

Resumen microeconometria (Cameron & Trivedi)

Santiago Alonso-Díaz, PhD

Profesor Visitante
Departamento de Economía
Universidad Javeriana

Prólogo (ver Judea Pearl)

¿Qué significa $P(\text{Lluvia}|\text{Sol})$?

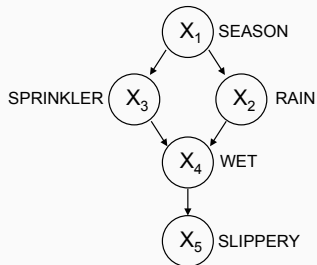
¿Correlación es causalidad?

¿Probabilidad condicional es causalidad?

Por ejemplo, si en país $P(\text{Lluvia}|\text{Sol}) = 0$ ¿Hay causalidad? ¿Sol causa no lluvia?

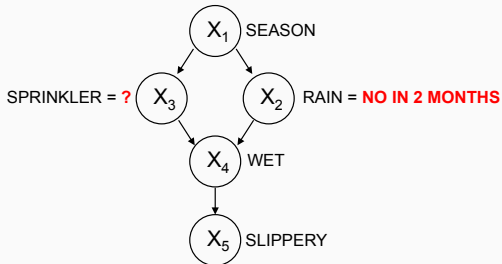
Modelo causal

¿Es X_2 independiente de X_3 ? ¿Si, no, depende?



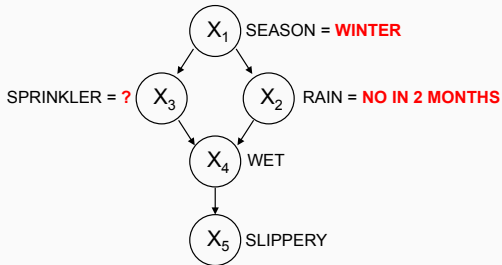
Modelo causal

¿Es X_2 independiente de X_3 ? ¿Si, no, depende?



Modelo causal

¿Es X_2 independiente de X_3 ? ¿Si, no, depende?

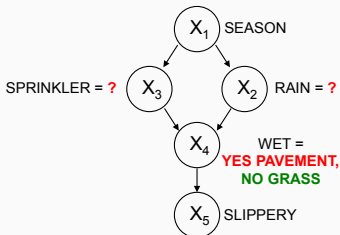


Modelo causal

¿Es X_2 independiente de X_3 ? ¿Si, no, depende?

Explaining away: una de las causas (X_2 o X_3) se vuelve menos probable dada la consecuencia (X_4)

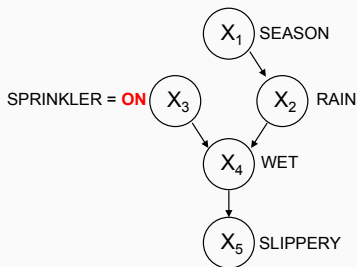
"Paradoja": NO es necesario un link directo entre X_2 y X_3



Modelo causal

¿Es X_2 independiente de X_3 ? ¿Si, no, depende?

Intervenciones tipo $\text{do}(X_3) = \text{ON}$ rompen el diagrama y pueden independizar variables



¿Qué intervención? ¿Medicar o no? Paradoja de Simpson: el efecto **grupal** es diferente al **individual**

Mujer & Hombre	Curado	No Curado	Total	Ratio Curación
Medicamento	20	20	40	50%
No medicamento	16	24	40	40%
	36	44	80	

Hombre	Curado	No Curado	Total	Ratio Curación
Medicamento	18	12	30	60%
No medicamento	7	3	10	70%
	25	15	40	

Mujer	Curado	No Curado	Total	Ratio Curación
Medicamento	2	8	10	20%
No medicamento	9	21	30	30%
	11	29	40	

¿INTERVENCION = CAUSALIDAD?

