

ECONOMETRÍA I

Tema 1: La naturaleza de la econometría y los datos econométricos

Patricia Moreno
Juan Manuel Rodríguez Poo
Alexandra Soberon
Departamento de Economía

- **Econometría**: ciencia basada en el desarrollo de métodos estadísticos que se utilizan para el estudio de relaciones económicas o la contrastación de teorías económicas:
 - determinar los efectos de un programa de formación sobre la productividad o el salario de los trabajadores.
 - estudiar si los rendimientos derivados de diversas estrategias de inversión a corto plazo cumplen lo establecido por las teorías económicas.
 - determinar los efectos de una campaña publicitaria.
- La econometría combina elementos de: teoría económica, matemáticas y estadística.

- La economía nos sugiere ciertas relaciones cualitativas entre variables o factores pero no las magnitudes cuantitativas de sus **efectos causales**.
- Así, la econometría nos permite determinar los efectos causales:
 - ¿Cuál es el efecto cuantitativo sobre el rendimiento de los estudiantes de reducir el tamaño de las aulas?
 - ¿Cuál es el efecto sobre los salarios de un año más de educación?
 - ¿Cuál es la elasticidad precio de los cigarrillos?
- En este curso aprenderemos ciertos métodos que nos permiten estimar estos efectos causales usando datos observacionales.

- La mayoría de las teorías económicas tratan de determinar el posible efecto causal de una variable (educación) sobre otra variable (productividad de los trabajadores).
- **Problema:** Rara vez podemos establecer una relación concluyente entre dos variables a menos que podamos establecer causalidad.
- Si en el modelo a analizar hemos controlado por suficientes variables, el efecto estimado *ceteris paribus* puede ser considerado como causal.
- La causalidad puede ser difícil de establecer. Aislar una variable concreta puede parecer una tarea imposible.

- **Efecto de los fertilizantes sobre el rendimiento de un cultivo:**
Incremento experimentado en la producción de un cultivo de la población analizada si, manteniendo todo lo demás constante, aplicamos distintas cantidades de fertilizante.
- Para determinar el efecto causal de la cantidad de fertilizante sobre el cultivo debemos realizar un experimento:
 - elegir varias parcelas de un acre.
 - aplicar de manera aleatoria a cada parcela una cantidad distinta de fertilizante y medir su rendimiento.
 - emplear métodos estadísticos para medir la relación entre rendimiento y cantidad de fertilizante.

- **Efecto causal de la educación en el salario**: Representa el incremento salarial que conseguiría un individuo de la población objeto de estudio si, manteniendo constantes sus demás características, tuviera un mayor nivel de educación.
- Para determinar el efecto causal de un año más de educación realizamos un experimento:
 - dividimos a la población de estudio en diversos grupos.
 - asignamos, de manera aleatoria, una cantidad de educación distinta a cada uno de estos grupos.
 - emplear métodos estadísticos para medir la relación entre la cantidad de educación y los salarios de los individuos.

- **Datos experimentales:** se recopilan en entornos controlados (laboratorios). Podemos interpretar un co-movimiento o correlación entre variables como un efecto causal pero en las ciencias sociales no son habituales.
- **Datos no experimentales (observacionales):** tanto los valores de la variable objeto de estudio como de los factores susceptibles de afectarla están fuera de nuestro control. No podemos interpretar un co-movimiento o correlación entre variables como un efecto causal.

Sección cruzada: muestra aleatoria sobre individuos, hogares, empresas, países, etc., en un momento de tiempo dado. Características:

- La ordenación no importa.
- En general, cabe esperar que las observaciones sean independientes entre sí.
- Si los datos no son una muestra aleatoria podemos encontrarnos con un problema de selección muestral.

Ejemplo: Muestra de 526 individuos de Cantabria en 2014.

Individuo (i)	$Salario_i$	$Educacion_i$	$Experiencia_i$	$Sexo_i$	$Civil_i$
1	3.10	11	2	1	0
2	3.24	12	22	1	1
3	3.00	11	2	0	0
4	6.00	8	44	0	1
5	5.30	12	7	0	1
⋮	⋮	⋮	⋮	⋮	⋮
525	1.56	16	5	0	1
526	3.50	14	5	1	0

Fuente: Wooldridge (2009).

Series temporales: observación de una o varias variables a lo largo del tiempo. Características:

- El orden cronológico y la frecuencia (anual, trimestral, mensual, diaria) de los datos son importantes.
- Rara vez puede suponerse que las observaciones son independientes en el tiempo. Al contrario, cabe esperar que exista dependencia entre observaciones.
- Dado que no es una muestra aleatoria diversos problemas de estimación han de ser considerados.
- Las tendencias y la estacionalidad son aspectos relevantes en las series temporales.

