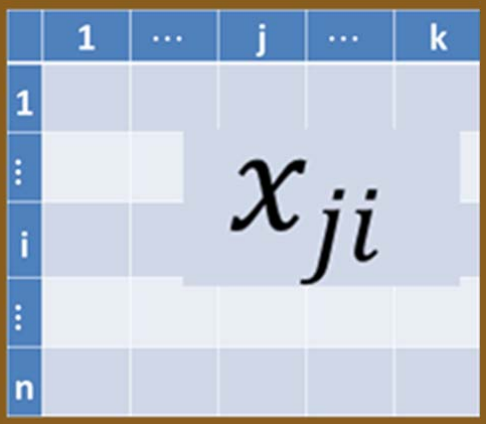


Planteo de una tabla de datos

PIE 2

ALFREDO MARIO BARONIO
ANA MARIA VIANCO

Cuadernos de Econometría

Planteo de una Tabla de Datos - PIE 2. Cuadernos de Econometría. Alfredo Mario Baronio y Ana María Vianco. 1ª edición. 2015

Xxx p; xxx cm

ISBN

1. Variables, 2 Observación 3 Probabilidad. 4 Estadística Aplicada

Fecha de ctalogación: xxx 2015

Planteo de una Tabla de Datos - PIE 2. Cuadernos de Econometría.
Alfredo Mario Baronio y Ana María Vianco.

Primera edición xxx 2015

ISBN XXXXXXX

Tirada xxx ejemplares

Esta edición es financiada con subsidios otorgados al proyecto Producción de Datos y Econometría Aplicada por la Secretaría de Ciencia y Técnica de la UNRC y el Instituto de Investigaciones de la UNVM.

Contenido

La Tabla de Datos resume las principales "operaciones de campo" que comporta todo proceso de observación. La definición de la investigación, planteada en el primer cuaderno, determina el conjunto de características a estudiar sobre unidades de observación (tiempo o individuos) que pertenecen a un espacio determinado. Las características de ese espacio se miden sobre una muestra de unidades de observación (cuyas propiedades y diseño se estudiarán en el tercer cuaderno) y los resultados de esa medición se disponen en una tabla de datos. Esto constituye la segunda etapa del proceso de investigación econométrica, que tiene en cuenta: 1. Enumeración de todas las variables relevantes; 2. Indicación del tipo de observaciones que van a utilizarse y 3. Elaboración del instrumento de recolección de datos.

1. DATOS, UNIDADES DE OBSERVACIÓN Y VARIABLES

En el proceso de investigación econométrica, se necesita plantear una tabla de datos, diseñar las fuentes de información y recolectar los datos, para luego pasar a la etapa de análisis. En lo que sigue se estudia uno de los elementos constitutivos y fundamentales de la investigación: la tabla de datos y las partes que le dan contenido -los datos, las unidades de observación y las variables-. Más adelante se verá como seleccionar las unidades de observación, como obtener los datos y, por último, como analizar información con el uso de ecuaciones a partir de la formulación de modelos.

1.1. Los tres pasos de observación de la realidad

El método econométrico requiere observar la realidad, en un espacio y tiempo determinado. Ello significa que necesariamente hay que transitar tres caminos para alcanzar los objetivos.

a) Primer Paso: Planteamiento de la Tabla de Datos. La definición de la investigación, 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 y los resultados de esa medición se disponen en una tabla de datos.

b) Segundo Paso: El Origen de los Datos. Para el diseño de

fuentes de datos se considera el espacio y tiempo en que se realiza la recolección de datos y el tipo de muestreo a utilizar. La información puede provenir de fuentes primarias, o la consulta y sistematización de fuentes secundarias para las unidades de observación. Para el primer caso habrá que considerar las propiedades estadísticas del diseño considerado.

c) Tercer Paso: La Recolección de Datos. Este paso incluye la organización de las tareas de relevamiento, la administración del instrumento de recolección de datos, el cálculo de la validez y confiabilidad del instrumento de medición, la codificación los datos y la creación de un archivo que los contenga, entre otros.

Estos tres pasos de "observación de la realidad" sintetizan el método a aplicar en la investigación econométrica, ex ante a la especificación de los modelos.

Al completar estos tres pasos, el método econométrico requiere definir los niveles de resolución a aplicar sobre la tabla de datos que, a esta altura, se encuentra completa y constituye un sistema. Así, un bajo nivel de resolución de un sistema socioeconómico se alcanza con la descripción de las características de los individuos. Un nivel de resolución más elevado puede considerar los individuos por bloques, las relaciones entre ellos y la explicación de su comportamiento.

A cada nivel de resolución corresponde, evidentemente, un modelo con un grado de desagregación diferente. Esta representación de los individuos en el espacio de las variables genera una nube de *puntos – individuos* que permitirá estudiar las relaciones entre las variables. De esto se ocupa la Econometría que, como quedó evidenciado, se nutre de la estadística, fundamentalmente, de la estadística matemática.

Con las unidades de observación, las variables y los datos, se puede analizar el comportamiento de los sujetos de la actividad económica a la luz de una teoría dada. Para esto se utiliza la econometría: comprobar si la teoría se cumple en espacio y tiempo determinado desde la observación hecha sobre la realidad. Al utilizarla, se usarán modelos matemáticos que, si

bien constituyen la forma más estricta de conocimiento científico de una realidad, ello no debe suponer, como dice Georgescu-Roegen, cit. por Pulido (1993), que "*su utilización indiscriminada "asfixie" toda elaboración teórica no directamente "matematizable" o, lo que es al menos tan perjudicial, encubra bajo su halo protector un conocimiento falso de la realidad aunque estrictamente planteado. En la ciencia en general, y en la economía en particular, siempre hay un límite a lo que se pueda hacer con números, así como hay un límite a lo que no se pueda hacer sin ellos"* (op. cit. pp 33). Por lo que, dice Pulido (1993), *"ser partidario de los modelos matemáticos no debe llevar el rechazo absoluto de todo tipo de modelo verbal o teoría expuesta en forma literaria"* (op. cit. pp 32).

La Tabla de Datos

Se puede establecer una interesante comparación entre el trabajo de un artista plástico con el de un investigador econométrico. Ambos necesitan algún recurso básico con el cual trabajar. Para el plástico, ese recurso es algún tipo de material. Para el investigador, el recurso básico son los datos. Así como el artista plástico utiliza diversos implementos para crear una obra con el material seleccionado, el investigador 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 trabajo econométrico se apreciará en un sentido extrínseco por como auxiliará a la solución de diversos problemas de administración, economía, mercados y otros relacionados con las ciencias humanas y sociales. Ahora bien, tanto para el artista como para el investigador, la calidad del producto final depende de la calidad de la materia prima utilizada. Por ello, el correcto planteo de la tabla de datos es importante en el proceso de investigación econométrica.

La Tabla de Datos es una tabla rectangular de $\boldsymbol{n}$ *individuos* por $\boldsymbol{k}$ *características* observadas de los mismos. Estas características

pueden ser *variables cuantitativas* o *variables cualitativas*. Las primeras tendrán propiedades numéricas las segundas no.

En síntesis, el proceso de investigación econométrica incorpora una tabla de datos que se construye siguiendo los lineamientos generales que se observan en la Figura 1.1.

Figura 1.1. Observación de la realidad

¿Para qué se construye una tabla de datos?

La tabla de datos se construye para, en la etapa de análisis de la información, evaluar: (1) La semejanza entre los individuos a través de los atributos seleccionados, por ejemplo obteniendo tipologías por medio de la aplicación de análisis factorial; y (2) La asociación entre las variables seleccionadas y observadas sobre el conjunto de unidades de observación, por ejemplo aplicando modelos econométricos u otros métodos de análisis confirmatorio.

¿Cómo relacionar las variables?

A través de ecuaciones o funciones, que pueden o no definir modelos económicos o econométricos.

¿Cómo queda representada la tabla de datos?

En las filas se incorporan las unidades de observación, ya sean individuos o tiempo. En las columnas se disponen los atributos o variables seleccionadas, sean estas cuantitativas o cualitativas. La Figura 1.2 muestra la ubicación de individuos y variables en la tabla de datos.

$U.O.$	X_1	X_2	$\cdots X_j \cdots$	X_k
1				
i			x_{ji}	
n				

- la $\boldsymbol{i}-$ésima línea contiene los valores que resultan de la "medición" de los k atributos sobre el $\boldsymbol{i}-$ésimo individuo o unidad de observación.
- la $\boldsymbol{j}-$ésima columna contiene los $\boldsymbol{n}$ valores que resultan de la "medición" del $\boldsymbol{j}-$ésimo atributo sobre las $\boldsymbol{n}$ unidades de observación.

Figura 1.2 La tabla de datos

1.2 Los datos

En realidad existe un vacío en la elaboración científica del dato que da origen a la información para la creación del conocimiento en las ciencias sociales. La fuente de los datos, ya sea primaria o secundaria, no ha sido metodológicamente ordenada para

facilitar el accionar de los investigadores en la econometría aplicada.

Existen datos que se generan a diario, pero todos se obtienen *in situ* sin contar con una guía metodológica general sobre la producción, disposición y posterior tratamiento de esa información. Es el principal *material* en la investigación econométrica y, por lo dicho anteriormente, queda determinado por la existencia de unidades de observación y variables, sin ellas no existe el dato.

Técnicamente, el dato es el valor que asume una unidad de observación para determinada variable. En general, la bibliografía especializada no le da al dato el lugar que merece, incluso se suelen encontrar confusiones importantes; por ejemplo, cuando se escribe sobre la media aritmética de los datos, en lugar de la media aritmética de las variables; o cuando se refiere a la dispersión de los datos, en lugar de la dispersión de las unidades de observación.

Según el diccionario de la Real Academia Española, *dato* es antecedente necesario para llegar al conocimiento exacto de una cosa o para deducir las consecuencias legítimas de un hecho o detalle que sirve de base a un razonamiento o a una investigación. Asimismo, se podría decir que el dato es una representación simbólica (numérica, alfabética, etc.), atributo o característica de una entidad. El dato no tiene valor semántico (sentido) en sí mismo, pero convenientemente tratado (procesado) se puede utilizar en la realización de cálculos o toma de decisiones.

Un dato por sí mismo no constituye información, es su inclusión en una tabla de datos lo que proporciona información. Se considera que un dato es una expresión mínima de contenido sobre un tema. La información representa un conjunto de datos relacionados que constituyen una estructura de mayor complejidad y que se define, en este cuaderno, como la tabla de datos.

Ejemplo 1.1. Datos.

-en Río Cuarto para el tercer trimestre de 2013, el ingreso promedio de un asalariado formal alcanza el nivel de $7000 mensuales,

-el título público BOGAR 18 cotizó a $141 el 7 de octubre de 2014,

-en 2013, el producto bruto geográfico de la Provincia de Córdoba alcanzó la cifra de $191212 millones de pesos a valores corrientes,

-al momento del Censo de Población y Vivienda de 2010, el departamento Juarez Celman tenía 61078 habitantes agrupados en 19745 hogares

El investigador econométrico, en primer lugar, tiene un problema de disponibilidad de información para poder realizar su investigación y, por último, tiene un problema de cómo analiza la información para poder tomar una decisión respecto a su problema a investigar. En el recorrido desde su problema inicial a su problema final tiene una situación de difícil o fácil resolución, según su experiencia, porque debe tratar con la dificultad de obtener información a partir de los datos. Ahora bien ¿cuándo el investigador logra superar este inconveniente? Cuando plantea correctamente su tabla de datos. En ese momento, *sus* datos se transforman en información útil para el análisis de su problema a investigar.

Kinnear y Taylor (1993), definen a la información como “los datos que disminuyen la incertidumbre en una situación de decisión”, continúan diciendo que la distinción se puede aclarar con un ejemplo, “… el capitán de un barco se enfrenta al problema de timonear su barco a un puerto difícil en la noche. Para ayudar el proceso de la toma de decisiones, el capitán ha recibido por radio los siguientes datos: (1) profundidad del canal, (2) velocidad y dirección del viento, (3) resultado del partido de béisbol local, (4) velocidad y dirección de la marea. Dado el problema del capitán, ¿cuáles datos podrían denominarse información?” La mayoría de los investigadores (capitanes) encuentran esta distinción relativamente fácil, pero en el caso de una situación de toma de decisiones típicas de las ciencias

económicas, la distinción entre datos e información se vuelve un desafío significativamente mayor.

La calidad de las decisiones depende, en gran parte, de la información para la persona que debe tomarlas. Es función de la investigación suministrar datos adecuados para esta toma de decisiones. Un investigador que no sepa usar o evaluar datos se encuentra muy limitado en sus habilidades para cumplir eficazmente su trabajo.

Por lo tanto, la producción de una tabla de datos no es un problema menor para un investigador, construirla significa pasar de tener datos a tener información y, para ello, deben transitar las etapas de planteamiento de la misma, determinación del origen de los datos, su recolección y su procesamiento.

En otras palabras, el dato se transforma en información cuando pertenece al recorrido de una variable de interés para el investigador y se refiere a alguna unidad de observación.

En este PIE se le da una importancia sustancial a los datos, desde la definición misma de 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, comprobadas a partir del estudio de la semejanza entre unidades y la relación entre variables, en un espacio y tiempo específico.

En palabras de Padua (1996), estos problemas subyacen en todas las Ciencias Sociales, "… la tradición de la investigación social ha prestado por lo general poca atención a los aspectos técnicos y a los instrumentos de los que se sirve el investigador para describir, explicar, predecir e interpretar su realidad. Esa desatención obedece a una serie de causas complejas, entre las que destacan el uso incorrecto del aparato técnico, la asociación entre las técnicas y cierto tipo de práctica profesional vinculada a modelos teóricos que no gozan de mucho prestigio en el área, el bajo nivel de profesionalismo en la práctica de la investigación, y

los prejuicios sobre un supuesto contenido ideológico intrínseco en las técnicas".

Se puede establecer que 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 en un espacio y tiempo determinado. También, de acuerdo con Padua (1996), el acopio de datos mediante relaciones directas o indirectas se utiliza en el contexto de distintos modelos teóricos. Los modelos o técnicas no pueden definirse a priori, sino que dependen del problema y de los tipos de interrogantes que la investigación plantea, del estado de avance de la teoría sustantiva, del planteamiento teórico general de la investigación. "Es la problemática teórica la que define tanto el objeto como los métodos con que se construye y apropia el objeto". Y en estos niveles, las técnicas de recolección de datos y la estadística son instrumentos valiosos.

El método del proceso de investigación econométrica consiste en describir los aspectos a tener en cuenta para definir, de manera adecuada, el problema bajo estudio y presentarlo en una tabla de datos que sea susceptible de tratamientos estadísticos para obtener nueva información.

Datos, información y conocimiento

Davenport y Prusak [1999] consideran que "los datos son la mínima unidad semántica, y se corresponden con elementos primarios de información que, por sí solos, son irrelevantes como apoyo a la toma de decisiones". También se pueden ver como un conjunto discreto de valores, que no dicen sobre el porqué de las cosas y no son orientativos para la acción.

Un número telefónico o el nombre de una persona, por ejemplo, son datos que, sin un propósito, una utilidad o un contexto no sirven como base para apoyar una investigación. Los datos pueden ser una colección de hechos almacenados en algún lugar físico como un papel, un dispositivo electrónico o la mente de

una persona. En este sentido, las tecnologías de la información han aportado mucho a la tarea de recopilación de datos.

La información se puede definir como un conjunto de datos procesados y que tienen un significado (relevancia, propósito y contexto), y que por lo tanto son de utilidad para quién debe tomar decisiones o realizar una investigación, al disminuir su incertidumbre. Los datos se pueden transformar en información añadiéndoles valor a partir de:

- Contextualizar: se sabe en qué contexto y para qué propósito se generaron
- Categorizar: se conocen las unidades de observación y las variables a las que pertenecen, lo que ayuda a interpretarlos
- Calcular: los datos pueden haber sido procesados matemática o estadísticamente
- Corregir: se han eliminado errores e inconsistencias de los datos
- Condensar: se han podido resumir en una tabla de datos

Por tanto, la información es la comunicación de conocimientos o inteligencia, y es capaz de cambiar la forma en que el receptor percibe algo, impactando sobre sus juicios de valor y sus comportamientos.

El conocimiento es una mezcla de experiencia, valores, información y *know-how* que sirve como marco para la incorporación de nuevas experiencias e información, es útil para la acción y se origina y aplica en la mente de los investigadores.

El conocimiento se deriva de la información, así como la información se deriva de los datos. Para que la información se convierta en conocimiento es necesario realizar acciones como:

- Comparación con otros elementos (estudio de la semejanza entre las unidades observadas)
- Búsqueda de conexiones (estudio de la asociación entre las variables)

- Comparación con otros portadores de conocimiento (elaboración de modelos y comparación de resultados)
- Predicción de consecuencias (predicción, pronósticos, análisis de sensibilidad, etc)

Uso incorrecto de los datos

El uso incorrecto de los datos puede proceder, por una parte, de una defectuosa interpretación de los requerimientos teóricos de la investigación. En la historia económica sobran los ejemplos de la mala utilización de los datos. Para dimensionarlo en su justa medida considérese la crítica que un famoso economista (Stöwe) realizaba a otro (Fisher) en los albores de la década del ’50 del siglo pasado. La crítica se dirige principalmente a probar que el período de tiempo utilizado no es correcto. Fisher aplica el modelo simplificado de Hicks a la economía estadounidense y obtiene un acelerador significativamente menor que la unidad. Según Stöwe debió haber utilizado otro período más adecuado; en efecto, al hacer la corrección oportuna prueba que la hipótesis de Hicks de que el acelerador es mayor que la unidad, puede aceptarse.

Por otra parte, los errores en el uso de los datos pueden deberse al insuficiente conocimiento de estos. Una vez definido el problema, en el que figura un cierto número de variables, hay que proceder a identificar los datos temporales o de corte transversal. Este proceso de identificación requiere con frecuencia someter los datos primarios a algunas operaciones, tales como obtener datos “per-cápita”, pasar de pesos corrientes a pesos constantes, etc. Unas veces, la teoría en que se apoya el investigador es lo suficientemente explícita como para poder efectuar dicha identificación de una manera correcta, pero otras, no. Entonces es posible que, por los motivos señalados, se introduzcan errores ajenos por completo a la metodología de investigación.

Es por todo esto que los objetivos de este trabajo se fundamentan en que, una vez planteado el problema, el investigador se enfrenta con la tarea de localizar u obtener los

datos necesarios y de optar por la naturaleza exacta de las variables y de las unidades de observación que se van a utilizar. De ello dependerá, entre otras cosas, el método de análisis de información que se empleará. Por lo que, las decisiones en este terreno han de basarse principalmente en la experimentación.

Por último, el conocimiento de los datos debe entenderse en el sentido de tener, al menos, una idea de los errores de medida que contienen. Datos con grandes errores –cosa bastante frecuente en muchas estadísticas–, aparte de las implicaciones que originan, conducen a tomar la decisión: *quizás no abandonarlos, pero solo utilizarlos como simple orientación*.

Tipos de datos en econometría

Los datos se pueden clasificar de la siguiente manera.

De acuerdo al tipo de variables de la tabla de datos:

a) Datos reales
b) Datos categóricos

Los datos reales pueden asumir valores en el campo de los números reales (puede ser números reales o números enteros).

Los datos categóricos a su vez se pueden clasificar en *datos binarios* (marcan ausencia o presencia de un atributo sobre el individuo observado) o en *códigos condensados* (definen con un número entero la modalidad que asume el individuo observado).