OPTOGENETICS VIDEO KONRAD PAPER

La pregunta que deben mantener en este curso (y en su carrera de economistas) es si funciones de este tipo se pueden interpretar como x causa y,

$$y = \beta x + u + \epsilon$$

y: outcome de interés
x: predictores del outcome
u: unobserved variables
 ϵ : errores estocásticos

The Three Layer Causal Hierarchy

Level (Symbol)	Typical Activity	Typical Questions	Examples
1. Association $P(y x)$	Seeing	What is? How would seeing X change my belief in Y ?	What does a symptom tell me about a disease? What does a survey tell us about the election results?
2. Intervention $P(y do(x), z)$	Doing Intervening	What if? What if I do X ?	What if I take aspirin, will my headache be cured? What if we ban cigarettes?
3. Counterfactuals $P(y_x x', y')$	Imagining, Retrospection	Why? Was it X that caused Y ? What if I had acted differently?	Was it the aspirin that stopped my headache? Would Kennedy be alive had Oswald not shot him? What if I had not been smok- ing the past 2 years?

Figure 1: The Causal Hierarchy. Questions at level i can only be answered if information from level i or higher is available.

Pearl (2018, <https://arxiv.org/pdf/1801.04016.pdf>)

Capítulo 1

- Microeconometria: análisis de datos a nivel individual:
 - Individuos
 - Hogares
 - Establecimientos (e.g. empresas)
- Objetivo: encontrar patrones de comportamiento económico
- Ejemplo tipo de datos hoy en día:
 - Datos de compras en supermercados
 - Viajes en aerolíneas
 - Redes sociales

- Fortalezas:

- ★ Datos discretos, no lineales (e.g. probit), y heterogeneos.

No son suaves como en datos macro donde la agregación reduce ruido.

Por ejemplo, consumo de carne promedio en una semana en una ciudad puede ser suave. Pero el consumo de carne de una persona no, puede variar mucho.

Por ejemplo 2: gasto en vacaciones en un país es en promedio positivo, pero a nivel individual no, muchos puede que gasten zero.

Por ejemplo 3: gasto en alcohol o cigarrillos en un país puede ser positivo pero muchos no fuman ni toman

- Fortalezas:

- ★ Más realistas que modelos macro.

Modelos macro usan agentes promedios o representativos (esconden variabilidad).

Por ejemplo, promedio matematicas exámenes Saber

Los modelos microeconomicos no asumen agentes promedio

- Fortalezas:

- ★ Más información que modelos macro.

Por ejemplo, entrevistas transversales independientes en cada corte.
Esto es, hay zero correlación (ver diapositivas de Ralf Haefner sobre correlación e información)

Para estudiar aspectos intertemporales se puede usar data de paneles o transición.

- Fortalezas:

- ★ Basada en fundamentos microeconómicos.

Por ejemplo, preferencias o relaciones tecnológicas.

Tiene técnicas para manejar heterogeneidad individual (e.g. Fixed vs random effects.)

Por ejemplo, $\text{ingreso} = \text{educacion} + \text{error} \dots$ el beta para educacion puede ser positivo pero sesgado por variabilidad de una variable no incluida como habilidad.

Capítulo 2: Causal Models

Cortes transversales
vs.
Longitudinal

Descripción (e.g. medias)

vs.

Causalidad (relaciones estructurales)

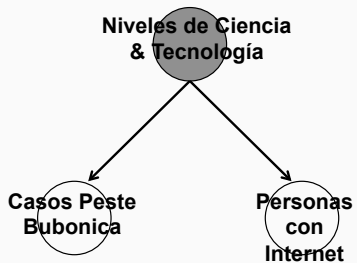
Correlación (descripción)

Correlación -

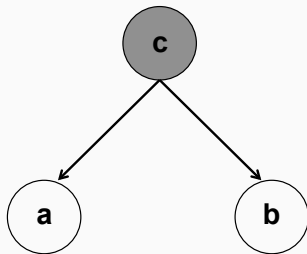
**Casos Peste
Bubonica**

**Personas
con
Internet**

Causalidad

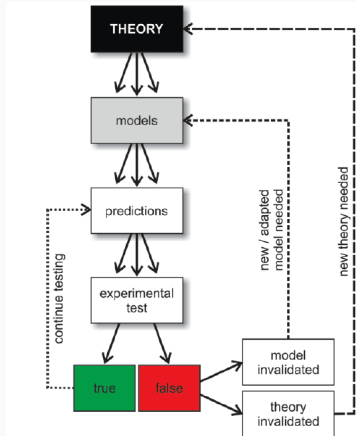


Causalidad



¿Qué es un modelo?

