoga. Por ejemplo, la media móvil de orden 6 para el período 10 se calcularía como  $(6,5+5,4+5,6+5,6+4,8+5,6)/6 = 5,58$ .

En Excel la fórmula de la media móvil para el primer período posible, el 6, sería  $(C22+C21+C20+C19+C18+C17)/6$ . La predicción del período 26 se formularía como  $=E41$ , y la del período 27 como  $=(E42+C41+C40+C39+C38+C37)/6$ .

**c) MEDIAS MÓVILES PONDERADAS DE TRES TÉRMINOS.**

El cálculo es semejante, pero ahora atribuimos a cada valor de la serie y la ponderación correspondiente. Por ejemplo, la media móvil de orden 3 para el período 10 se calcularía como el cociente entre el resultado de  $(6,5 \times 3)+(5,4 \times 2)+(5,6 \times 1)$ , y la suma de las ponderaciones, es decir  $(3+2+1)$ . Por tanto el valor para el período 10 sería igual a  $(35,9/6)=5,98$ . En lenguaje Excel anotaríamos en la celda F19 la fórmula  $=(3*C19+2*C18+1*C17)/(3+2+1)$ , y arrastraríamos el resultado hasta la celda F41.

Para la predicción se actúa en forma análoga a las anteriores. La predicción para el período 26 sería la media móvil obtenida en el período 25 ( $=F41$ ), mientras que para la predicción del período 27, se calcularía la medida móvil ponderada incluyendo el valor predicho para el período 26,  $=(3*F42+2*C41+1*C40)/(3+2+1)$ .

#### d) DOBLE MEDIA MÓVIL DE TRES Y TRES TÉRMINOS.

Se distinguen dos cálculos: en el primero se calcula la media móvil simple de orden 3 (punto a). En el segundo cálculo, se elabora una media móvil de orden 3 a partir de la serie obtenida por media móvil simple.

Por ejemplo, el valor asignado a la doble media móvil de orden 3, para el período 10 viene dada por los valores de la media móvil simple de orden 3 para el período 10, 9 y 8. En definitiva, se calcularía como  $(5,8+5,5+5,3)/3= 5,53$ .

Por tanto, en lenguaje Excel anotaríamos en la celda G20 la fórmula  $=(D21+D20+D19)/3$ , arrastrando hasta G41. La predicción para el período 26 se obtendría como  $=(G41+D41+D40)/3$ , mientras que la predicción del período 27 se obtiene a partir de la fórmula  $=(G42+G41+D41)/3$ .

En definitiva, los resultados que obtenga deben ser similares a los recogidos en la siguiente tabla de la hoja Resultados. Además recogemos el error cuadrático medio (ECM) y el porcentaje de error absoluto medio (PEAM) de cada variante de predicción.

#### B.4. Alisado exponencial sin tendencia: el alisado simple

##### Resumen:

- Se trata de una media móvil con ponderaciones decrecientes en forma de progresión geométrica:

$$M_t = \alpha \cdot y_t + \alpha \cdot (1 - \alpha)^1 \cdot y_{t-1} + \alpha \cdot (1 - \alpha)^2 \cdot y_{t-2} + \dots$$

Donde alpha es una constante que varía entre cero y uno, y la suma de los coeficientes de ponderación es uno.

- El valor otorgado a la constante alpha determina el número de términos de la media móvil.
- Una vez elegido el valor de alpha y calculada la serie, podemos pasar a realizar predicción.
- Tal y como se ha descrito, esta técnica sólo es utilizable cuando la serie no presenta tendencia ni estacionalidad.

#### **¿Cuál deber ser la constante alfa?**

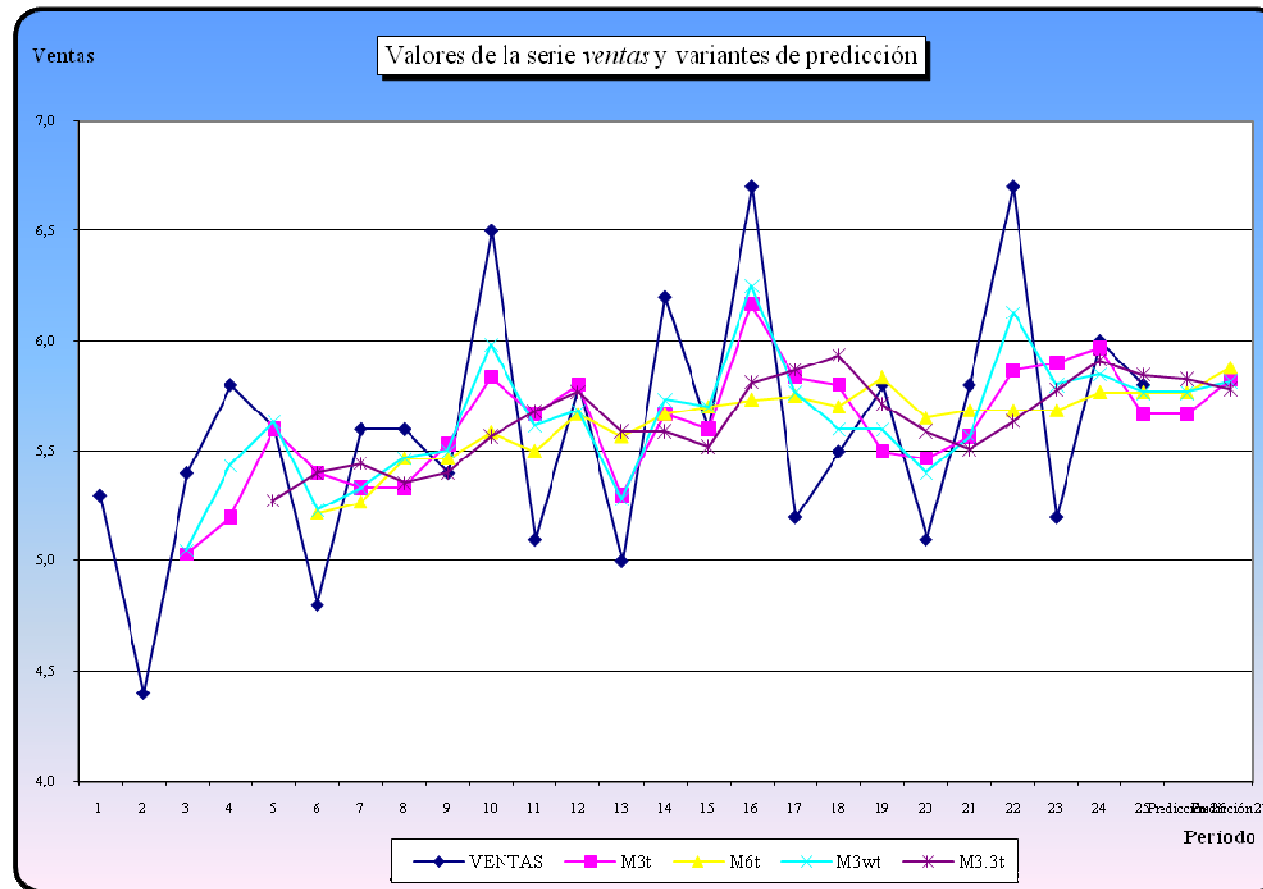
Al coeficiente alfa se le debe atribuir un determinado valor. Tal valor está comprendido entre cero y uno, y su elección tiene importantes consecuencias.

En primer lugar, si el coeficiente está entre cero y la unidad, entonces los coeficientes que acompañan a cada valor de la serie original en el cálculo de la serie transformada, se convierten en coeficientes de ponderación cuya suma vale uno. Se demuestra que:

$$\alpha + \alpha \cdot (1 - \alpha)^1 + \alpha \cdot (1 - \alpha)^2 + \dots = \alpha \sum_{s=0}^{\infty} (1 - \alpha)^s = \frac{\alpha}{1 - (1 - \alpha)} = 1$$



Por último, se incluye un gráfico que refleja el modo en que estas variantes de predicción transforman la serie original. Obsérvese, en especial, el modo en que las técnicas "alisan" los valores originales, obteniendo una serie con menor variabilidad.



Ello implica que cuanto mayor sea el valor del coeficiente alfa, menor será el número de términos incluido en la media móvil. Cuando  $\alpha=1$ , el valor de la media móvil coincide con el valor de la serie en el período. Cuando alfa se aproxima a cero, las ponderaciones son muy pequeñas y, por tanto, se incluyen gran número de términos. La relación entre el número de términos y el valor del coeficiente alfa es aproximadamente:

$$s = \frac{(2 - \alpha)}{\alpha}$$

La elección del parámetro alfa depende de las características de la serie objeto de estudio. Hoy día los programas de ordenador permiten el cálculo automático del valor óptimo de alfa, en el sentido de elegir aquel que minimiza el error cuadrático medio. En general se considera que un alfa alto es indicativo de fuertes oscilaciones o de tendencia en la serie, lo que conlleva un reducido alisamiento para un mejor ajuste a esas oscilaciones. Por el contrario, para una serie con pequeñas oscilaciones irregulares se aconseja un alfa reducido (entre 0,01 y 0,4) que implica un fuerte alisado de la serie, al incluir un elevado número de términos.

### ***¿Cómo predecir con alisado exponencial?***

Como fórmula de predicción la media deberá empezar a calcularse comenzando por el último dato disponible, es decir:

$$\hat{y}_{t+1} = \alpha \cdot y_t + \alpha \cdot (1 - \alpha)^1 \cdot y_{t-1} + \alpha \cdot (1 - \alpha)^2 \cdot y_{t-2} + \dots$$

por sustituciones sucesivas, podríamos llegar a una expresión alternativa resumida:

$$\hat{y}_{t+1} = \alpha \cdot y_t + (1 - \alpha) \cdot \hat{y}_t$$

lo que nos permite interpretar la predicción con alisado como una media ponderada de los valores previos anteriores reales y de predicción.

También podemos expresar la predicción en función del término de error  $e_t = y_t - \hat{y}_t$ :

$$\hat{y}_{t+1} = \alpha \cdot y_t + (1 - \alpha) \cdot \hat{y}_t = \alpha \cdot (\hat{y}_t + e_t) + (1 - \alpha) \cdot \hat{y}_t = \alpha \cdot e_t$$

lo cual nos muestra que las variaciones del valor de predicción  $\hat{y}_{t+1} - \hat{y}_t$  son una proporción,  $\alpha$ , del error del período anterior:

$$\hat{y}_{t+1} - \hat{y}_t = \alpha \cdot e_t$$

Ello implica que las predicciones sucesivas serán necesariamente iguales, al no disponer de los correspondientes errores y suponerse un valor nulo.

De todas las expresiones anteriores a efectos de cálculo suele utilizarse

$$\hat{y}_{t+1} = \alpha \cdot y_t + (1 - \alpha) \cdot \hat{y}_t$$

que exige disponer de un valor para  $\alpha$  y de un valor inicial de  $\hat{y}_t$ . Para esto último se suele adoptar el **primer valor de la serie**,  $\hat{y}_1 = y_1$  o utilizando **la media de un número reducido de las primeras observaciones**.

### ¿Cómo influye la tendencia en las medias móviles?

El alisado exponencial, como cualquier otra media móvil, sea ponderada o no, sólo es utilizable cuando la serie económica original no presenta tendencia, lo cual, como sabemos, no suele ocurrir en la mayor parte de las series económicas.

En el caso de una media móvil simple (sin ponderación), los valores estimados estarán, por ejemplo, sistemáticamente por debajo de los reales cuando existe una tendencia creciente. Para la misma serie, pero utilizando alisado exponencial, el valor óptimo de  $\alpha$  será la unidad, ya que, aunque incorporando un sesgo sistemático, éste sería el más reducido posible. Calculando los valores de predicción históricos con  $\alpha = 1$ , resulta que

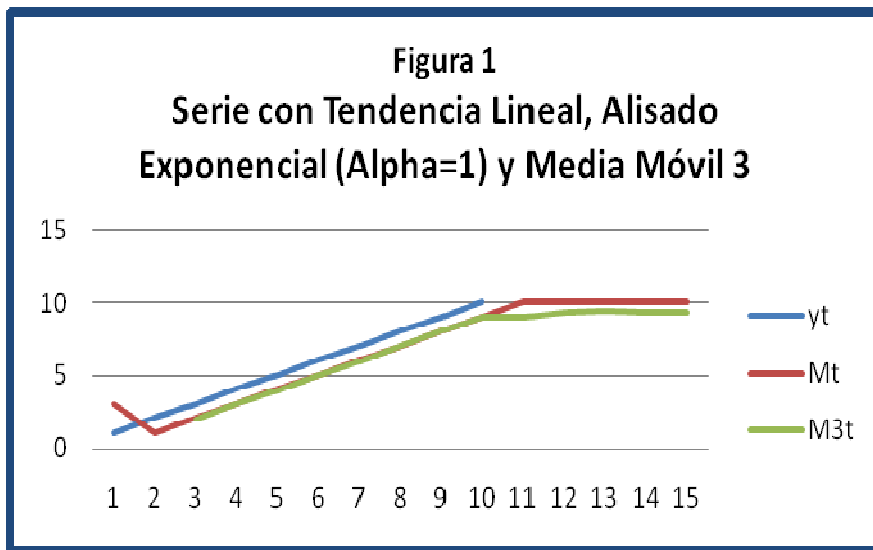
$$\hat{y}_t = y_{t-1}$$

con el correspondiente sesgo en una serie con tendencia. Así, para una serie tal como los diez primeros enteros, una media móvil de orden 3 y el mencionado alisado darían los siguientes resultados:

$y_t$	$M_t$	$M_t^3$
1	3	---
2	1	---
3	2	2
4	3	3
5	4	4
6	5	5
7	6	6
8	7	7
9	8	8
10	9	9
---	10	9
---	10	9,3
---	10	9,4
---	10	9,3
---	10	9,3

En la figura 1 puede verse la evolución real y de predicción (que se corresponde con los datos en cursiva de la tabla), con la característica estabilidad de las predicciones al nivel del último dato de la serie y con la posible perturbación inicial del valor seleccionado para el primer dato de alisado (en este caso  $\hat{y} = 3$ , media de los cinco primeros datos de la serie original).



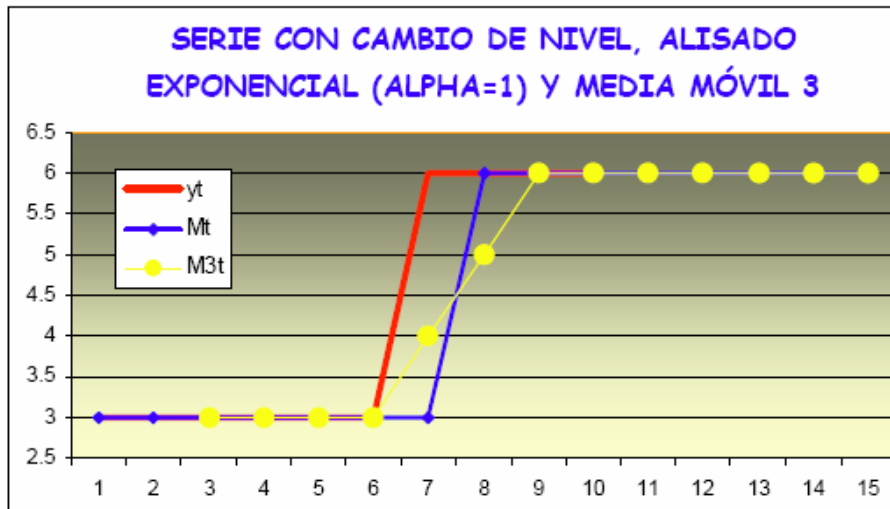


El problema es similar ante un cambio en el nivel de la serie. Como vemos en la siguiente tabla, con una media móvil simple, se centra el sesgo en los periodos inmediatos al cambio:

$y_t$	$M_t$	$M_t^3$
3	3	
3	3	
3	3	3
3	3	3
3	3	3
3	3	3
6	3	4
6	6	5
6	6	6
6	6	6
	6	6.0
	6	6.0
	6	6.0
	6	6.0
	6	6.0

Con un alisado exponencial (nuevamente, el valor óptimo será  $\alpha = 1$ , para adaptarse lo más rápidamente posible a la variación) el gráfico de la figura 2 muestra sesgo sólo en el periodo de cambio.

Figura 2

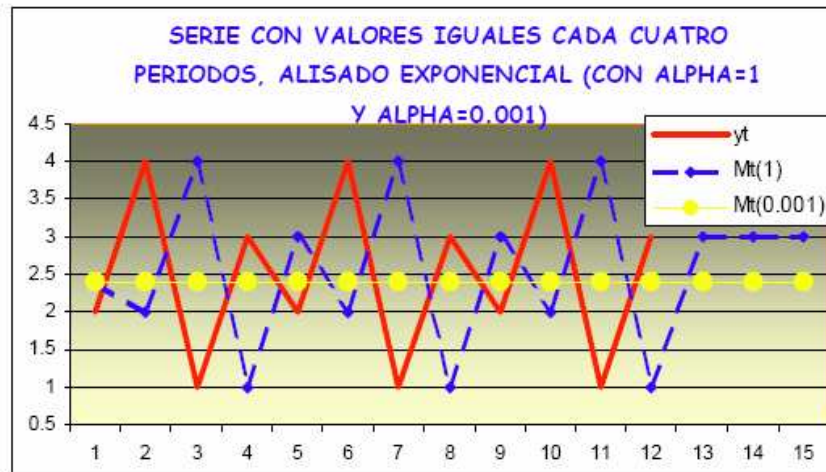


Para una serie con estacionalidad, el resultado es relativamente sorprendente. Valores de  $\alpha$  cercanos a la unidad, proporcionan una serie histórica estimada, con un comportamiento similar a la real, aunque desfasada un período. Sin embargo, el resultado óptimo, al presentar un ECM mínimo, puede dar un  $\alpha = 0$ , lo que supone un valor constante de predicción y unas diferencias predicción / realización menores que las que suponen correcciones sistemáticamente atrasadas. Esta situación es calificada por algunos especialistas de *sobrealisado*, al llegarse a un alisado máximo por una inadecuación del propio método para recoger estas oscilaciones sistemáticas.

En la figura 3.3 hemos representado una serie supuestamente trimestral con valores iguales cada cuatro periodos y sus predicciones de alisado para  $\alpha = 0,001$  y  $\alpha = 1$ .

$y_t$	$M_t^1$	$M_t^{0.001}$
2	2.4	2.4
4	2	2.4
1	4	2.4
3	1	2.4
2	3	2.4
4	2	2.4
1	4	2.4
3	1	2.4
2	3	2.4
4	2	2.4
1	4	2.4
3	1	2.4
	3	2.4
	3	2.4
	3	2.4

Figura 3



En resumen, la variante hasta aquí expuesta de alisado exponencial deberá aplicarse a una serie sin tendencia (o eliminada previamente ésta, por ejemplo, tomando diferencias de los valores iniciales); no recoge variaciones estacionales (aunque, posteriormente, puede incorporarse un esquema adicional para su tratamiento); además, sólo proporciona predicciones estables a más de un periodo, por lo que interesará revisar periódicamente las predicciones cada vez que se disponga de un nuevo dato.

### B.5. Alisados con tendencia

#### Resumen:

- ✚ Para las series que presentan tendencia, el desfase sistemático entre valores históricos y media móvil puede corregirse.
- ✚ Pero a efectos de predicción óptima, se necesitan alternativas como el alisado exponencial doble de Brown (también conocido como alisado doble de Brown o alisado exponencial lineal con parámetro único) o el alisado de Holt-Winters (conocido también como alisado exponencial lineal con doble parámetro), que eliminan el sesgo de la predicción de una serie con tendencia.

### **Tendencia y valores históricos**

Se puede demostrar que en una serie con tendencia los errores históricos (diferencia entre valores de la serie y su media móvil) se igualan a las diferencias entre la media móvil y la doble media móvil. Es decir, si definimos la media móvil de orden  $s$  como:

$$M_t^s = \frac{\sum_{t-s+1}^{t-1} y_t}{s}$$

y la doble media móvil como:

$$M_t^{s \cdot s} = \frac{\sum_{t-s+1}^{t-1} M_t^s}{s}$$

se tiene que se igualan los errores de ambos, es decir,

$$e_t = y_t - M_t^s = M_t^s - M_t^{s \cdot s} = e_t'$$

Por ello, para obtener una estimación sin sesgo sistemático de la variable  $y_t$  se utiliza:

$$M_t^s + e_t' = M_t^s + (M_t^s - M_t^{s \cdot s}) = 2M_t^s - M_t^{s \cdot s}$$

Ahora bien, esto tan sólo corrige el desfase sistemático existente en los valores históricos de la serie, y no el funcionamiento de cara a la predicción, ya que los valores predichos se generan sin tendencia. Por ello, se utilizan como alternativas la propuesta por Brown (Alisado exponencial doble) o por Holt-Winters.

### **El alisado exponencial doble de Brown**

Esta técnica (también conocida como alisado exponencial con parámetro único) es adecuada cuando la serie presenta una tendencia lineal. En ese caso se puede demostrar que se obtiene predicciones óptimas con

$$\hat{y}_{t+h} = a_t + b_t \cdot h$$

donde

$$a_t = M_t^s + (M_t^s - M_t^{s \cdot s}); \quad b_t = \frac{2}{s-1} (M_t^s - M_t^{s \cdot s})$$

Este planteamiento general fue adoptado por Brown en el caso del alisado exponencial. Definido el doble alisado exponencial como:

$$M_t^\alpha = \alpha y_t + (1-\alpha)M_{t-1}^\alpha$$

$$M_t^{\alpha \cdot \alpha} = \alpha M_t^\alpha + (1-\alpha)M_{t-1}^{\alpha \cdot \alpha}$$

la ecuación de predicción vendría dada por:

$$\hat{y}_{t+h} = a_t + b_t \cdot h$$

donde

$$a_t = M_t^\alpha + (M_t^\alpha - M_t^{\alpha \cdot \alpha}); \quad b_t = \frac{\alpha}{1-\alpha} (M_t^\alpha - M_t^{\alpha \cdot \alpha})$$

expresión que elimina el sesgo de la predicción de una serie con tendencia. Es habitual tomar como valor inicial de la media el primer valor de la variable

$M_1^\alpha = M_1^{\alpha \cdot \alpha} = y_1$ , o bien una media de los primeros valores, por ejemplo  $M_1^\alpha = M_1^{\alpha \cdot \alpha} = 1/3(y_1 + y_2 + y_3)$ . Un parámetro  $\alpha$  elevado es habitual en series de acusada tendencia. Lo más aconsejable es optar por el  $\alpha$  elegido por el programa, normalmente aquel que minimiza el error cuadrático medio.

### **La técnica Holt-Winters**

Se trata de una variante, también conocida como alisado exponencial lineal con doble parámetro, que también consigue eliminar el sesgo de la predicción de una serie con tendencia, a través de la inclusión en la media móvil de un componente de tendencia:

$$M_t = \alpha y_t + (1 - \alpha)(M_{t-1} + b_{t-1})$$

donde  $b$  es un factor de variación definido a partir de otra nueva constante de alisado para la tendencia,  $\beta$ :

$$b_t = \beta(M_t - M_{t-1} + (1 - \beta)b_{t-1})$$

de forma que la ecuación de predicción adoptaría la forma:

$$\hat{y}_{t+h} = M_t + b_t \cdot h$$

Existen varias posibilidades para la adopción de los valores iniciales. La más sencilla consistiría en hacer

$$M_1 = y_1; \quad b_1 = 0$$

Alternativamente, podemos tomar

$$b_1 = [(y_2 - y_1)/2] + [(y_4 - y_3)/2]$$

o también

$$M_2 = y_2; \quad b_2 = y_2 - y_1$$

Igualmente es aconsejable optar por los valores de los parámetros seleccionados por el programa, en general, elegidos como los que minimizan los errores cuadráticos medios. Como pauta general, los parámetros  $\alpha$  y  $\beta$  tomarán valores elevados (por encima de 0,7) en series de acusada tendencia.

Obviamente, en el caso particular en el que  $\beta = 0$ ;  $b_1 = 0$  el alisado con doble parámetro queda reducido al alisado exponencial simple.

## Ejercicio 2

En la tabla inferior hemos recogido una serie-índice con datos sobre costes laborales unitarios en la industria manufacturera en España, obtenidos de la antigua Dirección General de Previsión y Coyuntura del Ministerio de Economía y Hacienda, a la que denominamos COSLA.

- Introduzca en EViews esta serie. En este caso, para practicar todas las posibilidades introduzca la serie dato a dato (picado de datos).
- Analice las características de la serie, en especial la acusada tendencia.
- Aplique a la serie el alisado exponencial doble de Brown y Holt-Winters, con los parámetros óptimos. Compruebe con cuál se obtiene un menor error cuadrático medio. Compare la predicción de ambos.
- Aplique el alisado exponencial simple, y compare los resultados con los obtenidos por las dos versiones del alisado doble.
- Aplique el alisado exponencial doble con cualquier otro valor de alpha.

	COSLA
1	100,0000
2	110,5739
3	118,1201
4	126,3980
5	131,8828
6	137,6763
7	144,2019
8	149,4524
9	159,3335
10	176,4203
11	189,4843
12	205,1086
13	215,6812
14	210,4928
15	208,6748
16	219,5314
17	222,6703

## SOLUCIÓN

### A) INTRODUCCIÓN DE DATOS.

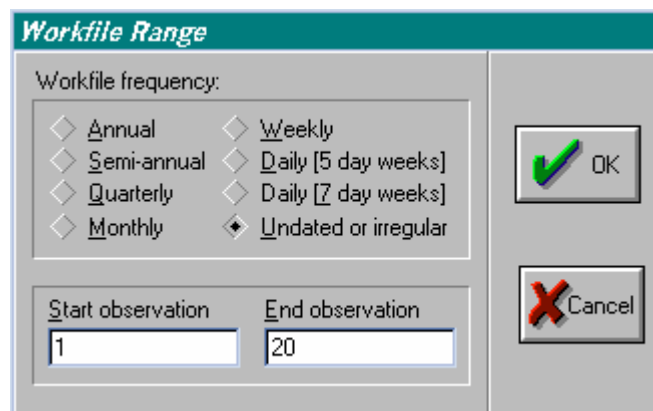
Contamos con una serie-índice de costes laborales unitarios en la industria manufacturera en España obtenidos de la antigua Dirección General de Previsión y Coyuntura del Ministerio de Economía y Hacienda. Se trata de una serie sin estacionalidad, pero con acusada tendencia.

Para preparar la introducción de datos, lo primero será crear un fichero de trabajo. Optamos por definirlo para datos sin referencia temporal específica de fecha, es decir, ordenados de 1 en adelante y para un máximo de 20 (17

observaciones muestrales y 3 datos de predicción, correspondiendo, realmente, el primer dato al año 1981 y el último a 1997).

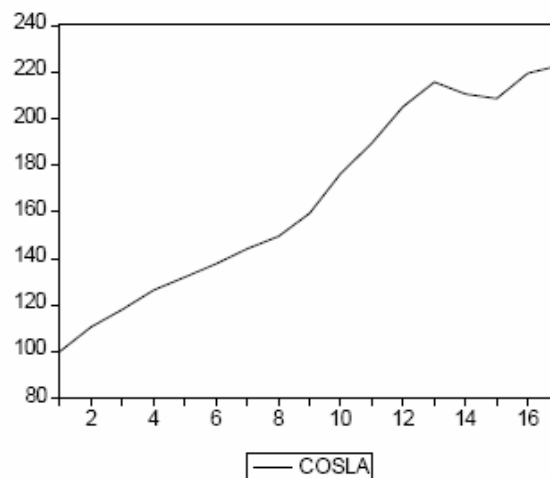
Para ello, en EViews, seleccionamos en el menú principal: FILE / NEW / WORKFILE / UNDATED OR IRREGULAR, y aquí indicamos que trabajaremos desde 1 a 20 datos.

A continuación, con la instrucción OBJECTS del menú principal, seleccionamos NEW OBJECT / SERIES e introducimos un nombre para nuestra serie (COSLA). Una vez creada esta serie, aparecerá como tal en la ventana del workfile, donde seleccionamos SHOW y, dentro de aquí, EDIT y procedemos a teclear los datos de la serie COSLA.



A continuación, con la instrucción OBJECTS del menú principal, seleccionamos NEW OBJECT / SERIES e introducimos un nombre para nuestra serie (COSLA). Una vez creada esta serie, aparecerá como tal en la ventana del *workfile*, donde seleccionamos SHOW y, dentro de aquí, EDIT y procedemos a teclear los datos de la serie COSLA:

1	100.0000
2	110.5739
3	118.1201
4	126.3980
5	131.8828
6	137.6763
7	144.2019
8	149.4524
9	159.3335
10	176.4203
11	189.4843
12	205.1086
13	215.6812
14	210.4928
15	208.6748
16	219.5314
17	222.6703



## B) CARACTERÍSTICAS DE LA SERIE

A continuación, seleccionamos el periodo muestral (SMPL) correspondiente a los datos históricos, 1 a 17 datos, comprobamos con SHOW los datos que hemos introducido, y dentro de ésta ventana, con la opción VIEW / LINE GRAPH, visualizamos un gráfico de la serie, donde comprobamos la acusada tendencia que presenta.

## C) ALISADO EXPONENCIAL DOBLE DE BROWN Y HOLT-WINTERS.

Procedemos ya a la estimación del alisado exponencial doble de Brown y de Holt-Winters, recogiendo los resultados, incluido predicciones de las nuevas series generadas en el proceso y que hemos denominado, respectivamente, PCOSLAD y PCOSLAN.

**Exponential Smoothing**

Smoothing Method:    # Parameters    Smoothed Series:

Single                    1                    PCOSLAD

Double                    1                    Series name for smoothed and forecasted values.

Holt-Winters - No seasonal                    3

Holt-Winters - Additive                    3

Holt-Winters - Multiplicative                    3

Smoothing Parameters:

Alpha: (mean)    E                    Enter number between 0 and 1, or E to estimate.

Beta: (trend)    E

Gamma: (seasonal)    E

Estimation Sample:    1 17

Forecasts begin in period following estimation endpoint.

Cycle for Seasonal:    5

OK    Cancel

Series: COSLA    Workfile: COSLA

View   Proc   Objects   Print   Name   Freeze   Edit+/-   Smp+/-   Label+/-   Wide+/-   InsDel   Title   Sa

Date: 03/26/99    Time: 03:32

Sample: 1 17

Included observations: 17

Method: Double Exponential

Original Series: COSLA

Forecast Series: PCOSLAD

---

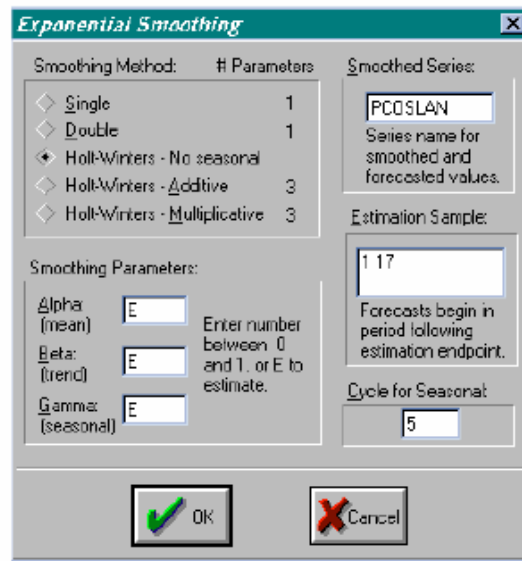
Parameters: Alpha	0.8240
Sum of Squared Residuals	638.1644
Root Mean Squared Error	5.126915

---

End of Period Levels:	Mean	222.7728
	Trend	4.581998

---

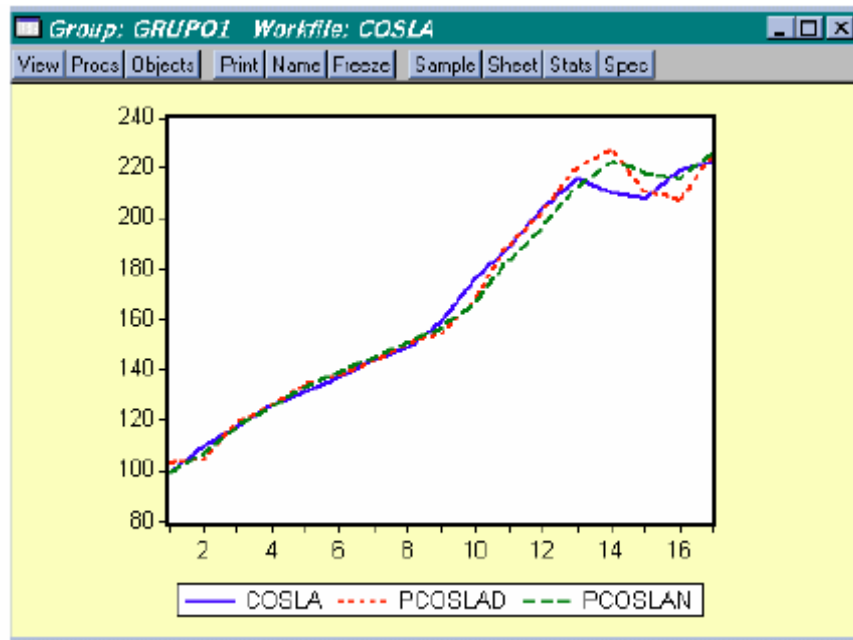




Series: COSLA Workfile: COSLA		
View Procs Objects Print Name Freeze Edit+/- Smp+/- Label+/- Wider+ InsDel Title Ser		
Date: 03/26/99 Time: 03:34		
Sample: 1 17		
Included observations: 17		
Method: Holt-Winters No Seasonal		
Original Series: COSLA		
Forecast Series: PCOSLAN		
<hr/>		
Parameters:	Alpha	1.0000
	Beta	0.0000
	Sum of Squared Residuals	505.6794
	Root Mean Squared Error	5.453975
<hr/>		
End of Period Levels:	Mean	222.6703
	Trend	7.416688
<hr/>		

Tanto con doble alisado de Brown (PCOSLAD) como con Holt-Winters (PCOSLAN), los parámetros que minimizan el error cuadrático medio (ECM) toman un valor alto, incluso unitario en el segundo caso. La suma de cuadrados de los residuos y la raíz cuadrada del ECM dan valores más reducidos en el alisado de Holt-Winters. Se nos indica además, que el nivel de la serie en su último dato (periodo 17, que, en realidad, corresponde al año 1997) es de 222, con un incremento respecto al dato anterior (tendencia en ese punto) de unos 6 puntos.

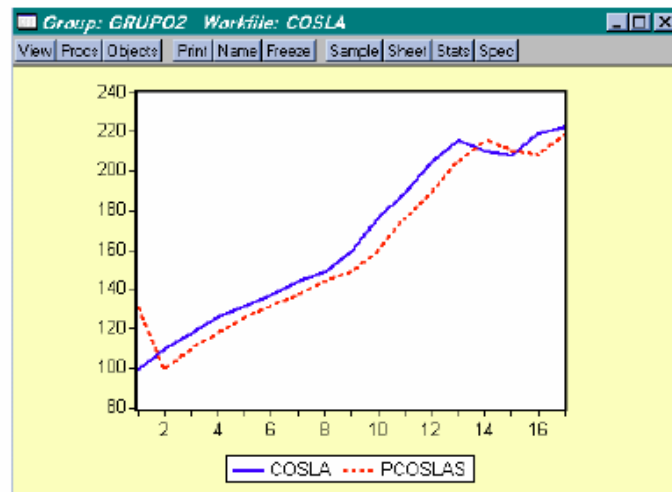
Los valores de predicción se ajustarían muy estrechamente a los datos en el periodo histórico y mantendrían una tendencia similar en ambas técnicas a efectos de predicción. Podemos representar gráficamente la serie original (COSLA) y las dos series obtenidas en cada alisado (PCOSLAD y PCOSLAN), abarcando tanto el periodo histórico como el de predicción.



#### D) ALISADO EXPONENCIAL SIMPLE.

Naturalmente, los resultados son muy distintos a los que se obtendrían con un alisado exponencial simple, donde comprobamos cómo el error de estimación ha aumentado considerablemente.

Series: COSLA Workfile: COSLA												
View	Proc	Object	Print	Name	Freeze	Sample	Genr	Sheet	Stats	Ident	Line	Bar
Date: 03/26/99 Time: 03:41												
Sample: 1 17												
Included observations: 17												
Method: Single Exponential												
Original Series: COSLA												
Forecast Series: PCOSLAS												
<hr/>												
Parameters: Alpha 0.9590												
Sum of Squared Residuals 2398.323												
Root Mean Squared Error 11.87762												
<hr/>												
End of Period Levels: Mean 222.6672												
<hr/>												



### E) ALISADO EXPONENCIAL DOBLE CON VALORES NO ÓPTIMOS.

Por último, podemos probar con valores prefijados (no óptimos) de los parámetros, a fin de analizar la incidencia de estos cambios. Aplicamos este análisis al alisado exponencial doble, con valores de alpha de 0,1, 0,5 y 0,9, generando las variables PCOSLAD1, PCOSLAD5 y PCOSLAD9.

Series: COSLA Workfile: COSLA												
View	Proc	Objects	Print	Name	Freeze	Sample	Genr	Sheet	Stats	Ident	Line	Bar
Date: 03/26/99 Time: 03:46												
Sample: 1 17												
Included observations: 17												
Method: Double Exponential												
Original Series: COSLA												
Forecast Series: PCOSLAD1												
-----												
Parameters: Alpha 0.1000												
Sum of Squared Residuals 1270.822												
Root Mean Squared Error 8.645052												
-----												
End of Period Levels: Mean 227.7626												
Trend 7.485984												

Series: COSLA Workfile: COSLA												
View	Proc	Objects	Print	Name	Freeze	Sample	Genr	Sheet	Stats	Ident	Line	Bar
Date: 03/26/99 Time: 03:48												
Sample: 1 17												
Included observations: 17												
Method: Double Exponential												
Original Series: COSLA												
Forecast Series: PCOSLAD5												
-----												
Parameters: Alpha 0.5000												
Sum of Squared Residuals 773.6535												
Root Mean Squared Error 6.746038												
-----												
End of Period Levels: Mean 223.0064												
Trend 4.842193												

Series: COSLA Workfile: COSLA												
View	Proc	Objects	Print	Name	Freeze	Sample	Gen	Sheet	Stats	Ident	Line	Bar
Date: 03/25/99 Time: 03:51												
Sample: 1 17												
Included observations: 17												
Method: Double Exponential												
Original Series: COSLA												
Forecast Series: PCOSLAD9												
-----												
Parameters: Alpha 0.9000												
Sum of Squared Residuals 645.2200												
Root Mean Squared Error 6.160691												
-----												
End of Period Levels: Mean 222.7218												
Trend 4.245406												
-----												

## RESULTADOS

En la tabla siguiente comparamos los resultados obtenidos en cada alisado exponencial, en la que también incluimos el alisado exponencial de Holt-Winters, aparte de los datos originales y las predicciones supuestamente óptimas, en la que se aprecia cómo el error de estimación (ECM) aumenta con las series alisadas mediante un parámetro no óptimo.

	COSLA	PCOSLAD	PCOSLAD1	PCOSLAD5	PCOSLAD9	PCOSLAN
1	100,0000	103,0213	103,0213	103,0213	103,0213	100,0000
2	110,5739	104,9990	109,3739	106,9569	104,5398	107,4167
3	118,1201	119,0919	116,5406	116,7754	119,9108	117,9906
4	126,3980	126,1811	123,7951	125,2259	126,0848	125,5368
5	131,8828	134,5694	131,2701	133,8399	134,5953	133,8147
6	137,6763	138,3199	138,3731	139,6177	137,9132	139,2994
7	144,2019	143,6131	145,2203	144,9219	143,4900	145,0929
8	149,4524	150,5003	151,9963	150,9622	150,5828	151,6186
9	159,3335	155,0901	158,4569	156,0327	154,9362	156,8691
10	176,4203	167,6885	165,5763	165,5364	168,3238	166,7502
11	189,4843	190,5649	174,6978	183,4483	191,9317	183,8369
12	205,1086	203,1992	184,7163	199,2333	203,1188	196,9010
13	215,6812	220,0273	196,0038	216,3666	220,3104	212,5252
14	210,4928	227,8427	207,3523	228,4080	227,1995	223,0979
15	208,6748	211,2768	215,5902	223,0483	208,5994	217,9095
16	219,5314	207,2353	221,8483	216,7515	206,6747	216,0915
17	222,6703	225,9794	228,9569	224,0148	227,8175	226,9481
18	NA	227,3548	235,2484	227,8486	226,9672	230,0870
19	NA	231,9368	242,7344	232,6908	231,2126	237,5037
20	NA	236,5188	250,2204	237,5330	235,4580	244,9204
<b>ECM</b>		<b>638,1644</b>	<b>127,0822</b>	<b>773,6535</b>	<b>645,2200</b>	<b>505,6794</b>

GRAFICO2: Comparativa entre la serie original (COSLA), serie de alisado doble (PCOSLAD) y serie de alisado de Holt-Winters (PCOSLAN)

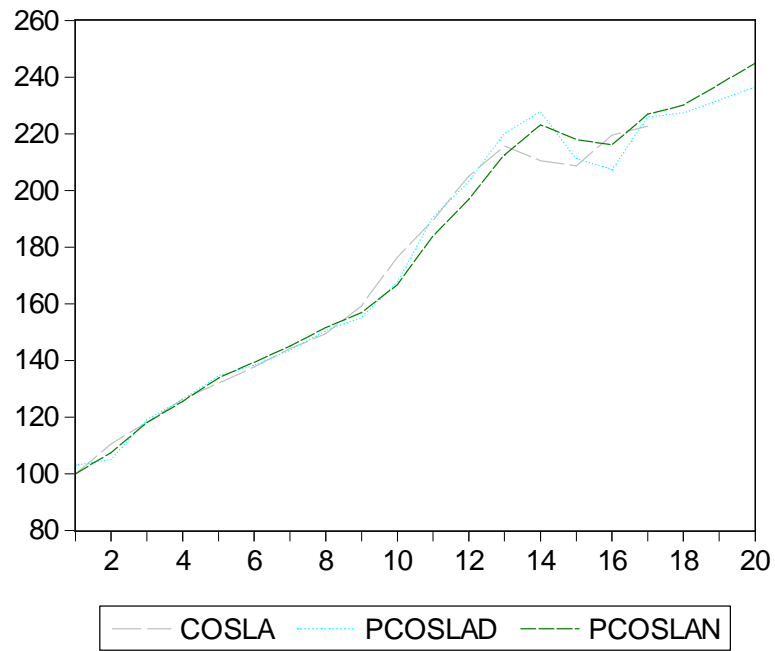
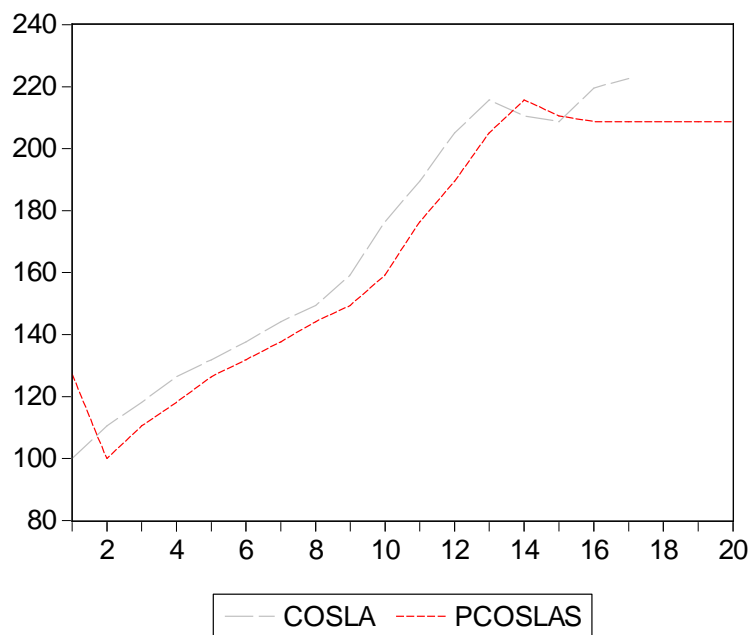


GRAFICO3: Comparativa entre la serie original (COSLA) y serie de alisado simple (PCOSLAS)



## B.6. Ajuste con funciones matemáticas

### Resumen:

- ✚ Este procedimiento nos permite tanto calcular el componente tendencia de una serie de forma alternativa, como su aprovechamiento a efectos de predicción.
- ✚ El procedimiento es útil en series con tendencia, pero sin estacionalidad.
- ✚ Consiste en ajustar a los datos una función matemática del tiempo. Son varias las formulaciones matemáticas posibles: recta, parábola, exponenciales ...
- ✚ Estos ajustes son similares a la regresión econométrica habitual (ver la unidad correspondiente), sólo que ahora las variables explicativas son el momento de tiempo, o funciones del tiempo.
- ✚ En general el cálculo de estos ajustes es sencillo. La dificultad estriba en la elección de la fórmula funcional más adecuada para describir el fenómeno.
- ✚ Una vez obtenidos los coeficientes de la formulación elegida, la predicción se obtiene por la simple sustitución de  $t$  por  $n + 1$ ,  $n + 2, \dots$ ,  $n + h$ .

### **Alternativas en el cálculo de la tendencia**

Recuérdese que inicialmente consideramos las técnicas de las medias móviles o de alisados como aptas para obtener el componente tendencia de una serie económica. La principal característica distintiva de algunos de los más utilizados ajustes de tendencia es que en aquellos casos obtenemos una tendencia variable en el tiempo, eso sí, tanto más estable cuanto mayor sea el número de términos incluidos en la media móvil, o cuanto mayor sea la memoria del proceso de alisado. Como alternativa a estos métodos podemos utilizar los ajustes de tendencia mediante funciones matemáticas, más o menos complejas, del tiempo.

#### *Fórmulas alternativas de ajuste de tendencia, transformación lineal e instrucciones para su estimación en EViews*

Fórmula de ajuste	Transformación lineal	Instrucciones EViews
<i>Recta</i> $y_t = a + bt$	--	LS Y C T
<i>Parábola de 2º grado</i> $y_t = a + bt + ct^2$	$y_t = a + bt + ct_2$ donde $t_2 = t^2$	GENR T2=T^2 LS Y C T T2
<i>Exponencial</i> $y_t = ab^t$	$\text{Log } y_t = \log a + t \log b = c + dt$	GENR LY=LOG(Y) LS LY C T
<i>Potencial</i> $y_t = at^b$	$\text{Log } y_t = \log a + b \log t = c + b \log t$	GENR LY=LOG(Y) GENR LT=LOG(T) LS LY C LT
<i>Exponencial modificada simple</i> $y_t = a + br^t ; (r < 1)$	--	NLS Y=c(1) + c(2)*c(3)^T
<i>Logística</i> $y_t = 1/(a + br^t) ; (r < 1)$	--	NLS Y=1/(c(1)+c(2)*c(3)^T)
<i>Gompertz</i> $\text{Log } y_t = a + br^t ; (r < 1)$	--	GENR LY=LOG(Y) NLS LY=c(1)+c(2)*c(3)^T

### Ejercicio 3: Cálculo de tendencia mediante funciones matemáticas en EViews

En la parte inferior se recoge una serie, llamada DEUDA, con datos anuales de la deuda pública viva de las AA.PP., para el período 1981-1998, obtenido del Banco de Datos CEPREDE. Se pide:

- Introduzca los datos, y ajuste los datos a una tendencia a través de una recta. Realice la predicción para el año 1999.
- Ajuste los datos a una parábola de segundo grado, y prediga el valor de la deuda viva para 1999.

Periodo	DEUDA
1981	4113,30
1982	6015,60
1983	8892,50
1984	11759,90
1985	14491,20
1986	16759,40
1987	18519,30
1988	19069,00
1989	22029,00
1990	25483,00
1991	28459,00
1992	32451,00
1993	40600,00
1994	45069,00
1995	50957,00
1996	56903,00
1997	61101,62
1998	64373,39

### Solución

#### A) INTRODUCCIÓN DE DATOS Y AJUSTE A UNA RECTA.

En primer lugar, y como siempre a la hora de trabajar en EViews, definimos un Workfile con datos de serie temporal y de frecuencia anual, de 1981 a 1998. Hecho esto, procedemos a introducir los datos, bien sea con la función de importación de Excel, o con el "picado" de datos. Si optamos por esta última posibilidad, seleccionaremos en la ventana general de instrucciones del programa OBJECTS / NEW OBJETC /SERIES.

Procederemos a darle nombre a la serie, por ejemplo DEUDA. Creada la serie, activaremos la introducción de datos con la instrucción EDIT, y ya podemos "picar" cada uno de los datos. Se remite al lector interesado a anteriores ejercicios para una mayor descripción.

Una vez introducidos los datos de nuestra variable, podemos ya realizar el ajuste pedido a una recta. Como es conocido, la expresión matemática de la recta adopta la forma:

$$y = a + b \cdot x$$

En nuestro caso, la variable dependiente  $y$ , sería la DEUDA, mientras que la variable independiente  $x$ , sería el momento de tiempo  $t$ , es decir:

$$DEUDA = a + b \cdot t$$

Conocidos los valores de la variable *DEUDA*, y otorgados valores a la variable *tiempo*, podemos estimar por métodos estadísticos sencillos los parámetros  $a$  y  $b$ , de forma que las estimaciones obtenidas cumplan ciertas propiedades adecuadas. Entraremos con más detalle en todo ello, en el tema correspondiente.

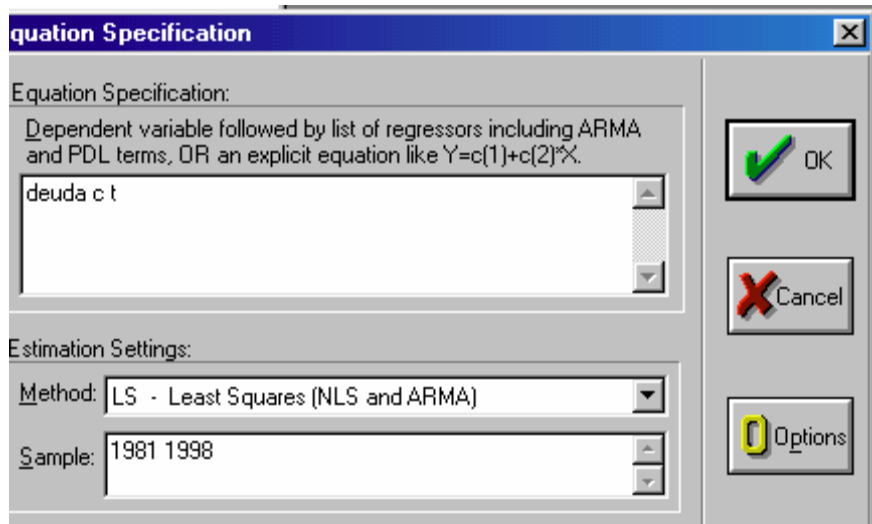
En todo caso, vamos a proceder a realizar una regresión entre nuestra variable *DEUDA*, un término independiente  $a$  (que el programa EViews identifica con la letra  $c$ ) y una variable que refleja el momento del tiempo. En primer lugar, debemos dotar de contenido cuantitativo a la variable tiempo. Para ello, creamos un nuevo objeto SERIE, en forma análoga a la serie *DEUDA*, al que denominaremos T. En este caso, y para recoger estos efectos temporales, la práctica habitual nos llevará a generar valores empezando con el valor 1 para 1981, 2 para 1982, y así sucesivamente. Se puede demostrar que los resultados que se obtienen son independientes de los valores dados a esta variable.

The screenshot shows the EViews software interface for a series object named 'T' in a workfile titled 'TENDENCIA'. The data is displayed in a table with columns for year and value. The values for years 1981 through 1998 are as follows:

Year	Value
1981	1.000000
1982	2.000000
1983	3.000000
1984	4
1985	NA
1986	NA
1987	NA
1988	NA
1989	NA
1990	NA
1991	NA
1992	NA
1993	NA
1994	NA
1995	NA
1996	NA
1997	NA
1998	NA



Contamos ya con toda la información para la estimación de la tendencia a través de una recta. La regresión en EViews puede llevarse a cabo de varias formas. Por ejemplo, en el menú principal del programa podemos seleccionar QUICK / ESTIMATE EQUATION, apareciendo una ventana como la siguiente:



En esta ventana distinguimos 3 áreas: en la inferior nos indica el período muestral que utilizará para realizar la regresión. En el área media se señala el método de estimación, en nuestro caso utilizaremos el método de los Mínimos Cuadrados Ordinarios (LS-Least Squares, en inglés). En el área superior indicamos las variables sobre las que tiene que realizar la regresión. Escribiremos en primer lugar el nombre de la variable dependiente, en nuestro caso *DEUDA*, seguida de la lista de variables independientes separadas por espacios en blanco. Si además queremos que la ecuación incluya término independiente (en el caso de una recta, abscisa en el origen), incluiremos como una variable más la letra *c*, procedimiento que EViews interpreta para la inclusión del intercepto. En definitiva, en nuestro caso escribiremos en el espacio señalado la siguiente expresión:

DEUDA C T

tal y como se muestra en la ventana anterior.

Pulsando OK, EViews nos muestra la *salida de regresión*.

Equation: RECTA Workfile: TENDENCIA				
View Procs Objects Print Name Freeze Estimate Forecast Stats Resids				
Dependent Variable: DEUDA				
Method: Least Squares				
Date: 08/10/03 Time: 20:48				
Sample: 1981 1998				
Included observations: 18				
Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	-4537.138	2077.394	-2.184052	0.0442
T	3559.735	191.9185	18.54816	0.0000
R-squared	0.955560	Mean dependent var	29280.34	
Adjusted R-squared	0.952782	S.D. dependent var	19440.67	
S.E. of regression	4224.388	Akaike info criterion	19.63959	
Sum squared resid	2.86E+08	Schwarz criterion	19.73851	
Log likelihood	-174.7562	F-statistic	344.0342	
Durbin-Watson stat	0.184709	Prob(F-statistic)	0.000000	

Entraremos con más detalle en el tema correspondiente sobre la interpretación de esta salida del programa, pero de momento es suficiente con que nos quedemos con los *coeficientes*. De ellos podemos concluir que la recta estimada es:

$$DEUDA = -4537,138 + 3559,735 \cdot t$$

Con ello sólo nos resta realizar la predicción. Si queremos predecir el valor de la deuda pública viva de las AA.PP. para el año 1999, tan sólo tendríamos que dar valor a la variable  $t$  (19 en nuestro caso) y sustituir valores:

$$DEUDA_{1999} = -4537,138 + 3559,735 \cdot 19 = 63097,827$$

## B) AJUSTE A UNA PARÁBOLA DE SEGUNDO GRADO

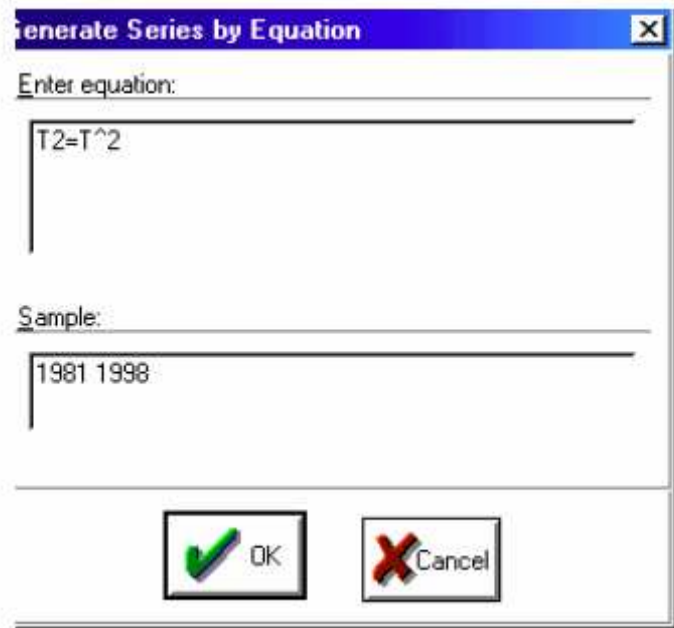
La parábola de segundo grado en función del tiempo se formularía, para nuestra variable DEUDA como

$$DEUDA_t = a + bt + ct^2$$

Para la estimación bastará en primer lugar con crear una variable, que denominaremos T2, que sea el cuadrado de la variable T, tal y como nos indica la expresión formal de la parábola. Esta variable puede ser generada directamente en EViews. Para ello seleccionamos dentro de la ventana general del programa el comando PROCS / GENERATE SERIES.

En la ventana que nos aparece deberemos anotar la formulación propia de EViews. En nuestro caso, la fórmula para introducir el cuadrado de la variable es, tal y como se muestra en la ventana anterior, la siguiente:

$$T2 = T^2$$



Una vez creada la variable T2, ya podemos pasar a la estimación en la misma forma que hicimos con el ajuste mediante recta. En este caso, la formulación a introducir será:

DEUDA C T T2

Lo que nos generará una nueva estimación de la tendencia, en la cual:

$$DEUDA_t = 5023,84 + 691,43 \cdot t + 150,96 \cdot t^2$$

Para la predicción en 1999 sustituiríamos  $t$  por 19, y  $t^2$  por 361(19<sup>2</sup>), obteniendo como predicción

$$DEUDA_{1999} = 5023,84 + 691,43 \cdot 19 + 150,96 \cdot 361 = 72657,57$$

### B.7. Predicción en series con componente estacional

#### Resumen:

- La estacionalidad en una serie presenta peculiaridades tales que exige un tratamiento singular de cara a la predicción.
- Los procedimientos de predicción de una serie con estacionalidad pasan por aislar tal componente. Por ello, las técnicas de predicción que veremos son una continuación del proceso previo de desestacionalización.
- Para la obtención del componente de estacionalidad se procede a separar el componente tendencial, y a continuación se aísla el componente estacional.
- Entre los procedimientos disponibles para la obtención de la estacionalidad (y de la posterior predicción) distinguimos dos: la técnica de relación a la media móvil (o Census X-11) y la de Holt-Winters.

- Para la predicción, se procede aplicando los coeficientes estacionales, calculados de una u otra forma, sobre la tendencia determinada (en el caso de X-11), o multiplicándolos por la fórmula inicial de Holt-Winters.

### **Obtención de la tendencia**

El primer paso para la descomposición de la serie en tendencia y estacionalidad consiste en calcular la tendencia de la serie original, separando el movimiento regular a largo plazo del conjunto de oscilaciones de la serie. Para ello se dispone de varias alternativas, por ejemplo:

- Calculando las diferencias entre valores de la serie original (diferenciación sucesiva).
- Ajuste de funciones matemáticas.
- Media móvil simple, ponderada, o alisado exponencial.

De entre estas alternativas nos resta por analizar con detalle la primera. Consiste en obtener una nueva serie, llamémosla  $z_t$ , (que ya no presente componente estacional) a partir de la serie original  $y_t$ , en la que para cada momento del tiempo  $t$  la serie  $z_t$  toma el siguiente valor:

$$z_t = y_t - y_{t-1} = \nabla y_t$$

Si al representar gráficamente esta serie,  $z_t$ , comprobamos que oscila alrededor del mismo valor, entonces tendríamos que se trata de la serie original sin tendencia. Si  $z_t$  crece o decrece a largo plazo, entonces aún presenta tendencia y habría que volver a diferenciar, es decir, proseguiríamos calculando ahora:

$$\nabla z_t = z_t - z_{t-1} = (y_t - y_{t-1}) - (y_{t-1} - y_{t-2}) = \nabla^2 y_t$$

y, en caso necesario, seguiríamos diferenciando, teniendo en cuenta que en cada diferenciación perdemos una observación. En cualquier caso, lo habitual es que con una primera diferencia la serie ya fluctúe en torno al mismo valor, es decir, ya no presente tendencia y, como mucho, se realizaría una segunda diferencia.

Un ejemplo sirve para mostrar la sencillez del procedimiento. Tal y como vimos en el ejercicio 4, la serie deuda presentaba claramente tendencia. Calcularíamos las primeras diferencias de la serie la que llamaremos Ddeuda, restando al valor de cada año, el valor tomado en el año anterior. Si la serie en primeras diferencias aún presentase tendencia, procederíamos a calcular la segunda diferencia de la serie deuda, para lo cual procederíamos a calcular la primera diferencia de la serie Ddeuda, ya diferenciada una vez. Todo ello se muestra en la siguiente tabla:

Una vez eliminada la tendencia de la serie original, se trata ahora de distinguir, aislando, el patrón de oscilación estacional más común. Para ello, se destacan dos métodos: Relación a la media móvil o CensusX-11 y Holt-Winters.

Serie deuda, en primeras y segundas diferencias			
Año	deuda	Ddeuda	DDdeuda
1981	4113,30	--	--
<b>1982</b>	<b>6015,60</b>	<b>1902,30</b>	--
<b>1983</b>	<b>8892,50</b>	<b>2876,90</b>	<b>974,60</b>
<b>1984</b>	<b>11759,90</b>	<b>2867,40</b>	<b>-9,50</b>
<b>1985</b>	<b>14491,20</b>	<b>2731,30</b>	<b>-136,10</b>
<b>1986</b>	<b>16759,40</b>	<b>2268,20</b>	<b>-463,10</b>
<b>1987</b>	<b>18519,30</b>	<b>1759,90</b>	<b>-508,30</b>
<b>1988</b>	<b>19069,00</b>	<b>549,70</b>	<b>-1210,20</b>
<b>1989</b>	<b>22029,00</b>	<b>2960,00</b>	<b>2410,30</b>
<b>1990</b>	<b>25483,00</b>	<b>3454,00</b>	<b>494,00</b>
<b>1991</b>	<b>28459,00</b>	<b>2976,00</b>	<b>-478,00</b>
<b>1992</b>	<b>32451,00</b>	<b>3992,00</b>	<b>1016,00</b>
<b>1993</b>	<b>40600,00</b>	<b>8149,00</b>	<b>4157,00</b>
<b>1994</b>	<b>45069,00</b>	<b>4469,00</b>	<b>-3680,00</b>
<b>1995</b>	<b>50957,00</b>	<b>5888,00</b>	<b>1419,00</b>
<b>1996</b>	<b>56903,00</b>	<b>5946,00</b>	<b>58,00</b>
<b>1997</b>	<b>61101,62</b>	<b>4198,62</b>	<b>-1747,38</b>
<b>1998</b>	<b>64373,39</b>	<b>3271,77</b>	<b>-926,85</b>

### ***Estacionalidad y predicción con la técnica de relación a la media móvil (Census X-11)***

La técnica de *relación a la media móvil*, y la versión más refinada conocida como *Census X-11*, es uno de los procedimientos de más difusión ya que, a su propio planteamiento se une el desarrollo de un software adecuado de gran extensión, y que todavía goza de una amplia utilización.

El esquema original de la técnica puede resumirse en tres fases:

1) Obtención del componente tendencia mediante una media móvil. Por ejemplo, para datos mensuales se sugiere una media móvil de 12 términos, como valor de esa tendencia para la media de ese período móvil:

$$M_t^{12} = \frac{1}{12}(y_{t-6} + y_{t-5} + \dots + y_t + y_{t+1} + \dots + y_{t+5})$$

Puesto que se trata de un número par de términos, se promedia con la media siguiente:

$$\bar{M}_t^{12} = \frac{1}{2}(M_t^{12} + M_{t+1}^{12}) = \frac{1}{12}\left(\frac{1}{2}y_{t-6} + y_{t-5} + \dots + y_t + y_{t+1} + \dots + y_{t+5} + \frac{1}{2}y_{t+6}\right)$$

2) Establecimiento de los índices o porcentajes de los datos originales respecto a las medias móviles centradas correspondientes:

$$(y_T / \bar{M}_t^{12}) * 100$$

3) Cálculo de la media para cada mes, de estos porcentajes (de todos los eneros, de todos los febreros...) que se tomarán como coeficiente de estacionalidad del esquema multiplicativo:

$$s_i = \frac{1}{n} [(y_{1i} / \bar{M}_{1i}^{12}) * 100 + (y_{2i} / \bar{M}_{2i}^{12}) * 100 + \dots + (y_{ni} / \bar{M}_{ni}^{12}) * 100], \quad i = 1, \dots, 12$$

donde el primer subíndice hace referencia al año (n en total) y el segundo a los meses de referencia, bien entendido que debe cumplirse que:

$$\sum_{i=1}^{12} s_i = 1200$$

Entre los perfeccionamientos de la técnica propuestos por Shiskin en el *Bureau of the Census* de los EE.UU. en la década de 1950, está el de considerar como una primera estimación estos coeficientes de estacionalidad anteriormente calculados. **En este caso se añaden aún tres nuevas etapas.**

4) Cálculo de los valores desestacionalizados de la serie original:

$$y_{it}^s = \frac{y_{it}}{s_i} * 100$$

5) Establecimiento de una nueva media móvil sobre estos datos desestacionalizados. Se suele proponer una media simple de orden 5 o ponderada de Spencer de orden 15.

