

Naturaleza del análisis de regresión

Como se mencionó en la introducción, la regresión es una herramienta fundamental de la econometría; en este capítulo se considera muy brevemente la naturaleza de este instrumento.

1.1 Origen histórico del término *regresión*

Francis Galton acuñó el término *regresión*. En un famoso ensayo, Galton planteó que, a pesar de la tendencia de los padres de estatura alta a procrear hijos altos y los padres de estatura baja, hijos bajos, la estatura promedio de los niños de padres de una estatura determinada tendía a desplazarse, o “regresar”, a la estatura promedio de la población total.¹ En otras palabras, la estatura de los hijos de padres inusualmente altos o inusualmente bajos tiende a dirigirse a la estatura promedio de la población. La *ley de regresión universal* de Galton fue confirmada por su amigo Karl Pearson, quien reunió más de mil registros de estaturas de miembros de grupos familiares.² Pearson descubrió que la estatura promedio de los hijos de un grupo de padres de estatura alta era menor que la estatura de sus padres, y que la estatura promedio de los hijos de un grupo de padres de estatura baja era mayor que la estatura de sus padres; es decir, se trata de un fenómeno mediante el cual los hijos altos e hijos bajos “regresan” por igual a la estatura promedio de todos los demás. En palabras de Galton, se trata de una “regresión a la mediocridad”.

1.2 Interpretación moderna de la regresión

La interpretación moderna de la regresión es, sin embargo, muy diferente. En términos generales, se afirma que:

El análisis de regresión trata del estudio de la dependencia de una variable (*variable dependiente*) respecto de una o más variables (*variables explicativas*) con el objetivo de estimar o predecir la media o valor promedio poblacional de la primera en términos de los valores conocidos o fijos (en muestras repetidas) de las segundas.

¹ Francis Galton, “Family Likeness in Stature”, *Proceedings of Royal Society*, Londres, vol. 40, 1886, pp. 42-72.

² K. Pearson y A. Lee, “On the Laws of Inheritance”, *Biometrika*, vol. 2, noviembre de 1903, pp. 357-462.

La importancia trascendental de este enfoque del análisis de regresión se verá claramente sobre la marcha, pero algunos ejemplos sencillos aclararán este concepto básico.

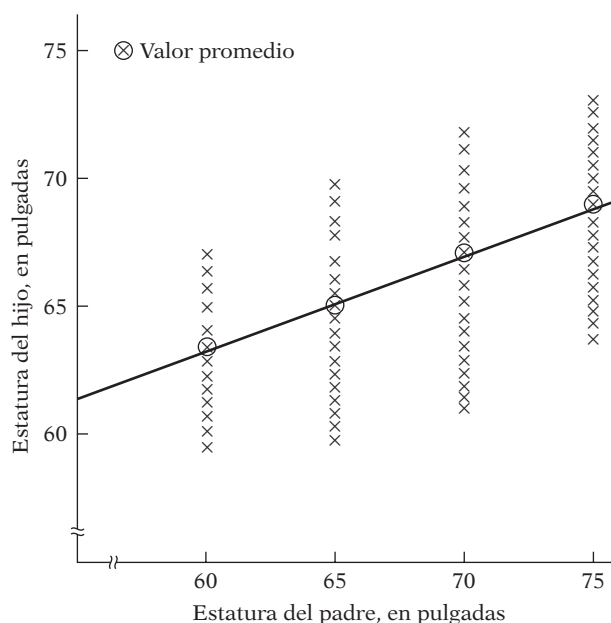
Ejemplos

1. Considere de nuevo la ley de regresión universal de Galton. A él le interesaba averiguar las razones de la estabilidad en la distribución de estaturas dentro de una población. En el enfoque moderno, la preocupación no es esta explicación, sino averiguar cómo cambia la estatura *promedio* de los hijos dada la estatura de los padres. En otras palabras, lo que interesa es predecir la estatura promedio de los hijos a partir de la estatura de sus padres. Para ver cómo hacerlo, considere la figura 1.1, que corresponde a un **diagrama de dispersión**. La figura muestra la distribución de las estaturas de los hijos en una población hipotética, correspondiente al conjunto de valores dados o *fijos* de las estaturas de los padres. Observe que, para cualquier estatura de un padre, existe un *rango* (distribución) de estaturas de los hijos. Sin embargo, observe también que, a pesar de la variabilidad de la estatura de los hijos conforme al valor de la estatura de los padres, la estatura promedio de los hijos aumenta, por lo general, en la medida en que lo hace la estatura de los padres. Para demostrar esto con claridad, las cruces dentro de los círculos en la figura indican la estatura *promedio* de los hijos que corresponde a una estatura determinada de los padres. Estos promedios se conectan para obtener la línea recta de la figura. Esta línea, como veremos, se conoce como **recta de regresión**. Dicha recta muestra que el *promedio* de la estatura de los hijos aumenta conforme crece la de los padres.³

2. Considere el diagrama de dispersión en la figura 1.2, que presenta la distribución de una población hipotética de estaturas de niños en edades *fijas*. Observe que existe un rango (distribución) de estaturas correspondiente a cada edad. Es obvia la improbabilidad de que todos los niños de una edad determinada tengan estaturas idénticas. Pero, *en promedio*, la estatura se incrementa con la edad (por supuesto, hasta cierta edad), que se ve con claridad al trazar una recta (la recta de

FIGURA 1.1

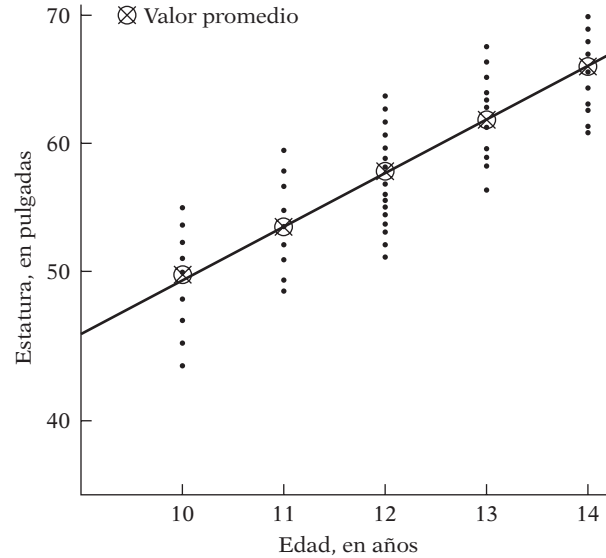
Distribución hipotética de las estaturas de los hijos correspondientes a las estaturas de los padres.



³ En esta etapa de estudio del tema, denominaremos a esta recta de regresión simplemente *recta que conecta el valor de la media, o promedio, de la variable dependiente (la estatura de los hijos) que corresponde a un valor dado de la variable explicativa (la estatura de los padres)*. Observe que esta recta tiene una pendiente positiva; pero la pendiente es menor que 1, lo cual está de acuerdo con el concepto de Galton de regresión a la mediocridad. (¿Por qué?)

FIGURA 1.2

Distribución hipotética de estaturas correspondientes a edades seleccionadas.



regresión) por los puntos dentro de los círculos, los cuales representan la estatura promedio de determinadas edades. Por consiguiente, si se conoce la edad, se predice la estatura promedio de dicha edad mediante la recta de regresión.

3. Al considerar lo referente a la economía, a un economista quizá le interese estudiar la dependencia del consumo personal respecto del ingreso personal neto disponible (después de impuestos). Con un análisis de este tipo se calcula la propensión marginal a consumir (PMC), es decir, el cambio promedio del consumo ante un cambio, digamos, de un dólar en el ingreso real (ver la figura 1.3).

4. Un monopolista que puede fijar el precio o la producción (pero no ambos factores) tal vez desee conocer la demanda de un producto con diversos precios. Tal experimento permite estimar la **elasticidad del precio** (es decir, la respuesta a variaciones del precio) de la demanda del producto y permite determinar el precio que maximiza las ganancias.

5. Un economista laboral quizá desee estudiar la tasa de cambio de los salarios monetarios o nominales en relación con la tasa de desempleo. Las cifras históricas aparecen en el diagrama de dispersión de la figura 1.3. La curva de esta figura es un ejemplo de la célebre *curva de Phillips*, que relaciona los cambios en los salarios nominales con la tasa de desempleo. Un diagrama de dispersión de este tipo permite al economista laboral predecir el cambio promedio en los salarios nominales con una cierta tasa de desempleo. Tal conocimiento sirve para establecer supuestos sobre el proceso inflacionario en una economía, pues es probable que los incrementos en los salarios monetarios se reflejen en incrementos de precios.