Modelos estructurales



[Blohm et al., 2017]

Cosas prácticas

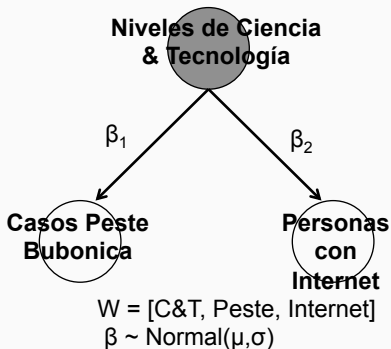
- ▶ Objetivo/Pregunta (e.g. AI en niveles de empleos en bancos)
- ▶ Literatura previa
- ▶ Herramientas que necesito (e.g. tipo de regresión)
- ▶ Variables a medir y variables latentes
- ▶ HIPÓTESIS que relacionen (matemáticamente) variables del paso anterior
- ▶ Implementar y evaluar el modelo (e.g. R^2 , AIC, BIC, WAIC, etc.)

¿Qué es una estructura?

Estructura

- ▶ Set de variables W (... partidas en $[Y\ Z]$)
- ▶ Distribución de probabilidad conjunta $p(W)$
- ▶ Relaciones hipotéticas de causa y efecto en W
- ▶ Formas funcionales y restricciones en los parámetros del modelo.

Ejemplo



$$y = f(z_i, \mu_i | \pi)$$

y: observable variables
z: explanatory variables
 μ : random disturbance
 π : model parameters

$$Peste = \beta_1 Ciencia + \mu$$

Exógeno (variables que el modelo no explica e.g. género)

vs

Endógeno (variables explicadas e.g. Ingreso)

Definición formal ...

... pero primero recordemos que buscamos la probabilidad conjunta de $W = [YZ]$...

... es decir, que ocurran tanto los datos Y & las variables explicativas Z

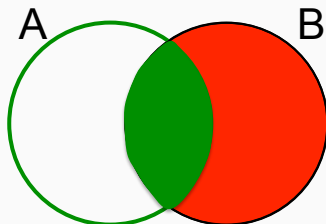
Probabilidad conjunta (definición)

$$p(A, B) = p(A|B) \times p(B)$$

Visualización

$$p(A|B) = \frac{p(A \cap B)}{p(B)}$$

$$= \frac{\text{Green Square}}{\text{Red Square}}$$



La probabilidad conjunta de $W = [Y \ Z]$

$$p(W|\theta) = p(Y|Z, \theta) \times p(Z|\theta)$$

θ : vector de parámetros

Ahora partamos el espacio de parámetros en sets INDEPENDIENTES $\theta : [\theta_1, \theta_2]$ de tal forma que,

$$p(W|\theta) = p(Y|Z, \theta_1) \times p(Z|\theta_2)$$

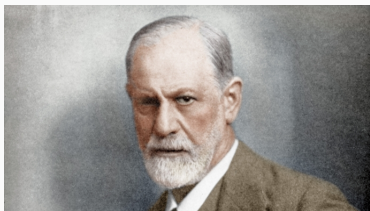
Definición: como θ_1 , el vector de parametros para Y , es independiente de θ_2 , decimos que Z es exogena.

En palabras simples: Cualquiera que sea la fuente de Z (i.e. de $p(Z|\theta_2)$), eso no afecta mis estimativos de los parámetros de mi modelo θ_1 .

Otra forma de decirlo: Mis variables explicativas Z no son afectadas por los parámetros (θ_1) que explican la variable de interés.

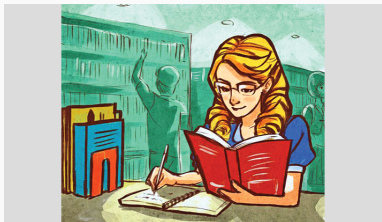
¿Por qué importa?

¿Citaciones dependen de prestigio o al revés?



$$Citaciones = \beta_1 Prestigio + \mu$$

¿Notas dependen de prestigio o al revés?



$$Notas = \beta_1 Prestigio + \mu$$

¿Comisiones dependen de prestigio o al revés?



$$\text{Comisiones} = \beta_1 \text{Prestigio} + \mu$$

La falta de exogeneidad afecta la estimación de parámetros del modelo

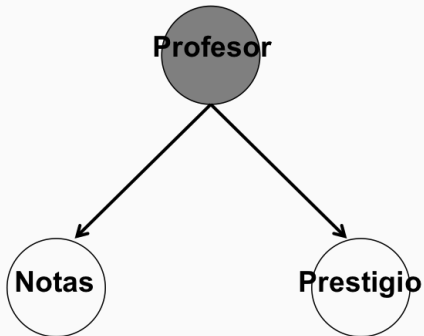
Tal vez más importante, limita cualquier conclusión de causalidad, el modelo se acerca a ser más descriptivo