De acuerdo a la forma de escoger a las unidades de observación:

c) Datos de corte transversal
d) Datos de series temporales
e) Datos fusionados de sección cruzada (pool de datos)
f) Datos de panel

El Ejemplo 1.2 presenta las características que permiten identificar a cada uno de estos tipos de datos.

Ejemplo 1.2. Sea el modelo de consumo definido por la función

$$C_t = \alpha + \beta Y_t + \mu_t; \quad \forall\, t = 1, \dots, 100 \tag{1.1}$$

Este modelo se puede utilizar a cuatro niveles distintos:

a. A nivel agregado, en cuyo caso las variables C_t e Y_t serán indicadores del nivel de consumo y renta agregados, respectivamente. 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	1300	1400
2	1974.II	1330	1420
3	1974.III	1370	1460
4	1974.IV	890	975
⋮	⋮	⋮	⋮
100	1998.IV	1421	2175

Figura 1.3. Tabla de datos de serie de tiempo

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

$$C_i = \alpha + \beta Y_i + \mu_i; \quad \forall\, i = 1, \dots, 500 \tag{1.2}$$

En donde los subíndices indican que cada observación pertenece 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 corte transversal.

Se supone que estos datos se han obtenido de una muestra aleatoria de la población, en un espacio y tiempo específico. Si obtenemos la información de las 500 familias, se puede decir que se cuenta con una muestra aleatoria de toda la población que tiene un ingreso. El muestreo aleatorio y la forma de seleccionar una muestra se enseña en el PIE 3 y es de utilidad para el análisis de corte transversal. En economía los datos de corte transversal son principalmente utilizados en el análisis microeconómico.

MUESTRA	C_i	Y_i
Familia 1	2212	2500
Familia 2	2165	2800
Familia 3	1800	2000
⋮	⋮	⋮
Familia 500	2975	3300

Figura 1.4. Tabla de datos de corte transversal

Familia	Tiempo	C_{it}	Y_{it}
	1974.I	1300	1400
1	1974.II	1380	1400
⋮	⋮	⋮	⋮
	1998.IV	1550	1600
	1974.I	1000	2000
2	1974.II	1110	2050
⋮	⋮	⋮	⋮
	1998.IV	1189	2200
	⋮	⋮	⋮
	1974.I	2330	3420
500	1974.II	2550	3800
⋮	⋮	⋮	⋮
	1998.IV	3000	3850

Figura 1.5. Tabla de datos de panel

c. A nivel de datos fusionados de sección cruzada, donde tienen características tanto de datos de corte transversal como de datos de series temporales. Se utilizan por ejemplo, para analizar efectos de políticas gubernamentales. El procedimiento consiste en recopilar datos anteriores y posteriores a un cambio de políticas y luego analizar la información como si se tuvieran datos

de corte transversal; la diferencia es que no se considera el valor de la variable analizada, sino el valor de la diferencia observada entre la variable medida antes del cambio y posterior al cambio implementado. Con esto se puede observar cómo una relación clave ha cambiado con el tiempo.

d. 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.

$C_{it} = \alpha + \beta Y_{it} + \mu_{it}; \forall\, i = 1, \dots, 500; \ \forall\, t = 1974.I, \dots 1998.IV$ (1.3)

La característica de los datos de panel, que lo diferencia de los datos fusionados de sección cruzada, es que se mantiene un registro de las mismas unidades de sección cruzada, en este caso familias, durante un período de tiempo específico.

1.3 Unidades de observación

Las *unidades de observación* son aquellas de las que se obtienen los datos conforme a los atributos sobre ellas requeridos. Pueden representar a toda la población, en cuyo caso se denominan *elementos* y constituyen la base de un *censo;* o pueden representar parte de la población y constituyen la base de una *muestra*. Las unidades de observación pueden ser de tiempo, de corte trasversal, de fusión y de panel.

El espacio que contiene las unidades de observación, de la tabla de datos planteada, puede consistir en una población de individuos; por ejemplo, los estudiantes de la Facultad de Ciencias Económicas de la Universidad Nacional de Río Cuarto. A estos individuos se los suele denominar *elementos* de una población, en tanto y en cuanto son las unidades de observación elementales que forman la población acerca de la cual se van a realizar descripciones o representaciones estadísticas; son las unidades del análisis y su naturaleza se determina mediante los objetivos de la investigación.

Las unidades de observación se pueden representar gráficamente de acuerdo a la cantidad de atributos o variables sobre ellas observadas, las cuales definen las dimensiones del análisis. Si la tabla de datos es de $nx1$, esto es n unidades de observación y 1 variable, el gráfico es una recta; si es de $nx2$, el gráfico es un plano; si es de $nx3$, el gráfico representa al espacio. Cuando se tienen más de 3 variables, no se puede representar gráficamente, aunque, como se verá, las dimensiones se pueden reducir al plano aplicando análisis factoriales.

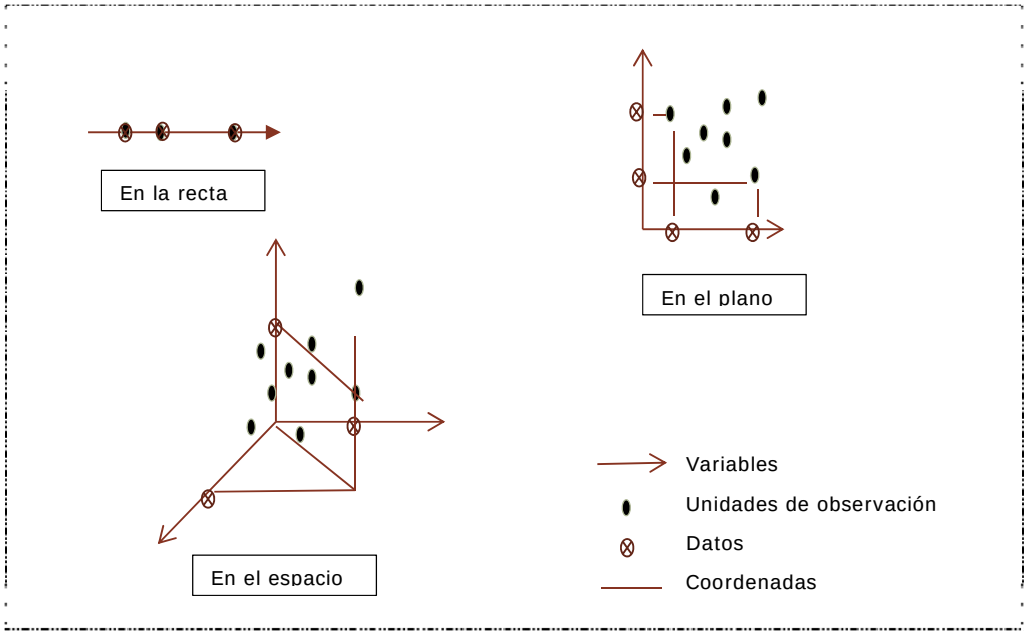


Figura 1.6. Gráficos de Dispersión de las unidades de observación

En la Figura 1.6 se representan simbólicamente los *gráficos de dispersión* de las unidades de observación según se observen sobre ellas una variable o atributo (en la recta, $\mathbb{R}$), dos variables (en el plano, $\mathbb{R}^2$), o tres variables (en el espacio $\mathbb{R}^3$). Para dicha representación se tienen en cuenta todos los elementos integrantes de la tabla de datos. La intersección de datos y variables da origen a las coordenadas que definen a las unidades de observación, en la recta, en el plano o en el espacio.

¿Cómo obtener las unidades de observación?

El objetivo de seleccionar unidades de observación es obtener una muestra de la población para producir los datos correspondientes a cada característica o variable estudiada sobre dicha población. Por lo tanto, cada unidad de observación (individuos o tiempo) va a tener asociada una cantidad de datos igual a la cantidad de variables que sobre ella se estudien.

De esta forma, seleccionar unidades de observación de una población se corresponde directamente a la obtención de datos para las variables estudiadas. En Ciencias Sociales, existen tres *métodos básicos* con los cuales el investigador puede obtener los datos deseados y que se corresponden a fuentes de información primarias o secundarias.

1) **uso de bases de datos ya publicados** (denominadas fuentes secundarias)

   Datos publicados por organismos oficiales o privados en páginas web o en medios de comunicación -como pueden ser diarios, revistas o libros- que constituyen bases de datos para períodos regulares de medición. Generalmente, las bases de datos ya publicados se corresponden con unidades de observación temporales, aunque es posible encontrar también datos referidos a individuos (empresas, regiones, países, etc.).

2) **elaboración de una encuesta.**

   Este método es el de mayor aplicación en una investigación econométrica que utiliza fuentes de información primarias.

3) **Experimentación.**

   Proporciona datos experimentales obtenidos de unidades de observación que han sido controladas por ciertos factores variables, para determinar qué efecto ejercerán esos factores en los datos.

Por lo tanto, la obtención de las unidades de observación dependerá de si son de tiempo, de corte transversal o de panel. La obtención de las unidades de tiempo es bastante sencilla, ya que basta con elegir el período de tiempo y el espacio de observación donde será medida la variable en estudio para obtener los datos. Estos, necesariamente, estarán disponibles en una base de datos predeterminada; es decir, fuentes secundarias. A lo sumo se deberán realizar transformaciones algebraicas de las variables, según el tipo de datos que se quiera utilizar; por ejemplo, valores corrientes o valores constantes de determinada variable.

Ejemplo 1.3. Organismos que presentan información estadística en sus portales de internet. Este listado no es exhaustivo sino sólo a modo de ejemplo.

Instituto Nacional de Estadísticas y Censos (INDEC), www.indec.gov.ar

Ministerio de Economía, www.mecon.gov.ar

Ministerio del Interior, www.mininterior.gov.ar

Gobierno de Córdoba, www.cba.gov.ar

Presidencia de la Nación, www.presidencia.gov.ar

Centro de Estudios para América Latina (CEPAL), www.cepal.org

Consejo Latinoamericano de Ciencias Sociales (CLACSO), www.clacso.org

Fundación de Investigaciones Económicas Latinoamericanas (FIEL), www.fiel.org

---

Cuando se trabaja con unidades de observación de corte transversal, el trabajo previo es importante porque pueden estar disponibles en fuentes secundarias (como sucede cuando se trabaja con regiones o países). Pero, cuando el investigador econométrico debe seleccionarlas y luego entrevistarlas, el trabajo que requiere la preparación previa de la investigación es considerablemente mayor.

Instrumento de recolección de información

Entre las cuestiones que el investigador debe agregar, la principal es elaborar el instrumento de recolección de datos, denominado cuestionario. Este trabajo lo hará al momento de plantear la tabla de datos para conocer cómo medir las variables necesarias sobre las unidades que va a observar. La tarea adicional que le queda al investigador es cómo seleccionar dichas unidades y como recolectar los datos. Estos, que son los pasos segundo y tercero de observación de la realidad, se estudiarán en los cuadernos correspondientes a dichos temas.

En este paso de la investigación se trata de construir un instrumento que sirva para medir las variables que se han seleccionado. Los métodos de recolección más utilizados son la observación y la entrevista.

La *observación* se aplica, preferentemente, en aquellas situaciones en las que se trata de detectar aspectos conductuales, 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; 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.

Diseño del cuestionario

El cuestionario está compuesto de un espacio para registrar la encuesta como unidad, espacio para las preguntas y espacio para la respuesta a esas preguntas; su elaboración es un arte que mejora con la experiencia.

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. 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 abiertas del tipo:

"¿Por qué eligió asistir a esta facultad en vez de otras facultades?_____”

Requiere que el sujeto responda libremente en forma de composición; suelen ser poco adecuadas, pues la persona 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 se cree conveniente poseer una respuesta a la elección de una Facultad en lugar de otra, se debe pensar en establecer alternativas de respuesta que hagan referencia a diferentes aspectos, por ejemplo:

"nivel de enseñanza"

"prestigio"

"situación socioeconómica"

“tipo de carrera”

“cercanía al lugar de origen”

“otras, ¿cuál?_____

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

También se cuenta con las que utilizan grados de acuerdo y se centran en el estímulo más que en el entrevistado:

Completamente de Acuerdo

De Acuerdo

Indiferente

En Desacuerdo

Completamente en Desacuerdo

Algunas escalas han sido elaboradas con criterio psicosocial por el autor Lickert, a la que deben su nombre. Estas escalas no son únicas, según la clase de respuestas que necesita el investigador para llevar a cabo sus mediciones se diseña la adecuada.

Las preguntas semicerradas tienen una pregunta abierta en la última categoría del tipo

“Otra, ¿cuál? _____”

De esta manera, el investigador no está obligado a listar todas las posibles alternativas de respuesta a la pregunta sino que

explicita las potencialmente más frecuentes y permite que una categoría, no advertida con anterioridad, adquiera relevancia. En la pregunta sobre las razones por las cuales se elige la Facultad, sólo se han mencionado algunas de todas las alternativas posibles; incorporar esta alternativa en último lugar permite detectar motivaciones desconocidas entre los potenciales estudiantes.

Ejemplo 1.4. El caso de estudio 6, muestra el cuestionario utilizado en una localidad para el estudio del perfil financiero de la población.

Hay tres formas de realizar entrevistas:

1) las entrevistas personales a través de encuestadores, son las más eficientes. Entre otras razones, 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.
2) Las encuestas por correo consisten en el envío del cuestionario vía postal o electrónica para ser respondido en un plazo determinado. Este tipo de encuestas requieren de la buena voluntad de la persona que recibe la correspondencia para dar respuesta. En general, no se tiene población de referencia y presentan un alto número de no respuesta.
3) Las encuestas telefónicas se han popularizado en los últimos años; en algunas oportunidades son encuestadores que relevan la información a través de un cuestionario y en otras es un sistema informático diseñado especialmente donde una persona virtual realiza la entrevista y el encuestado responde con números del teclado telefónico.

1.4 Las Variables

Una variable es una colección de observaciones sobre individuos o unidades de observación pertenecientes a una población o

muestra. Las variables son elementos fundamentales en una investigación econométrica y forman parte de la tabla de datos. En principio, las variables son características que están presentes en los individuos que forman la población o muestra.

Desde un punto de vista, un fenómeno que se puede medir o contar es una *variable*; es una característica o atributo, que puede ser medida, adoptando diferentes valores para cada unidad de observación.

Debido a que cada variable puede asumir valores distintos, debe ser representada por un símbolo en lugar de un número específico. En general, esta representación viene dada por letras como X, Y, Z.

Desde otro punto de vista, una variable es una dimensión que representa la coordenada de un individuo en el espacio. Por ejemplo, un individuo se puede caracterizar por tener determinada edad, determinado ingreso y determinado nivel de educación; cada una de estas son variables que caracterizan a un individuo particular, este individuo estará representado en el espacio como un punto sostenido por esas tres coordenadas donde cada una de ellas representa una dimensión.

Si se representa a *todos* los individuos sobre una recta en la dimensión ingresos, se tiene lo que ilustra la Figura 1.7: los ingresos ordenados de menor a mayor y todos ubicados en la misma recta. Esta recta puede asumir valores desde 0 a infinito, obviamente cada punto sobre la recta representará un individuo de la población o muestra y el intervalo que agrupa los valores, desde el mínimo al máximo, se denomina *recorrido* de la variable, siendo su *rango* la diferencia entre el valor máximo y el valor mínimo.

Figura 1.7. Representación de unidades de observación en la recta

Se debe distinguir entre *constantes, variables y parámetros*. Por ejemplo, el ingreso personal de un conjunto de individuos es una variable. En cambio, el coeficiente angular de una *recta dada* en un sistema de coordenadas es una constante. La constante es una magnitud que no cambia, como tal es lo contrario a la variable. Cuando la constante se une a una variable se suele denominar coeficiente, este puede ser simbólico en lugar de numérico; por ejemplo, si el coeficiente angular es general, no perteneciente a una *recta dada*, puede asumir distintos valores de acuerdo a la recta de que se trate. Esta forma particular de constantes, que al no tener un valor específico puede tomar prácticamente cualquiera, es denominada constante paramétrica o simplemente parámetro.

Las constantes son entonces números específicos y así serán representados, 2, 1, 29, etc., en cambio los parámetros serán representados por letras minúsculas (a, b, c) o sus contrapartes en el alfabeto griego (α, β, γ). Existe una clase especial de parámetros, los que se obtienen de calcular medidas de posición o de dispersión sobre los valores de la variable en la población; por ejemplo, la media aritmética o esperanza matemática, representada por μ o el desvío estándar representado por σ.

Clasificación de las variables

En una investigación econométrica, las variables se pueden clasificar *según la escala de medición* en*:*

- Variables cualitativas: son aquellas que toman un número limitado (o no) de modalidades y a cada modalidad corresponde una categoría de unidad de observación, tiempo o individuos. Estas categorías forman una partición de la población. En general, son variables que expresan distintas cualidades o características que no pueden cuantificarse o a las que no puede asignarse un valor. Cada modalidad que se presenta se denomina atributo o categoría y la medición consiste en una clasificación de dichos atributos. Las variables cualitativas son dicotómicas cuando sólo pueden tomar dos valores

posibles (sí y no, hombre y mujer) o politómicas cuando pueden adquirir tres o más valores. Dentro de ellas se distinguen:

- Variable cualitativa ordinal: La variable puede tomar distintos valores ordenados siguiendo una escala establecida, aunque no es necesario que el intervalo entre mediciones sea uniforme, por ejemplo, leve, moderado, grave.
- Variable cualitativa nominal: En esta variable los valores no pueden ser sometidos a un criterio de orden como por ejemplo los colores o el lugar de residencia.

Generalmente, las variables cualitativas se codifican con números, pero esto no las habilita a ser consideradas como variables cuantitativas; los cálculos de media o de varianza sobre estos valores no tienen sentido porque carecen de propiedades numéricas.

A veces el interés se centra en una categoría particular, $\boldsymbol{A}$, de una variable cualitativa que se denomina *indicadora* de $\boldsymbol{A}$; esta es una variable que toma el valor 1 para cualquier individuo que pertenece a $\boldsymbol{A}$ y 0 si no pertenece. La proporción de $\boldsymbol{A}$ con respecto a la población se simboliza π, y la proporción de $\boldsymbol{A}$ con respecto a la muestra $\hat{\pi}$ o p.

Variables cuantitativas: son aquellas que toman valores reales para los cuales se pueden calcular resúmenes numéricos como la media, varianza y desviación estándar. Dada una variable cuantitativa particular, $\boldsymbol{Y}$, se tienen la media y la varianza de $\boldsymbol{Y}$ en la población, $\boldsymbol{\mu}$ y $\boldsymbol{\sigma^2}$; la media y la varianza de $\boldsymbol{Y}$ en la muestra, $\bar{Y}$ y S^2.

De esta forma las variables cuantitativas asumen distintos valores, que la identifican como tal y definen el recorrido de la variable asociado a las unidades observadas. Este recorrido, constituido por datos, especifica diversas medidas que le darán ciertas propiedades a las mismas. Estas propiedades la distinguen de otra u otras variables cuantitativas observadas sobre las mismas unidades. Así la variable para el conjunto de las unidades de observación tendrá un *valor máximo* y un *valor*

mínimo en su recorrido. Aunque las características más importantes se estudian a través de sus medidas de posición y dispersión.

La medida de posición describe o resume el recorrido de los datos de la variable. Toda variable cuantitativa 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.