6. En la economía monetaria se sabe que, si se mantienen constantes otros factores, cuanto mayor sea la tasa de inflación π , menor será la proporción k del ingreso que la gente deseará mantener en forma de dinero, como se deduce de la figura 1.4. La pendiente de esta recta representa el cambio en k con un cambio en la tasa de inflación. Un análisis cuantitativo de esta relación permite al economista predecir la cantidad de dinero, como proporción del ingreso, que la gente deseará mantener con diversas tasas de inflación.

7. El director de marketing de una compañía tal vez quiera conocer la relación entre la demanda del producto de su compañía con el gasto de publicidad, por ejemplo. Un estudio de este tipo es de gran ayuda para encontrar la **elasticidad de la demanda** respecto de los gastos publicitarios, es decir, el cambio porcentual de la demanda en respuesta a un cambio de 1 por ciento, por ejemplo, en el presupuesto de publicidad. Saber esto sirve para determinar el presupuesto “óptimo” de publicidad.

FIGURA 1.3
Curva hipotética de Phillips.

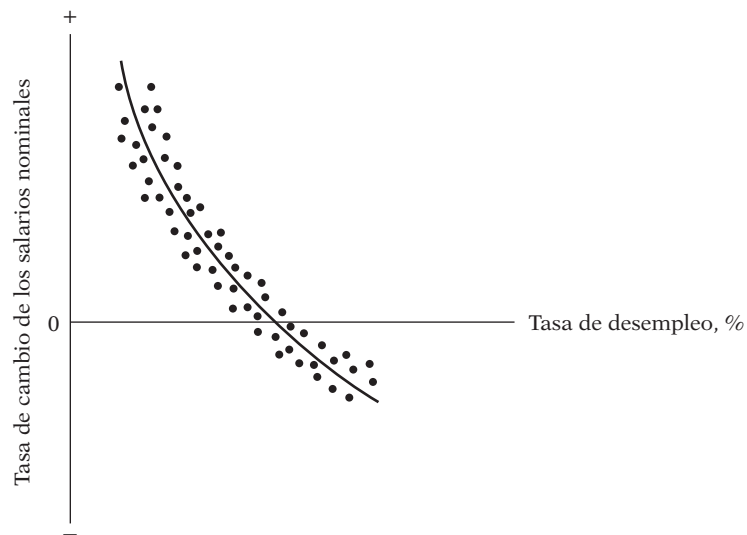
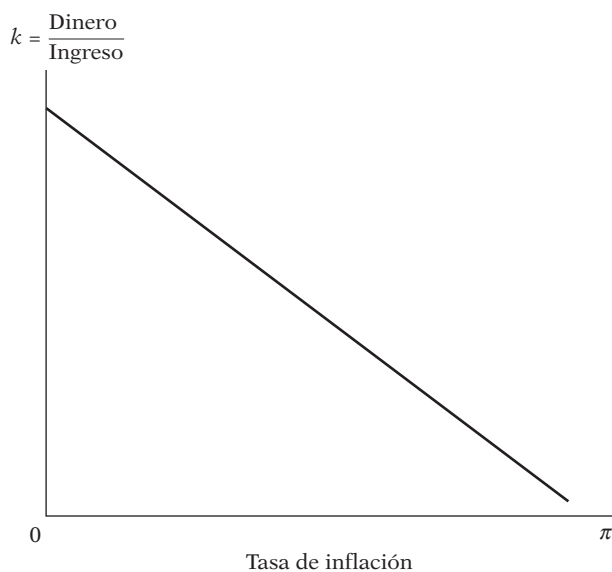


FIGURA 1.4
Tenencia de dinero en relación con la tasa de inflación π .



8. Por último, un agrónomo tal vez se interese en estudiar la relación entre el rendimiento de un cultivo, digamos de trigo, y la temperatura, lluvia, cantidad de sol y fertilizantes. Un análisis de dependencia de ese tipo facilitaría la predicción o el pronóstico del rendimiento medio del cultivo según la información sobre las variables explicativas.

El lector puede proporcionar una amplia gama de ejemplos similares de la dependencia de una variable respecto de otra o más variables. Las técnicas del análisis de regresión que se explican en este texto están diseñadas especialmente para estudiar dicha dependencia entre variables.

1.3 Relaciones estadísticas y relaciones deterministas

En los ejemplos de la sección 1.2 se observa que en el análisis de regresión interesa lo que se conoce como dependencia *estadística* entre variables, no así la *funcional* o *determinista*, propia de la física clásica. En las relaciones estadísticas entre variables se analizan, en esencia, variables **aleatorias** o **estocásticas**,⁴ es decir, variables con distribuciones de probabilidad. Por otra parte, en la dependencia funcional o determinista también se manejan variables, pero no son aleatorias o estocásticas.

Por ejemplo, el rendimiento de un cultivo depende de la temperatura, lluvia, Sol y fertilizantes, y dicha dependencia es de naturaleza estadística porque las variables explicativas, si bien son importantes, no permiten al agrónomo predecir en forma exacta el rendimiento del cultivo debido a los errores propios de la medición de estas variables y a otra serie de factores (variables) que en conjunto afectan el rendimiento pero son difíciles de identificar individualmente. De esta manera, habrá alguna variabilidad “intrínseca” o aleatoria en la variable dependiente, el rendimiento del cultivo, que no puede explicarse en su totalidad sin importar cuántas variables explicativas se consideren.

Los fenómenos deterministas, por otra parte, implican relaciones como la ley de la gravedad de Newton, la cual establece que toda partícula en el universo atrae a cualquier otra partícula con una fuerza directamente proporcional al producto de sus masas e inversamente proporcional al cuadrado de la distancia entre ellas. En términos matemáticos, $F = k(m_1 m_2 / r^2)$, donde F = fuerza, m_1 y m_2 son las masas de las dos partículas, r = distancia y k = constante de proporcionalidad. Otro ejemplo es el de la ley de Ohm, la cual postula que para conductores metálicos dentro de un intervalo limitado de temperatura, la corriente C es proporcional al voltaje V ; es decir, $C = (\frac{1}{k})V$, donde $\frac{1}{k}$ es la constante de proporcionalidad. Otros ejemplos de relaciones deterministas son la ley de los gases de Boyle, la ley de la electricidad de Kirchhoff y la ley del movimiento de Newton.

En este texto no interesan tales relaciones deterministas. Por supuesto, de haber errores de medición, por ejemplo, en la k de la ley de la gravedad de Newton, la relación que de otra forma habría sido determinista se convierte en una relación estadística. En esta situación es posible predecir la fuerza en forma aproximada sólo a partir de un valor dado de k (y m_1 , m_2 y r), el cual contiene errores. La variable F se convierte en este caso en aleatoria.

1.4 Regresión y causalidad

A pesar de que el análisis de regresión tiene que ver con la dependencia de una variable respecto de otras variables, esto no implica causalidad necesariamente. En palabras de Kendall y Stuart: “Una relación estadística, por más fuerte y sugerente que sea, nunca podrá establecer una conexión causal: nuestras ideas de causalidad deben provenir de estadísticas externas y, en último término, de una u otra teoría”.⁵

⁴ La palabra *estocástico* viene de la voz griega *stokhos*, que significa “centro del blanco”. El resultado de lanzar dardos sobre un tablero es un proceso estocástico, es decir, un proceso lleno de tiros fallidos.

⁵ M. G. Kendall y A. Stuart, *The Advanced Theory of Statistics*, Charles Griffin Publishers, Nueva York, 1961, vol. 2, cap. 26, p. 279.

En el ejemplo del rendimiento del cultivo citado, no hay una *razón estadística* para suponer que la lluvia no depende del rendimiento del cultivo. Considerar que el rendimiento del cultivo depende de la lluvia (entre otras cosas) se debe a cuestiones no estadísticas: el sentido común indica que la relación no puede ser a la inversa, pues no es posible controlar la lluvia mediante el rendimiento del cultivo.

