

Introducción al Análisis Econométrico

Econometría I

Pavel Solís

Universidad Anáhuac

2026

Índice

¿Qué es la Econometría?

Pasos del Análisis Empírico

Estructura de los Datos Económicos

Causalidad

¿Qué es la Econometría?

¿Qué es la econometría?

Ej. 1. Investigador quiere evaluar efectividad de programa financiado por gobierno para capacitar sobre uso de computadoras en manufactura

- *Objetivo:* Determinar efecto del programa en el salario
- *Implicación:* ¿Mantener financiamiento del programa de capacitación?

Ej. 2. Inversionista analiza rendimientos de estrategias con bonos del gobierno

- *Objetivo:* Determinar si estrategias están en línea con teoría económica
- *Implicación:* ¿Se pudieron obtener mayores rendimientos?

Ej. 3. Empleado de una empresa recaba información de las ventas en el país

- *Objetivo:* Pronosticar las ventas en los siguientes meses
- *Implicación:* ¿Empresa podría pedir crédito a un banco?

¿Qué es la econometría?

La **econometría** desarrolla métodos estadísticos que permiten:

- Estimar relaciones entre variables económicas
- Probar teorías económicas
- Analizar políticas públicas (evaluar programas gubernamentales)
- Implementar estrategias empresariales

Además de otras aplicaciones:

- Pronosticar variables macroeconómicas (tasas de interés, inflación, PIB)
- Influencia de gastos de campaña en resultados de votación
- Efectos del gasto público en educación sobre el desempeño de estudiantes

¿Qué es la econometría?

Los datos para el análisis empírico pueden venir de 2 fuentes:

- **Experimentales:** Se recolectan en laboratorios mediante experimentos
 - Comunes en ciencias naturales, difíciles de obtener en ciencias sociales
- **No experimentales** (observados, retrospectivos): No se recolectan mediante experimentos controlados porque involucran agentes económicos
 - Investigador es un recolector pasivo de datos

La econometría desarrolla técnicas para abordar complejidades inherentes a la recolección y análisis de datos económicos no experimentales

- Partió de estadística matemática pero el enfoque y la interpretación difieren

Historia breve

Newton (1700) obtuvo una ecuación del método de mínimos cuadrados ordinarios en su trabajo sobre equinoccios

Legendre (1805) y **Gauss** (1809) desarrollaron la regresión lineal por mínimos cuadrados para predecir el movimiento planetario

Quetelet (1831, 1835) tuvo gran influencia al introducir métodos estadísticos (incluido el método de mínimos cuadrados) en las ciencias sociales

Galton (1886) encontró que la altura de los hijos adultos tendía a desviarse menos de la altura media que la de sus padres, lo que llamó *regresión a la media*

Frisch fue uno de los cofundadores de la *Econometric Society* (1930) y fue el primer editor de la revista *Econometrica* (1933-1954)

Pasos del Análisis Empírico en Economía

Pasos del análisis empírico

1. Formular la pregunta de interés
2. Especificar el modelo econométrico
 - 2.1 Desarrollar el modelo teórico
 - 2.2 Especificar el modelo empírico
3. Formular la hipótesis de interés
4. Recolectar y *procesar* los datos para las variables de interés
5. Estimar los parámetros del modelo y probar la hipótesis de interés

Relevantes para cualquier área de economía aplicada (micro y macro)

Paso 2. Especificar modelo econométrico

Paso 2.1. Desarrollar modelo económico teórico

Cuando se quieren probar teorías económicas, se puede construir un modelo económico teórico (ecuaciones que describen comportamiento)

Ejemplo. Micro intermedia: Maximización de utilidad

Decisiones de consumo para maximizar utilidad sujeto a restricciones de recursos

cantidad demandada = $f(\text{precio}, \text{pr sustitutos}, \text{pr complementos}, \text{ingreso}, \text{gustos})$

Ejemplo. Macro intermedia: Modelo IS-LM

Relación: $Y, P, i, M \rightarrow \text{DA-OA}$

Paso 2. Especificar modelo econométrico

Paso 2.1. Desarrollar modelo económico teórico

La forma funcional específica se puede obtener de un modelo económico

- Derivaciones formales pueden dar conclusiones no intuitivas
- Ecuaciones del modelo pueden servir de base para el análisis econométrico

Ejemplo. Horas dedicadas al crimen

$$\text{crimen} = f(x_1, x_2, x_3, x_4, x_5, x_6, x_7)$$

Aunque es común utilizar la teoría económica de forma menos formal

Pregunta

¿Qué factores explican el sueldo (por hora) de una persona?