6) Con estos nuevos datos se repite el cálculo de índices para cada período y, finalmente, se determinan los coeficientes estacionales definitivos para cada mes.

En el *software* informático X-11 se añaden otras posibilidades de corrección de los datos, como por número de días laborables, por festividades, huelgas, etc.

En cualquier caso, obtenemos unos coeficientes estacionales que son constantes en el tiempo para cada mes. El mismo procedimiento nos sirve, obviamente, para obtener, junto a los coeficientes de estacionalidad, la propia serie desestacionalizada.

De cara a la predicción usaremos la serie desestacionalizada (pero con tendencia) a la que aplicaremos cualquiera de los métodos adecuados (alisado exponencial doble, Holt-Winters con dos parámetros,...). La predicción así obtenida debe ser afectada por los índices de estacionalidad

obtenidos por relación a la media móvil o X-11, obteniendo la predicción final.

### **Estacionalidad y predicción con Holt-Winters**

La técnica de Holt-Winters en su variante con tratamiento de la estacionalidad, también conocida como alisado exponencial con triple parámetro, calcula igualmente un índice de estacionalidad para cada dato por relación a la media móvil ( $y_t / M_t$ ), pero no se intenta establecer un coeficiente medio significativo, sino un procedimiento de actualización de esa estacionalidad estimada. Para ello, y con datos mensuales, la formulación utilizada es:

$$s_t = \delta(y_t / M_t) + (1 - \delta) \cdot s_{t-12}$$

expresión que se añade a las dos anteriormente utilizadas en la variante de Holt-Winters para series sólo con tendencia, es decir:

$$M_t = \alpha(y_t / s_{t-12}) + (1 - \alpha) \cdot (M_{t-1} + b_{t-1})$$

$$b_t = \beta(M_t - M_{t-1}) + (1 - \beta) \cdot b_{t-1}$$

con el único cambio de sustituir en el primer sumando del alisado  $y_t$  por  $y_t / s_{t-12}$ . Obviamente, a efectos de predicción, se multiplica la fórmula inicial de Holt-Winters por el factor de estacionalidad.

$$\hat{y}_{t+h} = (M_t + b_t h) \cdot s_{t-12+h}$$

### **Ejercicio 4: Análisis y descomposición de una serie temporal en EViews**

Disponemos de una serie temporal con datos sobre ventas del comercio al por menor (IVENTAS), indicadas en el año 1994, para el período comprendido entre enero de 1995 y diciembre de 1998, obtenidos del INE. A partir de los datos recogidos en la tabla inferior, se pide:

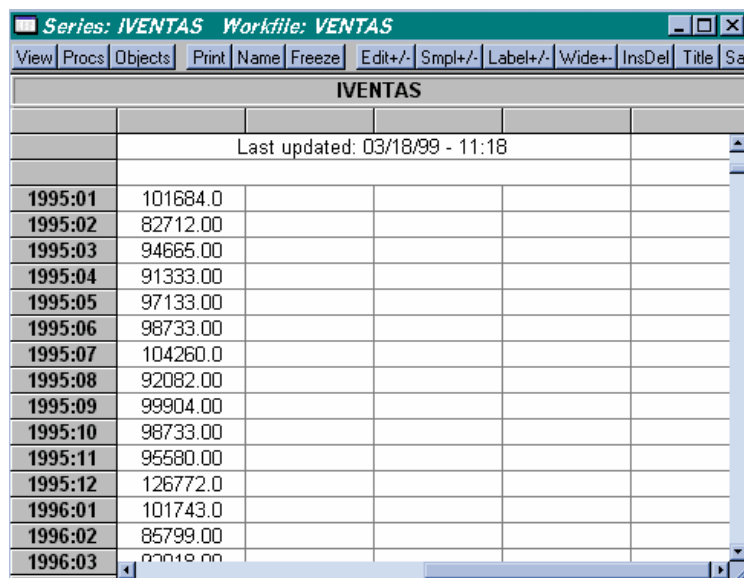
- Introduzca en EViews los datos de la serie. Analice con un gráfico la presencia clara de los componentes teóricos de una serie.
- Obtenga el componente estacional a partir del método de la relación a la media móvil.
- Compruebe con un gráfico la presencia de tendencia en la serie corregida de estacionalidad, y aíslela a partir de la aplicación de diferencias sucesivas.
- Obtenga alternativamente la componente estacional con el método X-11, y compare los resultados.

## Solución

### A) INTRODUCCIÓN DE DATOS Y ANÁLISIS DE LOS COMPONENTES.

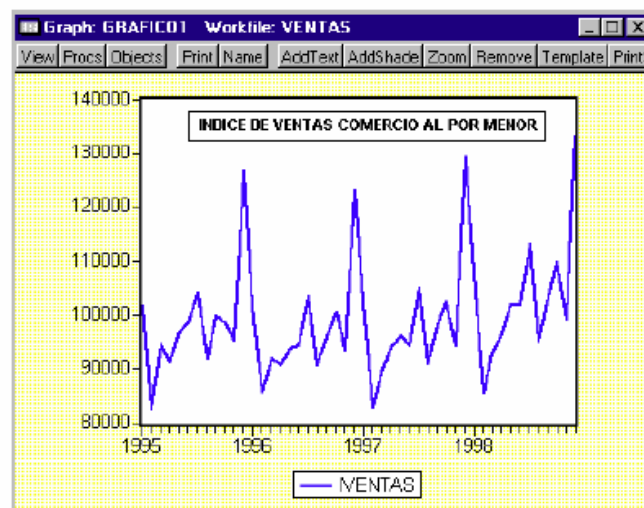
En primer lugar definimos el archivo de trabajo o workfile, definiendo una frecuencia mensual con un rango de enero de 1995 a diciembre de 1998. Una vez definido el archivo de trabajo, creamos un objeto serie, e introducimos los datos, bien sea dato a dato (picado de datos) o importando la serie de Excel a través de la función explícita de EViews.

En el menú del workfile podemos seleccionar la opción SHOW para abrir la ventana del objeto serie, o bien podemos pulsar dos veces sobre la misma con el botón izquierdo del ratón. La apariencia que tiene la ventana de la serie IVENTAS es similar a la de una hoja de cálculo, con la información dispuesta en celdas. Como vemos, tenemos ahora un nuevo menú específico de esta ventana. Si, por ejemplo, elegimos la opción EDIT podemos editar la hoja de datos e introducir modificaciones en la misma.



IVENTAS	
Last updated: 03/18/99 - 11:18	
1995:01	101684.0
1995:02	82712.00
1995:03	94665.00
1995:04	91333.00
1995:05	97133.00
1995:06	98733.00
1995:07	104260.0
1995:08	92082.00
1995:09	99904.00
1995:10	98733.00
1995:11	95580.00
1995:12	126772.0
1996:01	101743.0
1996:02	85799.00
1996:03	97018.00

Para visualizar un gráfico de la misma, accedemos a la opción VIEW / LINE GRAPH que nos permite obtener un gráfico de evolución temporal de la serie en el período muestral indicado previamente.





<b>Periodo</b>	<b>IVENTAS</b>
ene-95	101684,0
feb-95	82712,0
mar-95	94665,0
abr-95	91333,0
may-95	97133,0
jun-95	98733,0
jul-95	104260,0
ago-95	92082,0
sep-95	99904,0
oct-95	98733,0
nov-95	95580,0
dic-95	126772,0
ene-96	101743,0
feb-96	85799,0
mar-96	92018,0
abr-96	90851,0
may-96	93858,0
jun-96	94407,0
jul-96	103603,0
ago-96	90698,0
sep-96	96096,0
oct-96	100762,0
nov-96	93643,0
dic-96	123248,0
ene-97	102661,0
feb-97	82978,0
mar-97	89527,0
abr-97	94226,0
may-97	96261,0
jun-97	94382,0
jul-97	105156,0
ago-97	90961,0
sep-97	98168,0
oct-97	102545,0
nov-97	94367,0
dic-97	129456,0
ene-98	108149,0
feb-98	85679,0
mar-98	93303,0
abr-98	96863,0
may-98	102144,0
jun-98	101962,0
jul-98	113154,0
ago-98	95141,0
sep-98	103557,0
oct-98	109875,0
nov-98	99527,0
dic-98	133554,0

La observación del gráfico de la serie IVENTAS nos permite apreciar la presencia de una componente de estacionalidad clara en la serie, identificada en las "subidas" que se presentan en los meses de diciembre respecto al comportamiento medio de los meses restantes. Esto significa que en el último mes del año se comprueba una mayor compra de bienes y servicios en comparación a la que cabría esperar en meses anteriores. Por otra parte, la evolución general de la serie, es decir, a medio/largo plazo nos muestra una tendencia creciente. De hecho, puede imaginarse una línea recta con pendiente creciente en torno a la cual fluctúan los valores de la serie.

Para la aplicación de determinadas técnicas de predicción, en particular las denominadas elementales y en situaciones con historia (medias móviles y algunos tipos de alisados exponenciales) debemos proceder, previamente, a la transformación de los datos de la serie, para aislar la tendencia y estacionalidad.

## **B) ESTACIONALIDAD CON RELACIÓN A LA MEDIA MÓVIL.**

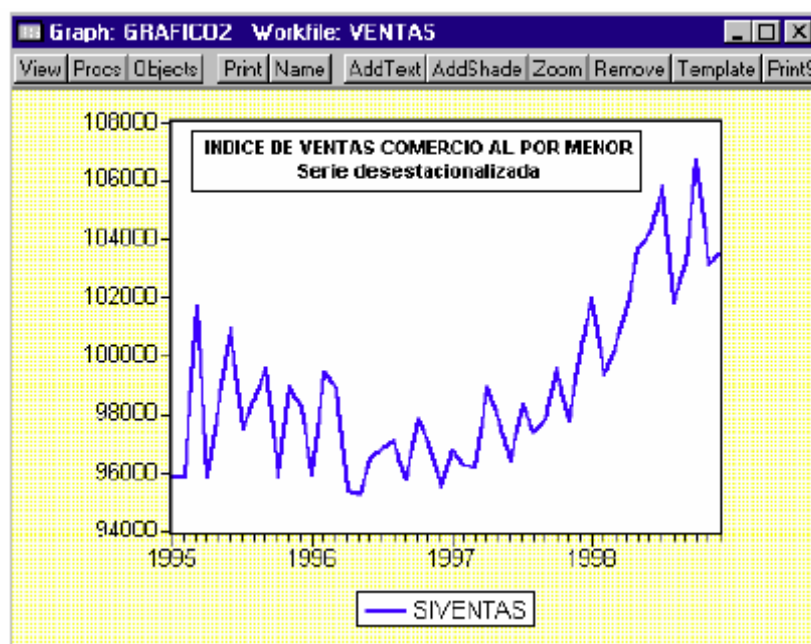
Dado que nuestra serie presenta, como hemos comentado, tendencia y estacionalidad vamos a proceder a la transformación de la misma. En primer lugar, comenzamos con la desestacionalización de la serie, procedimiento que consiste en la obtención de un factor de estacionalidad propio de cada mes. Para ello suponemos un planteamiento multiplicativo de descomposición de la serie. De esta forma, la serie original (IVENTAS) será igual a una serie desestacionalizada (SIVENTAS) multiplicada por un factor (FACTOR):

$IVENTAS = SIVENTAS * FACTOR$  lo que significa que obtendremos la serie desestacionalizada (SIVENTAS) dividiendo los datos de la serie original entre un factor que contiene doce datos mensuales (uno diferente para cada mes).

Para realizar esta operación, en el menú principal de EViews, seleccionamos QUICK / SERIES STATISTICS / SEASONAL ADJUSTMENT, e indicamos que vamos a trabajar sobre la variable IVENTAS. A continuación, seleccionamos el método de desestacionalización (Adjustment Method), indicamos el nombre para la variable desestacionalizada (Adjusted Series) y para el factor (Factors).



Dado que EViews ha generado la nueva serie desestacionalizada (SIVENTAS) podemos visualizar un gráfico de la misma, en forma similar a como hicimos con la serie IVENTAS.



### C) TENDENCIA CON DIFERENCIAS.

El gráfico de la serie SIVENTAS muestra fluctuaciones que ya no se corresponden con la estacionalidad, puesto que ésta ha sido corregida y como es lógico sigue presentando tendencia. Una forma sencilla, y la más empleada, para corregir la tendencia consiste en la aplicación de primeras diferencias de la serie.

Si al representar gráficamente esta serie,  $z_t$ , comprobamos que oscila alrededor del mismo valor, entonces tendríamos que se trata de la serie original sin tendencia. Si  $z_t$  crece o decrece a largo plazo, entonces aún presenta tendencia y habría que volver a diferenciar y, en caso necesario, seguiríamos diferenciando, teniendo en cuenta que en cada diferenciación perdemos una observación. En cualquier caso, lo habitual es que con una primera diferencia la serie ya fluctúe en torno al mismo valor, es decir, ya no presente tendencia y, como mucho, se realizaría una segunda diferencia.

Para eliminar la tendencia a la serie SIVENTAS, ya corregida de estacionalidad, seleccionamos en el menú del workfile la opción GENR y escribimos la operación a realizar, generando una nueva serie DSIVENTAS a partir de las primeras diferencias de SIVENTAS:



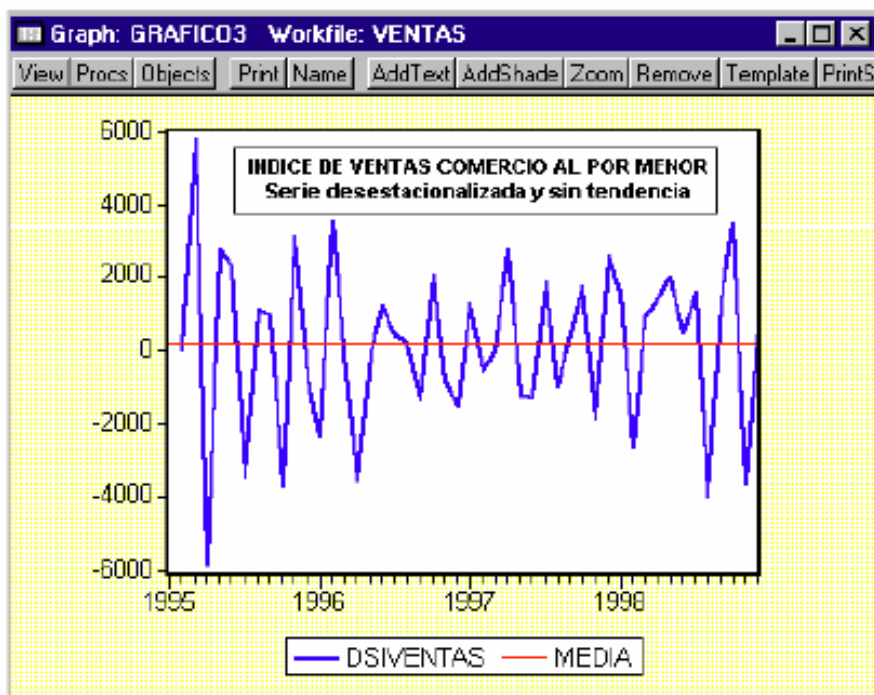
Podemos ahora visualizar un gráfico de la nueva serie DSIVENTAS, serie sin estacionalidad y sin tendencia. Para comprobar que los datos de la serie transformada fluctúan en torno a un mismo valor, calculamos la media de esta serie generando (GENR) una variable denominada MEDIA, tal que:

$MEDIA = @MEAN(DSIVENTAS)$

En el gráfico representamos la serie DSIVENTAS y su valor medio (MEDIA). Para ello, elegimos la opción SHOW en el menú de la ventana del workfile e indicamos que queremos ver las dos series mencionadas. EViews considera que se trata de un "grupo" de series y nos presenta ambas series en una ventana denominada GROUP a la que podemos dar un nombre (NAME) si así lo queremos.

Group: UNTITLED Workfile: VENTAS					
View	Proc	Objects	Print	Name	Freeze
obs	DSIVENTAS	MEDIA			
1995:01	NA	162.9506			
1995:02	45.56531	162.9506			
1995:03	5817.469	162.9506			
1995:04	-5840.364	162.9506			
1995:05	2721.189	162.9506			
1995:06	2299.126	162.9506			
1995:07	-3412.626	162.9506			
1995:08	1096.686	162.9506			
1995:09	976.2658	162.9506			
1995:10	-3685.198	162.9506			
1995:11	3123.061	162.9506			
1995:12	-741.3181	162.9506			
1996:01	-2344.211	162.9506			
1996:02	3570.839	162.9506			
1996:03	-608.9307	162.9506			
1996:04	-3501.105	162.9506			
1996:05	-98.53923	162.9506			
1996:06					

Para ver el gráfico con la representación de la serie DSIVENTAS y su MEDIA, seleccionamos VIEW / GRAPH / LINE. El gráfico nos sugiere una oscilación de los valores de la serie DSIVENTAS en torno al valor medio, por lo que podemos considerarla ya como una serie sin tendencia. Es importante comentar aquí que el menú de la ventana de una serie y el menú de un grupo de series difieren, como es lógico, en su contenido. Las diferencias se encuentran, fundamentalmente, en el contenido de los submenús VIEW y PROCS.

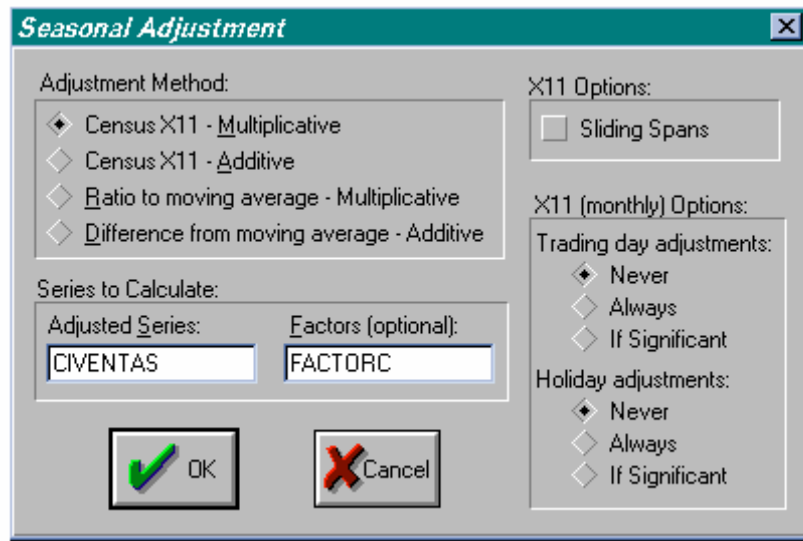


#### D) ESTACIONALIDAD CON X-11.

Finalmente, y a modo de comparación, hemos desestacionalizado la serie original, IVENTAS, con otro procedimiento de los disponibles en EViews: el método CensusX11.

Como ya hicimos con anterioridad, seleccionamos en el menú principal QUICK / SERIES STATISTICS / SEASONAL ADJUSTMENT. Nuestra serie original, la que vamos a desestacionalizar, es IVENTAS, y la serie desestacionalizada se llamará CIVENTAS, siendo FACTORC el nombre elegido para el factor estacional con el método CensusX11-Multiplicativo.

Una vez obtenida la nueva serie, CIVENTAS, desestacionalizada con el procedimiento Census X11, procedemos a representarla en un gráfico junto con la serie SIVENTAS, desestacionalizada con el método de la media móvil. Con la opción SHOW, del menú del workfile, abrimos un grupo que contenga ambas series y en el menú de esta ventana de grupo elegimos VIEW / GRAPH / LINE.

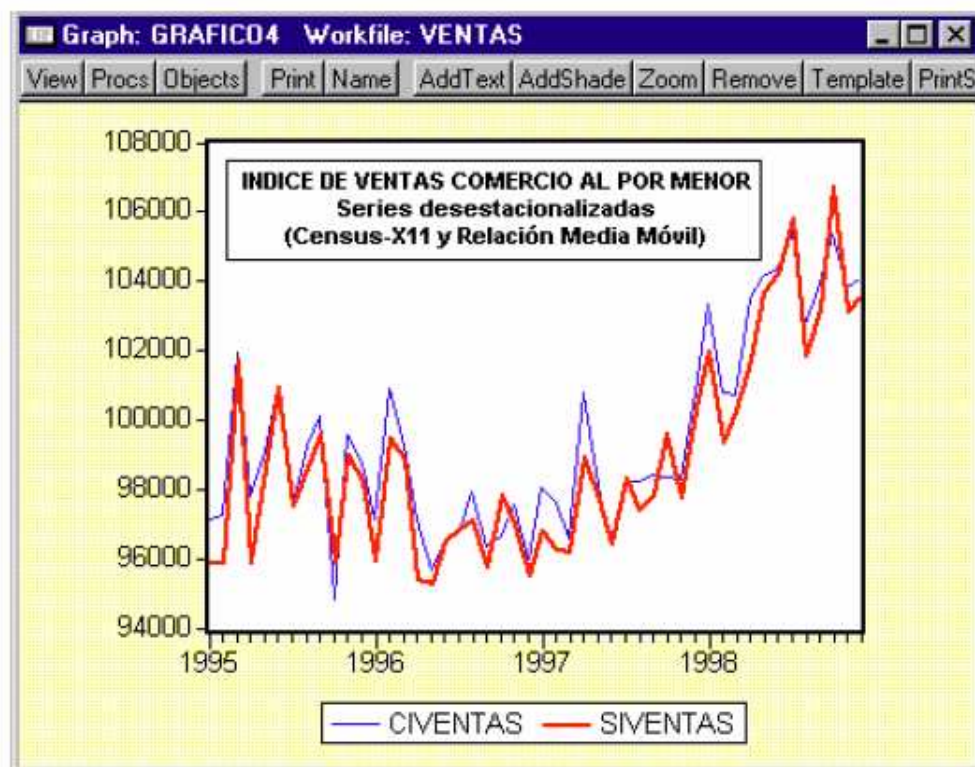


La evolución que presentan ambas series es similar, aunque parece que la serie SIVENTAS, representada con un trazo grueso, suaviza más las oscilaciones de los valores de las ventas del comercio al por menor corregidas de estacionalidad. Posiblemente sea más ilustrativo comprobar conjuntamente los valores de los dos factores de estacionalidad que han dado lugar a ambas series, donde:

$$\text{SIVENTAS} = \text{IVENTAS} / \text{FACTOR}$$

Mientras que,

$$\text{CIVENTAS} = \text{IVENTAS} / \text{FACTORC}$$



Comparamos en la siguiente tabla los valores de los dos factores:

	Relación a la Media Móvil FACTOR	Census X-11 FACTORC
Enero	106,03	104,68
Febrero	86,21	85,00
Marzo	93,02	92,62
Abril	95,21	93,57
Mayo	98,47	98,08
Junio	97,81	97,69
Julio	106,90	107,33
Agosto	93,36	92,60
Septiembre	100,30	99,63
Octubre	102,93	104,32
Noviembre	96,51	95,88
Diciembre	128,96	128,36

La principal diferencia entre el factor calculado por el método de Relación a la media móvil y el del método CensusX11 es que en este segundo caso el factor puede variar año a año, mientras que en el primer caso se asume un valor constante. En este cuadro hemos seleccionado los datos de cada factor para el año 1998, seleccionando el SMPL 1998:01 1998:12. En líneas generales, ambos factores son muy parecidos, simplemente se aprecia un valor menor en los datos calculados para cada mes por el factor del método CensusX11 (FACTORC). Realmente, podríamos utilizar indistintamente cualquiera de los dos procedimientos disponibles en EViews, mientras que si trabajamos con Micro-TSP sólo podemos aplicar el de Relación a la media móvil.

Como hemos señalado, el factor de estacionalidad del método de Relación a la media móvil (FACTOR) es el mismo en cada año. Sin embargo, si visualizamos los datos del factor de estacionalidad el método CensusX11 (FACTORC) para todo el periodo muestral (1995:01 - 1998:12) comprobaremos que no permanecen constantes.

COMPARACIÓN DE FACTORES MENSUALES DE DESESTACIONALIZACIÓN (FACTORC) DEL MÉTODO CENSUS-X11				
	1995	1996	1997	1998
Enero	104,69	104,77	104,73	104,68
Febrero	85,06	85,05	85,02	85,00
Marzo	92,84	92,72	92,69	92,62
Abril	93,40	93,52	93,50	93,57
Mayo	98,03	98,12	98,11	98,08
Junio	97,86	97,89	97,88	97,69
Julio	106,90	107,09	107,09	107,33
Agosto	92,77	92,60	92,61	92,60
Septiembre	99,81	99,75	99,77	99,63
Octubre	104,09	104,26	104,28	104,32
Noviembre	96,02	95,98	96,01	95,88
Diciembre	128,36	128,39	128,45	128,36



### B.8. ¿Qué técnica utilizar?

#### Resumen:

- ✚ Las técnicas de medias móviles y el alisado exponencial simple, así como los métodos *naïve*, sólo son utilizables en series sin tendencia y sin estacionalidad. Para aplicar estas técnicas a otras series es necesario haber eliminado previamente ambos componentes. Su aplicación se centra en el corto plazo.
- ✚ El alisado exponencial doble (o de Brown), el alisado Holt-Winters con dos parámetros, y el ajuste de funciones matemáticas, puede emplearse en series con tendencia, pero sin estacionalidad. Si se parte de una serie desestacionalizada, es habitual afectar las predicciones por los factores de estacionalidad. Tan sólo el ajuste de funciones matemáticas podría utilizarse para predicción a medio plazo, o incluso en algunas expresiones, para el largo.
- ✚ El alisado exponencial Holt-Winters con tres parámetros puede aplicarse directamente a series con tendencia y estacionalidad. Es recomendable su utilización sólo para predicción a corto plazo.

*Posibilidades de aplicación de técnicas elementales de predicción en función de los componentes de la serie*

Serie	Aplicación directa			Alternativa
Con tendencia y estacionalidad	AHW3P			<i>Eliminación de tendencia, desestacionalización y posterior afectación del índice de estacionalidad:</i> ING                      MM                      AES
				<i>Desestacionalización y posterior afectación del índice de estacionalidad:</i> AED                      AHW2P                      AFM
Con estacionalidad	AHW3P			<i>Desestacionalización y posterior afectación del índice de estacionalidad:</i> ING                      MM                      AES
Con tendencia	AED	AHW2P	AFM	<i>Eliminación de tendencia</i> ING                      MM                      AES
Sin tendencia ni estacionalidad	ING	MM	AES	

ING=Métodos ingenuos o *naï ve*; MM=Medias móviles; AES=Alisado exponencial simple; AED=Alisado exponencial doble (o de Brown); AHW2P=Alisado Holt-Winters con doble parámetro; AFM=Ajuste de funciones matemáticas; AHW3P=Alisado Holt-Winters con triple parámetro.

### ACTIVIDAD: Localización de información y análisis de indicadores adelantados.

A lo largo de la Unidad hemos puesto de manifiesto la importancia de la disponibilidad de información estadística. En la Lectura Adicional 1 hemos descrito las fuentes de localización de información estadística. Además, en la Unidad hemos descrito a los indicadores adelantados como una de las técnicas que se utilizan para llevar a cabo la predicción de una variable

cuando no se dispone de información histórica de la misma, o se renuncia a la misma. En algunos casos, existen magnitudes económicas estrechamente relacionadas con otras, de las que sí es posible disponer de información estadística, o de mejor calidad: frecuencia, actualización, disponibilidad electrónica...

Además algunos indicadores presentan, como vimos, una característica clave a efectos de predicción; y es que algunos de ellos presentan la propiedad de "adelantar" el comportamiento de algunas variables que describen el comportamiento de algunos sectores. Por ejemplo, la evidencia nos muestra que una caída hoy en el indicador conocido como consumo aparente de cemento, anticipa una contracción en magnitudes que describen la actividad económica global del sector de la construcción, como es el Valor Añadido Bruto. Téngase en cuenta, que este sector representa una parte muy significativa de la economía de un país, y el conocimiento sobre su comportamiento, puede ayudar en la anticipación sobre la evolución de magnitudes más agregadas, como el PIB.

Para tomar conocimiento de los principales indicadores que describen de manera desagregada el funcionamiento de una economía, le sugerimos que realice un breve informe sobre los indicadores de un sector. Ello, además, le permitirá ganar soltura para la obtención de la información estadística necesaria. Elija el sector de la economía en el que esté más interesado (industria, construcción, energía,...). Para tal sector, seleccione los indicadores que describen la actividad económica del mismo. Para ello, le sugerimos que revise anuarios estadísticos, informes patronales, páginas Web, ... Aproveche los comentarios realizados en la Lectura Adicional 1. De esa revisión habrá obtenido una buena batería de indicadores. Busque información que analice la posibilidad de que alguno de esos indicadores sea adelantado.

Por último le sugerimos que realice un análisis descriptivo de tales indicadores: analice las tasas de variación, su evolución histórica, busque coincidencias entre los mismos y, sobre todo, y en la medida de las disponibilidades estadísticas, analice las coincidencias con la magnitud. Aunque no sea más que de un modo intuitivo, ¿se atrevería a decir que son indicadores adelantados? ¿Se arriesgaría a realizar con ellos una predicción de la magnitud?

### **CASO DE APLICACIÓN: Predicción en una serie con componente estacional: el consumo de energía eléctrica.**

El consumo de electricidad es un indicador parcial de actividad económica de gran relevancia en predicción macroeconómica y sectorial, al considerarse como ejemplo de indicador adelantado. En general, el consumo eléctrico mes a mes puede ser explicado en forma bastante operativa por el efecto conjunto de cuatro tipos de causas:

- a) Estacionalidad propia del mes de que se trata.
- b) Condiciones climatológicas excepcionales.
- c) Variaciones en el número de días laborales o especiales.

d) Situación económica del país en su conjunto o de los principales sectores demandantes de electricidad.

En este caso de aplicación trabajaremos con un indicador de consumo de energía obtenido a partir de la transformación de la serie original de Red Eléctrica de España para el periodo comprendido entre enero de 1981 y octubre de 1998. Estas transformaciones realizadas pretenden tener presente aspectos como el número de días de cada mes, efectos de laboralidad, temperatura, situación económica, ...

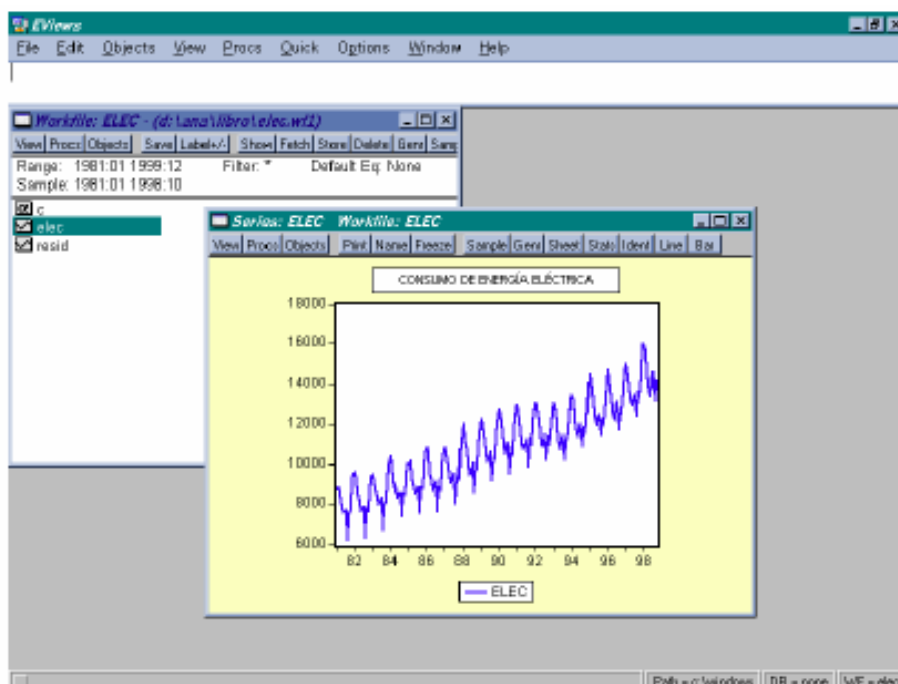
En este caso utilizaremos el programa EViews, y aplicaremos las distintas técnicas de desestacionalización para la posterior predicción de los valores futuros de este indicador.

## **Solución**

### **Predicción con una serie con componente estacional: el consumo de energía eléctrica**

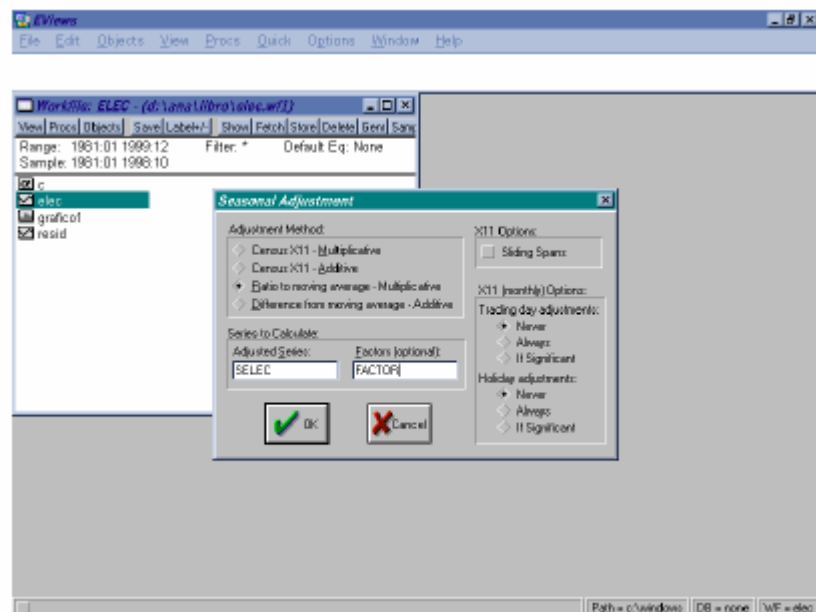
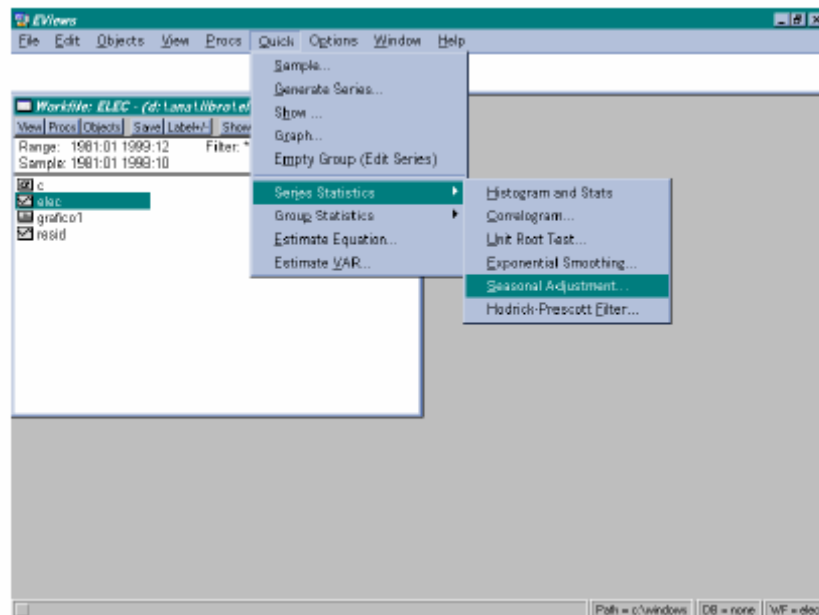
En primer lugar, creamos una sesión de trabajo en Econometric Views, para datos mensuales desde 1981:01 hasta 1999:12, abarcando así el periodo de predicción. En el menú principal seleccionamos sucesivamente: FILE / NEW / WORKFILE. Una vez creada la sesión de trabajo, cargamos la serie del workfile mes.

Como ya sabemos, se crea una ventana propia para el trabajo (*workfile*), en ésta seleccionamos SHOW para ver los datos de la serie, y dentro de esta ventana la opción VIEW / LINE GRAPH, para visualizar un gráfico de la serie, como el que aparece a continuación. Se aprecia una componente de tendencia en la serie, creciente a lo largo del tiempo, y una acusada *estacionalidad*, es decir, un movimiento regular que se repite año tras año con la misma intensidad y en el mismo

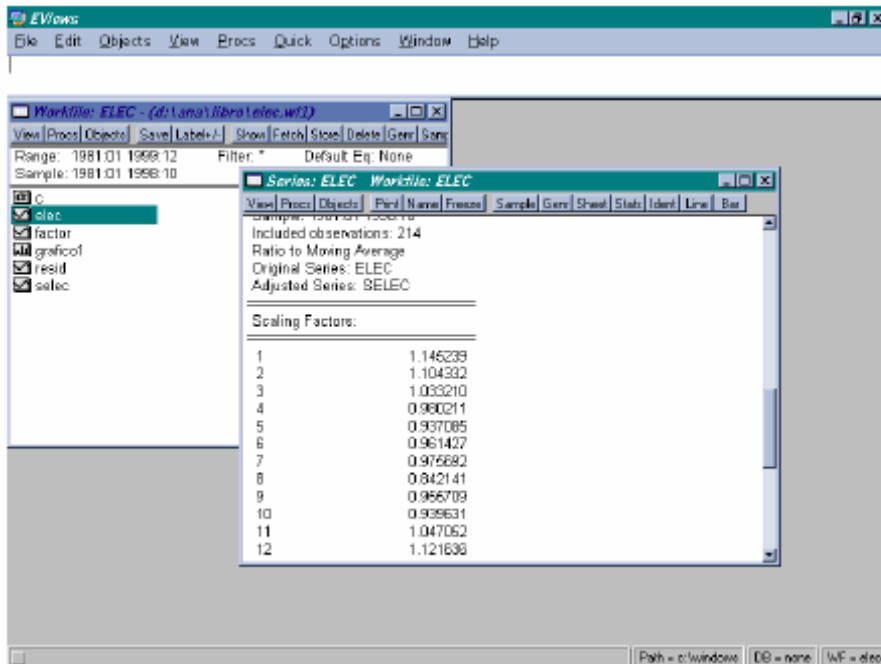


Año/Mes	Elec	Año/Mes	Elec	Año/Mes	Elec	Año/Mes	Elec	Año/Mes	Elec
1981M01	8794	1984M11	9363	1988M09	10065	1992M07	11535	1996M05	12255
1981M02	8880	1984M12	10039	1988M10	9809	1992M08	10257	1996M06	12933
1981M03	8577	1985M01	10053	1988M11	10973	1992M09	11416	1996M07	13119
1981M04	7859	1985M02	10207	1988M12	11987	1992M10	10964	1996M08	11796
1981M05	7614	1985M03	9288	1989M01	12305	1992M11	12071	1996M09	12542
1981M06	7679	1985M04	8882	1989M02	11965	1992M12	12661	1996M10	12542
1981M07	7744	1985M05	8545	1989M03	11180	1993M01	13059	1996M11	13690
1981M08	6239	1985M06	8751	1989M04	10636	1993M02	12680	1996M12	14658
1981M09	7524	1985M07	8882	1989M05	10117	1993M03	12062	1997M01	15027
1981M10	7703	1985M08	7621	1989M06	10386	1993M04	11184	1997M02	14366
1981M11	9313	1985M09	8905	1989M07	10633	1993M05	10782	1997M03	13866
1981M12	9604	1985M10	8836	1989M08	9259	1993M06	11224	1997M04	13279
1982M01	9654	1985M11	9786	1989M09	10369	1993M07	11491	1997M05	12937
1982M02	9265	1985M12	10578	1989M10	10337	1993M08	10252	1997M06	13212
1982M03	8588	1986M01	10876	1989M11	11572	1993M09	11240	1997M07	13486
1982M04	8123	1986M02	10612	1989M12	12495	1993M10	11283	1997M08	12426
1982M05	7681	1986M03	9585	1990M01	12747	1993M11	12334	1997M09	13619
1982M06	7815	1986M04	9244	1990M02	12485	1993M12	13193	1997M10	13349
1982M07	7894	1986M05	8729	1990M03	11595	1994M01	13447	1997M11	14430
1982M08	6370	1986M06	9033	1990M04	11008	1994M02	13180	1997M12	16020
1982M09	7829	1986M07	9173	1990M05	10519	1994M03	12333	1998M01	15996
1982M10	7672	1986M08	7665	1990M06	11076	1994M04	11841	1998M02	15645
1982M11	8693	1986M09	8907	1990M07	11051	1994M05	11361	1998M03	14559
1982M12	9421	1986M10	8738	1990M08	9587	1994M06	11893	1998M04	13956
1983M01	9530	1986M11	9854	1990M09	10883	1994M07	12083	1998M05	13430
1983M02	9100	1986M12	10481	1990M10	10507	1994M08	10920	1998M06	14111
1983M03	8787	1987M01	10862	1990M11	11813	1994M09	11915	1998M07	14659
1983M04	8352	1987M02	10615	1990M12	12830	1994M10	11606	1998M08	13196
1983M05	7990	1987M03	9874	1991M01	12946	1994M11	12880	1998M09	14224
1983M06	8174	1987M04	9536	1991M02	12634	1994M12	13674	1998M10	13870
1983M07	8323	1987M05	9114	1991M03	11756	1995M01	14527		
1983M08	6752	1987M06	9437	1991M04	11050	1995M02	13664		
1983M09	8044	1987M07	9600	1991M05	10826	1995M03	13117		
1983M10	8104	1987M08	8037	1991M06	11076	1995M04	12433		
1983M11	9318	1987M09	9421	1991M07	11252	1995M05	12018		
1983M12	9916	1987M10	9309	1991M08	9900	1995M06	12399		
1984M01	10400	1987M11	10474	1991M09	11331	1995M07	12516		
1984M02	10000	1987M12	11667	1991M10	10999	1995M08	11497		
1984M03	9181	1988M01	12039	1991M11	12169	1995M09	12366		
1984M04	8772	1988M02	11214	1991M12	12952	1995M10	12042		
1984M05	8338	1988M03	10702	1992M01	13066	1995M11	13181		
1984M06	8471	1988M04	10078	1992M02	12896	1995M12	14124		
1984M07	8611	1988M05	9494	1992M03	12173	1996M01	14786		
1984M08	7442	1988M06	9655	1992M04	11589	1996M02	13758		
1984M09	8560	1988M07	10069	1992M05	11002	1996M03	13348		
1984M10	8480	1988M08	8600	1992M06	10930	1996M04	12687		

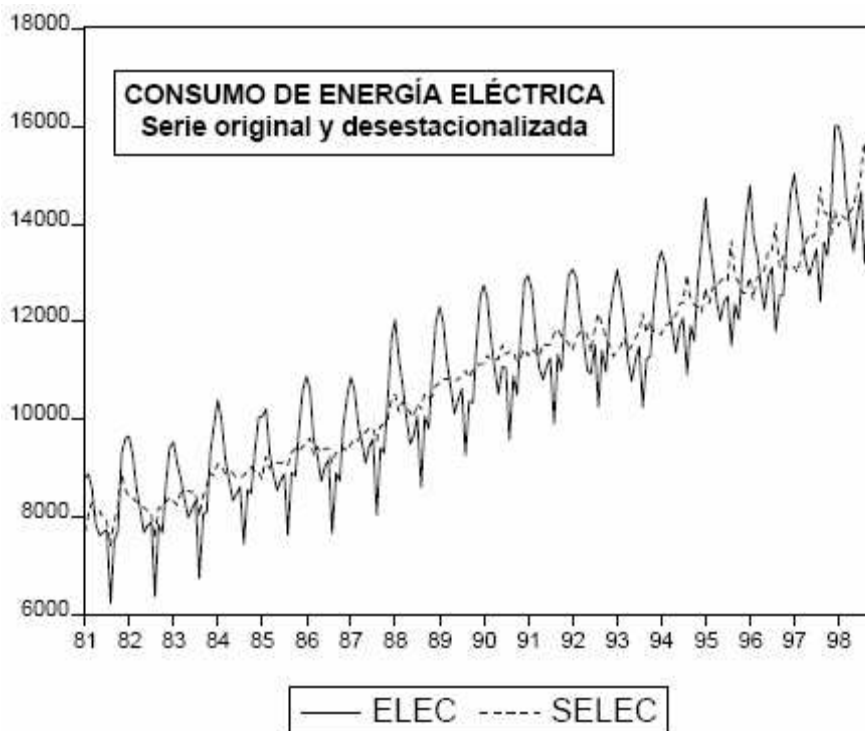
Procedemos al cálculo de los factores de estacionalidad de la serie con un procedimiento que, según las convenciones del X-11, incorpora EViews bajo la opción *Seasonal Adjustment* (menú principal: QUICK / SERIES STATISTICS / SEASONAL ADJUSTMENTS), y que permite generar tanto los índices estacionales de cada mes (FACTOR) como la variable corregida de estacionalidad (SELEC).



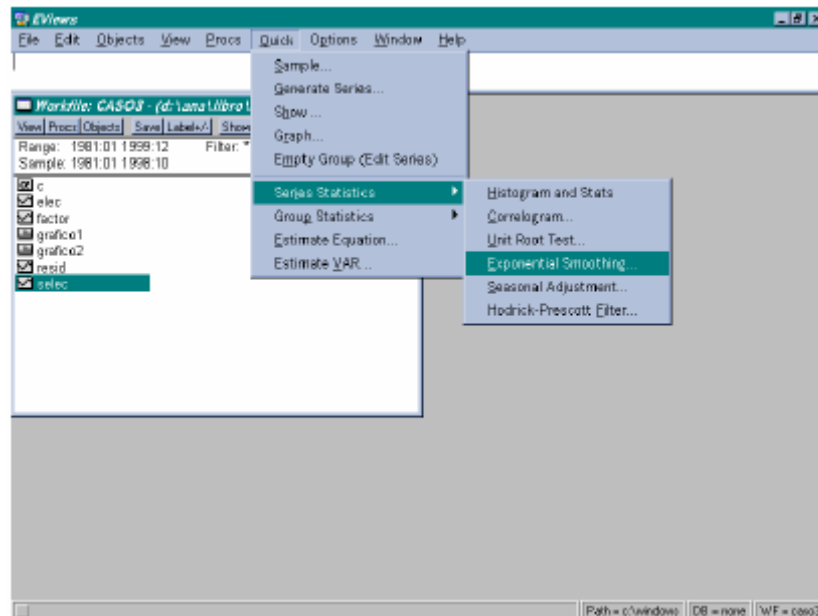
Siendo el factor resultante el que aparece a continuación, con valores en torno a la unidad.



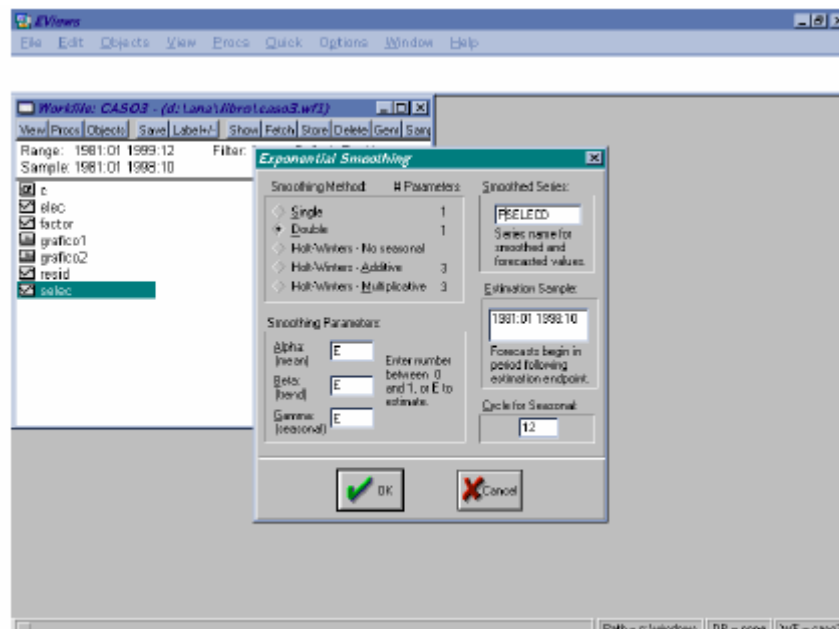
La serie resultante, a la que hemos denominado SELEC, está ya corregida de estacionalidad, a través de un factor multiplicativo de forma que la serie con estacionalidad (ELEC) es igual a una serie desestacionalizada (SELEC) multiplicada por un factor (FACTOR), es decir,  $ELEC = SELEC * FACTOR$ . El siguiente gráfico en el que se muestran ambas series, nos permite apreciar los cambios realizados:



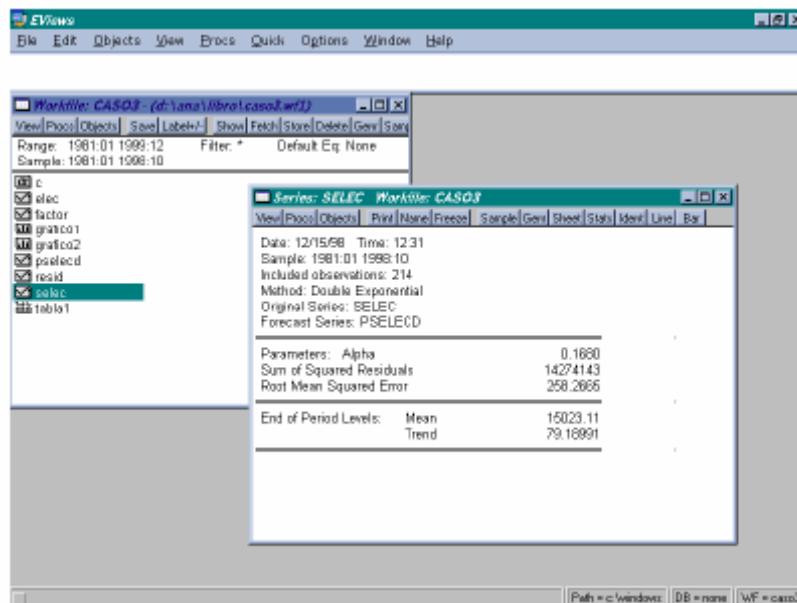
A la serie desestacionalizada, hemos aplicado las dos variantes de alisado exponencial explicadas (doble alisado y Holt-Winters con doble parámetro) que pueden tratar series con tendencia pero no con estacionalidad. Para hacer el alisado exponencial, seleccionamos en el menú principal: QUICK / SERIES STATISTICS / EXPONENTIAL SMOOTHING



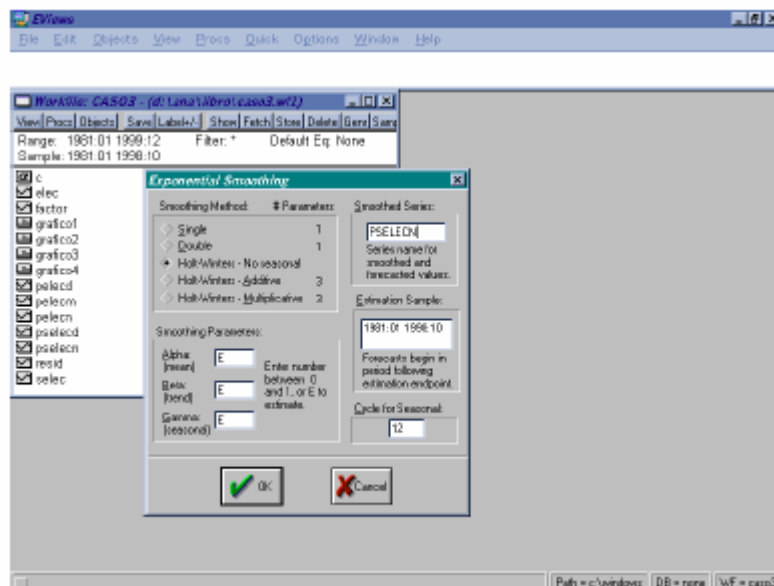
En primer lugar, hacemos el alisado exponencial doble de Brown:



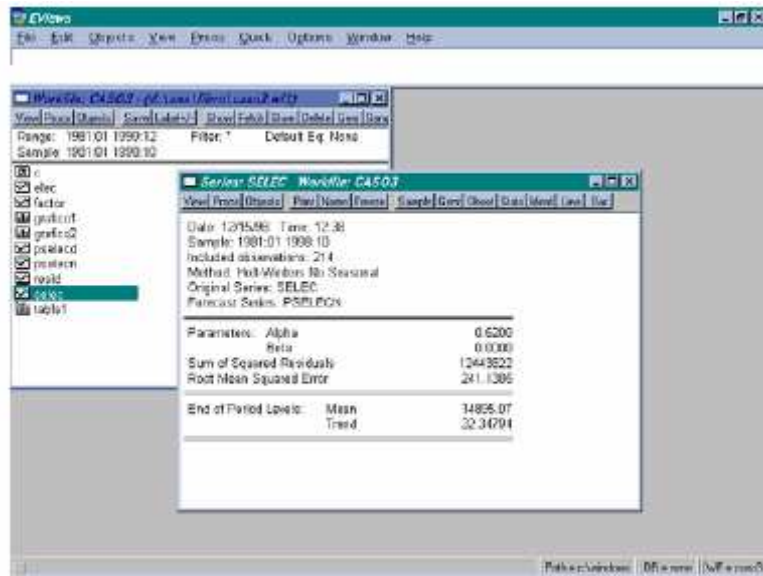
Que nos ofrece los siguientes resultados: el cálculo de la serie de alisado (PSELECD) a partir de la serie corregida de estacionalidad (SELEC), con un parámetro  $\alpha$  óptimo que toma el valor 0,1680:



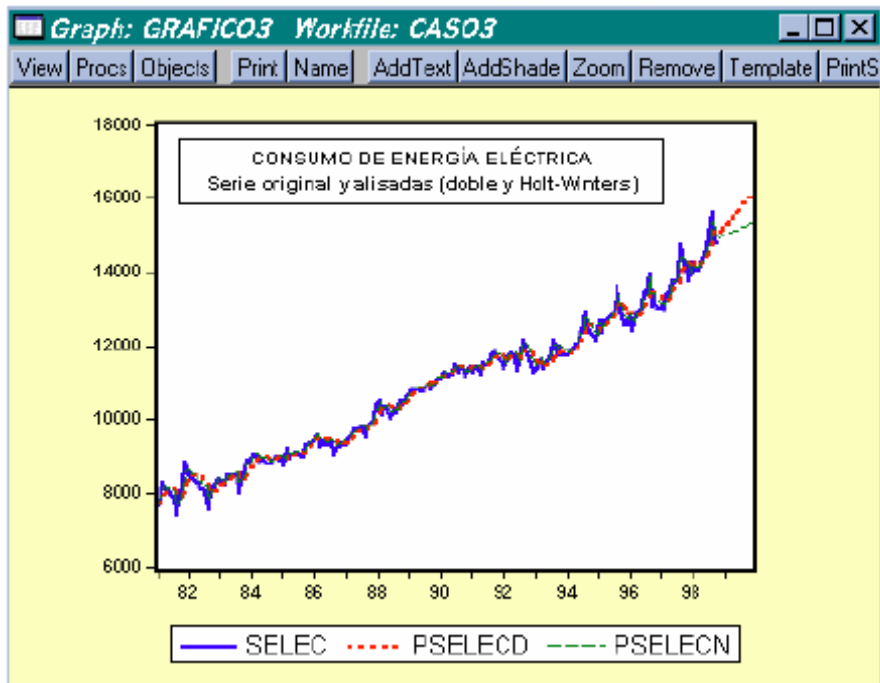
A continuación, procedemos a repetir la operación, pero esta vez seleccionamos el alisado doble de Holt-Winters, que calcula la serie de alisado a partir de dos parámetros,  $\alpha$  y  $\beta$ , que, en este caso, toman como valores óptimos 0,62 y 0,00, indicando este último dato que la serie presenta una tendencia rígida, siempre creciente en todo el periodo histórico.



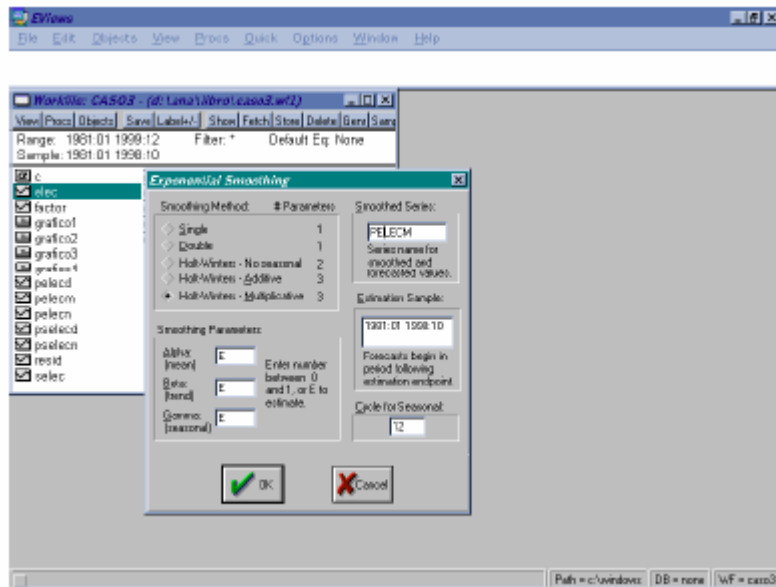




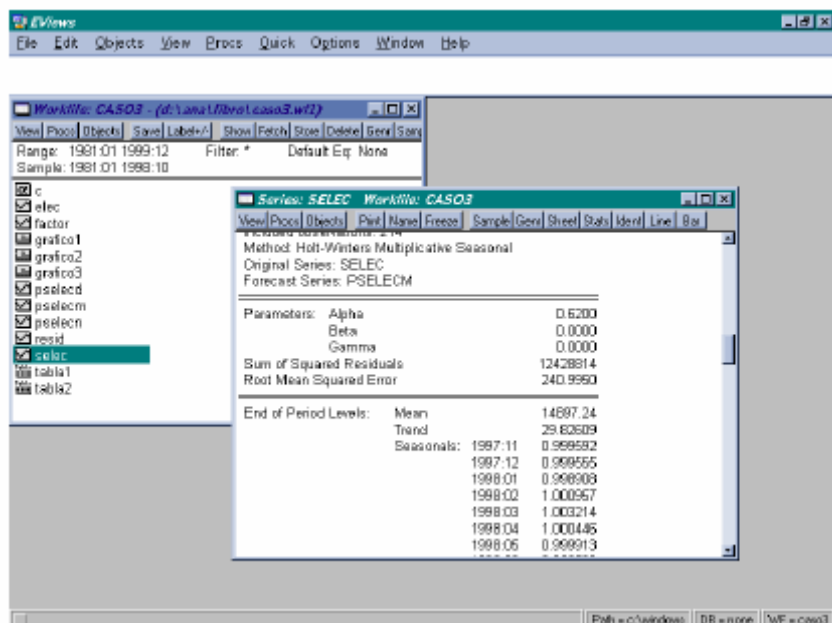
Las series generadas (que hemos notado por PSELECD y PSELECN, la letra inicial indicando que se trata de predicciones y la última referida a la letra de control en EViews del procedimiento de alisado utilizado) se ajustan bastante a la serie original desestacionalizada, como puede verse en el siguiente gráfico.



Directamente puede aplicarse a la serie original (ELEC) la técnica de alisado exponencial de Holt-Winters con tres parámetros.

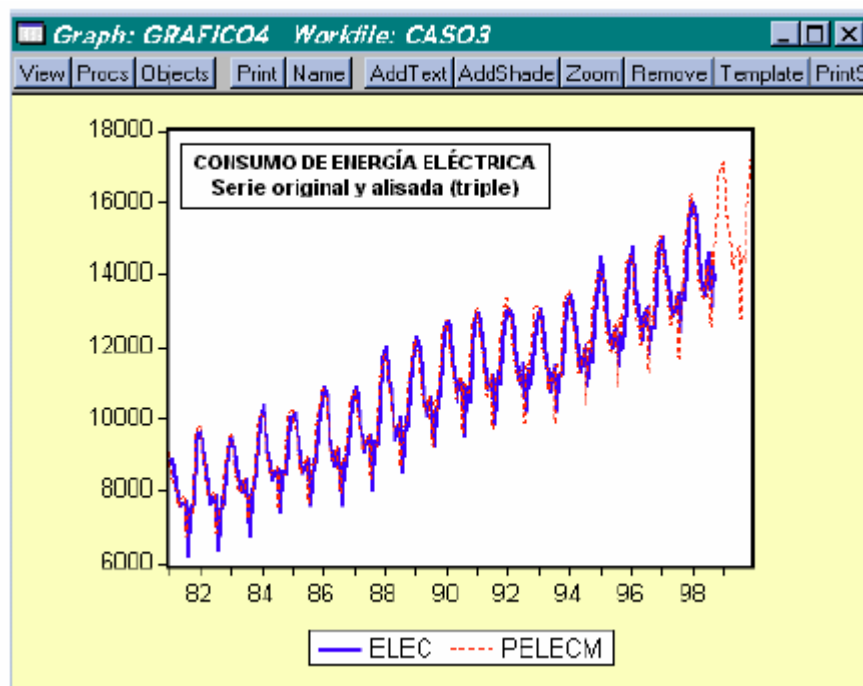


La nulidad del parámetro estacional en la estimación indica la existencia de una *estacionalidad rígida*, que no es preciso corregir con la propia evolución de la serie. Así, cuando EViews proporciona valores cero para los parámetros  $\beta$  y  $\gamma$ , como sucede en este caso, podemos considerar que los componentes de tendencia y estacionalidad se estiman como fijos y no cambiantes.



En la parte inferior de esta salida del alisado, aparecen los valores de media (*mean*) y tendencia (*trend*) al final de periodo de estimación, que se utilizan para los valores alisados de predicción.