En todos los ejemplos de la sección 1.2, lo que se debe notar es que **una relación estadística por sí misma no puede, por lógica, implicar causalidad**. Para aducir causalidad se debe acudir a consideraciones *a priori* o teóricas. Así, en el tercer ejemplo, es posible recurrir a la teoría económica para afirmar que el consumo depende del ingreso real.⁶

1.5 Regresión y correlación

El **análisis de correlación** se relaciona de manera estrecha con el de regresión, aunque conceptualmente los dos son muy diferentes. En el análisis de correlación, el objetivo principal es medir la *fuerza* o el *grado de asociación lineal* entre dos variables. El **coeficiente de correlación**, que veremos en detalle en el capítulo 3, mide esta fuerza de asociación (lineal): por ejemplo, si se desea encontrar la correlación (coeficiente) entre el hábito de fumar y el cáncer del pulmón; entre las calificaciones en exámenes de estadística y en exámenes de matemáticas; entre las calificaciones de bachillerato y de la universidad, y así sucesivamente. En el análisis de regresión, como ya mencionamos, no interesa ese tipo de medición. En cambio, se trata de estimar o predecir el valor promedio de una variable con base en los valores fijos de otras. Así, quizá se desee predecir el promedio de las calificaciones en un examen de estadística a partir de la calificación de un estudiante en un examen de matemáticas.

La regresión y la correlación presentan diferencias fundamentales que vale la pena mencionar. En el análisis de regresión hay una asimetría en el tratamiento a las variables dependientes y explicativas. Se supone que la variable dependiente es estadística, aleatoria o estocástica, es decir, que tiene una distribución de probabilidad. Por otra parte, se asume que las variables explicativas tienen valores fijos (en muestras repetidas),⁷ lo cual es explícito en la definición de regresión de la sección 1.2. Así, en la figura 1.2 se supuso que la variable de edad era fija en los niveles dados y se obtuvieron medidas de estatura en esos niveles. En el análisis de correlación, por otra parte, se tratan dos variables cualesquiera en forma simétrica; no hay distinción entre las variables dependiente y explicativa. Después de todo, la correlación entre las calificaciones de los exámenes de matemáticas y de estadística es la misma que la existente entre calificaciones de exámenes de estadística y de matemáticas. Además, las dos variables se consideran aleatorias. Como veremos, la mayor parte de la teoría de correlación parte del supuesto de aleatoriedad de las variables, mientras que la mayor parte de la teoría de regresión que expondremos en este texto está condicionada al supuesto de que la variable dependiente es estocástica y que las variables explicativas son fijas o no estocásticas.⁸

⁶ Pero, como veremos en el capítulo 3, el análisis clásico de regresión se basa en el supuesto de que el modelo del análisis es el correcto. Por consiguiente, la dirección de la causalidad puede estar implícita en el modelo postulado.

⁷ Es de crucial importancia notar que las variables explicativas pueden ser intrínsecamente estocásticas, pero, para fines del análisis de regresión, suponemos que sus valores son fijos en el muestreo repetido (es decir, que X toma los mismos valores en diversas muestras), de modo que, en efecto, no resultan aleatorias ni estocásticas. Hay más sobre este tema en la sección 3.2 del capítulo 3.

⁸ En el tratamiento avanzado de econometría se puede ser más flexible con el supuesto de que las variables explicativas son no estocásticas (ver la introducción de la segunda parte).

1.6 Terminología y notación

Antes de proceder al análisis formal de la teoría de regresión abordaremos brevemente la terminología y la notación. En las publicaciones especializadas, los términos *variable dependiente* y *variable explicativa* se definen de varias maneras; a continuación se presenta una lista representativa:

Variable dependiente	Variable explicativa
↕	↕
Variable explicada	Variable independiente
↕	↕
Predicha	Predictora
↕	↕
Regresada	Regresora
↕	↕
Respuesta	Estímulo
↕	↕
Endógena	Exógena
↕	↕
Resultado	Covariante
↕	↕
Variable controlada	Variable de control

Aunque es cuestión de preferencia personal y tradición, en este texto se utiliza la terminología de variable dependiente/variable explicativa, o la más neutral de regresada y regresora.

Si se estudia la dependencia de una variable respecto de una única variable explicativa, como el consumo que depende del ingreso real, dicho estudio se conoce como **análisis de regresión simple**, o **con dos variables**. Sin embargo, si se estudia la dependencia de una variable respecto de más de una variable explicativa, como el rendimiento de un cultivo, la lluvia, la temperatura, el Sol y los fertilizantes, se trata de un **análisis de regresión múltiple**. En otras palabras, en una regresión de dos variables sólo hay una variable explicativa, mientras que en la regresión múltiple hay más de una variable explicativa.

El término **aleatorio** es sinónimo de **estocástico**. Como ya vimos, una variable aleatoria o estocástica es la que toma cualquier conjunto de valores, positivos o negativos, con una probabilidad dada.⁹

A menos que se indique lo contrario, la letra Y representa la variable dependiente, y las X (X_1, X_2, \dots, X_k), las variables explicativas, con X_k como la k -ésima variable explicativa. Los subíndices i o t denotan la observación o valor i -ésimo o t -ésimo. X_{ki} (o X_{kt}) denota la i -ésima (o la t -ésima) observación de la variable X_k . N (o T) representa el número total de observaciones o valores en la población, y n (o t), el número total de observaciones en una muestra. Por convención, se utiliza el subíndice de observación i para los **datos transversales** (es decir, información recopilada en un momento determinado), y el subíndice t , para **datos de series de tiempo** (es decir, información reunida a lo largo de un periodo). La naturaleza de datos transversales y de series de tiempo, así como el importante tema de la naturaleza y las fuentes de datos para el análisis empírico, se estudian en la siguiente sección.

⁹ Hay una definición formal y más detalles en el **apéndice A**.

1.7 Naturaleza y fuentes de datos para el análisis económico¹⁰

El éxito de todo análisis econométrico depende a final de cuentas de la disponibilidad de los datos recopilados. Por consiguiente, es muy importante dedicar algún tiempo a estudiar la naturaleza, las fuentes y las limitaciones de los datos para el análisis empírico.

Tipos de datos

Hay tres tipos de datos disponibles para el análisis empírico: **series de tiempo**, **series transversales** e **información combinada** (combinación de series de tiempo y transversales).

Datos de series de tiempo

Los datos de la tabla 1.1 son un ejemplo de datos de series de tiempo. Una *serie de tiempo* es un conjunto de observaciones sobre los valores de una variable en diferentes momentos. Tal información debe recopilarse en intervalos regulares, es decir, en forma **diaria** (precios de acciones, informes del tiempo, etc.), **semanal** (como cifras de oferta monetaria), **mensual** (tasa de desempleo, Índice de Precios al Consumidor [IPC], etc.), **trimestral** (como el PIB), **anual** (como los presupuestos del gobierno), **quinquenal** (como el censo de la industria manufacturera), o **decenal** (como los censos de población). Algunas veces los datos están disponibles por trimestre y por año, como los datos del PIB y del consumo. Con las computadoras de alta velocidad, ahora se recopilan datos en intervalos muy breves, por ejemplo, precios de acciones, que se obtienen literalmente de manera continua (o *cotización en tiempo real*).

Si bien los datos de series de tiempo se utilizan mucho en estudios econométricos, presentan algunos problemas especiales para los econométricos. Como veremos en los capítulos sobre **econometría de series de tiempo**, la mayor parte del trabajo empírico con datos de series de tiempo supone que éstas son **estacionarias**. Aunque es muy pronto para introducir el significado técnico preciso de estacionariedad, *en términos generales, una serie de tiempo es estacionaria si su media y varianza no varían sistemáticamente con el tiempo*. Para entender esto, observe, en la figura 1.5, el comportamiento de la oferta de dinero M1 en Estados Unidos durante el periodo del primero de enero de 1959 a septiembre de 1999. (Los datos reales se proporcionan en el ejercicio 1.4.) Como se observa, la oferta de dinero M1 presenta una **tendencia** ascendente constante, así como variabilidad con el transcurso de los años, lo cual indica que la serie de tiempo M1 no es estacionaria.¹¹ En el capítulo 21 se analiza a fondo este tema.