Paso 2. Especificar modelo econométrico

Paso 2.2. Especificar modelo empírico

En la práctica, el análisis econométrico inicia con el paso 2.2

- Variables a incluir con base en teoría económica y disponibilidad de datos
- Razonamiento económico y sentido común son guías al seleccionar variables

Al especificar un modelo econométrico:

- ¿Qué forma funcional utilizamos?
- ¿Qué hacemos con variables que no se pueden observar (ej. sueldo ilegal)?
- ¿Qué pasa con todos los otros factores que pueden influir?

Paso 2. Especificar modelo econométrico

Paso 2.2. Especificar modelo empírico

Ejemplo 1. Horas dedicadas al crimen

$$\text{crimen} = \beta_0 + \beta_1 \text{leg} + \beta_2 \text{ileg} + \beta_3 \text{otroin} + \beta_4 \text{prarres} + \beta_5 \text{prcárc} + \beta_6 \text{tcárc} + \beta_7 \text{edad} + u$$

u es el **término de error** o de perturbación, y contiene factores no observados

- Carácter, hermanos, educación de padres, errores de medición, etc.
- Podemos agregar más variables pero nunca eliminaremos u por completo
- Tratamiento de u es parte más importante de cualquier análisis econométrico

Las constantes $\beta_0, \beta_1, \dots, \beta_7$ son los **parámetros** del modelo econométrico

- Describen la dirección y qué tan fuerte es la relación entre las variables

Paso 2. Especificar modelo econométrico

Paso 2.2. Especificar modelo empírico

Ejemplo 2. Sueldo por hora

$$\text{sueldo} = \beta_0 + \beta_1 \text{educación} + \beta_2 \text{experiencia} + \beta_3 \text{capacitación} + u$$

Pregunta

¿Qué contiene u ?

Paso 3. Formular la hipótesis

Después de especificar el modelo econométrico, formulamos la hipótesis de interés
La hipótesis se puede expresar en términos de los parámetros del modelo

- En ejemplo 1, si creemos que el sueldo legal no afecta la actividad criminal,
 $\beta_1 = 0$
- En ejemplo 2, si nos interesa el efecto de la capacitación, β_3 sería el parámetro de interés

Los parámetros del modelo no son conocidos

- Usamos datos (paso 4) y métodos econométricos (paso 5) para estimarlos

Paso 4. Recolectar y procesar datos

Los datos se pueden recolectar de una o varias fuentes

- Públicas: INEGI, FRED, Banxico, FMI, BM
- Privadas: Corporativas (ej. MasterCard, Facebook), Bloomberg

Se recomienda seguir buenas prácticas

- Documentar la recolección
- Guardar archivos originales por separado

Es responsabilidad del investigador revisar la base de datos

- Generalmente requieren un procesamiento previo al análisis

Paso 5. Estimar parámetros y probar hipótesis de interés

Usamos datos y métodos econométricos para estimar los parámetros del modelo

- Los signos y valores de los parámetros son importantes

Después de la estimación, probamos formalmente la hipótesis de interés

- Comprobar teoría económica
- Evaluar el impacto de un programa social
- Validar importancia de una variable para pronosticar otra
- ...

Estructura de los Datos Económicos

Estructura de los datos económicos

Hay 4 tipos de estructuras de datos:

- Datos de corte transversal (CT)
- Datos de series de tiempo (ST)
- Datos de CT independientes (CTI)
- Datos panel o longitudinales (DP)

El análisis empírico explota las características especiales de cada estructura

- Algunos métodos econométricos se pueden aplicar a diferentes tipos de datos

Datos de corte transversal (CT)

Consisten en una muestra de unidades tomada en un punto en el tiempo

- Individuos, empresas, hogares, ciudades, estados, países

A veces no corresponden exactamente al mismo tiempo

- Ej. Encuesta familiar durante 1 año
- Ignoramos pequeñas diferencias de tiempo en la recolección de datos

Supuesto: Datos de CT se obtienen de muestreo aleatorio de población objetivo

- Ej. Población ocupada (salarios, educación, experiencia)