Estas variables, que se expresan mediante cantidades numéricas, pueden clasificarse en:

Variable discreta: Es la variable que presenta separaciones o interrupciones en la escala de valores que puede tomar, estas separaciones o interrupciones indican la ausencia de valores entre los distintos valores específicos que la variable pueda asumir. La operación que da origen al valor que asume la variable es un proceso de conteo, pueden tomar valores enteros o fracciones pero sin continuidad. Un ejemplo es el número de hijos.

Variable continua: Es la variable que puede adquirir cualquier valor dentro de un intervalo especificado de valores. La operación que da origen al valor que asume la variable, para las distintas unidades de observación, se origina en un proceso de medición; entre estas se encuentran las medidas de capacidad, distancia y tiempo. Asumen valores dentro del campo de los números reales. Por ejemplo, el peso o la altura, que solamente está limitado por la precisión del aparato de medición, en teoría permiten que siempre exista un valor entre dos cualesquiera.

Entonces, la diferencia entre los dos tipos de variables estudiados es que las cualitativas arrojan respuestas categóricas; mientras que, las variables cuantitativas dan respuestas numéricas (en el campo de los números reales) que pueden, a su vez, clasificarse en continuas o discretas. El detalle se presenta en la Figura 1.8.

Tipos de variables	Tipos de preguntas	Respuestas
Cualitativas	¿Vive Ud. en las residencias estudiantiles?	Si No
Cuantitativas: Continuas Discretas	¿Cuál es su estatura? ¿A cuántas revistas está suscripto?	_____(cts.) _____(cantidad)

Figura 1.8. Tipos de variables usadas en econometría

Ejemplo 1.5. Tipo de variables

Variables Cualitativas

Lugar de procedencia de los estudiantes de primer año

Máximo nivel de estudio alcanzado por el entrevistado

Obra Social o Mutual prepaga a la que pertenece

Idioma oficial del país

Moneda de circulación

Alimento de preferencia

Variables Cuantitativas Continuas

Producto Bruto Geográfico de Córdoba

Reservas en el Banco Central

Producción de maní en granos

Volumen de ventas en el sector automotriz

Variables Cuantitativas Discretas

Unidades vendidas de máquinas cosechadoras

Número de camiones ingresados a puerto

Existencia de ganado lanar

Cantidad total de habitantes

Variables en los modelos económicos

Las variables se relacionan entre sí para definir ecuaciones o modelos matemáticos que pueden ser considerados modelos económicos cuando son representaciones de la teoría económica.

Toda ecuación es una relación matemática, entre un conjunto de *variables*, que se verifica para determinados valores numéricos, los que tienen significado económico. Según la influencia que se asigne a unas variables sobre otras, se tiene:

Variables independientes o exógenas: Son las que el investigador selecciona para establecer agrupaciones en la investigación, clasificando intrínsecamente a los casos del estudio.

Variables dependientes o endógenas: Son las variables de respuesta que se observan en la investigación y que podrían estar influidas por los valores de las variables independientes.

Las *variables expectativas*, son variables no observables y su introducción exige el enunciado de un postulado adicional en el que se especifica su comportamiento en función de variables observables.

La Figura 1.9 ilustra la clasificación de las variables.

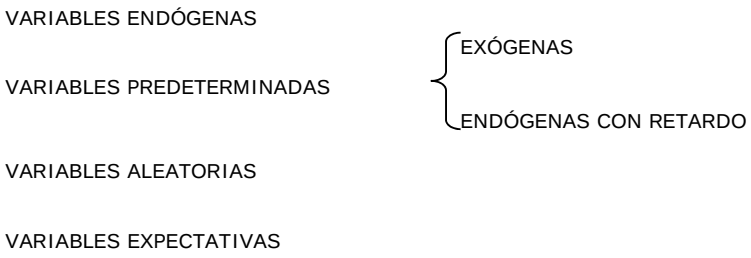


Figura 1.9. Tipos de variables en modelos económicos

Ejemplo 1.6: Sea C consumo e Y ingreso, la primera ecuación expresa que el incremento del consumo entre el periodo $t-1$ y el periodo t es una proporción d de la diferencia entre el consumo observado en el periodo $t-1$ y el consumo normal esperado $C^*{}_t$ en el periodo t, $C^*{}_t$ es variable expectativa. A su vez, la segunda ecuación expresa al consumo normal esperado en el periodo t como función lineal del ingreso Y.

$$C_t - C_{t-1} = d(C_t^* - C_{t-1}); \quad 0 < d < 1 \qquad (1.4)$$

$$C_t^* = \alpha + \beta Y_t \qquad (1.5)$$

Resolviendo, el sistema resulta:

$$C_t = \alpha d(1-d)C_{t-1} + \beta d Y_t \qquad (1.6)$$

Esta última es una ecuación en función sólo de variables observables.

Las *variables predeterminadas*, son aquellas que no se obtienen por solución del modelo, sino que provienen desde *afuera* y contribuyen a explicar el comportamiento de las variables endógenas, con la condición de no ser explicadas por el modelo mismo. Se pueden clasificar en exógenas y endógenas con retardo- La diferencia entre ambas radica en que estas últimas se determinan en el mismo modelo pero en un momento anterior.

Ejemplo 1.7 Variables predeterminadas

En el modelo de consumo

$$C_t = \alpha_0 + \alpha_1 Y_t + \alpha_2 Y_{t-1} + \cdots + \mu_t \qquad (3.7)$$

El consumo nacional puede explicarse por los niveles de ingreso nacional de todos los periodos precedentes incluido el actual, donde Y_{t-1} es una variable económica predeterminada exógena y las α_i

definen las propensiones marginales *parciales* a consumir, y $\alpha = \sum \alpha_i;\ 0 < \alpha_i < 1 \ \wedge \ \alpha < 1$ define la propensión marginal *total* a consumir.

En el modelo de demanda

$$D_t = \alpha_0 + \alpha_1 P_t + \alpha_2 P_{t-1} + \mu_t \qquad (3.8)$$

El P_t es una variable endógena que debe ser explicada por otra ecuación del modelo y P_{t-1} es una variable económica predeterminada endógena con retardo.

Si las unidades de observación fueron seleccionadas a través de una muestra aleatoria, las variables que sobre ellas se observen se denominarán *variables aleatorias* y las unidades de observación constituirán lo que técnicamente se denomina una *muestra*. Este nombre se debe a que constituyen una parte seleccionada aleatoriamente de una *población*. La cantidad de unidades de observación de una muestra se simboliza con una n, mientras que los elementos de una población se simbolizan con una N.

Las *variables aleatorias*, son variables no observables y su introducción caracteriza a los modelos estocásticos o probabilísticos (econométricos) en oposición a los modelos deterministas o exactos (economía matemática). Estas variables se incluyen en los modelos econométricos por omisión de variables explicativas, por errores de especificación y/o errores de medida (u observación). Estas variables siguen una distribución de probabilidad que no necesita ser especificada en el proceso de estimación pero sí cuando se realiza el contraste del modelo.

La clasificación de las variables en los *modelos de decisión* es necesaria cuando, a partir de la especificación del modelo, se desea lograr un objetivo como puede ser, por ejemplo, querer alcanzar un crecimiento del producto bruto nacional.

Cuando se fija el objetivo a alcanzar las variables son *endógenas objetivo*; cuando actúan sobre ella los sujetos de la actividad

económica con fines de control se denominan variables *exógenas controlables*, por ejemplo, los impuestos son variables controlables por el sector público.

Cuando son elegidas como medio o instrumento sobre el que se actuará para lograr los objetivos programados, las variables son *exógenas controlables instrumentales*.

La Figura 1.10 presenta en un esquema las variables que pueden encontrarse en un modelo de decisión

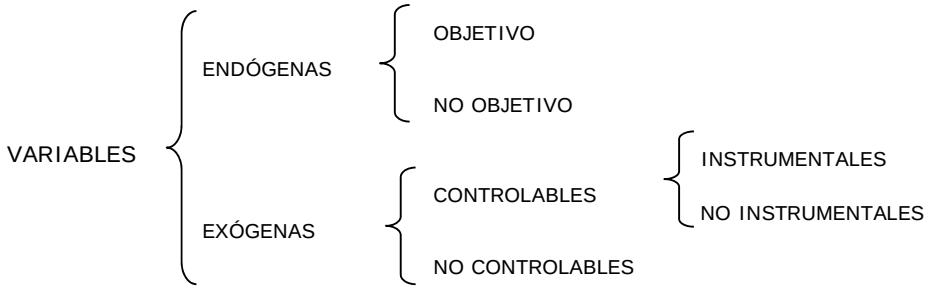


Figura 1.10. Tipos de variables en modelos de decisión

Ejemplo 1.6. Sea el siguiente modelo de decisión

$$\begin{cases} C_t = \alpha + \beta Y_t + \mu_t \\ Y_t = C_t + I_t + Z_t \end{cases} \qquad (3.9)$$

Que tiene como variables endógenas el consumo (C) y el ingreso (Y), y sus variables exógenas son la inversión privada (I) y la inversión pública (Z).

Si el modelo se utiliza con fines de decisión es necesario reclasificar las variables; bajo el supuesto de que el sujeto de la macro-decisión (gobierno, en este caso) fija como objetivo alcanzar una determinada tasa de crecimiento anual de ingreso nacional, entonces ahora es:

Y la variable endógena objetivo;

C la variable endógena no objetivo;

Z la variable exógena controlable e instrumental e

I la variable exógena no controlable por el sector público.

Ejemplo 1.7. Modelo macroeconómico del crecimiento

Dagum-Dagum (1971, pp. 35) presentan el modelo

(1) $C_t = \alpha_0 + \alpha_1 Y_t + \alpha_2 C_{t-1} + \mu_{1t}; \quad 0 < \alpha_1 < 1; \quad 0 < \alpha_2 < 1$

(2) $I_t = \beta_0 + \beta_1 B_t + \beta_2 K_{t-1} + \mu_{2t}; \quad \beta_1 > 0$

(3) $W_t = \phi_0 + \phi_1 Y_t + \phi_2 t + \mu_{3t}; \quad \phi_1 > 0$

(4) $Y_t = C_t + I_t + G_t$

(5) $B_t = Y_t - W_t$

(6) $K_t = K_{t-1} + I_t$

siendo:

C_t Consumo nacional

I_t Inversión Neta

Y_t Ingreso nacional

G_t Gasto Público en bienes y servicios

C_t Ingreso de los asalariados

B_t Ingreso de los no asalariados

K_t Existencias de capital al finalizar el periodo t

t Tiempo

Es un modelo multiecuacional, donde

(1), (2) y (3) son ecuaciones de comportamiento porque reflejan el modo de actuar de los consumidores, los inversores y los salarios, respectivamente.

(4), (5) y (6) son ecuaciones de definición o identidad

C_t, I_t, B_t, Y_t, W_t y K_t son variables endógenas

C_{t-1} y K_{t-1} son variables predeterminadas endógenas con retardo

G_t variable predeterminada exógena con sentido económico

t variable predeterminada exógena sin sentido económico

μ_1, μ_2 y μ_3 variables aleatorias

α_0, α_1 y α_2 son parámetros estructurales que representan el consumo autónomo, la propensión marginal a consumir debido al ingreso del periodo t y el efecto sobre el consumo debido al nivel de consumo alcanzado en el periodo anterior, respectivamente.

β_0, β_1 y β_2 son parámetros estructurales que miden la inversión autónoma, la velocidad de cambio en la inversión ante cambio en el ingreso de los no asalariados y la velocidad de cambio en la inversión ante cambio en el ingreso de los no asalariados.

φ_0, φ_1 y $\varphi_2$son parámetros estructurales que identifican el salario autónomo, la velocidad de cambio de los salarios totales ante cambios en el ingreso nacional y el impacto de la componente tendencial en la determinación de los salarios.

Además, la ecuación (1) puede deducirse de una ECUACION CON DISTRIBUCIÓN DE RETARDO siguiendo la ley geométrica. Procediendo a restituir C_{t-1} en forma recurrente se tiene:

$$C_t = \frac{\alpha_0}{1-\alpha_2} + \alpha_1 Y_t + \alpha_1\alpha_2 Y_{t-1} + \alpha_1\alpha_2^2 Y_{t-2} + \Lambda + v_t$$

donde:

$$C_{t-1} = \frac{\alpha_0}{1-\alpha_2} + \alpha_1 Y_{t-1} + \alpha_1\alpha_2 Y_{t-2} + \alpha_1\alpha_2^2 Y_{t-3}$$

La PROPENSION MARGINAL A CONSUMIR debida a Y_t se deduce derivando parcialmente C_t respecto a Y_t en la ecuación con distribución de retardo.

La PROPENSION MARGINAL TOTAL A CONSUMIR o derivada, a largo plazo, del consumo respecto al ingreso, se deduce obteniendo las derivadas parciales con respecto a Y_{t-1}, Y_{t-2}, Λ y sumándolas, es decir

$$\sum_{i=0}^{\infty} \frac{dC_t}{dY_{t-i}} = \frac{\alpha_1}{1-\alpha_2}$$

1.5 Parámetros y Estadísticos

Uno de los principales recursos en los que se basa el proceso de investigación econométrica es la *inferencia estadística*. Esta consiste en utilizar *estadísticos* -media aritmética, desviación estándar y proporción- que se obtienen sobre las variables con los datos de la muestra, para estimar su verdadero valor en la población.

Una *población* puede definirse como el conjunto de todos los elementos posibles en el espacio definido. Relacionado con este concepto, aparece el de *muestra*, que es un conjunto de unidades de observación seleccionadas de los elementos que constituyen la población.

Interesan las muestras probabilísticas o sea, aquellas obtenidas por medio de un mecanismo casual determinado. Una *muestra aleatoria* es una muestra probabilística en donde cada unidad de observación de la población tiene la misma probabilidad de ser elegida o seleccionada.

Cuando no es posible obtener la información buscada a través de la observación directa se utiliza la *encuesta*. Este es el proceso por el cual se recogen datos sobre las variables consideradas en el estudio, solicitando la información a otras personas.

Los datos, de los cuales se obtiene información por medio de la observación, pueden estar contenidos en cualquier *población* o *muestra* extraída de la misma.

Ejemplo 1.8. El conjunto de Estados de Resultados mensuales de los ejercicios económicos de la Empresa “A” obtenidos entre enero de 1980 –fecha de puesta en marcha- y el último ejercicio mensual cerrado por la Empresa “A” constituyen una población. Cualquier *subconjunto* de Estados de Resultados mensuales de la Empresa “A” seleccionados al azar entre enero de 1980 y la actualidad constituye una muestra.

El investigador, a través de los datos recolectados 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 de unidades de observación de tamaño determinado se selecciona aleatoriamente entre la población. Las unidades de observación se van a seleccionar mediante, por ejemplo, una tabla de números aleatorios.

En este sentido, la investigación tiene por objetivo *estimar* valores específicos de la población, denominados *parámetros*. Un parámetro 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 de un parámetro de la población, es la *media* de una variable. Algunos valores de la población muy relacionados con la media son la *proporción*, *mediana*, –entre otros *cuantiles*-, y el *total* o valor agregado de la población. Algunos valores de la población miden relaciones; los más comunes son la *diferencia entre dos medias*, el *coeficiente de correlación* y el *coeficiente de determinación*.

El valor de la muestra, o *estadístico*, es una estimación, del parámetro correspondiente a una variable, que se calcula a partir de las n unidades de observación en la muestra. En sí mismo 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 parámetro depende de los N valores de la variable en dicha población. Es una constante independiente de las fluctuaciones de la selección, aunque por lo general se desconoce.

Por lo tanto, en la investigación de cualquier fenómeno es necesario observar y recoger algunas características (variables

cuantitativas o cualitativas) de las unidades de observación. Para cada unidad de observación se deben obtener datos sobre las variables estudiadas. Para obtenerlos se realizan observaciones, a partir de una muestra y se pueden calcular diversos *estadísticos*, que son estimadores de los parámetros poblacionales y que se clasifican en medidas de posición y medidas de dispersión.

Entre las *medidas de posición* se encuentran:

a) La media aritmética es el promedio que surge de sumar todos los valores (datos) que asumen las unidades de observación para una variable y dividirlos por el total de unidades observadas. 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.

$$\bar{X} = \frac{1}{n}\sum_{i=1}^{n} x_i = \frac{x_1 + x_2 + \cdots + x_n}{n} \qquad (1.10)$$

donde:

$\bar{X}$, media aritmética de la variable X en la muestra

n, tamaño de la muestra

x_i, dato para la i-ésima unidad de observación de la variable X

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

Las propiedades de la media son:

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

Suma de cuadrados totales: $\sum_{i=1}^{n}(x_i - \bar{X})^2 = SCT$

Valor total: $N\bar{X}$, siendo: N, tamaño de la población;$\bar{X}$, media aritmética de la muestra

b) La *mediana* es el valor que aparece en el medio de una secuencia ordenada de datos correspondientes a una variable y no se ve afectada por ninguna observación extrema. Para calcular su valor, primero hay que 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 a una unidad de observación en el punto de posicionamiento; mientras que, si el tamaño de la muestra es par, el punto de posicionamiento cae entre dos unidades de observación, el dato promedio de estas dos observaciones es la mediana.

c) 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.

d) La *media ponderada* de un conjunto de números es el resultado de multiplicar cada uno de los números por un valor particular para cada uno de ellos, llamado su *peso*, obteniendo a continuación la suma de estos productos, y dividiendo el resultado por la suma de los pesos. Este *peso* depende de la importancia o significancia de cada uno de los valores.

Para una variable X con datos $x_1, x_2, \dots, x_n$

a la que corresponden los pesos $w_1, w_2, \dots, w_n$

la media ponderada se calcula como:

$$\bar{X} = \frac{\sum_{i=1}^{n} x_i w_i}{\sum_{i=1}^{n} w_i}$$

Un ejemplo es la obtención de la media ponderada de las notas de una prueba de oposición en la que se asigna distinta

importancia (*peso*) a cada uno de los elementos de que consta el examen.

e) La *media geométrica*, de n valores $x_1,\ x_2,\cdots,x_n$ de una variable X, es la raíz n-ésima del producto de todos los valores.

$$G = \sqrt[n]{\prod_{i=1}^{n} x_i} = \sqrt[n]{x_1.x_2.\cdots.x_n}$$

La media geométrica es relevante si todos los valores son positivos. Si uno de ellos es 0, entonces el resultado es 0. Si hubiera un número negativo (o una cantidad impar de ellos) entonces la media geométrica sería o bien negativa, o bien inexistente en los números reales.

Para el cálculo de la *media geométrica ponderada* deben introducirse ponderaciones $w_i, \forall i = 1,\cdots,n$ como exponentes:

$$G = \left(x_1^{w_1}.x_2^{w_2}\cdots x_n^{w_n}\right)^{\frac{1}{w_1+w_2+\cdots+w_n}}$$

f) La *media armónica* de una cantidad finita de valores observados de una variable X; su valor es igual a la inversa de la media aritmética de los recíprocos de dichos valores. Así, dados n valores $x_1,\ x_2,\cdots,x_n$ de una variable X, observados con la misma frecuencia, la media armónica será igual a:

$$Media\ armónica = \frac{n}{\sum_{i=1}^{n}\frac{1}{x_i}}$$

La media armónica resulta poco influida por la existencia de determinados valores mucho más grandes que el conjunto de los otros, siendo en cambio sensible a valores mucho más pequeños que el conjunto. No está definida en el caso de la existencia en el conjunto de valores nulos.

g) El *rango medio* es el promedio de las observaciones menores y mayores de una serie de datos. A menudo es usado como medida 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 valores menores y mayores de una sucesión, el rango medio se distorsiona como una medida de resumen de tendencia central si está presente una observación extrema.

h) 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 expresiones:

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

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

a) El *recorrido*, rango o amplitud es la diferencia entre el valor más grande y el más pequeño de los datos que asume la variable X, en una tabla de datos:

$$R = x_n - x_1$$

b) 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 dato correspondiente al recorrido de la variable X y su media

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

Los grados de libertad hacen referencia al número de cuadrados independientes.