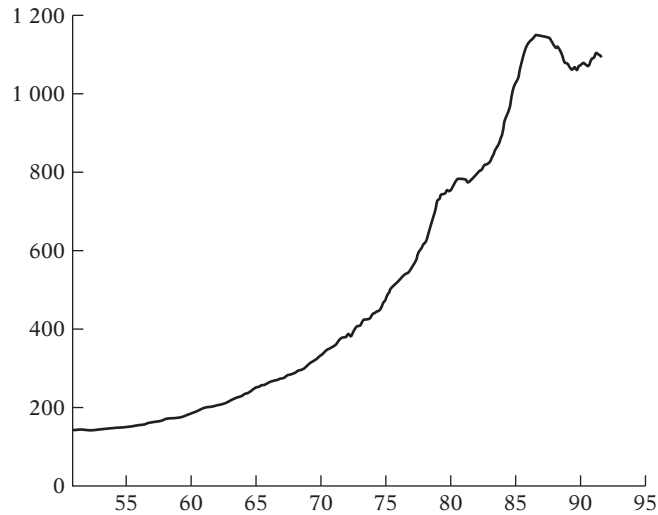
Datos transversales

Los datos transversales consisten en datos de una o más variables recopilados *en el mismo punto del tiempo*, como el censo de población realizado por la Oficina del Censo de Estados Unidos cada 10 años (el último fue en 2000), las encuestas de gastos del consumidor levantadas por la Universidad de Michigan y, sin duda, las encuestas de opinión de Gallup y diversas empresas especializadas. Un ejemplo concreto de datos transversales se presenta en la tabla 1.1, con datos sobre la producción y precios del huevo en Estados Unidos para los 50 estados durante 1990 y

¹⁰ En Michael D. Intriligator, *Econometric Models, Techniques, and Applications*, Prentice Hall, Englewood Cliffs, Nueva Jersey, 1978, cap. 3, hay una explicación ilustrativa.

¹¹ Para que resulte más claro, los datos se dividen en cuatro periodos: enero de 1951 a diciembre de 1962; enero de 1963 a diciembre de 1974; enero de 1975 a diciembre de 1986; y enero de 1987 a septiembre de 1999. En estos subperiodos, los valores de la media para la oferta de dinero (con sus correspondientes desviaciones estándar, entre paréntesis) fueron, respectivamente: 165.88 (23.27), 323.20 (72.66), 788.12 (195.43) y 1 099 (27.84); todas las cifras representan miles de millones de dólares. Ésta es una indicación aproximada de que la oferta de dinero no fue estacionaria a lo largo de todo el periodo.

FIGURA 1.5
Oferta de dinero M1:
Estados Unidos, enero de
1951-septiembre de 1999.



1991. Para cada año, los datos sobre los 50 estados son transversales. Así, en la tabla 1.1 aparecen dos muestras de corte transversal.

Así como los datos de series de tiempo crean problemas especiales (por la estacionariedad), los datos transversales también tienen sus propios problemas, en concreto, el de la *heterogeneidad*. En los datos de la tabla 1.1 se observa que hay algunos estados que producen grandes cantidades de huevo (como Pensilvania) y otros que producen muy poco (por ejemplo, Alaska). Cuando se incluyen unidades heterogéneas en un análisis estadístico, debe tenerse presente el **efecto de tamaño o de escala** con el fin de no mezclar manzanas con naranjas. Para ver esto con claridad, en la figura 1.6 se representan gráficamente los datos sobre la producción y los precios del huevo en los 50 estados de Estados Unidos en 1990. Esta figura muestra la amplia dispersión de las observaciones. En el capítulo 11 veremos que el efecto de escala puede ser importante al evaluar las relaciones entre variables económicas.

Datos combinados

Los datos combinados reúnen elementos de series de tiempo y transversales. Los datos de la tabla 1.1 son datos combinados. Hay 50 observaciones transversales por año, y dos observaciones de series de tiempo sobre precios y producción de huevo por estado: un total de 100 observaciones combinadas. De igual forma, los datos del ejercicio 1.1 son combinados, pues el índice de precios al consumidor de cada país de 1980 a 2005 representa datos de series de tiempo, en tanto que los datos del IPC de los siete países correspondientes a un solo año son transversales. Los datos combinados consisten en 182 observaciones: 26 observaciones anuales para cada uno de los siete países.

Datos en panel, longitudinales o en micropanel

Hay un tipo especial de datos combinados en el cual se estudia a través del tiempo la *misma* unidad transversal (por ejemplo, una familia o una empresa). Por ejemplo, el Departamento de Comercio de Estados Unidos realiza un censo de vivienda en intervalos periódicos. En cada encuesta periódica se entrevista a la misma unidad familiar (o a la gente que vive en la misma dirección) para averiguar si ha habido algún cambio en las condiciones de vivienda o financieras de esa unidad familiar desde la última encuesta. Los datos en panel que se obtienen de las entrevistas periódicas de la misma unidad familiar proporcionan información muy útil sobre la dinámica del comportamiento de las unidades familiares, como veremos en el capítulo 16.

FIGURA 1.6
Relación entre la producción y los precios del huevo, 1990.

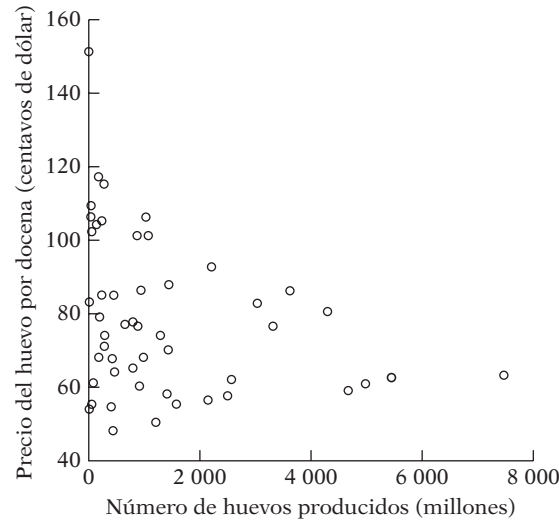


TABLA 1.1 Producción de huevo en Estados Unidos

Estado	Y_1	Y_2	X_1	X_2	Estado	Y_1	Y_2	X_1	X_2
AL	2 206	2 186	92.7	91.4	MT	172	164	68.0	66.0
AK	0.7	0.7	151.0	149.0	NE	1 202	1 400	50.3	48.9
AZ	73	74	61.0	56.0	NV	2.2	1.8	53.9	52.7
AR	3 620	3 737	86.3	91.8	NH	43	49	109.0	104.0
CA	7 472	7 444	63.4	58.4	NJ	442	491	85.0	83.0
CO	788	873	77.8	73.0	NM	283	302	74.0	70.0
CT	1 029	948	106.0	104.0	NY	975	987	68.1	64.0
DE	168	164	117.0	113.0	NC	3 033	3 045	82.8	78.7
FL	2 586	2 537	62.0	57.2	ND	51	45	55.2	48.0
GA	4 302	4 301	80.6	80.8	OH	4 667	4 637	59.1	54.7
HI	227.5	224.5	85.0	85.5	OK	869	830	101.0	100.0
ID	187	203	79.1	72.9	OR	652	686	77.0	74.6
IL	793	809	65.0	70.5	PA	4 976	5 130	61.0	52.0
IN	5 445	5 290	62.7	60.1	RI	53	50	102.0	99.0
IA	2 151	2 247	56.5	53.0	SC	1 422	1 420	70.1	65.9
KS	404	389	54.5	47.8	SD	435	602	48.0	45.8
KY	412	483	67.7	73.5	TN	277	279	71.0	80.7
LA	273	254	115.0	115.0	TX	3 317	3 356	76.7	72.6
ME	1 069	1 070	101.0	97.0	UT	456	486	64.0	59.0
MD	885	898	76.6	75.4	VT	31	30	106.0	102.0
MA	235	237	105.0	102.0	VA	943	988	86.3	81.2
MI	1 406	1 396	58.0	53.8	WA	1 287	1 313	74.1	71.5
MN	2 499	2 697	57.7	54.0	WV	136	174	104.0	109.0
MS	1 434	1 468	87.8	86.7	WI	910	873	60.1	54.0
MO	1 580	1 622	55.4	51.5	WY	1.7	1.7	83.0	83.0

Notas: Y_1 = huevos producidos en 1990 (millones).
 Y_2 = huevos producidos en 1991 (millones).
 X_1 = precio por docena (centavos de dólar) en 1990.
 X_2 = precio por docena (centavos de dólar) en 1991.
Fuente: *World Almanac*, 1993, p. 119. Los datos provienen del Economic Research Service, U.S. Department of Agriculture.

Como ejemplo concreto considere los datos proporcionados en la tabla 1.2. Los datos de la tabla, recopilados originalmente por Y. Grunfeld, se refieren a la inversión real, al valor real de la empresa y al capital accionario real de cuatro empresas estadounidenses, a saber, General Electric (GM), U.S. Steel (US), General Motors (GM) y Westinghouse (WEST), de 1935 a 1954.¹² En virtud de que los datos corresponden a varias empresas y se recopilaron a lo largo de varios años, se trata de un ejemplo clásico de datos en panel. En esta tabla, el número de observaciones de cada empresa es el mismo, pero no siempre ocurre así. Si todas las empresas tienen el mismo número de observaciones, se tiene lo que se conoce como **panel balanceado**. Si el número de observaciones no es igual para cada compañía, se llama **panel desbalanceado**. En el capítulo 16, Modelos de regresión con datos en panel, examinaremos estos datos y cómo estimar estos modelos.