Es interesante observar, para este caso, la similitud entre los coeficientes de estacionalidad iniciales, obtenidos por relación a la media móvil, contenidos en la serie FACTOR, y los ahora calculados por Holt-Winters con estacionalidad multiplicativa. La diferencia no es superior a una centésima para ningún mes. La correspondencia entre la serie original (ELEC) y la alisada con este procedimiento (PELECM) se muestran en el gráfico siguiente:

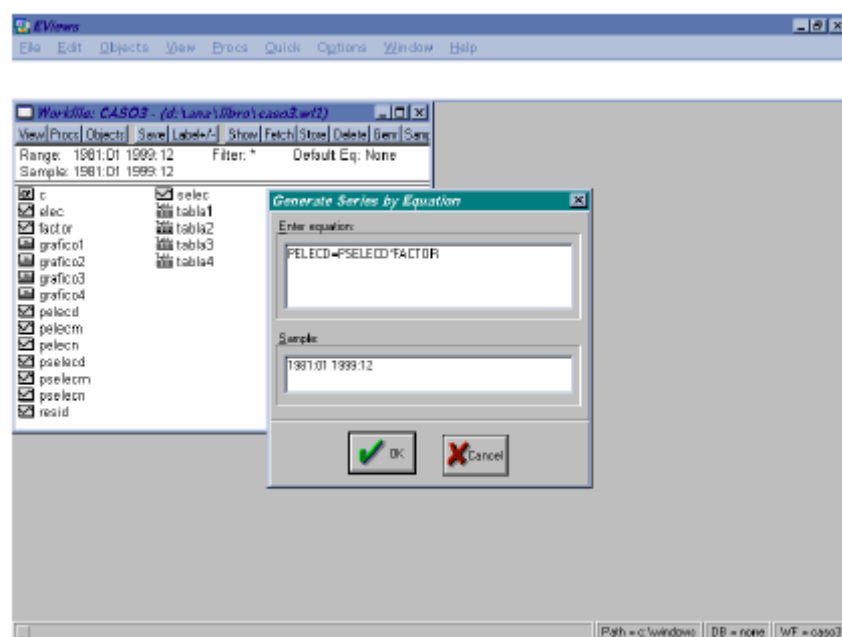


Finalmente, los resultados obtenidos se muestran en la siguiente tabla, en la que hemos seleccionado el periodo comprendido entre enero de 1998 hasta diciembre de 1999, que como sabemos abarca diez meses de historia (hasta octubre de 1998) y catorce de predicción (desde noviembre de 1998 hasta finales de 1999). Como cada serie alisada se ha obtenido a partir de la serie original desestacionalizada, las predicciones que automáticamente calcula Econometric Views no llevan incorporada la estacionalidad que originalmente presentaba la serie.

obs	SELEC	PSELECD	PSELECN
	Serie desestacionalizada	Alisado exponencial doble	Alisado exponencial de Holt-Winters
1998:01	13967.39	14299.47	14211.02
1998:02	14166.94	14252.19	14092.32
1998:03	14091.03	14278.47	14170.93
1998:04	14237.75	14268.02	14153.74
1998:05	14331.68	14305.08	14238.17
1998:06	14677.15	14360.39	14328.49
1998:07	15024.21	14513.95	14577.00
1998:08	15669.59	14741.45	14886.61
1998:09	14883.19	15123.76	15404.39
1998:10	14761.12	15139.59	15113.60
1998:11	NA	15102.30	14927.42
1998:12	NA	15181.49	14959.76
1999:01	NA	15260.68	14992.11
1999:02	NA	15339.87	15024.46
1999:03	NA	15419.06	15056.81
1999:04	NA	15498.25	15089.16
1999:05	NA	15577.44	15121.50
1999:06	NA	15656.63	15153.85
1999:07	NA	15735.82	15186.20
1999:08	NA	15815.01	15218.55
1999:09	NA	15894.20	15250.90
1999:10	NA	15973.39	15283.24
1999:11	NA	16052.58	15315.59
1999:12	NA	16131.77	15347.94

Todas las predicciones realizadas pueden también presentarse incluyendo variaciones estacionales. Para ello, basta con generar para doble alisado y Holt-Winters con dos parámetros, unas nuevas series (PELECD y PELECN, respectivamente) que tomen los valores estimados de tendencia y los multipliquen por el factor de escala de estacionalidad. Para ello, utilizamos la instrucción GENR de la ventana del Workfile que puede verse en la página siguiente.

En el cuadro incluimos, como ejemplo, los resultados para los años 1998 y 1999 completos de la evolución real, la de doble alisado, Holt-Winters con dos parámetros y Holt-Winters con tres parámetros. Como puede observarse, las diferencias entre las distintas variantes no son especialmente significativas en el presente caso.



obs	ELEC Serie original	PELECD Alisado exponencial doble	PELECN Alisado exponencial de Holt-Winters	PELECM Alisado exponencial triple parámetro
1998:01	15996.00	16376.31	16275.02	16253.85
1998:02	15645.00	15739.15	15562.60	15588.89
1998:03	14559.00	14752.66	14641.55	14680.51
1998:04	13956.00	13985.67	13873.66	13849.45
1998:05	13430.00	13405.07	13342.37	13322.25
1998:06	14111.00	13806.46	13775.80	13744.42
1998:07	14659.00	14161.14	14222.67	14214.79
1998:08	13196.00	12414.38	12536.62	12528.38
1998:09	14224.00	14453.92	14722.12	14700.99
1998:10	13870.00	14225.63	14201.20	14199.05
1998:11	NA	15812.90	15629.79	15630.95
1998:12	NA	17028.11	16779.41	16777.34
1999:01	NA	17477.13	17169.55	17149.40
1999:02	NA	16940.31	16591.99	16603.72
1999:03	NA	15931.13	15556.85	15600.59
1999:04	NA	15191.56	14790.56	14788.92
1999:05	NA	14597.38	14170.13	14158.85
1999:06	NA	15052.70	14569.32	14535.34
1999:07	NA	15353.32	14817.06	14794.84
1999:08	NA	13318.47	12816.16	12799.09
1999:09	NA	15190.23	14575.42	14553.63
1999:10	NA	15009.09	14360.61	14335.53
1999:11	NA	16807.89	16036.23	16002.96
1999:12	NA	18093.98	17214.80	17175.85

## C. La simulación en ausencia de información histórica

■ Existen determinadas ocasiones en las que carecemos de información histórica, es decir, de series estadísticas, sobre un determinado fenómeno. En otras, sencillamente renunciamos a utilizar tal información histórica. Ello no menoscaba la posibilidad de afrontar la simulación. En este acápite desarrollamos dos posibilidades que nos permiten la realización de simulaciones sencillas. Aún así, se trata de técnicas que pueden exigir de un cierto nivel de medios, materiales y humanos, para llevarlas a cabo. Analizaremos con ejemplos las posibilidades que nos ofrecen la simulación con fórmulas recursivas y el método Delfos.

■ Por tanto, vamos a desarrollar 2 apartados. Finalizaremos con una actividad propuesta, 2 Casos de Aplicación Práctica y el test de autoevaluación. Hemos estimado en 7 horas la dedicación que esta unidad puede necesitar para un aprovechamiento óptimo. El estudio de los conceptos teóricos puede implicar unas 2 horas, la realización de la actividad propuesta otras 2 horas, y el seguimiento de los Casos de Aplicación unas 3 horas.

### C.1. Simulación mediante fórmulas recursivas

#### Resumen:

- ✚ De entre las opciones posibles dentro de la predicción/simulación con fórmulas matemáticas, vamos a analizar aquí la más frecuente: los modelos recursivos o en cadena causal. En estos modelos existe un orden definido para su resolución. Posteriormente analizaremos los [modelos interdependientes](#).
  - ✚ Las fórmulas individuales que componen el modelo pueden ser de diferentes [tipos](#): identidades, ecuaciones de evolución temporal, ecuaciones autorregresivas o relaciones causales.
  - ✚ El primer antecedente relevante en la utilización de modelos multiecuacionales en situaciones sin historia es la llamada *dinámica de sistemas*, enfoque propuesto por Jay Forrester desde el Massachusetts Institute of Technology a principios de la década de 1960. La dinámica de sistemas permite simular el comportamiento de un sistema social a través de múltiples relaciones entre las variables que lo componen. Para ello crean un lenguaje de programación propio, el DYNAMO, en el que intervienen varios [elementos](#).
  - ✚ La versatilidad de la dinámica de sistemas se ha mostrado por la variedad y relevancia de algunas de las aplicaciones realizadas, que van desde planteamientos empresariales, hasta temas como la obsolescencia técnica, el deterioro urbano o los límites del crecimiento económico (el famoso informe Meadows, patrocinado por el Club de Roma en la década de 1970).
  - ✚ Desarrollaremos todas estas ideas con un [Caso de Aplicación](#).
-

### ***Elementos de la dinámica de sistemas***

- 1.- Variables de nivel o de estado: Son variables acumulativas o de *stock* y recogen, por ello, información sobre el pasado del sistema.
- 2.- Variables de decisión o tasa: Son variables de flujo actual y afectan a la variación de los niveles entre períodos.
- 3.- Variables auxiliares: Son componentes en los que se han subdividido las variables anteriores porque tienen significado independiente a efectos de tratamiento del sistema.
- 4.- Constantes o parámetros: Son datos para la resolución del sistema y pueden referirse a valores fijos de una variable real o de expectativa, desfases fijos en la adaptación entre variables, etc.
- 5.- Operadores o funciones especiales: Son procesos de cálculo definidos a priori sobre las variables del sistema, tales como tablas de interpolación, funciones matemáticas de transformación de variables, distribuciones aleatorias, etc.
- 6.- Operadores retardo o funciones de demora: Dada la importancia de los retardos o desfases temporales en la iteración entre variables, se definen unas funciones especiales para su tratamiento, generalmente según un esquema exponencial.

### **Tipos y ejemplos de fórmulas utilizables en modelos**

**1.- Identidades**: Establecen que una variable es igual a otra o a una transformación de otras (suma, producto, etc.) sin necesidad de coeficientes. Por ejemplo:

$$Y_1 = Y_2$$

$$Y = X_1 + X_2 + \dots + X_k$$

$$Y = X_1 * X_2 * \dots * X_k$$

$$Y = X_1 / X_2$$

**2.- Ecuaciones de evolución temporal**: La única variable explicativa es el tiempo, con muy diferentes variantes. Ejemplos:

$$Y = a + b * t$$

$$Y = a * b^t$$

$$Y = a * t^b$$

$$Y = a * e^{b+ct}$$

**3.- Ecuaciones autorregresiva**: La variable explicativa es ella misma en períodos anteriores. Una variante muy utilizada es la definición de ecuaciones en tasas de variación, bien constantes o incluso variables en el

tiempo. En los denominados modelos de difusión o de contagio, los incrementos pueden depender de las diferencias, en cada momento, a un techo o nivel potencial.

$$Y = Y_0 * (1 + r)^n$$

$$\Delta Y = a * (\bar{Y} - Y_{-1}) \quad (\text{con } \bar{Y} = \text{techo o nivel potencial})$$

$$\Delta Y = a * Y * (\bar{Y} - Y)$$

$$\Delta Y = a * Y * (\bar{Y} - Y) + b(\bar{Y} - Y)$$

$$Y = c_0 + c_1 Y_{-1} + c_2 Y_{-2} + \dots + c_k Y_{-k}$$

**4.- Relaciones causales:** Una variable depende de otra u otras variables explicativas. En el caso más simple se supondrá una relación constante entre dos variables. En otros casos, los coeficientes que ponderan el efecto de cada una de las causas explicativas pueden ser deducidos por analogías, encuestas o experimentos previos, o incluso deducirse de la propia historia, cuando ésta sea utilizable y se considere conveniente.

$$Y / X = b$$

$$\Delta Y / \Delta X = b$$

$$Y = a + bX$$

$$Y = c_0 + c_1 X_1 + \dots + c_k X_k$$

## CASO DE APLICACIÓN

*Simulación de estados financieros mediante fórmulas recursivas.*

*En este caso de aplicación plantearemos la utilización de fórmulas recursivas para abordar la simulación bajo distintos escenarios (normales y alternativos) de los estados financieros (cuenta de pérdidas y ganancias) de una empresa periodística. Tendremos en cuenta variables clave como la tirada y la publicidad, junto a otras como la circulación efectiva o el desarrollo económico del país. El caso se resuelve en una hoja de Excel. Se recomienda a los usuarios del curso que una vez seguido el caso de aplicación repliquen el mismo con un tema de su elección.*

En este caso de aplicación plantearemos la utilización de fórmulas recursivas para abordar la simulación bajo distintos **escenarios** (normales y alternativos) de los estados financieros (cuenta de pérdidas y ganancias) de una empresa periodística. Suponemos que una empresa periodística viene observando durante los últimos años un mayor incremento de sus



capítulos de gastos con relación a sus ingresos, con su correspondiente deterioro de la cuenta de resultados.

Para definir una nueva política integral de la empresa se realiza una previsión de estados financieros con un horizonte de medio plazo. Aunque interesa conocer la proyección de la empresa en "circunstancias normales", el análisis deberá considerar la posibilidad de diseñar estados financieros alternativos, según diferentes hipótesis y medidas de actuación. En particular, aquí vamos a centrarnos en la elaboración de cuentas previsionales de pérdidas y ganancias para los tres próximos ejercicios y, para mayor simplicidad vamos a suponer que la empresa limita su actividad a un único periódico de carácter diario.

Dado el papel esencial que, en este tipo de empresas, juegan la tirada y la publicidad, se realiza un análisis previo de estas variables en función de la información histórica de los últimos catorce años.

La simulación planteada se centrará en torno a algunas relaciones lógicas de cálculo tales como:

- La tirada efectiva (circulación) del periódico (variable CIRCUL) se considera dependerá, básicamente, en una primera aproximación, de los hábitos de lectura de la población de la zona de difusión del periódico, medida por la circulación total de prensa diaria (variable WCIRCUL):

$$\text{CIRCUL} = f(\text{WCIRCUL})$$

- A su vez, la circulación total (WCIRCUL) se hace depender del desarrollo económico del país, medido por el PIB a precios de mercado en unidades monetarias constantes.

$$\text{WCIRCUL} = f(\text{PIB})$$

Naturalmente, la simplicidad de este planteamiento es evidente. Por una parte, la cuota de participación del periódico analizado puede verse afectada por su propia estrategia comercial y la de sus competidores o incluso por el nacimiento de nuevos periódicos en su zona geográfica de influencia. Por otra parte, la circulación total no tiene que evolucionar sólo en razón de la prosperidad económica del país. Sin embargo, el planteamiento se considera como una primera aproximación suficientemente orientativa.

- Con relación al montante total de publicidad (COLUM) a conseguir, se ha planteado una ecuación en que ésta (medida en número de columnas) se hace depender de la circulación del periódico y del PIB del país:

$$\text{COLUM} = f(\text{CIRCUL}, \text{PIB})$$

La primera variable explicativa intenta recoger el atractivo publicitario del periódico en función de su circulación; la segunda, la propia evolución del esfuerzo publicitario de los anunciantes a través del crecimiento económico.

**¡ATENCIÓN! EN UNA HOJA EXCEL, COPIE LOS SIGUIENTES DATOS Y SIGA LOS PASOS**

Años	CIRCUL	COLUM	WCIRCUL	PBI
1	1240	51	5110	30524
2	1340	52	5480	31322
3	1350	51	5550	32324
4	1360	51	5610	34148
5	1370	62	5650	35910
6	1390	64	5740	37611
7	1410	63	5810	39018
8	1420	63	5970	39903
9	1440	78	6100	40177
10	1500	83	6240	39710
11	1670	88	6480	40604
12	1690	93	7310	41707
13	1720	97	7400	42715
14	1750	101	7520	44224

Para mayor simplicidad se han estimado relaciones lineales utilizando Excel, que será el programa de ordenador donde se simularán los resultados globales del análisis. Para ello, seleccionamos el rango en el que queremos que aparezca el resultado de la regresión con el ratón o con la tecla de mayúsculas y el cursor, con una extensión de cinco filas y tantas columnas como variables explicativas incluya la regresión (contando también el término independiente). A continuación, sin pulsar la tecla **ENTER** de forma que el rango seleccionado aparezca en color oscuro, escribimos la siguiente fórmula:

**=ESTIMACION.LINEAL(RANGO Y;RANGO X;VERDADERO;VERDADERO)**

En *rango y* indicamos la posición de los datos de la variable dependiente; en *rango x*, la ubicación de los datos de la/s variable/s explicativa/s; *verdadero* indica que queremos que se incluya el término independiente y el segundo *verdadero* se especifica para que en el resultado se incluyan los principales estadísticos. Para finalizar pulsamos simultáneamente las teclas

**MAYÚSCULAS + CTRL + ENTER**

Aparecerán los resultados de la regresión en el rango de salida seleccionado al principio.

En la siguiente ilustración mostramos el orden en que se presentan en Excel los resultados de la regresión:

Coefficiente $x_k$	Coefficiente $x_{k-1}$	...	Coefficiente $x_2$	Coefficiente $x_1$	Constante o término independiente
Desviación Típica coef. $x_k$	Desviación Típica coef. $x_{k-1}$	...	Desviación Típica coef. $x_2$	Desviación Típica coef. $x_1$	Desviación típica constante
Coefficiente $R^2$	Desviación típica y estimada				
Estadístico F	Grados de libertad				
Suma de regresión de cuadrados	Suma residual de cuadrados				

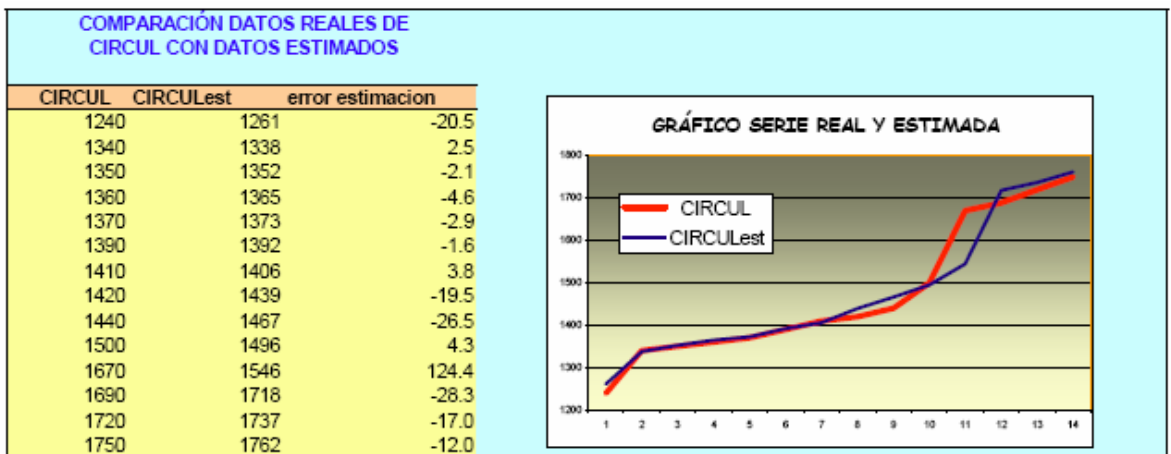
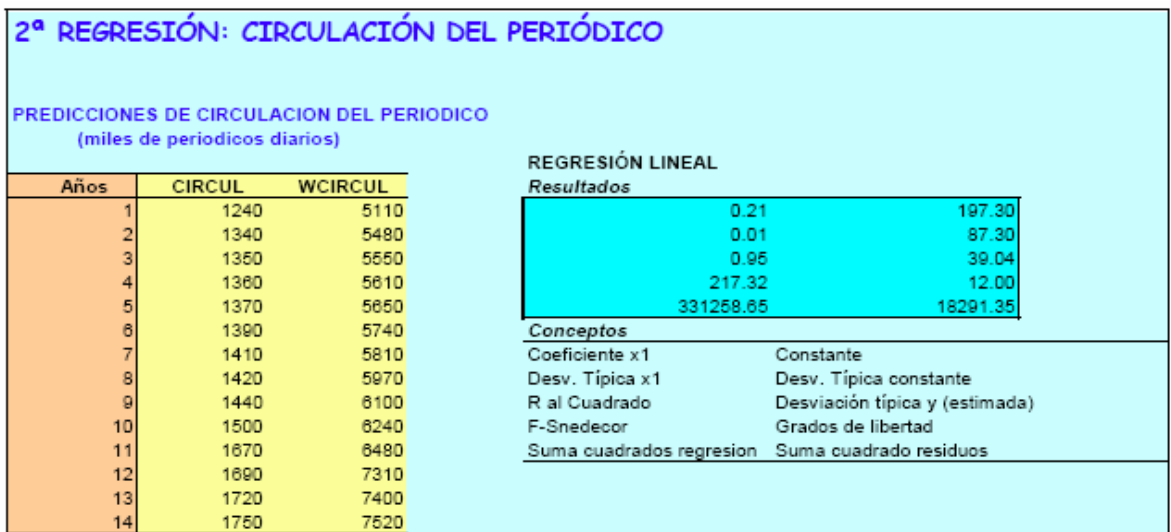
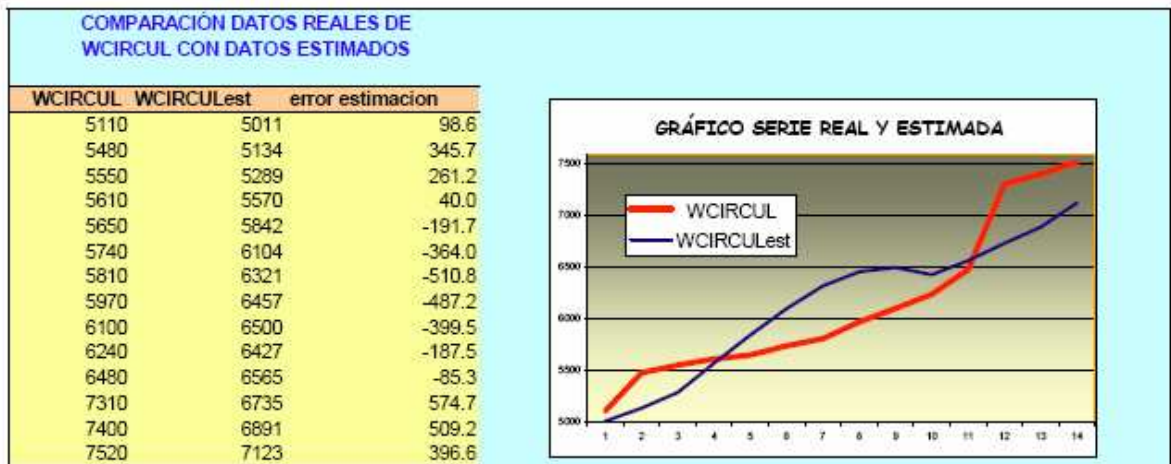
Dentro del libro de EXCEL que recoge el Caso de Aplicación deberemos tener una hoja que contiene los correspondientes resultados de las regresiones planteadas. En cada regresión comprobamos que los signos que acompañan a los coeficientes de las variables son coherentes con lo esperado (por ejemplo, un coeficiente positivo para la variable PIB indica que una mejora en la economía incentiva una mayor circulación de periódicos). También comprobamos que el ajuste realizado presenta una bondad adecuada, es decir, un coeficiente de determinación ( $R^2$ ) al menos superior a 0,7.

1ª REGRESIÓN: CIRCULACIÓN TOTAL		
PREDICIONES DE CIRCULACION TOTAL (miles de periodicos diarios)		
Años	WCIRCUL	PIB
1	5110	30524
2	5480	31322
3	5550	32324
4	5610	34148
5	5650	35910
6	5740	37611
7	5810	39018
8	5970	39903
9	6100	40177
10	6240	39710
11	6480	40604
12	7310	41707
13	7400	42715
14	7520	44224

REGRESIÓN LINEAL		
Resultados		
	0.15	305.79
	0.02	945.21
	0.76	388.96
	38.57	12.00
	5835978.08	1815514.77
Conceptos		
Coefficiente $x_1$	Constante	
Desv. Típica $x_1$	Desv. Típica constante	
R al Cuadrado	Desviación típica y (estimada)	
F-Snedecor	Grados de libertad	
Suma cuadrados regresion	Suma cuadrado residuos	

Con la correspondiente comparación de los datos reales de WCIRCUL y los obtenidos con la regresión (WCIRCULest).



3ª REGRESIÓN: COLUMNAS DE PUBLICIDAD			
PREDICCIONES DE COLUMNAS DE PUBLICIDAD (número de columnas)			
Años	COLUM	CIRCUL	PIB
1	51	1240	30524
2	52	1340	31322
3	51	1350	32324
4	51	1360	34148
5	62	1370	35910
6	64	1390	37611
7	63	1410	39018
8	63	1420	39903
9	78	1440	40177
10	83	1500	39710
11	88	1670	40604
12	93	1690	41707
13	97	1720	42715
14	101	1750	44224

REGRESIÓN LINEAL		
Resultados		
	0.00	0.08
	0.00	0.02
	0.94	5.03
	80.51	11.00
	4081.53	278.83
		#N/A
		#N/A
		#N/A

Conceptos		
Coefficiente x2	Coefficiente x1	Constante
Desv. Típica x2	Desv. Típica x1	Desv. Típica con
R al Cuadrado	Desviación típica y (estimada)	
F-Snedecor	Grados de libertad	
Suma cuadrados regresion	Suma cuadrado residuos	

COMPARACIÓN DATOS REALES DE COLUM CON DATOS ESTIMADOS		
COLUM	COLUMest	error estimacion
51	44	7.4
52	52	-0.3
51	54	-3.4
51	58	-6.6
62	61	1.4
64	64	-0.4
63	68	-4.8
63	70	-6.7
78	72	6.4
83	76	7.4
88	90	-1.8
93	93	0.3
97	96	0.7
101	101	0.4

A efectos de predicción, en la hoja de cálculo se recogen los valores supuestos para el PIB en el año actual (se considera que éste es 1998) y los tres del horizonte de simulación considerado (1999-2001). En función de los parámetros anteriormente estimados en las regresiones se calculan

$$WCIRCUL = 305,79 + 0,15 * PIB$$

$$CIRCUL = 197,30 + 0,21 * WCIRCUL$$

$$COLUM = -91,30 + 0,08 * CIRCUL + 0,001 * PIB$$

Predicciones Previas	1998	1999	2000	2001
Producto Interior Bruto (PBI)	45891.26	47405.74	48675.29	50083.26
Circulación Total de Periódicos (WCIRCUL)	7380.37	7613.84	7809.56	8026.61
Circulación del Periódico (CIRCUL)	1732.94	1781.51	1822.24	1867.40
Columnas de Publicidad (COLUM)	101.49	107.20	111.98	117.28

A partir de estos resultados previos y utilizando una hoja de cálculo del mismo libro Excel con que viene trabajando, resuelva:

- a) calcule una primera cuenta de pérdidas y ganancias, de acuerdo con las siguientes relaciones para 1999-2001. Para 1998 incluya los datos conocidos del último ejercicio:

<b>CUENTA DE PÉRDIDAS Y GANANCIAS (miles de pesos)</b>	<b>1998</b>
<b>INGRESOS</b>	
INGRESOS PUBLICIDAD	65,97
INGRESOS CIRCULACION	784,15
OTROS INGRESOS	61,58
<b>TOTAL INGRESOS</b>	<b>911,70</b>
<b>GASTOS</b>	
GASTOS DE PERSONAL	41,18
GASTOS DIVERSOS DE EXPLOTACION	13,70
GASTOS DE PAPEL	301,11
OTROS MATERIALES	3,99
COMISIONES Y PRIMAS	10,55
AMORTIZACIONES	100,00
<b>TOTAL DE GASTOS</b>	<b>470,53</b>
<b>RESULTADO SOLUCIÓN BÁSICA</b>	<b>441,17</b>

#### **INGRESOS**

- *Ingresos por publicidad*: número de columnas de publicidad, por su precio actual (650.00 pesos), con una tasa de variación anual de tarifas del 10 por 100.
- *Ingresos por circulación*: número de periódicos vendidos (circulación media \* 362 días anuales con prensa) por precio actual (1,25 pesos), con una tasa de variación anual de precios del 7 por 100.
- Otros Ingresos: Incremento anual del 8%.

#### **GASTOS**

- *Gastos de personal*: con un incremento anual del 10 por 100.
- *Gastos diversos de explotación*: con un incremento del 5 por 100.
- *Gastos de papel*: número diario de periódicos vendidos por 362 días, por 60 páginas de extensión media, por su peso en papel (20 gr. / pág.) incluidas mermas y periódicos no vendidos, por el precio actual del kilogramo de papel-prensa (0.4 pesos), actualizado a futuro con un incremento del 12 por 100 anual, basado en estudios especiales sobre el mercado del papel.
- *Gastos en otros materiales*: crecimiento anual del 10 por 100.
- *Comisiones y primas*: estimadas en un 16 por 100 de la cifra facturada en publicidad.
- *Amortizaciones*: calculadas según un informe especial sobre el tema en 1,20, 1,35 y 1,55 millones, respectivamente, para los años de predicción.

b) Exprese los resultados en porcentajes de variación, de un año respecto del año anterior.

c) Considere ahora la siguiente hipótesis alternativa:

- Un mayor crecimiento de las columnas de publicidad, basado en una reestructuración del departamento correspondiente:

Columnas de publicidad			
	1999	2000	2001
Solución básica	107	112	117
Solución alternativa	110	120	135

- Consecuencia de esta nueva estrategia, se supone un aumento de la comisión por publicidad del 16 al 18 por 100.
- Un plan especial de jubilaciones anticipadas y no renovación de vacantes, que según un informe especial del departamento de personal supondría reducir el crecimiento de los gastos de un 10 por 100 a sólo un 7 por 100 acumulativo.

Calcule la nueva cuenta de pérdidas y ganancias, aceptada la viabilidad de estos supuestos.

d) Compare los resultados con la solución básica.

e) Grafique las dos alternativas

## C.2. El método Delfos

### Resumen:

- ✚ El método Delfos (*Delphi*, en inglés) es un método diseñado para obtener y tratar la información de expertos consultados sobre un asunto determinado. Inicialmente pensado para estrategia militar, lo cierto es que es de aplicación a múltiples campos, entre ellos la economía.
  - ✚ Como técnica especial de encuesta a expertos presenta las siguientes características o rasgos distintivos: 1) Selección subjetiva de participantes; 2) Anonimato de las respuestas; 3) Retroacción controlada; 4) Preguntas con respuesta cuantitativa; y 5) Respuesta estadística de grupo.
  - ✚ Respecto al Delfos planteado en su forma convencional, existen otras variantes como son: Delfos en teleconferencia; Delfos con autoevaluación de expertos; Delfos con información cualitativa; Impactos cruzados.
  - ✚ Algunos ejemplos sobre Delfos nos pueden ayudar a comprender la técnica. Además desarrollaremos un Caso de Aplicación.
-

### **Características de la técnica Delfos**

- a) Selección subjetiva de participantes en función de su dominio del tema objeto de estudio. El número de participante suele ser reducido: entre 10 y 30 es lo más frecuente.
- b) Anonimato de las respuestas, por lo que ningún miembro del grupo conoce las respuestas individuales del resto, a fin de evitar cuestiones de liderazgo o condicionamientos.
- c) Retroacción controlada, con lo que la respuesta inicial de los expertos puede ser corregida en posteriores rondas según los resultados globales del grupo en la ronda precedente. Lo habitual es trabajar sólo a dos rondas: la inicial y una posterior revisión.
- d) Preguntas con respuesta cuantitativa, sean éstas un valor, una probabilidad de ocurrencia, una fecha de realización, etc.
- e) Respuesta estadística de grupo, consistente en una indicación de respuesta media y dispersión del grupo, en cada una de las rondas. Las medidas estadísticas utilizadas son la mediana (respuesta que deja igual número de expertos con opiniones por encima y debajo de la misma) y el recorrido intercuartílico (diferencia entre la respuesta que deja sólo un cuarto de expertos por encima y la que deja un cuarto por debajo).

### **Variantes del Delfos convencional**

- 1) Delfos en teleconferencia, en el que el monitor o responsable de la experiencia es sustituido por un ordenador programado para compilar los resultados del grupo y transmitirlos a los participantes en tiempo real.
- 2) Delfos con autoevaluación de expertos, en el que a los participantes se les pide una nota sobre su dominio del tema objeto del estudio. La nota puede utilizarse a efectos de ponderar respuestas o eliminar las de ciertos expertos poco seguros de sus valoraciones.
- 3) Delfos con información cualitativa, en el que, junto a respuestas numéricas, se incluyen otras valoraciones cualitativas o incluso preguntas abiertas. Es frecuente utilizar esta opción en una segunda o tercera ronda, para que aquellos expertos que se sitúan fuera del recorrido intercuartílico puedan exponer las razones de su discrepancia con la opinión más generalizada, y esta información puede ser transmitida a los restantes expertos para su consideración.
- 4) Impactos cruzados, en el que se consideran los condicionamientos relativos entre las probabilidades asignadas a sucesos que no siempre son independientes y que, sin embargo, pueden ser incongruentes en la valoración de determinados expertos.



### ***Algunos ejemplos de aplicación del método Delfos***

1) *Diseño de una política de futuro sobre consumo de drogas*: Aplicación realizada en 1974 en EE.UU. Participaron 38 expertos (investigadores, administradores de programas conexos con el consumo de drogas y políticos), a doble ronda con autoevaluación de expertos. Entre las cuestiones planteadas, los dos grupos más importantes eran los de objetivos y políticas. En las preguntas sobre objetivos se pedía una nota de 1 a 5 sobre posibilidad de alcanzarlos y relevancia. Con respecto a políticas, se pedía una valoración sobre si debería o no aplicarse cada una de las medidas de una lista previamente elaborada, en un horizonte a tres años.

2) *Competitividad de una región*. Estudio patrocinado por el Gobierno de la Provincia de Santa Fe y el CFI sobre competitividad y perspectivas de cinco ciudades seleccionadas. Realizado en 2007 se logró la "Estimación del Indicador de Competitividad para Algunas Ciudades Importantes de la Provincia de Santa Fe: la aplicación a Santa Fe, Rafaela, Avellaneda, Reconquista y Venado Tuerto". Las preguntas planteaban cuestiones a contestar en términos de probabilidad, porcentaje, orden de prioridad o nota de pertinencia, siguiendo el modelo WEF para la selección de factores.

3) *Futuro de una gran ciudad*. Delfos realizado en grandes ciudades como París o Madrid. La aplicación española, patrocinada por la Cámara de Comercio e Industria de Madrid a finales de 1988, plantaba diversas cuestiones en un horizonte año 2000. Los expertos fueron seleccionados entre urbanistas, economistas, sociólogos, etc., de la administración pública central o local, centros privados y Universidad. Se pedía autoevaluación para cada grupo de preguntas. Las cuestiones planteadas hacían referencia a valoración de problemas, importancia de factores para la localización de empresas u órganos de la Administración, estructura porcentual de la producción por sectores en el año 2000, crecimientos medios por actividades, alternativas de inversión, nota de calificación sobre coordinación de políticas, etc.

### **CASO DE APLICACIÓN 4; Una aplicación DELFOS al escenario socio-económico.**

A fin de asignar valores a las variables exógenas de un modelo econométrico de predicción de la economía española a medio plazo, parece aconsejable disponer de un escenario socioeconómico que sirva como punto de referencia básico. A tal efecto, periódicamente, en principio cada año, se revisan las probabilidades asignadas por un panel de expertos (economistas y empresarios de gran prestigio) a una lista de acontecimientos, fija su mayor parte, aunque con la posible incorporación de nuevos temas de actualidad. El caso ha sido diseñado con dos facetas distintas. Inicialmente, se presenta la experiencia tal como podría conocerla uno de los expertos participantes. Posteriormente, se detallan los resultados, en la forma en que son manejados por el responsable del DELFOS.

Del cuestionario, enviado por escrito un panel previamente contactado, con una carta e instrucciones explicativas sobre su cumplimentación, hemos

seleccionado siete acontecimientos, de los cuales se solicita la probabilidad de ocurrencia, valor éste comprendido entre 0 y 1. Al terminar de asignar las correspondientes probabilidades, el experto (si se trabaja con un ordenador en tiempo real) conoce la respuesta estadística del grupo. Si no, recibirá posteriormente esta información una vez tratada por el responsable DELFOS.

<b>DELFO S SOBRE ESCENARIO 2000</b>	
<b>Ronda: 1ª</b>	<b>Experto: 13</b>
A.1 <i>¿Qué probabilidad asigna a una reducción a la baja del barril de petróleo, para situarse por debajo de los 10\$ en 2000?</i>	
PROBABILIDAD =	<input type="text" value="0.2"/> (valor entre 0 y 1)
A.2 <i>¿Qué probabilidad asigna a un agudizamiento de los problemas de deuda exterior en los países latinoamericanos y africanos, poniendo en peligro el sistema monetario internacional?</i>	
PROBABILIDAD =	<input type="text" value="0.3"/>
A.3 <i>¿Qué probabilidad asigna a una depreciación del dólar frente al euro y con horizonte 2000?</i>	
PROBABILIDAD =	<input type="text" value="0.1"/>
A.4 <i>¿Qué probabilidad concede a un aumento del proteccionismo en los países desarrollados, como consecuencia del aumento de la competitividad y desequilibrios internacionales?</i>	
PROBABILIDAD =	<input type="text" value="0.4"/>
A.5 <i>¿Qué probabilidad asigna a la existencia real de una zona europea ampliada después del 2000?</i>	
PROBABILIDAD =	<input type="text" value="0.7"/>
A.6 <i>¿Qué probabilidad piensa que tiene un gobierno con mayoría absoluta popular en la próxima legislatura?</i>	
PROBABILIDAD =	<input type="text" value="0.5"/>
A.7 <i>¿Qué probabilidad asigna a una reducción significativa de las cuotas empresariales a la Seguridad Social antes del 2000?</i>	
PROBABILIDAD =	<input type="text" value="0.5"/>

1ª RONDA		OPINION DE CONSENSO DE LOS EXPERTOS						
	A.1	A.2	A.3	A.4	A.5	A.6	A.7	
Mediana	0.2	0.3	0.1	0.4	0.7	0.5	0.6	
Q1/4	0.1	0.2	0.1	0.3	0.5	0.5	0.5	
Q3/4	0.3	0.4	0.2	0.5	0.8	0.6	0.7	
Q3/4-Q1/4	0.2	0.2	0.1	0.2	0.3	0.1	0.2	

DEL FOS SOBRE ESCENARIO 2000	
<b>Ronda: 2ª</b>	<b>Experto: 13</b>
<p>A.1 ¿Qué probabilidad asigna a una reducción a la baja del barril de petróleo, para situarse por debajo de los 10\$ en 2000?</p> <p>PROBABILIDAD = <input type="text" value="0.3"/> (valor entre 0 y 1)</p>	
<p>A.2 ¿Qué probabilidad asigna a un agudizamiento de los problemas de deuda exterior en los países latinoamericanos y africanos, poniendo en peligro el sistema monetario internacional?</p> <p>PROBABILIDAD = <input type="text" value="0.3"/></p>	
<p>A.3 ¿Qué probabilidad asigna a una depreciación del dólar frente al euro y con horizonte 2000?</p> <p>PROBABILIDAD = <input type="text" value="0.1"/></p>	
<p>A.4 ¿Qué probabilidad concede a un aumento del proteccionismo en los países desarrollados, como consecuencia del aumento de la competitividad y desequilibrios internacionales?</p> <p>PROBABILIDAD = <input type="text" value="0.5"/></p>	
<p>A.5 ¿Qué probabilidad asigna a la existencia real de una zona europea ampliada después del 2000?</p> <p>PROBABILIDAD = <input type="text" value="0.7"/></p>	
<p>A.6 ¿Qué probabilidad piensa que tiene un gobierno con mayoría absoluta popular en la próxima legislatura?</p> <p>PROBABILIDAD = <input type="text" value="0.5"/></p>	
<p>A.7 ¿Qué probabilidad asigna a una reducción significativa de las cuotas empresariales a la Seguridad Social antes del 2000?</p> <p>PROBABILIDAD = <input type="text" value="0.6"/></p>	

En nuestro caso, por ejemplo, conocemos que para el acontecimiento A.1, 0.2 es la probabilidad mediana; 0.1 y 0.3, el primer y tercer cuartil, y 0.2 el recorrido intercuartílico (diferencia entre el tercer y el primer cuartil). Con los trece expertos consultados significa que seis opinan que la probabilidad de una bajada del precio del barril de petróleo por debajo de los 10 dólares es menor del 20%, y otros 6 expertos coincidieron en que es mayor (realmente en ambos extremos puede haber otras contestaciones también iguales a 0.2). Tan sólo 3 expertos consideran una probabilidad inferior al 10% y sólo otros 3 superiores al 30%.

La dispersión de opiniones, medida por el recorrido intercuartílico, muestra una coincidencia muy acusada de respuestas en los sucesos A.3 y A.6 y la mayor discrepancia con relación al acontecimiento A.5.

El experto, al conocer ahora la opinión del grupo, después de esta primera ronda de respuestas, habitualmente revisa alguna de sus estimaciones más discrepantes; en principio, aquellas en las que tiene menos seguridad. Supongamos que, en una segunda ronda de respuestas, nuestro experto número 13 eleva ligeramente su probabilidad de ocurrencia de los sucesos A.1 (de 0.2 a 0.3), A.4 (de 0.4 a 0.5) y A.7 (de 0.5 a 0.6).

Al igual que antes, al terminar la segunda ronda de respuestas es posible conocer la respuesta estadística del grupo:

2º RONDA		OPINION DE CONSENSO DE LOS EXPERTOS						
Mediana		0.2	0.3	0.1	0.4	0.7	0.5	0.6
Q1/4		0.2	0.2	0.1	0.3	0.5	0.5	0.5
Q3/4		0.3	0.4	0.2	0.5	0.8	0.6	0.7
Q3/4-Q1/4		0.1	0.2	0.1	0.2	0.3	0.1	0.2

El alto grado de acuerdo general de los expertos ya en la primera ronda ha provocado que, en la segunda, presenten sólo muy pequeños cambios. Únicamente en A.1 se ha reducido la dispersión de respuestas, medida por el recorrido intercuartílico.

Hasta aquí hemos descrito la experiencia tal y como la percibe un participante. Por su parte, el responsable Delfos conoce la respuesta individualizada de cada experto. Puede observarse aquí como una corrección de valores extremos (realizada al pasar de la primera a la segunda ronda) no afecta a los resultados estadísticos, según venimos midiéndolos. Por ejemplo, en A.2 las respuestas se mueven, en la primera ronda, entre 0.1 y 0.7, pero sólo entre 0.2 y 0.6 en la segunda. Sin embargo, al ordenarlas de menor a mayor dan un mismo valor de mediana y cuartiles.

A.2													
1ª RONDA	0.1	0.1	0.2	0.2	0.2	0.3	0.3	0.3	0.3	0.4	0.5	0.6	0.7
2ª RONDA	0.2	0.2	0.2	0.2	0.3	0.3	0.3	0.3	0.3	0.4	0.5	0.6	0.6
				Q 1/4			Me			Q 3/4			

El detalle de las tablas de respuestas de cada una de las dos rondas se recoge en las dos tablas siguientes:

1ª RONDA												
RESPUESTA DE LOS EXPERTOS												
	A.1	A.2	A.3	A.4	A.5	A.6	A.7					
EXPERTO 1	0.1	0.3	0.0	0.5	0.7	0.5	0.7					
EXPERTO 2	0.2	0.2	0.1	0.2	0.9	0.4	0.7					
EXPERTO 3	0.3	0.5	0.1	0.6	0.8	0.6	0.6					
EXPERTO 4	0.2	0.6	0.3	0.7	0.5	0.5	0.8					
EXPERTO 5	0.5	0.7	0.4	0.3	0.8	0.5	0.7					
EXPERTO 6	0.1	0.4	0.2	0.2	0.4	0.7	0.5					
EXPERTO 7	0.1	0.3	0.1	0.5	0.6	0.5	0.4					
EXPERTO 8	0.4	0.3	0.1	0.4	0.8	0.3	0.5					
EXPERTO 9	0.2	0.2	0.0	0.4	0.7	0.6	0.7					
EXPERTO 10	0.3	0.1	0.1	0.3	0.4	0.6	0.3					
EXPERTO 11	0.2	0.1	0.2	0.6	0.4	0.4	0.6					
EXPERTO 12	0.0	0.2	0.2	0.2	0.6	0.6	0.6					
EXPERTO 13	0.2	0.3	0.1	0.4	0.7	0.5	0.5					

2ª RONDA												
RESPUESTA DE LOS EXPERTOS												
	A.1	A.2	A.3	A.4	A.5	A.6	A.7					
EXPERTO 1	0.2	0.3	0.1	0.5	0.7	0.5	0.7					
EXPERTO 2	0.2	0.3	0.1	0.3	0.8	0.6	0.7					
EXPERTO 3	0.3	0.5	0.1	0.6	0.8	0.6	0.6					
EXPERTO 4	0.2	0.6	0.3	0.7	0.5	0.5	0.7					
EXPERTO 5	0.3	0.6	0.3	0.3	0.8	0.5	0.7					
EXPERTO 6	0.2	0.4	0.2	0.3	0.5	0.6	0.5					
EXPERTO 7	0.1	0.3	0.1	0.5	0.6	0.5	0.5					
EXPERTO 8	0.4	0.3	0.1	0.4	0.8	0.3	0.5					
EXPERTO 9	0.2	0.2	0.1	0.4	0.7	0.6	0.7					
EXPERTO 10	0.3	0.2	0.1	0.4	0.4	0.6	0.5					
EXPERTO 11	0.2	0.2	0.2	0.6	0.5	0.4	0.6					
EXPERTO 12	0.2	0.2	0.2	0.3	0.6	0.6	0.6					
EXPERTO 13	0.3	0.3	0.1	0.5	0.7	0.5	0.6					

Después de las tablas de respuestas, se muestran éstas ordenadas de mayor a mayor para facilitar el cálculo de los estadísticos mencionados.

1ª RONDA								
RESPUESTA DE LOS EXPERTOS ORDENADAS								
	A.1	A.2	A.3	A.4	A.5	A.6	A.7	
	0.0	0.1	0.0	0.2	0.4	0.3	0.3	0.3
	0.1	0.1	0.0	0.2	0.4	0.4	0.4	0.4
	0.1	0.2	0.1	0.2	0.4	0.4	0.5	0.5
	0.1	0.2	0.1	0.3	0.5	0.5	0.5	0.5
	0.2	0.2	0.1	0.3	0.6	0.5	0.5	0.5
	0.2	0.3	0.1	0.4	0.6	0.5	0.6	0.6
	0.2	0.3	0.1	0.4	0.7	0.5	0.6	0.6
	0.2	0.3	0.1	0.4	0.7	0.5	0.6	0.6
	0.2	0.3	0.2	0.5	0.7	0.6	0.7	0.7
	0.3	0.4	0.2	0.5	0.8	0.6	0.7	0.7
	0.3	0.5	0.2	0.6	0.8	0.6	0.7	0.7
	0.4	0.6	0.3	0.6	0.8	0.6	0.7	0.7
	0.5	0.7	0.4	0.7	0.9	0.7	0.8	0.8

2ª RONDA								
RESPUESTA DE LOS EXPERTOS ORDENADAS								
	A.1	A.2	A.3	A.4	A.5	A.6	A.7	
	0.1	0.2	0.1	0.3	0.4	0.3	0.5	0.5
	0.2	0.2	0.1	0.3	0.5	0.4	0.5	0.5
	0.2	0.2	0.1	0.3	0.5	0.5	0.5	0.5
	0.2	0.2	0.1	0.3	0.5	0.5	0.5	0.5
	0.2	0.3	0.1	0.4	0.6	0.5	0.6	0.6
	0.2	0.3	0.1	0.4	0.6	0.5	0.6	0.6
	0.2	0.3	0.1	0.4	0.7	0.5	0.6	0.6
	0.2	0.3	0.1	0.5	0.7	0.6	0.6	0.6
	0.3	0.3	0.2	0.5	0.7	0.6	0.7	0.7
	0.3	0.4	0.2	0.5	0.8	0.6	0.7	0.7
	0.3	0.5	0.2	0.6	0.8	0.6	0.7	0.7
	0.3	0.6	0.3	0.6	0.8	0.6	0.7	0.7
	0.4	0.6	0.3	0.7	0.8	0.6	0.7	0.7

El caso puede utilizarse cambiando las probabilidades asignadas, en la primera parte, por el experto 13 o bien en la tabla de respuesta de expertos (sin ordenar). Sin embargo, la tabla de respuestas ordenadas siempre deberá recalcularse de nuevo para el cómputo adecuado de los estadísticos mediana y cuartiles. Para ello, se utiliza la instrucción del menú principal de EXCEL: DATOS / ORDENAR, donde se define el rango a ordenar como los doce valores numéricos de una columna, con la opción de definir una segunda y tercera clave de ordenación, eligiendo como orden de clasificación (A), ascendente.

De acuerdo a los resultados de Delfos, puede definirse el escenario más probable (sin tener en cuenta, en este caso, los impactos cruzados), como aquel compuesto por acontecimientos con probabilidad de ocurrencia superior al 50%. En nuestra aplicación, el escenario sería (0, 0, 0, 0, 1, 0, 1); es decir, la supuesta no ocurrencia de A.1 a A.4 y A.6, así como la posible ocurrencia de A.5 y A.7.

ESCENARIO MAS PROBABLE PARA EL AÑO 2000						
A.1	A.2	A.3	A.4	A.5	A.6	A.7
0	0	0	0	1	0	1
<ul style="list-style-type: none"> <li>* NO REDUCCION A LA BAJA DEL BARRIL DE PETROLEO</li> <li>* NO AGUDIZAMIENTO MAYOR DE LOS PROBLEMAS DE DEUDA EXTERIOR</li> <li>* NO APRECIACION DEL DOLAR FRENTE AL EURO</li> <li>* NO AUMENTO DEL PROTECCIONISMO</li> <li>* POSIBLE ZONA EURO AMPLIADA</li> <li>* GOBIERNO SIN MAYORIA ABSOLUTA POPULAR</li> <li>* POSIBLE REDUCCION DE LAS CUOTAS EMPRESARIALES A LA SEGURIDAD SOCIAL</li> </ul>						

Tal y como ha sido definida la hoja de cálculo, el resultado del escenario más probable ha sido introducido directamente en función de las probabilidades calculadas. Sin embargo, el programa podría generalizarse para que este resultado se mostrase automáticamente en pantalla, dependiendo de que las probabilidades medianas fueran superiores o no a una cantidad prefijada.

---



---

## CASOS DE ESTUDIO, PREGUNTAS Y PROBLEMAS

### **ACTIVIDAD 1: Visitar páginas Web de instituciones de predicción.**

Hemos comprobado la pluralidad de instituciones que realizan predicciones en el campo de la economía y de la gestión de empresas, cada una de las cuales está elaborada con criterios, planteamientos metodológicos y técnicas de predicción distintas. Además, la interpretación que cada institución hace de la realidad económica que le rodea es bien distinta, puesto que está basada en sus propias percepciones. Por ello, con frecuencia las distintas predicciones realizadas sobre la marcha de la economía son muy distintas. Sin embargo, todas ellas son de utilidad para el analista económico o empresarial.

Como actividad le sugerimos que realice una visita a las páginas Web de algunas de las instituciones que realizan predicciones, para alcanzar un doble objetivo: En primer lugar, es probable que, desde el punto de vista de su propia práctica profesional, en algún momento deba hacerse una idea del entorno económico actual y las perspectivas que sobre el mismo tengan las diversas instituciones. Para ello deberá consultar los Informes de Predicción del mayor número posible de instituciones. Hoy día son varios los centros de predicción que proveen esta información gratuitamente a través de Internet.

Pues bien, para tener habilidad en la recopilación de esta información, le sugerimos que visite las páginas Web de algunas de estas Instituciones. Localice el lugar concreto de la página en el que la Institución coloca sus informes de coyuntura y predicción (le sugerimos que agregue las direcciones en su carpeta de favoritos), y deténgase a analizar con detalle la estructura de los distintos informes: tipo de información, variables sobre las que se realiza predicción, forma de estructurarla... Observe también las diferencias entre las distintas instituciones en cuanto a la forma de presentar la información. Analice también los comentarios que los autores de los informes realizan a partir de la predicción, puesto que en su categoría de expertos seguidores

de la realidad económica sus comentarios dan pistas interesantes sobre las perspectivas de la economía.

En segundo lugar, y como complemento a lo anterior, le sugerimos que elabore un breve informe de situación y perspectivas a partir de la información contenida en los informes de estas instituciones. Realice su propia clasificación de la información, sus propias tablas agrupando los datos, realice tablas de discrepancias entre instituciones y, por último, sintetice la información con unos breves comentarios. Con todo ello, podrá tener una visión de conjunto de la situación económica actual y su propio informe de predicción.

### **ACTIVIDAD 2: Diseño de un proceso de predicción y formulación de las etapas de predicción o simulación.**

Imagine una situación (preferiblemente económica o empresarial) en la que pueda ser necesaria la obtención de predicciones. Adapte las etapas genéricas del proceso de predicción/simulación para ese caso hipotético, en forma análoga a lo realizado con nuestros 2 ejemplos sobre la inversión en infraestructura para reducir la contaminación, y la reestructuración del sector siderúrgico. Explique con detalle cómo abordaría cada una de las etapas, qué técnica le parecería más adecuada a la vista de la situación de predicción presentada, cómo obtendría la información...

### **TEST DE AUTOEVALUACIÓN**

- 1) En el proceso de toma de decisiones en Economía y Gestión de Empresas. Marque la casilla CORRECTA.
  - a) No se debe tener en cuenta la información actual o pasada, sino la futura.
  - b) La información pasada debe tenerse en cuenta, pero no debe ser la única base para la decisión.
  - c) La información pasada nos sirve sólo si queremos realizar una única predicción.
  - d) La revisión de las predicciones se realiza sólo en caso de disponer de predicciones alternativas.
- 2) Marque la casilla INCORRECTA.
  - a) La demografía o la ecología son campos en los que se utiliza predicción.
  - b) Las predicciones macroeconómicas son necesarias para abordar la predicción empresarial.
  - c) Aunque los problemas sean distintos, la predicción en Economía se aborda de forma conjunta.
  - d) La predicción sólo es utilizable en muy escasas y exclusivas áreas, entre las que destaca la Economía y gestión de empresas.
- 3) Marque la casilla INCORRECTA.
  - a) Los modelos ARIMA se encuadran en el enfoque de la predicción según el análisis aislado de series.
  - b) En el enfoque según las relaciones entre variables se incluyen técnicas como los modelos econométricos o los modelos de simulación determinista.
  - c) El enfoque de predicción según información subjetiva, incluye técnicas de relación causal entre variables.
  - d) Las técnicas Delfos se incluyen en el enfoque de información subjetiva.



- 4) ¿Cómo tipificaría los modelos econométricos uniecuacionales?
- Situaciones con historia, horizonte medio y medios normales.
  - Situaciones con historia, horizonte medio y medios altamente profesionalizados.
  - Situaciones sin historia, horizonte corto y medios escasos.
  - Encajan en varias tipificaciones.
- 5) La técnica conocida como alisado exponencial se tipifica dentro de las técnicas como...
- Situaciones con historia, medios sofisticados y horizonte corto de predicción.
  - Situaciones con historia, medios sofisticados y horizonte largo de predicción.
  - Situaciones con historia, medios escasos y horizonte largo de predicción.
  - Situaciones con historia, medios escasos y horizonte corto de predicción.
- 6) ¿Cuál de las siguientes situaciones NO es frecuente en la práctica de la predicción profesional?
- Las predicciones de distintas instituciones suelen presentar una discrepancia.
  - Las predicciones de las instituciones oficiales se valoran de forma distinta a las del resto de instituciones, atendiendo a su origen.
  - Las predicciones de la misma institución permanecen estables a lo largo del horizonte de predicción.
  - Los supuestos de la predicción se utilizan para justificar las diferencias entre instituciones.
- 7) ¿Cuál de la siguiente afirmación es CORRECTA?
- Las predicciones son neutrales, y no influye el origen de las mismas.
  - Sobre los datos económicos oficiales que utilizamos para la predicción hay que establecer cautelas o escepticismo moderado.
  - La unanimidad en las predicciones de distintas instituciones evidencia los riesgos que hay detrás de la predicción.
  - A la hora de elegir entre predicciones alternativas, es recomendable preferir las de las instituciones oficiales.
- 8) Señale cuál de las siguientes NO forma parte de las etapas genéricas del proceso de predicción/simulación.
- Planteamiento del problema en términos de predicción.
  - Elección de la técnica de predicción.
  - Adaptación de la técnica al problema de predicción
  - Revisión del proceso.
- 9) ¿Cuáles de las siguientes técnicas precisan de medios altamente especializados?
- Alisado exponencial.
  - X-11 ARIMA.
  - Input-output.
  - Modelos VAR.
- 10) Señale cual de los siguientes factores NO caracteriza la *situación de predicción*.
- Objetivo de predicción.
  - Situación de predicción.
  - Medios disponibles para la predicción.
  - Tipo de información.

**TEST DE AUTOEVALUACIÓN****Marque la casilla de la respuesta que considere correcta:**

1) Qué ocurre si aplicamos una media móvil a una serie sobre la que ya se aplicó el mismo procedimiento.

- a) Que indirectamente hemos aplicado el Alisado exponencial doble de Brown.
- b) Que indirectamente hemos aplicado el Alisado Holt-Winters con doble parámetro.
- c) Que hemos aplicado una doble media móvil.
- d) Que hemos eliminado la tendencia de la serie.

2) Señale la respuesta INCORRECTA

- a) Una serie económica se puede dividir en tendencia, ciclo, estacionalidad y componente irregular.
- b) Según el esquema mixto de descomposición, a efectos prácticos se considera que la serie queda reducida a tendencia y ciclo.
- c) El ciclo y la tendencia conforman los componentes del largo plazo de la serie económica.
- d) La estacionalidad se relaciona con los movimientos a corto plazo en la serie.

3) La tendencia...

- a) Se aísla con el método Census X-11-
- b) Se relaciona con los movimientos a corto plazo de la serie
- c) Se elimina con diferenciación adecuada.
- d) Se analiza conjuntamente con el componente estacional.

4) Señale la respuesta INCORRECTA sobre las medias móviles:

- a) El alisado exponencial doble de Brown es una técnica de media móvil.
- b) Se considera una técnica de alisado y *naïve*.
- c) Realiza predicciones óptimas para series con tendencia.
- d) Precisa de información histórica.

5) La expresión:

$$M_t = \alpha \cdot y_t + \alpha \cdot (1-\alpha)^1 \cdot y_{t-1} + \alpha \cdot (1-\alpha)^2 \cdot y_{t-2} + \dots$$

Se corresponde con:

- a) Una doble media móvil.
- b) Un alisado simple.
- c) Un alisado Holt-Winters con un parámetro.
- d) Un ajuste exponencial de funciones matemáticas.

6) Señale la respuesta INCORRECTA en relación con el ajuste de funciones matemáticas:

- a) Nos permite realizar predicción en series con tendencia.
- b) Se realiza una regresión de los datos sobre el tiempo.
- c) Constituye el primer paso del Alisado de Holt-Winters con triple parámetro.
- d) Nos permite obtener una expresión de la tendencia de una serie.

7) Deseamos predecir una serie con estacionalidad. Aplicamos el método de relación a la media móvil. ¿Es suficiente?

- a) No. Ahora debemos aplicar la Técnica de Holt-Winters con triple parámetro.
- b) No. Debemos predecir la tendencia con cualquiera de los métodos adecuados para ello, y afectar por el coeficiente de estacionalidad.
- c) Sí. EViews realiza una predicción automática a partir de este método.
- d) Sí. Porque la serie ya está desestacionalizada, que es lo que introduce dificultades para predecir.

8) Señale la respuesta CORRECTA en relación con la predicción de una serie con estacionalidad:

- a) Se puede aplicar directamente el alisado doble de Brown con triple parámetro.
- b) Se puede aplicar el alisado de Holt-Winters con triple parámetro, y se afecta el resultado por los coeficientes de estacionalidad.
- c) Se puede aplicar un ajuste de funciones matemáticas, previa desestacionalización y posterior afectación del índice de estacionalidad.
- d) Bastaría con aplicar directamente el método Census X-11.

9) Realizamos un Alisado exponencial doble de Brown, con el parámetro alpha óptimo elegido por el programa estadístico. Este alpha ha resultado ser elevado, lo cual es señal de:

- a) Una acusada tendencia.
- b) Una acusada estacionalidad
- c) Una serie muy errática.
- d) Ninguna de las anteriores. El coeficiente alpha sólo determina el número de iteraciones de la técnica y se elige aleatoriamente.

10) Cuando no se dispone de información estadística (histórica) la predicción...

- a) No se puede llevar a cabo. Las técnicas de predicción precisan de información estadística.
- b) Se dificulta extraordinariamente, por lo que se recomienda recurrir a la generación de series estadísticas propias.
- c) Reviste el carácter de subjetivo, pero se puede llevar a cabo.
- d) Se borda con técnicas específicas que no necesitan de información histórica.

## LA SIMULACIÓN EN AUSENCIA DE INFORMACIÓN HISTÓRICA

### TEST DE AUTOEVALUACIÓN

Marque la casilla de la respuesta que considere correcta:

1) ¿Cuál de las siguientes NO es una variante del Delfos convencional?

- a) Sustitución del responsable de la experiencia por un procesado automático de las respuestas.

- b) Conocimiento de la autoría de cada respuesta.
- c) Valoración por cada experto de su grado de conocimiento del tema.
- d) Establecimiento de respuestas cualitativas o abiertas.

**2) Señale la respuesta INCORRECTA en relación con la simulación mediante fórmulas recursivas:**

- a) Permiten simular el funcionamiento de un sistema a través de la definición de las relaciones entre las variables que lo componen.
- b) Un modelo de ecuaciones autorregresivas es un caso de aplicación de fórmulas recursivas.
- c) La dinámica de sistemas constituye un sistema completo de simulación mediante ecuaciones.
- d) A pesar de lo ingenioso de su planteamiento, no ha gozado de una gran aplicabilidad por la dificultad de llevarlo a la práctica.

**3) Señale cuál de las siguientes NO es una característica del método Delfos.**

- a) Se encuesta a expertos de un determinado tema.
- b) La selección de los expertos es aleatoria.
- c) Los expertos no conocen el nombre de los autores de las respuestas.
- d) Los expertos disponen de la información estadística fundamental de las respuestas del conjunto de expertos.

**4) ¿Cuál de los siguientes NO es un elemento de la dinámica de sistemas?**

- a) Parámetros.
- b) Perturbación aleatoria.
- c) Operador retardo.
- d) Tasas.

---

---

### Bibliografía

- Pulido San Román, A. (2004) Curso combinado de Predicción y Simulación. [www.uam.es/predysim](http://www.uam.es/predysim). Universidad Autónoma de Madrid. Capítulo 1, 2 y 3

*Capítulo 11. ANÁLISIS EXPLORATORIO DE DATOS ..... 437*

11.1. Investigación estadística y Análisis exploratorio de datos.. 437

11.2 Análisis de Componentes Principales ..... 438

11.3 Análisis Factorial de correspondencias ..... 438

Origen ..... 438

Justificación ..... 438

Interpretación de los factores ..... 439

Ejemplo de tratamiento de la tabla múltiple: La encuesta de un diario regional ..... 439

1. *Análisis Simultáneo* ..... 441

2. *Análisis factorial de correspondencias múltiples de las secciones.* ..... 442

a. *Lectura de la tabla de datos* ..... 442

*Individuos*..... 445

*Modalidades*..... 446

*Un nuevo referencial de representación* ..... 447

*Centro de gravedad* ..... 447

*Inercia* ..... 447

*El cálculo de la inercia* ..... 449

b. *El plano de los dos primeros factores* ..... 450

*Las variables activas* ..... 450

*Grupos de secciones* ..... 453

*Las secciones mal representadas en el primer plano*..... 454

*Los individuos* ..... 456

*Las variables suplementarias: la caracterización* ..... 457

c. *Clasificación* ..... 458

*Clasificación jerárquica* ..... 459

*Coordenadas y valores test*..... 460

*Dendograma* ..... 460

d. *Partición de la nube de puntos*..... 461

*Descripción de la partición* ..... 463

e. *Tabla de Burt* ..... 466

3. *Análisis factorial de correspondencias múltiples de la caracterización*..... 470

a. *Las modalidades activas: la caracterización*..... 471

*Contradicción aparente entre los dos análisis* ..... 475

<i>Pérdida de las dispersiones específicas de los perfiles de lectura</i> .....	475
<i>b. Clasificación</i> .....	476
<i>c. Participación de la nube de puntos</i> .....	479
<i>d. Descripción de la partición</i> .....	480
4. <i>Análisis factorial de correspondencias múltiples con el conjunto de las secciones y de la caracterización</i> .....	482
<i>a. Lectura de la tabla de datos</i> .....	482
<i>Centro de gravedad</i> .....	485
<i>Inercia</i> .....	485
<i>El cálculo de la inercia</i> .....	486
<i>b. El plano de los dos primeros factores</i> .....	488
<i>Las variables activas</i> .....	488
<i>Grupos de secciones</i> .....	491
<i>Las modalidades mal representadas en el primer plano</i> .	491
<i>c. Clasificación jerárquica</i> .....	494
<i>Coordenadas y valores test</i> .....	494
<i>Dendograma</i> .....	495
<i>d. Partición de la nube de puntos</i> .....	496
<i>Descripción de la partición</i> .....	498
<i>e. Tabla de Burt</i> .....	500
<i>CASOS DE ESTUDIO, PREGUNTAS Y PROBLEMAS</i> .....	511
Caso 11.1: Perfil socioeconómico de lectores .....	511
Caso 11.2: Las razas de perros .....	511
<i>Bibliografía</i> .....	520

## Capítulo 11. ANÁLISIS EXPLORATORIO DE DATOS

### 11.1. Investigación estadística y Análisis exploratorio de datos

Cualquier investigador, al encontrarse frente a un trabajo empírico, debe seguir ciertos lineamientos lógicos en el desarrollo de su tarea. Esto lo conduce a seguir los pasos de una investigación:

- I. Definición de la Investigación
- II. Planteamiento de una Tabla de Datos
- III. Diseño de las Fuentes de Información
- IV. Recolección de Datos
- V. Análisis de la Información

Una vez realizados correctamente los pasos I, II, III y IV siguiendo una metodología de trabajo adecuada, si la investigación se basa en el estudio de grandes tablas de datos con muchos individuos y muchas características, para el de Análisis de la Información (paso V) se puede definir una Tabla de Datos (individuos x variables) que permita realizar análisis exploratorio.