OBSERVACIÓN: El número total de cuadrados en la expresión $\sum_{i=1}^{n}(x_i - \bar{X})^2$ es n, pero sólo hay $n-1$ cuadrados independientes; porque, una vez calculados los $n-1$ primeros cuadrados, el valor del último queda determinado automáticamente.

La razón de ello es la presencia de $\bar{X}$; una de las propiedades de $\bar{X}$ es que $\sum_{i=1}^{n}(x_i - \bar{X}) = 0$. Esto representa una restricción que debe cumplirse. Por ejemplo, si se tienen tres cuadrados y 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 asumir cualquier valor, por ejemplo $(-5)^2$; pero esto no puede suceder, porque en este caso se tiene

$$\sum_{i=1}^{3}(x_i - \bar{X}) = 2 + 4 - 5$$

cuya suma es (-1) y no cero, como tiene que ser. En efecto si los dos primeros cuadrados son 2^2 y 4^2, el tercero tiene que ser $(-6)^2$, porque sólo en este caso se cumple con la propiedad de la media aritmética $2 + 4 - 6 = 0$.

c) El *desvío estándar* se calcula como 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:

$$S_{total} = NS$$

d) 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 1.7. 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 B ha sido casi dos veces más variable (tiene mayor volatilidad) que la acción A con respecto al precio promedio para cada una de las dos.

	Acción A	Acción B
Desvío	500	300
Media	1500	5000
Coeficiente de variación	3.33%	6%

Figura 1.11. Cálculo del coeficiente de variación

2. VARIABLES ALEATORIAS

En la investigación econométrica es de interés una clase especial de variables que se denominan variables aleatorias. En principio, se dará un tratamiento general al término variable aleatoria, más adelante, se estudiarán variables aleatorias específicas de los modelos econométricos. Asimismo, se presentan las distribuciones de probabilidad de las variables aleatorias, necesarias para poder realizar las estimaciones y contrastes estadísticos de los parámetros poblacionales.

2.1. Modelos Probabilísticos

Anteriormente quedó establecido el insumo básico del proceso de investigación econométrica: el planteo de una tabla de datos. Es un proceso empírico y por lo tanto las variables que se observan para el conjunto de individuos seleccionados son aleatorias y susceptibles de ser estudiadas sus distribuciones de probabilidad en forma empírica. Estas variables del proceso de investigación econométrica serán denominadas en forma genérica como X_j, de tal manera que su estudio empírico determine que pueda ser aplicado de igual manera para las k variables que conforman la tabla de datos definida. Entonces cuando se refiera a la variable genérica X_j se estará refiriendo de igual manera a cualquiera de las $j = 1, \dots, k$ variables que se definen dicha tabla.

Suponiendo que un experimento consiste en observar la variable ingreso X_j, donde X_j puede ser cualquiera de las k variables dispuestas en una tabla de datos –es decir, $(\forall j = 1, \dots, k)$- y otras $k-1$ variables sobre un conjunto de n individuos -tal que,

$1 \leq i \leq n$- seleccionados al azar de una población de tamaño N. El investigador está interesado en conocer la distribución de los ingresos anuales de la población y estudiar si las otras $k-1$ variables determinan las variaciones de dichos ingresos aunque esto último, por ahora, no será tema de estudio. Para ello plantea una tabla de datos como se muestra en la Figura 2.1.

	X_1	⋯	X_j	⋯	X_k
1	x_{11}		x_{j1}		x_{k1}
2	x_{12}		x_{j2}		x_{k2}
3	x_{13}		x_{j3}		x_{k3}
⋮	⋮		⋮		⋮
i	x_{1i}		x_{ji}		x_{ki}
$i+r$	x_{1i+r}		x_{ki+r}		x_{ki+r}
⋮	⋮		⋮		⋮
n	x_{1n}		x_{jn}		x_{kn}
Sumas	$\sum_{i=1}^{n} x_{1i}$		$\sum_{i=1}^{n} x_{ji}$		$\sum_{i=1}^{n} x_{ki}$
Media aritmética	$\bar{X}_1 = \frac{\sum_{i=1}^{n} x_{1i}}{n}$		$\bar{X}_j = \frac{\sum_{i=1}^{n} x_{ji}}{n}$		$\bar{X}_k = \frac{\sum_{i=1}^{n} x_{ki}}{n}$

Figura 2.1. Tabla de datos

Para conocer la probabilidad de que un individuo posea ingresos menores a determinada cantidad, el investigador debe transformar la tabla de datos, en una tabla de frecuencias, de la siguiente manera:

1) Seleccionar la variable para la que quiera realizar la tabla de frecuencias, suponga que es la variable ingresos (X_j);
2) Ordenar el recorrido de la variable aleatoria X_j de menor a mayor;
3) Agrupar los individuos en l intervalos de clase.

Por lo tanto, la tabla de frecuencias quedará dispuesta de la siguiente manera:

Intervalo de Clase	Punto Medio X	Cantidad de individuos f_a	Frecuencia Relativa f_r	$X \cdot f_r$
1	x_1	n_1	n_1/n	$x_1 \cdot n_1/n$
2	x_2	n_2	n_2/n	$x_2 \cdot n_2/n$
3	x_3	n_3	n_3/n	$x_3 \cdot n_3/n$
⋮	⋮	⋮	⋮	⋮
i	x_i	n_i	n_i/n	$x_i \cdot n_i/n$
⋮	⋮	⋮	⋮	⋮
l	x_{l0}	n_l	n_l/n	$x_l \cdot n_l/n$
SUMAS		n	1	$\bar{X}$

Figura 2.2 Distribución de frecuencias de la Variable aleatoria Ingresos

Entonces

$$\bar{X}=\sum_{i=1}^{l} x_i \cdot n_i/n \tag{2.1}$$

representa el promedio de los ingresos de los individuos que conforman el experimento. Dónde x_i es el valor probable del i-ésimo intervalo de clase que agrupa los valores probables que representan a las unidades de observación ubicadas en ese intervalo.

Ahora bien, el *valor probable* para una unidad de observación correspondiente a una variable responde a un *modelo*

matemático probabilístico -no determinístico o estocástico-; dicho valor es una variable aleatoria. Esto es un principio básico de la Econometría Aplicada; si fuera no estocástico, los instrumentos de la econometría no se aplicarían ya que se especificaría un modelo determinístico y no tendría sentido la corroboración de teorías económicas en espacio y tiempo determinado. Se verá que una parte del modelo econométrico –la representada por variables exógenas– se considerará no estocástica; pero esto hace a la naturaleza misma del modelo probabilístico que se desarrolla en este acápite, ya que una vez ejecutado el experimento el *valor probable* se transforma en un *valor exacto* para una unidad de observación y será un valor que no cambia si se mantienen las condiciones en que se llevó a cabo el experimento. Como se verá, esto se desprende del concepto de variable aleatoria.

En este cuaderno se tratará un tipo especial de fenómeno. Para investigarlo se formulará un modelo matemático. Al principio es importante distinguir entre el fenómeno observable en sí mismo y el modelo matemático para dicho fenómeno. Evidentemente, no se puede influir sobre lo que se observa; sin embargo, al elegir un modelo, sí se puede aplicar un juicio crítico. Neyman (1954) escribió: “Cada vez que utilizamos las matemáticas con el objeto de estudiar fenómenos observables es indispensable empezar por construir un modelo matemático (determinístico o probabilístico) para esos fenómenos. Necesariamente, este modelo debe simplificar las cosas y permitir la omisión de ciertos detalles. El éxito del modelo depende de si los detalles que se omitieron tienen o no importancia en el desarrollo de los fenómenos estudiados. La solución del problema matemático puede ser correcta y aun así estar muy en desacuerdo con los datos observados, debido sencillamente a que no estaba probada la validez de las suposiciones básicas que se hicieron. Corrientemente, es bastante difícil afirmar con certeza si un modelo matemático es adecuado o no, antes de obtener algunos datos mediante la observación. Para verificar la validez del modelo, debemos deducir un cierto número de consecuencias del mismo y luego comparar con las observaciones esos resultados predichos”.

En la naturaleza, hay ejemplos de experimentos para los cuales los modelos determinísticos son apropiados. Así, las leyes gravitacionales describen exactamente lo que sucede a un cuerpo que cae bajo ciertas condiciones. Pero también, las identidades contables, en un modelo económico, definen una relación que determina unívocamente la cantidad del primer miembro de la ecuación si se dan las del segundo miembro.

En muchos casos el modelo matemático determinístico descrito anteriormente es suficiente. Sin embargo, hay fenómenos que necesitan un modelo matemático distinto para su investigación. Esos son los llamados *modelos no determinísticos o probabilísticos*. En este tipo de modelos, las condiciones experimentales determinan el comportamiento probabilístico de los resultados observables (más específicamente, la distribución de probabilidad); mientras que, en los modelos determinísticos se supone que el resultado real (sea numérico o no) está determinado por las condiciones bajo las cuales se efectúa el experimento o procedimiento.

Retomando el ejemplo del ingreso, la distribución de probabilidad será una función $P = f(X)$. Donde P depende de los valores que asume X, teniendo en cuenta que dichos valores ocurren con determinada frecuencia. En este sentido, P es función del recorrido de X, (R_x); es decir, dado R_x, P asume valores entre 0 y 1, representados en este caso como frecuencias relativas. Así como hay funciones lineales o cuadráticas hay funciones de probabilidad y dentro de ellas, por ejemplo, la binomial, la poisson y la normal. La Figura 2.4 muestra una función para la distribución de probabilidad de los puntos de clase de ingresos de los hogares de Río Cuarto para el mes de Mayo de 2003.

Ejemplo 2.1. En la Figura 2.3 se clasifica a los hogares de Río Cuarto por intervalos de clase según el nivel de ingresos declarados a la Encuesta Permanente de Hogares en Mayo de 2003.

Nivel de ingreso mensual del hogar	Punto Medio (X)	f_a	f_r	$f_r \cdot X$
Hasta 150	75	84	0,1303	9,7725
Entre 151 y 300	225	84	0,1321	29,7225
Entre 301 y 450	375	85	0,1341	50,2875
Entre 451 y 600	525	85	0,1341	70,4025
Entre 601 y 750	675	74	0,1164	78,57
Entre 751 y 1000	875	85	0,1321	115,5875
Entre 1001 y 1250	1125	47	0,0749	84,2625
Entre 1251 y 1500	1375	26	0,0414	56,925
Entre 1501 y 2000	1750	31	0,0493	86,275
Entre 2001 y 3000	2500	20	0,0316	79
Más de 3000	4500	15	0,0237	106,65
SUMAS		636	1	767,455

FUENTE: INDEC (2003). Encuesta Permanente de Hogares

Figura 2.3. Distribución Ingresos de los hogares de Río Cuarto

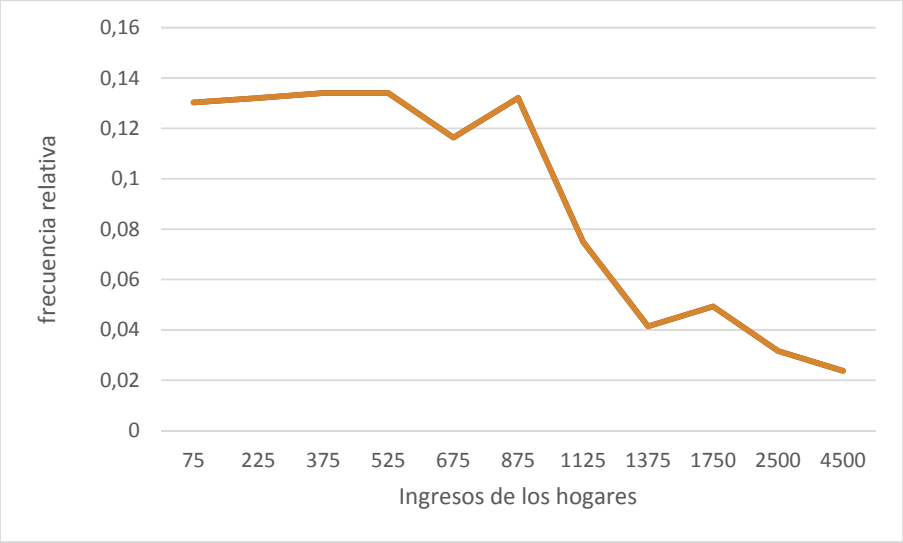


Figura 2.4 Distribución de frecuencia de los ingresos de los hogares

Espacio muestral

Meyer (1974) establece que los *experimentos aleatorios* tienen en común una serie de aspectos que los hacen sencillos de identificar:

a) Es posible repetir cada experimento indefinidamente sin cambiar esencialmente las condiciones
b) Se puede describir el conjunto de todos los resultados posibles del experimento, aunque en general no se pueda indicar cuál será un resultado particular
c) A medida que el experimento se repite, los resultados individuales parecen ocurrir en forma “caprichosa”. Sin embargo, como el modelo se repite un gran número de veces, aparece un modelo definido de regularidad. Esta regularidad hace posible la construcción de un modelo matemático preciso con el cual analizar el experimento.

En general, cuando se plantea la tabla de datos como consecuencia del proceso de investigación econométrica (procedimiento experimental), el investigador sabe que con la ocurrencia de un solo fenómeno se pueden calcular “varios” valores numéricos diferentes. Por ejemplo, si se escoge una persona entre un gran grupo de personas (y la elección se hace siguiendo el procedimiento experimental indicado previamente), se puede estar interesado en sexo, peso, ingreso anual, cantidad de hijos, etc. de la persona. En la mayoría de los casos el investigador que siga el procedimiento de investigación econométrica sabe, antes de comenzar el experimento, las características numéricas que le interesan.

El *espacio muestral* se define como el conjunto de todos los resultados posibles del experimento. Este conjunto se lo denomina S.

OBSERVACIÓN: En la tabla de datos estarán los resultados posibles para los elementos de una población. Estos elementos pueden ser de corte transversal o de tiempo. Cuando se utilizan elementos de tiempo, los resultados posibles provienen de experimentos repetidos en el tiempo en un espacio determinado.

El resultado de un experimento no necesita ser un número. Por ejemplo, en el experimento de seleccionar una persona entre un gran número de personas hay un resultado posible que no es numérico: el sexo. Aún más, el resultado puede ser un vector o

una función. Además, el número de resultados de un espacio muestral puede ser finito o infinito (numerable o no numerable).

Cuando el espacio muestral S es finito o infinito numerable, *todo* subconjunto se puede considerar como un *evento* E (si *S* tiene n elementos, hay exactamente 2^n subconjuntos o eventos). En la tabla de datos los valores probables para las unidades de observación (tiempo o individuos) constituirán los eventos.

> OBSERVACIÓN: Se advierte aquí que, aun cuando esto no tenga incidencia en los estudios econométricos, cuando el espacio muestral es infinito no numerable todo subconjunto de S no es necesariamente un evento. Asimismo, tanto el espacio muestral, como el conjunto vacío pueden ser considerados eventos. También, cualquier resultado individual se puede considerar un evento.

De acuerdo a los fundamentos anteriores, para el estudio econométrico, cada celda de la tabla de datos x_{ji}, estará representando un evento. Esto es, el valor probable que asuma para una determinada unidad de información i, $(\forall i = 1, 2, \dots, n)$, cualquier variable X_j, $(\forall\, j = 1, 2, \dots, k)$, bajo estudio; por ejemplo, si se observa la variable ingreso para un conjunto de individuos, puede interesar el evento de que ocurra para uno de ellos la pertenencia al primer intervalo de clase.

Cada evento del mundo real se halla relacionado con un conjunto infinito de otros hechos. Toda ciencia estudia solamente cierto número finito de vínculos. De tal modo se establecen las regularidades fundamentales de los eventos estudiados que reflejan las conexiones internas principales inherentes a estos últimos. En principio, es imposible llegar a conocer la infinita diversidad de relaciones existentes con cualquier evento dado.

En cada etapa del conocimiento humano, siempre quedan sin estudiar una multitud infinita de vínculos propios a un cierto evento. En consecuencia, cada regularidad puede reflejar solamente un número finito de relaciones fundamentales, debido a lo cual las leyes se cumplen sin precisión, con ciertas desviaciones. Las desviaciones de lo regular, originadas por una

infinidad de vínculos no previstos en el evento dado, se llaman *eventos aleatorios*. De este modo, la causalidad existe objetivamente en el mundo, porque en principio no es posible revelar todos los nexos accidentales entre el evento estudiado y la multiplicidad infinita de otros eventos.

"A medida que va desarrollándose la ciencia, dice Pugachev (1973), llegan a ser conocidas nuevas leyes, o sea, las conexiones que tiene el fenómeno estudiado con distintos factores. Por eso, las fronteras entre lo regular o casual no permanecen inalterables sino que cambian a medida que crece el conocimiento humano. Lo que en un período del desarrollo de la ciencia es accidental puede hacerse regular en otro período. Y al contrario, en los fenómenos considerados como estrictamente regulares en una etapa del desarrollo de la ciencia, a consecuencia de los perfeccionamientos en la técnica del experimento y las exigencias más rigurosas en lo que se refiere a la precisión durante el examen de las dependencias, se descubren desviaciones accidentales de las leyes y surge la necesidad de tenerlas en cuenta" (pp. 11-12).

Si el evento dado se observa una sola vez, no se puede predecir cuál será la desviación accidental de lo regular. Así, por ejemplo, al efectuar una medición sobre el ingreso de cierto residente en una ciudad, para estudiar el ingreso agregado de dicha comunidad, es imposible prever cual será el error de la misma. Sin embargo, si el número de observaciones del evento dado se hace grande, en las mismas desviaciones accidentales, se descubren ciertas regularidades -que pueden ser estudiadas y utilizadas para determinar la influencia de las desviaciones mencionadas sobre el curso de los eventos por estudiar-. Entonces, aparece la posibilidad de investigar eventos aleatorios frecuentes; es decir, observar eventos aleatorios un número ilimitado de veces en las mismas condiciones. La *teoría de la Probabilidad* es, precisamente, la ciencia que estudia las regularidades en los eventos aleatorios frecuentes.

Cierto evento E puede ocurrir como resultado de un experimento; por ejemplo, al observar determinado individuo, al que se le ha preguntado sobre la variable ingreso ¿resultará un

ingreso dentro del primer intervalo de clase? Ese resultado es, o se lo supone dicotómico; esto significa que existe un evento alternativo al que se designa como $\bar{E}$. Un experimento tiene, generalmente, varios resultados posibles; se denomina $x_{ji}, \forall i = 1, 2, \ldots, n \; y \; \forall \; j = 1, 2, \ldots, k)$ a cualquier resultado posible en un experimento, para la variable aleatoria X_j.

Cada ciencia se basa sobre algunos hechos experimentales que sirven para formar los conceptos fundamentales de la misma. Estos hechos y conceptos se utilizan para obtener determinadas conclusiones prácticas. En la ciencia económica, las conclusiones se comprueban por la experiencia. La coincidencia, de las conclusiones científicas con los resultados de la experiencia, es el criterio para comprobar una teoría científica.