El propósito de Grunfeld cuando recopiló estos datos fue investigar cómo depende la inversión bruta real (I) del valor real de la empresa (F) un año antes y del capital accionario real (C) un año antes. Como las compañías de esta muestra operan en el mismo mercado de capital, Grunfeld las estudió en conjunto para averiguar si tenían funciones de inversión parecidas.

Fuentes de datos¹³

Los datos para el análisis empírico pueden provenir de una dependencia gubernamental (por ejemplo, el Departamento de Comercio), un organismo internacional (el Fondo Monetario Internacional [FMI] o el Banco Mundial), una organización privada (por ejemplo, Standard & Poor's) o un particular. Hay miles de agencias de este tipo que recopilan datos para uno u otro fin.

Internet

Internet revolucionó la labor de recopilación de datos. Si uno “navega” por la red en los motores de búsqueda con sólo una palabra o frase (por ejemplo, tipos de cambio), se verá inundado con todo tipo de fuentes de datos. En el **apéndice E** se mencionan algunos sitios Web que suministran todo tipo de información financiera y económica, y que se visitan con mayor frecuencia. La mayoría de los datos se descarga con un costo mínimo. Conviene incluir en la lista de Favoritos, los sitios Web que brinden datos económicos útiles.

Los datos recopilados por estas organizaciones pueden ser de naturaleza **experimental** o **no experimental**. En los datos experimentales, frecuentes en las ciencias naturales, el investigador suele recabar los datos con algunos factores constantes, con el fin de evaluar el efecto de otros en un fenómeno dado. Por ejemplo, al estimar el efecto de la obesidad en la presión arterial, el investigador recopilaría los datos y mantendría constantes los hábitos de las personas respecto de comer, fumar y beber para reducir la influencia de estas variables en la presión arterial.

En las ciencias sociales, los datos por lo general son de naturaleza no experimental, es decir, no están sujetos al control del investigador.¹⁴ Por ejemplo, el investigador no ejerce ningún control directo sobre los datos del PIB, desempleo, precios de acciones, etc. Como veremos, esta falta de control a menudo genera problemas especiales para el investigador al identificar la causa o causas precisas que afectan una situación particular. Por ejemplo, ¿es la oferta monetaria la que determina el PIB (nominal) o la relación es inversa?

¹² Y. Grunfeld, “The Determinants of Corporate Investment”, tesis doctoral inédita, Departamento de Economía, Universidad de Chicago, 1958. Estos datos son ya una herramienta excelente para ilustrar los modelos de regresión con datos en panel.

¹³ Para mayor claridad, ver Albert T. Somers, *The U.S. Economy Demystified: What the Major Economic Statistics Mean and Their Significance for Business*, D.C. Heath, Lexington, Massachusetts, 1985.

¹⁴ También en las ciencias sociales, con mucha frecuencia, hay experimentos controlados; se da un ejemplo en el ejercicio 1.6.

TABLA 1.2 Datos de inversión de cuatro empresas, 1935-1954

Observación	I	F_{-1}	C_{-1}	Observación	I	F_{-1}	C_{-1}
GE				US			
1935	33.1	1 170.6	97.8	1935	209.9	1 362.4	53.8
1936	45.0	2 015.8	104.4	1936	355.3	1 807.1	50.5
1937	77.2	2 803.3	118.0	1937	469.9	2 673.3	118.1
1938	44.6	2 039.7	156.2	1938	262.3	1 801.9	260.2
1939	48.1	2 256.2	172.6	1939	230.4	1 957.3	312.7
1940	74.4	2 132.2	186.6	1940	361.6	2 202.9	254.2
1941	113.0	1 834.1	220.9	1941	472.8	2 380.5	261.4
1942	91.9	1 588.0	287.8	1942	445.6	2 168.6	298.7
1943	61.3	1 749.4	319.9	1943	361.6	1 985.1	301.8
1944	56.8	1 687.2	321.3	1944	288.2	1 813.9	279.1
1945	93.6	2 007.7	319.6	1945	258.7	1 850.2	213.8
1946	159.9	2 208.3	346.0	1946	420.3	2 067.7	232.6
1947	147.2	1 656.7	456.4	1947	420.5	1 796.7	264.8
1948	146.3	1 604.4	543.4	1948	494.5	1 625.8	306.9
1949	98.3	1 431.8	618.3	1949	405.1	1 667.0	351.1
1950	93.5	1 610.5	647.4	1950	418.8	1 677.4	357.8
1951	135.2	1 819.4	671.3	1951	588.2	2 289.5	341.1
1952	157.3	2 079.7	726.1	1952	645.2	2 159.4	444.2
1953	179.5	2 371.6	800.3	1953	641.0	2 031.3	623.6
1954	189.6	2 759.9	888.9	1954	459.3	2 115.5	669.7
GM				WEST			
1935	317.6	3 078.5	2.8	1935	12.93	191.5	1.8
1936	391.8	4 661.7	52.6	1936	25.90	516.0	0.8
1937	410.6	5 387.1	156.9	1937	35.05	729.0	7.4
1938	257.7	2 792.2	209.2	1938	22.89	560.4	18.1
1939	330.8	4 313.2	203.4	1939	18.84	519.9	23.5
1940	461.2	4 643.9	207.2	1940	28.57	628.5	26.5
1941	512.0	4 551.2	255.2	1941	48.51	537.1	36.2
1942	448.0	3 244.1	303.7	1942	43.34	561.2	60.8
1943	499.6	4 053.7	264.1	1943	37.02	617.2	84.4
1944	547.5	4 379.3	201.6	1944	37.81	626.7	91.2
1945	561.2	4 840.9	265.0	1945	39.27	737.2	92.4
1946	688.1	4 900.0	402.2	1946	53.46	760.5	86.0
1947	568.9	3 526.5	761.5	1947	55.56	581.4	111.1
1948	529.2	3 245.7	922.4	1948	49.56	662.3	130.6
1949	555.1	3 700.2	1 020.1	1949	32.04	583.8	141.8
1950	642.9	3 755.6	1 099.0	1950	32.24	635.2	136.7
1951	755.9	4 833.0	1 207.7	1951	54.38	732.8	129.7
1952	891.2	4 924.9	1 430.5	1952	71.78	864.1	145.5
1953	1 304.4	6 241.7	1 777.3	1953	90.08	1 193.5	174.8
1954	1 486.7	5 593.6	2 226.3	1954	68.60	1 188.9	213.5

Notas: $Y = I$ = inversión bruta = adiciones a planta y equipo más mantenimiento y reparaciones, en millones de dólares deflacionados por P_1 .

$X_2 = F$ = valor de la empresa = precio de las acciones comunes y preferentes al 31 de diciembre (o precio promedio del 31 de diciembre y 31 de enero del año siguiente) multiplicado por el número de acciones comunes y preferentes en circulación más el valor total de la deuda en libros al 31 de diciembre, en millones de dólares deflacionados por P_2 .

$X_3 = C$ = existencias de planta y equipo = suma acumulada de adiciones netas a planta y equipo deflacionadas por P_1 menos provisión para depreciación deflacionada por P_3 en estas definiciones.

P_1 = factor de deflación implícito de los precios de equipo duradero de los productores (1947 = 100).

P_2 = factor de deflación implícito de los precios del PIB (1947 = 100).

P_3 = factor de deflación del gasto de depreciación = promedio móvil de 10 años del índice de precios al mayoreo de metales y productos metálicos (1947 = 100).

Fuente: Reproducido de H. D. Vinod y Aman Ullah, *Recent Advances in Regression Methods*, Marcel Dekker, Nueva York, 1981, pp. 259-261.

Precisión de los datos¹⁵

Si bien se dispone de numerosos datos para la investigación económica, su calidad no siempre es adecuada, y por múltiples razones.