¿Para qué se construye una Tabla "individuos x variables"?

- Para comparar todos los individuos con la finalidad de evaluar el grado de semejanza que existe entre ellos,
- Para evaluar el nivel de asociación que existe entre las características seleccionadas y observadas sobre el conjunto de individuos.

Si la tabla de datos contiene sólo variables cuantitativas, se debe realizar un "análisis de componentes principales"; mientras que si las variables son cualitativas, el análisis a realizar es el "factorial de correspondencias". Este puede ser "análisis factorial de correspondencias simple" (AFCS), o "análisis factorial de correspondencias múltiples" (AFCM).

Si sólo nos interesamos en dos atributos cualitativos (variables cualitativas) observados (es decir en dos columnas de la tabla de códigos condensados) se hace necesario el **análisis de la información** basado en una tabla de contingencia sobre la cual aplicaremos AFCS.

Si el interés se centra en el análisis de la totalidad de atributos observados se hace necesario el **análisis de la información** basado en una tabla de contingencia sobre la cual aplicaremos AFCM.

## 11.2 Análisis de Componentes Principales

El método de análisis de componentes principales (ACP) y la diagonalización de matrices se desarrollarán en Econometría.

## 11.3 Análisis Factorial de correspondencias

El análisis factorial es una técnica que se utiliza para el estudio y la interpretación de las correlaciones halladas entre un grupo de variables con el objeto de descubrir los posibles factores comunes a todas ellas.

### Origen

Se puede remontar al año 1904 en el que el estadístico Ch.Spearman publicó el primer trabajo sobre análisis factorial, aunque limitado a la búsqueda de un único factor común entre varias variables. Su extensión a varios factores se debe a L. Thurstone quien en 1941 publicó la obra "Factorial Studies on Intelligence". Esta técnica, que se suele inscribir dentro del conjunto de técnicas que forman parte del análisis multivariable, es bastante empleada en Psicología y Psicología Social. También es empleada en Sociología y en general en las ciencias sociales. De acuerdo con Blalock (1977) como fundamento del análisis factorial se encuentra "la idea de que si tenemos un gran número de índices o variables correlacionadas entre sí, estas relaciones recíprocas pueden deberse a la presencia de una o más variables o factores subyacentes relacionados en grado diverso con aquellos. Se supone que las interrelaciones altas dentro de un grupo de variables se deben a uno o varios factores o variables generales a las que representa el grupo. Precisamente uno de los objetivos principales del análisis factorial es identificar estos factores o variables comunes y por tanto más generales que los datos. Con ello frecuentemente reciben significación muchos conjuntos de correlaciones que, de otra manera, parecía que carecían de sentido.

### Justificación

El análisis factorial está sobre todo justificado, cuando, como dice Schuessler (1971), nos encontramos con un conjunto de intercorrelaciones significativas entre variables sociales y económicas muy distintas, por ejemplo: riqueza, empleo, desarrollo, divorcio, criminalidad, escolarización, etc. con referencia a ciudades o núcleos de población, que dan la impresión de constituir una mezcla de cifras sin un sentido coherente.

- oportunidad: utilizar el análisis factorial para determinar si dicho conjunto de intercorrelaciones aparentemente inconexas se deben a uno



o varios factores o variables no explícitas con los cuales las variables iniciales se hallan fuertemente correlacionadas.

- lógica: se puede asimilar al método de las concordancias de Stuart Mill. De acuerdo con este método, cuando se descubre un factor que es común a una serie de variables, que en otro caso no tendrían relación entre sí, tal factor se puede considerar causa común de la intercorrelación observada entre las variables dispares.

Asimismo, el análisis factorial es adecuado, en principio, para el estudio de las intercorrelaciones de variables cuantitativas. Sin embargo, también se admite su utilización respecto a variables cualitativas, de escala nominal u ordinal, siempre que su empleo pueda proporcionar una interpretación significativa de las correlaciones entre estas variables.

#### Interpretación de los factores

El análisis factorial revela estadísticamente los factores comunes más generales que unen las variables dadas en el análisis, pero no los identifica, es decir, no da su nombre, sino únicamente da, como es comprensible, un resultado numérico.

- Para identificar o determinar el nombre de estos factores es necesario interpretar estos resultados numéricos, lo que se realiza viendo qué variables iniciales de las correlacionadas tienen valoraciones altas en cada factor.
- También hay que tener en cuenta si estas valoraciones son positivas o negativas o ambas cosas a la vez.

El examen de las variables iniciales que tienen valoraciones altas en los factores y de su signo debe conducir a identificar y nombrar los factores comunes buscados, de acuerdo con estas características. En cierto modo es un problema de conjeturar el carácter del factor común descubierto y darle el nombre apropiado.

Por ejemplo, supongamos que mediante este análisis se descubre en una investigación sobre relaciones familiares que un factor tiene correlaciones positivas con varias variables referentes a los padres y otras varias negativas con variables relativas a los hijos, entonces a este factor se le podría llamar factor padres versus hijo.

#### Ejemplo de tratamiento de la tabla múltiple: La encuesta de un diario regional

Un diario regional realizó en 1999 una encuesta a 505 hogares en la región donde el diario tenía influencia.

- El objetivo general de esta encuesta era "conocer mejor a sus lectores" saber lo que se leía en el diario y responder a la pregunta: "¿Quién lee qué?".

Para ello fueron seleccionadas 20 secciones que aparecían diariamente y que cubrían aproximadamente la totalidad de los temas abordados en el diario.

Estas secciones vienen indicadas en la Tabla 1.

<b>Tabla 1: Secciones del diario regional (efectivos y peso de sus lectores)</b>		
Variable	efectivos	peso
Noticias internacionales	128	39.75
Clasificados	175	54.35
Noticias sobre sexo, violencia, drogas	124	38.51
Noticias policiales	91	28.26
Noticias nacionales económicas	74	22.98
Noticias nacionales políticas	90	27.95
Información sobre la corrupción política y social	104	32.30
Deportes	122	37.89
Comentarios científicos	166	51.55
Información gremial	179	55.59
Notas editoriales	168	52.17
Comentarios firmados por analistas económicos y políticos	141	43.79
Comentarios firmados por analistas locales sobre aspectos regionales	103	31.99
Noticias y comentarios agropecuarios	154	47.83
Noticias regionales	52	16.15
Investigaciones periodísticas	126	39.13
Información sobre espectáculos	145	45.03
Sociales	121	37.58
Chistes, horóscopo y entretenimiento en general	123	38.20
Suplemento y revista dominical	117	36.34

A cada encuestado se lo consultaba, en primer término, si leía el diario (322). Si la respuesta era positiva entonces se le pedía que indicara si leía esa sección del diario, si no leía, o si no la conocía (Ns/Nc).

- En consecuencia existían, para cada sección (variable cualitativa), tres respuestas posibles (modalidades).
- Las encuestas, como en el caso del diario regional, además del tema particular de investigación contiene, casi siempre, un cuestionario relativo a la caracterización del encuestado, (Tabla 2). En efecto, cada encuestado no interesa en sí mismo sino como representante de ciertas categorías de población.

<b>Tabla 2: La caracterización de los efectivos y pesos de sus modalidades</b>			
Variable	Modalidad	Efectivos	Peso
Sexo	Hombre	162	50.31
	Mujer	160	49.69
Es jefe de familia?	SI	152	47.20
	NO	170	52.80
Actividad que desarrolla	Desocupado	8	2.48
	Estudiante	28	8.70
	Ama de casa	55	17.08
	Jubilado	39	12.11
	Empleadores	10	3.11
	Trabajo no especializado o changarín	14	4.35
	Comerciante sin personal	29	9.01
	Técnico	13	4.04
	Profesional independiente y Otros autónomos	25	7.76
	Relación de dependencia	12	3.73
	Obrero, calificado y no calificado, capataz	12	3.73
	Empleado sin jerarquía	67	20.81
	Jefe intermedio	10	3.11
Nivel de Educación	Primario incompleto	23	7.14
	Primario completo	89	27.64
	Secundario incompleto	67	20.81
	Secundario completo	77	23.91
	Terciario incompleto	13	4.04
	Terciario completo	25	7.76
	Universitario incompleto	17	5.28
	Universitario completo	5	1.55
	Ns/Nc	6	1.86
Nivel de Ingresos	Menos de \$300.00	49	15.22
	De \$300.00 a \$500.00	95	29.50
	De \$501.00 a \$700.00	68	21.12
	De \$701.00 a \$1000.00	45	13.98
	De \$1001.00 a \$1500.00	28	8.70
	De \$1501.00 a \$2500.00	19	5.90
	De \$2501.00 a \$4000.00	6	1.86
	Más de \$4000.00	2	0.62
	Ns/Nc	10	3.11
Número de personas en el hogar	Una persona	10	3.11
	Dos personas	40	12.42
	De 3 a 4 personas	145	45.03
	De 5 a 6 personas	84	26.09
	Más de 7 personas	43	13.35
Tiene Computadora?	Si	42	13.04
	No	280	86.96
Tiene Televisor	Si	316	98.14
	No	6	1.86
Tiene horno a microondas	Si	41	12.73
	No	281	87.27
Tiene radio	Si	315	97.83
	No	7	2.17
Tiene auto	Si	151	46.89
	No	171	53.11
Tiene video	Si	102	31.68
	No	220	68.32
Nivel de la vivienda	Media superior	29	9.01
	Media-Media	114	35.40
	Media Baja	118	36.65
	Muy modesta o Marginal	61	18.94

### 1. Análisis Simultáneo

El caso de esta encuesta ilustra una situación muy corriente: *las variables definidas sobre un conjunto de individuos no constituyen un conjunto homogéneo, sino que están estructuradas en varios grupos.*

Del amplio uso de la técnica de los elementos suplementarios ha derivado la técnica siguiente: *una o varias tablas sirven de base al análisis, las restantes se colocan como suplementarias.*

En la encuesta del diario regional que contiene dos grupos de variables cualitativas, son posibles dos soluciones de este tipo:

- Un AFCM del conjunto de las secciones (como principal) y de la caracterización (como suplementario)
- Un AFCM del conjunto de la caracterización (como principal) y de las secciones (como suplementario)

A estos dos análisis parciales en los cuales uno de los grupos es privilegiado y sirve de base de referencia, se debe añadir:

- Un AFCM con el conjunto de las secciones y de la caracterización como principal, que es el análisis que brindará la respuesta a la pregunta "¿quién lee qué?"

## 2. Análisis factorial de correspondencias múltiples de las secciones.

De este análisis se puede esperar

- Una tipología de los individuos según su perfil de lectura: dos individuos son próximos si se interesan o no se interesan por las mismas secciones del diario
- Un estudio de las relaciones entre la lectura (o la no lectura) de las diferentes secciones: si varias secciones son, a menudo, leídas por los mismos lectores constituyen un grupo que será puesto en evidencia. Si, a la inversa, hay fenómenos de exclusión (los lectores de la sección A no leen jamás la sección B) serán detectados

Con los elementos suplementarios, un estudio de la relación entre cada una de las variables de caracterización, tomadas por separado, y los principales factores de variabilidad de los perfiles lectura.

### a. Lectura de la tabla de datos

Las modalidades utilizadas en las variables cualitativas son categorías mutuamente excluyentes entre sí. De modo que la  $j$ -ésima característica observada está compuesta de  $k_j$  modalidades mutuamente excluyentes.

Por ejemplo la variable Información internacional se compone de las modalidades: lee la sección, no lee la sección pero la conoce y no conoce la sección. Si se observa la Tabla 3 se encontrará el número efectivo de individuos (eff) que presentan una modalidad determinada dentro de cada variable considerada. Este número de individuos son los puntos (poids) considerados para el análisis.

**Tabla 3. Variables activas y sus modalidades**

ANALYSE DES CORRESPONDANCES MULTIPLES						
APUREMENT DES MODALITES ACTIVES						
SEUIL (PCMIN) :	2.00 %	POIDS :	6.44			
AVANT APUREMENT :	20 QUESTIONS ACTIVES			60 MODALITES ASSOCIEES		
APRES :	20 QUESTIONS ACTIVES			60 MODALITES ASSOCIEES		
POIDS TOTAL DES INDIVIDUS ACTIFS :		322.00				
TRI-A-PLAT DES QUESTIONS ACTIVES						
IDENT	MODALITES LIBELLE	AVANT APUREMENT EFF. POIDS	EFF. POIDS	APRES APUREMENT POIDS	HISTOGRAMME DES POIDS RELATIFS	
3 . Noticias Internacionales						
Int1 - Int1		128 128.00	128	128.00	*****	
Int2 - Int2		167 167.00	167	167.00	*****	
Int3 - Int3		27 27.00	27	27.00	*****	
4 . Clasificados						
Clal - Cla1		175 175.00	175	175.00	*****	
Cla2 - Cla2		91 91.00	91	91.00	*****	
Cla3 - Cla3		56 56.00	56	56.00	*****	
5 . Noticias de Violencia, sexo, droga						
Vio1 - Vio1		124 124.00	124	124.00	*****	
Vio2 - Vio2		146 146.00	146	146.00	*****	
Vio3 - Vio3		52 52.00	52	52.00	*****	
6 . Noticias policiales						
Pol1 - Pol1		91 91.00	91	91.00	*****	
Pol2 - Pol2		208 208.00	208	208.00	*****	
Pol3 - Pol3		23 23.00	23	23.00	*****	
7 . Noticias nacionales económicas						
Nae1 - Nae1		74 74.00	74	74.00	*****	
Nae2 - Nae2		228 228.00	228	228.00	*****	
Nae3 - Nae3		20 20.00	20	20.00	*****	
8 . Noticias nacionales políticas						
Nap1 - Nap1		90 90.00	90	90.00	*****	
Nap2 - Nap2		207 207.00	207	207.00	*****	
Nap3 - Nap3		25 25.00	25	25.00	*****	
9 . Información sobre corrupción política y social						
Cor1 - Cor1		104 104.00	104	104.00	*****	
Cor2 - Cor2		172 172.00	172	172.00	*****	
Cor3 - Cor3		46 46.00	46	46.00	*****	
10 . Deportes						
Dep1 - Dep1		122 122.00	122	122.00	*****	
Dep2 - Dep2		174 174.00	174	174.00	*****	
Dep3 - Dep3		26 26.00	26	26.00	*****	
11 . Comentarios Científicos						
Coc1 - Coc1		166 166.00	166	166.00	*****	
Coc2 - Coc2		97 97.00	97	97.00	*****	
Coc3 - Coc3		59 59.00	59	59.00	*****	
12 . Información Gremial						
Gre1 - Gre1		179 179.00	179	179.00	*****	
Gre2 - Gre2		69 69.00	69	69.00	*****	
Gre3 - Gre3		74 74.00	74	74.00	*****	
13 . Notas Editoriales						
Edi1 - Edi1		168 168.00	168	168.00	*****	
Edi2 - Edi2		85 85.00	85	85.00	*****	
Edi3 - Edi3		69 69.00	69	69.00	*****	
14 . Comentarios económicos y políticos						
Coel - Coe1		141 141.00	141	141.00	*****	
Coe2 - Coe2		129 129.00	129	129.00	*****	
Coe3 - Coe3		52 52.00	52	52.00	*****	
15 . Comentarios regionales						
Cor1 - Cr1		103 103.00	103	103.00	*****	
Cor2 - Cr2		162 162.00	162	162.00	*****	
Cor3 - Cr3		57 57.00	57	57.00	*****	
16 . Noticias Agropecuarias						
Agr1 - Agr1		154 154.00	154	154.00	*****	
Agr2 - Agr2		106 106.00	106	106.00	*****	
Agr3 - Agr3		62 62.00	62	62.00	*****	
17 . Noticias Regionales						
Reg1 - Reg1		52 52.00	52	52.00	*****	
Reg2 - Reg2		221 221.00	221	221.00	*****	
Reg3 - Reg3		49 49.00	49	49.00	*****	
18 . Investigación periodística						
Inv1 - Inv1		126 126.00	126	126.00	*****	
Inv2 - Inv2		138 138.00	138	138.00	*****	
Inv3 - Inv3		58 58.00	58	58.00	*****	
19 . Información sobre Espectáculos						
Esp1 - Esp1		145 145.00	145	145.00	*****	
Esp2 - Esp2		123 123.00	123	123.00	*****	
Esp3 - Esp3		54 54.00	54	54.00	*****	

Tabla 3 (Continuación)					
20 . Sociales					
Soc1 - Soc1	121	121.00	121	121.00	*****
Soc2 - Soc2	141	141.00	141	141.00	*****
Soc3 - Soc3	60	60.00	60	60.00	*****
21 . Chistes, Horoscopos y entretenimiento					
Chi1 - Chi1	123	123.00	123	123.00	*****
Chi2 - Chi2	166	166.00	166	166.00	*****
Chi3 - Chi3	33	33.00	33	33.00	*****
22 . Suplemento y revista dominical					
Sup1 - Sup1	117	117.00	117	117.00	*****
Sup2 - Sup2	153	153.00	153	153.00	*****
Sup3 - Sup3	52	52.00	52	52.00	*****

Cuando se observa la j-ésima característica sobre el i-ésimo individuo, se puede afectar la k-ésima modalidad de esa j-ésima característica al i-ésimo individuo.

En la Tabla 4, por ejemplo, el individuo 7 asume la modalidad 2 en noticias policiales, esto equivale a decir "no leo la sección pero la conozco".

La Tabla 4 es una tabla de códigos condensados que se construye a partir de la codificación de la información.

Tabla 4: Tabla de datos o Tabla de códigos condensados (npx)								
Observaciones	Características							
	Individuo	Noticia internacional	Clasificados	Noticias sobre sexo, drogas, violencia	Noticias policial	Noticias nacionales económicas	Noticias nacionales políticas	.....
1	2	2	2	2	2	2	2	4
2	1	1	1	2	2	2	2	4
4	2	2	2	2	2	2	2	4
5	2	2	2	2	2	2	2	4
6	2	1	1	1	2	2	2	4
7	2	2	2	2	2	2	2	4
10	2	1	1	1	2	2	2	3
11	2	2	2	2	2	2	2	3
12	1	1	1	1	1	1	1	4
....								
....								
....								
....								
....								
503	2	1	2	2	2	2	2	2

- Cada línea contiene todos los códigos correspondientes a las modalidades atribuidas a un individuo, para cada una de las características observadas. Los valores que figuran en la intersección de los individuos con las modalidades de las variables son códigos y,

precisamente por tratarse de una tabla de códigos, la tabla carece de propiedad numéricas. Esta tabla tiene tantas filas como individuos se consideren y tantas columnas como la suma de la totalidad de modalidades contempladas en todas las variables que forman parte del análisis.

El análisis factorial de correspondencias de esta tabla de códigos tiene por objetivo

- Hallar semejanza de individuos
- Estudiar relación entre variables
- Resumir las características observadas en un pequeño número de variables
- Comparar modalidades de diferentes variables

En el proceso de análisis esta tabla de códigos condensados se convierte en una tabla lógica (Tabla 5), donde la existencia o ausencia de una modalidad se denota con 1 o 0.

Tabla 5: Tabla lógica (nxp)										
Observaciones	Características									
	Noticias Internacionales			Clasificados...				....	Nivel de vivienda	
Individuo	Lee	No lee	No conoce	Lee	No lee	No conoce	....	Media Media	Media Baja	Muy modesta o marginal
1	0	1	0	0	1	0		0	0	1
2	1	0	0	1	0	0		0	0	1
4	0	1	0	0	1	0		0	0	1
5	0	1	0	0	1	0		0	0	1
6	0	1	0	1	0	0		0	0	1
7	0	1	0	0	1	0		0	0	1
10	0	1	0	1	0	0		0	1	0
11	0	1	0	0	1	0		0	1	0
12	1	0	0	1	0	0		0	0	1
....										
....										
....										
....										
....										
....										
503	0	1	0	1	0	0		1	0	0

#### Individuos

Esta tabla de códigos se representa en un espacio de  $np$  dimensiones ( $n$  individuos y  $p$  características observadas) donde las coordenadas del individuo  $i$ -ésimo viene dada por

$$\frac{x_{ij}}{p\sqrt{\frac{n_j}{np}}}$$

donde  $x_{ij}$  representa el individuo  $i$  asociado a una modalidad  $j$  de una de las  $p$  características posibles, asume el valor 1 si la modalidad se presenta y el valor cero si no se presenta

$p$  es el número de características medidas

$n_j$  es el número de individuos que asumen la característica  $j$

$np$  es el número de individuos por las  $p$  características

- Dónde radica la importancia de conocer las coordenadas de los individuos?

En que podemos conocer la distancia entre ambos y esta distancia nos permitirá conocer si dos individuos son semejantes o no de modo que podamos afirmar que pertenecen a un mismo grupo de individuos o no.

Esta distancia se conoce como distancia del Chi<sup>2</sup> y se define como

$$d_{(i,i')}^2 = \frac{1}{p} \sum_{j=1}^K \frac{n}{n_j} (x_{ij} - x_{i'j})^2$$

#### En resumen:

- cada término de la sumatoria  $(x_{ij} - x_{i'j})^2$  no puede valer más que 1 o 0. Vale 1 si dos individuos considerados no presentan simultáneamente la misma modalidad y vale 0 si se presenta la misma modalidad
- las distancias entre los individuos crece a medida que aumentan las diferencias de modalidades
- cada modalidad interviene en el cálculo de la distancia con la inversa del peso de la modalidad es decir  $\frac{n}{n_j}$
- la distancia del Chi<sup>2</sup> respeta el criterio de comparación adoptado

#### *Modalidades*

De igual manera que con los individuos debemos proceder a conocer la distancia entre modalidades. La coordenada de la modalidad  $j$  se define como



$$\frac{\frac{x_{nj}}{np}}{\frac{n_j}{np} \sqrt{\frac{p}{np}}}$$

donde  $x_{nj}$  representa la existencia de un individuo con esa modalidad, puede valer 1 o 0 de acuerdo a que la modalidad se presente o no

$np$  es el número de individuos por las  $p$  características

$p$  es el número de características medidas

$n_j$  es el número de individuos que asumen la característica  $j$

*Un nuevo referencial de representación*

Ahora bien, lo que tenemos es la información de los individuos y las modalidades en un espacio de  $np$  dimensiones.

Esto no puede observarse gráficamente y es necesario, para continuar con el análisis referirlo a otro sistema de coordenadas de manera tal que el primer plano (las coordenadas de los dos primeros ejes) brinde la mejor representación de la tabla de datos.

*Centro de gravedad*

Para esto es necesario en primer término conocer el centro de gravedad de cada nube de puntos.

La nube de puntos individuos se representan en el espacio de las modalidades y tiene un centro de gravedad cuyas coordenadas son

$$\left[ \sqrt{\frac{n_1}{np}}; \dots; \sqrt{\frac{n_j}{np}}; \dots; \sqrt{\frac{n_K}{np}} \right] \quad \forall j = 1, \dots, K.$$

La nube de puntos modalidades se representan en el espacio de los individuos y tiene un centro de gravedad cuyas coordenadas vienen dadas por la raíz cuadrada de la inversa del número total de individuos y tendremos un eje por cada individuo, es decir, tendremos un espacio de 322 dimensiones.

Las coordenadas son  $\left[ \sqrt{\frac{1}{n}}; \dots; \sqrt{\frac{1}{n}}; \dots; \sqrt{\frac{1}{n}} \right] \quad \forall i = 1, \dots, n,$

*Inercia*

Una vez conocido el centro de gravedad es necesario referir las coordenadas de todos los individuos y de todas las modalidades con su respectivo centro de gravedad.

- De esta manera obtenemos las denominadas matrices de inercia.

La inercia mide la variabilidad o dispersión de una nube de puntos respecto de un punto en común: el baricentro o centro de gravedad de la nube.

La inercia de la nube de puntos respecto al baricentro se define como la suma del producto del cuadrado de la distancia de todos los puntos respecto del baricentro por el peso asociado a cada punto. En términos algebraicos se define como

$$I_{GC} = \sum_{j=1}^K p_j d_{(j,G_C)}^2 = \sum_{j=1}^K \frac{n_j}{np} d_{(j,G_C)}^2 = \left[ \sum_{j=1}^K \frac{1}{p} \left( 1 - \frac{n_j}{n} \right) \right]$$

Esto significa que si la modalidad es rara, de modo que  $n_j$  tienda a 0, la contribución a la inercia va a ser importante.

La contribución de cada modalidad se define como

$$Contr_j \text{ a } I_{GC} = \frac{1}{p} \left( 1 - \frac{n_j}{n} \right)$$

En definitiva la inercia no es otra cosa que la suma de las contribuciones de cada punto respecto del baricentro ponderados por su peso en la nube de puntos.

La inercia global respecto de  $G_C$  de la nube de puntos de modalidades será

$$\sum_{j=1}^K Contr_j \text{ a } I_{GC} = \left[ \sum_{j=1}^K \frac{1}{p} \left( 1 - \frac{n_j}{n} \right) \right] = \sum_{j=1}^K \frac{1}{p} - \sum_{j=1}^K \frac{n_j}{np}$$

Es decir que

$$I_{GC} = \frac{K}{p} - 1$$

- De modo que la inercia total dependerá del número de características observadas ( $p$ ) y del número total de modalidades ( $K$ ) que presentaron esas características

*Habíamos dicho dos cosas:*

- *que necesitábamos un nuevo referencial de manera que pudiéramos resumir la información y proyectarla en un espacio de dos dimensiones*
- *que el nuevo referencial lo hallaríamos a partir de las matrices de inercia*

La matriz de inercia se obtiene al centrar todas las coordenadas de las modalidades (obtener la distancia respecto del baricentro). Al diagonalizar la matriz de inercia hallaremos los valores propios de la matriz que dan origen a las direcciones principales de la nube de puntos. Es indistinto hallar las direcciones principales de la matriz de inercia de las modalidades o de los individuos, en general se trabaja con la de modalidades.

#### *El cálculo de la inercia*

Habíamos dicho que la inercia total de la nube de puntos era la diferencia entre la cantidad total de modalidades que formaban parte del análisis (K) dividida la cantidad de variables consideradas (p) y 1.

$$\text{De modo que, } I_{gc}^{N_j} = \frac{K}{p} - 1 = \frac{60}{20} - 1 = 2.$$

Si observamos en la Tabla 6 veremos que la traza de la matriz de inercia antes de la diagonalización es igual a la suma de los valores propios: 2; siendo la suma de los valores propios la inercia total de la nube. Precisamente descomponer la inercia implica encontrar los valores propios.

<b>Tabla 6. Valores propios de la matriz de inercia</b>				
APERCU DE LA PRECISION DES CALCULS : TRACE AVANT DIAGONALISATION . . . 2.0000				
SOMME DES VALEURS PROPRES . . . . 2.0000				
HISTOGRAMME DES 40 PREMIERES VALEURS PROPRES				
NUMERO	VALEUR PROPRE	POURCENT.	POURCENT. CUMULE	
1	0.6276	31.38	31.38	*****
2	0.3612	18.06	49.44	*****
3	0.0914	4.57	54.01	*****
4	0.0759	3.80	57.80	*****
5	0.0716	3.58	61.38	*****
6	0.0622	3.11	64.49	*****
7	0.0520	2.60	67.09	*****
8	0.0476	2.38	69.47	*****
9	0.0432	2.16	71.63	*****
10	0.0395	1.98	73.61	*****
11	0.0365	1.82	75.43	*****
12	0.0353	1.76	77.20	*****
13	0.0333	1.67	78.86	*****
14	0.0310	1.55	80.41	*****
15	0.0300	1.50	81.91	*****
16	0.0286	1.43	83.34	*****
17	0.0266	1.33	84.67	*****
18	0.0238	1.19	85.86	*****
19	0.0232	1.16	87.02	*****
20	0.0222	1.11	88.13	*****
21	0.0216	1.08	89.21	*****
22	0.0197	0.99	90.20	*****
23	0.0182	0.91	91.11	*****
24	0.0178	0.89	92.00	*****
25	0.0175	0.87	92.87	*****
26	0.0151	0.75	93.63	*****
27	0.0144	0.72	94.35	*****
28	0.0136	0.68	95.03	*****
29	0.0134	0.67	95.70	*****
30	0.0127	0.63	96.33	*****
31	0.0112	0.56	96.89	*****
32	0.0100	0.50	97.39	*****
33	0.0098	0.49	97.88	*****
34	0.0088	0.44	98.32	*****
35	0.0078	0.39	98.71	*****
36	0.0077	0.38	99.09	*****
37	0.0055	0.27	99.37	*****
38	0.0049	0.24	99.61	*****
39	0.0046	0.23	99.84	*****
40	0.0032	0.16	100.00	*****

El método de diagonalización asegura que el primer valor propio es el más importante, por ende el factor asociado a él será la dirección principal de la nube de puntos.

La tabla informa:

- el valor asumido por el valor propio, 0.6276, y
- el peso que tiene en el total, 31.38. La columna de porcentaje acumulado indica cuánta información estoy representando al incorporar un eje adicional en el análisis.

Con los dos primeros valores propios obtengo el plano factorial en el cual tengo representada el 49.44% del total de la inercia. En la tabla se muestra el histograma de los 40 primeros valores propios, para este último se alcanza el 100% de la información.

Una pregunta común en esta etapa es, cuántos ejes conservo en el análisis? La respuesta viene dada por el cociente entre la inercia y el total de valores propios. El resultado de ese cociente se compara con los valores de los autovalores y deberán conservarse todos los ejes que se encuentren por encima de él. En este caso los 7 primeros valores propios. Ahora bien, la información que se agrega a partir del tercer eje es despreciable, en general se analiza el plano y puede en algún caso estudiarse el tercer eje. El valor asumido por el valor propio indica la inercia contenida en ese eje.

El nuevo referencial presenta tres ventajas importantes

- Permite analizar completamente la forma de la nube de puntos perfiles
- Permite la representación de la forma de la nube de puntos perfiles cualquiera que sean las dimensiones de las nubes de puntos
- Permite producir una representación objetiva de esas nubes de puntos independientes del punto de vista del análisis

#### *b. El plano de los dos primeros factores*

##### *Las variables activas*

Los puntos en el espacio que representan las modalidades de las variables utilizadas para el análisis deben referirse al nuevo sistema de coordenadas.

- Esto significa que deben proyectarse sobre las direcciones principales halladas al descomponer la inercia.

Cada modalidad tiene un peso relativo en la nube de puntos que viene dado por la cantidad de individuos que se identifican con esa modalidad respecto del producto del total de observaciones y la cantidad de variables que componen la nube de puntos.

Por ejemplo el peso relativo de la modalidad lee noticias internacionales será

$$\frac{\sum_{i=1}^n x_{ij}}{np} * 100 = \frac{128}{322 * 20} * 100 = 1.99$$

La distancia está referida al centro de gravedad y se obtiene haciendo

$$d_{(j,Gc)}^2 = \frac{n}{n_j} - 1 = \frac{322}{128} - 1 = 1.52$$

Las coordenadas indican la representación de cada modalidad sobre el nuevo sistema de referencia. Cuánto más alto es el valor de las coordenadas significa que más alejado se encuentra el punto del origen del sistema y cuánto más cerca se encuentre del eje factorial más contribuirá a su formación.

Así la modalidad no lee de noticias internacionales tiene sobre el primer eje factorial coordenada igual a  $-0.09$  está muy cercana al origen del sistema y no contribuye a la formación del eje, mientras que la modalidad no conoce la sección de noticias internacionales tienen coordenada  $-2.07$  y la contribución es de  $2.9$  la cual es muy alta si tenemos en cuenta que la variable noticias internacionales contribuye en el  $3.8\%$  a la formación del factor 1.

La contribución de la variable a la formación del factor surge de la suma de la contribución individual de cada modalidad perteneciente a la variable. Luego la suma de todas las contribuciones acumuladas de las variables consideradas es igual a  $100$ , es decir, el  $100\%$ .

La contribución, por ejemplo de la modalidad no conoce la sección de noticias internacionales, se calcula haciendo

$$\frac{\text{coordenada}^2 * \text{peso}}{\text{autovalor}} = \frac{-2.07^2 * 0.42}{0.6276} = 2.9$$

El coseno cuadrado mide la calidad de representación de la distancia del punto  $i$  al origen del eje factorial. Medir la calidad de representación significa conocer si el punto individuo o modalidad está bien representado. El coseno cuadrado varía entre  $0$  y  $1$ . Cuanto mayor sea el coseno cuadrado mejor representada tendré a esa modalidad en ese eje.

Su cálculo para la misma modalidad anterior se realiza haciendo

$$\frac{\text{coordenada}^2}{d_{(j,Gc)}^2} = \frac{-2.07^2}{10.93} = 0.39$$

Con todos estos elementos vamos ahora a leer la Tabla 7

En primer lugar buscamos aquellas variables que tengan contribución acumulada más alta en cada eje porque esto significa que son las que más aportan a su construcción.



Para el eje factorial 1 podemos separar Clasificados (contribución 5.7), Noticias de violencia, sexo y drogas (5.3), Información sobre corrupción política y social (5.0), Comentarios científicos (6.1), Información gremial (6.6), Notas editoriales (6.5), Noticias agropecuarias (5.9), Noticias regionales (5.5). Estas variables son las que más peso ejercen sobre el eje, y son en las que se concentra la mayor cantidad de inercia.

- Nótese que las modalidades de mayor peso dentro de cada variable son "lee la sección" ubicadas en el semieje positivo, y "no conoce la sección" en el semieje negativo.
- Qué significa esto? que en el primer eje factorial que reúne el 31.38% de la inercia total de la nube de puntos se oponen las personas que leen diferentes secciones del diario habitualmente y aquellas que ni siquiera la conocen.

En el factor 2 el análisis es similar las variables relevantes son Noticias nacionales económicos (3.9), Noticias nacionales políticos (3.8), Comentarios económicos y políticos (7.6), Comentarios regionales (6.4), Investigación periodística (7.7), Información sobre espectáculos (6.0), Sociales (6.5) y Suplemento y revista dominical (6.6).

- El segundo eje que reúne el 18.86% de la inercia total de la nube de puntos opone a las personas que leen determinadas secciones del diario de aquellas que no leen dichas secciones.

Puede observarse en el gráfico que se agrupan las modalidades que indican a individuos que leen secciones del diario y cercana a ellas los que no leen pero sí conocen que existen, más alejadas se encuentran aquellos que no conocen las secciones del diario.

#### *Grupos de secciones*

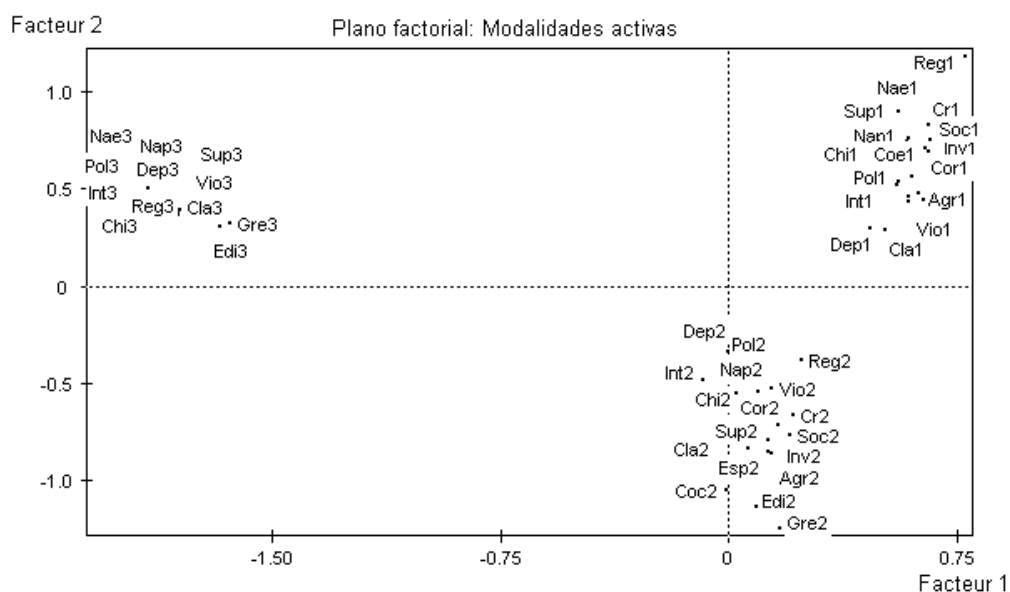
Algunas reagrupaciones son visibles sobre el primer plano, en particular el de comentarios científicos, información gremial, notas editoriales, comentarios económicos y políticos, comentarios regionales, noticias agropecuarias e investigación. Las modalidades leídas de estas secciones se sitúan todas en el primer cuadrante del plano, la modalidad no leída en el cuarto cuadrante y la modalidad no conocida mayormente en el segundo cuadrante.

El punto común de estas diferentes secciones es su aspecto relativamente intelectual y de información general. Esto implica que interesan y no interesan a las mismas subpoblaciones. Observemos que la lectura de una de ellas se encuentra muy cercana a la lectura de otras por ejemplo comentarios económicos y políticos y comentarios regionales. El interés de un análisis multidimensional estriba en poner en evidencia tales fenómenos.

*Las secciones mal representadas en el primer plano*

La calidad de representación de la modalidad no leída de la sección policiales es muy baja conjuntamente con las modalidades de leída y no leída de la sección deportes. Esto indica que su posición en el conjunto de las secciones está mal expresada sobre este plano. En realidad la variable policiales, con todas sus modalidades está bien representada en el quinto eje factorial mientras que deportes en otro eje más allá del quinto.

El factor 1 opone el conocimiento o no de las secciones del diario, mientras que el eje 2 opone la lectura de la sección a la no lectura. Si tuviéramos que darle un nombre diríamos que el eje 1 es "conocer" y el eje 2 es "leer".



En la [Tabla 8](#) tenemos los valores test o valores de prueba.

Estos valores nos van a indicar si la ubicación de las observaciones en el plano factorial es aleatoria o no. Lo que se hace es una prueba de hipótesis donde la hipótesis nula indica la aleatoriedad en la construcción de cada eje factorial,

- rechazar la hipótesis nula significa que la proyección de la nube de puntos se ha realizado de manera no aleatoria.

El valor crítico es 2 que se asocia a un nivel de probabilidad de la distribución normal de 0.9544. Si los valores test están por encima de 2 significa que se rechaza la hipótesis nula.

Se observa en la tabla que, ha excepción de la modalidad 2, para la mayoría de las variables en el eje 1 todas las modalidades tienen valores test altos y en el eje 2 todas están por encima del nivel crítico con lo cual habla de una muy buena proyección de la nube de puntos sobre el plano factorial.







- por otra parte, un individuo se sitúa en el centro de gravedad de las modalidades lee, no lee o no conoce de las secciones que lee, no lee o no conoce.

Esto significa que si un individuo tiene coordenadas positivas en el primer plano será asiduo lector de editoriales, artículos científicos, noticias gremiales, comentarios económicos y políticos e investigaciones.

Pero la posición de tal o cual punto no nos interesa en particular: el único interés del gráfico es ver que los individuos se reparten bastante uniformemente sobre el plano y que no hay clases de perfiles de lectura muy marcados.

- Por el contrario la posición de los encuestados nos interesa para representar las tendencias de quién en la pregunta quién lee qué?.

Es aquí donde interviene la caracterización de los encuestados ya que se conoce para cada individuo su sexo, su nivel de estudios, su nivel socioeconómico, etc.

Para observar mejor como estas categorías están relacionadas con los modos de lectura, es posible representar los dos baricentros de varones y mujeres, los nueve baricentros de nivel de estudios, los trece de la actividad desarrollada, los cuatro de nivel socioeconómico de la vivienda, etc. Este punto de vista del análisis de los individuos se confunde con el estudio de las modalidades suplementarias

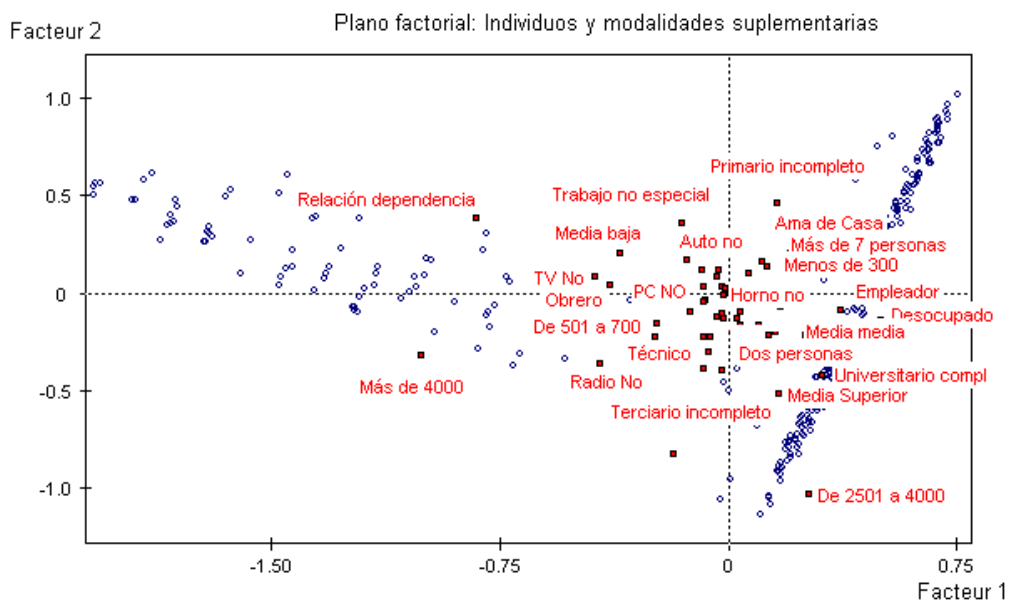
#### *Las variables suplementarias: la caracterización*

La proyección de las modalidades de estas variables sobre el plano indica que la variable relacionada con el primer factor es empleo y con el segundo educación.

El nivel de estudios está muy ligado al perfil de lectura. Los lectores cuyo nivel de estudios es débil se oponen a los encuestados de nivel de estudios elevado. Estos últimos leen las informaciones "intelectuales" y pasan sin detenerse por las páginas de información gremial, de editoriales, y de otras informaciones más anecdóticas. Es al menos lo que ellos declaran, ya que no hay que olvidar que son las respuestas y no los comportamientos de lectura lo que se analiza; pero este es otro problema.

El respeto del orden de los niveles de estudios muestra que cuanto más instruido se es más "selectiva" se vuelve la lectura del diario regional: las personas de baja instrucción leen todo el diario, las de alta eligen qué leer. Bajo nivel de estudio se relaciona con noticias y comentarios regionales, suplementos y noticias nacionales económicas. A medida que se avanza en el nivel de estudios adquieren más relevancia las noticias políticas nacionales y las noticias internacionales.

Que los puntos representantes del nivel de estudios estén sobre una misma línea indica que, en la dispersión de la nube de los perfiles de lectura, existe otra dimensión independiente del nivel de estudios. Esta dimensión representada sobre el plano por la primera bisectriz de los dos ejes opone los individuos que leen un gran número de secciones a los que leen pocas.



El análisis del resto de las variables queda para el lector.

### c. Clasificación

El objetivo de hacer análisis factorial es encontrar tipologías de individuos y asociación de variables o características relacionadas con esos individuos.

Existe un procedimiento que se denomina clasificación y que consiste en reunir a los individuos de acuerdo a su similitud, es decir, formar grupos.

Se conocen dos grandes familias de métodos estadísticos que permiten clasificar un conjunto de observaciones:

- los métodos que fraccionan un conjunto de unidades de observación en subconjuntos homogéneos.
- los procedimientos que distribuyen o designan los elementos de un conjunto de observaciones en clases preestablecidas

Cualquiera de los métodos está compuesto de procedimientos destinados a definir clases de individuos lo más semejantes posibles. Las clases son subconjuntos de individuos que se caracterizan por la proximidad en el espacio.

Cómo los métodos de análisis factorial, los métodos de clasificación jerárquica son destinados a producir una representación gráfica de la información contenida en la tabla de datos. Las clasificaciones jerárquicas tienen por objeto representar de manera sintética, el resultado de las comparaciones entre los objetos de una tabla de observaciones.

Tabla 9. Clasificación de la nube de puntos

CLASSIFICATION HIERARCHIQUE (VOISINS RECIPROQUES) SUR LES 10 PREMIERS AXES FACTORIELS						
DESCRIPTION DES 50 NOEUDS D'INDICES LES PLUS ELEVES						
NUM.	AINE	BENJ	EFF.	POIDS	INDICE	HISTOGRAMME DES INDICES DE NIVEAU
594	551	478	8	8.00	0.00233	*
595	488	527	12	12.00	0.00234	*
596	539	565	7	7.00	0.00234	*
597	517	555	5	5.00	0.00243	*
598	577	124	4	4.00	0.00244	*
599	522	534	7	7.00	0.00258	*
600	564	480	11	11.00	0.00260	*
601	588	540	11	11.00	0.00260	*
602	521	487	6	6.00	0.00268	*
603	448	574	10	10.00	0.00269	*
604	547	572	8	8.00	0.00289	*
605	570	519	20	20.00	0.00289	*
606	598	553	6	6.00	0.00306	*
607	589	578	18	18.00	0.00308	*
608	417	581	6	6.00	0.00321	*
609	562	582	14	14.00	0.00326	*
610	513	579	15	15.00	0.00336	*
611	558	597	8	8.00	0.00374	*
612	610	595	27	27.00	0.00407	*
613	593	590	9	9.00	0.00409	*
614	605	536	30	30.00	0.00425	*
615	608	557	9	9.00	0.00426	*
616	550	511	29	29.00	0.00491	*
617	587	544	11	11.00	0.00493	*
618	573	599	17	17.00	0.00533	*
619	585	583	31	31.00	0.00544	*
620	592	602	14	14.00	0.00582	*
621	596	580	9	9.00	0.00633	*
622	615	584	14	14.00	0.00687	*
623	614	591	46	46.00	0.00757	**
624	600	607	29	29.00	0.00768	**
625	603	604	18	18.00	0.00793	**
626	613	611	17	17.00	0.00813	**
627	609	626	31	31.00	0.00835	**
628	625	594	26	26.00	0.00840	**
629	576	606	11	11.00	0.00850	**
630	620	619	45	45.00	0.00995	**
631	586	618	33	33.00	0.01043	**
632	601	630	56	56.00	0.01310	**
633	621	617	20	20.00	0.01495	***
634	616	632	85	85.00	0.01561	***
635	627	629	42	42.00	0.02102	***
636	631	612	60	60.00	0.02243	****
637	628	624	55	55.00	0.02601	****
638	633	622	34	34.00	0.03474	*****
639	623	636	106	106.00	0.05335	*****
640	638	635	76	76.00	0.06902	*****
641	634	637	140	140.00	0.07247	*****
642	639	641	246	246.00	0.25261	*****
643	640	642	322	322.00	0.56270	*****
SOMME DES INDICES DE NIVEAU = 1.47214						

### Clasificación jerárquica

Una clasificación jerárquica es una serie de particiones encajadas. Por cada nudo que se forma se calcula el índice de unión que permite evaluar la distancia entre los objetos clasificados.

Existen diferentes algoritmos de clasificación entre los que se encuentran el método del vecino más próximo, el método de los centroides de la distancia media y el método basado en el crecimiento mínimo del momento de orden dos en las clases de particiones indexadas.

El método del *vecino más próximo* consiste en evaluar las distancias entre todos los individuos. Si la distancia tiende a cero significa que estos individuos tienen similitud. En función de estas distancias los individuos comienzan a agruparse formando los nudos. Estos grupos de individuos se van agrupando con otro grupo u otro individuo.

En la [Tabla 9](#) tenemos el resultado de los 50 nudos, o agrupaciones, más importantes de la clasificación jerárquica realizada sobre los 10 primeros ejes factoriales aplicando el método del vecino más próximo.

Por ejemplo, el nudo 594 se forma con el nudo 551 (primogénito) y el 478 (benjamín) quienes reúnen 8 individuos. El nudo 625 (18 individuos) está compuesto por el 603 (10 individuos) y el 604 (8 individuos). Luego el 625 y el 594 forman el 628 (26 individuos = 8+18). Sucesivamente se

van agrupando de acuerdo a la similitud existente entre ellos hasta llegar al nudo 643 que reúne a los 322 individuos.

#### Coordenadas y valores test

Cada nudo formado, y cada individuo, es sometido a una d6cima a efectos de conocer su aleatoriedad. El tener valor test muy alto nos conduce, como en casos anteriores, a rechazar la hip6tesis de aleatoriedad. Este an6lisis se realiza a los elementos ubicados en la base del dendograma y puede observarse en la [Tabla 10](#).

ELEMENTS				VALEURS-TEST					COORDONNEES				
NUM	IDENT	POIDS	EFF	1	2	3	4	5	1	2	3	4	5
1	578	6.00	6	1.11	-1.09	1.38	-0.69	-1.48	0.45	-0.44	0.56	-0.28	-0.60
2	589	12.00	12	1.33	-2.07	1.99	-1.20	-2.68	0.38	-0.59	0.57	-0.34	-0.76
3	480	4.00	4	0.80	-1.45	-0.24	-0.10	1.82	0.40	-0.72	-0.12	-0.05	0.91
4	564	7.00	7	1.12	-1.90	-1.48	0.75	1.91	0.42	-0.71	-0.55	0.28	0.71
5	594	8.00	8	1.88	1.37	-0.63	-0.46	-5.45	0.66	0.48	-0.22	-0.16	-1.91
6	572	4.00	4	1.13	-0.36	-2.67	1.07	2.48	0.56	-0.18	-1.33	0.53	1.23
7	547	4.00	4	1.34	0.79	-2.56	1.22	-0.44	0.67	0.39	-1.27	0.61	-0.22
8	574	7.00	7	1.57	0.13	-3.30	1.93	-3.92	0.59	0.05	-1.24	0.72	-1.47
9	448	3.00	3	1.24	0.98	-3.96	2.43	-0.90	0.72	0.56	-2.28	1.40	-0.52
10	583	15.00	15	1.05	-4.98	-0.01	-0.17	-2.32	0.27	-1.26	0.00	-0.04	-0.59
11	585	16.00	16	0.93	-5.17	0.29	-0.63	-0.47	0.23	-1.26	0.07	-0.15	-0.11
12	487	2.00	2	0.61	-0.76	-1.33	0.64	-0.95	0.43	-0.53	-0.94	0.45	-0.67
13	521	4.00	4	0.86	-1.55	-3.54	1.35	-3.10	0.43	-0.77	-1.76	0.67	-1.54
14	592	8.00	8	0.93	-3.33	-2.03	1.02	1.47	0.32	-1.16	-0.71	0.36	0.51
15	540	4.00	4	0.71	-2.13	-1.32	0.42	4.07	0.36	-1.06	-0.66	0.21	2.03
16	588	7.00	7	0.82	-2.39	0.61	-0.36	5.30	0.31	-0.89	0.23	-0.13	1.99
17	511	4.00	4	0.61	-2.10	1.72	-1.41	-0.44	0.31	-1.04	0.86	-0.70	-0.22
18	550	25.00	25	0.75	-9.42	-1.34	-0.37	0.60	0.14	-1.81	-0.26	-0.07	0.12
19	527	5.00	5	1.27	0.49	3.77	-2.66	-1.81	0.56	0.22	1.68	-1.18	-0.81
20	488	7.00	7	1.58	1.28	4.45	-3.25	-2.68	0.59	0.48	1.67	-1.22	-1.00
21	579	6.00	6	1.59	1.74	2.75	-1.99	-2.65	0.64	0.70	1.11	-0.80	-1.07
22	513	9.00	9	2.02	2.06	4.08	-2.75	-3.22	0.66	0.68	1.34	-0.91	-1.06
23	534	4.00	4	0.95	-0.32	1.54	-1.07	3.22	0.47	-0.16	0.77	-0.53	1.60
24	522	3.00	3	1.03	0.30	1.53	-0.73	2.61	0.59	0.17	0.88	-0.42	1.50
25	573	10.00	10	2.16	1.94	1.78	-1.60	2.66	0.67	0.60	0.56	-0.50	0.83
26	586	16.00	16	3.12	4.43	3.54	-2.19	5.42	0.76	1.08	0.86	-0.54	1.32
27	591	16.00	16	3.38	5.08	-2.38	1.30	-4.03	0.83	1.24	-0.58	0.32	-0.98
28	536	10.00	10	2.75	3.96	-4.01	2.32	1.17	0.86	1.23	-1.25	0.72	0.37
29	519	9.00	9	2.70	4.71	-1.36	1.26	2.17	0.89	1.55	-0.45	0.41	0.71
30	570	11.00	11	2.90	4.83	-1.62	1.32	3.01	0.86	1.43	-0.48	0.39	0.89
31	553	2.00	2	-2.21	0.72	-0.05	2.54	-1.60	-1.56	0.51	-0.03	1.79	-1.13
32	200	1.00	1	-1.01	0.36	1.35	1.96	2.25	-1.01	0.36	1.35	1.96	2.25
33	577	3.00	3	-2.22	-0.21	2.76	0.56	-1.10	-1.28	-0.12	1.59	0.32	-0.63
34	576	5.00	5	-4.01	0.36	2.29	0.69	1.86	-1.78	0.16	1.02	0.31	0.82
35	555	2.00	2	-0.55	-0.59	0.50	0.57	0.73	-0.39	-0.42	0.35	0.40	0.52
36	517	3.00	3	-1.67	-0.69	1.59	2.15	-0.44	-0.96	-0.40	0.91	1.24	-0.26
37	558	3.00	3	-1.75	-0.48	0.83	0.13	-1.76	-1.01	-0.27	0.48	0.07	-1.01
38	590	5.00	5	-2.87	0.13	2.45	3.99	0.99	-1.27	0.06	1.09	1.77	0.44
39	593	4.00	4	-2.83	1.08	-1.99	4.60	0.14	-1.41	0.54	-0.99	2.29	0.07
40	582	6.00	6	-3.11	0.49	2.69	4.12	-1.56	-1.26	0.20	1.09	1.67	-0.63
41	562	8.00	8	-4.21	0.23	1.66	4.47	0.27	-1.47	0.08	0.58	1.56	0.10
42	584	5.00	5	-4.05	0.73	-3.64	-6.60	0.88	-1.80	0.32	-1.62	-2.93	0.39
43	557	3.00	3	-4.27	1.34	-3.09	-4.94	0.17	-2.46	0.77	-1.78	-2.84	0.10
44	581	4.00	4	-4.23	1.36	-4.36	-6.26	0.11	-2.10	0.68	-2.17	-3.11	0.05
45	417	2.00	2	-3.51	1.06	-1.34	-1.72	1.10	-2.48	0.75	-0.95	-1.21	0.78
46	544	4.00	4	-4.80	1.47	1.79	2.22	1.49	-2.39	0.73	0.89	1.10	0.74
47	587	7.00	7	-5.82	1.65	2.55	2.25	-1.10	-2.18	0.62	0.95	0.84	-0.41
48	580	2.00	2	-1.91	-0.18	-2.57	-1.76	-0.59	-1.35	-0.13	-1.82	-1.24	-0.41
49	565	4.00	4	-4.46	1.56	-1.31	-1.19	-0.60	-2.22	0.77	-0.65	-0.59	-0.30
50	539	3.00	3	-4.09	1.15	-1.18	-2.24	-0.81	-2.36	0.66	-0.68	-1.29	-0.46

#### Dendograma

El *dendograma* es una representaci3n gr6fica del proceso de agrupamiento que sufren los individuos.

Cuando tenemos muchos puntos, como en este caso, solo se observa gr6ficamente los grupos m6s representativos, es decir, las 6ltimas 50 formaciones.

En la [Tabla 11](#) se describe la formaci3n de cada nudo del dendograma.

**Tabla 11: Descripción de los nodos de la jerarquía**  
(INDICES EN POURCENTAGE DE LA SOMME DES INDICES : 1.33943)

NOEUD NUMERO	INDICE	SUCESSEURS		EFFECT.	POIDS	COMPOSITION	
		AINE	BENJ			PREMIER	DERNIER
51	0.17	20	19	12	12.00	19	20
52	0.17	50	49	7	7.00	49	50
53	0.18	36	35	5	5.00	35	36
54	0.18	33	32	4	4.00	32	33
55	0.19	24	23	7	7.00	23	24
56	0.19	4	3	11	11.00	3	4
57	0.19	16	15	11	11.00	15	16
58	0.20	13	12	6	6.00	12	13
59	0.20	9	8	10	10.00	8	9
60	0.22	7	6	8	8.00	6	7
61	0.22	30	29	20	20.00	29	30
62	0.23	54	31	6	6.00	31	33
63	0.23	2	1	18	18.00	1	2
64	0.24	45	44	6	6.00	44	45
65	0.24	41	40	14	14.00	40	41
66	0.25	22	21	15	15.00	21	22
67	0.28	37	53	8	8.00	35	37
68	0.30	66	51	27	27.00	19	22
69	0.31	39	38	9	9.00	38	39
70	0.32	61	28	30	30.00	28	30
71	0.32	64	43	9	9.00	43	45
72	0.37	18	17	29	29.00	17	18
73	0.37	47	46	11	11.00	46	47
74	0.40	25	55	17	17.00	23	25
75	0.41	11	10	31	31.00	10	11
76	0.43	14	58	14	14.00	12	14
77	0.47	52	48	9	9.00	48	50
78	0.51	71	42	14	14.00	42	45
79	0.57	70	27	46	46.00	27	30
80	0.57	56	63	29	29.00	1	4
81	0.59	59	60	18	18.00	6	9
82	0.61	69	67	17	17.00	35	39
83	0.62	65	82	31	31.00	35	41
84	0.63	81	5	26	26.00	5	9
85	0.63	34	62	11	11.00	31	34
86	0.74	76	75	45	45.00	10	14
87	0.78	26	74	33	33.00	23	26
88	0.98	57	86	56	56.00	10	16
89	1.12	77	73	20	20.00	46	50
90	1.17	72	88	85	85.00	10	18
91	1.57	83	85	42	42.00	31	41
92	1.67	87	68	60	60.00	19	26
93	1.94	84	80	55	55.00	1	9
94	2.59	89	78	34	34.00	42	50
95	3.98	79	92	106	106.00	19	30
96	5.15	94	91	76	76.00	31	50
97	5.41	90	93	140	140.00	1	18
98	18.86	95	97	246	246.00	1	30
99	42.01	96	98	322	322.00	1	50

El dendograma nos permite identificar la cantidad de clases en las que se divide la nube de puntos. En el ejemplo pueden identificarse 3 clases.

#### *d. Partición de la nube de puntos*

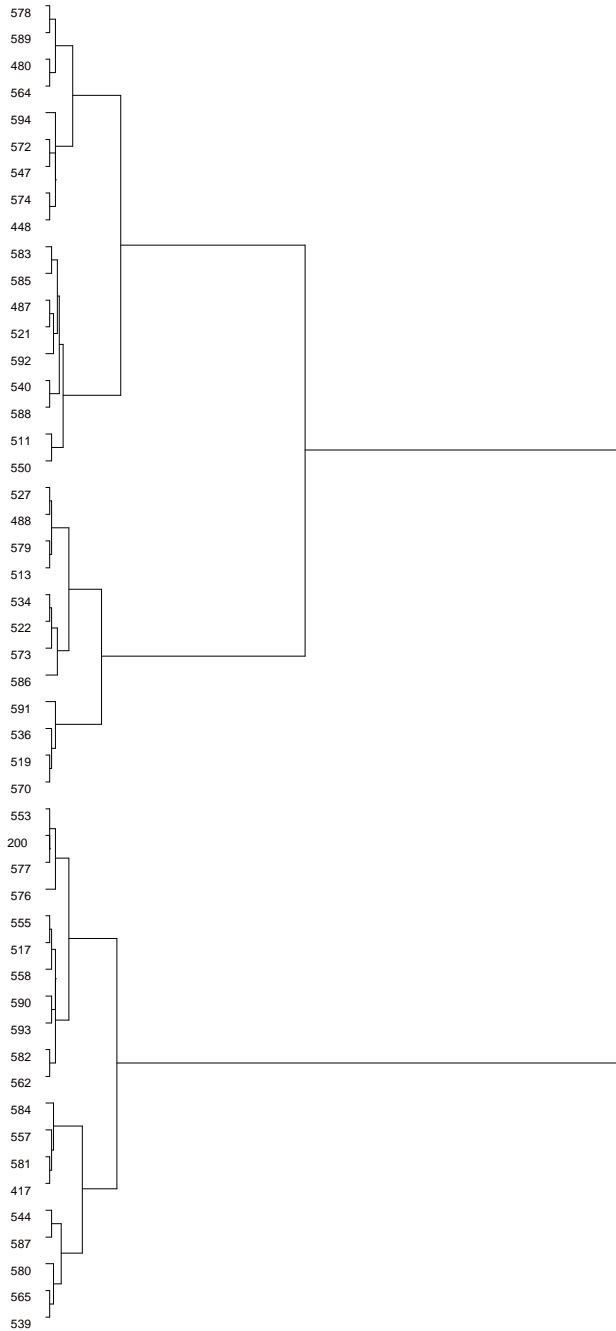
Cuando se realiza la partición encontramos, en primer lugar, la cantidad de individuos que forman parte de cada clase.

Así nos encontramos que:

- La clase 1 la forman 140 individuos que corresponden a los nudos 1 a 18, en la Figura son los primeros 18 nudos identificados en la base y en la [Tabla 12](#) los primeros 18 números.
- La segunda clase reúne a 106 individuos ubicados desde la posición 19 a la 30.
- La tercera clase corresponde a los 76 individuos reunidos desde la posición 31 a la 50.

Luego esta partición es evaluada en términos de la **inercia** que involucra a cada clase. Al realizar la consolidación de la partición se observa que las clases no quedan conformadas como en un comienzo se había visto.

Classification hierarchique directe



El reagrupamiento asegura una mejor representación en el primer plano. Esto se ve en los valores test antes de la consolidación y los valores test posteriores a la consolidación.

Obsérvese que antes de la consolidación las clases tenían buena representación en el tercer eje factorial, luego de la consolidación las tres clases se representan bien en el primer plano factorial únicamente. La inercia total es la misma (1.4721), cambia la composición y la inercia existente dentro de cada grupo.