Independencia condicional ... notas y prestigio pueden correlacionarse

Correlación +



Independencia condicional ... por las características del profesor se vuelven independientes



¿Cómo saber si una variable es exógena? Arte y ciencia

Por diseño experimental (e.g. tratado vs no tratado)

Por conocimiento del área (e.g. videojuegos y violencia)

Por estados de naturaleza (e.g. liberal/conservador; hombre/mujer)

Ejemplo: Simultaneous Equations Model (SEM)

Structural form (endógeno + exógeno - ruido = 0)

$$YB + Z\Gamma = U$$

with,

$$Y = \begin{bmatrix} y_1 \\ \vdots \\ y_N \end{bmatrix}, \quad Z = \begin{bmatrix} z_1 \\ \vdots \\ z_N \end{bmatrix}, \quad U = \begin{bmatrix} u_1 \\ \vdots \\ u_N \end{bmatrix}$$

Ejemplo: Simultaneous Equations Model (SEM)

Reduced form (variable endógena en función de exógenas y desviaciones)

$$\begin{aligned} YB + Z\Gamma &= U \\ Y + Z\Gamma B^{-1} &= UB^{-1} \\ Y &= Z\Pi + V \end{aligned}$$

with, $\Pi = -\Gamma B^{-1}$ and $V = UB^{-1}$

Ejemplo: Simultaneous Equations Model (SEM)

Relaciones causales en SEM (ejemplo)

Ecuaciones estructurales

$$y_1 = \gamma_1 + \beta_1 y_2 + u_1$$

$$y_2 = y_1 + z_1$$

Ejemplo: Simultaneous Equations Model (SEM)

$$\begin{aligned} \text{ingreso} &= \gamma_1 + \beta_1 \text{educacion}_{\text{nivel}} + u_1 \\ \text{educacion}_{\text{nivel}} &= f(\text{ingreso}) + \text{tratamiento} \end{aligned}$$

¿ β_1 mide efectos causales de *educacion* en *ingreso*?

Sí, pero como paso intermedio. Es el tratamiento (exógeno) la causa definitiva (e.g. gemelos separados al nacer con el mismo nivel de educación)

Ejemplo 2: Potential Outcome Model (POM)

- ▶ ¿Cuál es el efecto de una política pública en algún resultado?
- ▶ Experimento social (e.g. a un grupo política X al otro no)
- ▶ (quasi) experimento natural (e.g. mina en un pueblo X y en otro no)
- ▶ Contrafactuales (e.g. si no hubiera cambiado nada)

Ejemplo 2: Potential Outcome Model (POM)

Toda inferencia causal requiere comparar un factual contra un contrafactual

Factual: Transferencia de ingreso aumenta bienestar

Contrafactual: Sin transferencia de ingreso aumento bienestar

Ejemplo 2: Potential Outcome Model (POM)

Modelo Causal de Rubin

Cuando se recibe el tratamiento D (e.g. transferencia de ingreso) el resultado y (e.g. bienestar) es diferente

$$y_i = \begin{cases} y_{1i} & \text{if } D_i = 1 \\ y_{0i} & \text{if } D_i = 0 \end{cases}$$

El efecto promedio de tratamiento es,

$$E[y|D=1] - E[y|D=0]$$

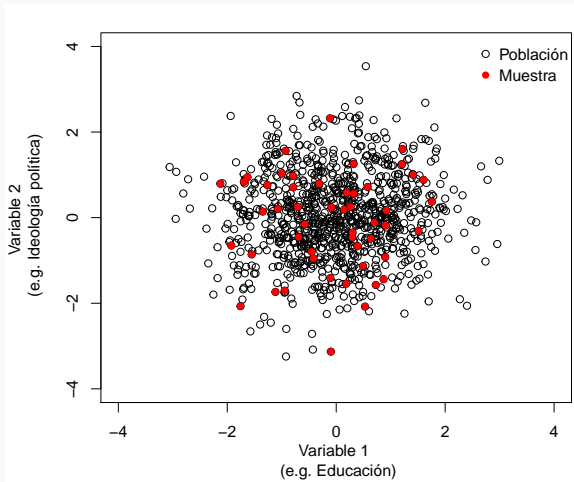
Para causalidad, pertenencia al grupo de tratamiento y no tratamiento debe ser ALEATORIO