1. Como ya vimos, en su mayoría, los datos de las ciencias sociales son de naturaleza no experimental. Por consiguiente, es posible incurrir en errores de observación, sea por acción u omisión.
2. Aun en datos reunidos experimentalmente surgen errores de medición debido a las aproximaciones o al redondeo.
3. En encuestas por cuestionarios, el problema de la falta de respuesta puede ser grave; un investigador tiene suerte si obtiene una tasa de respuesta de 40%. El análisis basado en dicha tasa de respuesta parcial quizá no refleje de verdad el comportamiento del 60% que no respondió, y ocasione, por consiguiente, un sesgo de selectividad (muestral). Además, existe el problema de quienes responden el cuestionario pero no todas las preguntas, sobre todo las que son delicadas por tratar cuestiones financieras, lo que genera un sesgo adicional de selectividad.
4. Los métodos de muestreo para obtención de datos llegan a variar tanto que a menudo es difícil comparar los resultados de las diversas muestras.
5. Las cifras económicas suelen estar disponibles en niveles muy agregados. Por ejemplo, la mayor parte de los macrodatos (como el PIB, empleo, inflación, desempleo) están disponibles para la economía en su conjunto, o, en el mejor de los casos, para algunas regiones geográficas muy amplias. Los datos con estos niveles tan elevados de agregación tal vez no ilustren mucho sobre los sujetos o las microunidades objeto de estudio.
6. Debido a su carácter confidencial, ciertos datos sólo pueden publicarse en forma muy agregada. En el caso de Estados Unidos, por ejemplo, la ley prohíbe al IRS (hacienda) revelar información sobre declaraciones de impuestos individuales; sólo puede revelar algunos datos generales. Por consiguiente, si se desea conocer el monto gastado en salud por los individuos con cierto nivel de ingresos, sólo es posible en un nivel muy agregado. Pero los macroanálisis de este tipo con frecuencia resultan insuficientes para revelar la dinámica del comportamiento de las microunidades. De igual forma, el Departamento de Comercio estadounidense, que levanta el censo de empresas cada cinco años, no tiene autorización para revelar información sobre producción, empleo, consumo de energía, gastos de investigación y desarrollo, etc., de las empresas. Así, es difícil estudiar las diferencias entre las empresas en estos aspectos.

Por estos problemas, y muchos más, **el investigador debe tener siempre en mente que el resultado de la investigación será tan bueno como lo sea la calidad de los datos**. Por tanto, si en algunas situaciones los investigadores concluyen que los resultados de la investigación son “insatisfactorios”, la causa puede ser la mala calidad de los datos y no un modelo equivocado. Por desgracia, debido a la naturaleza no experimental de los datos de la mayoría de los estudios de ciencias sociales, los investigadores con frecuencia no tienen más remedio que depender de la información disponible. Sin embargo, siempre deben tener presente que los datos pueden no ser los mejores y tratar de no ser muy dogmáticos sobre los resultados de un estudio dado, sobre todo cuando la calidad de los datos no es confiable.

Una observación sobre las escalas de medición de las variables¹⁶

Las variables que a menudo encontrará se clasifican en cuatro categorías generales: *escala de razón*, *escala de intervalo*, *escala ordinal* y *escala nominal*. Es importante comprender cada una.

¹⁵ Para un examen crítico, ver O. Morgenstern, *The Accuracy of Economic Observations*, 2a. ed., Princeton University Press, Princeton, Nueva Jersey, 1963.

¹⁶ El siguiente análisis se basa en gran medida en Aris Spanos, *Probability Theory and Statistical Inference: Econometric Modeling with Observational Data*, Cambridge University Press, Nueva York, 1999, p. 24.

Escala de razón

Para la variable X , al tomar dos valores (X_1 y X_2), la razón X_1/X_2 y la distancia ($X_2 - X_1$) son cantidades con un significado. Asimismo, hay un ordenamiento natural (ascendente o descendente) de los valores a lo largo de la escala. En virtud de lo anterior, son sensatas las comparaciones como $X_2 \leq X_1$ o $X_2 \geq X_1$. En su mayoría, las variables económicas pertenecen a esta categoría. Por consiguiente, no es descabellado preguntar a cuánto asciende el PIB de este año en comparación con el del año anterior. El ingreso personal, en dólares, es una variable de razón; alguien que gana 100 000 dólares recibe el doble que quien percibe 50 000 (antes de impuestos, desde luego).

Escala de intervalo

Una variable en escala de intervalo satisface las dos últimas propiedades de la variable en escala de razón, pero no la primera. Por tanto, la distancia entre dos periodos, (digamos 2000-1995), tiene significado, no así la razón de dos periodos (2000/1995). A las 11 de la mañana (hora de la costa del Pacífico de Estados Unidos) del 11 de agosto de 2007 se registró en Portland, Oregon, una temperatura de 60° Fahrenheit (15.5° Celsius), y en Tallahassee, Florida, de 90° F (32° C). La temperatura con esta escala no se mide en escala de razón pues no tiene sentido decir que en Tallahassee hizo 50% más calor que en Portland. Esto se debe sobre todo a que la escala Fahrenheit no usa 0° como base natural.

Escala ordinal

Una variable pertenece a esta categoría sólo si satisface la tercera propiedad de la escala de razón (es decir, el orden natural), como los sistemas de calificaciones por letras (A, B, C) o los niveles de ingresos alto, medio y bajo). Para estas variables hay un orden, pero las distancias entre las categorías no son cuantificables. Los estudiantes de economía recordarán las *curvas de indiferencia* entre dos bienes, en donde una curva superior de indiferencia señala un mayor nivel de utilidad, pero no se puede cuantificar en qué medida una curva de indiferencia es mayor que otra.

Escala nominal

Las variables de esta categoría no tienen ninguna característica de las variables en escala de razón. Las variables como el género (masculino y femenino) y el estado civil (casado, soltero, divorciado, separado) simplemente denotan categorías. *Pregunta:* ¿por qué no expresar dichas variables con las escalas de razón, intervalo u orden?

Como veremos, las técnicas econométricas adecuadas para las variables en escala de razón no resultarían pertinentes para las variables en escala nominal. En consecuencia, es importante tener en mente las diferencias entre los cuatro tipos de escalas de medición recién analizadas.

Resumen y conclusiones

1. La idea fundamental del análisis de regresión es la dependencia estadística de una variable, la dependiente, respecto de otra o más variables, las explicativas.
2. El objetivo de tal análisis es estimar o predecir la media o el valor promedio de la variable dependiente con base en los valores conocidos o fijos de las explicativas.
3. En la práctica, un buen análisis de regresión depende de la disponibilidad de datos apropiados. En este capítulo analizamos la naturaleza, fuentes y limitaciones de los datos disponibles para la investigación, en especial en las ciencias sociales.
4. En toda investigación se debe señalar con claridad las fuentes de los datos para el análisis, sus definiciones, sus métodos de recolección y cualquier laguna u omisión en ellos, así como toda revisión que se les haya aplicado. Tenga en cuenta que los datos macroeconómicos que publica el gobierno con frecuencia son objeto de revisión.
5. Como el lector tal vez no tenga tiempo, energía o recursos para llegar a la fuente original de los datos, tiene el derecho de suponer que el investigador los recopiló de manera apropiada, y que los cálculos y análisis son correctos.

EJERCICIOS

- 1.1. La tabla 1.3 proporciona datos sobre el índice de precios al consumidor de siete países industrializados, cuya base es 1982-1984 = 100.
- A partir de estos datos, calcule la tasa de inflación en cada país.¹⁷
 - Grafique la tasa de inflación de cada nación en función del tiempo (es decir, asigne el eje horizontal al tiempo, y el vertical, a la tasa de inflación).
 - ¿Qué conclusiones generales surgen respecto de la inflación en los siete países?
 - ¿Qué país tiene, al parecer, la tasa de inflación más variable? ¿Puede explicarlo?
- 1.2. *a)* Use la tabla 1.3 para trazar la gráfica de la tasa de inflación de Canadá, Francia, Alemania, Italia, Japón y Gran Bretaña, en comparación con la de Estados Unidos.
- Comente en términos generales el comportamiento de la tasa de inflación de los seis países, en comparación con la de Estados Unidos.
 - Si descubre que las tasas de inflación de esos seis países tienen la misma dirección que la de Estados Unidos, ¿esto indicaría que la inflación en Estados Unidos “provoca” la inflación en los demás países? ¿Por qué?

TABLA 1.3
IPC de siete países
industrializados,
1980-2005
(1982-1984 = 100)

Fuente: *Economic Report of the President*, 2007, tabla 108, p. 354.