Tabla 12. Partición del cuerpo del árbol jerárquico

COUPURE 'a' DE L'ARBRE EN 3 CLASSES														
FORMATION DES CLASSES (INDIVIDUS ACTIFS)														
DESCRIPTION SOMMAIRE														
CLASSE	EFFECTIF	POIDS	CONTENU											
aa1a	140	140.00	1 A 18											
aa2a	106	106.00	19 A 30											
aa3a	76	76.00	31 A 50											
COORDONNEES ET VALEURS-TEST AVANT CONSOLIDATION														
AXES 1 A 5														
CLASSES				VALEURS-TEST					COORDONNEES					DISTO.
IDEN - LIBELLE	EFF.	P.ABS	1	2	3	4	5	1	2	3	4	5		
COUPURE 'a' DE L'ARBRE EN 3 CLASSES														
aa1a - CLASSE 1 / 3	140	140.00	5.5	-13.8	-4.4	0.8	-1.9	0.28	-0.53	-0.09	0.01	-0.03	0.37	
aa2a - CLASSE 2 / 3	106	106.00	9.3	11.8	3.9	-2.8	2.1	0.59	0.57	0.09	-0.06	0.04	0.68	
aa3a - CLASSE 3 / 3	76	76.00	-16.8	3.1	0.9	2.2	-0.1	-1.33	0.19	0.03	0.06	0.00	1.82	
CONSOLIDATION DE LA PARTITION														
AUTOUR DES 3 CENTRES DE CLASSES, REALISEE PAR 10 ITERATIONS A CENTRES MOBILES														
PROGRESSION DE L'INERTIE INTER-CLASSES														
ITERATION	I.TOTALE	I.INTER	QUOTIENT											
0	1.47214	0.81531	0.55382											
1	1.47214	0.85452	0.58046											
2	1.47214	0.85684	0.58203											
3	1.47214	0.85719	0.58228											
4	1.47214	0.85726	0.58232											
ARRET APRES L'ITERATION 4 L'ACCROISSEMENT DE L'INERTIE INTER-CLASSES														
PAR RAPPORT A L'ITERATION PRECEDENTE N'EST QUE DE 0.008 %.														
DECOMPOSITION DE L'INERTIE														
CALCULEE SUR 10 AXES.														
INERTIES	INERTIES		EFFECTIFS		POIDS		DISTANCES							
	AVANT	APRES	AVANT	APRES	AVANT	APRES	AVANT	APRES						
INTER-CLASSES	0.8153	0.8573												
INTRA-CLASSE														
CLASSE 1 / 3	0.2471	0.1747	140	124	140.00	124.00	0.3686	0.4711						
CLASSE 2 / 3	0.1528	0.1975	106	125	106.00	125.00	0.6840	0.6115						
CLASSE 3 / 3	0.2570	0.2426	76	73	76.00	73.00	1.8214	1.9339						
TOTALE	1.4721	1.4721												
QUOTIENT (INERTIE INTER / INERTIE TOTALE) : AVANT ... 0.5538														
APRES ... 0.5823														
COORDONNEES ET VALEURS-TEST APRES CONSOLIDATION														
AXES 1 A 5														
CLASSES				VALEURS-TEST					COORDONNEES					DISTO.
IDEN - LIBELLE	EFF.	P.ABS	1	2	3	4	5	1	2	3	4	5		
COUPURE 'a' DE L'ARBRE EN 3 CLASSES														
aa1a - CLASSE 1 / 3	124	124.00	4.0	-15.3	-1.7	-0.1	0.8	0.22	-0.65	-0.04	0.00	0.01	0.47	
aa2a - CLASSE 2 / 3	125	125.00	10.4	12.4	1.2	-1.6	-0.7	0.58	0.52	0.03	-0.03	-0.01	0.61	
aa3a - CLASSE 3 / 3	73	73.00	-16.8	3.3	0.7	2.0	-0.1	-1.37	0.20	0.02	0.06	0.00	1.93	

En la [Tabla 13](#) observamos los individuos característicos de cada clase. Aquí solo se muestran los 10 principales. La condición que debe cumplir el individuo para tener esta jerarquía es tener la menor distancia respecto del centro de gravedad de la clase.

Las clases de individuos que se formaron anteriormente tienen asociadas ciertas características que quedan expuestas en la descripción de la partición.

#### Descripción de la partición

En la [Tabla 14](#) se encuentra la **primer clase**. Los 124 individuos que la componen representan el 38.51% del total de individuos analizados. Se destacan por no leer las diferentes secciones del diario, son jefes de familia y pertenecen al nivel socioeconómico medio y medio superior.

Tabla 13. Individuos característicos de las clases

PARANGONS									
CLASSE 1/ 3									
EFFECTIF: 124									
RG	DISTANCE	IDENT.	RG	DISTANCE	IDENT.	RG	DISTANCE	IDENT.	
1	0.06893	256	2	0.06893	262	3	0.10484	99	
4	0.10484	257	5	0.10560	101	6	0.11096	295	
7	0.13495	30	8	0.15079	4	9	0.15780	384	
10	0.16321	302							
CLASSE 2/ 3									
EFFECTIF: 125									
RG	DISTANCE	IDENT.	RG	DISTANCE	IDENT.	RG	DISTANCE	IDENT.	
1	0.21815	441	2	0.22005	436	3	0.22208	425	
4	0.22396	69	5	0.23404	500	6	0.24538	493	
7	0.24544	462	8	0.24969	214	9	0.25253	350	
10	0.25335	371							
CLASSE 3/ 3									
EFFECTIF: 73									
RG	DISTANCE	IDENT.	RG	DISTANCE	IDENT.	RG	DISTANCE	IDENT.	
1	0.38003	407	2	0.39686	66	3	0.40600	76	
4	0.41637	248	5	0.53653	78	6	0.56249	207	
7	0.60496	128	8	0.60496	340	9	0.62496	170	
10	0.63522	219							

Tabla 14. Descripción de la partición. Clase 1.

V.TEST	PROBA	POURCENTAGES			MODALITES	DES VARIABLES	IDEN	POIDS
		CLA/MOD	MOD/CLA	GLOBAL	CARACTERISTIQUES			
				38.51	CLASSE 1 / 3			
13.12	0.000	78.26	87.10	42.86	2	Investigación periodística	aala 124	
13.08	0.000	77.30	87.90	43.79	2	Sociales	Inv2 138	
12.15	0.000	71.90	88.71	47.52	2	Suplemento y revista dominical	Soc2 141	
11.98	0.000	77.52	80.65	40.06	2	Comentarios económicos y políticos	Sup2 153	
11.04	0.000	67.28	87.90	50.31	2	Comentarios regionales	Coe2 129	
10.95	0.000	94.20	52.42	21.43	2	Información Gremial	Cor2 162	
10.84	0.000	75.61	75.00	38.20	2	Información sobre Espectáculos	Gre2 69	
10.78	0.000	87.06	59.68	26.40	2	Notas Editoriales	Esp2 123	
10.39	0.000	81.44	63.71	30.12	2	Comentarios Científicos	Edi2 85	
10.07	0.000	55.20	98.39	68.63	2	Noticias Regionales	Coc2 97	
9.50	0.000	75.47	64.52	32.92	2	Noticias Agropecuarias	Reg2 221	
9.38	0.000	62.65	83.87	51.55	2	Chistes, Horoscopos y entretenimiento	Agr2 106	
9.36	0.000	61.63	85.48	53.42	2	Información sobre corrupción política y social	Chi2 166	
7.47	0.000	53.14	88.71	64.29	2	Noticias nacionales políticas	Cor2 172	
7.27	0.000	60.27	70.97	45.34	2	Noticias de Violencia, sexo, droga	Nap2 207	
7.18	0.000	50.44	92.74	70.81	2	Noticias nacionales económicas	Vio2 146	
7.05	0.000	56.89	76.61	51.86	2	Noticias Internacionales	Nae2 228	
5.93	0.000	64.84	47.58	28.26	2	Clasificados	Int2 167	
5.26	0.000	49.04	82.26	64.60	2	Noticias policiales	Clas2 91	
3.58	0.000	47.70	66.94	54.04	2	Deportes	Pol2 208	
2.89	0.002	65.52	15.32	9.01	Media Superior	Dep2 174		
2.52	0.006	46.05	56.45	47.20	1	Nivel vivienda	Cl 29	
2.29	0.011	47.37	43.55	35.40	Media media	Es jefe de familia?	JefS 152	
1.91	0.028	53.85	16.94	12.11	Jubilado/pensionado	Nivel vivienda	C2 114	
1.75	0.040	52.50	16.94	12.42	Dos personas	Actividad desarrollada	Jub 39	
1.64	0.050	56.00	11.29	7.76	Terciario completo	Número de persoans en el hogar	Dos 40	
1.13	0.129	47.62	16.13	13.04	PC Si	Nivel de educación	Teco 25	
1.09	0.139	60.00	4.84	3.11	Una persona	Computadora	PCS 42	
1.05	0.148	46.67	16.94	13.98	De 701 a 1000	Número de persoans en el hogar	Una 10	
1.00	0.158	52.94	7.26	5.28	Universitario incompleto	Nivel de ingresos	850 45	
1.00	0.159	41.72	50.81	46.89	Auto Si	Nivel de educación	Unin 17	
0.93	0.175	46.34	15.32	12.73	Horno si	Auto	AutS 151	
0.87	0.191	53.85	5.65	4.04	Terciario incompleto	Horno a microondas	HMS 41	
0.77	0.222	42.86	26.61	23.91	Secundario completo	Nivel de educación	Tein 13	
0.70	0.241	46.43	10.48	8.70	Estudiante	Nivel de educación	Seco 77	
0.70	0.241	46.43	10.48	8.70	De 1001 a 1500	Actividad desarrollada	Est 28	
0.61	0.270	40.69	47.58	45.03	De 3 a 4 personas	Nivel de ingresos	1250 28	
0.55	0.292	41.18	33.87	31.68	Video Si	Número de persoans en el hogar	3a4 145	
0.45	0.328	50.00	4.03	3.11	Empleador	Video	VidS 102	
0.38	0.350	44.00	8.87	7.76	Profesional y otros	Actividad desarrollada	Empr 10	
0.30	0.380	46.15	4.84	4.04	Técnico	Actividad desarrollada	Prof 25	
0.03	0.490	38.89	50.81	50.31	1	Sexo	Tec 13	
							Sex1 162	

El valor test docima la aleatoriedad de esa modalidad en la clase. Si el valor test es superior a 2 con un nivel de confianza del 95.44%, significa que rechaza la hipótesis de aleatoriedad con lo cual esa modalidad es realmente una característica de la clase.

La modalidad *no lee* de la variable Investigación periodística tiene 138 puntos en total, es decir, el 42.86% del total de personas entrevistadas presentan esta modalidad.

Luego de los 138 que presentan la modalidad en el total el 78.26% (108 individuos) forman parte de esta clase lo que significa que el 87.10% de los individuos de esta clase *no lee* la sección Investigación periodística.

En la Tabla 15 se encuentra la **segunda clase** que reúne el 38.82% de los individuos que se caracterizan por leer las diferentes secciones del diario, ser ama de casa y tener un bajo nivel de instrucción.

Estas características deben interpretarse como predominantes y no exclusivas, es decir, la mayoría de los que *leen* investigación periodística están en esta clase (el 87.20%), lo cual no significa que pueda haber en este grupo alguna persona que no lea esta sección o que no la conozca.

V.TEST	PROBA	CLA/	MOD	MOD/CLA	GLOBAL	POURCENTAGES	MODALITES	DES VARIABLES	IDEN	POIDS
							CARACTERISTIQUES			
						38.82	CLASSE 2 / 3			
14.67	0.000	86.51	87.20	39.13	1		Investigación periodística		aa2a	125
14.18	0.000	86.78	84.00	37.58	1		Sociales		Invl	126
14.09	0.000	72.62	97.60	52.17	1		Notas Editoriales		Socl	121
13.94	0.000	72.89	96.80	51.55	1		Comentarios Científicos		Edil	168
13.76	0.000	79.43	89.60	43.79	1		Comentarios económicos y políticos		Cocl	166
13.48	0.000	68.72	98.40	55.59	1		Información Gremial		Coel	141
13.06	0.000	76.55	88.80	45.03	1		Información sobre Espectáculos		Grel	179
12.77	0.000	83.76	78.40	36.34	1		Suplemento y revista dominical		Espl	145
11.38	0.000	83.50	68.80	31.99	1		Comentarios regionales		Supl	117
11.10	0.000	69.48	85.60	47.83	1		Noticias Agropecuarias		Corl	103
10.99	0.000	76.42	75.20	38.20	1		Chistes, Horoscopos y entretenimiento		Agri	154
10.70	0.000	80.77	67.20	32.30	1		Información sobre corrupción política y social		Chil	123
10.18	0.000	72.66	74.40	39.75	1		Noticias Internacionales		Corl	104
9.49	0.000	96.15	40.00	16.15	1		Noticias Regionales		Intl	128
8.84	0.000	77.78	56.00	27.95	1		Noticias nacionales políticas		Regl	52
8.62	0.000	68.55	68.00	38.51	1		Noticias de Violencia, sexo, droga		Napl	90
8.43	0.000	59.43	83.20	54.35	1		Clasificados		Viol	124
8.11	0.000	79.73	47.20	22.98	1		Noticias nacionales económicas		Cial	175
6.88	0.000	69.23	50.40	28.26	1		Noticias policiales		Nael	74
5.45	0.000	58.20	56.80	37.89	1		Deportes		Poll	91
2.15	0.016	52.73	23.20	17.08	1		Ama de Casa		Depl	122
2.01	0.022	60.87	11.20	7.14	1		Primario incompleto	Actividad desarrollada	Ama	55
1.77	0.038	47.19	33.60	27.64	1		Primario completo	Nivel de educación	Prin	23
1.73	0.042	51.02	20.00	15.22	1		Menos de 300	Nivel de educación	Prco	89
1.69	0.046	49.18	24.00	18.94	1		Muy modesta/Marginal	Nivel de ingresos	300	49
1.41	0.080	45.26	34.40	29.50	1		De 300 a 500	Nivel vivienda	D2	61
1.28	0.101	48.84	16.80	13.35	2		Más de 7 personas	Nivel de ingresos	400	95
1.03	0.151	41.76	56.80	52.80	2		Es jefe de familia?	Número de persoans en el hogar	+7	43
1.02	0.153	62.50	4.00	2.48	2		Desocupado	Es jefe de familia?	JefN	170
1.02	0.155	42.98	39.20	35.40	2		Media media	Actividad desarrollada	Des	8
0.75	0.225	42.86	28.80	26.09	2		De 5 a 6 personas	Nivel vivienda	C2	114
0.72	0.237	42.16	34.40	31.68	2		Video Si	Número de persoans en el hogar	5a6	84
0.56	0.289	47.37	7.20	5.90	2		De 1501 a 2500	Video	VidS	102
0.48	0.314	40.35	55.20	53.11	2		Auto no	Nivel de ingresos	2000	19
0.48	0.316	39.50	88.80	87.27	2		Horno no	Auto	AutN	171
0.35	0.363	44.00	8.80	7.76	2		Profesional y otros	Horno a microondas	HMN	281
0.26	0.396	39.29	88.00	86.96	2		PC NO	Actividad desarrollada	Prof	25
0.14	0.443	40.30	21.60	20.81	2		Secundario incom	Computadora	PCN	280
0.14	0.444	39.51	51.20	50.31	1		Radio Sí	Nivel de educación	Sein	67
0.14	0.445	39.05	98.40	97.83	1		Trabajo no especial	Sexo	Sexl	162
0.06	0.478	42.86	4.80	4.35	1			Radio	RS	315
								Actividad desarrollada	Noes	14

La **tercera clase** se describe en la Tabla 16. El 22.67% del total de individuos están en esta clase y se caracterizan por no conocer las diferentes secciones del diario, pertenecer a un nivel socioeconómico bajo y con ingresos promedios del hogar de \$600.

Información gremial y notas editoriales son las secciones del diario que menos conoce la población estudiada.

Tabla 16. Descripción de la partición. Clase 3.

V.TEST	PROBA	POURCENTAGES			MODALITES	DES VARIABLES	IDEN	POIDS
		CLA/MOD	MOD/CLA	GLOBAL	CARACTERISTIQUES			
				22.67	CLASSE 3 / 3		aa3a	73
16.13	0.000	93.24	94.52	22.98	3	Información Gremial	Gre3	74
15.38	0.000	94.20	89.04	21.43	3	Notas Editoriales	Edi3	69
14.77	0.000	98.33	80.82	18.63	3	Sociales	Soc3	60
14.34	0.000	95.16	80.82	19.25	3	Noticias Agropecuarias	Agr3	62
14.20	0.000	98.25	76.71	17.70	3	Comentarios regionales	Cor3	57
14.15	0.000	96.61	78.08	18.32	3	Comentarios Científicos	Coc3	59
13.74	0.000	100.00	71.23	16.15	3	Suplemento y revista dominical	Sup3	52
13.58	0.000	96.43	73.97	17.39	3	Clasificados	Cl3	56
13.56	0.000	94.83	75.34	18.01	3	Investigación periodística	Inv3	58
13.26	0.000	98.08	69.86	16.15	3	Noticias de Violencia, sexo, droga	Vio3	52
13.21	0.000	96.30	71.23	16.77	3	Información sobre Espectáculos	Esp3	54
12.84	0.000	96.15	68.49	16.15	3	Comentarios económicos y políticos	Coe3	52
12.71	0.000	97.96	65.75	15.22	3	Noticias Regionales	Reg3	49
11.72	0.000	95.65	60.27	14.29	3	Información sobre corrupción política y social	Cor3	46
10.26	0.000	100.00	45.21	10.25	3	Chistes, Horoscopos y entretenimiento	Chi3	33
9.09	0.000	100.00	36.99	8.39	3	Noticias Internacionales	Int3	27
8.33	0.000	96.15	34.25	8.07	3	Deportes	Dep3	26
8.27	0.000	100.00	31.51	7.14	3	Noticias policiales	Pol3	23
8.12	0.000	96.00	32.88	7.76	3	Noticias nacionales políticas	Nap3	25
7.00	0.000	95.00	26.03	6.21	3	Noticias nacionales económicas	Nae3	20
4.83	0.000	38.14	61.64	36.65	Media baja	Nivel vivienda	DI	118
1.95	0.026	32.35	30.14	21.12	De 501 a 700	Nivel de ingresos	600	68
1.85	0.032	50.00	8.22	3.73	Relación dependencia	Actividad desarrollada	Dep	12
1.77	0.039	37.93	15.07	9.01	Comerciante s/pers	Actividad desarrollada	Com	29
1.63	0.052	25.45	76.71	68.32	Video No	Video	VID	220
1.59	0.056	26.47	61.64	52.80	2	Es jefe de familia?	JefN	170
1.34	0.090	28.57	32.88	26.09	De 5 a 6 personas	Número de persoans en el hogar	5a6	84
1.08	0.139	28.36	26.03	20.81	Secundario incom	Nivel de educación	Sein	67
0.96	0.169	40.00	5.48	3.11	Ns/Nc	Nivel de ingresos	Nsnc	10
0.88	0.189	35.71	6.85	4.35	Trabajo no especial	Actividad desarrollada	Noes	14
0.79	0.215	23.57	90.41	86.96	PC NO	Computadora	PCN	280
0.70	0.241	24.83	49.32	45.03	De 3 a 4 personas	Número de persoans en el hogar	3a4	145
0.59	0.278	33.33	5.48	3.73	Obrero	Actividad desarrollada	Obr	12
0.46	0.322	23.98	56.16	53.11	Auto no	Auto	AutN	171
0.44	0.331	23.95	54.79	51.86	2	Noticias Internacionales	Int2	167
0.33	0.372	23.75	52.05	49.69	2	Sexo	Sex2	160
0.29	0.385	23.13	89.04	87.27	Horno no	Horno a microondas	HMN	281
0.12	0.453	23.88	21.92	20.81	Empleado s/jerarquía	Actividad desarrollada	Hmpl	67
0.11	0.457	25.00	9.59	8.70	Estudiante	Actividad desarrollada	Est	28
0.01	0.497	28.57	2.74	2.17	Radio No	Radio	RN	7

e. Tabla de Burt

- Esta tabla está formada por una yuxtaposición de tablas de contingencia; definida por las variables de los dos grupos. La Tabla de Burt consiste en la confección de la totalidad de tablas de contingencia posibles. Se construye de dos maneras: de puntos absolutos y de perfiles horizontales. En la tabla de valores absolutos (Tabla 17) definiendo una modalidad (leen la sección noticias internacionales, Int1) podemos leer hacia abajo que 102 personas leen los clasificados (Cl1). Es decir, 102 personas leen conjuntamente noticias internacionales y clasificadas.

Tabla 17. Tabla de Burt (puntos absolutos)

	Int1	Int2	Int3	Cl1	Cl2	Cl3	Vio1	Vio2	Vio3	Pol1	Pol2	Pol3	Nae1	Nae2	Nae3	Nap1	Nap2	Nap3	Cor1	Cor2	Cor3	
Int1	128	0	0																			
Int2	0	167	0																			
Int3	0	0	27																			
Cl1	102	73	0	175	0	0																
Cl2	21	69	1	0	91	0																
Cl3	5	25	26	0	0	56																
Vio1	75	49	0	98	22	4	124	0	0													
Vio2	49	88	9	74	62	10	0	146	0													
Vio3	4	30	18	3	7	42	0	0	52													
Pol1	52	38	1	74	12	5	84	5	2	91	0	0										
Pol2	72	119	17	100	77	31	39	139	30	0	208	0										
Pol3	4	10	9	1	2	20	1	2	20	0	0	23										
Nae1	57	16	1	58	10	6	52	20	2	37	36	1	74	0	0							
Nae2	67	147	14	116	78	34	72	122	34	54	165	9	0	228	0							
Nae3	4	4	12	1	3	16	0	4	16	0	7	13	0	0	20							
Nap1	65	25	0	70	16	4	55	31	4	37	52	1	70	20	0	90	0	0				
Nap2	59	135	13	104	70	33	67	109	31	52	143	12	2	201	4	0	207	0				
Nap3	4	7	14	1	5	19	2	6	17	2	13	10	2	7	16	0	0	25				
Cor1	70	34	0	82	19	3	72	31	1	51	53	0	56	48	0	66	38	0	104	0	0	
Cor2	53	113	6	87	67	18	48	106	18	36	127	9	13	153	6	20	147	5	0	172	0	
Cor3	5	20	21	6	5	35	4	9	33	4	28	14	5	27	14	4	22	20	0	0	46	











- Un estudio de las relaciones entre las diferentes variables de la caracterización
- Con los elementos suplementarios, un estudio de la relación entre el perfil característico general y la lectura de cada sección.

a. Las modalidades activas: la caracterización

LECTURE DE LA BASE DE DONNEES	
LECTURE DU FICHIER BASE	
NOM DE LA BASE : D:-1.SBA	
NOMBRE D'INDIVIDUS : 322	
NOMBRE DE VARIABLES : 34	
SELECTION DES INDIVIDUS ET DES VARIABLES UTILES	
VARIABLES NOMINALES ACTIVES	
7 VARIABLES	45 MODALITES ASSOCIEES
-----	
1 . Sexe	( 2 MODALITES )
23 . Es jefe de familia?	( 2 MODALITES )
24 . Actividad desarrollada	( 13 MODALITES )
25 . Nivel de educación	( 10 MODALITES )
26 . Nivel de ingresos	( 9 MODALITES )
27 . Número de personas en el hogar	( 5 MODALITES )
34 . Nivel vivienda	( 4 MODALITES )
-----	
VARIABLES NOMINALES ILLUSTRATIVES	
26 VARIABLES	72 MODALITES ASSOCIEES
-----	
3 . Noticias Internacionales	( 3 MODALITES )
4 . Clasificados	( 3 MODALITES )
5 . Noticias de Violencia, sexo, droga	( 3 MODALITES )
6 . Noticias policiales	( 3 MODALITES )
7 . Noticias nacionales económicas	( 3 MODALITES )
8 . Noticias nacionales políticas	( 3 MODALITES )
9 . Información sobre corrupción política y social	( 3 MODALITES )
10 . Deportes	( 3 MODALITES )
11 . Comentarios Científicos	( 3 MODALITES )
12 . Información Gremial	( 3 MODALITES )
13 . Notas Editoriales	( 3 MODALITES )
14 . Comentarios económicos y políticos	( 3 MODALITES )
15 . Comentarios regionales	( 3 MODALITES )
16 . Noticias Agropecuarias	( 3 MODALITES )
17 . Noticias Regionales	( 3 MODALITES )
18 . Investigación periodística	( 3 MODALITES )
19 . Información sobre Espectáculos	( 3 MODALITES )
20 . Sociales	( 3 MODALITES )
21 . Chistes, Horoscopos y entretenimiento	( 3 MODALITES )
22 . Suplemento y revista dominical	( 3 MODALITES )
28 . Computadora	( 2 MODALITES )
29 . Televisor	( 2 MODALITES )
30 . Horno a microondas	( 2 MODALITES )
31 . Radio	( 2 MODALITES )
32 . Auto	( 2 MODALITES )
33 . Video	( 2 MODALITES )
-----	
VARIABLES CONTINUES ILLUSTRATIVES	
1 VARIABLES	
-----	
2 . Edad	( CONTINUE )
-----	
INDIVIDUS	
-----	
POIDS DES INDIVIDUS: Poids des individus, uniforme egal a 1.	UNIF
RETENUS ..... NITOT = 322 PITOT = 322.000	
ACTIFS ..... NIACT = 322 PIACT = 322.000	
SUPPLEMENTAIRES ... NISUP = 0 PISUP = 0.000	
-----	

Tabla 21. Variables activas y sus modalidades

ANALYSE DES CORRESPONDANCES MULTIPLES					
APUREMENT DES MODALITES ACTIVES					
SEUIL (PCMIN) :	2.00 %	POIDS:	6.44		
AVANT APUREMENT :	7 QUESTIONS ACTIVES	45 MODALITES ASSOCIEES			
APRES :	7 QUESTIONS ACTIVES	40 MODALITES ASSOCIEES			
POIDS TOTAL DES INDIVIDUS ACTIFS :	322.00				
TRI-A-PLAT DES QUESTIONS ACTIVES					
IDENT	MODALITES LIBELLE	AVANT APUREMENT EFF. POIDS	APRES APUREMENT EFF. POIDS	HISTOGRAMME DES POIDS RELATIFS	
-----					
1 . Sexe					
Sex1 - 1		162 162.00	162 162.00	*****	
Sex2 - 2		160 160.00	160 160.00	*****	
-----					
23 . Es jefe de familia?					
JefS - Jefs		152 152.00	152 152.00	*****	
JefN - JefN		170 170.00	170 170.00	*****	
-----					
24 . Actividad desarrollada					
Des - Desocupado		8 8.00	8 8.00	**	
Est - Estudiante		28 28.00	28 28.00	*****	
Ama - Ama de Casa		55 55.00	55 55.00	*****	
Jub - Jubilado/pensionado		39 39.00	39 39.00	*****	
Empr - Empleador		10 10.00	10 10.00	**	
Noes - Trabajo no especial		14 14.00	14 14.00	***	
Com - Comerciante s/pers		29 29.00	29 29.00	*****	
Tec - Técnico		13 13.00	13 13.00	***	
Prof - Profesional y otros		25 25.00	25 25.00	*****	
Dep - Relación dependencia		12 12.00	12 12.00	***	
Obr - Obrero		12 12.00	12 12.00	***	
Empl - Empleado s/jerarquía		67 67.00	67 67.00	*****	
Jefe - Jefe intermedio		10 10.00	10 10.00	**	
-----					

Tabla 21. (Continuación)

IDENT	MODALITES LIBELLE	AVANT APUREMENT		APRES APUREMENT		HISTOGRAMME DES POIDS RELATIFS
		EFF.	POIDS	EFF.	POIDS	
25 . Nivel de educación						
SIN	- Sin instrucción	0	0.00			
Prin	- Primario incompleto	23	23.00	25	25.00	*****
Prco	- Primario completo	89	89.00	91	91.00	*****
Sein	- Secundario incom	67	67.00	69	69.00	*****
Seco	- Secundario completo	77	77.00	78	78.00	*****
Tein	- Terciario incompleto	13	13.00	15	15.00	***
Teco	- Terciario completo	25	25.00	27	27.00	*****
Unin	- Universitario incomp	17	17.00	17	17.00	****
Unco	- Universitario compl	5	5.00			=== VENTILEE ===
25_	- reponse manquante	6	6.00			=== VENTILEE ===
26 . Nivel de ingresos						
300	- Menos de 300	49	49.00	49	49.00	*****
400	- De 300 a 500	95	95.00	96	96.00	*****
600	- De 501 a 700	68	68.00	69	69.00	*****
850	- De 701 a 1000	45	45.00	46	46.00	*****
1250	- De 1001 a 1500	28	28.00	30	30.00	*****
2000	- De 1501 a 2500	19	19.00	19	19.00	****
3250	- De 2501 a 4000	6	6.00			=== VENTILEE ===
4000	- Más de 4000	2	2.00			=== VENTILEE ===
Nsmc	- Ns/Nc	10	10.00	13	13.00	***
27 . Número de personas en el hogar						
Una	- Una persona	10	10.00	10	10.00	**
Dos	- Dos personas	40	40.00	40	40.00	*****
3a4	- De 3 a 4 personas	145	145.00	145	145.00	*****
5a6	- De 5 a 6 personas	84	84.00	84	84.00	*****
+7	- Más de 7 personas	43	43.00	43	43.00	*****
34 . Nivel vivienda						
C1	- Media Superior	29	29.00	29	29.00	*****
C2	- Media media	114	114.00	114	114.00	*****
D1	- Media baja	118	118.00	118	118.00	*****
D2	- Muy modesta/Marginal	61	61.00	61	61.00	*****

Tabla 22. Valores Propios de la matriz de inercia

APERCU DE LA PRECISION DES CALCULS : TRACE AVANT DIAGONALISATION .. 4.7143			
SOMME DES VALEURS PROPRES .... 4.7143			
HISTOGRAMME DES 33 PREMIERES VALEURS PROPRES			
NUMERO	VALEUR PROPRE	POURCENT.	POURCENT. CUMULE
1	0.3399	7.21	7.21
2	0.2913	6.18	13.39
3	0.2344	4.97	18.36
4	0.2179	4.62	22.98
5	0.2082	4.42	27.40
6	0.1902	4.03	31.44
7	0.1865	3.96	35.39
8	0.1790	3.80	39.19
9	0.1711	3.63	42.82
10	0.1622	3.44	46.26
11	0.1598	3.39	49.65
12	0.1539	3.27	52.91
13	0.1520	3.22	56.14
14	0.1465	3.11	59.25
15	0.1443	3.06	62.31
16	0.1414	3.00	65.31
17	0.1358	2.88	68.19
18	0.1330	2.82	71.01
19	0.1303	2.76	73.77
20	0.1253	2.66	76.43
21	0.1182	2.51	78.94
22	0.1124	2.38	81.32
23	0.1092	2.32	83.64
24	0.1039	2.20	85.84
25	0.1010	2.14	87.98
26	0.0935	1.98	89.97
27	0.0899	1.91	91.87
28	0.0824	1.75	93.62
29	0.0764	1.62	95.24
30	0.0630	1.34	96.58
31	0.0587	1.24	97.83
32	0.0556	1.18	99.01
33	0.0468	0.99	100.00

Tabla 23. Coordenadas, Contribuciones y Cosenos cuadrados de las modalidades activas

IDEN - LIBELLE	P.REL	DISTO	COORDONNEES					CONTRIBUTIONS					COSINUS CARRES				
			1	2	3	4	5	1	2	3	4	5	1	2	3	4	5
1 . Sexo																	
Sex1 - 1	7.19	0.99	-0.48	-0.53	-0.29	0.08	-0.13	4.9	6.8	2.5	0.2	0.6	0.24	0.28	0.08	0.01	0.02
Sex2 - 2	7.10	1.01	0.49	0.53	0.29	-0.08	0.13	5.0	6.9	2.6	0.2	0.6	0.24	0.28	0.08	0.01	0.02
CONTRIBUTION CUMULEE = 9.9 13.8 5.1 0.4 1.2																	
23 . Es jefe de familia?																	
JefS - JefeS	6.74	1.12	-0.50	-0.67	0.03	0.10	0.05	4.9	10.3	0.0	0.3	0.1	0.22	0.40	0.00	0.01	0.00
JefN - JefN	7.54	0.89	0.44	0.60	-0.03	-0.09	-0.05	4.4	9.2	0.0	0.3	0.1	0.22	0.40	0.00	0.01	0.00
CONTRIBUTION CUMULEE = 9.3 19.5 0.1 0.5 0.2																	
24 . Actividad desarrollada																	
Des - Desocupado	0.35	39.25	0.43	-0.58	-0.02	0.89	-0.61	0.2	0.4	0.0	1.3	0.6	0.00	0.01	0.00	0.02	0.01
Est - Estudiante	1.24	10.50	-0.03	1.51	-0.62	0.59	-2.05	0.0	9.8	2.0	2.0	25.1	0.00	0.22	0.04	0.03	0.40
Ama - Ama de Casa	2.44	4.85	1.36	0.49	0.37	-0.44	0.29	13.3	2.0	1.4	2.2	1.0	0.38	0.05	0.03	0.04	0.02
Jub - Jubilado/pensionado	1.73	7.26	0.01	-0.96	1.43	0.51	-0.50	0.0	5.5	15.0	2.1	2.1	0.00	0.13	0.28	0.04	0.03
Empr - Empleador	0.44	31.20	-1.38	-0.12	0.29	-2.50	0.10	2.5	0.0	0.2	12.7	0.0	0.06	0.00	0.00	0.20	0.00
Noes - Trabajo no especial	0.62	22.00	0.35	-0.75	-1.64	-0.29	0.45	0.2	1.2	7.1	0.2	0.6	0.01	0.03	0.12	0.00	0.01
Com - Comerciante s/pers	1.29	10.10	-0.21	-0.28	-0.65	-0.07	0.24	0.2	0.4	2.3	0.0	0.3	0.00	0.01	0.04	0.00	0.01
Tec - Técnico	0.58	23.77	-1.05	-0.74	-0.13	0.05	-0.05	1.9	1.1	0.0	0.0	0.0	0.05	0.02	0.00	0.00	0.00
Prof - Profesional y otros	1.11	11.88	-0.91	-0.35	0.04	-0.25	0.09	2.7	0.5	0.0	0.3	0.0	0.07	0.01	0.00	0.01	0.00
Dep - Relación dependencia	0.53	25.83	0.07	-0.04	-1.02	1.09	0.41	0.0	0.0	2.4	2.9	0.4	0.00	0.00	0.04	0.05	0.01
Obr - Obrero	0.53	25.83	0.12	-1.22	-0.65	-0.64	-0.10	0.0	2.7	1.0	1.0	0.0	0.00	0.06	0.02	0.02	0.00
Empl - Emplead s/jerarquia	2.97	3.81	-0.19	0.24	-0.02	0.23	0.63	0.3	0.6	0.0	0.7	5.7	0.01	0.02	0.00	0.01	0.10
Jefe - Jefe intermedio	0.44	31.20	-1.57	1.00	0.28	-0.40	0.39	3.2	1.5	0.2	0.3	0.3	0.08	0.03	0.00	0.01	0.00
CONTRIBUTION CUMULEE = 24.5 25.6 31.6 25.8 36.4																	





Tabla 24. (Continuación)

MODALITES				VALEURS-TEST					COORDONNEES					DISTO.
IDEN	LIBELLE	EFF.	P. ABS	1	2	3	4	5	1	2	3	4	5	
17 . Noticias Regionales														
Reg1	- Reg1	52	52.00	0.6	0.2	0.1	-2.1	-0.4	0.08	0.02	0.01	-0.27	-0.05	5.19
Reg2	- Reg2	221	221.00	-0.4	0.2	1.3	-0.2	-1.4	-0.02	0.01	0.05	-0.01	-0.05	0.46
Reg3	- Reg3	49	49.00	-0.1	-0.4	-1.8	2.4	2.2	-0.01	-0.05	-0.23	0.31	0.30	5.57
18 . Investigación periodística														
Inv1	- Inv1	126	126.00	2.8	-1.0	0.3	-1.0	0.1	0.20	-0.07	0.02	-0.07	0.01	1.56
Inv2	- Inv2	138	138.00	-2.7	1.1	1.6	0.1	-0.2	-0.18	0.07	0.10	0.01	-0.01	1.33
Inv3	- Inv3	58	58.00	-0.1	-0.2	-2.4	1.1	0.2	-0.01	-0.02	-0.28	0.13	0.02	4.55
19 . Información sobre Espectáculos														
Esp1	- Esp1	145	145.00	0.6	-3.0	0.2	-1.6	0.6	0.04	-0.18	0.02	-0.10	0.04	1.22
Esp2	- Esp2	123	123.00	-1.7	3.3	2.0	1.1	-1.1	-0.12	0.23	0.14	0.08	-0.08	1.62
Esp3	- Esp3	54	54.00	1.5	-0.3	-2.9	0.8	0.7	0.18	-0.04	-0.36	0.10	0.08	4.96

Tabla 25. Correlación entre las variables continuas y los factores

VARIABLES			CARACTERISTIQUES				CORRELATIONS				
NUM . (IDEN)	LIBELLE	COURT	EFF.	P. ABS	MOYENNE	EC. TYPE	1	2	3	4	5
2 . (Edad)	Edad		322	322.00	41.09	15.85	0.00	-0.45	0.44	0.03	0.06

#### *Contradicción aparente entre los dos análisis*

Cada análisis centra la atención sobre un aspecto de los datos y estos puntos de vista se complementan mutuamente. En el primer análisis, la tipología de individuos traduce comportamientos de lectura ante el conjunto de secciones. Este comportamiento general está muy relacionado con numerosas variables de la caracterización especialmente educación. En este segundo análisis donde las secciones van como elementos suplementarios, ellas aparecen separadas. De donde se concluye que la caracterización está muy relacionada con el perfil general de lectura pero poco con la lectura de una sección particular. Un nivel social elevado implica muy frecuentemente una lectura "intelectual" del diario que privilegia este tipo de secciones pero no sistemáticamente una de ellas. Este ejemplo ilustra una vez más la riqueza del análisis multidimensional que permite descubrir relaciones que el estudio por separado de cada dimensión (en este caso cada sección) no puede revelar.

#### *Pérdida de las dispersiones específicas de los perfiles de lectura*

La oposición entre lectura y no lectura válida para las 20 secciones es una de las características más sobresalientes del primer análisis. Traduce una dispersión de los perfiles de lectura relacionada con el número total de secciones leídas. Esto no aparece en este análisis donde la tipología de los individuos se efectúa únicamente sobre su caracterización: toda dispersión que no esté ligada a la misma es forzosamente invisible.



Tabla 27. (Continuación)

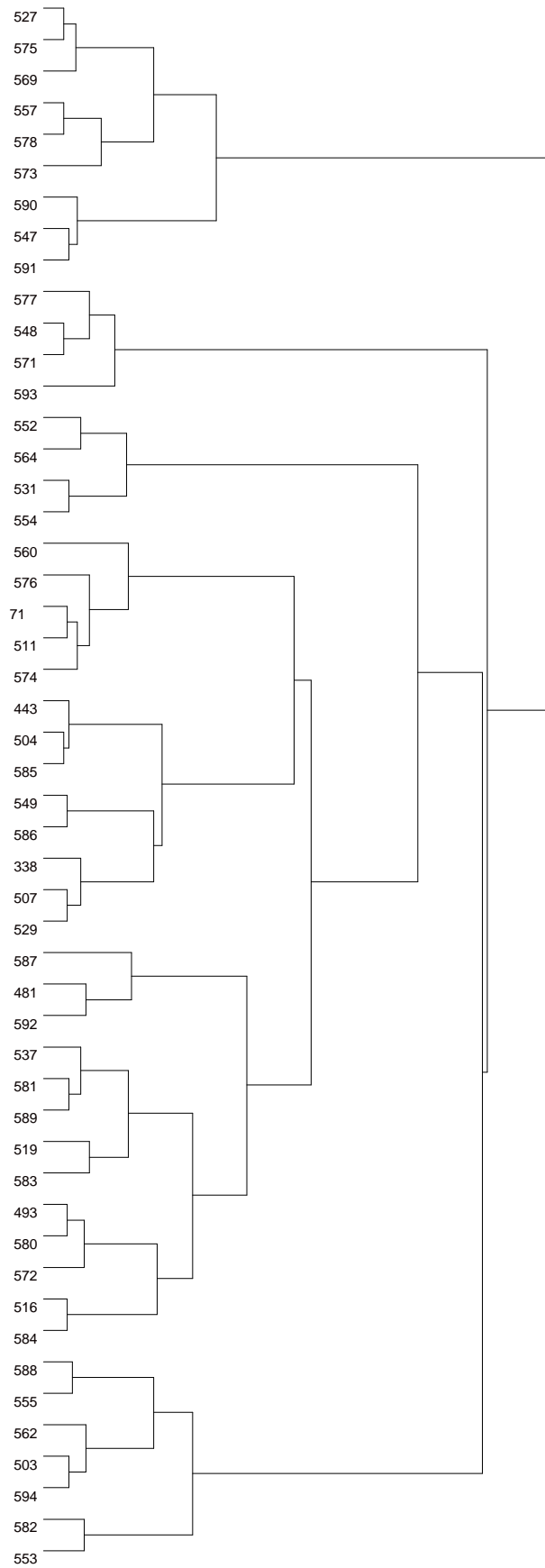
ELEMENTS				VALEURS-TEST					COORDONNEES				
NUM .	IDENT	POIDS	EFF	1	2	3	4	5	1	2	3	4	5
36 .	589	15.00	15	-2.49	-2.60	-2.23	1.43	2.14	-0.63	-0.66	-0.56	0.36	0.54
37 .	519	9.00	9	2.63	1.50	0.00	1.13	2.05	0.87	0.49	0.00	0.37	0.67
38 .	583	7.00	7	-1.30	0.86	0.55	2.00	3.99	-0.49	0.32	0.21	0.75	1.49
39 .	493	3.00	3	-1.35	1.01	0.64	1.74	2.75	-0.78	0.58	0.37	1.00	1.58
40 .	580	18.00	18	0.42	2.05	-1.69	1.98	2.93	0.10	0.47	-0.39	0.45	0.67
41 .	572	10.00	10	1.01	2.65	0.31	2.04	2.54	0.31	0.83	0.10	0.64	0.79
42 .	516	4.00	4	-1.28	0.59	-1.96	3.28	0.86	-0.64	0.29	-0.97	1.63	0.43
43 .	584	12.00	12	-2.72	-1.34	-2.48	4.37	2.59	-0.77	-0.38	-0.70	1.24	0.73
44 .	588	6.00	6	-1.54	2.53	4.22	-0.63	1.03	-0.62	1.02	1.71	-0.26	0.42
45 .	555	5.00	5	-1.55	3.64	3.46	-1.93	3.69	-0.69	1.62	1.54	-0.86	1.64
46 .	562	4.00	4	-4.73	0.85	2.11	-6.18	0.32	-2.35	0.42	1.05	-3.08	0.16
47 .	503	3.00	3	-1.07	1.08	0.34	-2.11	0.98	-0.62	0.62	0.20	-1.22	0.56
48 .	594	6.00	6	-4.98	0.61	1.31	-4.03	1.06	-2.02	0.25	0.53	-1.63	0.43
49 .	582	5.00	5	-3.83	2.27	0.72	-0.88	2.30	-1.70	1.01	0.32	-0.39	1.02
50 .	553	2.00	2	-3.09	2.59	1.27	-1.39	-1.70	-2.18	1.83	0.90	-0.98	-1.20

Tabla 28. Descripción de los nodos de la jerarquía

(INDICES EN POURCENTAGE DE LA SOMME DES INDICES : 1.71659)

NOEUD NUMERO	INDICE	SUCESSEURS		EFFECT.	POIDS	COMPOSITION	
		AINE	BENJ			PREMIER	DERNIER
51	0.41	25	24	12	12.00	24	25
52	0.41	12	11	8	8.00	11	12
53	0.43	2	1	12	12.00	1	2
54	0.43	5	4	14	14.00	4	5
55	0.44	21	20	3	3.00	20	21
56	0.46	40	39	21	21.00	39	40
57	0.47	30	29	10	10.00	29	30
58	0.47	27	26	13	13.00	26	27
59	0.48	43	42	16	16.00	42	43
60	0.51	36	35	27	27.00	35	36
61	0.51	51	23	16	16.00	23	25
62	0.53	9	8	23	23.00	8	9
63	0.53	48	47	9	9.00	47	48
64	0.53	17	16	10	10.00	16	17
65	0.57	45	44	11	11.00	44	45
66	0.66	3	53	19	19.00	1	3
67	0.71	22	55	6	6.00	20	22
68	0.72	62	7	32	32.00	7	9
69	0.73	57	28	12	12.00	28	30
70	0.73	15	14	15	15.00	14	15
71	0.77	60	34	31	31.00	34	36
72	0.81	41	56	31	31.00	39	41
73	0.84	50	49	7	7.00	49	50
74	0.85	33	32	10	10.00	32	33
75	0.87	63	46	13	13.00	46	48
76	0.91	52	10	14	14.00	10	12
77	0.94	38	37	16	16.00	37	38
78	0.94	67	19	10	10.00	19	22
79	1.16	6	54	22	22.00	4	6
80	1.43	13	76	25	25.00	10	13
81	1.67	64	70	25	25.00	14	17
82	1.73	77	71	47	47.00	34	38
83	1.74	78	18	16	16.00	18	22
84	1.80	74	31	17	17.00	31	33
85	2.25	79	66	41	41.00	1	6
86	2.27	69	58	25	25.00	26	30
87	2.28	75	65	24	24.00	44	48
88	2.34	59	72	47	47.00	39	43
89	2.47	86	61	41	41.00	23	30
90	3.08	73	87	31	31.00	44	50
91	3.11	88	82	94	94.00	34	43
92	3.57	68	85	73	73.00	1	9
93	4.18	91	84	111	111.00	31	43
94	5.16	89	83	57	57.00	18	30
95	5.55	93	94	168	168.00	18	43
96	7.75	95	81	193	193.00	14	43
97	9.08	90	96	224	224.00	14	50
98	9.20	97	80	249	249.00	10	50
99	10.52	98	92	322	322.00	1	50

Classification hierarchique directe





## c. Participación de la nube de puntos

Tabla 29. Partición del cuerpo del árbol jerárquico															
COUPURE 'a' DE L'ARBRE EN 5 CLASSES															
FORMATION DES CLASSES (INDIVIDUS ACTIFS)															
DESCRIPTION SOMMAIRE															
CLASSE	EFFECTIF	POIDS	CONTENU												
aa1a	73	73.00	1 A 9												
aa2a	25	25.00	10 A 13												
aa3a	25	25.00	14 A 17												
aa4a	168	168.00	18 A 43												
aa5a	31	31.00	44 A 50												
COORDONNEES ET VALEURS-TEST AVANT CONSOLIDATION															
AXES 1 A 5															
CLASSES				VALEURS-TEST					COORDONNEES					DISTO.	
IDEN - LIBELLE	EFF.	P.ABS			1	2	3	4	5	1	2	3	4	5	DISTO.
COUPURE 'a' DE L'ARBRE EN 5 CLASSES															
aa1a - CLASSE 1 / 5	73	73.00			9.9	-2.0	-5.1	-7.8	1.6	0.60	-0.11	-0.25	-0.38	0.08	0.62
aa2a - CLASSE 2 / 5	25	25.00			-0.1	8.1	-3.3	2.8	-11.2	-0.01	0.84	-0.31	0.25	-0.99	1.88
aa3a - CLASSE 3 / 5	25	25.00			1.8	-8.0	6.5	1.1	-5.2	0.20	-0.83	0.60	0.10	-0.46	1.63
aa4a - CLASSE 4 / 5	168	168.00			-4.4	-1.5	-0.7	8.4	5.4	-0.14	-0.04	-0.02	0.21	0.13	0.11
aa5a - CLASSE 5 / 5	31	31.00			-8.2	5.3	5.5	-6.7	3.5	-0.82	0.49	0.46	-0.53	0.27	1.62
CONSOLIDATION DE LA PARTITION															
AUTOUR DES 5 CENTRES DE CLASSES, REALISEE PAR 10 ITERATIONS A CENTRES MOBILES															
PROGRESSION DE L'INERTIE INTER-CLASSES															
ITERATION	I.TOTALE	I.INTER	QUOTIENT												
0	2.18081	0.62728	0.28764												
1	2.18081	0.69689	0.31955												
2	2.18081	0.70697	0.32418												
3	2.18081	0.71602	0.32833												
4	2.18081	0.71811	0.32928												
5	2.18081	0.71861	0.32951												
6	2.18081	0.71879	0.32960												
ARRET APRES L'ITERATION 6 L'ACCROISSEMENT DE L'INERTIE INTER-CLASSES															
PAR RAPPORT A L'ITERATION PRECEDENTE N'EST QUE DE 0.026 %.															
DECOMPOSITION DE L'INERTIE															
CALCULEE SUR 10 AXES.															
INERTIES	INERTIES		EFFECTIFS		POIDS		DISTANCES								
	AVANT	APRES	AVANT	APRES	AVANT	APRES	AVANT	APRES							
INTER-CLASSES	0.6273	0.7188													
INTRA-CLASSE															
CLASSE 1 / 5	0.2584	0.2919	73	85	73.00	85.00	0.6159	0.5973							
CLASSE 2 / 5	0.0942	0.1397	25	32	25.00	32.00	1.8803	1.5560							
CLASSE 3 / 5	0.0818	0.0766	25	23	25.00	23.00	1.6286	1.9347							
CLASSE 4 / 5	0.9227	0.6671	168	144	168.00	144.00	0.1136	0.1949							
CLASSE 5 / 5	0.1965	0.2866	31	38	31.00	38.00	1.6196	1.5350							
TOTALE	2.1808	2.1808													
QUOTIENT (INERTIE INTER / INERTIE TOTALE) : AVANT ... 0.2876															
APRES ... 0.3296															
COORDONNEES ET VALEURS-TEST APRES CONSOLIDATION															
AXES 1 A 5															
CLASSES				VALEURS-TEST					COORDONNEES					DISTO.	
IDEN - LIBELLE	EFF.	P.ABS			1	2	3	4	5	1	2	3	4	5	DISTO.
COUPURE 'a' DE L'ARBRE EN 5 CLASSES															
aa1a - CLASSE 1 / 5	85	85.00			12.7	2.0	-1.0	-6.2	2.3	0.69	0.10	-0.05	-0.27	0.10	0.60
aa2a - CLASSE 2 / 5	32	32.00			-0.4	8.3	-3.9	2.6	-11.3	-0.04	0.75	-0.32	0.20	-0.87	1.56
aa3a - CLASSE 3 / 5	23	23.00			1.9	-8.3	7.4	1.4	-5.3	0.22	-0.90	0.72	0.13	-0.49	1.93
aa4a - CLASSE 4 / 5	144	144.00			-5.3	-5.3	-3.8	8.4	6.8	-0.19	-0.18	-0.11	0.24	0.19	0.19
aa5a - CLASSE 5 / 5	38	38.00			-10.4	4.3	5.0	-8.1	1.1	-0.92	0.36	0.37	-0.58	0.08	1.53

Tabla 30. Individuos característicos de las clases

PARANGONS									
CLASSE 1/ 5									
EFFECTIF: 85									
RG	DISTANCE	IDENT.	RG	DISTANCE	IDENT.	RG	DISTANCE	IDENT.	
1	0.20798	286	2	0.25098	472	3	0.27867	41	
4	0.29065	333	5	0.29065	382	6	0.32721	328	
7	0.35555	212	8	0.40287	441	9	0.49344	77	
10	0.49344	324							
CLASSE 2/ 5									
EFFECTIF: 32									
RG	DISTANCE	IDENT.	RG	DISTANCE	IDENT.	RG	DISTANCE	IDENT.	
1	0.33634	427	2	0.51358	177	3	0.63525	440	
4	0.70363	223	5	0.88887	75	6	0.89027	182	
7	1.02921	70	8	1.03503	203	9	1.16664	32	
10	1.17149	266							
CLASSE 3/ 5									
EFFECTIF: 23									
RG	DISTANCE	IDENT.	RG	DISTANCE	IDENT.	RG	DISTANCE	IDENT.	
1	0.33432	305	2	0.36739	417	3	0.45769	139	
4	0.49155	34	5	0.51512	125	6	0.56911	197	
7	0.59156	375	8	0.63584	162	9	0.64157	178	
10	0.67288	150							
CLASSE 4/ 5									
EFFECTIF: 144									
RG	DISTANCE	IDENT.	RG	DISTANCE	IDENT.	RG	DISTANCE	IDENT.	
1	0.18543	410	2	0.18918	345	3	0.19382	422	
4	0.32790	179	5	0.39655	446	6	0.40943	138	
7	0.45814	405	8	0.46313	40	9	0.48797	391	
10	0.49499	299							
CLASSE 5/ 5									
EFFECTIF: 38									
RG	DISTANCE	IDENT.	RG	DISTANCE	IDENT.	RG	DISTANCE	IDENT.	
1	0.74989	307	2	0.79162	20	3	1.30296	465	
4	1.31059	16	5	1.33150	195	6	1.48664	140	
7	1.49988	241	8	1.54862	176	9	1.62326	302	
10	1.64490	10							

## d. Descripción de la partición

Tabla 31. Descripción de la partición. Clase 1.

CLASSE 1 / 5										
V. TEST	PROBA	POURCENTAGES			MODALITES		DES VARIABLES		IDEN	POIDS
		CLA/MOD	MOD/CLA	GLOBAL	CARACTERISTIQUES					
				26.40	CLASSE 1 / 5				aala	85
12.80	0.000	98.18	63.53	17.08	Ama de Casa	Actividad desarrollada			Ama	55
7.88	0.000	44.12	88.24	52.80	Jefn	Es jefe de familia?			JefN	170
7.66	0.000	45.00	84.71	49.69	2	Sexo			Sex2	160
6.84	0.000	63.93	45.88	18.94	Muy modesta/Marginal	Nivel vivienda			D2	61
6.02	0.000	51.69	54.12	27.64	Primario completo	Nivel de educación			Prco	89
5.50	0.000	61.22	35.29	15.22	Menos de 300	Nivel de ingresos			300	49
3.78	0.000	42.22	44.71	27.95	Napl	Noticias nacionales políticas			Napl	90
3.60	0.000	29.64	97.65	86.96	PC NO	Computadora			PCN	280
3.52	0.000	29.54	97.65	87.27	Horno no	Horno a microondas			HMN	281
3.49	0.000	43.24	37.65	22.98	Nael	Noticias nacionales económicas			Nael	74
3.43	0.000	37.70	54.12	37.89	Depl	Deportes			Depl	122
3.42	0.000	34.50	69.41	53.11	Auto no	Auto			AutN	171
3.20	0.001	31.82	82.35	68.32	Video No	Video			VidN	220
3.12	0.001	35.46	58.82	43.79	Coel	Comentarios económicos y políticos			Coel	141
3.11	0.001	38.95	43.53	29.50	De 300 a 500	Nivel de ingresos			400	95
2.75	0.003	35.16	52.94	39.75	Intl	Noticias Internacionales			Intl	128
2.68	0.004	36.54	44.71	32.30	Cor1	Información sobre corrupción política y social			Cor1	104

Tabla 32. Descripción de la partición. Clase 2.

CLASSE 2 / 5									
V.TEST	PROBA	POURCENTAGES			MODALITES	DES VARIABLES		IDEN	POIDS
		CLA/MOD	MOD/CLA	GLOBAL	CARACTERISTIQUES				
		9.94			CLASSE 2 / 5		aa2a		32
11.94	0.000	96.43	84.38	8.70	Estudiante	Actividad desarrollada		Est	28
7.00	0.000	92.31	37.50	4.04	Terciario incompleto	Nivel de educación		Tein	13
4.61	0.000	17.06	90.62	52.80	JefN	Es jefe de familia?		JefN	170
3.88	0.000	47.06	25.00	5.28	Universitario incompleto	Nivel de educación		Unin	17

Tabla 33. Descripción de la partición. Clase 3

CLASSE 3 / 5									
V.TEST	PROBA	POURCENTAGES			MODALITES	DES VARIABLES		IDEN	POIDS
		CLA/MOD	MOD/CLA	GLOBAL	CARACTERISTIQUES				
		7.14			CLASSE 3 / 5		aa3a		23
7.14	0.000	60.87	60.87	7.14	Primario incompleto	Nivel de educación		Prin	23
6.67	0.000	41.03	69.57	12.11	Jubilado/pensionado	Actividad desarrollada		Jub	39
6.39	0.000	90.00	39.13	3.11	Una persona	Número de personas en el hogar		Una	10
3.83	0.000	12.28	91.30	53.11	Auto no	Auto		AutN	171
3.39	0.000	12.50	82.61	47.20	Jefs	Es jefe de familia?		JefS	152
2.72	0.003	18.37	39.13	15.22	Menos de 300	Nivel de ingresos		300	49
2.39	0.008	9.55	91.30	68.32	Video No	Video		VidN	220

Tabla 34. Descripción de la partición. Clase 4.

CLASSE 4 / 5									
V.TEST	PROBA	POURCENTAGES			MODALITES	DES VARIABLES		IDEN	POIDS
		CLA/MOD	MOD/CLA	GLOBAL	CARACTERISTIQUES				
		44.72			CLASSE 4 / 5		aa4a		144
6.82	0.000	63.58	71.53	50.31	1	Sexo		Sex1	162
6.70	0.000	64.47	68.06	47.20	Jefs	Es jefe de familia?		JefS	152
5.30	0.000	71.43	38.19	23.91	Secundario completo	Nivel de educación		Seco	77
4.87	0.000	71.64	33.33	20.81	Empleado s/jerarquía	Actividad desarrollada		Empl	67
4.36	0.000	75.56	23.61	13.98	De 701 a 1000	Nivel de ingresos		850	45
4.15	0.000	67.65	31.94	21.12	De 501 a 700	Nivel de ingresos		600	68
3.54	0.000	54.02	65.28	54.04	Dep2	Deportes		Dep2	174
3.40	0.000	57.89	45.83	35.40	Media media	Nivel vivienda		C2	114
2.96	0.002	72.41	14.58	9.01	Comerciante s/pers	Actividad desarrollada		Com	29
2.73	0.003	55.08	45.14	36.65	Media baja	Nivel vivienda		D1	118
2.68	0.004	63.27	21.53	15.22	Reg3	Noticias Regionales		Reg3	49

Tabla 35. Descripción de la partición. Clase 5.

CLASSE 5 / 5									
V.TEST	PROBA	POURCENTAGES			MODALITES	DES VARIABLES		IDEN	POIDS
		CLA/MOD	MOD/CLA	GLOBAL	CARACTERISTIQUES				
		11.80			CLASSE 5 / 5		aa5a		38
9.99	0.000	86.21	65.79	9.01	Media Superior	Nivel vivienda		C1	29
6.05	0.000	57.14	42.11	8.70	De 1001 a 1500	Nivel de ingresos		1250	28
5.20	0.000	41.46	44.74	12.73	Horno si	Horno a microondas		HMS	41
5.05	0.000	52.00	34.21	7.76	Terciario completo	Nivel de educación		Teco	25
4.87	0.000	21.19	84.21	46.89	Auto Si	Auto		AutS	151
4.82	0.000	80.00	21.05	3.11	Jefe intermedio	Actividad desarrollada		Jefe	10
4.81	0.000	25.49	68.42	31.68	Video Si	Video		VidS	102
4.71	0.000	38.10	42.11	13.04	PC Si	Computadora		PCS	42
4.34	0.000	52.63	26.32	5.90	De 1501 a 2500	Nivel de ingresos		2000	19
3.85	0.000	18.56	81.58	51.86	Int2	Noticias Internacionales		Int2	167
3.46	0.000	60.00	15.79	3.11	Empleador	Actividad desarrollada		Empr	10
3.24	0.001	19.38	65.79	40.06	Coe2	Comentarios económicos y políticos		Coe2	129
3.14	0.001	15.35	92.11	70.81	Nae2	Noticias nacionales económicas		Nae2	228
2.97	0.001	41.18	18.42	5.28	Universitario incompleto	Nivel de educación		Unin	17
2.94	0.002	17.28	73.68	50.31	Cr2	Comentarios regionales		Cor2	162
2.67	0.004	15.46	84.21	64.29	Nap2	Noticias nacionales políticas		Nap2	207
2.54	0.006	16.28	73.68	53.42	Cor2	Información sobre corrupción política y social		Cor2	172
2.51	0.006	19.78	47.37	28.26	Poll	Noticias policiales		Poll	91
2.51	0.006	18.87	52.63	32.92	Agr2	Noticias Agropecuarias		Agr2	106

Tabla 36. Caracterización de la partición para las variables continuas.

CARACTERISATION PAR LES CONTINUES DES CLASSES OU MODALITES DE COUPURE 'a' DE L'ARBRE EN 5 CLASSES CLASSE 1 / 5							
V. TEST	PROBA	MOYENNES CLASSE GENERALE		ECARTS TYPES CLASSE GENERAL		NUM. LIBELLE	VARIABLES CARACTERISTIQUES
IDEN							
CLASSE 2 / 5							
V. TEST	PROBA	MOYENNES CLASSE GENERALE		ECARTS TYPES CLASSE GENERAL		NUM. LIBELLE	VARIABLES CARACTERISTIQUES
IDEN							
CLASSE 2 / 5 ( POIDS = 32.00 EFFECTIF = 32 )							
-6.86	0.000	22.81	41.09	8.25	15.85	2.Edad	Edad
CLASSE 3 / 5							
V. TEST	PROBA	MOYENNES CLASSE GENERALE		ECARTS TYPES CLASSE GENERAL		NUM. LIBELLE	VARIABLES CARACTERISTIQUES
IDEN							
CLASSE 3 / 5 ( POIDS = 23.00 EFFECTIF = 23 )							
7.45	0.000	64.87	41.09	14.43	15.85	2.Edad	Edad
CLASSE 4 / 5							
V. TEST	PROBA	MOYENNES CLASSE GENERALE		ECARTS TYPES CLASSE GENERAL		NUM. LIBELLE	VARIABLES CARACTERISTIQUES
IDEN							
CLASSE 5 / 5							
V. TEST	PROBA	MOYENNES CLASSE GENERALE		ECARTS TYPES CLASSE GENERAL		NUM. LIBELLE	VARIABLES CARACTERISTIQUES
IDEN							

#### 4. Análisis factorial de correspondencias múltiples con el conjunto de las secciones y de la caracterización

El Análisis Factorial Múltiple, es un método de análisis de tablas compuestas de varios grupos de variables y aporta una solución más satisfactoria al problema del equilibrio de los grupos que los anteriores análisis.

Comprende indicadores de ayudas a la interpretación relativos a los grupos y algunos otros indicadores que permiten comparar los grupos entre sí.

De este análisis se puede esperar

- Una tipología de los individuos según su perfil de lectura y su caracterización: dos individuos son próximos si se interesan o no se interesan por las mismas secciones del diario y si tienen similares características socioeconómicas.
- Con los elementos suplementarios, un estudio de la relación entre las variables activas y suplementarias.

##### a. Lectura de la tabla de datos

En la Tabla 37 se encontrarán las variables consideradas y el número de modalidades de cada una de ellas, las variables suplementarias y los individuos que forman parte del análisis.

En la Tabla 38 se encuentra una descripción de cada variable donde se consigna el número de individuos (eff) que presentan una modalidad

determinada dentro de cada variable considerada. Este número de individuos son los puntos (poids) considerados para el análisis.