Enfoque frecuencial de la probabilidad

Es necesario recordar el concepto de probabilidad en forma heurística. La probabilidad de un evento es la frecuencia relativa de su ocurrencia en un gran número de intentos. Esto, en principio, no se contradice con la definición axiomática de la probabilidad que se considera como una función que satisface axiomas que provienen de propiedades elementales de las frecuencias relativas. Más adelante, se asocia el concepto de *distribución de frecuencias* a la distribución del recorrido muestral de las variables aleatorias y el de *distribución de probabilidad* a la distribución del recorrido poblacional de las mismas.

> OBSERVACIÓN: La teoría de probabilidad es una teoría desarrollada para describir los eventos aleatorios. Se puede ver la teoría de probabilidad desde dos puntos de vista: sin teoría de medidas o con teoría de medidas. El primer punto de vista es lo que se enseña primero, y entonces se introduce la teoría de medidas. En este cuaderno se reseñan los conceptos básicos de la probabilidad. La probabilidad es la característica de un evento del que existen razones para creer que se realizará.

Los eventos tienden a ser una frecuencia relativa del número de veces que se realiza el experimento. La probabilidad de aparición de un evento E de un total de M casos posibles igualmente factibles es la razón entre el número de ocurrencias r de dicho evento y el número total de casos posibles M. La probabilidad es un valor numérico entre 0 y 1. Cuando el evento es imposible se dice que su probabilidad es 0, un evento cierto es aquel que ocurre siempre y su probabilidad es 1. La probabilidad de no ocurrencia de un evento está dada por $q = P(\bar{E}) = 1 - (r/M)$. Simbólicamente el espacio de resultados, que normalmente se denota por S, es el espacio que consiste en todos los resultados que son posibles. Los resultados, que se denotan por $x_{ji}, (i = 1,2, \dots, n,\ j = 1, 2, \dots, k)$, son elementos del espacio muestral S, para la variable aleatoria j.

Según Spiegel (1976) la definición de la probabilidad estimada o empírica está basada en la frecuencia relativa de aparición de un evento E cuando S es muy grande. La probabilidad de un evento es una cantidad, se escribe como $p(E)$ y mide con qué frecuencia ocurre algún evento si se hace algún experimento indefinidamente. La definición anterior es complicada de representar matemáticamente ya que S debiera ser infinito. Otra manera de definir la probabilidad es de forma axiomática, estableciendo las relaciones o propiedades que existen entre los conceptos y operaciones que la componen.

En términos generales y bajo condiciones constantes, si se considera que M es el número de experimentos que se llevan a cabo -el número de veces que la tabla de datos planteada puede ser aplicada para el estudio de un mismo fenómeno económico-, entonces la probabilidad $p(E)$ del evento E se define como

$$p(E) = \lim_{M \to \infty} \frac{r}{M} \qquad (2.2)$$

donde r es el número de ocurrencias del evento E y M es el número de experimentos. Por ejemplo, la probabilidad de que una persona tenga un ingreso perteneciente al segundo intervalo de clase en el ejemplo 2.1.

La probabilidad definida en (2.2) proviene evidentemente (y es una extensión) del concepto de frecuencia relativa r/M, donde M es finito.

Aun cuando r/M, con M finito, es inestable en casi todos los fenómenos observables, se requiere que el límite (2.2) exista; aunque para muchos fenómenos puede no existir. La estabilidad final de la frecuencia relativa está explícita en la definición de probabilidad. Por lo tanto, las probabilidades se definen de manera completamente legítima y se usan provechosamente en aquellos campos en los cuales M puede ser grande y donde, para M grande, la estabilidad de las frecuencias relativas de eventos es un hecho empírico.

No implica necesariamente que se lleven a cabo $M \to \infty$ intentos, sino que infinitos intentos son conceptualmente posibles.

Definición axiomática de la probabilidad

Las frecuencias relativas finitas satisfacen, con pocas excepciones, los axiomas sobre probabilidades. Con respecto a un experimento, sea S un conjunto numerable de puntos, cada uno de los cuales representa un posible resultado de un intento; S es el conjunto de todos los resultados posibles de un intento, esto es, el conjunto de todos los x_{ji} posibles. Sea E cualquier subconjunto de puntos en el conjunto S, p indicará una función que asegura a cada evento E un número real $p(E)$, llamado la probabilidad de E. En la teoría de la probabilidad, generalmente, se encuentra el siguiente conjunto mínimo de *axiomas*:

1) $p(E) \geq 0$
2) $p(S) = 1$
3) $p(E + F) = p(E) + p(F)$, siendo E y F mutuamente excluyentes.

Para tratar con eventos E y F que no son mutuamente excluyentes, se tratará de expresarlos en términos de eventos

que sí lo son, puesto que solamente sobre ellos existe un axioma.

Nótese que al contrario de la definición (2.2), los axiomas no sugieren cómo estimar $p(E)$. Muchas estimaciones numéricas de probabilidades satisfacen estos axiomas; es necesario, entonces, aprender a seleccionar aquellos que son los más adecuados para el problema particular de que se trate.

Los eventos a los que se han asignado probabilidades pertenecen a una clase elemental. Se desea introducir ahora una manera nueva de referirse a algunos de los eventos para los que se pueden asignar probabilidades. Mediante la introducción de una función $X(E)$ se asocia con cada punto E perteneciente a S un número real $x_{ji}, (\forall i = 1, 2, \dots, n \; y \; \forall \; j = 1, 2, \dots, k)$, esto es, se construye una correspondencia entre los puntos en el espacio S de eventos y los números reales $x_{j1}, x_{j2}, \dots, x_{jn}$, correspondiente a la variable aleatoria X_j. La relación no necesita ser biunívoca, se pueden representar varios puntos de S con números reales x_{ji}. Entonces, el espacio original S puede verse como si estuviera representado en un nuevo espacio de eventos que consiste en puntos sobre la línea real; los eventos particulares E de interés en S, esto es, subconjuntos particulares de puntos de S, se convierten ahora en colecciones particulares de puntos sobre la línea real.

Ejemplo 2.2. En el experimento de observar la variable ingreso sobre un conjunto de individuos, agrupados en intervalos de clase, asigna un valor a cada individuo; es decir, mediante la función $X(E)$ se le asigna al grupo de individuos del primer intervalo de clase el valor medio x_{j1} que asume para esa clase la variable j, en este caso ingresos, al segundo grupo de individuos, intervalo de clase 2 el valor x_{j2}, y así siguiendo…, sobre la línea real. Por ejemplo, si se tiene $x_{j1} = 75$, $x_{j2} = 225$, puede interpretarse convenientemente como el valor del ingreso para el primero y segundo grupo de individuos, observado en un experimento, representado en la investigación econométrica en una tabla de datos según lo observado en la Figura 2.5.

Cualquier característica cuantitativa del experimento se llama variable aleatoria. Por ejemplo, el resultado de $225 en la medición de la variable ingreso.

	X_1	...	X_j	...	X_k
1	x_{11}		**75**		x_{1k}
2	x_{21}		**225**		x_{2k}
3	x_{31}		x_{3j}		x_{3k}
⋮	⋮		⋮		⋮
i	x_{i1}		x_{ij}		x_{ik}
⋮	⋮		⋮		⋮
n	x_{n1}		x_{nj}		x_{nk}

Figura 2.5. Representación de un experimento en la investigación econométrica (tabla de datos)

2.2. Variables aleatorias

Al describir el espacio muestral de un experimento, no se especificó que necesariamente un resultado individual debe ser un número. Aunque esto no debe ser un inconveniente, ya que se podría medir algo y consignarlo como un número. Por ejemplo, el sexo podría ser nominado como 1 femenino y 2 masculino. Lo importante aquí es que, en muchas situaciones experimentales, se desea asignar un número real x a cada uno de los elementos s del espacio muestral S. Esto es, $x = X(s)$ es el valor de una función X del espacio muestral a los números reales. Teniendo esto presente, se puede expresar la siguiente definición formal:

Una *variable aleatoria* es el resultado numérico de un experimento aleatorio. Matemáticamente, es una aplicación $X: S \to \mathbb{R}$ que da un valor numérico (X), del conjunto de los reales (R), a cada evento en el espacio (S) de los resultados posibles del experimento.

Una variable es un elemento de una fórmula, proposición o algoritmo que puede adquirir o ser sustituido por un valor

cualquiera; los valores están definidos dentro de un rango. Mientras que, *variable aleatoria*, es una variable que cuantifica los resultados de un experimento aleatorio, que puede tomar diferentes valores cada vez que ocurre un suceso; el valor sólo se conocerá determinísticamente una vez acaecido el evento.

Freeman (1979) establece que "la mayoría de las variables que nos son conocidas (longitud, peso, ingreso, precio, resistencia, densidad) se manejan cómodamente con representaciones sobre los números reales; otras, como la inteligencia, preferencia, sabor y color, se han cuantificado, y a menudo se puede hacerlo, para manejarse con ese mapeo sobre los números reales. La función $X(S)$ que efectúa este cambio del espacio de eventos anterior al nuevo se llama función aleatoria, o más comúnmente, variable aleatoria"; el adjetivo *aleatorio* se emplea para recordarnos que el dato x_{ji} de un experimento sobre X_j "es cierto y, por tanto, que el valor de la función (y no la función misma) es incierto. Una variable aleatoria es simplemente una función de valores reales definidos en S" (pag. 30 y ss.)

Al igual que las variables cuantitativas, las *variables aleatorias cuantitativas* se clasifican en, *variable aleatoria discreta*, variable que toma un número finito o infinito de valores numerables; y *variable aleatoria continua*, variable que puede asumir cualquier valor dentro de un intervalo de valores.

Ejemplo 2.3: Suponga que se lanzan dos monedas al aire y se establece que X es la variable aleatoria que identifica el número de caras obtenidas en el lanzamiento. El espacio muestral (S) es $S = \{cc, cs, sc, ss\}$ donde c identifica una cara y s un sello. El recorrido de X (R_X,) es 0, 1 o 2 ($R_X = \{0, 1, 2\}$) de acuerdo al número de caras obtenidas en el lanzamiento. Así será: $X: S \rightarrow \mathbb{R}\ cc \rightarrow 2\ cs, sc \rightarrow 1\ ss \rightarrow 0$, siendo X una variable aleatoria discreta. Otros experimentos como cantidad de hijos de una familia elegida al azar, definen variables aleatorias discretas.

Ejemplo 2.4. Sea X la variable aleatoria altura de un individuo elegido al azar en el intervalo $[1{,}40; 1{,}90]$. Entonces el Recorrido de X, será: $X: S \to \mathbb{R} \wedge 1{,}40 \leq x \leq 1{,}90$, siendo X una variable aleatoria continua y puede asumir cualquier valor en dicho intervalo, dependiendo de la precisión del instrumento de medición. Factores análogos como el ingreso de un individuo o la cantidad de lluvia caída pueden considerarse como variables aleatorias continuas.

Una variable aleatoria tiene una *distribución de probabilidades*. Ésta es un modelo teórico que describe la forma en que varían los resultados de un experimento aleatorio. Es decir, detalla los resultados de un experimento con las probabilidades que se esperarían ver asociadas con cada resultado. La función *de distribución* es la función que acumula probabilidades asociadas a una variable aleatoria.

Se denomina: *función de probabilidad* a la función que asigna probabilidades a cada uno de los valores de una variable aleatoria *discreta; y función de densidad* a la función que mide concentración de probabilidad alrededor de los valores de una variable aleatoria *continua*.

Dada una variable aleatoria X se pueden calcular diferentes medidas como la media aritmética, la media geométrica o la media ponderada y también el valor esperado y la varianza de la distribución de probabilidad de X.

OBSERVACIÓN: Nótese que los resultados posibles de un experimento son variables aleatorias, pero una vez que se ejecutó el experimento; es decir una vez que, por ejemplo, se seleccionó una persona y se midió su ingreso anual, se obtiene un valor específico para la variable. Esta distinción entre una variable aleatoria y su valor es muy importante. Este último es un valor no estocástico. No cambia si no cambia el experimento, o lo que es lo mismo, no cambia si no cambian las condiciones bajo las cuales se lleva a cabo el experimento. Es decir, entre las dos interpretaciones posibles que la bibliografía le suele dar al recorrido R_X de

una variable aleatoria, en econometría es evidente que como dice Meyer (1973) “el experimento aleatorio termina, de hecho, con la observación de s $\in$ S”.

Las variables aleatorias se escriben en mayúsculas $X_j, \forall\, j = 1, \ldots, k$ y los resultados posibles de una variable aleatoria en minúsculas $x_{ji}\ \forall\, i = 1, \ldots, n$.

2.3 Distribuciones de frecuencia y distribuciones de probabilidad

Una *distribución de frecuencias* representa una ordenación de los datos de tal forma que indique el *número de observaciones*, surgidas de una muestra, en cada clase de la variable. El número de observaciones en cada clase se denomina *frecuencia absoluta;* ésta se distingue de la *frecuencia relativa*, que indica la *proporción de observaciones* en cada clase.

En una distribución de frecuencias los datos pueden agruparse, como en el ejemplo 2.1, 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, el que puede ser determinado sumando los límites superior e inferior y dividiendo por dos.

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.

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

En el ejemplo 2.6 el 16,67% de *todos* los Estados de Resultados mensuales tuvieron una incidencia de gastos financieros inferior al 5%, esto equivale a afirmar que la probabilidad de seleccionar aleatoriamente un Estado de Resultado con gastos financieros inferiores al 5% es 0,1667.

Ejemplo 2.5. 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 la Figura 2.6.

% gasto financiero (variable)	N° de Estados de Resultados (frec.absoluta)	Proporción de Estados de Resultados (frec.relativa)
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

Figura 2.6. Muestra de los Estados de Resultados de la Empresa ***A***

Ejemplo 2.6. 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 la Figura 2.7.

% 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 %

Figura 2.7: Estados de Resultados de la Empresa A

Con la distribución muestral se puede inferir la poblacional, evaluando estadísticamente el error cometido a partir de la inferencia. 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. No siempre la muestra representa exactamente a la población, como en este caso; pero la ampliación sobre este tema se deja para el próximo cuaderno

En resumen, se utilizan las muestras para hacer juicios respecto de 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 los juicios basados en el estadístico de la muestra que recibe el nombre de *estimador*.

Ejemplo 2.7. 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 de interés (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 la variable gasto financiero de los últimos 12 ejercicios económicos de la Empresa A (variable X) y el interés radica en estimar el gasto financiero medio (un parámetro α) se usa el estadístico muestral a -donde a es un estimador de α-, y un valor concreto de a (obtenido a partir de una muestra determinada) es una estimación de α. Como el gasto financiero de una empresa es una variable continua que tiene valores positivos, la distribución muestral de a tendrá, posiblemente, una forma como la que se muestra en la Figura 2.8.

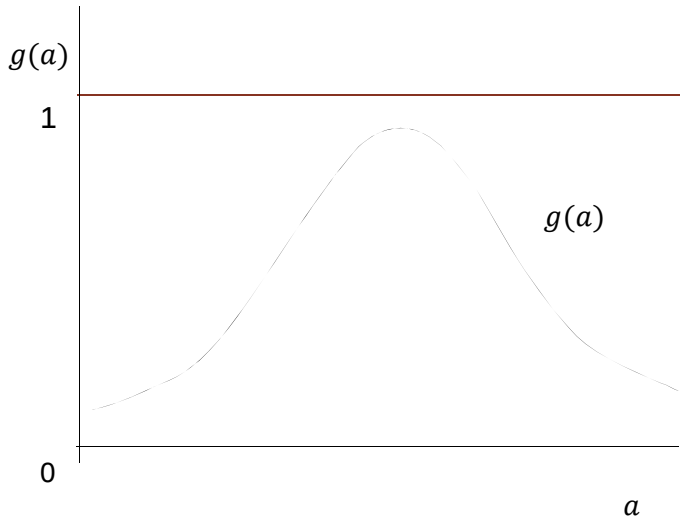


Figura 2.8. Distribución muestral

Sin mayor rigurosidad se puede decir que, el concepto de distribución muestral se entiende 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 2.8, las frecuencias relativas de a se miden a lo largo del eje $g(a)$.

Con el análisis anterior se puede ver la 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 aleatorias 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.

2.4. Distribuciones teóricas de probabilidad

La teoría de las Distribuciones Estadísticas es fundamental en el proceso de investigación econométrica. Es necesario distinguir entre las *distribuciones experimentales* y las distribuciones teóricas, teniendo en cuenta que estas últimas se basan en la teoría de la probabilidad.

La *función de densidad o función de probabilidad* de una variable aleatoria $X_j, \forall\, j = 1, \dots, k$, es una función a partir de la cual se obtiene la probabilidad de cada valor x_{ji} $(\forall\, i = 1, \dots, n)$ que toma la variable.

La *función de distribución de probabilidad* $F(X_j)$, es una función de la probabilidad que representa los resultados que se van obteniendo en un experimento aleatorio.

Para un valor dado x_{ji}, la probabilidad $p(X_j \leq x_{ji}) = F(X_j)$.

Con las anteriores definiciones se puede realizar una extensión de los axiomas de probabilidad de la siguiente manera:

Para dos números reales cualesquiera $a,\ b$ [tal que $(a < b)$ y que (a, b) pertenezcan a un intervalo, por ejemplo el del recorrido de la variable X_j, en la Figura 2.1, esto es $x_{j1} \leq a, b \leq x_{jn}$], los eventos $p(X_j \leq a)$ y $p\ (a < X_j \leq b)$ serán mutuamente excluyentes y su suma es el nuevo evento $p(X_j \leq b)$ por lo que:

$$p(X_j \leq b) = p(X_j \leq a) + p\ (a < X_j \leq b)$$

$$p\ (a < X_j \leq b) = p(X_j \leq b) - p(X_j \leq a)$$

Y, finalmente

$$p\ (a < X_j \leq b) = F(b) - F(a)$$

Por lo tanto, una vez conocida la función de distribución $F(X_j)$ para todos los valores de la variable aleatoria X_j se conocerá completamente la distribución de probabilidad de la variable.

Distribuciones de variable discreta

La *función de probabilidad* de una variable aleatoria discreta X_j, indicada como $f(X_j)$, se define como una regla que asigna a cada número real x_{ji} la probabilidad de que la variable X_j asuma el valor x_{ji}. Es decir,

$$f(X_j) = p(X_j = x_{ji}) \qquad (2.3)$$

En cambio, la *función de distribución de probabilidad* de X_j, indicada como $F(X_j)$, se define como una regla que asigna a cada número real x_{ji} la probabilidad de que la variable aleatoria X_j sea igual o menor al valor de x_{ji}. Es decir,

$$F(X_j) = p(X_j \leq x_{ji}) = \sum_{X_j \leq x_{ji}} f(x_{ji}) \qquad (2.4)$$

Ejemplo 2.8. Se define una variable aleatoria X como las unidades que constituyen la demanda de los productos de la Empresa A durante el próximo año. Se suponen posibles e igualmente probables cuatro niveles de venta: 10, 12, 15 ó 18 unidades. Como las probabilidades de estos cuatro resultados posibles deben sumar 1, la función de probabilidad de X está dada por:

$$f(X) = \begin{cases} P(X = 10) = \dfrac{1}{4} \\ P(X = 12) = \dfrac{1}{4} \\ P(X = 15) = \dfrac{1}{4} \\ P(X = 18) = \dfrac{1}{4} \end{cases} \qquad (2.5)$$

(2.5) indica que la probabilidad de que la demanda sea de 10, 12, 15 ó 18 unidades es cada una igual a ¼.