Datos de corte transversal (CT)

Muestreo aleatorio:

- Cada observación es independiente e idénticamente distribuida (iid)
- Tipo de muestreo común en estadística
- Simplifica el análisis de datos de CT

Violaciones del muestreo aleatorio:

- Problema de selección de muestra (ej. acumulación de riqueza)
- Población no es grande para asumir que observaciones son independientes

Datos de corte transversal (CT)

	wage	educ	exper	female	married
1.	3.1	11	2	1	0
2.	3.2	12	22	1	1
3.	3	11	2	0	0
4.	6	8	44	0	1
5.	5.3	12	7	0	1
199.	10	17	5	0	1
200.	11	12	30	0	1
201.	6.9	12	31	0	1
525.	12	16	5	0	1
526.	3.5	14	5	1	0

- Información de 526 individuos
- Número de observación no es característica del individuo
- Unidades de medición: wage (USD por hora), educ y exper (años)
- Variables binarias indican características cualitativas

Datos de corte transversal (CT)

Características de datos de CT:

- Orden de los datos no importa para el análisis econométrico
- Obtenidos de un muestreo aleatorio

Datos de CT se utilizan ampliamente en ciencias sociales

- Economía (micro aplicada)
 - Laboral, economía de la salud, organización industrial, economía urbana
- Ciencia política
 - Gastos de campaña, corrupción, votaciones parlamentarias

Datos de series de tiempo (ST)

Consisten en observaciones de una o varias variables a través del tiempo

- Precios de acciones, INPC, PIB, tasa homicidios, ventas de autos

El tiempo es una dimensión importante

- Eventos pasados pueden influir eventos futuros
- Rezago en comportamiento es común en ciencias sociales

Orden cronológico de observaciones puede dar información importante

- Correlación en el tiempo (a diferencia de CT)

Es importante considerar la *frecuencia* con que se recolectan los datos

- Diario, semanal, mensual, trimestral, anual

Datos de series de tiempo (ST)

	year	avgmin	prepop	prunemp	prgnp
1.	1950	.198	.47	15.4	878.7
2.	1951	.209	.449	16	925
3.	1952	.225	.434	14.8	1015.9
4.	1953	.311	.428	14.5	1081.3
19.	1968	1.104	.403	10.3	2455.3
20.	1969	1.149	.399	10.3	2684
36.	1985	3.35	.331	21.8	4172.8
37.	1986	3.35	.351	18.9	4281.6
38.	1987	3.35	.369	16.8	4496.7

- Variables económicas de un país 1950-1987
- Datos ordenados del más antiguo al más reciente

Datos de series de tiempo (ST)

Característica de datos de ST:

- Difícil asumir que observaciones económicas son independientes en el tiempo
 - Ej. Valor del PIB en periodo $t - 1$, nos dice posible rango en periodo t
- Datos tienen que estar ordenados cronológicamente para análisis de ST

Análisis de ST es diferente a CT

- Aunque varios métodos se pueden usar en ST y CT
- Pasos adicionales para especificar el modelo econométrico con ST

Técnicas desarrolladas para tomar en cuenta la dependencia en ST

- Ej. Tendencia, persistencia, efectos estacionales

Datos de CT independientes (CTI)

Algunos conjuntos de datos tienen características de CT y de ST

- Ej. 2 encuestas de CT a (distintos) hogares en años diferentes

Análisis de CTI sirve para:

- Aumentar el tamaño de la muestra
- Ver cómo cambia en el tiempo una relación clave

Orden no es esencial, pero identificar el periodo de cada observación es valioso

- Analizar efectos de política pública (ej. reducción de impuestos residenciales)

CTI se analizan como un CT estándar, pero podemos considerar dependencias en variables a través del tiempo

Datos de CT independientes (CTI)

	year	age	educ	female	exper
1.	78	25	12	0	8
2.	78	47	12	1	30
3.	78	49	6	0	38
550.	78	32	15	0	12
551.	85	43	10	0	27
552.	85	38	12	0	20
553.	85	22	12	1	4
1083.	85	32	12	0	14
1084.	85	27	12	0	9

- Encuesta poblacional en 1978 y 1985
- Individuos no son necesariamente los mismos
- Tamaños de muestra diferentes

Datos panel (DP)