Capítulo 3: Estructura de Datos

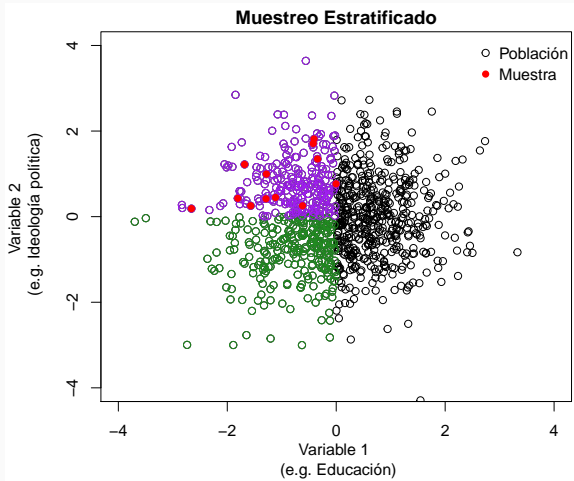
- ▶ Observacional: encuestas y censos
- ▶ Experimental: e.g. becas a un grupo y a otro no

Observacional

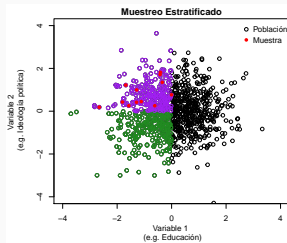
Muestra aleatoria para una encuesta



Muestra por etapas e.g. estratificado: samplear subpoblaciones

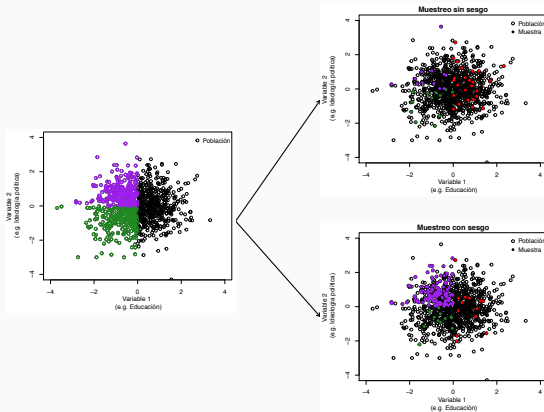


En muestreos estratificados hay que tener claridad que la muestra no es representativa de la población e.g. solo son morados



Observacional

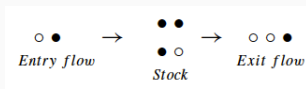
Preguntarse si hay sesgo i.e. distribución de la muestra ($f_w|\theta$) \neq la población ($F_w|\theta$). Problema: e.g. medias se sesgan.



w : variables de interés; θ : parámetros de la distribución

Fuentes de sesgo en el muestreo

- ▶ Muestreo por variable exógena (e.g. por género 50/50 pero nos interesa algo sobre ciclos menstruales)
- ▶ Muestreo por variable endógena (e.g. solo usuarios de transporte público cuando nos interesa qué transporte público se preferiría en la población de la ciudad)
- ▶ Muestreo por duración (e.g. desempleados con poca rotación i.e. puntos negros en stock)



- Slow movers
- Fast movers

Fuentes de sesgo en el muestreo

- ▶ Encuestados no responden (e.g. z : entrenamiento (0,1); y : productividad; x : ccs del trabajador. Solo hay prod. para $\text{entren.}=1$)
- ▶ Errores de medida (e.g. no comprensión de la pregunta; errores de procesamiento de la data)
- ▶ Perdida de muestra (e.g. en un estudio longitudinal de 20 años perder sujetos)

z : variable exógena; y : respuesta; x : características de la muestra

Tipos de data observacional

- ▶ Data transversal (e.g. ¿hoy qué comió? Los datos solo se recogen en solo UN punto del tiempo)
- ▶ Data transversal repetida (e.g. ¿hoy qué comió? En varios puntos de tiempo a DIFERENTES sujetos)
- ▶ Longitudinal o panel (e.g. ¿hoy qué comió? En varios puntos de tiempo a los MISMOS sujetos)