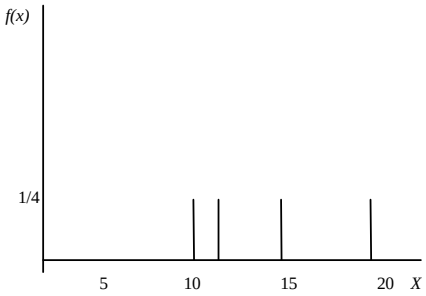
La función de distribución de probabilidad de X, estará dada por:

$$F(X) = \begin{cases} P(X \leq 10) = \dfrac{1}{4} \\ P(X \leq 12) = \dfrac{1}{2} \\ P(X \leq 15) = \dfrac{3}{4} \\ P(X \leq 18) = 1 \end{cases} \qquad (2.6)$$

(2.6) dice que hay una probabilidad de ¼ que la demanda sea igual o menor a 10 unidades, una probabilidad de ½ de que la demanda real sea menor o igual a 12 unidades, una probabilidad de ¾ de que la demanda sea menor o igual a 15 unidades y una probabilidad cierta (igual a 1) de que la demanda sea menor o igual a 18 unidades.

La Figura 2.9 representa la función de probabilidad dada por (2.5) y a la función de distribución de probabilidad dada por (2.6).

a) Función de Probabilidad



b) Función de Distribución de probabilidad

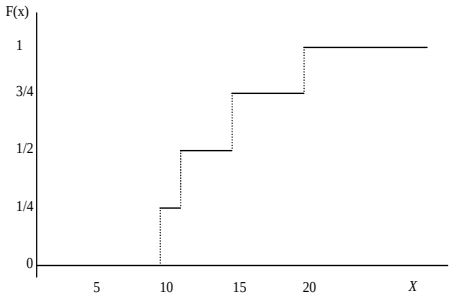


Figura 2.9 Pronóstico de Ventas

La Figura 2.9 muestra la gráfica del caso particular de la Distribución Binomial P(X=k) para k=10,12,15, 18.

Distribuciones de variable continua

El análisis anterior se ve modificado en el campo de las variables aleatorias continuas. La *función de densidad* de $X_j, \forall j = 1, \dots, k$ (donde X_j es, por ejemplo, la variable definida en la Figura 2.1, $\forall i = 1, \dots, n$ unidades de observación), es:

$$f(X_j) = p(x_{j,i} \leq X_j \leq x_{j,i+r}) = \int_{x_{j,i,}}^{x_{j,i+r}} f(X_j) dX_j \quad \text{donde } x_{j,i} < x_{j,i+r} \quad (24.7)$$

$f(X_j)$ es una función donde el área bajo la misma, entre $x_{j,i} \leq X_j \leq x_{j,i+r}$ es exactamente la probabilidad de que X_j asuma un valor entre $x_{j,i}$ y $x_{j,i+r}$. De la misma manera, la *función de distribución de probabilidad*, $F(X_j)$, está dada por la expresión:

$$F(X_j) = p(X_j < x_{j,i+r}) = \int_{-\infty}^{x_{j,i+r}} f(X_j)dX_j \qquad (2.8)$$

Es decir, para determinar la probabilidad acumulativa de que X_j sea igual o menor que $x_{j,i+r}$, se calcula el área bajo la función de densidad $f(X_j)$ entre $-\infty$ y $x_{j,i+r}$.

OBSERVACIÓN: La probabilidad de que la variable aleatoria continua X_j sea exactamente igual a cierto valor $x_{j,i}$ es cero.

Ejemplo 2.9. Una función de densidad de una variable aleatoria continua es la distribución de probabilidad normal estándar. La función de densidad y la función de distribución de probabilidad de una variable aleatoria X_j normal estándar son las que muestra la Figura 2.10.

a) Función de densidad

b) Función de distribución de probabilidad

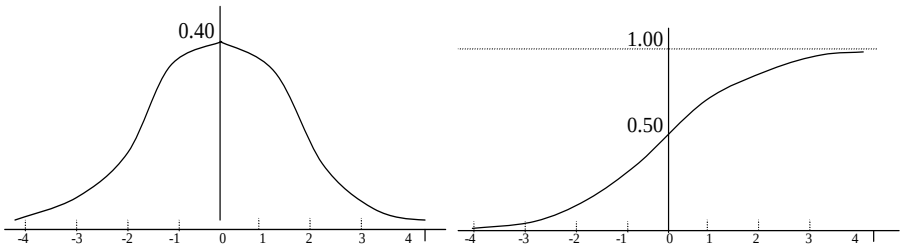


Figura 2.10 Distribución normal estándar

Esperanza matemática

La *esperanza matemática* de una variable aleatoria es la suma de la probabilidad de cada evento multiplicada por su valor. *Si todos los sucesos son de igual probabilidad, la esperanza es la media aritmética.*

La esperanza matemática o valor esperado de una variable aleatoria discreta X_j, indicado como $E(X_j)$, se define:

$$E(X_j) = \sum_{\forall i} X_j f(X_j) \qquad (2.9)$$

Si x_{ji} representa cualquier valor posible de X_j, y $f(X_j)$ es la probabilidad de que $X_j = x_{ji}$, entonces $E(X_j)$ es un promedio ponderado de todos los valores posibles de X_j, donde las ponderaciones son las respectivas probabilidades de estos valores.

La varianza de una variable aleatoria X_j, indicada por $V(X_j)$, se define:

$$V(X_j) = \sum_{i=1}^{n} \left(x_{ji} - E(X_j)\right)^2 f(X_j) \qquad (2.10)$$

donde, todos los términos responden a las definiciones anteriores. Es decir, $V(X_j)$ es un promedio ponderado de las desviaciones cuadráticas de los valores observados de X_j con respecto a su valor esperado, donde las ponderaciones son las respectivas probabilidades.

La desviación estándar de una variable aleatoria X_j, de suma utilidad práctica, se define como la raíz cuadrada de la $V(X_j)$.

El valor esperado y la varianza de variables aleatorias con distribución de probabilidad continua, se definen con las respectivas fórmulas como:

$$E(X_j) = \int_{-\infty}^{\infty} x_{ji} f(X_j) dX_j \qquad (2.11)$$

$$V(X_j) = \int_{-\infty}^{\infty} \left[x_{ji} - E(X_j)\right]^2 f(X_j) dX_j \qquad (2.12)$$

donde, $f(X_j)$ es la función de densidad de la variable aleatoria continua X_j.

2.5. FUNCIÓN GENERATRIZ DE MOMENTOS

Por lo visto anteriormente, la variable aleatoria tiene ciertas características que la distingue de otras variables como distribución de probabilidad, esperanza matemática y varianza. En esta sección se estudia la forma de derivar esas características.

La esperanza matemática y la varianza de una variable aleatoria se pueden obtener a partir del estudio de los momentos naturales y centrados. En términos generales se puede decir que la esperanza matemática es el momento natural (o respecto al origen) de orden 1; mientras que la varianza es el momento natural de orden 2 menos el momento natural de orden 1, elevado al cuadrado.

Momentos de una distribución

a) Momentos poblacionales con respecto al origen (naturales)

Los momentos de orden $r = 1, 2, \ldots$ con respecto al origen, de la variable aleatoria continua X_j con función de densidad $f(X_j)$, se definen por:

$$m'_r = E(X_j^r) = \int_{-\infty}^{\infty} X_j^r f(X_j) dX_j \qquad (2.13)$$

Los momentos de orden $r = 1, 2, \ldots$ con respecto al origen, de la variable aleatoria discreta X_j con función de probabilidad $f(X_j) = p(X_j = x_{ji})$, se definen por:

$$m'_r = E(X_j^r) = \sum_{i=1}^{N} X_{ji}^r \ f(X_j) \tag{2.14}$$

El primer momento m'_1 recibe el nombre de media de X_j.

b) Momentos poblacionales con respecto a la media (o centrados)

Cuando existe la esperanza matemática de una variable aleatoria X_j, pueden definirse momentos con respecto a dicha esperanza por medio de:

$$m_r = E\left[\left(X_j - m'_1\right)^r\right] = \int_{-\infty}^{\infty} \left(X_j - m'_1\right)^r f(X_j) d(X_j) \tag{2.15}$$

si la distribución de X_j es de tipo continuo, o de:

$$m_r = E\left[\left(X_j - m'_1\right)^r\right] = \sum_{i=1}^{N} \left(X_{ji} - m'_1\right)^r f(X_j) \tag{2.16}$$

si la distribución es discreta.

De esta forma, el momento centrado de orden 2 recibe el nombre de varianza de X_j, ya que en el caso continuo,

$$m_2 = E\left[\left(X_j - m'_1\right)^2\right] = \int_{-\infty}^{\infty} \left(X_j - m'_1\right)^2 f(X_j) d(X_j) =$$

$$= \int_{-\infty}^{\infty} [(X_j^2 - 2X_j m'_1 + (m'_1)^2] f(X_j) d(X_j) =$$

$$= \int_{-\infty}^{\infty} X_j^2 f(X_j) d(X_j) - 2m'_1 \int_{-\infty}^{\infty} X_j f(X_j) d(X_j) + (m'_1)^2 \underbrace{\int_{-\infty}^{\infty} f(X_j) d(X_j)}_{1} =$$

$$= m'_2 - 2m'_1 m'_1 + (m'_1)^2 =$$

$$= m'_2 - (m'_1)^2$$

El momento centrado de orden 2 (m_2) o varianza es la diferencia de los momentos naturales 2 (m_2') y el cuadrado del momento natural 1 (m_1').

Función generatriz de momentos

Los momentos de una función de densidad desempeñan un papel muy importante en estadística teórica y aplicada. Dicen Mood y Graybill (1969) que, en algunos casos, "si se conocen todos los momentos, puede determinarse la función de densidad" (pp 130) de cualquier variable aleatoria X_j.

Por lo tanto, es útil hallar una función que genere los momentos de una función de densidad o de probabilidad. Esta función es la función generatriz de momentos, que se define como el valor esperado de $e^{\theta X_j}$ si este valor existe para todo valor de θ de determinado intervalo $-h^2 < \theta < h^2$, y se puede representar por:

$$M_{X_j}(\theta) = E\left(e^{\theta X_j}\right) = \int_{-\infty}^{\infty} e^{\theta X_j} f(X_j)\, dX_j \qquad (2.17)$$

si la variable aleatoria X_j es continua, y por

$$M_{X_j}(\theta) = E\left(e^{\theta X_j}\right) = \sum_{X_j} e^{\theta X_j} f(X_j) \qquad (2.18)$$

si la variable aleatoria es discreta.

Si existe una función generatriz de momentos, $M_{X_j}(\theta)$ es independiente derivable en un entorno del origen. Si se deriva r veces respecto a θ, se obtiene

$$\frac{d^r}{d\theta^r} M_{X_j}(\theta) = \int_{-\infty}^{\infty} X_j^r e^{\theta X_j} f(X_j)\, dX_j$$

Y haciendo $\theta = 0$, se encuentra

$$\frac{d^r}{d\theta^r} M_{X_j}(\theta = 0) = \int_{-\infty}^{\infty} X_j^r f(X_j)\, dX_j = m'_r \qquad (2.19)$$

Propiedades

a) La *función generatriz de momentos de la suma de variables aleatorias* independientes en el parámetro θ es igual al *producto* de las respectivas funciones generatrices de momentos de dichas variables aleatorias.

$$M_{X_1+X_2+\cdots+X_n}(\theta) = E\left\{e^{\theta(X_1+X_2+\cdots+X_n)}\right\}$$

Aplicando la propiedad distributiva

$$E\left\{e^{\theta x_1} e^{\theta x_2} e^{\theta x_3} \ldots e^{\theta x_n}\right\}$$

Dado que las variables son independientes, lo anterior puede expresarse como:

$$E\left(e^{\theta x_1}\right) E\left(e^{\theta x_2}\right) E\left(e^{\theta x_3}\right) \ldots E\left(e^{\theta x_n}\right)$$

Por (2.17) se sabe que

$$E\left(e^{\theta x}\right) = M_x(\theta)$$

Teniendo en cuenta esto y reemplazando

$$M_{x_1}(\theta)\, M_{x_2}(\theta)\, M_{x_3}(\theta) \ldots\ M_{x_n}(\theta)$$

$$M_{x_1+x_2+\cdots+x_n}(\theta) = M_x(\theta)\, M_x(\theta) \ldots\ M_x(\theta)$$

b) La *función generatriz de momentos del producto de una constante por una variable* para el parámetro θ, $M_{cX_j}(\theta)$, es la función generatriz de momentos del producto del parámetro θ por la constante (c) para la variable X_j:

$$M_{cX_j}(\theta) = M_{X_j}(c\,\theta) \qquad (2.20)$$

2.6. ALGUNAS DISTRIBUCIONES DE PROBABILIDAD

No necesariamente todas las variables aleatorias bajo estudio responden a las distribuciones teóricas de probabilidad. Existen las distribuciones experimentales que, una vez obtenidas, pueden o no responder a las formas de las distribuciones teóricas.

En las Figuras 2.11 y 2.12 se presentan algunas distribuciones teóricas de probabilidad, tanto discretas como continuas. Se han incluido en el cuadro las funciones de probabilidad (también llamadas funciones de densidad cuando están asociadas con variables aleatorias continuas) y los principales parámetros (media, varianza) de las distribuciones. En las figuras, las fórmulas ilustran la distribución de una variable genérica X_j medida sobre unidades de observación poblacionales, esto es $1 \leq i \leq N$, con datos representados por x_{ji}.

Es importante comentar aquí que, a partir de las distribuciones muestrales se obtienen estimadores de los parámetros poblacionales. Estos estimadores, por provenir de una muestra aleatoria constituyen, en sí mismos, variables aleatorias sujetas a distribuciones de probabilidad y a distribuciones acumulativas de probabilidad. Esta es la verdadera naturaleza de la Inferencia Estadística.

Necesidad del uso de probabilidades

En cualquier circunstancia, en el ámbito macro o microeconómico, toda decisión tiene efecto durante un período de tiempo que se extiende hacia el futuro. Esta característica, que es común a todas las decisiones, probablemente se observe con mayor intensidad en las áreas empresariales. Sin embargo,

una decisión involucra aspectos del futuro, cualquiera sea la base sobre la que se tome.

Distribución de X_j	Función de Probabilidad (caso discreto)	Parámetros	
		$E(X_j)$	$V(X_j)$
Uniforme	$p(X_j = x_{ji}) = \frac{1}{N}$	$\sum_{i=1}^{N} \frac{x_{ji}}{N}$	$\frac{\sum_{i=1}^{N}(x_{ji} - E(X_j))^2}{N}$
Poisson	$P(x_{ji} = k) = \frac{e^{-\alpha}\alpha^k}{k!}; k = 0, 1, \dots;\ \alpha > 0$	α	α
Geométrica	$P(x_{ji} = k) = (1-p)^{k-1}p; k = 0, 1, 2, \dots$	$\frac{1}{p}$	$\frac{1-p}{p^2}$
Binomial	$P(x_{ji} = k) = \binom{n}{k} p^k (1-p)^{n-k} p; k = 0, 1, 2, \dots, n$	np	$np(1-p)$
Binomial Negativa	$b^*(x; k; \theta) = \binom{x-1}{k-1} \theta^k (1-\theta)^{x-k};$ $r > 0; k = 0, 1, \dots;\ x = k, k+1, k+2, \dots$	$\frac{k}{\theta}$	$\frac{k(1-\theta)}{\theta^2}$
Pascal	$P(x_{ji} = k) = \binom{k-1}{r-1} p^r (1-p)^{k-r}; k = r, r+1, \dots$	$\frac{r}{p}$	$\frac{r(1-p)}{p^2}$
Hipergeométrica	$P(x_{ji} = K) = \frac{\binom{r}{k}\binom{N-r}{n-k}}{\binom{N}{n}}; k = 0, 1, \dots$	np	$np(1-p)\frac{N-n}{N-1}$
Multinomial	$P(x_{j1} = n_1, x_{j2} = n_2, \dots, x_{jk} = n_e) =$ $= \begin{cases} \frac{n!\, p_1^{x_{j1}} \dots p_e^{x_{je}}}{n_1!\, n_2! \dots\, n_e!}; cuando \sum_{i=1}^{e} x_{ji} = N \\ 0;\ en\ otro\ caso \end{cases}$	np_i	$np_i(1-p_i);$ $i = 1, \dots e$

Figura 2.11: Algunas distribuciones teóricas de probabilidad de variable discreta

Teniendo en cuenta que al evaluar una propuesta se estará mirando hacia el futuro, ésta se traducirá en estimaciones de variables; por ejemplo, costos, ventas, precios, inversiones o impuestos, que estarán sujetas a cierto nivel de incertidumbre. Ante este nivel de incertidumbre en la estimación de variables importantes para la economía, ¿es suficiente trabajar con el valor sospechado, probable o experimental?, o ¿es más conveniente trabajar con la distribución de probabilidad de cada variable?

Hay que tener en cuenta que el riesgo es inseparable en la estimación de cualquier alternativa de decisión. Evidentemente, en el campo de la toma de decisiones, es más importante basarse en los métodos probabilísticos que en los subjetivos.

Distribución de X_j	Función de Densidad (Caso Continuo)	Parámetros	
		$E(X_j)$	$V(X_j)$
Distribución Uniforme	$f(X_j) = \begin{cases} \frac{1}{b-a} & a \leq X_j \leq b \\ 0 & resto \end{cases}$	$\frac{(a+b)}{2}$	$\frac{(b-a)^2}{12}$
Distribución Pareto	$f(X_j) = \frac{ab^a}{X_j^{a+1}} \quad X_j > b$	$\frac{ab}{a-1}$	$\frac{ab^2}{(a-1)^2(a-2)}$
Distribución beta	$f(X_j) = \frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)} X_j^{a-1}(1-X_j)^{b-1} \quad 0 < x < 1$	$\frac{a}{a+b}$	$\frac{ab}{(a+b+1)(a+b)^2}$
Normal	$f(X_j) = \frac{1}{\sqrt{2\pi}\sigma_j} \exp\left[-\frac{1}{2}\left(\frac{X_j-\mu_j}{\sigma_j}\right)^2\right]; -\infty < X_j < \infty$	μ_j	σ_j^2
Exponencial	$f(X_j) = \alpha e^{-\alpha X_j};\ X_j > 0$	$\frac{1}{\alpha}$	$\frac{1}{\alpha^2}$
Gamma	$f(X_j) = \frac{\alpha}{\Gamma(r)}(\alpha X_j)^{r-1} e^{-\alpha X_j};\ X_j > 0$	$\frac{r}{\alpha}$	$\frac{r}{\alpha^2}$

Figura 2.12: Algunas distribuciones teóricas de probabilidad de variable continua

2.7. DISTRIBUCIÓN NORMAL

Dentro del proceso de investigación econométrica, aunque proceso empírico, la distribución teórica más utilizada, para variable aleatoria continua, es la distribución normal. Quizás, la distribución de las alturas de los individuos de una población sea el ejemplo más utilizado en los libros de textos estadísticos para mostrar experimentalmente la distribución normal.

El uso que se hace en estos cuadernos de la distribución normal tiene que ver más con la necesidad de comprobar ciertas propiedades que exigen a la variable aleatoria seguir esta distribución teórica. Aunque también en inferencia estadística, más precisamente en la teoría del muestreo, se demuestra que el *estimador* de la media poblacional, obtenido de muestras lo suficientemente grandes o provenientes de una variable normal, sigue dicha distribución.