Tabla 37. Lectura de la base de datos

LECTURE DU FICHER BASE	
NOM DE LA BASE	: D~-1.SBA
NOMBRE D'INDIVIDUS	: 322
NOMBRE DE VARIABLES	: 34
SELECTION DES INDIVIDUS ET DES VARIABLES UTILES	
VARIABLES NOMINALES ACTIVES	
27 VARIABLES	105 MODALITES ASSOCIEES
-----	
1 . Sexe	( 2 MODALITES )
3 . Noticias Internacionales	( 3 MODALITES )
4 . Clasificados	( 3 MODALITES )
5 . Noticias de Violencia, sexo, droga	( 3 MODALITES )
6 . Noticias policiales	( 3 MODALITES )
7 . Noticias nacionales económicas	( 3 MODALITES )
8 . Noticias nacionales políticas	( 3 MODALITES )
9 . Información sobre corrupción política y social	( 3 MODALITES )
10 . Deportes	( 3 MODALITES )
11 . Comentarios Científicos	( 3 MODALITES )
12 . Información Gremial	( 3 MODALITES )
13 . Notas Editoriales	( 3 MODALITES )
14 . Comentarios económicos y políticos	( 3 MODALITES )
15 . Comentarios regionales	( 3 MODALITES )
16 . Noticias Agropecuarias	( 3 MODALITES )
17 . Noticias Regionales	( 3 MODALITES )
18 . Investigación periodística	( 3 MODALITES )
19 . Información sobre Espectáculos	( 3 MODALITES )
20 . Sociales	( 3 MODALITES )
21 . Chistes, Horoscopos y entretenimiento	( 3 MODALITES )
22 . Suplemento y revista dominical	( 3 MODALITES )
23 . Es jefe de familia?	( 2 MODALITES )
24 . Actividad desarrollada	( 13 MODALITES )
25 . Nivel de educación	( 10 MODALITES )
26 . Nivel de ingresos	( 9 MODALITES )
27 . Número de personas en el hogar	( 5 MODALITES )
34 . Nivel vivienda	( 4 MODALITES )
-----	
VARIABLES NOMINALES ILLUSTRATIVES	
6 VARIABLES	12 MODALITES ASSOCIEES
-----	
28 . Computadora	( 2 MODALITES )
29 . Televisor	( 2 MODALITES )
30 . Horno a microondas	( 2 MODALITES )
31 . Radio	( 2 MODALITES )
32 . Auto	( 2 MODALITES )
33 . Video	( 2 MODALITES )
-----	
INDIVIDUS	
-----	
POIDS DES INDIVIDUS: Poids des individus, uniforme egal a 1.	UNIF
RETENUS ..... NITOT = 322	PITOT = 322.000
ACTIFS ..... NIACT = 322	PIACT = 322.000
SUPPLEMENTAIRES .... NISUP = 0	PISUP = 0.000

Tabla 38. Variables activas y sus modalidades

ANALYSE DES CORRESPONDANCES MULTIPLES					
APUREMENT DES MODALITES ACTIVES					
SEUIL (PCMIN)	: 2.00 %	POIDS:	6.44		
AVANT APUREMENT :	27 QUESTIONS ACTIVES	105 MODALITES ASSOCIEES			
APRES :	27 QUESTIONS ACTIVES	100 MODALITES ASSOCIEES			
POIDS TOTAL DES INDIVIDUS ACTIFS : 322.00					
TRI-A-PLAT DES QUESTIONS ACTIVES					
IDENT	MODALITES LIBELLE	AVANT APUREMENT EFF. POIDS	APRES APUREMENT EFF. POIDS	HISTOGRAMME DES POIDS RELATIFS	
-----					
1 . Sexe					
Sex1 - 1		162 162.00	162 162.00	*****	
Sex2 - 2		160 160.00	160 160.00	*****	
-----					
3 . Noticias Internacionales					
Int1 - Int1		128 128.00	128 128.00	*****	
Int2 - Int2		167 167.00	167 167.00	*****	
Int3 - Int3		27 27.00	27 27.00	*****	
-----					
4 . Clasificados					
Cla1 - Cla1		175 175.00	175 175.00	*****	
Cla2 - Cla2		91 91.00	91 91.00	*****	
Cla3 - Cla3		56 56.00	56 56.00	*****	
-----					
5 . Noticias de Violencia, sexo, droga					
Vio1 - Vio1		124 124.00	124 124.00	*****	
Vio2 - Vio2		146 146.00	146 146.00	*****	
Vio3 - Vio3		52 52.00	52 52.00	*****	
-----					
6 . Noticias policiales					
Pol1 - Pol1		91 91.00	91 91.00	*****	
Pol2 - Pol2		208 208.00	208 208.00	*****	
Pol3 - Pol3		23 23.00	23 23.00	*****	
-----					

Tabla 38. (Continuación)

IDENT	MODALITES LIBELLE	AVANT APUREMENT		APRES APUREMENT		HISTOGRAMME DES POIDS RELATIFS
		EFF.	POIDS	EFF.	POIDS	
Noticias nacionales económicas						
Nae1 - Nae1		74	74.00	74	74.00	*****
Nae2 - Nae2		228	228.00	228	228.00	*****
Nae3 - Nae3		20	20.00	20	20.00	****
8 . Noticias nacionales políticas						
Nap1 - Nap1		90	90.00	90	90.00	*****
Nap2 - Nap2		207	207.00	207	207.00	*****
Nap3 - Nap3		25	25.00	25	25.00	*****
9 . Información sobre corrupción política y social						
Cor1 - Cor1		104	104.00	104	104.00	*****
Cor2 - Cor2		172	172.00	172	172.00	*****
Cor3 - Cor3		46	46.00	46	46.00	*****
10 . Deportes						
Dep1 - Dep1		122	122.00	122	122.00	*****
Dep2 - Dep2		174	174.00	174	174.00	*****
Dep3 - Dep3		26	26.00	26	26.00	****
11 . Comentarios Científicos						
Coc1 - Coc1		166	166.00	166	166.00	*****
Coc2 - Coc2		97	97.00	97	97.00	*****
Coc3 - Coc3		59	59.00	59	59.00	*****
12 . Información Gremial						
Gre1 - Gre1		179	179.00	179	179.00	*****
Gre2 - Gre2		69	69.00	69	69.00	*****
Gre3 - Gre3		74	74.00	74	74.00	*****
13 . Notas Editoriales						
Edi1 - Edi1		168	168.00	168	168.00	*****
Edi2 - Edi2		85	85.00	85	85.00	*****
Edi3 - Edi3		69	69.00	69	69.00	*****
14 . Comentarios económicos y políticos						
Coe1 - Coe1		141	141.00	141	141.00	*****
Coe2 - Coe2		129	129.00	129	129.00	*****
Coe3 - Coe3		52	52.00	52	52.00	*****
15 . Comentarios regionales						
Cor1 - Cr1		103	103.00	103	103.00	*****
Cor2 - Cr2		162	162.00	162	162.00	*****
Cor3 - Cr3		57	57.00	57	57.00	*****
16 . Noticias Agropecuarias						
Agr1 - Agr1		154	154.00	154	154.00	*****
Agr2 - Agr2		106	106.00	106	106.00	*****
Agr3 - Agr3		62	62.00	62	62.00	*****
17 . Noticias Regionales						
Reg1 - Reg1		52	52.00	52	52.00	*****
Reg2 - Reg2		221	221.00	221	221.00	*****
Reg3 - Reg3		49	49.00	49	49.00	*****
18 . Investigación periodística						
Inv1 - Inv1		126	126.00	126	126.00	*****
Inv2 - Inv2		138	138.00	138	138.00	*****
Inv3 - Inv3		58	58.00	58	58.00	*****
19 . Información sobre Espectáculos						
Esp1 - Esp1		145	145.00	145	145.00	*****
Esp2 - Esp2		123	123.00	123	123.00	*****
Esp3 - Esp3		54	54.00	54	54.00	*****
20 . Sociales						
Soc1 - Soc1		121	121.00	121	121.00	*****
Soc2 - Soc2		141	141.00	141	141.00	*****
Soc3 - Soc3		60	60.00	60	60.00	*****
21 . Chistes, Horoscopos y entretenimiento						
Chi1 - Chi1		123	123.00	123	123.00	*****
Chi2 - Chi2		166	166.00	166	166.00	*****
Chi3 - Chi3		33	33.00	33	33.00	*****
22 . Suplemento y revista dominical						
Sup1 - Sup1		117	117.00	117	117.00	*****
Sup2 - Sup2		153	153.00	153	153.00	*****
Sup3 - Sup3		52	52.00	52	52.00	*****
23 . Es jefe de familia?						
JefS - JefS		152	152.00	152	152.00	*****
JefN - JefN		170	170.00	170	170.00	*****
24 . Actividad desarrollada						
Des - Desocupado		8	8.00	8	8.00	**
Est - Estudiante		28	28.00	28	28.00	*****
Ama - Ama de Casa		55	55.00	55	55.00	*****
Jub - Jubilado/pensionado		39	39.00	39	39.00	*****
Empr - Empleador		10	10.00	10	10.00	**
Noes - Trabajo no especial		14	14.00	14	14.00	***
Com - Comerciante s/pers		29	29.00	29	29.00	*****
Tec - Técnico		13	13.00	13	13.00	***
Prof - Profesional y otros		25	25.00	25	25.00	*****
Dep - Relación dependencia		12	12.00	12	12.00	***
Obr - Obrero		12	12.00	12	12.00	***
Empl - Empleado s/jerarquía		67	67.00	67	67.00	*****
Jefe - Jefe intermedio		10	10.00	10	10.00	**

Tabla 38. (Continuación)

IDENT	MODALITES LIBELLE	AVANT APUREMENT		APRES APUREMENT		HISTOGRAMME DES POIDS RELATIFS
		EFF.	POIDS	EFF.	POIDS	
Nivel de educación						
SIN	- Sin instrucción	0	0.00			
Prin	- Primario incompleto	23	23.00	25	25.00	*****
Proco	- Primario completo	89	89.00	91	91.00	*****
Sein	- Secundario incom	67	67.00	69	69.00	*****
Seco	- Secundario completo	77	77.00	78	78.00	*****
Tein	- Terciario incompleto	13	13.00	15	15.00	***
Teco	- Terciario completo	25	25.00	27	27.00	*****
Unin	- Universitario incompl	17	17.00	17	17.00	****
Unco	- Universitario compl	5	5.00			=== VENTILEE ===
25_	- reponse manquante	6	6.00			=== VENTILEE ===
26 . Nivel de ingresos						
300	- Menos de 300	49	49.00	49	49.00	*****
400	- De 300 a 500	95	95.00	96	96.00	*****
600	- De 501 a 700	68	68.00	69	69.00	*****
850	- De 701 a 1000	45	45.00	46	46.00	*****
1250	- De 1001 a 1500	28	28.00	30	30.00	*****
2000	- De 1501 a 2500	19	19.00	19	19.00	****
3250	- De 2501 a 4000	6	6.00			=== VENTILEE ===
4000	- Más de 4000	2	2.00			=== VENTILEE ===
Nsnc	- Ns/Nc	10	10.00	13	13.00	***
Número de personas en el hogar						
Una	- Una persona	10	10.00	10	10.00	**
Dos	- Dos personas	40	40.00	40	40.00	*****
3a4	- De 3 a 4 personas	145	145.00	145	145.00	*****
5a6	- De 5 a 6 personas	84	84.00	84	84.00	*****
+7	- Más de 7 personas	43	43.00	43	43.00	*****
34 . Nivel vivienda						
C1	- Media Superior	29	29.00	29	29.00	*****
C2	- Media media	114	114.00	114	114.00	*****
D1	- Media baja	118	118.00	118	118.00	*****
D2	- Muy modesta/Marginal	61	61.00	61	61.00	*****

El análisis factorial de correspondencias de la tabla de códigos condensada tiene por objetivo, nuevamente,

- Hallar semejanza de individuos
- Estudiar relación entre variables
- Resumir las características observadas en un pequeño número de variables
- Comparar modalidades de diferentes variables

La representación de los individuos y de las modalidades se realiza de igual manera a como lo apreciábamos en el primer análisis en un espacio de  $np$  dimensiones.

Esto no puede observarse gráficamente y es necesario, para continuar con el análisis referirlo a otro sistema de coordenadas de manera tal que el primer plano (las coordenadas de los dos primeros ejes) brinde la mejor representación de la tabla de datos.

#### *Centro de gravedad*

Recordemos que para esto es necesario, en primer término, conocer el centro de gravedad de cada nube de puntos y se realiza de igual manera que en el primer análisis.

#### *Inercia*

Una vez conocido el centro de gravedad es necesario referir las coordenadas de todos los individuos y de todas las modalidades con su respectivo centro de gravedad.

- De esta manera obtenemos las denominadas matrices de inercia.

La inercia mide la variabilidad o dispersión de una nube de puntos respecto de un punto en común: el baricentro o centro de gravedad de la nube.

La matriz de inercia se obtiene al centrar todas las coordenadas de las modalidades (obtener la distancia respecto del baricentro). Al diagonalizar la matriz de inercia hallaremos los valores propios de la matriz que dan origen a las direcciones principales de la nube de puntos. Es indistinto hallar las direcciones principales de la matriz de inercia de las modalidades o de los individuos, en general se trabaja con la de modalidades.

#### *El cálculo de la inercia*

Si observamos en la Tabla 39 veremos que la traza de la matriz de inercia antes de la diagonalización es igual a la suma de los valores propios: 2.7037, siendo la suma de los valores propios la inercia total de la nube. Precisamente descomponer la inercia implica encontrar los valores propios.

El método de diagonalización asegura que el primer valor propio es el más importante, por ende el factor asociado a él será la dirección principal de la nube de puntos.

La tabla informa:

- el valor asumido por el valor propio, 0.4730, y
- el peso que tiene en el total, 17.50.

La columna de porcentaje acumulado indica cuánta información estoy representando al incorporar un eje adicional en el análisis. Con los dos primeros valores propios obtengo el plano factorial en el cual tengo representada el 27.73% del total de la inercia. En la tabla se muestra el histograma de los 73 primeros valores propios, para este último se alcanza el 100% de la información.

Cuántos ejes conservo en el análisis? Los 22 primeros valores propios. Ahora bien, la información que se agrega a partir del tercer eje es despreciable, en general se analiza el plano y puede en algún caso estudiarse el tercer eje. El valor asumido por el valor propio indica la inercia contenida en ese eje. Además, hay una diferencia importante entre el segundo y el tercer valor propio, el decrecimiento en los valores de este son muy pronunciados.



Tabla 39. Valores Propios

APERCU DE LA PRECISION DES CALCULS : TRACE AVANT DIAGONALISATION .. 2.7037				
SOMME DES VALEURS PROPRES ..... 2.7037				
HISTOGRAMME DES 73 PREMIERES VALEURS PROPRES				
NUMERO	VALEUR PROPRE	POURCENT.	POURCENT. CUMULE	
1	0.4730	17.50	17.50	*****
2	0.2767	10.24	27.73	*****
3	0.1035	3.83	31.56	*****
4	0.0895	3.31	34.87	*****
5	0.0718	2.66	37.52	*****
6	0.0690	2.55	40.08	*****
7	0.0612	2.26	42.34	*****
8	0.0594	2.20	44.54	*****
9	0.0528	1.95	46.49	*****
10	0.0523	1.93	48.42	*****
11	0.0512	1.89	50.32	*****
12	0.0509	1.88	52.20	*****
13	0.0497	1.84	54.03	*****
14	0.0471	1.74	55.77	*****
15	0.0459	1.70	57.47	*****
16	0.0443	1.64	59.11	*****
17	0.0433	1.60	60.71	*****
18	0.0425	1.57	62.28	*****
19	0.0406	1.50	63.78	*****
20	0.0400	1.48	65.26	*****
21	0.0393	1.45	66.72	*****
22	0.0378	1.40	68.12	*****
23	0.0369	1.36	69.48	*****
24	0.0361	1.33	70.82	*****
25	0.0358	1.32	72.14	*****
26	0.0336	1.24	73.38	*****
27	0.0323	1.19	74.58	*****
28	0.0318	1.18	75.75	*****
29	0.0314	1.16	76.91	*****
30	0.0298	1.10	78.02	*****
31	0.0290	1.07	79.09	*****
32	0.0274	1.01	80.10	*****
33	0.0265	0.98	81.09	*****
34	0.0253	0.94	82.02	*****
35	0.0242	0.90	82.92	*****
36	0.0238	0.88	83.80	*****
37	0.0227	0.84	84.64	****
38	0.0223	0.82	85.46	****
39	0.0213	0.79	86.25	****
40	0.0211	0.78	87.03	****
41	0.0205	0.76	87.79	****
42	0.0200	0.74	88.53	****
43	0.0196	0.72	89.25	****
44	0.0180	0.67	89.92	****
45	0.0172	0.64	90.55	***
46	0.0167	0.62	91.17	***
47	0.0158	0.59	91.75	***
48	0.0153	0.57	92.32	***
49	0.0146	0.54	92.86	***
50	0.0136	0.50	93.36	***
51	0.0133	0.49	93.85	***
52	0.0130	0.48	94.33	***
53	0.0121	0.45	94.78	**
54	0.0118	0.44	95.22	**
55	0.0112	0.41	95.63	**
56	0.0108	0.40	96.03	**
57	0.0103	0.38	96.41	**
58	0.0100	0.37	96.78	**
59	0.0092	0.34	97.12	**
60	0.0089	0.33	97.45	**
61	0.0083	0.31	97.76	**
62	0.0082	0.30	98.07	**
63	0.0074	0.27	98.34	**
64	0.0067	0.25	98.59	**
65	0.0060	0.22	98.81	**
66	0.0056	0.21	99.01	*
67	0.0053	0.20	99.21	*
68	0.0051	0.19	99.40	*
69	0.0046	0.17	99.57	*
70	0.0035	0.13	99.70	*
71	0.0032	0.12	99.82	*
72	0.0030	0.11	99.93	*
73	0.0020	0.07	100.00	*

Recodemos las ventajas del nuevo referencial

- Permite analizar completamente la forma de la nube de puntos perfiles
- Permite la representación de la forma de la nube de puntos perfiles cualquiera que sean las dimensiones de las nubes de puntos
- Permite producir una representación objetiva de esas nubes de puntos independientes del punto de vista del análisis

*b. El plano de los dos primeros factores*

*Las variables activas*

Los puntos en el espacio que representan las modalidades de las variables utilizadas para el análisis deben proyectarse sobre las direcciones principales halladas al descomponer la inercia.

Cada modalidad tiene un peso relativo en la nube de puntos que viene dado por la cantidad de individuos que se identifican con esa modalidad respecto del producto del total de observaciones y la cantidad de variables que componen la nube de puntos.

Las coordenadas indican la representación de cada modalidad sobre el nuevo sistema de referencia. Cuánto más alto es el valor de las coordenadas significa que más alejado se encuentra el punto del origen del sistema y cuánto más cerca se encuentre del eje factorial más contribuirá a su formación.

Así la modalidad lee de la variable deportes tiene sobre el primer eje factorial coordenada igual a  $-0.47$ , está muy cercana al origen del sistema y contribuye muy poco a la formación del eje, mientras que la modalidad no conoce de la sección información gremial tiene coordenada  $1.63$  y la contribución es de  $4.8$  la cual es muy alta si tenemos en cuenta que la variable información gremial contribuye con el  $6.4\%$  a la formación del factor 1.

La contribución de la variable a la formación del factor surge de la suma de la contribución individual de cada modalidad perteneciente a la variable. Luego la suma de todas las contribuciones acumuladas de las variables consideradas es igual a  $100$ , es decir, el  $100\%$ .

El coseno cuadrado mide la calidad de representación de la distancia del punto  $i$  al origen del eje factorial. Medir la calidad de representación significa conocer si el punto individuo o modalidad está bien representado. El coseno cuadrado varía entre  $0$  y  $1$ . Cuanto mayor sea el coseno cuadrado mejor representada tendrá a esa modalidad en ese eje.

Con todos estos elementos vamos ahora a leer la Tabla 40

En primer lugar buscamos aquellas variables que tengan contribución acumulada más alta en cada eje porque significa que son las que más aportan a su construcción.

Para el eje factorial 1 podemos separar Clasificados (contribución  $5.6$ ), Noticias de violencia, sexo y drogas ( $5.2$ ), Información sobre corrupción política y social ( $4.9$ ), Información gremial ( $6.4$ ), Noticias agropecuarias ( $5.8$ ), Noticias regionales ( $5.4$ ). Estas variables son las que más peso ejercen sobre el eje, y son en las que se concentra la mayor cantidad de inercia.

- Nótese que las modalidades de mayor peso dentro de cada variable son "no conoce la sección" ubicadas en el semieje positivo, y "lee la sección" en el semieje negativo.





socioeconómica, medida a través de la vivienda, baja, en el semieje positivo.

Puede observarse en el gráfico que se agrupan las modalidades que indican a individuos que leen secciones del diario y cercana a ellas los que no leen pero sí conocen que existen, más alejadas se encuentran aquellos que no conocen las secciones del diario. De igual manera, el alto nivel socioeconómico, medido por la vivienda, nivel de ingreso y nivel educativo se asocia a la no lectura de ciertas secciones del diario lo cual lleva a pensar en una lectura mucho más selectiva. Por el contrario, el bajo nivel de ingresos, estudios y socioeconómico se relaciona con la lectura indiscriminada.

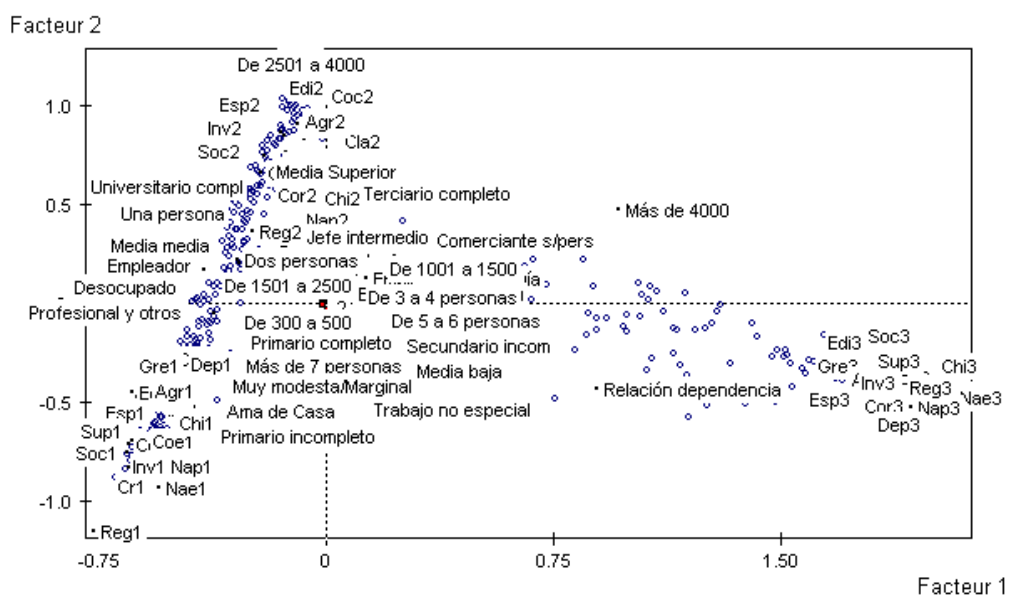
#### *Grupos de secciones*

Algunas reagrupaciones son visibles sobre el primer plano, en particular el de comentarios científicos, información gremial, notas editoriales, comentarios económicos y políticos, comentarios regionales, noticias agropecuarias, investigación, espectáculos y sociales. Las modalidades no conocida de estas secciones se sitúan todas en el segundo cuadrante del plano, la modalidad no leída en el cuarto cuadrante y la modalidad no leída mayormente en el segundo cuadrante.

#### *Las modalidades mal representadas en el primer plano*

La calidad de representación de la modalidad no leída de la sección policiales y deportes es muy baja. Las variables de caracterización tienen, todas sus modalidades, mala representación en el primer plano. El plano formado por el eje 2 y el eje 3 reúne una mejor representación.

El factor 1 el conocimiento o no de las secciones del diario, mientras que el eje 2 opone la lectura de la sección a la no lectura. Si tuviéramos que darle un nombre diríamos que el eje 1 es "conocer" y el eje 2 es "leer".



En la Tabla 41 tenemos los valores test o valores de prueba.

Se observa en la tabla que para la mayoría de las variables secciones del diario tienen buenos valores tests en los dos primeros ejes lo cual habla de una muy buena proyección de la nube de puntos sobre el plano factorial. Por su parte las variables de caracterización tiene valores test que pasan la prueba de hipótesis recién en el tercer y cuarto eje.

MODALITES			VALEURS-TEST					COORDONNEES					DISTO.
IDEN - LIBELLE	EFF.	P.ABS	1	2	3	4	5	1	2	3	4	5	
1 . Sexe													
Sex1 - 1	162	162.00	-0.2	0.3	-10.6	-2.4	-9.5	-0.01	0.01	-0.59	-0.13	-0.53	0.99
Sex2 - 2	160	160.00	0.2	-0.3	10.6	2.4	9.5	0.01	-0.01	0.60	0.13	0.54	1.01
3 . Noticias Internacionales													
Int1 - Int1	128	128.00	-7.9	-7.8	3.7	1.1	-1.9	-0.54	-0.53	0.25	0.08	-0.13	1.52
Int2 - Int2	167	167.00	1.5	9.0	-3.5	-0.9	2.6	0.08	0.49	-0.19	-0.05	0.14	0.93
Int3 - Int3	27	27.00	11.2	-2.5	-0.2	-0.3	-1.4	2.07	-0.47	-0.04	-0.06	-0.25	10.93
4 . Clasificados													
Clal1 - Clal1	175	175.00	-10.0	-5.4	-1.5	1.4	2.3	-0.51	-0.28	-0.08	0.07	0.12	0.84
Clal2 - Clal2	91	91.00	-1.5	8.6	1.5	-2.5	-1.7	-0.13	0.76	0.14	-0.23	-0.15	2.54
Clal3 - Clal3	56	56.00	14.9	-3.1	0.2	1.1	-1.0	1.81	-0.37	0.02	0.14	-0.12	4.75
5 . Noticias de Violencia, sexo, droga													
Vio1 - Vio1	124	124.00	-8.3	-6.3	-3.2	6.1	-1.5	-0.59	-0.44	-0.23	0.43	-0.10	1.60
Vio2 - Vio2	146	146.00	-2.4	8.2	2.9	-6.9	1.3	-0.15	0.50	0.18	-0.43	0.08	1.21
Vio3 - Vio3	52	52.00	14.2	-2.8	0.3	1.3	0.2	1.81	-0.35	0.04	0.17	0.03	5.19
6 . Noticias policiales													
Pol1 - Pol1	91	91.00	-6.2	-5.5	-5.1	6.9	-0.9	-0.55	-0.49	-0.45	0.62	-0.08	2.54
Pol2 - Pol2	208	208.00	0.2	6.5	5.3	-8.3	1.5	0.01	0.27	0.22	-0.34	0.06	0.55
Pol3 - Pol3	23	23.00	10.5	-2.3	-1.0	3.4	-1.3	2.10	-0.46	-0.20	0.68	-0.25	13.00
7 . Noticias nacionales económicas													
Nae1 - Nae1	74	74.00	-5.4	-9.1	5.7	4.3	-4.9	-0.55	-0.93	0.58	0.44	-0.50	3.35
Nae2 - Nae2	228	228.00	-0.1	9.6	-5.6	-5.3	5.5	0.00	0.34	-0.20	-0.19	0.20	0.41
Nae3 - Nae3	20	20.00	9.5	-2.2	0.7	2.5	-1.9	2.07	-0.47	0.16	0.54	-0.40	15.10
8 . Noticias nacionales políticas													
Nap1 - Nap1	90	90.00	-6.5	-8.7	5.9	3.1	-4.3	-0.58	-0.78	0.53	0.28	-0.38	2.58
Nap2 - Nap2	207	207.00	0.4	9.7	-5.6	-3.7	5.6	0.02	0.40	-0.23	-0.15	0.23	0.56
Nap3 - Nap3	25	25.00	10.0	-2.8	0.1	1.4	-2.8	1.93	-0.53	0.02	0.28	-0.54	11.88
9 . Información sobre corrupción política y social													
Cor1 - Cor1	104	104.00	-8.0	-8.8	2.6	2.5	-2.1	-0.65	-0.71	0.21	0.21	-0.17	2.10
Cor2 - Cor2	172	172.00	-1.9	10.6	-2.4	-2.6	2.9	-0.10	0.55	-0.12	-0.14	0.15	0.87
Cor3 - Cor3	46	46.00	13.4	-3.4	-0.2	0.3	-1.3	1.83	-0.46	-0.02	0.05	-0.18	6.00
10 . Deportes													
Dep1 - Dep1	122	122.00	-6.5	-4.3	4.5	1.6	7.9	-0.47	-0.31	0.32	0.11	0.56	1.64
Dep2 - Dep2	174	174.00	0.8	5.6	-4.5	-1.8	-8.7	0.04	0.29	-0.23	-0.09	-0.45	0.85
Dep3 - Dep3	26	26.00	10.1	-2.6	0.2	0.4	2.0	1.91	-0.48	0.05	0.08	0.37	11.38
11 . Comentarios Científicos													
Coc1 - Coc1	166	166.00	-11.4	-8.8	-0.5	-0.1	1.0	-0.62	-0.48	-0.02	0.00	0.06	0.94
Coc2 - Coc2	97	97.00	0.0	12.3	0.4	1.0	-2.2	0.00	1.05	0.03	0.09	-0.19	2.32
Coc3 - Coc3	59	59.00	14.8	-3.2	0.1	-1.1	1.3	1.74	-0.38	0.01	-0.13	0.16	4.46
12 . Información Gremial													
Gre1 - Gre1	179	179.00	-12.2	-7.0	-0.9	0.7	1.6	-0.61	-0.35	-0.04	0.03	0.08	0.80
Gre2 - Gre2	69	69.00	-1.6	11.6	0.8	-0.4	-2.9	-0.17	1.24	0.09	-0.05	-0.31	3.67
Gre3 - Gre3	74	74.00	16.0	-3.1	0.2	-0.4	0.9	1.63	-0.31	0.02	-0.04	0.09	3.35
13 . Notas Editoriales													
Edi1 - Edi1	168	168.00	-11.9	-8.5	-1.0	-1.9	1.2	-0.63	-0.45	-0.05	-0.10	0.06	0.92
Edi2 - Edi2	85	85.00	-1.0	12.2	0.9	2.6	-2.8	-0.10	1.14	0.09	0.24	-0.26	2.79
Edi3 - Edi3	69	69.00	15.6	-2.8	0.2	-0.4	1.6	1.66	-0.30	0.02	-0.05	0.17	3.67
14 . Comentarios económicos y políticos													
Coe1 - Coe1	141	141.00	-9.3	-11.0	1.0	-1.9	-1.1	-0.59	-0.69	0.06	-0.12	-0.07	1.28
Coe2 - Coe2	129	129.00	-1.4	13.3	-0.9	1.3	1.1	-0.09	0.91	-0.06	0.09	0.08	1.50
Coe3 - Coe3	52	52.00	14.4	-2.9	-0.2	0.8	0.0	1.83	-0.37	-0.03	0.10	0.00	5.19
15 . Comentarios regionales													
Cor1 - Cr1	103	103.00	-8.0	-10.2	-0.6	0.0	-0.2	-0.65	-0.83	-0.05	0.00	-0.02	2.13
Cor2 - Cr2	162	162.00	-3.8	11.9	0.1	-0.1	0.8	-0.21	0.66	0.01	-0.01	0.04	0.99
Cor3 - Cr3	57	57.00	14.7	-3.1	0.6	0.2	-0.8	1.77	-0.37	0.07	0.02	-0.09	4.65
16 . Noticias Agropecuarias													
Agr1 - Agr1	154	154.00	-10.1	-7.6	2.1	1.6	-0.3	-0.59	-0.44	0.12	0.09	-0.02	1.09
Agr2 - Agr2	106	106.00	-1.8	10.9	-3.1	-1.7	-0.2	-0.15	0.87	-0.25	-0.14	-0.01	2.04
Agr3 - Agr3	62	62.00	15.0	-3.3	1.1	0.0	0.6	1.71	-0.38	0.13	0.00	0.07	4.19
17 . Noticias Regionales													
Reg1 - Reg1	52	52.00	-6.0	-9.1	-2.3	2.5	1.2	-0.77	-1.16	-0.29	0.32	0.15	5.19
Reg2 - Reg2	221	221.00	-6.4	9.7	2.6	-1.9	-1.5	-0.24	0.37	0.10	-0.07	-0.05	0.46
Reg3 - Reg3	49	49.00	14.5	-3.2	-1.1	-0.1	0.7	1.90	-0.42	-0.14	-0.01	0.09	5.57
18 . Investigación periodística													
Inv1 - Inv1	126	126.00	-9.4	-10.8	-1.2	-1.4	0.0	-0.66	-0.75	-0.08	-0.10	0.00	1.56
Inv2 - Inv2	138	138.00	-2.1	13.1	1.9	1.7	-0.3	-0.13	0.84	0.12	0.11	-0.02	1.33
Inv3 - Inv3	58	58.00	14.6	-3.1	-0.9	-0.4	0.4	1.74	-0.37	-0.11	-0.05	0.04	4.55



c. Clasificación jerárquica

En la Tabla 42 tenemos el resultado de los 50 nudos, o agrupaciones, más importantes de la clasificación jerárquica realizada sobre los 10 primeros ejes factoriales aplicando el método del vecino más próximo.

Por ejemplo, el nudo 629 se forma con el nudo 589 (primogénito) y el 577 (benjamín) quienes reúnen 17 individuos. El nudo 631 (64 individuos) está compuesto por el 624 (36 individuos) y el 623 (28 individuos). Luego el 629 y el 631 forman el 635 (81 individuos). Sucesivamente se van agrupando de acuerdo a la similitud existente entre ellos hasta llegar al nudo 643 que reúne a los 322 individuos.

Tabla 42. Clasificación de la nube de puntos

CLASSIFICATION HIERARCHIQUE (VOISINS RECIPROQUES) SUR LES 10 PREMIERS AXES FACTORIELS						
DESCRIPTION DES 50 NOEUDS D'INDICES LES PLUS ELEVES						
NUM.	AINE	BENJ	EFF.	POIDS	INDICE	HISTOGRAMME DES INDICES DE NIVEAU
594	466	494	4	4.00	0.00292	*
595	540	522	9	9.00	0.00297	*
596	573	497	17	17.00	0.00304	*
597	490	180	3	3.00	0.00305	*
598	591	481	10	10.00	0.00306	*
599	536	584	15	15.00	0.00341	*
600	557	539	6	6.00	0.00362	*
601	460	574	7	7.00	0.00379	*
602	530	578	12	12.00	0.00385	*
603	576	600	11	11.00	0.00411	*
604	593	560	23	23.00	0.00424	*
605	548	587	13	13.00	0.00432	*
606	596	553	23	23.00	0.00442	*
607	504	558	7	7.00	0.00462	*
608	583	594	12	12.00	0.00470	*
609	562	572	16	16.00	0.00482	*
610	595	585	20	20.00	0.00485	*
611	563	538	16	16.00	0.00514	*
612	581	586	5	5.00	0.00536	**
613	602	575	21	21.00	0.00563	**
614	610	542	24	24.00	0.00582	**
615	555	597	9	9.00	0.00583	**
616	551	599	21	21.00	0.00602	**
617	601	566	13	13.00	0.00622	**
618	588	579	25	25.00	0.00623	**
619	580	605	16	16.00	0.00655	**
620	607	500	9	9.00	0.00770	**
621	615	614	33	33.00	0.00796	**
622	613	570	31	31.00	0.00823	**
623	609	608	28	28.00	0.00868	**
624	606	582	36	36.00	0.00933	**
625	618	590	36	36.00	0.00935	**
626	517	619	21	21.00	0.00947	**
627	603	598	21	21.00	0.00981	**
628	604	550	30	30.00	0.01073	***
629	589	577	17	17.00	0.01128	***
630	616	611	37	37.00	0.01185	***
631	624	623	64	64.00	0.01221	***
632	621	612	38	38.00	0.01278	***
633	625	592	41	41.00	0.01368	***
634	620	626	30	30.00	0.01459	***
635	629	631	81	81.00	0.01681	****
636	633	630	78	78.00	0.01894	****
637	632	617	51	51.00	0.02026	****
638	636	628	108	108.00	0.02723	*****
639	635	622	112	112.00	0.03943	*****
640	639	634	142	142.00	0.04437	*****
641	637	627	72	72.00	0.04757	*****
642	638	640	250	250.00	0.19809	*****
643	641	642	322	322.00	0.42318	*****
SOMME DES INDICES DE NIVEAU = 1.30921						

*Coordenadas y valores test*

Cada nudo formado, y cada individuo, es sometido a una dócima a efectos de conocer su aleatoriedad. El tener valor test muy alto nos conduce, como en casos anteriores, a rechazar la hipótesis de aleatoriedad. Este análisis se realiza a los elementos ubicados en la base del dendograma y puede observarse en la Tabla 43.



Tabla 43. Coordenadas y valores test de la base del dendograma

ELEMENTS			VALEURS-TEST					COORDONNEES					
NUM .	IDENT	POIDS	EFF	1	2	3	4	5	1	2	3	4	5
1 .	587	9.00	9	-0.82	2.45	-2.22	-2.15	-0.53	-0.27	0.81	-0.73	-0.71	-0.18
2 .	548	4.00	4	-0.89	1.27	-2.50	0.69	-1.18	-0.44	0.63	-1.25	0.35	-0.59
3 .	580	3.00	3	-0.64	2.12	-4.65	1.43	-0.18	-0.37	1.22	-2.67	0.82	-0.10
4 .	517	5.00	5	-0.83	2.36	-0.49	3.01	4.07	-0.37	1.05	-0.22	1.34	1.81
5 .	500	2.00	2	1.28	-0.05	-3.36	0.94	2.58	0.91	-0.03	-2.37	0.67	1.82
6 .	558	4.00	4	-1.26	-0.27	-3.13	2.81	4.50	-0.63	-0.13	-1.56	1.40	2.24
7 .	504	3.00	3	-1.38	-1.41	-4.25	1.91	0.21	-0.79	-0.81	-2.45	1.10	0.12
8 .	570	10.00	10	-1.38	1.23	3.48	-0.52	4.14	-0.43	0.38	1.09	-0.16	1.29
9 .	575	9.00	9	-1.32	1.72	1.61	1.80	-2.01	-0.43	0.57	0.53	0.59	-0.66
10 .	578	8.00	8	-1.74	-0.66	5.08	1.50	-1.17	-0.61	-0.23	1.78	0.53	-0.41
11 .	530	4.00	4	-1.16	-0.06	3.31	3.82	-0.82	-0.58	-0.03	1.65	1.90	-0.41
12 .	594	4.00	4	-0.72	1.99	-1.48	1.67	-1.31	-0.36	0.99	-0.74	0.83	-0.65
13 .	583	8.00	8	-0.78	3.80	-0.71	2.55	3.04	-0.27	1.33	-0.25	0.89	1.06
14 .	572	11.00	11	-1.07	4.13	2.53	0.06	1.22	-0.32	1.23	0.75	0.02	0.36
15 .	562	5.00	5	-0.76	2.19	1.55	3.54	-0.34	-0.34	0.97	0.69	1.57	-0.15
16 .	582	13.00	13	-1.07	2.76	-1.20	-3.00	-1.46	-0.29	0.75	-0.33	-0.82	-0.40
17 .	553	6.00	6	-0.53	3.44	-1.69	-0.05	-0.55	-0.21	1.39	-0.68	-0.02	-0.22
18 .	497	6.00	6	-0.55	3.71	-1.47	0.39	-1.10	-0.22	1.50	-0.59	0.16	-0.44
19 .	573	11.00	11	-0.55	5.71	-0.63	0.96	-4.20	-0.16	1.69	-0.19	0.28	-1.25
20 .	577	5.00	5	-0.45	3.00	0.42	-3.92	0.26	-0.20	1.33	0.19	-1.74	0.11
21 .	589	12.00	12	-0.70	5.32	4.48	-2.38	-0.55	-0.20	1.51	1.27	-0.68	-0.16
22 .	550	7.00	7	-2.22	-3.55	1.84	4.55	-1.57	-0.83	-1.33	0.69	1.70	-0.59
23 .	560	9.00	9	-2.49	-3.82	3.39	-1.08	-0.31	-0.82	-1.26	1.11	-0.36	-0.10
24 .	593	14.00	14	-3.18	-4.86	3.87	0.22	2.37	-0.83	-1.27	1.01	0.06	0.62
25 .	538	5.00	5	-1.56	-1.80	0.96	-2.24	-2.47	-0.70	-0.80	0.42	-0.99	-1.10
26 .	563	11.00	11	-2.97	-4.66	-1.82	0.96	-2.60	-0.88	-1.38	-0.54	0.29	-0.77
27 .	584	11.00	11	-2.26	-2.41	-4.34	1.55	0.10	-0.67	-0.72	-1.29	0.46	0.03
28 .	536	4.00	4	-1.66	-2.52	-1.35	2.08	0.76	-0.83	-1.26	-0.67	1.04	0.38
29 .	551	6.00	6	-1.97	-3.43	-2.07	0.06	-3.62	-0.80	-1.39	-0.84	0.02	-1.47
30 .	592	5.00	5	-1.87	-2.60	-1.06	-3.61	1.11	-0.83	-1.16	-0.47	-1.61	0.49
31 .	590	11.00	11	-1.87	-0.85	-0.92	-2.86	-2.64	-0.55	-0.25	-0.27	-0.85	-0.78
32 .	579	10.00	10	-2.00	-1.71	1.21	-1.69	4.39	-0.62	-0.53	0.38	-0.53	1.37
33 .	588	15.00	15	-2.51	-2.58	-3.68	-4.64	1.02	-0.63	-0.65	-0.93	-1.17	0.26
34 .	481	2.00	2	2.44	0.10	-0.39	0.46	-1.01	1.72	0.07	-0.28	0.33	-0.71
35 .	591	8.00	8	6.27	-1.61	0.68	3.33	0.43	2.19	-0.56	0.24	1.17	0.15
36 .	539	2.00	2	3.31	-0.83	0.04	2.10	0.74	2.34	-0.59	0.03	1.48	0.52
37 .	557	4.00	4	4.18	-1.22	-0.08	1.32	-1.45	2.08	-0.61	-0.04	0.66	-0.72
38 .	576	5.00	5	5.02	-1.93	-0.11	-1.42	-2.38	2.23	-0.86	-0.05	-0.63	-1.06
39 .	566	6.00	6	5.78	-1.51	-1.69	0.40	-0.78	2.34	-0.61	-0.68	0.16	-0.32
40 .	574	5.00	5	4.69	-0.85	-0.59	-1.64	1.93	2.09	-0.38	-0.26	-0.73	0.86
41 .	460	2.00	2	3.33	-0.94	-0.93	-2.78	0.75	2.35	-0.66	-0.66	-1.96	0.53
42 .	586	2.00	2	1.89	0.24	0.30	1.41	-2.16	1.34	0.17	0.21	0.99	-1.52
43 .	581	3.00	3	2.01	0.46	0.03	2.76	1.03	1.15	0.26	0.02	1.59	0.60
44 .	542	4.00	4	1.95	0.84	-0.11	-1.91	3.05	0.97	0.42	-0.05	-0.95	1.52
45 .	585	11.00	11	4.79	-0.48	-0.25	0.07	2.03	1.42	-0.14	-0.08	0.02	0.60
46 .	522	3.00	3	1.73	0.35	0.32	-2.43	-1.31	1.00	0.20	0.19	-1.40	-0.75
47 .	540	6.00	6	3.55	-0.31	0.72	-1.50	-1.62	1.44	-0.13	0.29	-0.61	-0.66
48 .	271	1.00	1	1.54	-0.64	0.69	1.87	-2.14	1.54	-0.64	0.69	1.87	-2.14
49 .	490	2.00	2	1.66	-0.87	1.39	0.03	-2.21	1.18	-0.62	0.98	0.02	-1.56
50 .	555	6.00	6	4.11	-1.49	2.93	-2.03	2.06	1.66	-0.60	1.19	-0.82	0.84

### Dendograma

Recordemos que el *dendograma* es una representación gráfica del proceso de agrupamiento que sufren los individuos. Cuando tenemos muchos puntos, como en este caso, solo se observa gráficamente los grupos más representativos, es decir, las últimas 50 formaciones.

En la [Tabla 44](#) se describe la formación de cada nudo del dendograma. De acuerdo a la figura del dendograma podemos deducir que la partición óptima puede ser la de tres clases.

Tabla 44. Descripción de los nodos de la jerarquía

(INDICES EN POURCENTAGE DE LA SOMME DES INDICES : 1.10920)

NOEUD NUMERO	INDICE	SUCSESSEURS		EFFECT.	POIDS	COMPOSITION	
		AINE	BENJ			PREMIER	DERNIER
51	0.27	47	46	9	9.00	46	47
52	0.27	19	18	17	17.00	18	19
53	0.28	49	48	3	3.00	48	49
54	0.28	35	34	10	10.00	34	35
55	0.31	28	27	15	15.00	27	28
56	0.33	37	36	6	6.00	36	37
57	0.34	41	40	7	7.00	40	41
58	0.35	11	10	12	12.00	10	11
59	0.37	38	56	11	11.00	36	38
60	0.38	24	23	23	23.00	23	24
61	0.39	2	1	13	13.00	1	2
62	0.40	52	17	23	23.00	17	19
63	0.42	7	6	7	7.00	6	7

NOEUD		SUCESSEURS				COMPOSITION	
NUMERO	INDICE	AINE	BENJ	EFFECT.	POIDS	PREMIER	DERNIER
64	0.42	13	12	12	12.00	12	13
65	0.43	15	14	16	16.00	14	15
66	0.44	51	45	20	20.00	45	47
67	0.46	26	25	16	16.00	25	26
68	0.48	43	42	5	5.00	42	43
69	0.51	58	9	21	21.00	9	11
70	0.53	66	44	24	24.00	44	47
71	0.53	50	53	9	9.00	48	50
72	0.54	29	55	21	21.00	27	29
73	0.56	57	39	13	13.00	39	41
74	0.56	33	32	25	25.00	32	33
75	0.59	3	61	16	16.00	1	3
76	0.69	63	5	9	9.00	5	7
77	0.72	71	70	33	33.00	44	50
78	0.74	69	8	31	31.00	8	11
79	0.78	65	64	28	28.00	12	15
80	0.84	62	16	36	36.00	16	19
81	0.84	74	31	36	36.00	31	33
82	0.85	4	75	21	21.00	1	4
83	0.88	59	54	21	21.00	34	38
84	0.97	60	22	30	30.00	22	24
85	1.02	21	20	17	17.00	20	21
86	1.07	72	67	37	37.00	25	29
87	1.10	80	79	64	64.00	12	19
88	1.15	77	68	38	38.00	42	50
89	1.23	81	30	41	41.00	30	33
90	1.32	76	82	30	30.00	1	7
91	1.52	85	87	81	81.00	12	21
92	1.71	89	86	78	78.00	25	33
93	1.83	88	73	51	51.00	39	50
94	2.45	92	84	108	108.00	22	33
95	3.55	91	78	112	112.00	8	21
96	4.00	95	90	142	142.00	1	21
97	4.29	93	83	72	72.00	34	50
98	17.86	94	96	250	250.00	1	33
99	38.15	97	98	322	322.00	1	50

#### d. Partición de la nube de puntos

Cuando se realiza la partición encontramos, en primer lugar, la cantidad de individuos que forman parte de cada clase.

Así nos encontramos que:

La clase 1 la forman 142 individuos que corresponden a los nudos 1 a 21, en la Figura son los primeros 21 nudos identificados en la base y en la [Tabla 45](#) los primeros 21 números.

La segunda clase reúne a 108 individuos ubicados desde la posición 22 a la 33.

La tercera clase corresponde a los 72 individuos reunidos desde la posición 34 a la 50.

Luego esta partición es evaluada en términos de la **inercia** que involucra a cada clase. Al realizar la consolidación de la partición, las clases se reagrupan de manera de asegura una mejor representación en el primer plano. Esto se ve en los valores test antes de la consolidación y los valores test posteriores a la consolidación.

Obsérvese que antes de la consolidación las clases tenían buena representación en el tercer eje factorial, luego de la consolidación las tres clases se representan bien en el primer plano factorial únicamente. La inercia total es la misma (0.4745), cambia la composición y la inercia existente

dentro de cada grupo: clase 1 125 individuos, clase 2 124 individuos y clase 3 73 individuos.

Classification hierarchique directe

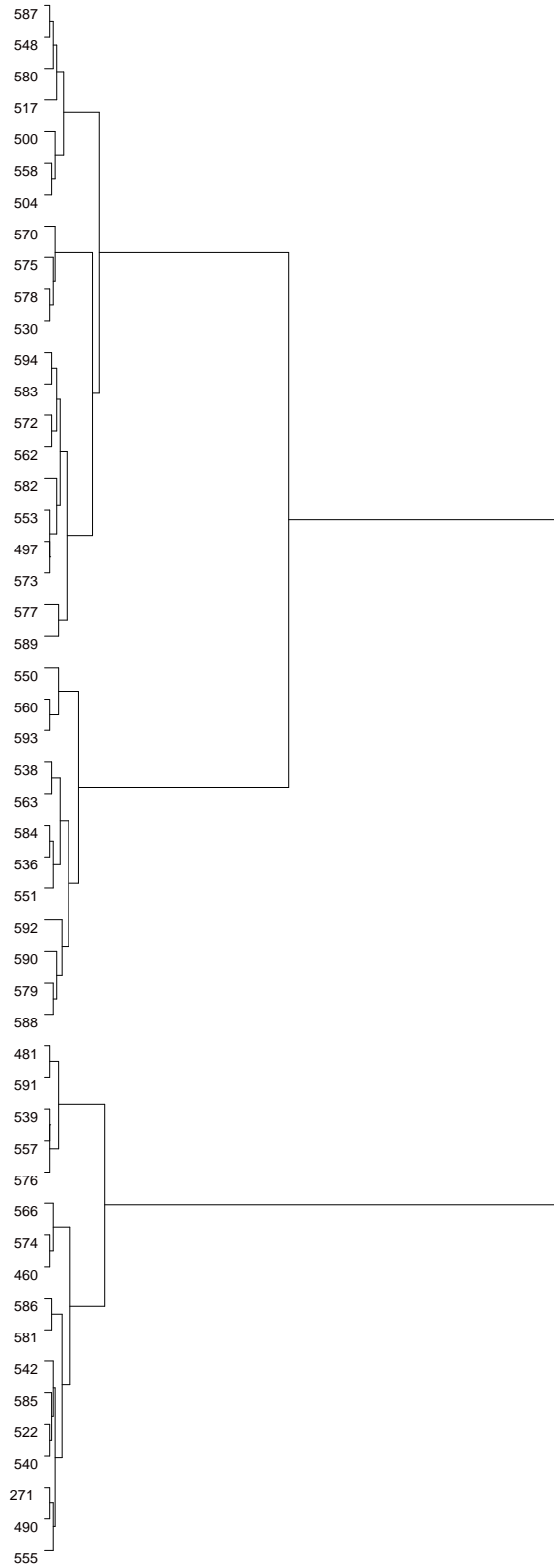


Tabla 45. Partición del cuerpo del árbol jerárquico														
COUPURE 'a' DE L'ARBRE EN 3 CLASSES														
FORMATION DES CLASSES (INDIVIDUS ACTIFS)														
DESCRIPTION SOMMAIRE														
CLASSE	EFFECTIF	POIDS	CONTENU											
aa1a	142	142.00	1 A 21											
aa2a	108	108.00	22 A 33											
aa3a	72	72.00	34 A 50											
COORDONNEES ET VALEURS-TEST AVANT CONSOLIDATION														
AXES 1 A 5														
CLASSES				VALEURS-TEST					COORDONNEES					
IDEN - LIBELLE	EFF.	P. ABS	1	2	3	4	5	1	2	3	4	5	DISTO.	
COUPURE 'a' DE L'ARBRE EN 3 CLASSES														
aa1a - CLASSE 1 / 3	142	142.00	-5.2	14.1	0.7	2.8	0.5	-0.22	0.47	0.01	0.05	0.01	0.27	
aa2a - CLASSE 2 / 3	108	108.00	-9.4	-12.1	-1.4	-2.7	-0.7	-0.51	-0.50	-0.04	-0.06	-0.01	0.51	
aa3a - CLASSE 3 / 3	72	72.00	16.8	-3.1	0.8	-0.2	0.1	1.20	-0.17	0.03	-0.01	0.00	1.47	
CONSOLIDATION DE LA PARTITION														
AUTOUR DES 3 CENTRES DE CLASSES, REALISEE PAR 10 ITERATIONS A CENTRES MOBILES														
PROGRESSION DE L'INERTIE INTER-CLASSES														
ITERATION	I. TOTALE	I. INTER	QUOTIENT											
0	1.30921	0.62126	0.47453											
1	1.30921	0.64725	0.49438											
2	1.30921	0.64815	0.49507											
3	1.30921	0.64821	0.49512											
4	1.30921	0.64821	0.49512											
ARRRET APRES L'ITERATION 4 L'ACCROISSEMENT DE L'INERTIE INTER-CLASSES														
PAR RAPPORT A L'ITERATION PRECEDENTE N'EST QUE DE 0.000 %.														
DECOMPOSITION DE L'INERTIE														
CALCULEE SUR 10 AXES.														
INERTIES	INERTIES		EFFECTIFS		POIDS		DISTANCES							
	AVANT	APRES	AVANT	APRES	AVANT	APRES	AVANT	APRES						
INTER-CLASSES	0.6213	0.6482												
INTRA-CLASSE														
CLASSE 1 / 3	0.3121	0.2289	142	125	142.00	125.00	0.2725	0.3566						
CLASSE 2 / 3	0.1797	0.2279	108	124	108.00	124.00	0.5144	0.4662						
CLASSE 3 / 3	0.1961	0.2041	72	73	72.00	73.00	1.4694	1.4568						
TOTALE	1.3092	1.3092												
QUOTIENT (INERTIE INTER / INERTIE TOTALE) : AVANT ... 0.4745														
APRES ... 0.4951														
COORDONNEES ET VALEURS-TEST APRES CONSOLIDATION														
AXES 1 A 5														
CLASSES				VALEURS-TEST					COORDONNEES					
IDEN - LIBELLE	EFF.	P. ABS	1	2	3	4	5	1	2	3	4	5	DISTO.	
COUPURE 'a' DE L'ARBRE EN 3 CLASSES														
aa1a - CLASSE 1 / 3	125	125.00	-4.2	15.2	0.9	1.2	-0.8	-0.20	0.56	0.02	0.02	-0.01	0.36	
aa2a - CLASSE 2 / 3	124	124.00	-10.3	-12.5	-1.0	-1.3	0.3	-0.50	-0.46	-0.02	-0.03	0.01	0.47	
aa3a - CLASSE 3 / 3	73	73.00	16.8	-3.2	0.1	0.2	0.5	1.19	-0.17	0.00	0.00	0.01	1.46	

En la [Tabla 46](#) observamos los individuos característicos de cada clase. Aquí solo se muestran los 10 principales. La condición que debe cumplir el individuo para tener esta jerarquía es tener la menor distancia respecto del centro de gravedad de la clase.

#### Descripción de la partición

En la [Tabla 47](#) se encuentra la **primer clase**. Los 125 individuos que la componen representan el 38.82% del total de individuos analizados. Se destacan por no leer las diferentes secciones del diario, son jefes de familia y pertenecen al nivel socioeconómico medio y medio superior.

La modalidad *no lee* de la variable Investigación periodística tiene 138 puntos en total, es decir, el 42.86% del total de personas entrevistadas presentan esta modalidad.

Luego de los 138 que presentan la modalidad en el total el 78.26% (108 individuos) forman parte de esta clase lo que significa que el 86.40% de los individuos de esta clase *no lee* la sección Investigación periodística.

Tabla 46. Individuos característicos de las clases

PARANGONS									
CLASSE 1/ 3									
EFFECTIF: 125									
RG	DISTANCE	IDENT.	RG	DISTANCE	IDENT.	RG	DISTANCE	IDENT.	
1	0.18277	295	2	0.18659	298	3	0.19524	448	
4	0.19904	299	5	0.20574	321	6	0.22106	391	
7	0.22509	446	8	0.22781	489	9	0.25348	183	
10	0.25455	181							
CLASSE 2/ 3									
EFFECTIF: 124									
RG	DISTANCE	IDENT.	RG	DISTANCE	IDENT.	RG	DISTANCE	IDENT.	
1	0.12285	69	2	0.22579	421	3	0.23280	268	
4	0.23372	292	5	0.24159	441	6	0.24185	371	
7	0.24905	234	8	0.25340	472	9	0.29632	390	
10	0.29780	310							
CLASSE 3/ 3									
EFFECTIF: 73									
RG	DISTANCE	IDENT.	RG	DISTANCE	IDENT.	RG	DISTANCE	IDENT.	
1	0.23975	248	2	0.30814	242	3	0.35160	86	
4	0.35455	170	5	0.39192	37	6	0.40056	40	
7	0.41066	402	8	0.41462	76	9	0.43324	244	
10	0.43402	245							

Tabla 47. Descripción de la partición. Clase 1.

V.TEST	PROBA	POURCENTAGES			MODALITES	DES VARIABLES	IDEN	POIDS
		CLA/MOD	MOD/CLA	GLOBAL	CARACTERISTIQUES			
				38.82	CLASSE 1 / 3		aala	125
12.99	0.000	78.26	86.40	42.86	2	Investigación periodística	Inv2	138
12.66	0.000	76.60	86.40	43.79	2	Sociales	Soc2	141
12.65	0.000	79.84	82.40	40.06	2	Comentarios económicos y políticos	Coe2	129
11.72	0.000	71.24	87.20	47.52	2	Suplemento y revista dominical	Sup2	153
11.68	0.000	69.14	89.60	50.31	2	Comentarios regionales	Cor2	162
11.24	0.000	95.65	52.80	21.43	2	Información Gremial	Gre2	69
11.00	0.000	88.24	60.00	26.40	2	Notas Editoriales	Edi2	85
10.85	0.000	83.51	64.80	30.12	2	Comentarios Científicos	Coc2	97
10.47	0.000	74.80	73.60	38.20	2	Información sobre Espectáculos	Esp2	123
10.14	0.000	55.66	98.40	68.63	2	Noticias Regionales	Reg2	221
9.66	0.000	76.42	64.80	32.92	2	Noticias Agropecuarias	Agr2	106
8.97	0.000	62.05	82.40	51.55	2	Chistes, Horoscopos y entretenimiento	Chi2	166
8.95	0.000	61.05	84.00	53.42	2	Información sobre corrupción política y social	Cor2	172
7.15	0.000	60.27	70.40	45.34	2	Noticias de Violencia, sexo, droga	Vio2	146
7.00	0.000	52.66	87.20	64.29	2	Noticias nacionales políticas	Nap2	207
6.67	0.000	56.29	75.20	51.86	2	Noticias Internacionales	Int2	167
6.64	0.000	50.00	91.20	70.81	2	Noticias nacionales económicas	Nae2	228
6.36	0.000	67.03	48.80	28.26	2	Clasificados	Cla2	91
4.83	0.000	48.56	80.80	64.60	2	Noticias policiales	Pol2	208
3.92	0.000	48.85	68.00	54.04	2	Deportes	Dep2	174
3.25	0.001	68.97	16.00	9.01	Media Superior	Nivel vivienda	C1	29
2.68	0.004	49.12	44.80	35.40	Media media	Nivel vivienda	C2	114
2.63	0.004	46.71	56.80	47.20	1	Es jefe de familia?	Jefs	152

En la Tabla 48 se encuentra la **segunda clase** que reúne el 38.51% de los individuos que se caracterizan por leer las diferentes secciones del diario, ser ama de casa y tener un bajo nivel de instrucción.

Estas características deben interpretarse como predominantes y no exclusivas, es decir, la mayoría de los que *leen* investigación periodística están en esta clase (el 87.10%), lo cual no significa que pueda haber en este grupo alguna persona que no lea esta sección o que no la conozca.