Ejemplo: Datos anuales de salario mínimo, desempleo y datos relevantes para un país determinado (Puerto Rico).

Año	Salario minimo	Tasa media cobertura	Desempleo	PIB
1950	0.20	20.1	15.4	878.7
1951	0.21	20.7	16.0	925.0
1952	0.23	22.6	14.8	1015.9
⋮	⋮	⋮	⋮	⋮
1986	3.35	58.1	18.9	4281.6
1987	3.35	58.2	16.8	4496.7

Fuente: Wooldridge (2009).

Datos de panel: base de datos de sección cruzada donde disponemos de una serie de tiempo para cada individuo. Características:

- Nos permiten seguir a la misma unidad individual aleatoria a lo largo del tiempo (datos longitudinales).
- Por cada miembro de la sección cruzada el orden cronológico es importante.
- No se puede confundir los datos de panel con las series temporales de secciones cruzadas (cohortes).

Ejemplo: muestra que contiene los principales estadísticos criminales de 150 ciudades durante el periodo 1986 – 1990.

Ciudad	Año	Asesinatos	Población	desempleo	Policías
1	1986	5	350000	8.7	440
⋮	⋮	⋮	⋮	⋮	⋮
1	1990	8	359200	7.2	471
2	1986	2	64300	5.4	75
⋮	⋮	⋮	⋮	⋮	⋮
2	1990	1	65100	5.5	75
⋮	⋮	⋮	⋮	⋮	⋮
150	1986	25	543000	4.3	520
⋮	⋮	⋮	⋮	⋮	⋮
150	1990	32	546200	5.2	493

Fuente: Wooldridge (2009).

- **Objetivo:** utilizar datos para cuantificar una relación o contrastar una teoría.
- Estructura de un trabajo empírico:
 - Modelo económico: La teoría económica propone modelos que explican el comportamiento de una o varias variables, Y_1, \dots, Y_m , en función de otras variables, X_1, \dots, X_k determinadas fuera del modelo.
 - Modelo econométrico: permite cuantificar y contrastar las relaciones entre variables establecidas por los modelos económicos a partir de los datos.
 - Datos.

- Formular de manera cuidadosa el problema a estudiar. Ej: determinar el comportamiento criminal de los individuos.
- Construir un modelo económico formal que nos permita expresar Y a través de una ecuación del tipo, Becker (1968):

$$Y = f(X_1, X_2, X_3, \dots, X_7),$$

donde $f(\cdot)$ representa la función de utilidad de los individuos,

- Y = horas invertidas en actividades delictivas
- X_1 = "salario" por cada hora invertida en la actividad delictiva
- X_2 = salario por hora en un empleo legal
- X_3 = otro ingreso
- X_4 = probabilidad de ser atrapado
- X_5 = probabilidad de ser declarado culpable una vez atrapado
- X_6 = pena prevista si es declarado culpable
- X_7 = edad.

- Establecer el modelo econométrico. Para ello hay que
 - determinar cómo tratar el problema de las variables que, aunque postuladas por el modelo económico, no pueden ser observadas.
 - establecer la naturaleza estocástica que gobierna las relaciones entre variables.
 - especificar la forma de la función $f(\cdot)$ estableciendo una relación entre Y y X_1, \dots, X_k a través de parámetros constantes.
- **Modelo econométrico:**

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + \beta_4 X_4 + \beta_5 X_5 + \beta_6 X_6 + \beta_7 X_7 + u,$$

donde u contiene la heterogeneidad individual y los factores no observados (salario por realizar la actividad criminal, moralidad, aprendizaje familiar, así como errores a la hora de medir aspectos como la actividad criminal o la probabilidad de ser arrestado).

- Establecer las hipótesis de interés en relación con los parámetros desconocidos:

$$H_0 : \beta_1 = 0.$$

- Determinar los datos a emplear.
- Una vez recolectados los datos, los métodos econométricos son usados para:
 - estimar los parámetros desconocidos.
 - evaluar las hipótesis de interés.
 - realizar predicciones.

Conceptos previos: valor esperado

- **Valor esperado:** si X es una variable aleatoria, $E(X)$ es un promedio ponderado de todos los posibles valores de X ,
 - Variable discreta: $E(X) = \sum_{j=1}^k x_j f(x_j)$.
 - Variable continua: $E(X) = \int_{-\infty}^{\infty} x f(x) dx$.
- Propiedades:
 - Para cualquier constante c , $E(c) = c$.
 - Para cualquier constante a y b , $E(aX + b) = aE(X) + b$.
 - Si a_1, a_2, \dots, a_n son constantes y X_1, X_2, \dots, X_n son variables aleatorias,

$$E(a_1X_1 + a_2X_2 + \dots + a_nX_n) = a_1E(X_1) + a_2E(X_2) + \dots + a_nE(X_n).$$