Consisten en una ST para cada miembro de la sección transversal

- También se conocen como datos longitudinales
- Ej. Salario, educación, experiencia para mismos individuos cada 5 años
- Ej. Datos financieros de mismas empresas por 10 años

Características de DP:

- Mismas unidades en CT se siguen por un periodo de tiempo
- Son más difíciles de obtener que CTI

Estructura de DP se puede ver como CTI pero con mismas unidades observadas

- DP permiten contestar preguntas que CTI no

Datos panel (DP)

	state	year	nghtfat	minage	sl65	seatbelt
1.	1	1987	499	21	.458	0
2.	1	1988	423	21	1	0
3.	3	1987	428	21	.75	0
4.	3	1988	386	21	1	0
93.	50	1987	377	21	.542	0
94.	50	1988	381	21	1	2
95.	51	1987	59	18	.667	0
96.	51	1988	69	20	1	0

- Datos de manejo en 48 estados durante 2 años
- Cada estado, dos observaciones
- Cada estado, un identificador
 - Número y orden irrelevantes
- Años adyacentes para cada estado, orden preferido para DP
 - Para transformar datos por estado

Datos panel (DP)

Observar mismas unidades en el tiempo, tiene ventajas sobre CT y CTI

- Permite controlar por características no observadas (individuos, empresas)
 - Se agrega una constante por unidad
- Tener más de una observación puede facilitar inferencia causal en situaciones en que inferir causalidad sería difícil con un solo CT
- Permite estudiar la importancia de los rezagos en el comportamiento o como resultado de la toma de decisiones
 - Cambios en políticas públicas pueden no tener efecto inmediato
- Análisis DP simple, no muy diferente a CT y permite un gran avance

Énfasis en CT

Curso se enfocará en CT porque:

- Plantea menores dificultades conceptuales y técnicas
- Ilustra muchos de los temas clave en el análisis econométrico
- Mismos métodos e intuición se utilizarán en siguientes semestres

No ST para no contaminar un tratamiento apropiado de sus particularidades

- Tendencias, persistencia, estacionalidad

CTI y DP simples son extensiones sencillas de CT

Causalidad

Correlación \neq Causalidad

Una relación (positiva o negativa) entre dos variables no implica causalidad

- Inferir causalidad conlleva retos (endogeneidad, aleatoriedad)

En general, correlación:

- Pronosticar ✓
- Inferir un efecto causal ✗

Pero si podemos encontrar variación exógena, entonces correlación es probablemente causalidad

- ¿Qué significa variación exógena?
- ¿Cómo distinguir aleatoriedad de causalidad?

Causalidad

Si variables A (observada) y B (de interés) varían, ¿ A causa B ? o

- Aleatoriedad: ¿Relación existe? Ej. divorcios y margarina
- Causalidad inversa: ¿ B causa A ? Ej. policías y delitos, camionetas y bebés
- Variable omitida (endogeneidad): ¿ C causa B ? Ej. shorts y helados

Una variable es **endógena** si sus cambios se relacionan con variables en u

- Otras variables que determinan a la variable de interés

Una variable es **exógena** si sus cambios no están relacionados con variables en u

Es difícil saber si una variable es endógena o exógena porque u no se observa

- Difícil establecer causalidad para variables endógenas

Experimentos aleatorios

Experimentos con un proceso de aleatorización (RCT) crean exogeneidad

- Para que la variable independiente (**tratamiento**) sea exógena
- El **grupo de control** no recibe la intervención

Ayudan a resolver el problema de endogeneidad (Premio Nobel 2019)

- Ej. Bolsas de arroz, desparasitación

Sin embargo, no siempre son posibles ni perfectos (Deaton, 2018)

- Pueden no ser generalizables (¿aplica más allá de la muestra?)
- Pueden no tener validez interna (¿inferencia sesgada?)
- Pueden no tener validez externa (¿aplica fuera del contexto?)

Ceteris paribus

Expresión latina que significa 'manteniendo todo lo demás constante'

- Ej. Curva de demanda fija ingreso, gustos, precios relacionados, etc.

Estudiaremos métodos para replicar estándar ideal de experimentos aleatorios

- Efecto parcial: Cambio esperado en una variable manteniendo fijas las otras

En ambientes no experimentales podemos replicar un ambiente controlado

- Damos una interpretación ceteris paribus
- Como si mantuviéramos otros factores fijos
- Aunque los datos no se hayan recolectado de esa forma