Año	Estados Unidos	Canadá	Japón	Francia	Alemania	Italia	Gran Bretaña
1980	82.4	76.1	91.0	72.2	86.7	63.9	78.5
1981	90.9	85.6	95.3	81.8	92.2	75.5	87.9
1982	96.5	94.9	98.1	91.7	97.0	87.8	95.4
1983	99.6	100.4	99.8	100.3	100.3	100.8	99.8
1984	103.9	104.7	102.1	108.0	102.7	111.4	104.8
1985	107.6	109.0	104.2	114.3	104.8	121.7	111.1
1986	109.6	113.5	104.9	117.2	104.6	128.9	114.9
1987	113.6	118.4	104.9	121.1	104.9	135.1	119.7
1988	118.3	123.2	105.6	124.3	106.3	141.9	125.6
1989	124.0	129.3	108.0	128.7	109.2	150.7	135.4
1990	130.7	135.5	111.4	132.9	112.2	160.4	148.2
1991	136.2	143.1	115.0	137.2	116.3	170.5	156.9
1992	140.3	145.3	117.0	140.4	122.2	179.5	162.7
1993	144.5	147.9	118.5	143.4	127.6	187.7	165.3
1994	148.2	148.2	119.3	145.8	131.1	195.3	169.3
1995	152.4	151.4	119.2	148.4	133.3	205.6	175.2
1996	156.9	153.8	119.3	151.4	135.3	213.8	179.4
1997	160.5	156.3	121.5	153.2	137.8	218.2	185.1
1998	163.0	157.8	122.2	154.2	139.1	222.5	191.4
1999	166.6	160.5	121.8	155.0	140.0	226.2	194.3
2000	172.2	164.9	121.0	157.6	142.0	231.9	200.1
2001	177.1	169.1	120.1	160.2	144.8	238.3	203.6
2002	179.9	172.9	119.0	163.3	146.7	244.3	207.0
2003	184.0	177.7	118.7	166.7	148.3	250.8	213.0
2004	188.9	181.0	118.7	170.3	150.8	256.3	219.4
2005	195.3	184.9	118.3	173.2	153.7	261.3	225.6

¹⁷ Reste del IPC del año en cuestión el IPC del año anterior, divida la diferencia entre el IPC del año anterior y multiplique el resultado por 100. Así, la tasa de inflación de Canadá en 1981 fue de $[(85.6 - 76.1)/76.1] \times 100 = 12.48\%$ (aproximadamente).

- 1.3. En la tabla 1.4 se presentan los tipos de cambio de nueve países industrializados correspondientes a 1985-2006. Excepto para Gran Bretaña, el tipo de cambio se define como las unidades de la divisa equivalentes a un dólar estadounidense; para ese país, se define como el número de dólares estadounidenses que se cambian por una libra inglesa.
 - a) Grafique los tipos de cambio en función del tiempo y comente el comportamiento general de los tipos de cambio durante el periodo.
 - b) Se dice que el dólar tiene una *apreciación* si se compran más unidades de una divisa. Por el contrario, se dice que sufre una *depreciación* si se adquieren menos divisas. En el periodo 1985-2006, en general, ¿qué comportamiento tuvo el dólar estadounidense? A propósito, busque en un texto de macroeconomía o economía internacional los factores que determinan la apreciación o depreciación de una moneda.
- 1.4. En la tabla 1.5 aparecen los datos en los que se basa la oferta de dinero M1 de la figura 1.5. ¿Puede explicar por qué se incrementó la oferta de dinero durante el periodo que se muestra en la tabla?
- 1.5. Suponga que va a crear un modelo económico de actividades delictivas en el que considere las horas invertidas en ellas (por ejemplo, en la venta de drogas). ¿Qué variables tomaría en cuenta para crear dicho modelo? Vea si su modelo se asemeja al del economista ganador del premio Nobel, Gary Becker.¹⁸

TABLA 1.4 Tipos de cambio de nueve países: 1985-2006

Año	Australia	Canadá	R. P. de China	Japón	México	Corea del Sur	Suecia	Suiza	Gran Bretaña
1985	0.7003	1.3659	2.9434	238.47	0.257	872.45	8.6032	2.4552	1.2974
1986	0.6709	1.3896	3.4616	168.35	0.612	884.60	7.1273	1.7979	1.4677
1987	0.7014	1.3259	3.7314	144.60	1.378	826.16	6.3469	1.4918	1.6398
1988	0.7841	1.2306	3.7314	128.17	2.273	734.52	6.1370	1.4643	1.7813
1989	0.7919	1.1842	3.7673	138.07	2.461	674.13	6.4559	1.6369	1.6382
1990	0.7807	1.1668	4.7921	145.00	2.813	710.64	5.9231	1.3901	1.7841
1991	0.7787	1.1460	5.3337	134.59	3.018	736.73	6.0521	1.4356	1.7674
1992	0.7352	1.2085	5.5206	126.78	3.095	784.66	5.8258	1.4064	1.7663
1993	0.6799	1.2902	5.7795	111.08	3.116	805.75	7.7956	1.4781	1.5016
1994	0.7316	1.3664	8.6397	102.18	3.385	806.93	7.7161	1.3667	1.5319
1995	0.7407	1.3725	8.3700	93.96	6.447	772.69	7.1406	1.1812	1.5785
1996	0.7828	1.3638	8.3389	108.78	7.600	805.00	6.7082	1.2361	1.5607
1997	0.7437	1.3849	8.3193	121.06	7.918	953.19	7.6446	1.4514	1.6376
1998	0.6291	1.4836	8.3008	130.99	9.152	1 400.40	7.9522	1.4506	1.6573
1999	0.6454	1.4858	8.2783	113.73	9.553	1 189.84	8.2740	1.5045	1.6172
2000	0.5815	1.4855	8.2784	107.80	9.459	1 130.90	9.1735	1.6904	1.5156
2001	0.5169	1.5487	8.2770	121.57	9.337	1 292.02	10.3425	1.6891	1.4396
2002	0.5437	1.5704	8.2771	125.22	9.663	1 250.31	9.7233	1.5567	1.5025
2003	0.6524	1.4008	8.2772	115.94	10.793	1 192.08	8.0787	1.3450	1.6347
2004	0.7365	1.3017	8.2768	108.15	11.290	1 145.24	7.3480	1.2428	1.8330
2005	0.7627	1.2115	8.1936	110.11	10.894	1 023.75	7.4710	1.2459	1.8204
2006	0.7535	1.1340	7.9723	116.31	10.906	954.32	7.3718	1.2532	1.8434

Fuente: *Economic Report of the President*, 2007, tabla B-110, p. 356.

¹⁸ G. S. Becker, "Crime and Punishment: An Economic Approach", *Journal of Political Economy*, vol. 76, 1968, pp. 169-217.

TABLA 1.5

**Oferta de dinero
M1 ajustada por
estacionalidad: enero de
1959-julio de 1999 (miles
de millones de dólares)**

Fuente: Board of Governors,
Federal Reserve Bank, Estados
Unidos.