- **Varianza:** distancia esperada de X respecto de su media:

$$\text{Var}(X) = E \left[(X - \mu)^2 \right] = E \left(X^2 \right) - \mu^2.$$

- Propiedades:

- $\text{Var}(X) = 0$ si y sólo si existe una c tal que $P(X = c) = 1$, en cuyo caso $E(X) = c$.
- Para cualquier constante a y b , $\text{Var}(aX + b) = a^2 \text{Var}(X)$.
- Para cualquier constante a y b ,

$$\text{Var}(aX + bY) = a^2 \text{Var}(X) + b^2 \text{Var}(Y) + 2ab \text{Cov}(X, Y).$$

- **Desviación estándar:** Raíz cuadrada de la varianza.

$$sd(X) = \sqrt{Var(X)}$$

- Propiedades:
 - Para cualquier constante c , $sd(c) = 0$.
 - Para cualquier constante a y b , $sd(aX + b) = |a|sd(X)$.

- Dada una variable aleatoria X , definimos una nueva variable aleatoria restándole su media y dividiéndola por su desviación estándar:

$$Z = \frac{X - \mu}{\sigma}$$

- Propiedades de la variable estandarizada (Z): si $Z = aX - b$, donde $a = (1/\sigma)$ y $b = (\mu/\sigma)$,
 - $E(Z) = aE(X) - b = 0$.
 - $Var(Z) = Var(aX - b) = a^2 Var(X) = 1$.

- **Covarianza:** medida de la dependencia lineal entre dos variables aleatorias (X, Y)

$$\text{Cov}(X, Y) = E[(X - \mu_X)(Y - \mu_Y)] = E(XY) - \mu_X\mu_Y.$$

- Propiedades:

- Si X e Y son independientes, $\text{Cov}(X, Y) = 0$.

Proof: Si $X \perp Y$, $E(XY) = E(X)E(Y)$ y

$$\text{Cov}(X, Y) = E(XY) - \mu_X\mu_Y = \mu_X\mu_Y - \mu_X\mu_Y = 0.$$

- Para cualquier constante a_1, a_2, b_1, b_2 ,

$$\text{Cov}(a_1X + b_1, a_2X + b_2) = a_1a_2\text{Cov}(X, Y).$$

- Desigualdad de Cauchy-Schwarz: $|\text{Cov}(X, Y)| \leq \text{sd}(X)\text{sd}(Y)$.

Conceptos previos: coeficiente de correlación

- **Coeficiente de correlación:** medida sobre la dependencia lineal entre dos variables aleatorias (X, Y) que no tiene en cuenta las unidades de medida:

$$\text{Corr}(X, Y) = \frac{\text{Cov}(X, Y)}{\text{sd}(X)\text{sd}(Y)} = \frac{\sigma_{XY}}{\sigma_X\sigma_Y}.$$

- Propiedades:
 - Al ser positivas, $\text{Cov}(X, Y)$ y $\text{Corr}(X, Y)$ siempre si tienen el mismo signo.
 - $-1 \leq \text{Corr}(X, Y) \leq 1$.
 - Para las constantes (a_1, a_2, b_1, b_2) con $a_1, a_2 > 0$,

$$\text{Corr}(a_1X + b_1, a_2Y + b_2) = \text{Corr}(X, Y).$$

Conceptos previos: esperanza condicional

- **Esperanza condicional:** a diferencia de la covarianza y de la correlación, la esperanza condicional de Y dado X nos permite explicar una variable Y en términos de otra variable X o establecer si Y se relaciona con X de un modo no lineal.
 - Variable discreta: $E(Y|X) = \sum_{j=1}^m Y_j f_{Y|X}(Y_j|X)$.
 - Variable continua: $E(Y|X) = \int_{-\infty}^{\infty} Y f_{Y|X}(Y|X)$.
- Propiedades:
 - $E[c(X)|X] = c(X)$, para cualquier función $c(X)$.
 - Para las funciones $a(X)$ y $b(X)$

$$E[a(X)Y + b(X)|X] = a(X)E(Y|X) + b(X).$$

- Si X e Y son independientes, $E(Y|X) = E(Y)$.
- Ley de esperanzas iteradas: $E[E(Y|X)] = E(Y)$ o $E(Y|X) = E[E(Y|X, Z)|Z]$.

- **Varianza condicional:** varianza asociada a la distribución condicional de Y , dado X .

$$\text{Var}(Y|X) = E(Y^2|X) - [E(Y|X)]^2.$$

- Propiedades:
 - Si X e Y son independientes, $\text{Var}(Y|X) = \text{Var}(Y)$.

- Wooldridge, J. (2009) Capítulo 1 y Apéndices A y D.
- Stock, J.H. y Watson, M.M. (2012) Capítulo 1.