Tabla 50. (Continuación)

	Cor1	Cor2	Cor3	Dep1	Dep2	Dep3	Coc1	Coc2	Coc3	Gre1	Gre2	Gre3	Edi1	Edi2	Edi3	Coe1	Coe2	Coe3	Cor1	Cor2	Cor3	
Una	2	7	1	5	4	1	3	6	1	3	4	3	4	3	3	3	6	1	3	6	1	
Dos	11	28	1	18	19	3	21	14	5	24	11	5	21	15	4	18	22	0	14	26	0	
3a4	43	78	24	50	80	15	70	46	29	74	35	36	71	41	33	56	62	27	43	73	29	
5a6	25	44	15	32	46	6	46	18	20	49	12	23	42	19	23	40	25	19	28	36	20	
+7	23	15	5	17	25	1	26	13	4	29	7	7	30	7	6	24	14	5	15	21	7	
C1	4	22	3	14	14	1	13	14	2	18	8	3	11	15	3	4	21	4	6	19	4	
C2	35	73	6	43	67	4	63	41	10	68	32	14	63	38	13	52	53	9	40	66	8	
D1	36	52	30	40	60	18	56	28	34	56	18	44	59	18	41	54	34	30	33	51	34	
D2	29	25	7	25	33	3	34	14	13	37	11	13	35	14	12	31	21	9	24	26	11	
	Agr1	Agr2	Agr3	Reg1	Reg2	Reg3	Inv1	Inv2	Inv3	Esp1	Esp2	Esp3	Soc1	Soc2	Soc3	Chi1	Chi2	Chi3	Sup1	Sup2	Sup3	
Agr1	154	0	0																			
Agr2	0	106	0																			
Agr3	0	0	62																			
Reg1	48	4	0	52	0	0																
Reg2	105	94	22	0	221	0																
Reg3	1	8	40	0	0	49																
Inv1	101	20	5	47	76	3	126	0	0													
Inv2	50	76	12	4	128	6	0	138	0													
Inv3	3	10	45	1	17	40	0	0	58													
Esp1	104	37	4	47	94	4	110	32	3	145	0	0										
Esp2	48	63	12	4	107	12	15	97	11	0	123	0										
Esp3	2	6	46	1	20	33	1	9	44	0	0	54										
Soc1	93	24	4	44	75	2	96	22	3	107	13	1	121	0	0							
Soc2	59	73	9	7	128	6	28	103	10	34	98	9	0	141	0							
Soc3	2	9	49	1	18	41	2	13	45	4	12	44	0	0	60							
Chi1	86	28	9	40	75	8	86	28	9	98	18	7	92	24	7	123	0	0				
Chi2	68	71	27	11	138	17	40	104	22	46	99	21	29	115	22	0	166	0				
Chi3	0	7	26	1	8	24	0	6	27	1	6	26	0	2	31	0	0	33				
Sup1	92	21	4	44	68	5	90	21	6	99	14	4	92	21	4	93	24	0	117	0	0	
Sup2	61	76	16	7	134	12	34	107	12	43	102	8	29	113	11	25	126	2	0	153	0	
Sup3	1	9	42	1	19	32	2	10	40	3	7	42	0	7	45	5	16	31	0	0	52	
JefS	63	65	24	23	105	24	58	69	25	75	55	22	65	63	24	61	73	18	60	71	21	
JefN	91	41	38	29	116	25	68	69	33	70	68	32	56	78	36	62	93	15	57	82	31	
Des	6	2	0	0	8	0	4	4	0	4	4	0	4	4	0	2	6	0	4	4	0	
Est	18	4	6	5	20	3	9	14	5	7	17	4	6	15	7	7	18	3	7	17	4	
Ama	34	13	8	13	36	6	32	16	7	28	18	9	20	28	7	23	27	5	26	23	6	
Jub	19	14	6	5	31	3	15	20	4	19	16	4	16	18	5	14	21	4	13	22	4	
Empr	4	6	0	3	7	0	3	6	1	6	3	1	5	4	1	5	4	1	5	4	1	
Noes	5	5	4	3	9	2	6	5	3	7	3	4	6	5	3	7	6	1	7	4	3	
Com	9	12	8	2	22	5	8	13	8	12	10	7	9	14	6	8	17	4	6	14	9	
Tec	3	8	2	1	10	2	3	7	3	7	4	2	7	3	3	6	5	2	4	6	3	
Prof	18	5	2	5	18	2	11	12	2	15	9	1	12	10	3	14	11	0	11	13	1	
Dep	3	3	6	1	5	6	3	3	6	5	3	4	4	3	5	3	6	3	3	5	4	
Obr	5	4	3	1	7	4	5	4	3	5	4	3	5	3	4	5	4	3	3	5	4	
Empl	26	26	15	10	42	15	23	30	14	25	28	14	23	30	14	24	37	6	24	32	11	
Jefe	4	4	2	3	6	1	4	4	2	5	4	1	4	4	2	5	4	1	4	4	2	
Prin	16	5	4	6	15	4	12	8	5	15	8	2	14	8	3	14	8	3	14	9	2	
Prco	46	28	17	17	64	10	43	34	14	49	26	16	38	39	14	37	48	6	44	35	12	
Sein	33	17	19	12	43	14	28	26	15	30	21	18	26	27	16	22	38	9	23	30	16	
Seco	34	34	10	11	53	14	30	33	15	30	37	11	27	36	15	30	39	9	23	41	14	
Tein	8	4	3	3	11	1	4	8	3	6	6	3	3	9	3	2	11	2	2	11	2	
Teco	8	13	6	2	21	4	4	20	3	12	13	2	7	15	5	12	12	3	5	19	3	
Unin	9	5	3	1	14	2	5	9	3	3	12	2	6	7	4	6	10	1	6	8	3	
300	27	15	7	13	30	6	27	16	6	29	14	6	26	18	5	22	24	3	27	17	5	
400	49	29	18	9	79	8	42	38	16	45	33	18	32	47	17	35	55	6	38	41	17	
600	27	21	21	10	40	19	19	33	17	28	24	17	24	25	20	21	34	14	20	32	17	
850	24	16	6	8	30	8	16	22	8	17	23	6	14	27	5	17	25	4	11	31	4	
1250	13	11	6	6	20	4	11	14	5	8	18	4	9	14	7	15	11	4	10	15	5	
2000	9	9	1	4	12	3	9	8	2	14	5	0	11	5	3	10	8	1	8	9	2	
Nanc	5	5	3	2	10	1	2	7	4	4	6	3	5	5	3	3	9	1	3	8	2	
Una	3	5	2	1	8	1	2	6	2	5	3	2	5	3	2	4	5	1	3	6	1	
Dos	22	13	5	4	33	3	17	20	3	16	20	4	15	21	4	18	19	3	13	23	4	
3a4	61	51	33	22	98	25	53	64	28	63	55	27	46	68	31	46	79	20	54	65	26	
5a6	44	24	16	15	54	15	29	37	18	38	30	16	34	32	18	38	40	6	30	37	17	
+7	24	13	6	10	28	5	25	11	7	23	15	5	21	17	5	17	23	3	17	22	4	
C1	10	17	2	5	23	1	6	19	4	10	16	3	10	15	4	13	15	1	8	18	3	
C2	56	48	10	19	85	10	48	57	9	55	54	5	45	59	10	51	60	3	46	61	7	
D1	52	28	38	14	71	33	45	38	35	52	31	35	38	42	38	33	57	28	35	51	32	
D2	36	13	12	14	42	5	27	24	10	28	22	11	28	25	8	26	34	1	28	23	10	
	JefS	JefN	Des	Est	Ama	Jub	Empr	Noes	Com	Tec	Prof	Dep	Obr	Empl	Jefe	Prin	Prco	Sein	Seco	Tein	Teco	Unin
JefS	152	0																				
JefN	0	170																				
Des	3	5	8</																			

Tabla 50. (Continuación)

	JefS	JefN	Des	Est	Ama	Jub	Empr	Noes	Com	Tec	Prof	Dep	Obr	Empl	Jefe	Prin	Prco	Sein	Seco	Tein	Teco	Unin
Prin	14	11	0	0	5	10	0	0	2	1	2	1	1	3	0	25	0	0	0	0	0	0
Prco	45	46	2	0	28	12	2	6	10	4	5	2	6	14	0	0	91	0	0	0	0	0
Sein	30	39	3	10	14	5	1	5	6	1	4	4	2	12	2	0	0	69	0	0	0	0
Seco	39	39	3	2	7	9	5	2	8	4	10	4	3	20	1	0	0	0	78	0	0	0
Tein	4	11	0	7	0	0	0	0	1	0	1	0	0	5	1	0	0	0	0	15	0	0
Teco	13	14	0	1	0	2	1	1	2	2	3	0	0	11	4	0	0	0	0	0	27	0
Unin	7	10	0	8	1	1	1	0	0	1	0	1	0	2	2	0	0	0	0	0	0	17
300	22	27	2	2	18	7	2	3	2	1	1	1	2	8	0	9	27	10	2	1	0	0
400	38	58	4	8	19	11	1	7	10	2	7	4	4	18	1	8	36	21	19	5	4	3
600	37	32	1	5	9	11	1	2	6	3	4	3	3	17	4	5	18	15	18	3	6	4
850	25	21	0	7	4	3	0	2	4	3	5	4	1	12	1	0	7	15	15	0	6	3
1250	9	21	0	2	4	5	3	0	2	1	3	0	1	8	1	2	2	2	10	2	8	4
2000	13	6	0	1	0	1	3	0	3	1	3	0	1	4	2	0	1	3	8	3	3	1
Nsnc	8	5	1	3	1	1	0	0	2	2	2	0	0	0	1	1	0	3	6	1	0	2
Una	10	0	1	0	0	6	0	0	0	1	0	0	1	1	0	4	2	2	2	0	0	0
Dos	19	21	2	3	13	10	0	0	1	0	6	0	0	5	0	5	15	7	9	0	0	2
3a4	74	71	2	10	23	10	5	9	16	8	10	8	5	31	8	5	31	38	37	8	16	10
5a6	37	47	2	10	10	7	4	4	10	3	8	3	4	17	2	6	29	15	20	6	5	3
+7	12	31	1	5	9	6	1	1	2	1	1	1	2	13	0	5	14	7	10	1	4	2
C1	20	9	0	2	0	3	4	0	2	3	7	0	0	4	4	0	3	2	10	3	6	5
C2	54	60	2	13	17	18	2	3	9	3	7	5	2	29	4	6	28	19	35	5	12	9
D1	57	61	6	9	21	13	2	5	13	5	8	4	5	25	2	12	35	31	23	6	9	2
D2	21	40	0	4	17	5	2	6	5	2	3	3	5	9	0	7	25	17	10	1	0	1
	JefS	JefN	Des	Est	Ama	Jub	Empr	Noes	Com	Tec	Prof	Dep	Obr	Empl	Jefe	Prin	Prco	Sein	Seco	Tein	Teco	Unin
300	400	600	850	1250	2000	Nsnc	Una	Dos	3a4	5a6	+7	C1	C2	D1	D2							
300	49	0	0	0	0	0	0	0	0	0	0	0	0	0	0							
400	0	96	0	0	0	0	0	0	0	0	0	0	0	0	0							
600	0	0	69	0	0	0	0	0	0	0	0	0	0	0	0							
850	0	0	0	46	0	0	0	0	0	0	0	0	0	0	0							
1250	0	0	0	0	30	0	0	0	0	0	0	0	0	0	0							
2000	0	0	0	0	0	19	0	0	0	0	0	0	0	0	0							
Nsnc	0	0	0	0	0	0	13	0	0	0	0	0	0	0	0							
Una	4	4	1	0	0	1	0	10	0	0	0	0	0	0	0							
Dos	8	13	9	4	4	0	2	0	40	0	0	0	0	0	0							
3a4	19	44	34	23	14	5	6	0	0	145	0	0	0	0	0							
5a6	12	26	18	11	6	9	2	0	0	0	84	0	0	0	0							
+7	6	9	7	8	6	4	3	0	0	0	0	43	0	0	0							
C1	0	3	3	3	11	6	3	1	0	16	10	2	29	0	0	0						
C2	8	24	30	24	13	9	6	3	22	46	29	14	0	114	0	0						
D1	21	40	31	13	6	4	3	4	14	59	21	20	0	0	118	0						
D2	20	29	5	6	0	0	1	2	4	24	24	7	0	0	0	61						
300	400	600	850	1250	2000	Nsnc	Una	Dos	3a4	5a6	+7	C1	C2	D1	D2							

Tabla 51. Tabla de Burt (Perfiles horizontales)

LE TRI-A-PLAT DE CHAQUE QUESTION FIGURE SUR LA DIAGONALE CORRESPONDANTE  
TOUS LES NOMBRES SONT EXPRIMES EN POURCENTAGES

	Sex1	Sex2	Int1	Int2	Int3	Clal	Clal2	Clal3	Viol1	Vio2	Vio3	Pol1	Pol2	Pol3	Nae1	Nae2	Nae3	Nap1	Nap2	Nap3
Sex1	50.3	0.0	36.4	53.7	9.9	54.3	27.8	17.9	41.4	43.2	15.4	31.5	60.5	8.0	19.1	75.3	5.6	24.7	66.0	9.3
Sex2	0.0	49.7	43.1	50.0	6.9	54.4	28.8	16.9	35.6	47.5	16.9	25.0	68.8	6.2	26.9	66.2	6.9	31.2	62.5	6.2
Int1	46.1	53.9	39.8	0.0	0.0	79.7	16.4	3.9	58.6	38.3	3.1	40.6	56.2	3.1	44.5	52.3	3.1	50.8	46.1	3.1
Int2	52.1	47.9	0.0	51.9	0.0	43.7	41.3	15.0	29.3	52.7	18.0	22.8	71.3	6.0	9.6	88.0	2.4	15.0	80.8	4.2
Int3	59.3	40.7	0.0	0.0	8.4	0.0	3.7	96.3	0.0	33.3	66.7	3.7	63.0	33.3	3.7	51.9	44.4	0.0	48.1	51.9
Clal	50.3	49.7	58.3	41.7	0.0	54.3	0.0	0.0	56.0	42.3	1.7	42.3	57.1	0.6	33.1	66.3	0.6	40.0	59.4	0.6
Clal2	49.5	50.5	23.1	75.8	1.1	0.0	28.3	0.0	24.2	68.1	7.7	13.2	84.6	2.2	11.0	85.7	3.3	17.6	76.9	5.5
Clal3	51.8	48.2	8.9	44.6	46.4	0.0	0.0	17.4	7.1	17.9	75.0	8.9	55.4	35.7	10.7	60.7	28.6	7.1	58.9	33.9
Viol1	54.0	46.0	60.5	39.5	0.0	79.0	17.7	3.2	38.5	0.0	0.0	67.7	31.5	0.8	41.9	58.1	0.0	44.4	54.0	1.6
Vio2	47.9	52.1	33.6	60.3	6.2	50.7	42.5	6.8	0.0	45.3	0.0	3.4	95.2	1.4	13.7	83.6	2.7	21.2	74.7	4.1
Vio3	48.1	51.9	7.7	57.7	34.6	5.8	13.5	80.8	0.0	0.0	16.1	3.8	57.7	38.5	3.8	65.4	30.8	7.7	59.6	32.7
Pol1	56.0	44.0	57.1	41.8	1.1	81.3	13.2	5.5	92.3	5.5	2.2	28.3	0.0	0.0	40.7	59.3	0.0	40.7	57.1	2.2
Pol2	47.1	52.9	34.6	57.2	8.2	48.1	37.0	14.9	18.8	66.8	14.4	0.0	64.6	0.0	17.3	79.3	3.4	25.0	68.8	6.2
Pol3	56.5	43.5	17.4	43.5	39.1	4.3	8.7	87.0	4.3	8.7	87.0	0.0	0.0	7.1	4.3	39.1	56.5	4.3	52.2	43.5
Nae1	41.9	58.1	77.0	21.6	1.4	78.4	13.5	8.1	70.3	27.0	2.7	50.0	48.6	1.4	23.0	0.0	0.0	94.6	2.7	2.7
Nae2	53.5	46.5	29.4	64.5	6.1	50.9	34.2	14.9	31.6	53.5	14.9	23.7	72.4	3.9	0.0	70.8	0.0	8.8	88.2	3.1
Nae3	45.0	55.0	20.0	20.0	60.0	5.0	15.0	80.0	8.0	20.0	80.0	0.0	35.0	65.0	0.0	0.0	6.2	0.0	20.0	80.0
Nap1	44.4	55.6	72.2	27.8	0.0	77.8	17.8	4.4	61.1	34.4	4.4	41.1	57.8	1.1	77.8	22.2	0.0	28.0	0.0	0.0
Nap2	51.7	48.3	28.5	65.2	6.3	50.2	33.8	15.9	32.4	52.7	15.0	25.1	69.1	5.8	1.0	97.1	1.9	0.0	64.3	0.0
Nap3	60.0	40.0	16.0	28.0	56.0	4.0	20.0	76.0	8.0	24.0	68.0	8.0	52.0	40.0	8.0	28.0	64.0	0.0	0.0	7.8
Cor1	47.1	52.9	67.3	32.7	0.0	78.8	18.3	2.9	69.2	29.8	1.0	49.0	51.0	0.0	53.8	46.2	0.0	63.5	36.5	0.0
Cor2	50.6	49.4	30.8	65.7	3.5	50.6	39.0	10.5	27.9	61.6	10.5	20.9	73.8	5.2	7.6	89.0	3.5	11.6	85.5	2.9
Cor3	56.5	43.5	10.9	43.5	45.7	13.0	10.9	76.1	8.7	19.6	71.7	8.7	60.9	30.4	10.9	58.7	30.4	8.7	47.8	43.5
Dep1	27.9	72.1	55.7	44.3	0.0	72.1	25.4	2.5	54.1	40.2	5.7	36.1	63.9	0.0	33.6	66.4	0.0	41.0	59.0	0.0
Dep2	68.4	31.6	32.2	59.8	8.0	47.7	32.8	19.5	32.8	51.7	15.5	27.0	66.7	6.3	17.8	77.6	4.6	22.4	70.1	7.5
Dep3	34.6	65.4	15.4	34.6	50.0	15.4	11.5	73.1	3.8	26.9	0.0	53.8	46.2	7.7	46.2	46.2	3.8	50.0	46.2	
Coc1	50.6	49.4	64.5	35.5	0.0	79.5	19.3	1.2	58.4	40.4	1.2	42.2	57.8	0.0	36.7	63.3	0.0	44.6	54.8</	

Tabla 51. (Continuación)

	Sex1	Sex2	Int1	Int2	Int3	Clal	Cl2	Cl3	Viol	Vio2	Vio3	Poll	Pol2	Pol3	Nae1	Nae2	Nae3	Nap1	Nap2	Nap3
Edi1	51.8	48.2	60.1	39.9	0.0	78.6	20.8	0.6	57.1	42.9	0.0	41.7	58.3	0.0	36.9	63.1	0.0	44.6	54.8	0.6
Edi2	52.9	47.1	24.7	71.8	3.5	38.8	51.8	9.4	30.6	64.7	4.7	18.8	78.8	2.4	7.1	89.4	3.5	11.8	84.7	3.5
Edi3	43.5	56.5	8.7	56.5	34.8	14.5	17.4	68.1	2.9	27.5	69.6	7.2	62.3	30.4	8.7	66.7	24.6	7.2	62.3	30.4
Coel	51.8	48.2	69.5	30.5	0.0	78.7	17.7	3.5	58.9	36.2	5.0	44.7	55.3	0.0	43.3	56.7	0.0	51.8	46.8	1.4
Coe2	49.6	50.4	20.2	75.2	4.7	45.7	45.7	8.5	29.5	64.3	6.2	20.2	76.0	3.9	6.2	79.9	3.9	9.3	86.0	4.7
Coe3	48.1	51.9	7.7	51.9	40.4	9.6	13.5	76.9	5.8	23.1	71.2	3.8	61.5	34.6	9.6	61.5	28.8	9.6	57.7	32.7
Cor1	52.4	47.6	69.9	30.1	0.0	82.5	14.6	2.9	66.0	31.1	2.9	48.5	50.5	1.0	47.6	51.5	1.0	51.5	45.6	2.9
Cor2	50.0	50.0	32.1	64.2	3.7	52.5	40.7	6.8	33.3	61.1	5.6	23.5	72.8	3.7	11.7	85.2	3.1	19.1	78.4	2.5
Cor3	47.4	52.6	7.0	56.1	36.8	8.8	17.5	73.7	3.5	26.3	70.2	5.3	66.7	28.1	10.5	64.9	24.6	10.5	57.9	31.6
Agr1	50.0	50.0	60.4	37.7	1.9	75.3	22.1	2.6	57.1	40.3	2.6	40.9	58.4	0.6	39.0	60.4	0.6	48.1	50.0	1.9
Agr2	54.7	45.3	25.5	72.6	1.9	48.1	47.2	4.7	30.2	64.2	5.7	20.8	77.4	1.9	6.6	91.5	1.9	8.5	90.6	0.9
Agr3	43.5	56.5	12.9	51.6	35.5	12.9	11.3	75.8	6.5	25.8	67.7	9.7	58.1	32.3	11.3	61.3	27.4	11.3	54.8	33.9
Reg1	53.8	46.2	63.5	36.5	0.0	90.4	9.6	0.0	82.7	17.3	0.0	65.4	34.6	0.0	57.7	42.3	0.0	63.5	36.5	0.0
Reg2	49.8	50.2	41.2	55.2	3.6	56.1	35.3	8.6	36.2	57.5	6.3	24.9	74.2	0.9	18.6	79.6	1.8	24.4	72.4	3.2
Reg3	49.0	51.0	8.2	53.1	38.8	8.2	16.3	75.5	2.0	20.4	77.6	4.1	53.1	42.9	6.1	61.2	32.7	6.1	57.1	36.7
Inv1	53.2	46.8	69.0	31.0	0.0	80.2	17.5	2.4	59.5	38.9	1.6	46.0	53.2	0.8	43.7	55.6	0.8	50.8	47.6	1.6
Inv2	46.4	53.6	26.8	68.8	4.3	47.1	44.2	8.7	32.6	60.1	7.2	20.3	76.8	2.9	10.1	86.2	3.6	15.9	81.2	2.9
Inv3	53.4	46.6	6.9	56.9	36.2	15.5	13.8	70.7	6.9	24.1	69.0	8.6	60.3	31.0	8.6	67.2	24.1	6.9	60.3	32.8
Esp1	54.5	45.5	62.8	37.2	0.0	80.0	18.6	1.4	56.6	42.1	1.4	42.8	56.6	0.7	35.2	64.1	0.7	43.4	54.5	2.1
Esp2	45.5	54.5	25.2	69.1	5.7	42.3	48.0	9.8	30.9	56.9	12.2	18.7	75.6	5.7	13.8	80.5	5.7	18.7	76.4	4.9
Esp3	50.0	50.0	11.1	51.9	37.0	13.0	9.3	77.8	7.4	27.8	64.8	11.1	61.1	27.8	11.1	66.7	22.2	7.4	63.0	29.6
Soc1	59.5	40.5	62.8	36.4	0.8	79.3	19.8	0.8	59.5	38.8	1.7	44.6	54.5	0.8	36.4	62.0	1.7	45.5	52.1	2.5
Soc2	44.7	55.3	33.3	64.5	2.1	51.1	41.8	7.1	34.0	58.9	7.1	22.7	74.5	2.8	17.0	80.1	2.8	21.3	75.2	3.5
Soc3	45.0	55.0	8.3	53.3	38.3	11.7	13.3	75.0	6.7	26.7	66.7	8.3	61.7	30.0	10.0	66.7	23.3	8.3	63.3	28.3
Chi1	54.5	45.5	57.7	39.0	3.3	76.4	18.7	4.9	60.2	35.0	4.9	46.3	52.0	1.6	33.3	64.2	2.4	38.2	58.5	3.3
Chi2	46.4	53.6	33.1	62.0	4.8	45.8	39.8	14.5	28.9	55.4	15.7	19.9	75.3	4.8	18.7	75.9	5.4	24.7	68.7	6.6
Chi3	54.5	45.5	6.1	48.5	45.5	15.2	6.1	78.8	6.1	33.3	60.6	3.0	57.6	39.4	6.1	69.7	24.2	6.1	63.6	30.3
Sup1	56.4	43.6	62.4	35.9	1.7	76.9	19.7	3.4	59.8	35.9	4.3	44.4	54.7	0.9	39.3	59.0	1.7	46.2	51.3	2.6
Sup2	46.4	53.6	33.3	62.1	4.6	51.6	39.9	8.5	32.7	57.5	9.8	22.9	73.2	3.9	14.4	81.0	4.6	20.3	75.8	3.9
Sup3	48.1	51.9	7.7	57.7	34.6	11.5	13.5	75.0	7.7	30.8	61.5	7.7	61.5	30.8	11.5	67.3	21.2	9.6	59.6	30.8
JefS	74.3	25.7	32.9	57.9	9.2	56.6	29.6	13.8	39.5	48.7	11.8	29.6	63.2	7.2	12.5	81.6	5.9	16.4	75.7	7.9
JefN	28.8	71.2	45.9	46.5	7.6	52.4	27.1	20.6	37.6	42.4	20.0	27.1	65.9	7.1	32.4	61.2	6.5	38.2	54.1	7.6
Des	62.5	37.5	75.0	25.0	0.0	62.5	37.5	0.0	25.0	75.0	0.0	12.5	87.5	0.0	37.5	50.0	12.5	50.0	50.0	0.0
Est	39.3	60.7	50.0	39.3	10.7	64.3	17.9	17.9	39.3	39.3	21.4	28.6	60.7	10.7	35.7	53.6	10.7	39.3	50.0	10.7
Ama	1.8	98.2	58.2	34.5	7.3	63.6	25.5	10.9	40.0	50.9	9.1	23.6	74.5	1.8	38.2	58.2	3.6	45.5	50.9	3.6
Jub	56.4	43.6	38.5	59.0	2.6	66.7	23.1	10.3	33.3	56.4	10.3	20.5	79.5	0.0	12.8	87.2	0.0	17.9	71.8	10.3
Empr	70.0	30.0	10.0	90.0	0.0	70.0	30.0	0.0	70.0	30.0	0.0	50.0	50.0	0.0	20.0	80.0	0.0	20.0	80.0	0.0
Noes	78.6	21.4	28.6	57.1	14.3	50.0	14.3	35.7	57.1	21.4	21.4	35.7	57.1	7.1	35.7	57.1	7.1	21.4	71.4	7.1
Com	62.1	37.9	31.0	58.6	10.3	37.9	41.4	20.7	34.5	37.9	27.6	31.0	62.1	6.9	3.4	93.1	3.4	17.2	75.9	6.9
Tec	84.6	15.4	30.8	61.5	7.7	38.5	46.2	15.4	38.5	46.2	15.4	38.5	53.8	7.7	7.7	84.6	7.7	7.7	84.6	7.7
Prof	80.0	20.0	40.0	56.0	4.0	60.0	36.0	4.0	52.0	40.0	8.0	32.0	68.0	0.0	36.0	64.0	0.0	48.0	48.0	4.0
Dep	50.0	50.0	33.3	41.7	25.0	41.7	16.7	41.7	25.0	41.7	33.3	25.0	33.3	41.7	25.0	50.0	25.0	16.7	50.0	33.3
Obr	75.0	25.0	25.0	41.7	33.3	41.7	33.3	25.0	33.3	58.3	8.3	8.3	91.7	0.0	8.3	83.3	8.3	16.7	75.0	8.3
Empl	53.7	46.3	35.8	56.7	7.5	44.8	29.9	25.4	31.3	44.8	23.9	31.3	55.2	13.4	19.4	70.1	10.4	20.9	70.1	9.0
Jefe	50.0	50.0	20.0	80.0	0.0	60.0	20.0	20.0	50.0	40.0	10.0	40.0	50.0	10.0	0.01000	0.0	0.0	20.0	80.0	0.0
Prin	60.0	40.0	48.0	36.0	16.0	52.0	28.0	20.0	36.0	44.0	20.0	24.0	72.0	4.0	24.0	72.0	4.0	32.0	56.0	12.0
Prco	49.5	50.5	44.0	50.5	5.5	60.4	27.5	12.1	38.5	49.5	12.1	26.4	71.4	2.2	22.0	74.7	3.3	31.9	64.8	3.3
Sein	47.8	52.2	42.0	46.4	11.6	52.2	23.2	24.6	39.1	42.0	18.8	20.3	71.0	8.7	33.3	60.9	5.8	37.7	52.2	10.1
Seco	55.1	44.9	37.2	57.7	5.1	51.3	33.3	15.4	35.9	47.4	16.7	28.2	61.5	10.3	20.5	71.8	7.7	23.1	67.9	9.0
Tein	60.0	40.0	53.3	40.0	6.7	53.3	26.7	20.0	53.3	33.3	13.3	53.3	40.0	6.7	26.7	66.7	6.7	26.7	66.7	6.7
Teco	33.3	66.7	22.2	66.7	11.1	59.3	25.9	14.8	33.3	48.1	18.5	37.0	48.1	14.8	11.1	74.1	14.8	11.1	77.8	11.1
Unin	47.1	52.9	23.5	64.7	11.8	41.2	35.3	23.5	47.1	35.3	17.6	41.2	52.9	5.9	11.8	82.4	5.9	11.8	82.4	5.9
300	38.8	61.2	42.9	51.0	6.1	51.0	42.9	6.1	38.8	53.1	8.2	22.4	77.6	0.0	24.5	67.3	8.2	32.7	59.2	8.2
400	50.0	50.0	56.2	36.5	7.3	57.3	24.0	18.8	42.7	43.8	13.5	28.1	66.7	5.2	33.3	61.5	5.2	35.4	57.3	7.3
600	50.7	49.3	29.0	59.4	11.6	55.1	21.7	23.2	24.6	52.2	23.2	17.4	71.0	11.6	15.9	78.8	5.8	27.5	65.2	7.2
850	63.0	37.0	32.6	58.7	8.7	52.2	28.3	19.6	41.3	41.3	17.4	37.0	47.8	15.2	21.7	69.6	8.7	21.7	67.4	10.9
1250	30.0	70.0	30.0	60.0	10.0	46.7	33.3	20.0	40.0	40.0	20.0	36.7	56.7	6.7	10.0	80.0	10.0	13.3	76.7	10.0
2000	78.9	21.1	31.6	68.4	0.0	68.4	26.3	5.3												

Tabla 51. (Continuación)

	Cor1	Cor2	Cor3	Dep1	Dep2	Dep3	Coc1	Coc2	Coc3	Gre1	Gre2	Gre3	Edi1	Edi2	Edi3	Coel	Coe2	Coe3	Cor1	Cor2	Cor3
Pol1	56.0	39.6	4.4	48.4	51.6	0.0	76.9	18.7	4.4	80.2	13.2	6.6	76.9	17.6	5.5	69.2	28.6	2.2	54.9	41.8	3.3
Pol2	25.5	61.1	13.5	37.5	55.8	6.7	46.2	36.1	17.8	51.0	26.9	22.1	47.1	32.2	20.7	37.5	47.1	15.4	25.0	56.7	18.3
Pol3	0.0	39.1	60.9	0.0	47.8	52.2	0.0	21.7	78.3	0.0	4.3	95.7	0.0	8.7	91.3	0.0	21.7	78.3	4.3	26.1	69.6
Nae1	75.7	17.6	6.8	55.4	41.9	2.7	82.4	10.8	6.8	82.4	8.1	9.5	83.8	8.1	8.1	82.4	10.8	6.8	66.2	25.7	8.1
Nae2	21.1	67.1	11.8	35.5	59.2	5.3	46.1	37.7	16.2	51.8	26.8	21.5	46.5	33.3	20.2	35.1	50.9	14.0	23.2	60.5	16.2
Nae3	0.0	30.0	70.0	0.0	40.0	60.0	0.0	15.0	85.0	0.0	10.0	90.0	0.0	15.0	85.0	0.0	25.0	75.0	5.0	25.0	70.0
Nap1	73.3	22.2	4.4	55.6	43.3	1.1	82.2	13.3	4.4	83.3	10.0	6.7	83.3	11.1	5.6	81.1	13.3	5.6	58.9	34.4	6.7
Nap2	18.4	71.0	10.6	34.8	58.9	6.3	44.0	38.6	17.4	49.8	28.5	21.7	44.4	34.8	20.8	31.9	53.6	14.5	22.7	61.4	15.9
Nap3	0.0	20.0	80.0	0.0	52.0	48.0	4.0	20.0	76.0	4.0	4.0	92.0	4.0	12.0	84.0	8.0	24.0	68.0	12.0	16.0	72.0
Cor1	32.3	0.0	0.0	60.6	38.5	1.0	84.6	11.5	3.8	86.5	8.7	4.8	84.6	13.5	1.9	79.8	19.2	1.0	59.6	37.5	2.9
Cor2	0.0	53.4	0.0	32.0	61.6	6.4	44.2	43.0	12.8	51.2	32.6	16.3	45.9	37.8	16.3	31.4	59.3	9.3	22.1	68.6	9.3
Cor3	0.0	0.0	14.3	0.0	60.9	30.4	4.3	23.9	71.7	2.2	8.7	89.1	2.2	13.0	84.8	8.7	15.2	76.1	6.5	10.9	82.6
Dep1	51.6	45.1	3.3	37.9	0.0	0.0	74.6	18.0	7.4	77.9	11.5	10.7	73.0	17.2	9.8	63.9	32.8	3.3	49.2	44.3	6.6
Dep2	23.0	60.9	16.1	0.0	54.0	0.0	42.5	40.8	16.7	47.7	31.0	21.3	44.8	35.6	19.5	35.6	46.6	17.8	23.6	57.5	19.0
Dep3	3.8	42.3	53.8	0.0	0.0	8.1	3.8	15.4	80.8	3.8	3.8	92.3	3.8	7.7	88.5	3.8	30.8	65.4	7.7	30.8	61.5
Coc1	53.0	45.8	1.2	54.8	44.6	0.6	51.6	0.0	0.0	91.6	6.6	1.8	88.6	9.6	1.8	74.1	25.3	0.6	55.4	42.8	1.8
Coc2	12.4	76.3	11.3	22.7	73.2	4.1	0.0	30.1	0.0	27.8	57.7	14.4	21.6	64.9	13.4	13.4	77.3	9.3	8.2	79.4	12.4
Coc3	6.8	37.3	55.9	15.3	49.2	35.6	0.0	0.0	18.3	0.0	3.4	96.6	0.0	10.2	89.8	8.5	20.3	71.2	5.1	23.7	71.2
Gre1	50.3	49.2	0.6	53.1	46.4	0.6	84.9	15.1	0.0	55.6	0.0	0.0	86.6	13.4	0.0	67.6	31.3	1.1	50.3	48.6	1.1
Gre2	13.0	81.2	5.8	20.3	78.3	1.4	15.9	81.2	2.9	0.0	21.4	0.0	17.4	76.8	5.8	15.9	78.3	5.8	11.6	84.1	4.3
Gre3	6.8	37.8	55.4	17.6	50.0	32.4	4.1	18.9	77.0	0.0	0.0	23.0	1.4	10.8	87.8	12.2	25.7	62.2	6.8	23.0	70.3
Edi1	52.4	47.0	0.6	53.0	46.4	0.6	87.5	12.5	0.0	92.3	7.1	0.6	52.2	0.0	0.0	72.6	26.8	0.6	54.2	45.2	0.6
Edi2	16.5	76.5	7.1	24.7	72.9	2.4	18.8	74.1	7.1	28.2	62.4	9.4	0.0	26.4	0.0	12.9	80.0	7.1	9.4	83.5	7.1
Edi3	2.9	40.6	56.5	17.4	49.3	33.3	4.3	18.8	76.8	0.0	5.8	94.2	0.0	0.0	21.4	11.6	23.2	65.2	5.8	21.7	72.5
Coel	58.9	38.3	2.8	55.3	44.0	0.7	87.2	9.2	3.5	85.8	7.8	6.4	86.5	7.8	5.7	43.8	0.0	0.0	67.4	29.1	3.5
Coe2	15.5	79.1	5.4	31.0	62.8	6.2	32.6	58.1	9.3	43.4	41.9	14.7	34.9	52.7	12.4	0.0	40.1	0.0	5.4	90.7	3.9
Coe3	1.9	30.8	67.3	7.7	59.6	32.7	1.9	17.3	80.8	3.8	7.7	88.5	1.9	11.5	86.5	0.0	0.0	16.1	1.9	7.7	90.4
Cor1	60.2	36.9	2.9	58.3	39.8	1.9	89.3	7.8	2.9	87.4	7.8	4.9	88.3	7.8	3.9	92.2	6.8	1.0	32.0	0.0	0.0
Cor2	24.1	72.8	3.1	33.3	61.7	4.9	43.8	47.5	8.6	53.7	35.8	10.5	46.9	43.8	9.3	25.3	72.2	2.5	0.0	50.3	0.0
Cor3	5.3	28.1	66.7	14.0	57.9	28.1	5.3	21.1	73.7	3.5	5.3	91.2	1.8	10.5	87.7	8.8	8.8	82.5	0.0	0.0	17.7
Agr1	52.6	44.8	2.6	57.1	42.2	0.6	80.5	17.5	1.9	85.1	11.0	3.9	82.5	14.9	2.6	71.4	26.6	1.9	53.9	44.2	1.9
Agr2	17.9	77.4	4.7	23.6	74.5	1.9	36.8	52.8	10.4	42.5	46.2	11.3	36.8	51.9	11.3	22.6	69.8	7.5	15.1	76.4	8.5
Agr3	6.5	33.9	59.7	14.5	48.4	37.1	4.8	22.6	72.6	4.8	4.8	90.3	3.2	11.3	85.5	11.3	22.6	66.1	6.5	21.0	72.6
Reg1	73.1	23.1	3.8	69.2	28.8	1.9	96.2	1.9	1.9	98.1	0.0	1.9	96.2	1.9	1.9	88.5	11.5	0.0	88.5	11.5	0.0
Reg2	29.9	64.7	5.4	36.7	61.1	2.3	51.6	40.3	8.1	57.9	29.9	12.2	53.4	36.2	10.4	41.2	51.6	7.2	25.3	67.0	7.7
Reg3	0.0	34.7	65.3	10.2	49.0	40.8	4.1	14.3	81.6	0.0	6.1	93.9	0.0	8.2	91.8	8.2	18.4	73.5	2.0	16.3	81.6
Inv1	61.1	36.5	2.4	52.4	46.0	1.6	91.3	7.9	0.8	91.3	5.6	3.2	92.1	5.6	2.4	83.3	15.9	0.8	61.9	36.5	1.6
Inv2	18.1	76.1	5.8	34.1	61.6	4.3	34.8	57.2	8.0	44.9	43.5	11.6	37.0	50.7	12.3	22.5	69.6	8.0	15.9	75.4	8.7
Inv3	3.4	36.2	60.3	15.5	53.4	31.0	5.2	13.8	81.0	3.4	3.4	93.1	1.7	13.8	84.5	8.6	22.4	69.0	5.2	20.7	74.1
Esp1	51.7	46.9	1.4	46.9	51.0	2.1	85.5	11.7	2.8	83.4	11.7	4.8	84.8	10.3	4.8	74.5	24.1	1.4	56.6	41.4	2.1
Esp2	21.1	70.7	8.1	37.4	57.7	4.9	32.5	55.3	12.2	46.3	40.7	13.0	36.6	51.2	12.2	21.1	65.9	13.0	14.6	75.6	9.8
Esp3	5.6	31.5	63.0	14.8	53.7	31.5	3.7	22.2	74.1	1.9	3.7	94.4	0.0	13.0	87.0	13.0	24.1	63.0	5.6	16.7	77.8
Soc1	58.7	38.8	2.5	49.6	47.9	2.5	87.6	10.7	1.7	90.1	6.6	3.3	89.3	8.3	2.5	79.3	19.0	1.7	62.0	36.4	1.7
Soc2	22.0	71.6	6.4	36.9	61.0	2.1	41.1	51.1	7.8	48.9	41.8	9.2	42.6	48.2	9.2	27.7	66.7	7.7	18.4	75.2	6.4
Soc3	3.3	40.0	56.7	16.7	50.0	33.3	3.3	20.0	76.7	1.7	3.3	95.0	0.0	11.7	88.3	10.0	20.0	70.0	3.3	20.0	76.7
Chi1	51.2	44.7	4.1	49.6	47.2	3.3	76.4	15.4	8.1	80.5	10.6	8.9	80.5	12.2	7.3	71.5	22.8	5.7	61.0	34.1	4.9
Chi2	24.7	65.1	10.2	35.5	59.6	4.8	43.4	44.6	12.0	48.2	32.5	19.3	41.6	39.8	18.7	31.9	56.6	11.4	16.9	67.5	15.7
Chi3	0.0	27.3	72.7	6.1	51.5	42.4	0.0	12.1	87.9	0.0	6.1	93.9	0.0	12.1	87.9	0.0	21.2	78.8	0.0	24.2	75.8
Sup1	57.3	38.5	4.3	53.0	44.4	2.6	82.9	12.0	5.1	87.2	6.0	6.8	88.9	6.0	5.1	77.8	17.9	4.3	65.0	30.8	4.3
Sup2	22.2	72.5	5.2	35.3	61.4	3.3	44.4	47.7	7.8	49.7	39.2	11.1	41.8	47.1	11.1	29.4	62.7	7.8	15.7	75.8	8.5
Sup3	5.8	30.8	63.5	11.5	53.8	34.6	1.9	19.2	78.8	1.9	3.8	94.2	0.0	11.5	88.5	9.6	23.1	67.3	5.8	19.2	75.0
JefS	25.0	61.2	13.8	30.3	61.8	7.9	48.7	35.5	15.8	53.3	27.0	19.7	53.9	27.6	18.4	38.8	48.7	12.5	28.9	57.9	13.2
JefN	38.8	46.5	14.7	44.7	47.1	8.2	54.1	25.3	20.6	57.6	16.5	25.9	50.6	25.3	24.1	48.2	32.4	19.4	34.7	43.5	21.8
Des	37.5	62.5	0.0	50.0	50.0	0.0	62.5	37.5	0.0	75.0	25.0	0.0	75.0	25.0	0.0	62.5	37.5	0.0	25.0	75.0	0.0
Est	46.4	32.1	21.4	35.7	57.1	7.1	46.4	39.3	14.3	50.0	21.4	28.6	35.7	42.9	21.4	32.1	46.4	21.4	28.6	46.4	25.0
Ama	41.8	47.3	10.9	65.5	27.3	7.3	65.5	16.4	18.2	69.1	12.7	18.2	69.1	14.5	16.4	61.8	25.5	12.7	45.5	40.0	14.5
Jub	25.6	66.7	7.7	48.7	46.2	5.1	51.3	35.9	12.8	51.3	30.8	17.9	56.4	28.2	15.4	46.2	43.6	10.3	30.8	61.5	7.7
Empr	50.0	40.0	10.0	60.0	40.0	0.0	60.0	30.0	10.0	70.0	20.0	10.0	70.0	20.0	10.0	60.0	40.0	10.0	30.0	70.0	0.0
Noes	42.9	42.9	14.3	14.3	71.4	14.3	57.1	7.1	35.7	57.1	7.1	35.7	50.0	14.3	35.7	57.1	21.4	21.4	42.9	35.7	21.4
Com	27.6	51.7	20.7	17.2	72.4	10.3	41.4	34.5	24.1	41.4	20.7	37.9	34.5	37.9	27.6	31.0	51.7	17.2	24.1	55.2	20.7
Tec	38.5	53.8	7.7	15.4	84.6	0.0	30.8	46.2	23.1	61.5	15.4	23.1	46.2	38.5	15.4	30.8	46.2	23.1	15.4	61.5	23.1
Prof	40.0	56.0	4.0	40.0	56.0	4.0	64.0	28.0	8.0	72.0	16.0	12.0	60.0	28.0	12.0	48.0	44.0	8.0	40.0	48.0</	

Tabla 51. (Continuación)

	Cor1	Cor2	Cor3	Dep1	Dep2	Dep3	Coc1	Coc2	Coc3	Grel	Gre2	Gre3	Edil	Edi2	Edi3	Coel	Coe2	Coe3	Cor1	Cor2	Cor3
Una	20.0	70.0	10.0	50.0	40.0	10.0	30.0	60.0	10.0	30.0	40.0	30.0	40.0	30.0	30.0	30.0	60.0	10.0	30.0	60.0	10.0
Dos	27.5	70.0	2.5	45.0	47.5	7.5	52.5	35.0	12.5	60.0	27.5	12.5	52.5	37.5	10.0	45.0	55.0	0.0	35.0	65.0	0.0
3a4	29.7	53.8	16.6	34.5	55.2	10.3	48.3	31.7	20.0	51.0	24.1	24.8	49.0	28.3	22.8	38.6	42.8	18.6	29.7	50.3	20.0
5a6	29.8	52.4	17.9	38.1	54.8	7.1	54.8	21.4	23.8	58.3	14.3	27.4	50.0	22.6	27.4	47.6	29.8	22.6	33.3	42.9	23.8
+7	53.5	34.9	11.6	39.5	58.1	2.3	60.5	30.2	9.3	67.4	16.3	16.3	69.8	16.3	14.0	55.8	32.6	11.6	34.9	48.8	16.3
C1	13.8	75.9	10.3	48.3	48.3	3.4	44.8	48.3	6.9	62.1	27.6	10.3	37.9	51.7	10.3	13.8	72.4	13.8	20.7	65.5	13.8
C2	30.7	64.0	5.3	37.7	58.8	3.5	55.3	36.0	8.8	59.6	28.1	12.3	55.3	33.3	11.4	45.6	46.5	7.9	35.1	57.9	7.0
D1	30.5	44.1	25.4	33.9	50.8	15.3	47.5	23.7	28.8	47.5	15.3	37.3	50.0	15.3	34.7	45.8	28.8	25.4	28.0	43.2	28.8
D2	47.5	41.0	11.5	41.0	54.1	4.9	55.7	23.0	21.3	60.7	18.0	21.3	57.4	23.0	19.7	50.8	34.4	14.8	39.3	42.6	18.0
	Agri1	Agri2	Agri3	Reg1	Reg2	Reg3	Inv1	Inv2	Inv3	Esp1	Esp2	Esp3	Soc1	Soc2	Soc3	Chil	Chi2	Chi3	Sup1	Sup2	Sup3
Sex1	47.5	35.8	16.7	17.3	67.9	14.8	41.4	39.5	19.1	48.8	34.6	16.7	44.4	38.9	16.7	41.4	47.5	11.1	40.7	43.8	15.4
Sex2	48.1	30.0	21.9	15.0	69.4	15.6	36.9	46.2	16.9	41.2	41.9	16.9	30.6	48.8	20.6	35.0	55.6	9.4	31.9	51.2	16.9
Int1	72.7	21.1	6.2	25.8	71.1	3.1	68.0	28.9	3.1	71.1	24.2	4.7	59.4	36.7	3.9	55.5	43.0	1.6	57.0	39.8	3.1
Int2	34.7	46.1	19.2	11.4	73.1	15.6	23.4	56.9	19.8	32.3	50.9	16.8	26.3	54.5	19.2	28.7	61.7	9.6	25.1	56.9	18.0
Int3	11.1	7.4	81.5	0.0	29.6	70.4	0.0	22.2	77.8	0.0	25.9	74.1	3.7	11.1	85.2	14.8	29.6	55.6	7.4	25.9	66.7
Clai1	66.3	29.1	4.6	26.9	70.9	2.3	57.7	37.1	5.1	66.3	29.7	4.0	54.9	41.1	4.0	53.7	43.4	2.9	51.4	45.1	3.4
Clai2	37.4	54.9	7.7	5.5	85.7	8.8	24.2	67.0	8.8	29.7	64.8	5.5	26.4	64.8	8.8	25.3	72.5	2.2	25.3	67.0	7.7
Clai3	7.1	8.9	83.9	0.0	33.9	66.1	5.4	21.4	73.2	3.6	21.4	75.0	1.8	17.9	80.4	10.7	42.9	46.4	7.1	23.2	69.6
Vio1	71.0	25.8	3.2	34.7	64.5	0.8	60.5	36.3	3.2	66.1	30.6	3.2	58.1	38.7	3.2	59.7	38.7	1.6	56.5	40.3	3.2
Vio2	42.5	46.6	11.0	6.2	87.0	6.8	33.6	56.8	9.6	41.8	47.9	10.3	32.2	56.8	11.0	29.5	63.0	7.5	28.8	60.3	11.0
Vio3	7.7	11.5	80.8	0.0	26.9	73.1	3.8	19.2	76.9	3.8	28.8	67.3	3.8	19.2	76.9	11.5	50.0	38.5	9.6	28.8	61.5
Pol1	69.2	24.2	6.6	37.4	60.4	2.2	63.7	30.8	5.5	68.1	25.3	6.6	59.3	35.2	5.5	62.6	36.3	1.1	57.1	38.5	4.4
Pol2	43.3	39.4	17.3	8.7	78.8	12.5	32.2	51.0	16.8	39.4	44.7	15.9	31.7	50.5	17.8	30.8	60.1	9.1	30.8	53.8	15.4
Pol3	4.3	8.7	87.0	0.0	8.7	91.3	4.3	17.4	78.3	4.3	30.4	65.2	4.3	17.4	78.3	8.7	34.8	56.5	4.3	26.1	69.6
Nae1	81.1	9.5	9.5	40.5	55.4	4.1	74.3	18.9	6.8	68.9	23.0	8.1	59.5	32.4	8.1	55.4	41.9	2.7	62.2	29.7	8.1
Nae2	40.8	42.5	16.7	9.6	77.2	13.2	37.0	52.2	17.1	40.8	43.4	15.8	32.9	49.6	17.5	34.6	55.3	10.1	30.3	54.4	15.4
Nae3	5.0	10.0	85.0	0.0	20.0	80.0	5.0	25.0	70.0	5.0	35.0	60.0	10.0	20.0	70.0	15.0	45.0	40.0	10.0	35.0	55.0
Nap1	82.2	10.0	7.8	36.7	60.0	3.3	71.1	24.4	4.4	70.0	25.6	4.4	61.1	33.3	5.6	52.2	45.6	2.2	60.0	34.4	5.6
Nap2	37.2	46.4	16.4	9.2	77.3	13.5	29.0	54.1	16.9	38.2	45.4	16.4	30.4	51.2	18.4	34.8	55.1	10.1	29.0	56.0	15.0
Nap3	12.0	4.0	84.0	0.0	28.0	72.0	8.0	16.0	76.0	12.0	24.0	64.0	12.0	20.0	68.0	16.0	44.0	40.0	12.0	24.0	64.0
Cor1	77.9	18.3	3.8	36.5	63.5	0.0	74.0	24.0	1.9	72.1	25.0	2.9	68.3	29.8	1.9	60.6	39.4	0.0	64.4	32.7	2.9
Cor2	40.1	47.7	12.2	7.0	83.1	9.9	26.7	61.0	12.2	39.5	50.6	9.9	27.3	58.7	14.0	32.0	62.8	5.2	26.2	64.5	9.3
Cor3	8.7	10.9	80.4	4.3	26.1	69.6	6.5	17.4	76.1	4.3	21.7	73.9	6.5	19.6	73.9	10.9	37.0	52.2	10.9	17.4	71.7
Dep1	72.1	20.5	7.4	29.5	66.4	4.1	54.1	38.5	7.4	55.7	37.7	6.6	49.2	42.6	8.2	50.0	48.4	1.6	50.8	44.3	4.9
Dep2	37.4	45.4	17.2	8.6	77.6	13.8	33.3	48.9	17.8	42.5	40.8	16.7	33.3	49.4	17.2	33.3	56.9	9.8	29.9	54.0	16.1
Dep3	3.8	7.7	88.5	3.8	19.2	76.9	7.7	23.1	69.2	11.5	23.1	65.4	11.5	11.5	76.9	15.4	30.8	53.8	11.5	19.2	69.2
Coc1	74.7	23.5	1.8	30.1	68.7	1.2	69.3	28.9	1.8	74.7	24.1	1.2	63.9	34.9	1.2	56.6	43.4	0.0	58.4	41.0	0.6
Coc2	27.8	57.7	14.4	1.0	91.8	7.2	10.3	81.4	8.2	17.5	70.1	12.4	13.4	74.2	12.4	19.6	76.3	4.1	14.4	75.3	10.3
Coc3	5.1	18.6	76.3	1.7	30.5	67.8	1.7	18.6	79.7	6.8	25.4	67.8	3.4	18.6	78.0	16.9	33.9	49.2	10.2	20.3	69.5
Grel	73.2	25.1	1.7	28.5	71.5	0.0	64.2	34.6	1.1	67.6	31.8	0.6	60.9	38.5	0.6	55.3	44.7	0.0	57.0	42.5	0.6
Gre2	24.6	71.0	4.3	0.0	95.7	4.3	10.1	87.0	2.9	24.6	72.5	2.9	11.6	85.5	2.9	18.8	78.3	2.9	10.1	87.0	2.9
Gre3	8.1	16.2	75.7	1.4	36.5	62.2	5.4	21.6	73.0	9.5	21.6	68.9	5.4	17.6	77.0	14.9	43.2	41.9	10.8	23.0	66.2
Edi1	75.6	23.2	1.2	29.8	70.2	0.0	69.0	30.4	0.6	73.2	26.8	0.0	64.3	35.7	0.0	58.9	41.1	0.0	61.9	38.1	0.0
Edi2	27.1	64.7	8.2	1.2	94.1	4.7	8.2	82.4	9.4	17.6	74.1	8.2	11.8	80.0	8.2	17.6	77.6	4.7	8.2	84.7	7.1
Edi3	5.8	17.4	76.8	1.4	33.3	65.2	4.3	24.6	71.0	10.1	21.7	68.1	4.3	18.8	76.8	13.0	44.9	42.0	8.7	24.6	66.7
Coel	78.0	17.0	5.0	32.6	64.5	2.8	74.5	22.0	3.5	76.6	18.4	5.0	68.1	27.7	4.3	62.4	37.6	0.0	64.5	31.9	3.5
Coe2	31.8	57.4	10.9	4.7	88.4	7.0	15.5	74.4	10.1	27.1	62.8	10.1	17.8	72.9	9.3	21.7	72.9	5.4	16.3	74.4	9.3
Coe3	5.8	15.4	78.8	0.0	30.8	69.2	1.9	21.2	76.9	3.8	30.8	65.4	3.8	15.4	80.8	13.5	36.5	50.0	9.6	23.1	67.3
Cor1	80.6	15.5	3.9	44.7	54.4	1.0	75.7	21.4	2.9	79.6	17.5	2.9	72.8	25.2	1.9	72.8	27.2	0.0	73.8	23.3	2.9
Cor2	42.0	50.0	8.0	3.7	91.4	4.9	28.4	64.2	7.4	37.0	57.4	5.6	27.2	65.4	7.4	25.9	69.1	4.9	22.2	71.6	6.2
Cor3	5.3	15.8	78.9	0.0	29.8	70.2	3.5	21.1	75.4	5.3	21.1	73.7	3.5	15.8	80.7	10.5	45.6	43.9	8.8	22.8	68.4
Agri1	47.8	0.0	0.0	31.2	68.2	0.6	65.6	32.5	1.9	67.5	31.2	1.3	60.4	38.3	1.3	55.8	44.2	0.0	59.7	39.6	0.6
Agri2	0.0	32.9	0.0	3.8	88.7	7.5	18.9	71.7	9.4	34.9	59.4	5.7	22.6	68.9	8.5	26.4	67.0	6.6	19.8	71.7	8.5
Agri3	0.0	0.0	19.3	0.0	35.5	64.5	8.1	19.4	72.6	6.5	19.4	74.2	6.5	14.5	79.0	14.5	43.5	41.9	6.5	25.8	67.7
Reg1	92.3	7.7	0.0	16.1	0.0	0.0	90.4	7.7	1.9	90.4	7.7	1.9	84.6	13.5	1.9	76.9	21.2	1.9	84.6	13.5	1.9
Reg2	47.5																				

Tabla 51. (Continuación)

	Agr1	Agr2	Agr3	Reg1	Reg2	Reg3	Inv1	Inv2	Inv3	Esp1	Esp2	Esp3	Soc1	Soc2	Soc3	Chi1	Chi2	Chi3	Sup1	Sup2	Sup3	
Des	75.0	25.0	0.0	0.0100	0.0	0.0	50.0	50.0	0.0	50.0	50.0	0.0	50.0	50.0	0.0	25.0	75.0	0.0	50.0	50.0	0.0	
Est	64.3	14.3	21.4	17.9	71.4	10.7	32.1	50.0	17.9	25.0	60.7	14.3	21.4	53.6	25.0	25.0	64.3	10.7	25.0	60.7	14.3	
Ama	61.8	23.6	14.5	23.6	65.5	10.9	58.2	29.1	12.7	50.9	32.7	16.4	36.4	50.9	12.7	41.8	49.1	9.1	47.3	41.8	10.9	
Jub	48.7	35.9	15.4	12.8	79.5	7.7	38.5	51.3	10.3	48.7	41.0	10.3	41.0	46.2	12.8	35.9	53.8	10.3	33.3	56.4	10.3	
Empr	40.0	60.0	0.0	30.0	70.0	0.0	30.0	60.0	10.0	60.0	30.0	10.0	50.0	40.0	10.0	50.0	40.0	10.0	50.0	40.0	10.0	
Noes	35.7	35.7	28.6	21.4	64.3	14.3	42.9	35.7	21.4	50.0	21.4	28.6	42.9	35.7	21.4	50.0	42.9	7.1	50.0	28.6	21.4	
Com	31.0	41.4	27.6	6.9	75.9	17.2	27.6	44.8	27.6	41.4	34.5	24.1	31.0	48.3	20.7	27.6	58.6	13.8	20.7	48.3	31.0	
Tec	23.1	61.5	15.4	7.7	76.9	15.4	23.1	53.8	23.1	53.8	30.8	15.4	53.8	23.1	23.1	46.2	36.5	15.4	30.8	46.2	23.1	
Prof	72.0	20.0	8.0	20.0	72.0	8.0	44.0	48.0	8.0	60.0	36.0	4.0	48.0	40.0	12.0	56.0	44.0	0.0	44.0	56.0	4.0	
Dep	25.0	25.0	50.0	8.3	41.7	50.0	25.0	25.0	50.0	41.7	25.0	33.3	33.3	25.0	41.7	25.0	50.0	25.0	25.0	41.7	33.3	
Obr	41.7	33.3	25.0	8.3	58.3	33.3	41.7	33.3	25.0	41.7	33.3	25.0	41.7	25.0	33.3	41.7	33.3	25.0	25.0	41.7	33.3	
Empl	38.8	38.8	22.4	14.9	62.7	22.4	34.3	44.8	20.9	37.3	41.8	20.9	34.3	44.8	20.9	35.8	55.2	9.0	35.8	47.8	16.4	
Jefe	40.0	40.0	20.0	30.0	60.0	10.0	40.0	40.0	20.0	50.0	40.0	10.0	40.0	40.0	20.0	50.0	40.0	10.0	40.0	40.0	20.0	
Prin	64.0	20.0	16.0	24.0	60.0	16.0	48.0	32.0	20.0	60.0	32.0	8.0	56.0	32.0	12.0	56.0	32.0	12.0	56.0	36.0	8.0	
Prco	50.5	30.8	18.7	18.7	70.3	11.0	47.3	37.4	15.4	53.8	28.6	17.6	41.8	42.9	15.4	40.7	52.7	6.6	48.4	38.5	13.2	
Sein	47.8	24.6	27.5	17.4	62.3	20.3	40.6	37.7	21.7	43.5	30.4	26.1	37.7	39.1	23.2	31.9	55.1	13.0	33.3	43.5	23.2	
Seco	43.6	43.6	12.8	14.1	67.9	17.9	38.5	42.3	19.2	38.5	47.4	14.1	34.6	46.2	19.2	38.5	50.0	11.5	29.5	52.6	17.9	
Tein	53.3	26.7	20.0	20.0	73.3	6.7	26.7	53.3	20.0	40.0	40.0	20.0	20.0	60.0	20.0	13.3	73.3	13.3	13.3	73.3	13.3	
Teco	29.6	48.1	22.2	7.4	77.8	14.8	14.8	74.1	11.1	44.4	48.1	7.4	25.9	55.6	18.5	44.4	44.4	11.1	18.5	70.4	11.1	
Unin	52.9	29.4	17.6	5.9	82.4	11.8	29.4	52.9	17.6	17.6	70.6	11.8	35.3	41.2	23.5	35.3	58.8	5.9	35.3	47.1	17.6	
300	55.1	30.6	14.3	26.5	61.2	12.2	55.1	32.7	12.2	59.2	28.6	12.2	53.1	36.7	10.2	44.9	49.0	6.1	55.1	34.7	10.2	
400	51.0	30.2	18.8	9.4	82.3	8.3	43.8	39.6	16.7	46.9	34.4	18.8	33.3	49.0	17.7	36.5	57.3	6.2	39.6	42.7	17.7	
600	39.1	30.4	30.4	14.5	58.0	27.5	27.5	47.8	24.6	40.6	34.8	24.6	34.8	36.2	29.0	30.4	49.3	20.3	29.0	46.4	24.6	
850	52.2	34.8	13.0	17.4	65.2	17.4	34.8	47.8	17.4	37.0	50.0	13.0	30.4	58.7	10.9	37.0	54.3	8.7	23.9	67.4	8.7	
1250	43.3	36.7	20.0	20.0	66.7	13.3	36.7	46.7	16.7	26.7	60.0	13.3	30.0	46.7	23.3	50.0	36.7	13.3	33.3	50.0	16.7	
2000	47.4	47.4	5.3	21.1	63.2	15.8	47.4	42.1	10.5	73.7	26.3	0.0	57.9	26.3	15.8	52.6	42.1	5.3	42.1	47.4	10.5	
Nanc	38.5	38.5	23.1	15.4	76.9	7.7	38.5	38.5	30.8	30.8	46.2	23.1	38.5	38.5	23.1	23.1	69.2	7.7	23.1	61.5	15.4	
Una	30.0	50.0	20.0	10.0	80.0	10.0	20.0	60.0	20.0	50.0	30.0	20.0	50.0	30.0	20.0	40.0	50.0	10.0	30.0	60.0	10.0	
Dos	55.0	32.5	12.5	10.0	82.5	7.5	42.5	50.0	7.5	40.0	50.0	10.0	37.5	52.5	10.0	45.0	47.5	7.5	32.5	57.5	10.0	
3a4	42.1	35.2	22.8	15.2	67.6	17.2	36.6	44.1	19.3	43.4	37.9	18.6	31.7	46.9	21.4	31.7	54.5	13.8	37.2	44.8	17.9	
5a6	52.4	28.6	19.0	17.9	64.3	17.9	34.5	44.0	21.4	45.2	35.7	19.0	40.5	38.1	21.4	45.2	47.6	7.1	35.7	44.0	20.2	
+7	55.8	30.2	14.0	23.3	65.1	11.6	58.1	25.6	16.3	53.5	34.9	11.6	48.8	39.5	11.6	39.5	53.5	7.0	39.5	51.2	9.3	
C1	34.5	58.6	6.9	17.2	79.3	3.4	20.7	65.5	13.8	34.5	55.2	10.3	34.5	51.7	13.8	44.8	51.7	3.4	27.6	62.1	10.3	
C2	49.1	42.1	8.8	16.7	74.6	8.8	42.1	50.0	7.9	48.2	47.4	4.4	39.5	51.8	8.8	44.7	52.6	2.6	40.4	53.5	6.1	
D1	44.1	23.7	32.2	11.9	60.2	28.0	38.1	32.2	29.7	44.1	26.3	29.7	32.2	35.6	32.2	28.0	48.3	23.7	29.7	43.2	27.1	
D2	59.0	21.3	19.7	23.0	68.9	8.2	44.3	39.3	16.4	45.9	36.1	18.0	45.9	41.0	13.1	42.6	55.7	1.6	45.9	37.7	16.4	
JefS	JefN	Des	Est	Ama	Jub	Empr	Noes	Com	Tec	Prof	Dep	Obr	Empl	Jefe	Prin	Prco	Sein	Seco	Tein	Teco	Unin	
Sex1	69.8	30.2	3.1	6.8	0.6	13.6	4.3	6.8	11.1	6.8	12.3	3.7	5.6	22.2	3.1	9.3	27.8	20.4	26.5	5.6	5.6	4.9
Sex2	24.4	75.6	1.9	10.6	33.8	10.6	1.9	1.9	6.9	1.2	3.1	3.8	1.9	19.4	3.1	6.2	28.8	22.5	21.9	3.8	11.2	5.6
Int1	39.1	60.9	4.7	10.9	25.0	11.7	0.8	3.1	7.0	3.1	7.8	3.1	2.3	18.8	1.6	9.4	31.2	22.7	22.7	6.2	4.7	3.1
Int2	52.7	47.3	1.2	6.6	11.4	13.8	5.4	4.8	10.2	4.8	8.4	3.0	3.0	22.8	4.8	5.4	27.5	19.2	26.9	3.6	10.8	6.6
Int3	51.9	48.1	0.0	11.1	14.8	3.7	0.0	7.4	11.1	3.7	3.7	11.1	14.8	18.5	0.0	14.8	18.5	29.6	14.8	3.7	11.1	7.4
Cla1	49.1	50.9	2.9	10.3	20.0	14.9	4.0	4.0	6.3	2.9	8.6	2.9	2.9	17.1	3.4	7.4	31.4	20.6	22.9	4.6	9.1	4.0
Cla2	49.5	50.5	3.3	5.5	15.4	9.9	3.3	2.2	13.2	6.6	9.9	2.2	4.4	22.0	2.2	7.7	27.5	17.6	28.6	4.4	7.7	6.6
Cla3	37.5	62.5	0.0	8.9	10.7	7.1	0.0	8.9	10.7	3.6	1.8	8.9	5.4	30.4	3.6	8.9	19.6	30.4	21.4	5.4	7.1	7.1
Viol	48.4	51.6	1.6	8.9	17.7	10.5	5.6	6.5	8.1	4.0	10.5	2.4	3.2	16.9	4.0	7.3	28.2	21.8	22.6	6.5	7.3	6.5
Viol2	50.7	49.3	4.1	7.5	19.2	15.1	2.1	2.1	7.5	4.1	6.8	3.4	4.8	20.5	2.7	7.5	30.8	19.9	25.3	3.4	8.9	4.1
Viol3	34.6	65.4	0.0	11.5	9.6	7.7	0.0	5.8	15.4	3.8	3.8	7.7	1.9	30.8	1.9	9.6	21.2	25.0	25.0	3.8	9.6	5.8
Poll	49.5	50.5	1.1	8.8	14.3	8.8	5.5	5.5	9.9	5.5	8.8	3.3	1.1	23.1	4.4	6.6	26.4	15.4	24.2	8.8	11.0	7.7
Poll2	46.2	53.8	3.4	8.2	19.7	14.9	2.4	3.8	8.7	3.4	8.2	1.9	5.3	17.8	2.4	8.7	31.2	23.6	23.1	2.9	6.2	4.3
Poll3	47.8	52.2	0.0	13.0	4.3	0.0	0.0	4.3	8.7	4.3	0.0	21.7	0.0	39.1	4.3	4.3	8.7	26.1	34.8	4.3	17.4	4.3
Nae1	25.7	74.3	4.1	13.5	28.4	6.8	2.7	6.8	1.4	1.4	12.2	4.1	1.4	17.6	0.0	8.1	27.0	31.1	21.6	5.4	4.1	2.7
Nae2	54.4	45.6	1.8	6.6	14.0	14.9	3.5	3.5	11.8	4.8	7.0	2.6	4.4	20.6	4.4	7.9	29.8	18.4	24.6	4.4	8.8	6.1
Nae3	45.0	55.0	5.0	15.0	10.0	0.0	0.0	5.0	5.0	5.0	0.0	15.0	5.0	35.0	0.0	5.0	15.0	20.0	30.0	5.0	20.0	5.0
Nap1	27.8	72.2	4.4	12.2	27.8	7.8	2.2	3.3	5.6	1.1	13.3	2.2	2.2	15.6	2.2	8.9	32.2	28.9	20.0	4.4	3.3	2.2
Nap2	55.6	44.4	1.9	6.8	13.5	13.5	3.9	4.8	10.6	5.3	5.8	2.9	4.3	22.7	3.9	6.8	28.5	17.4	25.6	4.8	10.1	6.8
Nap3	48.0	52.0	0.0	12.0	8.0	16.0	0.0	4.0	8.0	4.0	4.0	16.0	4.0	24.0	0.0	12.0	12.0	28.0	28.0	4.0	12.0	4.0
Cor1	36.5	63.5	2.9	12.5	22.1	9.6	4.8	5.8	7.7	4.8	9.6	2.9	2.9	12.5	1.9	9.6	29.8	27.9	20.2	2.9	3.8	5.8
Cor2	54.1	45.9	2.9	5.2	15.1	15.1	2.3	3.5	8.7	4.1	8.1	2.9	3.5	23.8	4.7	5.8	28.5	16.3	27.3	5.8	11.6	4.7
Cor3	45.7	54.3	0.0	13.0	13.0	6.5	2.2	4.3	13.0	2.2	2.2	8.7	6.5	28.3	0.0	10.9	23.9	26.1	21.7	4.3	6.5	6.5
Dep1	37.7	62.3	3.3	8.2	29.5	15.6	4.9	1.6	4.1	1.6	8.2	2.5	0.8	17.2	2.5	10.7	27.9	22.1	22.1	4.1	7.4	5.7
Dep2	54.0	46.0	2																			