Experimentos Sociales

Experimentos sociales

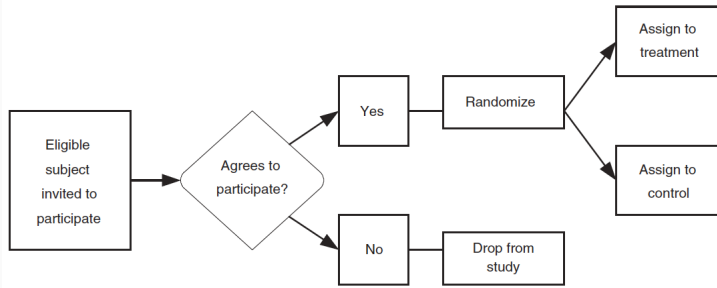


Table 3.1. *Features of Some Selected Social Experiments*

Experiment	Tested Treatments	Target Population
Rand Health Insurance Experiment (RHIE), 1974–1982	Health insurance plans with varying copayment rate and differing levels of maximum out-of-pocket expenses	Low- and moderate-level income persons and families
Negative Income Tax (NIT), 1968–1978	NIT plans with alternative income guarantees and tax rates	Low- and moderate-level income persons and families with nonaged head of household
Job Training Partnership Act (JTPA), (1986–1994)	Job search assistance, on-the-job training, classroom training financed under JTPA	Out-of-school youths and disadvantaged adults

Ventajas

- ▶ Aleatorización remueve correlaciones
- ▶ Exogenizar una política (e.g. uso de servicios de salud en personas con o sin seguro)

Limites

- ▶ Altos costos comprometen aleatorización (e.g. no poder muestrear estratos altos)
- ▶ Ética de seleccionar una muestra (e.g. dar becas a unos y otros no)
- ▶ Perder sujetos (attrition)
- ▶ Sujetos se adaptan por muchos motivos (e.g. efecto de beca puede ser por otro evento)
- ▶ Adaptación por etiquetar (e.g. recibir una beca cambia mi comportamiento normal)
- ▶ Generalizar a población no es tan fácil

Experimentos Naturales

Table 3.2. *Features of Some Selected Natural Experiments*

Experiment	Treatments Studied	Reference
Outcomes for identical twins with different schooling levels	Differences in returns to schooling through correlation between schooling and wages	Ashenfelter and Krueger (1994)
Transition to National Health Insurance in Canada as Saskatchewan moves to NHI and other states follow several years later	Labor market effects of NHI based on comparison of provinces with and without NHI	Gruber and Hanratty (1995)
New Jersey increases minimum wage while neighboring Pennsylvania does not	Minimum wage effects on employment	Card and Krueger (1994)

$$y = \beta_1 + \beta_2 x + u$$

Un experimento natural puede exogenizar x

$$Ingreso = \beta_1 + \beta_2 Educacion + u$$

Un experimento natural, gemelos idénticos con distintos niveles de educación, exogeniza educación

Diferencia en diferencias (D&D).

Primero las diferencias:

$$y_{it} = \alpha + \beta D_t + \epsilon_{it}$$

$i = 1, \dots, N$; $t = 0, 1$

D es una dummy. $D_t = 0$ (pre-intervention); $D_t = 1$ (post-intervention)

$\hat{\beta} = \bar{y}_1 - \bar{y}_0$ i.e. impacto de la política

IDENTIFICABILIDAD i.e. con observaciones infinitas obtengo el parámetro de verdad

En este ejemplo, para obtener $\hat{\beta}$ identificabilidad requiere que el grupo sea comparable en los dos periodos de tiempo

$$\hat{\beta} = \bar{y}_1 - \bar{y}_0$$

Si el grupo cambia, no es posible identificar el parámetro por que el efecto de la política puede causarse por ese cambio.

Ahora si diferencia en diferencias (D&D): Hay dos grupos uno tratado y otro no. Se miden ambos antes y después del tratamiento.