1959:01	138.8900	139.3900	139.7400	139.6900	140.6800	141.1700
1959:07	141.7000	141.9000	141.0100	140.4700	140.3800	139.9500
1960:01	139.9800	139.8700	139.7500	139.5600	139.6100	139.5800
1960:07	140.1800	141.3100	141.1800	140.9200	140.8600	140.6900
1961:01	141.0600	141.6000	141.8700	142.1300	142.6600	142.8800
1961:07	142.9200	143.4900	143.7800	144.1400	144.7600	145.2000
1962:01	145.2400	145.6600	145.9600	146.4000	146.8400	146.5800
1962:07	146.4600	146.5700	146.3000	146.7100	147.2900	147.8200
1963:01	148.2600	148.9000	149.1700	149.7000	150.3900	150.4300
1963:07	151.3400	151.7800	151.9800	152.5500	153.6500	153.2900
1964:01	153.7400	154.3100	154.4800	154.7700	155.3300	155.6200
1964:07	156.8000	157.8200	158.7500	159.2400	159.9600	160.3000
1965:01	160.7100	160.9400	161.4700	162.0300	161.7000	162.1900
1965:07	163.0500	163.6800	164.8500	165.9700	166.7100	167.8500
1966:01	169.0800	169.6200	170.5100	171.8100	171.3300	171.5700
1966:07	170.3100	170.8100	171.9700	171.1600	171.3800	172.0300
1967:01	171.8600	172.9900	174.8100	174.1700	175.6800	177.0200
1967:07	178.1300	179.7100	180.6800	181.6400	182.3800	183.2600
1968:01	184.3300	184.7100	185.4700	186.6000	187.9900	189.4200
1968:07	190.4900	191.8400	192.7400	194.0200	196.0200	197.4100
1969:01	198.6900	199.3500	200.0200	200.7100	200.8100	201.2700
1969:07	201.6600	201.7300	202.1000	202.9000	203.5700	203.8800
1970:01	206.2200	205.0000	205.7500	206.7200	207.2200	207.5400
1970:07	207.9800	209.9300	211.8000	212.8800	213.6600	214.4100
1971:01	215.5400	217.4200	218.7700	220.0000	222.0200	223.4500
1971:07	224.8500	225.5800	226.4700	227.1600	227.7600	228.3200
1972:01	230.0900	232.3200	234.3000	235.5800	235.8900	236.6200
1972:07	238.7900	240.9300	243.1800	245.0200	246.4100	249.2500
1973:01	251.4700	252.1500	251.6700	252.7400	254.8900	256.6900
1973:07	257.5400	257.7600	257.8600	259.0400	260.9800	262.8800
1974:01	263.7600	265.3100	266.6800	267.2000	267.5600	268.4400
1974:07	269.2700	270.1200	271.0500	272.3500	273.7100	274.2000
1975:01	273.9000	275.0000	276.4200	276.1700	279.2000	282.4300
1975:07	283.6800	284.1500	285.6900	285.3900	286.8300	287.0700
1976:01	288.4200	290.7600	292.7000	294.6600	295.9300	296.1600
1976:07	297.2000	299.0500	299.6700	302.0400	303.5900	306.2500
1977:01	308.2600	311.5400	313.9400	316.0200	317.1900	318.7100
1977:07	320.1900	322.2700	324.4800	326.4000	328.6400	330.8700
1978:01	334.4000	335.3000	336.9600	339.9200	344.8600	346.8000
1978:07	347.6300	349.6600	352.2600	353.3500	355.4100	357.2800
1979:01	358.6000	359.9100	362.4500	368.0500	369.5900	373.3400
1979:07	377.2100	378.8200	379.2800	380.8700	380.8100	381.7700
1980:01	385.8500	389.7000	388.1300	383.4400	384.6000	389.4600
1980:07	394.9100	400.0600	405.3600	409.0600	410.3700	408.0600
1981:01	410.8300	414.3800	418.6900	427.0600	424.4300	425.5000
1981:07	427.9000	427.8500	427.4600	428.4500	430.8800	436.1700
1982:01	442.1300	441.4900	442.3700	446.7800	446.5300	447.8900
1982:07	449.0900	452.4900	457.5000	464.5700	471.1200	474.3000
1983:01	476.6800	483.8500	490.1800	492.7700	499.7800	504.3500
1983:07	508.9600	511.6000	513.4100	517.2100	518.5300	520.7900
1984:01	524.4000	526.9900	530.7800	534.0300	536.5900	540.5400
1984:07	542.1300	542.3900	543.8600	543.8700	547.3200	551.1900
1985:01	555.6600	562.4800	565.7400	569.5500	575.0700	583.1700
1985:07	590.8200	598.0600	604.4700	607.9100	611.8300	619.3600
1986:01	620.4000	624.1400	632.8100	640.3500	652.0100	661.5200

(continúa)

TABLA 1.5
(continuación)

1986:07	672.2000	680.7700	688.5100	695.2600	705.2400	724.2800
1987:01	729.3400	729.8400	733.0100	743.3900	746.0000	743.7200
1987:07	744.9600	746.9600	748.6600	756.5000	752.8300	749.6800
1988:01	755.5500	757.0700	761.1800	767.5700	771.6800	779.1000
1988:07	783.4000	785.0800	784.8200	783.6300	784.4600	786.2600
1989:01	784.9200	783.4000	782.7400	778.8200	774.7900	774.2200
1989:07	779.7100	781.1400	782.2000	787.0500	787.9500	792.5700
1990:01	794.9300	797.6500	801.2500	806.2400	804.3600	810.3300
1990:07	811.8000	817.8500	821.8300	820.3000	822.0600	824.5600
1991:01	826.7300	832.4000	838.6200	842.7300	848.9600	858.3300
1991:07	862.9500	868.6500	871.5600	878.4000	887.9500	896.7000
1992:01	910.4900	925.1300	936.0000	943.8900	950.7800	954.7100
1992:07	964.6000	975.7100	988.8400	1 004.340	1 016.040	1 024.450
1993:01	1 030.900	1 033.150	1 037.990	1 047.470	1 066.220	1 075.610
1993:07	1 085.880	1 095.560	1 105.430	1 113.800	1 123.900	1 129.310
1994:01	1 132.200	1 136.130	1 139.910	1 141.420	1 142.850	1 145.650
1994:07	1 151.490	1 151.390	1 152.440	1 150.410	1 150.440	1 149.750
1995:01	1 150.640	1 146.740	1 146.520	1 149.480	1 144.650	1 144.240
1995:07	1 146.500	1 146.100	1 142.270	1 136.430	1 133.550	1 126.730
1996:01	1 122.580	1 117.530	1 122.590	1 124.520	1 116.300	1 115.470
1996:07	1 112.340	1 102.180	1 095.610	1 082.560	1 080.490	1 081.340
1997:01	1 080.520	1 076.200	1 072.420	1 067.450	1 063.370	1 065.990
1997:07	1 067.570	1 072.080	1 064.820	1 062.060	1 067.530	1 074.870
1998:01	1 073.810	1 076.020	1 080.650	1 082.090	1 078.170	1 077.780
1998:07	1 075.370	1 072.210	1 074.650	1 080.400	1 088.960	1 093.350
1999:01	1 091.000	1 092.650	1 102.010	1 108.400	1 104.750	1 101.110
1999:07	1 099.530	1 102.400	1 093.460			

1.6. *Experimentos controlados en economía:* El 7 de abril de 2000, el presidente Clinton convirtió en ley una propuesta aprobada por ambas cámaras legislativas estadounidenses mediante la cual se eliminaban las limitaciones de beneficios para los derechohabientes del sistema de seguridad social. Antes de esa ley, los derechohabientes de entre 65 y 69 años con percepciones mayores de 17 000 dólares al año perdían el equivalente a un dólar de las prestaciones de seguridad social por cada 3 dólares de ingresos que excedieran 17 000 dólares. ¿Cómo se planearía un estudio que evaluara el efecto de este cambio en la ley? *Nota:* En la ley derogada no había restricción de ingresos para los derechohabientes de más de 70 años.

1.7. Los datos de la tabla 1.6 se publicaron el primero de marzo de 1984 en el periódico *The Wall Street Journal*. Se refieren al presupuesto publicitario (en millones de dólares) de 21 empresas en 1983 y a los millones de impactos semanales (veces que los clientes ven los anuncios de los productos de dichas compañías por semana). La información se basa en una encuesta a 4 000 adultos en la que se pidió a los usuarios de los productos que mencionaran un comercial que hubieran visto en la semana anterior y que tuviera que ver con la categoría del producto.

- Trace una gráfica con los impactos en el eje vertical y el gasto publicitario en el horizontal.
- ¿Qué se puede decir sobre la relación entre ambas variables?
- Al observar la gráfica, ¿cree que es redituable el gasto en publicidad? Piense en todos los comerciales que se transmiten el domingo que se juega el Super Bowl o durante la Serie Mundial del beisbol estadounidense.

Nota: En los siguientes capítulos estudiaremos más a fondo los datos de la tabla 1.6.

TABLA 1.6
Efecto del gasto en
publicidad

Fuente: <http://lib.stat.cmu.edu/DASL/Datafiles/tvadsdat.html>.

Empresa	Impactos, millones	Gasto, millones de dólares de 1983
1. Miller Lite	32.1	50.1
2. Pepsi	99.6	74.1
3. Stroh's	11.7	19.3
4. Fed'l Express	21.9	22.9
5. Burger King	60.8	82.4
6. Coca-Cola	78.6	40.1
7. McDonald's	92.4	185.9
8. MCI	50.7	26.9
9. Diet Cola	21.4	20.4
10. Ford	40.1	166.2
11. Levi's	40.8	27.0
12. Bud Lite	10.4	45.6
13. ATT/Bell	88.9	154.9
14. Calvin Klein	12.0	5.0
15. Wendy's	29.2	49.7
16. Polaroid	38.0	26.9
17. Shasta	10.0	5.7
18. Meow Mix	12.3	7.6
19. Oscar Meyer	23.4	9.2
20. Crest	71.1	32.4
21. Kibbles 'N Bits	4.4	6.1