En general la variable aleatoria continua X_j que tiene una distribución normal se puede indicar como

$$X_j \sim N(\mu_j, \sigma_j^2) \qquad (2.21)$$

Esto significa que X_j, que puede asumir cualquier valor en el intervalo $(-\infty, \infty)$, sigue una *distribución* normal con *parámetros*, media y varianza, iguales a μ_j, σ_j^2, respectivamente.

La función de densidad de la distribución normal es

$$f(X_j) = \frac{1}{\sqrt{2\pi}\sigma_j} \exp\left[-\frac{1}{2}\left(\frac{X_j - \mu_j}{\sigma_j}\right)^2\right]; -\infty < X_j < \infty \qquad (2.22)$$

Donde, exp significa exponencial de $e = 2{,}71828\ldots$ (base del logaritmo neperiano), esto es e elevado a la potencia que se expresa entre corchetes, y $\pi = 3{,}14159\ldots$.

La función de densidad se puede representar como en la Figura 2.13, que muestra la simetría de la misma alrededor de su valor medio.

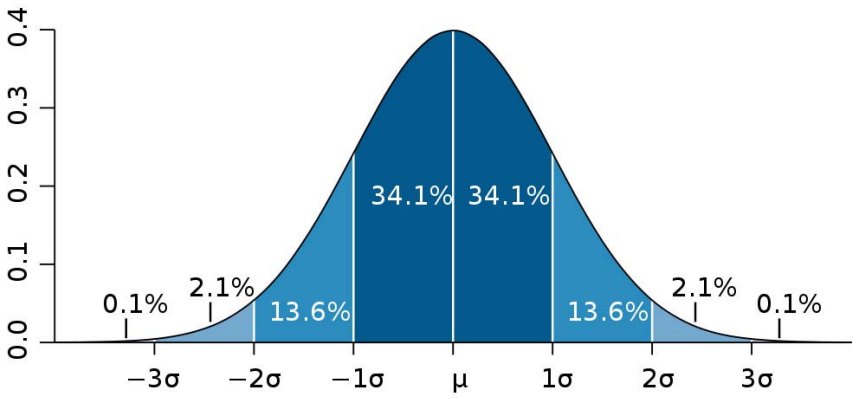


Figura 2.13. Áreas bajo la curva normal.

Fuente: http://www.flickr.com/photos/23925401@N06/6850657337/in/photostream/

La función de densidad alcanza su máximo en su valor medio y es asintótica en sus extremos. Esto se puede interpretar como

que la probabilidad de que la variable alcance un valor alejado de su valor medio es cada vez menor.

Aproximadamente, el 68% del área bajo la curva normal se encuentra entre los valores $\mu_j \pm \sigma_j$, el 95% del área se encuentra entre $\mu_j \pm 2\sigma_j$ y el 99,7% entre $\mu_j \pm 3\sigma_j$. Estas áreas se pueden utilizar como medidas de probabilidad.

Se demuestra a continuación que la integral entre $(-\infty, \infty)$ de la función de densidad de una variable con distribución normal es igual a 1. Esto significa que, la probabilidad de que $X_j \sim N(\mu_j, \sigma_j^2)$ asuma valores en el intervalo $(-\infty, \infty)$ es igual a uno, que es el valor del área bajo la curva normal.

Así, denominando I a la integral entre $(-\infty, \infty)$ de la función especificada en (2.22) se tiene:

$$I = \int_{-\infty}^{\infty} f(X_j) dX_j = \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}\sigma_j} e^{\left[-\frac{1}{2}\left(\frac{X_j - \mu_j}{\sigma_j}\right)^2\right]} dX_j =$$

$$= \frac{1}{\sqrt{2\pi}\sigma_j} \int_{-\infty}^{\infty} e^{\left[-\frac{1}{2}\left(\frac{X_j - \mu_j}{\sigma_j}\right)^2\right]} dX_j$$

Realizando la sustitución

$$z = \frac{X_j - \mu_j}{\sigma_j} \qquad (2.23)$$

Donde z se denomina variable aleatoria normal tipificada (o estandarizada) con distribución normal de media 0 y varianza 1

$$z \sim N(0, 1)$$

OBSERVACIÓN: la media y la varianza de z se obtiene de aplicar las propiedades de la esperanza matemática y de la varianza así

$$E(z) = \frac{1}{\sigma_j}\left[E(X_j) - E(\mu_j)\right] = \frac{1}{\sigma_j}\left[\mu_j - \mu_j\right] = 0$$

$$V(z) = \left(\frac{1}{\sigma_j}\right)^2 \left[V(X_j) - V(\mu_j)\right] = \frac{1}{\sigma_j^2}\left[{\sigma^2}_j - 0\right] = 1$$

Entonces,

$$I = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{\left[-\frac{1}{2}(z)^2\right]}\, dz$$

Se quiere demostrar que I es igual a 1.

Considerando el teorema de la adición para e^X y puesto que $e^{\left[-\frac{1}{2}(z)^2\right]}$ es par, la integral se puede factorizar como:

$$I^2 = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{\left[-\frac{1}{2}(z_1)^2\right]}\, dz_1 \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{\left[-\frac{1}{2}(z_2)^2\right]}\, dz_2$$

Siendo irrelevante el nombre de una variable ficticia, en este caso el remplazo de z, por z_1 y z_2.

Si $I = 1$, entonces se puede demostrar que $I^2 = 1$ y deducir inmediatamente que se cumple la propiedad. Por lo tanto,

$$I^2 = \frac{1}{2\pi} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} e^{\left[-\frac{1}{2}(z_1 + z_2)^2\right]} dz_1\, dz_2$$

Para resolver esta integral se realiza el cambio de variables expresadas en coordenadas cartesianas a *coordenadas polares*, mediante la sustitución

$$z_1 = r\, cos\, w$$

$$z_2 = r\, \mathrm{sen}\, w$$

OBSERVACIÓN: Las coordenadas cartesianas no es la única forma de describir la posición de un punto en el plano por dos números. Hay otras, entre ellas las coordenadas polares son particularmente útiles. Sea P un punto en el plano con coordenadas cartesianas (z_1, z_2). Supóngase que P no es el origen 0, por lo que la distancia euclidiana, $r = \sqrt{{z_1}^2 + {z_2}^2}$ puede reexpresarse como $r^2 = {z_1}^2 + {z_2}^2$

dando lugar a que la distancia de P a 0, es positiva. Entonces, es válido decir que $(z_1/r)^2 + (z_2/r)^2 = 1$. Así, $(z_1/r, z_2/r)$ son las coordenadas de un punto sobre la circunferencia unidad, y según la definición de la función seno y coseno, hay un número w tal que $z_1/r = cos\, w$ y $z_2/r = sen\, w$. Esta w es determinada sólo hasta un múltiplo de 2π. Los números (r, w) se llaman coordenadas polares de P. La relación entre coordenadas polares (r, w) y coordenadas cartesianas la dan las fórmulas o función de transformación,

$$z_1 = r\, cos\, w,\ z_2 = r\, sen\, w \qquad (2.24)$$

Estas también pueden verse en la Figura 2.14

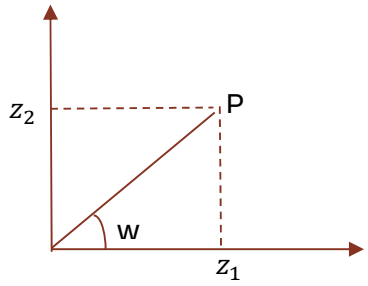


Figura 2.14 Transformación de coordenadas cartesianas a polares

Es conveniente ampliar la definición y considerar cada par de números (r, w) que satisfacen (2.24) como las coordenadas polares del punto con coordenadas cartesianas (z_1, z_2). Esta convención significa que

a) Si (r, w) son coordenadas polares de P, también lo son $(r, w \pm 2\pi)$;

b) Si (r, w) son coordenadas polares de P, también lo son $(-r, w \pm \pi)$;

c) Para cada w los números $(0, w)$ son coordenadas polares del origen.

Ejemplo 2.10. Hallar las coordenadas cartesianas del punto con coordenadas polares $(3, 3\pi/4)$. Se usa la ecuación (2.24) con $r = 3$ y $w = 3\pi/4$.

Entonces $z_1 = 3\,cos\left(\frac{3\pi}{4}\right) = 3\left(-\sqrt{2}/2\right) \cong -2{,}121$; $y\ z_2 = 3\,sen\,(3\pi/4) \cong 2{,}121$.

Hallar las coordenadas polares del punto con coordenadas cartesianas $(3, 4)$. Se usa la ecuación (2.24) despejando $cos\, w = z_1/r$; $sen\, w = z_2/r$.

Conociendo esto se tiene $r = \sqrt{3^2 + 4^2} = 5\ y\ w = tan\, arc\, 4/3 = tan\, arc\, 1.333\ldots \cong 53°$.

Teniendo en cuenta la observación anterior se procede a encontrar la solución de la integral, considerando el cambio de variables, para ello se tendrá en cuenta la función de transformación de coordenadas cartesianas a coordenadas polares. Esto es, $T(r, w) = (rcos\, w, rsen\, w)$ y el jacobiano de dicha transformación. En general el proceso de cambio de variables, en el plano, se puede sintetizar de la siguiente forma:

$$\int_{T(A)=D} f(x_1, x_2)dx_1dx_2 = \int_A f\big(T(v_1, v_2)\big)|\,J_T|\,(v_1, v_2)dv_1dv_2$$

Donde,

- $T: U \to R^2$ es la función de transformación (derivable y continua), $(x_1, x_2) = T(v_1, v_2)$ con Jacobiano $J_T(v_1, v_2) \neq 0$ en un conjunto abierto U
- Un subconjunto $A \subset U$ medible de forma que T sea biyectiva entre A y $T(A)$
- Una función $f: T(A) \to R$ acotada y continua

Entonces la integral se puede expresar como:

$$I^2 = \frac{1}{2\pi}\int_0^{2\pi}\int_0^{\infty} e^{\left[-\frac{1}{2}r^2\right]} r\, dr\, dw$$

Donde se realizaron los siguientes reemplazos:

- Los límites de integración se modificaron teniendo en cuenta la función de transformación dada en (2.24). Se pasa de un dominio de integración en coordenadas cartesianas a otro en coordenadas polares:

$$D = \{(z_1, z_2)/ -\infty \leq z_1 \leq \infty \ \wedge \ -\infty \leq z_2 \leq \infty\}$$

$$D = \{(r, w)/\ 0 \leq r \leq \infty \ \wedge \ 0 \leq w \leq 2\pi\}$$

Ya que como $z_1\, y\, z_2$ varían de $-\infty$ a ∞, se está abarcando todo el plano de coordenadas cartesianas, entonces se debe hacer lo mismo con el círculo. El radio r va de cero a ∞ y el ángulo w de cero a 2π. De esta forma, el círculo en coordenadas polares abarca también todo el plano.

a) El jacobiano de la transformación se obtuvo del hecho que,

$$J_T = \begin{vmatrix} \frac{\partial z_1}{\partial r} & \frac{\partial z_2}{\partial w} \\ \frac{\partial z_1}{\partial r} & \frac{\partial z_2}{\partial w} \end{vmatrix} = \begin{vmatrix} \cos w & -r\ sen\ w \\ sen\ w & r\cos w \end{vmatrix} = r(cos^2 w + sen^2 w) = r$$

De modo que el diferencial del área es

$$dz_1 dz_2 = dA \cong |J_T|\, drdw = rdrdw$$

Por lo tanto se procede a resolver la integral I^2. En primer lugar se resuelve

$$\int_0^{\infty} e^{\left[-\frac{1}{2}r^2\right]} r\, dr$$

Pero como,

$\int_0^{\infty} e^{f(X)} f'(X) dX = e^{f(X)} + C$, siendo C una constante de integración

Se tiene que

$$\int_0^{\infty} e^{-\frac{r^2}{2}} r dr = -\int_0^{\infty} e^{-\frac{r^2}{2}} r dr = -e^{-\frac{r^2}{2}}|_0^{\infty}$$

O sea que en el límite cuando B tiende a infinito, esto es

$$\lim_{B\to\infty} -e^{-\frac{r^2}{2}}|_0^B = \lim_{B\to\infty}\left[-\frac{1}{e^{\frac{B^2}{2}}} + \frac{1}{e^0}\right] = 1$$

Por lo tanto,

$$\int_0^{\infty} r e^{-\frac{1}{2}r^2}\, dr = 1$$

Ahora

$$I^2 = \frac{1}{2\pi}\int_0^{2\pi}\int_0^{\infty} e^{\left[-\frac{1}{2}r^2\right]} r\, dr\, dw = \frac{1}{2\pi}\int_0^{2\pi} dw = w|_0^{2\pi} = \frac{1}{2\pi}(2\pi - 0) = 1$$

Por lo que queda demostrado, ya que si I^2, entonces $I = 1$. Significa que la función de densidad de la distribución normal cumple con la propiedad de que su integral sea igual a 1.

Función generatriz de momentos de una variable con distribución normal

Si la función de densidad de una variable aleatoria $X_j \sim N(\mu_j, \sigma_j)$ es la función definida en (2.22) y si al tipificarla se obtiene una variable aleatoria normal centrada y reducida Z_j, denominada variable normal estándar dada en (2.23). La función de densidad de Z_j es

$$f(Z_j) = \frac{1}{\sqrt{2\pi}} \exp\left[-\frac{1}{2}(Z_j)^2\right]; -\infty < Z_j < \infty \qquad (2.25)$$

La función generatriz de momentos de la variable Z_j, usando (2.17), es

$$M_{Z_j}(\theta) = E(e^{\theta Z_j}) = \int_{-\infty}^{\infty} e^{\theta Z_j} f(Z_j)\, dZ_j \qquad (2.26)$$

Reemplazando $f(Z_j)$ por su igual

$$M_{Z_j}(\theta) = \int_{-\infty}^{\infty} e^{\theta Z_j} \frac{1}{\sqrt{2\pi}} e^{\left[-\frac{1}{2}(Z_j)^2\right]} dZ_j = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{\theta Z_j} e^{\left(-\frac{Z_j^2}{2}\right)} dZ_j \qquad (2.27)$$

En la última expresión se tiene un producto de exponentes de igual base

$$e^{\theta Z_j} e^{\left(-\frac{Z_j^2}{2}\right)}$$

el resultado de este producto es conservar la base y sumar los exponentes

$$e^{\left(\theta Z_j - \frac{Z_j^2}{2}\right)}$$

A fin de encontrar una expresión menos compleja, se suma y resta $\frac{1}{2}\theta^2$ en el exponente de e

$$\theta Z_j - \frac{Z_j^2}{2} + \frac{1}{2}\theta^2 - \frac{1}{2}\theta^2$$

Al ordenar convenientemente y sacar factor común $-\frac{1}{2}$

$$-\frac{1}{2}(\theta^2 - 2\theta Z_j + Z_j^2) + \frac{1}{2}\theta^2$$

Se tiene el desarrollo del cuadrado

$$-\frac{1}{2}\left(Z_j-\theta\right)^2+\frac{1}{2}\theta^2$$

OBSERVACIÓN. Definir el binomio $\left(\theta-Z_j\right)^2$ es equivalente a $\left(Z_j-\theta\right)^2$. Es conveniente adoptar la segunda expresión para que, en la sustitución de variables que se realiza en (2.29) dZ_j sea positivo.

En síntesis:

$$e^{\left(\theta Z_j-\frac{Z_j^2}{2}\right)}=e^{-\frac{1}{2}\left(Z_j-\theta\right)^2+\frac{1}{2}\theta^2}=e^{-\frac{1}{2}\left(Z_j-\theta\right)^2}e^{\frac{1}{2}\theta^2} \qquad (2.28)$$

Reemplazando el resultado de (2.28) en (2.27) se obtiene una nueva expresión de la función generatriz de momentos de la variable Z_j en el parámetro θ

$$M_{Z_j}(\theta)=\frac{1}{\sqrt{2\pi}}\int_{-\infty}^{\infty}e^{-\frac{1}{2}\left(Z_j-\theta\right)^2}e^{\frac{1}{2}\theta^2}\,dZ_j=\frac{e^{\frac{1}{2}\theta^2}}{\sqrt{2\pi}}\int_{-\infty}^{\infty}e^{-\frac{1}{2}\left(Z_j-\theta\right)^2}dZ_j \qquad (2.29)$$

La diferencia entre (2.29) y (2.27) radica en la expresión a integrar, (2.29) es de sencilla resolución a través de sustitución de variables. Para lo cual se define $Z_j-\theta=t$

Entonces si $Z_j-\theta=t$

$$dZ_j=dt \qquad \text{(Porque } \theta \text{ es constante)}$$

Reemplazando en (2.29)

$$M_{Z_j}(\theta)=\frac{e^{\frac{1}{2}\theta^2}}{\sqrt{2\pi}}\int_{-\infty}^{\infty}e^{-\frac{1}{2}(t)^2}dt$$

Esta última expresión tiene la ventaja de contener una integral notable cuyo resultado es $\sqrt{2\pi}$

$$\int_{-\infty}^{\infty} e^{-\frac{1}{2}(t)^2} dt = \sqrt{2\pi} \tag{2.30}$$

Entonces

$$M_{Z_j}(\theta) = \frac{e^{\frac{1}{2}\theta^2}}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{-\frac{1}{2}(t)^2} dt = \frac{e^{\frac{1}{2}\theta^2}}{\sqrt{2\pi}} \sqrt{2\pi} = e^{\left(\frac{1}{2}\theta^2\right)} \tag{2.31}$$

En síntesis, la función generatriz de momentos de la variable Z_j, en el parámetro θ, es una función exponencial con exponente $\frac{\theta^2}{2}$

$$M_{Z_j}(\theta) = e^{\left(\frac{\theta^2}{2}\right)} \tag{2.32}$$

Al derivar $M_{Z_j}(\theta)$, se encuentran los momentos naturales de orden 1 y 2 que permiten hallar la media y la varianza de una variable aleatoria con distribución normal tipificada,

$$m_1' = \frac{\partial M_{Z_j}(\theta)}{\partial \theta} = \frac{\partial \left(e^{\frac{\theta^2}{2}}\right)}{\partial \theta} = e^{\frac{\theta^2}{2}} \theta \Big|_{\theta = 0} = 0 \quad \rightarrow E(Z_j) = 0$$

$$m_2' = \frac{\partial^2 M_{Z_j}(\theta)}{\partial \theta^2} = \frac{\partial \left(e^{\frac{\theta^2}{2}} \theta\right)}{\partial \theta} = e^{\frac{\theta^2}{2}} \theta\theta + e^{\frac{\theta^2}{2}} 1 \Big|_{\theta = 0} = 1$$

$$V(Z_j) = m_2' - (m_1')^2 = 1 - 0 = 1$$

Retomando la variable normal estándar definida en (2.23), si se multiplica en ambos miembros por σ_j, la diferencia entre la variable y su media es la variable estándar por el desvío, que define a la *variable aleatoria normal centrada*:

$$X_j - \mu_j = Z\sigma_j \qquad (2.33)$$

La función generatriz de la variable aleatoria normal centrada $X_j - \mu_j$, en el parámetro θ, es la función generatriz de momentos de la variable estándar por el desvío en el parámetro θ

$M_{X_j-\mu_j}(\theta) = M_{Z_j\sigma_j}(\theta)$ (donde σ_j es constante)