Tabla 51. (Continuación)

	JefS	JefN	Des	Est	Ama	Jub	Empr	Noes	Com	Tec	Prof	Dep	Obr	Empl	Jefe	Prin	Prco	Sein	Seco	Tein	Teco	Unin
Inv1	46.0	54.0	3.2	7.1	25.4	11.9	2.4	4.8	6.3	2.4	8.7	2.4	4.0	18.3	3.2	9.5	34.1	22.2	23.8	3.2	3.2	4.0
Inv2	50.0	50.0	2.9	10.1	11.6	14.5	4.3	3.6	9.4	5.1	8.7	2.2	2.9	21.7	2.9	5.8	24.6	18.8	23.9	5.8	14.5	6.5
Inv3	43.1	56.9	0.0	8.6	12.1	6.9	1.7	5.2	13.8	5.2	3.4	10.3	5.2	24.1	3.4	8.6	24.1	25.9	25.9	5.2	5.2	5.2
Esp1	51.7	48.3	2.8	4.8	19.3	13.1	4.1	4.8	8.3	4.8	10.3	3.4	3.4	17.2	3.4	10.3	33.8	20.7	20.7	4.1	8.3	2.1
Esp2	44.7	55.3	3.3	13.8	14.6	13.0	2.4	2.4	8.1	3.3	7.3	2.4	3.3	22.8	3.3	6.5	21.1	17.1	30.1	4.9	10.6	9.8
Esp3	40.7	59.3	0.0	7.4	16.7	7.4	1.9	7.4	13.0	3.7	1.9	7.4	5.6	25.9	1.9	3.7	29.6	33.3	20.4	5.6	3.7	3.7
Soc1	53.7	46.3	3.3	5.0	16.5	13.2	4.1	5.0	7.4	5.8	9.9	3.3	4.1	19.0	3.3	11.6	31.4	21.5	22.3	2.5	5.8	5.0
Soc2	44.7	55.3	2.8	10.6	19.9	12.8	2.8	3.5	9.9	2.1	7.1	2.1	2.1	21.3	2.8	5.7	27.7	19.1	25.5	6.4	10.6	5.0
Soc3	40.0	60.0	0.0	11.7	11.7	8.3	1.7	5.0	10.0	5.0	5.0	8.3	6.7	23.3	3.3	5.0	23.3	26.7	25.0	5.0	8.3	6.7
Chi1	49.6	50.4	1.6	5.7	18.7	11.4	4.1	5.7	6.5	4.9	11.4	2.4	4.1	19.5	4.1	11.4	30.1	17.9	24.4	1.6	9.8	4.9
Chi2	44.0	56.0	3.6	10.8	16.3	12.7	2.4	3.6	10.2	3.0	6.6	3.6	2.4	22.3	2.4	4.8	28.9	22.9	23.5	6.6	7.2	6.0
Chi3	54.5	45.5	0.0	9.1	15.2	12.1	3.0	3.0	12.1	6.1	0.0	9.1	9.1	18.2	3.0	9.1	18.2	27.3	27.3	6.1	9.1	3.0
Sup1	51.3	48.7	3.4	6.0	22.2	11.1	4.3	6.0	5.1	3.4	9.4	2.6	2.6	20.5	3.4	12.0	37.6	19.7	19.7	1.7	4.3	5.1
Sup2	46.4	53.6	2.6	11.1	15.0	14.4	2.6	2.6	9.2	3.9	8.5	3.3	3.3	20.9	2.6	5.9	22.9	19.6	26.8	7.2	12.4	5.2
Sup3	40.4	59.6	0.0	7.7	11.5	7.7	1.9	5.8	17.3	5.8	1.9	7.7	7.7	21.2	3.8	3.8	23.1	30.8	26.9	3.8	5.8	5.8
JefS	47.2	0.0	2.0	1.3	4.6	19.1	5.3	5.3	10.5	7.2	11.2	3.9	5.9	19.7	3.9	9.2	29.6	19.7	25.7	2.6	8.6	4.6
JefN	0.0	52.8	2.9	15.3	28.2	5.9	1.2	3.5	7.6	1.2	4.7	3.5	1.8	21.8	2.4	6.5	27.1	22.9	22.9	6.5	8.3	5.9
Des	37.5	62.5	2.5	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	25.0	37.5	37.5	0.0	0.0	0.0
Est	7.1	92.9	0.0	8.7	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	35.7	7.1	25.0	3.6	28.6
Ama	12.7	87.3	0.0	0.0	17.1	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	9.1	50.9	25.5	12.7	0.0	0.0	1.8
Jub	74.4	25.6	0.0	0.0	0.0	12.1	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	25.6	30.8	12.8	23.1	0.0	5.1	2.6
Empr	80.0	20.0	0.0	0.0	0.0	0.0	3.1	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	20.0	10.0	50.0	0.0	10.0	10.0
Noes	57.1	42.9	0.0	0.0	0.0	0.0	0.0	4.3	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	42.9	35.7	14.3	0.0	7.1	0.0
Com	55.2	44.8	0.0	0.0	0.0	0.0	0.0	0.0	9.0	0.0	0.0	0.0	0.0	0.0	0.0	6.9	34.5	20.7	27.6	3.4	6.9	0.0
Tec	84.6	15.4	0.0	0.0	0.0	0.0	0.0	0.0	4.0	0.0	0.0	0.0	0.0	0.0	0.0	7.7	30.8	7.7	30.8	0.0	15.4	7.7
Prof	68.0	32.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	7.8	0.0	0.0	0.0	0.0	8.0	20.0	16.0	40.0	4.0	12.0	0.0
Dep	50.0	50.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	3.7	0.0	0.0	0.0	8.3	16.7	33.3	33.3	0.0	0.0	8.3
Obr	75.0	25.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	3.7	0.0	0.0	8.3	50.0	16.7	25.0	0.0	0.0	0.0
Empl	44.8	55.2	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	20.8	0.0	4.5	20.9	17.9	29.9	7.5	16.4	3.0
Jefe	60.0	40.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	3.1	0.0	0.0	20.0	10.0	10.0	40.0	20.0
Prin	56.0	44.0	0.0	0.0	20.0	40.0	0.0	0.0	8.0	4.0	8.0	4.0	4.0	12.0	0.0	7.8	0.0	0.0	0.0	0.0	0.0	0.0
Prco	49.5	50.5	2.2	0.0	30.8	13.2	2.2	6.6	11.0	4.4	5.5	2.2	6.6	15.4	0.0	0.0	28.3	0.0	0.0	0.0	0.0	0.0
Sein	43.5	56.5	4.3	14.5	20.3	7.2	1.4	7.2	8.7	1.4	5.8	5.8	2.9	17.4	2.9	0.0	0.0	21.4	0.0	0.0	0.0	0.0
Seco	50.0	50.0	3.8	2.6	9.0	11.5	6.4	2.6	10.3	5.1	12.8	5.1	3.8	25.6	1.3	0.0	0.0	0.0	24.2	0.0	0.0	0.0
Tein	26.7	73.3	0.0	46.7	0.0	0.0	0.0	0.0	6.7	0.0	6.7	0.0	0.0	33.3	6.7	0.0	0.0	0.0	0.0	4.7	0.0	0.0
Teco	48.1	51.9	0.0	3.7	0.0	7.4	3.7	3.7	7.4	7.4	11.1	0.0	0.0	40.7	14.8	0.0	0.0	0.0	0.0	0.0	8.4	0.0
Unin	41.2	58.8	0.0	47.1	5.9	5.9	5.9	0.0	0.0	5.9	0.0	5.9	0.0	11.8	11.8	0.0	0.0	0.0	0.0	0.0	0.0	5.3
300	44.9	55.1	4.1	4.1	36.7	14.3	4.1	6.1	4.1	2.0	2.0	2.0	4.1	16.3	0.0	18.4	55.1	20.4	4.1	2.0	0.0	0.0
400	39.6	60.4	4.2	8.3	19.8	11.5	1.0	7.3	10.4	2.1	7.3	4.2	4.2	18.8	1.0	8.3	37.5	21.9	19.8	5.2	4.2	3.1
600	53.6	46.4	1.4	7.2	13.0	15.9	1.4	2.9	8.7	4.3	5.8	4.3	4.3	24.6	5.8	7.2	26.1	21.7	26.1	4.3	8.7	5.8
850	54.3	45.7	0.0	15.2	8.7	6.5	0.0	4.3	8.7	6.5	10.9	8.7	2.2	26.1	2.2	0.0	15.2	32.6	32.6	0.0	13.0	6.5
1250	30.0	70.0	0.0	6.7	13.3	16.7	10.0	0.0	6.7	3.3	10.0	0.0	3.3	26.7	3.3	6.7	6.7	6.7	33.3	6.7	26.7	13.3
2000	68.4	31.6	0.0	5.3	0.0	5.3	15.8	0.0	15.8	5.3	15.8	0.0	5.3	21.1	10.5	0.0	5.3	15.8	42.1	15.8	15.8	5.3
Nsnc	61.5	38.5	7.7	23.1	7.7	7.7	0.0	0.0	15.4	15.4	15.4	0.0	0.0	0.0	7.7	7.7	0.0	23.1	46.2	7.7	0.0	15.4
Una	100.0	0.0	10.0	0.0	0.0	60.0	0.0	0.0	0.0	10.0	0.0	0.0	10.0	10.0	0.0	40.0	20.0	20.0	20.0	0.0	0.0	0.0
Dos	47.5	52.5	5.0	7.5	32.5	25.0	0.0	0.0	2.5	0.0	15.0	0.0	0.0	12.5	0.0	12.5	37.5	17.5	22.5	0.0	5.0	5.0
3a4	51.0	49.0	1.4	6.9	15.9	6.9	3.4	6.2	11.0	5.5	6.9	5.5	3.4	21.4	5.5	3.4	21.4	26.2	25.5	5.5	11.0	6.9
5a6	44.0	56.0	2.4	11.9	11.9	8.3	4.8	4.8	11.9	3.6	9.5	3.6	4.8	20.2	2.4	7.1	34.5	17.9	23.8	7.1	6.0	3.6
+7	27.9	72.1	2.3	11.6	20.9	14.0	2.3	2.3	4.7	2.3	2.3	2.3	4.7	30.2	0.0	11.6	32.6	16.3	23.3	2.3	9.3	4.7
C1	69.0	31.0	0.0	6.9	0.0	10.3	13.8	0.0	6.9	10.3	24.1	0.0	0.0	13.8	13.8	0.0	10.3	6.9	34.5	10.3	20.7	17.2
C2	47.4	52.6	1.8	11.4	14.9	15.8	1.8	2.6	7.9	2.6	6.1	4.4	1.8	25.4	3.5	5.3	24.6	16.7	30.7	4.4	10.5	7.9
D1	48.3	51.7	5.1	7.6	17.8	11.0	1.7	4.2	11.0	4.2	6.8	3.4	4.2	21.2	1.7	10.2	29.7	26.3	19.5	5.1	7.6	1.7
D2	34.4	65.6	0.0	6.6	27.9	8.2	3.3	9.8	8.2	3.3	4.9	4.9	8.2	14.8	0.0	11.5	41.0	27.9	16.4	1.6	0.0	1.6
JefS	400	600	Des	Est	Ama	Jub	Empr	Noes	Com	Tec	Prof	Dep	Obr	Empl	Jefe	Prin	Prco	Sein	Seco	Tein	Teco	Unin
Sex1	11.7	29.6	21.6	17.9	5.6	9.3	4.3	3.7	9.3	45.1	29.6	12.3	11.7	34.0	37.0	17.3	18.8	30.0	21.0	13.1	2.5	3.8
Sex2	18.8	30.0	12.0	10.6	13.1	2.5	3.8	2.5	15.6	45.0	22.5	14.4	6.2	36.9	36.2	20.6						
Int1	16.4	42.2	15.6	11.7	7.0	4.7	2.3	1.6	14.8	42.2	25.0	16.4	3.9	40.6	36.7	18.8						
Int2	15.0	21.0	24.6	16.2	10.8	7.8	4.8	4.2	11.4	46.1	25.7	12.6	13.8	34.7	32.3	19.2						
Int3	11.1	25.9	29.6	14.8	11.1	0.0	7.4	3.7	7.4	51.9	33.3	3.7	3.7	14.8	63.0	18.5						
Cla1	14.3	31.4	21.7	13.7	8.0	7.4	3.4	2.9	16.6	41.7	23.4	15.4	8.6	44.6	32.6	14.3						
Cla2	23.1	25.3	16.5	14.3	11.0	5.5	4.4	4.4	8.8	47.3	29.7	9.9	13.2	28.6	28.6	29.7						
Cla3	5.4	32.1	28.6	16.1	10.7	1.8	5.4	1.8	5.4	51.8	28.6	12.5	3.6	17.9	62.5	16.1						
Viol	15.3	33.1	13.7	15.3	9.7	9.7	3.2	1.6	13.7	45.2	26.6	12.9	12.1	37.9</								

Tabla 51. (Continuación)

	300	400	600	850	1250	2000	Nsnc	Una	Dos	3a4	5a6	+7	C1	C2	D1	D2
Coc1	17.5	31.9	18.7	13.9	9.0	6.6	2.4	1.8	12.7	42.2	27.7	15.7	7.8	38.0	33.7	20.5
Coc2	14.4	24.7	21.6	16.5	8.2	6.2	8.2	6.2	14.4	47.4	18.6	13.4	14.4	42.3	28.9	14.4
Coc3	10.2	32.2	28.8	11.9	11.9	3.4	1.7	1.7	8.5	49.2	33.9	6.8	3.4	16.9	57.6	22.0
Gre1	17.9	29.1	18.4	14.5	10.1	6.1	3.9	1.7	13.4	41.3	27.4	16.2	10.1	38.0	31.3	20.7
Gre2	13.0	31.9	23.2	14.5	5.8	7.2	4.3	5.8	15.9	50.7	17.4	10.1	11.6	46.4	26.1	15.9
Gre3	10.8	29.7	27.0	13.5	10.8	4.1	4.1	4.1	6.8	48.6	31.1	9.5	4.1	18.9	59.5	17.6
Edi1	18.5	31.0	20.2	13.1	8.3	6.5	2.4	2.4	12.5	42.3	25.0	17.9	6.5	37.5	35.1	20.8
Edi2	11.8	28.2	18.8	16.5	10.6	5.9	8.2	3.5	17.6	48.2	22.4	8.2	17.6	44.7	21.2	16.5
Edi3	11.6	29.0	27.5	14.5	10.1	4.3	2.9	4.3	5.8	47.8	33.3	8.7	4.3	18.8	59.4	17.4
Coel	20.6	36.9	19.1	9.9	5.7	5.7	2.1	2.1	12.8	39.7	28.4	17.0	2.8	36.9	38.3	22.0
Coe2	10.9	24.0	22.5	18.6	12.4	6.2	5.4	4.7	17.1	48.1	19.4	10.9	16.3	41.1	26.4	16.3
Coe3	11.5	25.0	25.0	15.4	11.5	5.8	5.8	1.9	0.0	51.9	36.5	9.6	7.7	17.3	57.7	17.3
Cor1	19.4	35.0	19.4	13.6	5.8	3.9	2.9	2.9	13.6	41.7	27.2	14.6	5.8	38.8	32.0	23.3
Cor2	13.6	27.8	20.4	14.2	13.0	7.4	3.7	3.7	16.0	45.1	22.2	13.0	11.7	40.7	31.5	16.0
Cor3	12.3	26.3	28.1	15.8	5.3	5.3	7.0	1.8	0.0	50.9	35.1	12.3	7.0	14.0	59.6	19.3
Agr1	17.5	31.8	17.5	15.6	8.4	5.8	3.2	1.9	14.3	39.6	28.6	15.6	6.5	36.4	33.8	23.4
Agr2	14.2	27.4	19.8	15.1	10.4	8.5	4.7	4.7	12.3	48.1	22.6	12.3	16.0	45.3	26.4	12.3
Agr3	11.3	29.0	33.9	9.7	9.7	1.6	4.8	3.2	8.1	53.2	25.8	9.7	3.2	16.1	61.3	19.4
Reg1	25.0	17.3	19.2	15.4	11.5	7.7	3.8	1.9	7.7	42.3	28.8	19.2	9.6	36.5	26.9	26.9
Reg2	13.6	35.7	18.1	13.6	9.0	5.4	4.5	3.6	14.9	44.3	24.4	12.7	10.4	38.5	32.1	19.0
Reg3	12.2	16.3	38.8	16.3	8.2	6.1	2.0	2.0	6.1	51.0	30.6	10.2	2.0	20.4	67.3	10.2
Inv1	21.4	33.3	15.1	12.7	8.7	7.1	1.6	1.6	13.5	42.1	23.0	19.8	4.8	38.1	35.7	21.4
Inv2	11.6	27.5	23.9	15.9	10.1	5.8	5.1	4.3	14.5	46.4	26.8	8.0	13.8	41.3	27.5	17.4
Inv3	10.3	27.6	29.3	13.8	8.6	3.4	6.9	3.4	5.2	48.3	31.0	12.1	6.9	15.5	60.3	17.2
Esp1	20.0	31.0	19.3	11.7	5.5	9.7	2.8	3.4	11.0	43.4	26.2	15.9	6.9	37.9	35.9	19.3
Esp2	11.4	26.8	19.5	18.7	14.6	4.1	4.9	2.4	16.3	44.7	24.4	12.2	13.0	43.9	25.2	17.9
Esp3	11.1	33.3	31.5	11.1	7.4	0.0	5.6	3.7	7.4	50.0	29.6	9.3	5.6	9.3	64.8	20.4
Soc1	21.5	26.4	19.8	11.6	7.4	9.1	4.1	4.1	12.4	38.0	28.1	17.4	8.3	37.2	31.4	23.1
Soc2	12.8	33.3	17.7	19.1	9.9	3.5	3.5	2.1	14.9	48.2	22.7	12.1	10.6	41.8	29.8	17.7
Soc3	8.3	28.3	33.3	8.3	11.7	5.0	5.0	3.3	6.7	51.7	30.0	8.3	6.7	16.7	63.3	13.3
Chi1	17.9	28.5	17.1	13.8	12.2	8.1	2.4	3.3	14.6	37.4	30.9	13.8	10.6	41.5	26.8	21.1
Chi2	14.5	33.1	20.5	15.1	6.6	4.8	5.4	3.0	11.4	47.6	24.1	13.9	9.0	36.1	34.3	20.5
Chi3	9.1	18.2	42.4	12.1	12.1	3.0	3.0	3.0	9.1	60.6	18.2	9.1	3.0	9.1	84.8	3.0
Sup1	23.1	32.5	17.1	9.4	8.5	6.8	2.6	2.6	11.1	46.2	25.6	14.5	6.8	39.3	29.9	23.9
Sup2	11.1	26.8	20.9	20.3	9.8	5.9	5.2	3.9	15.0	42.5	24.2	14.4	11.8	39.9	33.3	15.0
Sup3	9.6	32.7	32.7	7.7	9.6	3.8	3.8	1.9	7.7	50.0	32.7	7.7	5.8	13.5	61.5	19.2
JefS	14.5	25.0	24.3	16.4	5.9	8.6	5.3	6.6	12.5	48.7	24.3	7.9	13.2	35.5	37.5	13.8
JefN	15.9	34.1	18.8	12.4	12.4	3.5	2.9	0.0	12.4	41.8	27.6	18.2	5.3	35.3	35.9	23.5
Des	25.0	50.0	12.5	0.0	0.0	0.0	12.5	12.5	25.0	25.0	25.0	12.5	0.0	25.0	75.0	0.0
Est	7.1	28.6	17.9	25.0	7.1	3.6	10.7	0.0	10.7	35.7	35.7	17.9	7.1	46.4	32.1	14.3
Ama	32.7	34.5	16.4	7.3	7.3	0.0	1.8	0.0	23.6	41.8	18.2	16.4	0.0	30.9	38.2	30.9
Jub	17.9	28.2	28.2	7.7	12.8	2.6	2.6	15.4	25.6	25.6	17.9	15.4	7.7	46.2	33.3	12.8
Empr	20.0	10.0	10.0	0.0	30.0	30.0	0.0	0.0	0.0	50.0	40.0	10.0	40.0	20.0	20.0	20.0
Noes	21.4	50.0	14.3	14.3	0.0	0.0	0.0	0.0	0.0	64.3	28.6	7.1	0.0	21.4	35.7	42.9
Com	6.9	34.5	20.7	13.8	6.9	10.3	6.9	0.0	3.4	55.2	34.5	6.9	6.9	31.0	44.8	17.2
Tec	7.7	15.4	23.1	23.1	7.7	7.7	15.4	7.7	0.0	61.5	23.1	7.7	23.1	23.1	38.5	15.4
Prof	4.0	28.0	16.0	20.0	12.0	12.0	8.0	0.0	24.0	40.0	32.0	4.0	28.0	28.0	32.0	12.0
Dep	8.3	33.3	25.0	33.3	0.0	0.0	0.0	0.0	0.0	66.7	25.0	8.3	0.0	41.7	33.3	25.0
Obr	16.7	33.3	25.0	8.3	8.3	8.3	0.0	8.3	0.0	41.7	33.3	16.7	0.0	16.7	41.7	41.7
Empl	11.9	26.9	25.4	17.9	11.9	6.0	0.0	1.5	7.5	46.3	25.4	19.4	6.0	43.3	37.3	13.4
Jefe	0.0	10.0	40.0	10.0	10.0	20.0	10.0	0.0	0.0	80.0	20.0	0.0	40.0	40.0	20.0	0.0
Prin	36.0	32.0	20.0	0.0	8.0	0.0	4.0	16.0	20.0	20.0	24.0	20.0	0.0	24.0	48.0	28.0
Prco	29.7	39.6	19.8	7.7	2.2	1.1	0.0	2.2	16.5	34.1	31.9	15.4	3.3	30.8	38.5	27.5
Sein	14.5	30.4	21.7	21.7	2.9	4.3	4.3	2.9	10.1	55.1	21.7	10.1	2.9	27.5	44.9	24.6
Seco	2.6	24.4	23.1	19.2	12.8	10.3	7.7	2.6	11.5	47.4	25.6	12.8	12.8	44.9	29.5	12.8
Tein	6.7	33.3	20.0	0.0	13.3	20.0	6.7	0.0	0.0	53.3	40.0	6.7	20.0	33.3	40.0	6.7
Teco	0.0	14.8	22.2	22.2	29.6	11.1	0.0	0.0	7.4	59.3	18.5	14.8	22.2	44.4	33.3	0.0
Unin	0.0	17.6	23.5	17.6	23.5	5.9	11.8	0.0	11.8	58.8	17.6	11.8	29.4	52.9	11.8	5.9
300	15.2	0.0	0.0	0.0	0.0	0.0	0.0	8.2	16.3	38.8	24.5	12.2	0.0	16.3	42.9	40.8
400	0.0	29.8	0.0	0.0	0.0	0.0	0.0	4.2	13.5	45.8	27.1	9.4	3.1	25.0	41.7	30.2
600	0.0	0.0	21.4	0.0	0.0	0.0	0.0	1.4	13.0	49.3	26.1	10.1	4.3	43.5	44.9	7.2
850	0.0	0.0	0.0	14.3	0.0	0.0	0.0	0.0	8.7	50.0	23.9	17.4	6.5	52.2	28.3	13.0
1250	0.0	0.0	0.0	0.0	9.3	0.0	0.0	0.0	13.3	46.7	20.0	20.0	36.7	43.3	20.0	0.0
2000	0.0	0.0	0.0	0.0	0.0	5.9	0.0	5.3	0.0	26.3	47.4	21.1	31.6	47.4	21.1	0.0
Nsnc	0.0	0.0	0.0	0.0	0.0	0.0	4.0	0.0	15.4	46.2	15.4	23.1	23.1	46.2	23.1	7.7
Una	40.0	40.0	10.0	0.0	0.0	10.0	0.0	3.1	0.0	0.0	0.0	0.0	10.0	30.0	40.0	20.0
Dos	20.0	32.5	22.5	10.0	10.0	0.0	5.0	0.0	12.4	0.0	0.0	0.0	0.0	55.0	35.0	10.0
3a4	13.1	30.3	23.4	15.9	9.7	3.4	4.1	0.0	0.0	45.0	0.0	0.0	11.0	31.7	40.7	16.6
5a6	14.3	31.0	21.4	13.1	7.1	10.7	2.4	0.0	0.0	0.0	26.1	0.0	11.9	34.5	25.0	28.6
+7	14.0	20.9	16.3	18.6	14.0	9.3	7.0	0.0	0.0	0.0	0.0	13.4	4.7	32.6	46.5	16.3
C1	0.0	10.3	10.3	10.3	37.9	20.7	10.3	3.4	0.0	55.2	34.5	6.9	9.0	0.0	0.0	0.0
C2	7.0	21.1	26.3	21.1	11.4	7.9	5.3	2.6	19.3	40.4	2					



---



---

## CASOS DE ESTUDIO, PREGUNTAS Y PROBLEMAS

### *Caso 11.1: Perfil socioeconómico de lectores*

Con la información suministrada en las tablas Tabla 20 a Tabla 36, correspondientes a un AFCM de la base de datos provenientes del estudio para editar un nuevo diario regional, realice el análisis y elabore el perfil socioeconómico de los potenciales lectores del nuevo diario.

### *Caso 11.2: Las razas de perros*

A partir de los listados de resultados que se acompañan, identificar grupos de razas de perros que presentan características semejantes.

En particular se solicita analizar el primer plano factorial, identificar las coordenadas y contribuciones de las modalidades activas, seleccionar la mejor partición, identificar individuos característicos en cada grupo y caracterizar las clases por las modalidades y los individuos.

```
LECTURE DE LA BASE DE DONNEES
LECTURE DU FICHER BASE
NOM DE LA BASE      : E:213008-1.SBA
NOMBRE D'INDIVIDUS : 27
NOMBRE DE VARIABLES : 7
SELECTION DES INDIVIDUS ET DES VARIABLES UTILES
VARIABLES NOMINALES ACTIVES
  7 VARIABLES      19 MODALITES ASSOCIEES
-----
  1 . Taille      ( 3 MODALITES )
  2 . Poids      ( 3 MODALITES )
  3 . Vitesse     ( 3 MODALITES )
  4 . Intelligence ( 3 MODALITES )
  5 . Affectivite ( 2 MODALITES )
  6 . Agressevite ( 2 MODALITES )
  7 . Fonction    ( 3 MODALITES )
-----
INDIVIDUS
----- NOMBRE ----- POIDS ---
POIDS DES INDIVIDUS: Poids des individus, uniforme egal a 1.          UNIF
RETENUS ..... NITOT = 27      PITOT = 27.000
ACTIFS ..... NIACT = 27      PIACT = 27.000
SUPPLEMENTAIRES .... NISUP = 0      PISUP = 0.000
-----
STATISTIQUES USUELLES DES VARIABLES
TRIS A PLAT DES VARIABLES NOMINALES
----- EFFECTIFS -----
ABSOLU  %/TOTAL  %/EXPR.  HISTOGRAMME DES POIDS
-----
  1 . Taille
TP - Pequeño      7  25.93  25.93  *****
TM - Mediano     5  18.52  18.52  *****
TG - Grande     15  55.56  55.56  *****
ENSEMBLE      27 100.00 100.00
-----
  2 . Poids
PB - Bajo        8  29.63  29.63  *****
PM - Mediano    14  51.85  51.85  *****
PA - Alto        5  18.52  18.52  *****
ENSEMBLE      27 100.00 100.00
```

-----					
3 . Velocidad					
VB	- Baja	10	37.04	37.04	*****
VM	- Media	8	29.63	29.63	*****
VA	- Alta	9	33.33	33.33	*****
	ENSEMBLE	27	100.00	100.00	
-----					
4 . Inteligencia					
IB	- Baja	8	29.63	29.63	*****
IM	- Media	13	48.15	48.15	*****
IA	- Alta	6	22.22	22.22	*****
	ENSEMBLE	27	100.00	100.00	
-----					
5 . Afectividad					
AFL	- Leve	13	48.15	48.15	*****
AFA	- Alta	14	51.85	51.85	*****
	ENSEMBLE	27	100.00	100.00	
-----					
6 . Agresividad					
AGL	- Leve	14	51.85	51.85	*****
AGA	- Alta	13	48.15	48.15	*****
	ENSEMBLE	27	100.00	100.00	
-----					
7 . Función					
Com	- Compañía	10	37.04	37.04	*****
Caz	- Caza	9	33.33	33.33	*****
UTI	- Utilidad	8	29.63	29.63	*****
	ENSEMBLE	27	100.00	100.00	
-----					

LECTURE DE LA BASE DE DONNEES  
LECTURE DU FICHIER BASE  
NOM DE LA BASE : E:213008-1.SBA  
NOMBRE D'INDIVIDUS : 27  
NOMBRE DE VARIABLES : 7  
SELECTION DES INDIVIDUS ET DES VARIABLES UTILES  
VARIABLES NOMINALES ACTIVES  
7 VARIABLES 19 MODALITES ASSOCIEES

-----	
1 . Tamaño	( 3 MODALITES )
2 . Peso	( 3 MODALITES )
3 . Velocidad	( 3 MODALITES )
4 . Inteligencia	( 3 MODALITES )
5 . Afectividad	( 2 MODALITES )
6 . Agresividad	( 2 MODALITES )
7 . Función	( 3 MODALITES )
-----	

INDIVIDUS  
----- NOMBRE ----- POIDS ---  
POIDS DES INDIVIDUS: Poids des individus, uniforme egal a 1. UNIF  
RETENUS ..... NITOT = 27 PITOT = 27.000  
ACTIFS ..... NIACT = 27 PIACT = 27.000  
SUPPLEMENTAIRES .... NISUP = 0 PISUP = 0.000  
-----

ANALYSE DES CORRESPONDANCES MULTIPLES  
APUREMENT DES MODALITES ACTIVES  
SEUIL (PCMIN) : 2.00 % POIDS: 0.54  
AVANT APUREMENT : 7 QUESTIONS ACTIVES 19 MODALITES ASSOCIEES  
APRES : 7 QUESTIONS ACTIVES 19 MODALITES ASSOCIEES  
POIDS TOTAL DES INDIVIDUS ACTIFS : 27.00  
TRI-A-PLAT DES QUESTIONS ACTIVES

IDENT	MODALITES LIBELLE	AVANT APUREMENT		APRES APUREMENT		HISTOGRAMME DES POIDS RELATIFS
		EFF.	POIDS	EFF.	POIDS	
-----						
1 . Tamaño						
TP	- Pequeño	7	7.00	7	7.00	*****
TM	- Mediano	5	5.00	5	5.00	*****
TG	- Grande	15	15.00	15	15.00	*****
-----						
2 . Peso						
PB	- Bajo	8	8.00	8	8.00	*****
PM	- Mediano	14	14.00	14	14.00	*****
PA	- Alto	5	5.00	5	5.00	*****
-----						
3 . Velocidad						
VB	- Baja	10	10.00	10	10.00	*****
VM	- Media	8	8.00	8	8.00	*****
VA	- Alta	9	9.00	9	9.00	*****
-----						
4 . Inteligencia						
IB	- Baja	8	8.00	8	8.00	*****
IM	- Media	13	13.00	13	13.00	*****
IA	- Alta	6	6.00	6	6.00	*****
-----						
5 . Afectividad						
AFL	- Leve	13	13.00	13	13.00	*****
AFA	- Alta	14	14.00	14	14.00	*****
-----						
6 . Agresividad						
AGL	- Leve	14	14.00	14	14.00	*****
AGA	- Alta	13	13.00	13	13.00	*****
-----						
7 . Función						
Com	- Compañía	10	10.00	10	10.00	*****
Caz	- Caza	9	9.00	9	9.00	*****
UTI	- Utilidad	8	8.00	8	8.00	*****
-----						

TABLEAU DE BURST

	TP	TM	TG	PB	PM	PA	VB	VM	VA	IB	IM	IA	AFL	AFA	AGL	AGA	Com	Caz	UTI
TP	7	0	0																
TM	0	5	0																
TG	0	0	15																
PB	7	1	0	8	0	0													
PM	0	4	10	0	14	0													
PA	0	0	5	0	0	5													
VB	5	1	4	6	0	4	10	0	0										
VM	2	4	2	2	6	0	0	8	0										
VA	0	0	9	0	8	1	0	0	9										
IB	3	0	5	3	3	2	4	1	3	8	0	0							
IM	3	4	6	4	7	2	5	5	3	0	13	0							
IA	1	1	4	1	4	1	1	2	3	0	0	6							
AFL	7	5	1	8	5	0	6	6	1	3	7	3	13	0					
AFA	0	0	14	0	9	5	4	2	8	5	6	3	0	14					
AGL	5	3	6	5	8	1	5	5	4	3	8	3	8	6	14	0			
AGA	2	2	9	3	6	4	5	3	5	5	5	3	5	8	0	13			
Com	6	3	1	7	3	0	5	4	1	2	7	1	9	1	7	3	10	0	0
Caz	1	2	6	1	8	0	1	4	4	4	3	2	3	6	6	3	0	9	0
UTI	0	0	8	0	3	5	4	0	4	2	3	3	1	7	1	7	0	0	8

VALEURS PROPRES

APERCU DE LA PRECISION DES CALCULS : TRACE AVANT DIAGONALISATION .. 1.7143  
SOMME DES VALEURS PROPRES .... 1.7143

HISTOGRAMME DES 12 PREMIERES VALEURS PROPRES

NUMERO	VALEUR PROPRE	POURCENT.	POURCENT. CUMULE	
1	0.5464	31.87	31.87	*****
2	0.3548	20.70	52.57	*****
3	0.2126	12.40	64.98	*****
4	0.1546	9.02	73.99	*****
5	0.1449	8.45	82.44	*****
6	0.1137	6.63	89.08	*****
7	0.0704	4.11	93.19	*****
8	0.0561	3.27	96.46	*****
9	0.0292	1.70	98.16	*****
10	0.0194	1.13	99.30	***
11	0.0074	0.43	99.73	**
12	0.0047	0.27	100.00	*

RECHERCHE DE PALIERS (DIFFERENCES TROISIEMES)

PALIER ENTRE	VALEUR DU PALIER	
3-- 4	-69.75	*****
6-- 7	-41.63	*****
2-- 3	-35.86	*****

RECHERCHE DE PALIERS ENTRE (DIFFERENCES SECONDES)

PALIER ENTRE	VALEUR DU PALIER	
2-- 3	84.20	*****
1-- 2	49.33	*****
3-- 4	48.33	*****
6-- 7	28.98	*****
8-- 9	17.25	*****

COORDONNEES, CONTRIBUTIONS ET COSINUS CARRÉS DES MODALITES ACTIVES AXES 1 A 5

IDEN - LIBELLE	MODALITES		COORDONNEES					CONTRIBUTIONS					COSINUS CARRÉS					
	P.REL	DISTO	1	2	3	4	5	1	2	3	4	5	1	2	3	4	5	
1 . Tamaño																		
TP	- Pequeño	3.70	2.86	-1.28	0.71	0.63	0.27	-0.12	11.1	5.2	6.9	1.8	0.3	0.57	0.17	0.14	0.03	0.00
TM	- Mediano	2.65	4.40	-0.76	-1.10	-1.19	-0.16	0.62	2.8	9.1	17.6	0.4	7.0	0.13	0.28	0.32	0.01	0.09
TG	- Grande	7.94	0.80	0.85	0.04	0.10	-0.07	-0.15	10.5	0.0	0.4	0.3	1.3	0.90	0.00	0.01	0.01	0.03
CONTRIBUTION CUMULEE = 24.3 14.3 24.9 2.5 8.6																		
2 . Peso																		
PB	- Bajo	4.23	2.38	-1.26	0.66	0.38	0.20	-0.03	12.2	5.2	2.9	1.1	0.0	0.67	0.18	0.06	0.02	0.00
PM	- Mediano	7.41	0.93	0.35	-0.84	0.08	0.02	-0.03	1.7	14.6	0.2	0.0	0.1	0.13	0.75	0.01	0.00	0.00
PA	- Alto	2.65	4.40	1.02	1.29	-0.83	-0.37	0.15	5.1	12.3	8.7	2.4	0.4	0.24	0.38	0.16	0.03	0.01
CONTRIBUTION CUMULEE = 19.0 32.1 11.8 3.5 0.5																		
3 . Velocidad																		
VB	- Baja	5.29	1.70	-0.37	1.04	-0.10	-0.24	0.00	1.3	16.2	0.2	2.0	0.0	0.08	0.64	0.01	0.03	0.00
VM	- Media	4.23	2.38	-0.55	-0.94	-0.46	0.00	0.50	2.3	10.5	4.2	0.0	7.2	0.13	0.37	0.09	0.00	0.10
VA	- Alta	4.76	2.00	0.90	-0.32	0.52	0.27	-0.44	7.0	1.4	6.0	2.3	6.4	0.40	0.05	0.13	0.04	0.10
CONTRIBUTION CUMULEE = 10.7 28.1 10.4 4.3 13.6																		



CLASSIFICATION HIERARCHIQUE (VOISINS RECIPROQUES)  
SUR LES 10 PREMIERS AXES FACTORIELS

DESCRIPTION DES NOEUDS

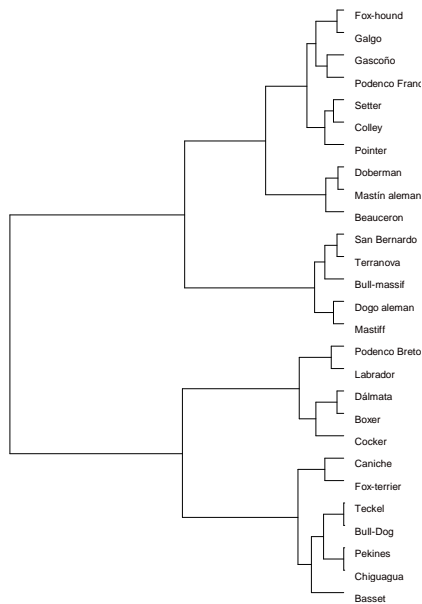
NUM.	AINE	BENJ	EFF.	POIDS	INDICE	HISTOGRAMME DES INDICES DE NIVEAU
28	20	7	2	2.00	0.00000	*
29	26	4	2	2.00	0.00000	*
30	11	18	2	2.00	0.00651	**
31	10	3	2	2.00	0.01026	**
32	13	15	2	2.00	0.01026	**
33	25	27	2	2.00	0.01026	**
34	24	9	2	2.00	0.01503	***
35	12	19	2	2.00	0.01505	***
36	21	17	2	2.00	0.01738	***
37	16	22	2	2.00	0.02496	****
38	30	2	3	3.00	0.02555	*****
39	33	5	3	3.00	0.02653	*****
40	6	14	2	2.00	0.02773	*****
41	34	23	3	3.00	0.02819	*****
42	29	28	4	4.00	0.02860	*****
43	32	37	4	4.00	0.04056	*****
44	31	8	3	3.00	0.04102	*****
45	39	35	5	5.00	0.04217	*****
46	42	1	5	5.00	0.04677	*****
47	43	41	7	7.00	0.05387	*****
48	36	44	5	5.00	0.06583	*****
49	40	46	7	7.00	0.06797	*****
50	47	38	10	10.00	0.11626	*****
51	50	45	15	15.00	0.23823	*****
52	48	49	12	12.00	0.24083	*****
53	51	52	27	27.00	0.50239	*****

SOMME DES INDICES DE NIVEAU = 1.70221

DESCRIPTION DES NOEUDS DE LA HIERACHIE  
(INDICES EN POURCENTAGE DE LA SOMME DES INDICES : 1.70221)

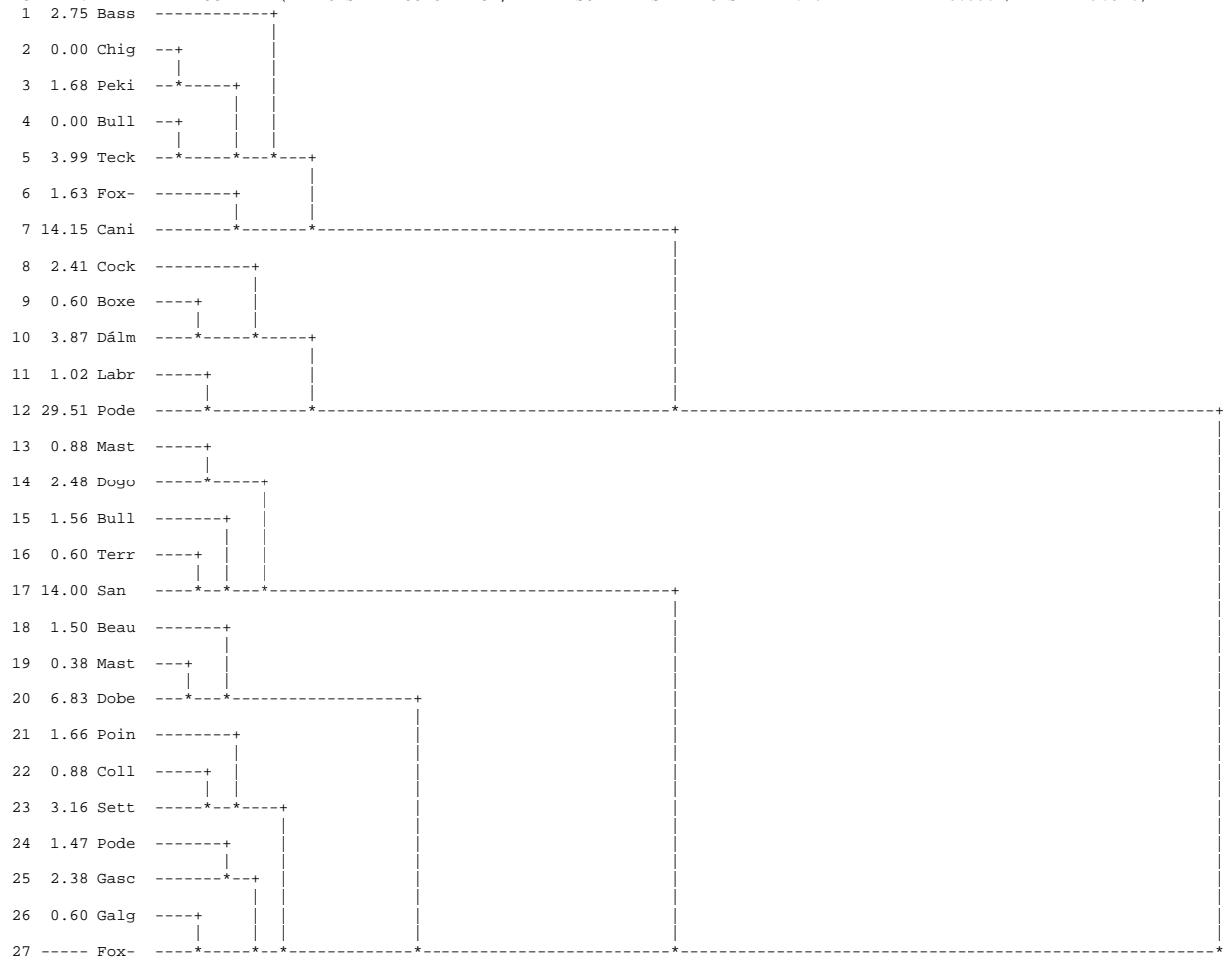
NOEUD NUMERO	INDICE	SUCESSEURS		EFFECT.	POIDS	COMPOSITION	
		AINE	BENJ			PREMIER	DERNIER
28	0.00	3	2	2	2.00	2	3
29	0.00	5	4	2	2.00	4	5
30	0.38	20	19	2	2.00	19	20
31	0.60	10	9	2	2.00	9	10
32	0.60	27	26	2	2.00	26	27
33	0.60	17	16	2	2.00	16	17
34	0.88	23	22	2	2.00	22	23
35	0.88	14	13	2	2.00	13	14
36	1.02	12	11	2	2.00	11	12
37	1.47	25	24	2	2.00	24	25
38	1.50	30	18	3	3.00	18	20
39	1.56	33	15	3	3.00	15	17
40	1.63	7	6	2	2.00	6	7
41	1.66	34	21	3	3.00	21	23
42	1.68	29	28	4	4.00	2	5
43	2.38	32	37	4	4.00	24	27
44	2.41	31	8	3	3.00	8	10
45	2.48	39	35	5	5.00	13	17
46	2.75	42	1	5	5.00	1	5
47	3.16	43	41	7	7.00	21	27
48	3.87	36	44	5	5.00	8	12
49	3.99	40	46	7	7.00	1	7
50	6.83	47	38	10	10.00	18	27
51	14.00	50	45	15	15.00	13	27
52	14.15	48	49	12	12.00	1	12
53	29.51	51	52	27	27.00	1	27

Classification hierarchique directe



## DENDROGRAMME

RANG IND. IDEN DENDROGRAMME (INDICES EN POURCENTAGE, DE LA SOMME DES INDICES : 1.70221 MIN = 0.00% / MAX = 29.51%)

PARTITION PAR COUPURE D'UN ARBRE HIERARCHIQUE  
RECHERCHE DES MEILLEURES PARTITIONS  
RECHERCHE DES PALIERS

PALIER ENTRE	VALEUR DU PALIER	
49-- 50	-55.95	*****
50-- 51	-27.53	*****

## LISTE DES 2 MEILLEURE(S) PARTITION(S) ENTRE 3 ET 10 CLASSES

- 1 - PARTITION EN 5 CLASSES
- 2 - PARTITION EN 4 CLASSES

COUPURE 'a' DE L'ARBRE EN 4 CLASSES  
FORMATION DES CLASSES (INDIVIDUS ACTIFS)  
DESCRIPTION SOMMAIRE

CLASSE	EFFECTIF	POIDS	CONTENU
aa1a	7	7.00	1 A 7
aa2a	5	5.00	8 A 12
aa3a	5	5.00	13 A 17
aa4a	10	10.00	18 A 27

COORDONNEES ET VALEURS-TEST AVANT CONSOLIDATION  
AXES 1 A 5

IDEN - LIBELLE	CLASSES		VALEURS-TEST					COORDONNEES					DISTO.
	EFF.	P. ABS	1	2	3	4	5	1	2	3	4	5	
COUPURE 'a' DE L'ARBRE EN 4 CLASSES													
aa1a - CLASSE 1 / 4	7	7.00	-3.9	2.1	1.9	0.8	-0.3	-0.95	0.42	0.29	0.11	-0.04	1.18
aa2a - CLASSE 2 / 4	5	5.00	-1.8	-2.7	-2.9	-0.4	1.5	-0.56	-0.66	-0.55	-0.06	0.24	1.16
aa3a - CLASSE 3 / 4	5	5.00	2.5	3.1	-2.0	-0.9	0.4	0.76	0.77	-0.38	-0.15	0.06	1.40
aa4a - CLASSE 4 / 4	10	10.00	3.0	-2.3	2.2	0.3	-1.2	0.56	-0.35	0.26	0.03	-0.12	0.55

CONSOLIDATION DE LA PARTITION  
AUTOUR DES 4 CENTRES DE CLASSES, REALISEE PAR 10 ITERATIONS A CENTRES MOBILES  
PROGRESSION DE L'INERTIE INTER-CLASSES

ITERATION	I. TOTALE	I. INTER	QUOTIENT
0	1.70221	0.98145	0.57657
1	1.70221	0.98145	0.57657
2	1.70221	0.98145	0.57657

ARRET APRES L'ITERATION 2 L'ACCROISSEMENT DE L'INERTIE INTER-CLASSES  
PAR RAPPORT A L'ITERATION PRECEDENTE N'EST QUE DE 0.000 %.  
DECOMPOSITION DE L'INERTIE  
CALCULEE SUR 10 AXES.

INERTIES	INERTIES		EFFECTIFS		POIDS		DISTANCES	
	AVANT	APRES	AVANT	APRES	AVANT	APRES	AVANT	APRES
INTER-CLASSES	0.9814	0.9814						
INTRA-CLASSE								
CLASSE 1 / 4	0.1711	0.1711	7	7	7.00	7.00	1.1789	1.1789
CLASSE 2 / 4	0.1345	0.1345	5	5	5.00	5.00	1.1572	1.1572
CLASSE 3 / 4	0.0940	0.0940	5	5	5.00	5.00	1.4012	1.4012
CLASSE 4 / 4	0.3212	0.3212	10	10	10.00	10.00	0.5455	0.5455
TOTALE	1.7022	1.7022						

QUOTIENT (INERTIE INTER / INERTIE TOTALE) : AVANT ... 0.5766  
APRES ... 0.5766

COORDONNEES ET VALEURS-TEST APRES CONSOLIDATION  
AXES 1 A 5

IDEN - LIBELLE	EFF.	P.ABS	VALEURS-TEST					COORDONNEES					DISTO.
			1	2	3	4	5	1	2	3	4	5	
COUPURE 'a' DE L'ARBRE EN 4 CLASSES													
aa1a - CLASSE 1 / 4	7	7.00	-3.9	2.1	1.9	0.8	-0.3	-0.95	0.42	0.29	0.11	-0.04	1.18
aa2a - CLASSE 2 / 4	5	5.00	-1.8	-2.7	-2.9	-0.4	1.5	-0.56	-0.66	-0.55	-0.06	0.24	1.16
aa3a - CLASSE 3 / 4	5	5.00	2.5	3.1	-2.0	-0.9	0.4	0.76	0.77	-0.38	-0.15	0.06	1.40
aa4a - CLASSE 4 / 4	10	10.00	3.0	-2.3	2.2	0.3	-1.2	0.56	-0.35	0.26	0.03	-0.12	0.55

COMPOSITION DE: COUPURE 'a' DE L'ARBRE EN 4 CLASSES

CLASSE 1 / 4  
Bass Bull Cani Chig Fox- Peki Teck  
CLASSE 2 / 4  
Boxe Cock Dálm Labr Pode  
CLASSE 3 / 4  
Bull Dogo Mast San Terr  
CLASSE 4 / 4  
Beau Coll Dobe Fox- Galg Gasc Mast Pode Poin Sett

PARANGONS

CLASSE 1 / 4  
EFFECTIF: 7

RG	DISTANCE	IDENT.	RG	DISTANCE	IDENT.
1	0.33032	Bull-Dog	2	0.33032	Teckel
3	0.35815	Chiguagua	4	0.35815	Pekines
5	0.94877	Fox-terrier	6	1.11084	Caniche
7	1.18240	Basset			

CLASSE 2 / 4  
EFFECTIF: 5

RG	DISTANCE	IDENT.	RG	DISTANCE	IDENT.
1	0.32291	Dálmata	2	0.40828	Boxer
3	0.48679	Labrador	4	1.04886	Podenco Breton
5	1.36425	Cocker			

CLASSE 3 / 4  
EFFECTIF: 5

RG	DISTANCE	IDENT.	RG	DISTANCE	IDENT.
1	0.26103	San Bernardo	2	0.30191	Mastiff
3	0.58457	Bull-massif	4	0.60302	Terranova
5	0.78773	Dogo aleman			

CLASSE 4 / 4  
EFFECTIF: 10

RG	DISTANCE	IDENT.	RG	DISTANCE	IDENT.
1	0.50696	Setter	2	0.63278	Fox-hound
3	0.64126	Galgo	4	0.70280	Pointer
5	0.74629	Beauceron	6	0.89358	Colley
7	0.93845	Doberman	8	1.04516	Podenco Frances
9	1.18665	Gascoño	10	1.37827	Mastin aleman

COUPURE 'b' DE L'ARBRE EN 5 CLASSES  
FORMATION DES CLASSES (INDIVIDUS ACTIFS)  
DESCRIPTION SOMMAIRE

CLASSE	EFFECTIF	POIDS	CONTENU
bb1b	7	7.00	1 A 7
bb2b	5	5.00	8 A 12
bb3b	5	5.00	13 A 17
bb4b	3	3.00	18 A 20
bb5b	7	7.00	21 A 27

COORDONNEES ET VALEURS-TEST AVANT CONSOLIDATION  
AXES 1 A 5

IDEN - LIBELLE	CLASSES		VALEURS-TEST					COORDONNEES					DISTO.
	EFF.	P.ABS	1	2	3	4	5	1	2	3	4	5	
COUPURE 'b' DE L'ARBRE EN 5 CLASSES													
bb1b - CLASSE 1 / 5	7	7.00	-3.9	2.1	1.9	0.8	-0.3	-0.95	0.42	0.29	0.11	-0.04	1.18
bb2b - CLASSE 2 / 5	5	5.00	-1.8	-2.7	-2.9	-0.4	1.5	-0.56	-0.66	-0.55	-0.06	0.24	1.16
bb3b - CLASSE 3 / 5	5	5.00	2.5	3.1	-2.0	-0.9	0.4	0.76	0.77	-0.38	-0.15	0.06	1.40
bb4b - CLASSE 4 / 5	3	3.00	1.9	-0.2	-0.6	2.6	-0.9	0.76	-0.06	-0.14	0.58	-0.19	1.21
bb5b - CLASSE 5 / 5	7	7.00	2.0	-2.4	2.9	-1.6	-0.7	0.48	-0.47	0.44	-0.20	-0.08	0.71

CONSOLIDATION DE LA PARTITION  
AUTOUR DES 5 CENTRES DE CLASSES, REALISEE PAR 10 ITERATIONS A CENTRES MOBILES  
PROGRESSION DE L'INERTIE INTER-CLASSES

ITERATION	I.TOTALE	I.INTER	QUOTIENT
0	1.70221	1.09771	0.64487
1	1.70221	1.09771	0.64487
2	1.70221	1.09771	0.64487

ARRET APRES L'ITERATION 2 L'ACCROISSEMENT DE L'INERTIE INTER-CLASSES  
PAR RAPPORT A L'ITERATION PRECEDENTE N'EST QUE DE 0.000 %.  
DECOMPOSITION DE L'INERTIE  
CALCULEE SUR 10 AXES.

INERTIES	INERTIES		EFFECTIFS		POIDS		DISTANCES	
	AVANT	APRES	AVANT	APRES	AVANT	APRES	AVANT	APRES
INTER-CLASSES	1.0977	1.0977						
INTRA-CLASSE								
CLASSE 1 / 5	0.1711	0.1711	7	7	7.00	7.00	1.1789	1.1789
CLASSE 2 / 5	0.1345	0.1345	5	5	5.00	5.00	1.1572	1.1572
CLASSE 3 / 5	0.0940	0.0940	5	5	5.00	5.00	1.4012	1.4012
CLASSE 4 / 5	0.0321	0.0321	3	3	3.00	3.00	1.2102	1.2102
CLASSE 5 / 5	0.1729	0.1729	7	7	7.00	7.00	0.7090	0.7090
TOTALE	1.7022	1.7022						

QUOTIENT (INERTIE INTER / INERTIE TOTALE) : AVANT ... 0.6449  
APRES ... 0.6449

COORDONNEES ET VALEURS-TEST APRES CONSOLIDATION  
AXES 1 A 5

IDEN - LIBELLE	CLASSES		VALEURS-TEST					COORDONNEES					DISTO.
	EFF.	P.ABS	1	2	3	4	5	1	2	3	4	5	
COUPURE 'b' DE L'ARBRE EN 5 CLASSES													
bb1b - CLASSE 1 / 5	7	7.00	-3.9	2.1	1.9	0.8	-0.3	-0.95	0.42	0.29	0.11	-0.04	1.18
bb2b - CLASSE 2 / 5	5	5.00	-1.8	-2.7	-2.9	-0.4	1.5	-0.56	-0.66	-0.55	-0.06	0.24	1.16
bb3b - CLASSE 3 / 5	5	5.00	2.5	3.1	-2.0	-0.9	0.4	0.76	0.77	-0.38	-0.15	0.06	1.40
bb4b - CLASSE 4 / 5	3	3.00	1.9	-0.2	-0.6	2.6	-0.9	0.76	-0.06	-0.14	0.58	-0.19	1.21
bb5b - CLASSE 5 / 5	7	7.00	2.0	-2.4	2.9	-1.6	-0.7	0.48	-0.47	0.44	-0.20	-0.08	0.71

COMPOSITION DE: COUPURE 'b' DE L'ARBRE EN 5 CLASSES

CLASSE 1 / 5  
Bass Bull Cani Chig Fox- Peki Teck  
CLASSE 2 / 5  
Boxe Cock Dalm Labr Pode  
CLASSE 3 / 5  
Bull Dogo Mast San Terr  
CLASSE 4 / 5  
Beau Dobe Mast  
CLASSE 5 / 5  
Coll Fox- Galg Gasc Pode Poin Sett

PARANGONS  
CLASSE 1 / 5  
EFFECTIF: 7

RG	DISTANCE	IDENT.	RG	DISTANCE	IDENT.
1	0.33032	Bull-Dog	2	0.33032	Teckel
3	0.35815	Chiguagua	4	0.35815	Pekines
5	0.94877	Fox-terrier	6	1.11084	Caniche
7	1.18240	Basset			



CLASSE 2/ 5  
EFFECTIF: 5

RG	DISTANCE	IDENT.	RG	DISTANCE	IDENT.
1	0.32291	Dálmata	2	0.40828	Boxer
3	0.48679	Labrador	4	1.04886	Podenco Breton
5	1.36425	Cocker			

CLASSE 3/ 5  
EFFECTIF: 5

RG	DISTANCE	IDENT.	RG	DISTANCE	IDENT.
1	0.26103	San Bernardo	2	0.30191	Mastiff
3	0.58457	Bull-massif	4	0.60302	Terranova
5	0.78773	Dogo aleman			

CLASSE 4/ 5  
EFFECTIF: 3

RG	DISTANCE	IDENT.	RG	DISTANCE	IDENT.
1	0.14145	Doberman	2	0.26430	Mastín aleman
3	0.45996	Beauceron			

CLASSE 5/ 5  
EFFECTIF: 7

RG	DISTANCE	IDENT.	RG	DISTANCE	IDENT.
1	0.33565	Setter	2	0.35930	Galgo
3	0.60140	Fox-hound	4	0.71358	Podenco Frances
5	0.75035	Pointer	6	0.91219	Colley
7	0.99501	Gascoño			

DESCRIPTION DE PARTITION(S)  
DESCRIPTION DE LA COUPURE 'a' DE L'ARBRE EN 4 CLASSES  
CARACTERISATION DES CLASSES PAR LES MODALITES  
CARACTERISATION PAR LES MODALITES DES CLASSES OU MODALITES  
DE COUPURE 'a' DE L'ARBRE EN 4 CLASSES  
CLASSE 1 / 4

V.TEST	PROBA	POURCENTAGES			MODALITES	DES VARIABLES		IDEN	POIDS
		CLA/	MOD/	GLOBAL	CARACTERISTIQUES				
				25.93	CLASSE 1 / 4			aa1a	7
4.73	0.000	100.00	100.00	25.93	Pequeño	Tamaño		TP	7
4.29	0.000	87.50	100.00	29.63	Bajo	Peso		PB	8
2.89	0.002	53.85	100.00	48.15	Leve	Afectividad		AFL	13
2.64	0.004	60.00	85.71	37.04	Compañía	Función		Com	10

CLASSE 2 / 4

V.TEST	PROBA	POURCENTAGES			MODALITES	DES VARIABLES		IDEN	POIDS
		CLA/	MOD/	GLOBAL	CARACTERISTIQUES				
				18.52	CLASSE 2 / 4			aa2a	5
4.22	0.000	100.00	100.00	18.52	Mediano	Tamaño		TM	5

CLASSE 3 / 4

V.TEST	PROBA	POURCENTAGES			MODALITES	DES VARIABLES		IDEN	POIDS
		CLA/	MOD/	GLOBAL	CARACTERISTIQUES				
				18.52	CLASSE 3 / 4			aa3a	5
4.22	0.000	100.00	100.00	18.52	Alto	Peso		PA	5
3.20	0.001	62.50	100.00	29.63	Utilidad	Función		UTI	8

CLASSE 4 / 4

V.TEST	PROBA	POURCENTAGES			MODALITES	DES VARIABLES		IDEN	POIDS
		CLA/	MOD/	GLOBAL	CARACTERISTIQUES				
				37.04	CLASSE 4 / 4			aa4a	10
3.68	0.000	71.43	100.00	51.85	Mediano	Peso		PM	14
3.59	0.000	88.89	80.00	33.33	Alta	Velocidad		VA	9
3.38	0.000	66.67	100.00	55.56	Grande	Tamaño		TG	15
2.73	0.003	64.29	90.00	51.85	Alta	Afectividad		AFA	14

DESCRIPTION DE LA COUPURE 'b' DE L'ARBRE EN 5 CLASSES  
CARACTERISATION DES CLASSES PAR LES MODALITES  
CARACTERISATION PAR LES MODALITES DES CLASSES OU MODALITES  
DE COUPURE 'b' DE L'ARBRE EN 5 CLASSES  
CLASSE 1 / 5

V.TEST	PROBA	POURCENTAGES			MODALITES	DES VARIABLES		IDEN	POIDS
		CLA/	MOD/	GLOBAL	CARACTERISTIQUES				
				25.93	CLASSE 1 / 5			bb1b	7
4.73	0.000	100.00	100.00	25.93	Pequeño	Tamaño		TP	7
4.29	0.000	87.50	100.00	29.63	Bajo	Peso		PB	8
2.89	0.002	53.85	100.00	48.15	Leve	Afectividad		AFL	13
2.64	0.004	60.00	85.71	37.04	Compañía	Función		Com	10

CLASSE 2 / 5										
V.TEST	PROBA	POURCENTAGES			MODALITES	DES VARIABLES	IDEN	POIDS		
		CLA/MOD	MOD/CLA	GLOBAL	CARACTERISTIQUES					
4.22	0.000	100.00	100.00	18.52	CLASSE 2 / 5	Tamaño		bb2b	5	5
				18.52	Mediano			TM		
CLASSE 3 / 5										
V.TEST	PROBA	POURCENTAGES			MODALITES	DES VARIABLES	IDEN	POIDS		
		CLA/MOD	MOD/CLA	GLOBAL	CARACTERISTIQUES					
4.22	0.000	100.00	100.00	18.52	CLASSE 3 / 5	Peso		bb3b	5	5
3.20	0.001	62.50	100.00	29.63	Alto	Función		PA	5	8
					Utilidad			UTI		
CLASSE 4 / 5										
V.TEST	PROBA	POURCENTAGES			MODALITES	DES VARIABLES	IDEN	POIDS		
		CLA/MOD	MOD/CLA	GLOBAL	CARACTERISTIQUES					
CLASSE 5 / 5										
V.TEST	PROBA	POURCENTAGES			MODALITES	DES VARIABLES	IDEN	POIDS		
		CLA/MOD	MOD/CLA	GLOBAL	CARACTERISTIQUES					
2.92	0.002	66.67	85.71	33.33	CLASSE 5 / 5	Función		bb5b	7	7
2.66	0.004	50.00	100.00	51.85	Caza	Peso		Caz	9	14
2.66	0.004	50.00	100.00	51.85	Mediano	Afectividad		PM	14	14
2.44	0.007	46.67	100.00	55.56	Alta	Tamaño		AFA	14	15
					Grande			TG		

---



---

## Bibliografía

- Crivisqui, E. (2002) Iniciación a los métodos estadísticos exploratorios multivariados. Université Libre de Bruxelles. Belgique.
- Crivisqui, E. (1993), Análisis factorial de Correspondencias: Un instrumento de Investigación en Ciencias sociales, Bruselas, LMTD Université Libre de Bruxelles, Asunción, Universidad Católica de Asunción.
- Escofier, B. y J. Pages (1992) Análisis Factoriales Simples y Múltiples: Objetivos, Métodos e Interpretación. Bilbao, Universidad del País Vasco.