Si se tiene en cuenta lo visto en (2.20), donde $M_{cX_j}(\theta) = M_{X_j}(c\theta)$, la función generatriz de momentos de la variable centrada $X_j - \mu_j$ en el parámetro θ, es la función generatriz de momentos de la variable estándar por el desvío en el parámetro θ, que es igual a la función generatriz de momentos de la variable estándar Z_j en el parámetro $\sigma_j\theta$

$$M_{X_j-\mu_j}(\theta) = M_{Z_j\sigma_j}(\theta) = M_{Z_j}(\sigma_j\theta)$$

Por (2.32) se sabe que la función generatriz de momentos de la variable Z_j en el parámetro θ es $M_{Z_j}(\theta) = e^{\left(\frac{\theta^2}{2}\right)}$

Entonces

$$M_{Z_j}(\sigma_j\theta) = e^{\left(\frac{\theta\sigma_j{}^2}{2}\right)} = e^{\left[\frac{1}{2}(\theta\sigma_j)^2\right]}$$

Es decir, la función generatriz de momentos de la variable normal aleatoria centrada $X_j - \mu_j$ en el parámetro θ es

$$M_{X_j-\mu_j}(\theta) = e^{\left[\frac{1}{2}(\theta\sigma_j)^2\right]} \qquad (2.34)$$

La variable aleatoria Z_j fue reexpresada de acuerdo a (2.33), si se suma μ_j en ambos miembros se tiene la expresión de la variable aleatoria X_j

$$X_j = Z\sigma_j + \mu_j \qquad (2.35)$$

La función generatriz de momentos de una variable aleatoria en el parámetro θ, teniendo en cuenta (2.35), es la función generatriz de momentos de la suma de un producto (la variable Z_j ponderada por el desvío) y una constante, en el parámetro θ

$$M_{X_j}(\theta) = M_{Z_j\sigma_j+\mu_j}(\theta) \qquad (2.36)$$

Pero, teniendo en cuenta una propiedad de función generatriz de momentos

$$M_{c+X_j}(\theta) = e^{c\theta}M_{X_j}(\theta) \qquad (2.37)$$

y la ya expresada en (2.20), $M_{cX_j}(\theta) = M_{X_j}(c\,\theta)$, la expresión (2.36) es igual a

$$M_{X_j}(\theta) = M_{Z_j\sigma_j+\mu_j}(\theta) = e^{\mu_j\theta}M_{Z_j}(\sigma_j\theta) \qquad (2.38)$$

La función generatriz de momentos de la variable X_j en el parámetro θ es el producto de una constante, $e^{\mu_j\theta}$, y la función generatriz de momentos de una variable Z_j en el parámetro $\sigma_j\theta$

Con el resultado alcanzado en (2.32), se reexpresa $M_{Z_j}(\sigma_j\theta)$

$$M_{Z_j}(\sigma_j\theta) = e^{\frac{(\sigma_j\theta)^2}{2}}$$

y se tiene

$$M_{X_j}(\theta) = e^{\mu_j\theta+\frac{1}{2}\sigma_j^2\theta^2} \quad (2.39)$$

(2.39) es la función generatriz de momentos de la variable aleatoria X_j normal de media μ_j y desviación estándar σ_j.

CASOS DE ESTUDIO, PREGUNTAS Y PROBLEMAS

Caso 1: Tabla de datos

Diseñe la tabla de datos que se corresponda con la investigación planteada en el Caso 2 del Cuaderno Materiales y Métodos de la Econometría PIE 1. 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 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.

Caso 3: Situación socioeconómica en países desarrollados

Dada la descripción de 14 países por medio de 7 variables socioeconómicas dispuesta en una tabla de datos, donde las variables son:

X_1: Densidad de Población

X_2: Porcentaje de personas empleadas en la agricultura

X_3: Renta per cápita

X_4: Inversiones de rendimiento de capital en maquinarias

X_5: Tasa de mortalidad infantil

X_6: Consumo de energía por 100 habitantes

X_7: Aparatos de TV por 100 habitantes

Realice un gráfico de dispersión entre dos variables, ubique a cada uno de los países e interprete el resultado alcanzado.

Países	X_1	X_2	X_3	X_4	X_5	X_6	X_7
Australia	2	6	8.4	10.1	12	5.2	36
Francia	97	9	10.7	9.2	10	3.7	28
Alemania	247	6	12.4	9.1	15	4.6	33
Grecia	72	31	4.1	8.1	19	1.7	12
Islandia	2	13	11.0	6.6	11	5.8	25
Italia	189	15	5.7	7.9	15	2.5	22
Japón	311	11	8.7	10.9	8	3.3	24
N. Zelanda	12	10	6.8	8.0	14	3.4	26
Portugal	107	31	2.1	5.5	39	1.1	9
España	74	19	5.3	6.9	15	2.0	21
Suecia	18	6	12.8	7.2	7	6.3	37
Turquía	56	61	1.6	8.8	153	0.7	5
Reino Unido	229	3	7.2	9.3	13	3.9	39
Estados Unidos	24	4	10.6	7.3	13	8.7	62

Figura Tabla de datos para 14 países seleccionados

Por ejemplo, el gráfico siguiente muestra que a medida que crece la renta per cápita, crece la cantidad de TV por cada 100 habitantes. Cada uno de los puntos representa un país distinto.

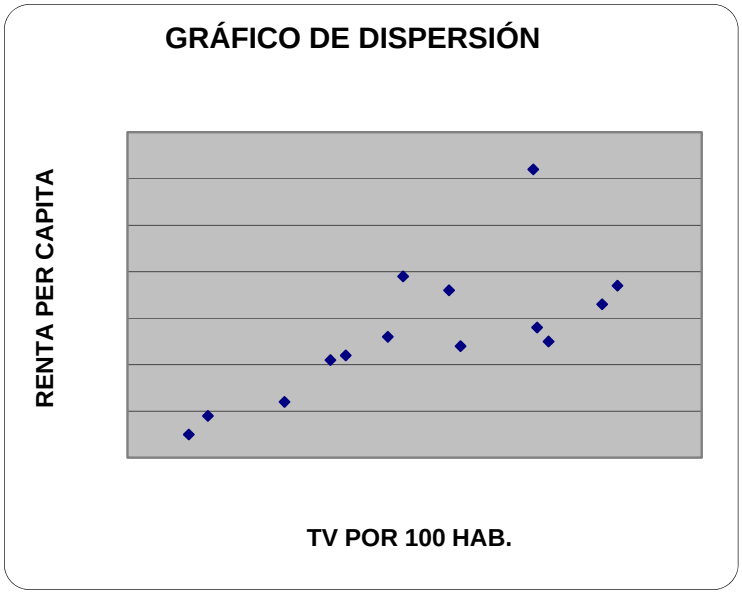


Figura . Gráfico de dispersión para 14 países en el plano

Caso 4: Modelo macroeconómico de demanda nacional

Dagum-Dagum (1971) adaptan el modelo de demanda nacional, de bienes alimenticios, especificado por M.A. Girshick y T. Haavelmo, pp. 36

(1) $C_t = \alpha_0 + \alpha_1 Y_t + \alpha_2 Y_{t-1} - \alpha_3 P_t + \alpha_4 t + \mu_{1t}; \quad 0 < \alpha_1 + \alpha_2 < 1; \quad \alpha_3 > 0$

(2) $S_t = \beta_0 + \beta_1 P_t + \beta_2 Z_t + \mu_{2t}; \qquad \beta_1; \beta_2 > 0$

(3) $C_t = S_t$

(4) $Y_t = L_0 + L_1 I_t + L_2 Y_{t-1} + \mu_{31t}; \qquad L_0 = \dfrac{\theta_0}{1-\theta_1}\ ; L_1 = \dfrac{1}{1-\theta_1}; \ L_2 = \dfrac{\theta_2}{1-\theta_1}$

(5) $C_t = O_0 + O_1 P_{Mt} + O_2 P_{M,t-1} + O_3 t + \mu_{4t}; \quad O_1 > 0$

(6) $P_{Mt} = w_0 + w_1 P_t + w_2 t + \mu_{5t}; \quad w_1 > 0$

donde:

C_t; consumo de productos alimenticios

Y_t; ingreso disponible

P_t; índice de precios al por menor de productos alimenticios

t; tiempo

S_t; oferta de productos alimenticios en el mercado al por menor

I_t; inversión neta

Z_t; producción de bienes alimenticios

P_{Mt}; índice de precios al por mayor

Identifique tipo de ecuaciones, tipos de variables y parámetros, e interprete el significado económico de estos últimos.

Caso 5. Cuestionario de Encuesta de Opinión a hogares

Lea detenidamente la encuesta Perfil del habitante de HHH, luego identifique las variables y clasifíquelas en cualitativas y cuantitativas

Suponga que se realizaron 387 encuestas, plantee una tablas de datos e indique cuáles son las unidades de observación y cuáles las variables.

Caso 6: Ingreso medio en los hogares

En el Cuadro 1 se clasifica a los hogares de una localidad de acuerdo al nivel de ingresos declarados a la Encuesta Permanente de Hogares en Mayo de 2003.

Cuadro 1 Ingreso de una localidad	
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

Con esta información, y teniendo en cuenta que la cantidad de hogares relevados fue de 636, calcule:

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

Preguntas

1. 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) comparar, con el método tradicional, los resultados de un nuevo método para entrenar vendedores de boletos de avión

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

2. Explique las diferencias entre

Variables aleatorias cualitativas y cuantitativas y de tres ejemplos de cada una de ellas,

Variables aleatorias discretas y continuas y de tres ejemplos de cada una de ellas.

3. Para los siguientes tipos de valores, identifique si corresponden a variables discretas (D), a variables continuas (C) o a variables cualitativas (Cu).

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) número de asistentes a la reunión anual de una compañía

m) vida 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

4. Explique las diferencias entre

Dato y unidades de observación

Distribuciones de frecuencias, distribuciones de frecuencias relativas y distribuciones de porcentajes.

Histogramas, polígonos y ojivas (polígono acumulado).

5. Dar la definición de los siguientes términos:

Variable aleatoria continua

Parámetros

Variable aleatoria discreta

Distribución Muestral

Encuesta

Distribución de Frecuencia

Experimento

Población

Polígono de Frecuencia

Distribución Teórica

Histograma

Variable Categórica

Muestra

Estimador

Distribución de Probabilidad

Problemas

1. El cuadro muestra, en intervalos de clase, los importes de alquiler cobrados por 200 departamentos

Cuadro. Departamentos según nivel de renta	
Renta Mensual (en miles de $)	Número de Departamentos
350-379,99	3
380-409,99	8
410-439,99	10
440-469,99	13
470-499,99	33
500-529,99	40
530-559,99	35
560-589,99	30
590-619,99	16
620-649,99	12
	Total 200

a. ¿Cuáles son los límites nominales inferior y superior de la primera clase?

b. ¿Cuáles son los límites exactos inferior y superior de la primera clase?

c. El intervalo de clase ¿es igual en todas las clases de la distribución?. ¿Cuál es el tamaño del intervalo?

d. ¿Cuál es el punto medio de la primera clase?

e. ¿Cuáles son los límites exactos inferior y superior de la clase en la que se tabuló la mayor cantidad de rentas de departamentos?

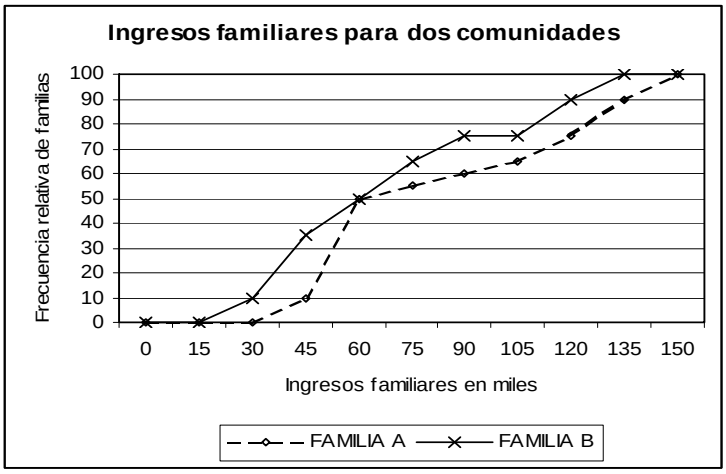
f. Si se reporta una renta de $439,50 identifique los límites nominales inferior y superior de la clase en la que se incluiría esta renta.

g. Construye un histograma, un polígono de frecuencias y una curva de frecuencias.

h. Describe la curva de frecuencias anterior desde el punto de vista de la asimetría.

i. Construye una distribución de frecuencias acumuladas y preséntala en forma gráfica mediante una ojiva.

2. La Figura contiene los polígonos de frecuencias relativas acumuladas de los ingresos familiares de dos muestras aleatorias (A y B) de 200 familias cada una, trazados en dos comunidades.



Con base en estos datos:

a) ¿Cuántas familias en la muestra A, tienen ingresos mayores a $120.000?

b) ¿Cuál es el porcentaje de familias en la muestra A, con ingresos menores de $90.000?

c) ¿Cuál muestra tiene un recorrido más amplio?

d) ¿Cuántas familias en la muestra B, tienen ingresos entre $90.000 y $105.000?

e) ¿Cuál de las muestras, muestra A o muestra B, tiene más ingresos familiares superiores a $60.000?

f) ¿Qué porcentaje de familias de la muestra A, ganan menos de $60.000?

g) ¿Qué porcentaje de familias de la muestra A, ganan más de $60.000?

h) ¿Qué muestra tiene más ingresos inferiores a $120.000?

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	Edad	SEXO		Último dígito de número cuestionario										Selección	
de Orden		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 Internet	3
4	Por Diarios Locales	4
5	Por Diarios Provinciales	5
6	Por Diarios Nacionales	6
7	Alternativamente por unos y otros	7
8	Por comentario en el ámbito de trabajo	8
9	Por comentario en otro ámbito (bar, etc)	9
10	Ns/Nc	10
11		11

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 naipe						
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
¿Por qué? _____(Ir a 9)
3. Ns/Nc

6. ¿Cuál es el diario que *lee más*? (E)
1. Local_____
2. Ámbito 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. más frecuentemente y en qué momento del día? (L)
Radio_____
Mañana Tarde Noche
(Puede marcar más de uno)

11. ¿Ud. mira televisión? (E)
1. SI
2. NO (IR A 16)
3. Ns/Nc

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

	Mañana	Tarde	Noche
1) América 2			
2) Canal 9			
3) Canal 13 Bs.As.			
4) Canal 11 Telefe			
5) Canal 7 TV pública			
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. Cablevisión
2. Multicanal
3. Telecentro
4. TV Pública Digital
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 ¿Por qué? _____
3. NS/Nc

21. ¿Cuál o cuáles tarjetas de crédito posee Ud.? (E)
1. No tengo tarjeta (IR a 36)
2. Argencard/Master
3. Visa/Visa Internacional
4. Diners
5. Bisel
6. Cabal
7. Naranja
8. Provencred
9. American Express
10. Otra. ¿Cuál? _____
11. Ns/Nc

22. Habitualmente ¿para qué 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
3. Ns/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
¿Por qué? _____
_____
(PASAR A 39)

36. (dijo no tener tarjeta) ¿Por qué 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 qué 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. ¿Qué plan de medicina prepaga tiene? (A)
_____

45. ¿Realiza Ud. aportes jubilatorios? (E)
1. Si
2. No
3. Ns/Nc

46. ¿Realiza aportes a cajas complementarias? (E)
1. Si
2. No
3. 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)
1. Propia (Seguir en 57)
2. Alquilada (Seguir en 53)
3. De familiares (Seguir en 55)
4. Otros (Seguir en 55)
5. NS/NC

53. ¿Cuánto paga de alquiler por mes? (E)
$_____
(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. ¿Qué tipo de vivienda prefiere Ud.

(Marcar lo que corresponda)		¿De cuántos dormitorios?
1. Casa		
2. Departamento		
3. 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
	Inactivo	
1	Desocupado	1
2	Estudiante	2
3	Ama de Casa	3
4	Jubilado/pensionado	4
	Autónomos	
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
	Empleadores	
11	1-5 empleados	11
12	6-20 empleados	12
13	más de 20 empleados	13
	Relación de dependencia	
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)

	SI	NO		SI	NO
Computadora			Gas por red		
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. Una última preguntita y terminamos, ¿cuántas personas viven en el hogar? ___
MUCHAS GRACIAS POR COLABORAR CON NOSOTROS

67. Antes de iniciar la próxima entrevista el encuestador deberá observar el nivel de vivienda (colocar número de nivel _____

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:
_____
_____
_____
_____
Referencias: A abierta, E espontánea, L leer, T tarjeta

Tabla de Contenido

Referencias

Bibliografía

- Baronio, A. M. (2001). *Metodología estadística para el estudio socioeconómico de regiones.* Tesis Doctorado en Ciencias Económicas, Universidad Nacional de Río Cuarto, Secretaría de Posgrado, Río Cuarto.
- Berenson, M., & Levine, D. M. (1996). *Estadística Básica en Administración.* México: Prentice Hall.
- Dagum, C., & Bee de Dagum, E. M. (1971). *Introducción a la econometría (10 ed.).* México: Editorial Siglo XXI.
- Davenport, T., & Prusak, L. (1999). *Conocimiento en Acción. Cómo las organizaciones manejan lo que saben.* México: Prentice Hall.
- Dixon, W. J., & Massey, F. J. (1957). *Introduction to Statistical Analysis.* Nueva York: McGraw-Hill.
- Freeman, H. (1963). *Introducción a la Inferencia Estadística.* México: Trillas.
- Gomez Villegas, M. A. (2005). *Inferencia Estadística.* España: Ediciones Díaz de Santo.
- Hernández Sampieri, R., Fernández Collado, C., & Baptista Lucio, M. d. (2010). *Metodología de la Investigación (Quinta ed.).* México: McGraw-Hill/ Interamericana editores, S.A. de C.V.
- Kazmier, L., & Díaz Mata, A. (1993). *Estaística aplicada a la Administración y a la Economía.* México: Mc. Graw Hill.
- Kinnear, T., & Taylor, J. (1993). *Investigación de Mercado.* Mc. Graw Hill.
- Mao, J. C. (1980). *Análisis Financiero (Cuarta ed.).* Buenos Aires: El Ateneo.
- Meyer, P. L. (1973). *Probabilidad y aplicaciones estadísticas.* México: Fondo Educativo Interamericano SA.
- Navarro, A. M. (1997). *Reflexiones sobre la relación entre economía, econometría y epistemología.* Anales.
- Padua, J. (1996). *Técnicas de Investigación Aplicadas a las Ciencias Sociales (Septima reinpresión).* México: Fondo de Cultura Económica.

- Pugachev, V. S. (1973). *Introducción a la Teoría de las Probabilidades.* Moscú: Editorial Mir.
- Pulido San Román, A. (1993). *25 años de experiencia en econometría aplicada.* (Vol. 0). Valladolid: Estudios de Economía Aplicada. .
- Pulido San Román, A. (1993). *Modelos econométricos.* Madrid: Editorial Pirámide.
- Sabino, C. (1996). *El proceso de investigación (Segunda ed.).*
- Spiegel, M. R. (1976). *Teoría y problemas de probabilidad y estadística.* México: Mc. Graw Hill.
- Tramutola, C. D. (1971). *Modelos probabilísticos y decisiones financieras.* Capital Federal: E.C.Moderan.
- Wackerly, D., Mendenhall, W., & Scheaffer, R. (2008). *Estadística Matemática con Aplicaciones (Séptima edición).* México: Cencage Learning.