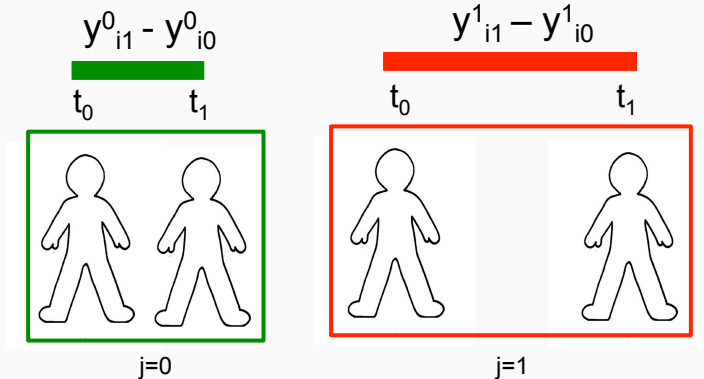
$$y_{it}^j = \alpha + \alpha_1 D_t + \alpha_1 D^j + \beta D_t^j + \epsilon_{it}^j$$

Tenemos 3 dummies. D_t pre y post intervención.

D^j tratado y no tratado.

D_t^j tratado post intervención y los demás.

Experimentos naturales



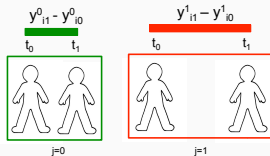
Experimentos naturales

$$(y_{i1}^1 - y_{i0}^1) - (y_{i1}^0 - y_{i0}^0) = \beta + (\epsilon_{i1}^1 - \epsilon_{i0}^1) - (\epsilon_{i1}^0 - \epsilon_{i0}^0)$$

Dado que el valor esperado de los errores es zero i.e.

$E[(\epsilon_{i1}^1 - \epsilon_{i0}^1) - (\epsilon_{i1}^0 - \epsilon_{i0}^0)] = 0$, tenemos que

$\beta = (y_{i1}^1 - y_{i0}^1) - (y_{i1}^0 - y_{i0}^0)$ i.e. impacto de la política = Rojo menos Verde



Experimentos naturales



[Ashenfelter and Krueger, 1994]

Experimentos naturales

TABLE 1—DESCRIPTIVE STATISTICS

Variable	Means (standard deviations in parentheses)		Population ^b
	Identical twins ^a	Fraternal twins ^a	
Self-reported education	14.11 (2.16)	13.72 (2.01)	13.14 (2.73)
Sibling-reported education	14.02 (2.14)	13.41 (2.07)	—
Hourly wage	\$13.31 (11.19)	\$12.07 (5.40)	\$11.10 (7.41)
Age	36.56 (10.36)	35.59 (8.29)	38.91 (12.53)
White	0.94 (0.24)	0.93 (0.25)	0.87 (0.34)
Female	0.54 (0.50)	0.48 (0.50)	0.45 (0.50)
Self-employed	0.15 (0.36)	0.10 (0.30)	0.12 (0.32)
Covered by union	0.24 (0.43)	0.30 (0.46)	—
Married	0.45 (0.50)	0.54 (0.50)	0.62 (0.48)
Age of mother at birth	28.27 (6.37)	29.38 (7.05)	—
Twins report same education	0.49 (0.50)	0.43 (0.50)	—
Twins studied together	0.74 (0.44)	0.38 (0.49)	—
Helped sibling find job	0.43 (0.50)	0.24 (0.43)	—
Sibling helped find job	0.35 (0.48)	0.22 (0.41)	—
Sample size	298	92	164,085

^aSource: Twinsburg Twins Survey, August 1991.

^bSource: 1990 Current Population Survey (Outgoing Rotation Groups File). Sample includes workers aged 18–65 with an hourly wage greater than \$1.00 per hour.

Ingreso (y) de gemelo 1 en familia i

$$y_{1i} = \alpha X_i + \beta Z_{1i} + \mu_i + \epsilon_{1i}$$

Ingreso (y) de gemelo 2 en familia i

$$y_{2i} = \alpha X_i + \beta Z_{2i} + \mu_i + \epsilon_{2i}$$

X: características familiares (e.g. salario padre y madre)

Z: características individuales de cada gemelo (e.g. nivel de educación de gemelo 1)

μ, ϵ : características no observadas de la familia o de cada gemelo, respectivamente

Diferencia en diferencias

$$(y_{1i} - y_{2i}) = \beta(Educacion_{1i} - Educacion_{2i}) + \epsilon_{1i} - \epsilon_{2i}$$

Experimentos naturales

El estimativo de diferencia en diferencias (iii o iv en la tabla) indica que mayor educación lleva a mayores salarios en 1994 cuando se publicó el estudio

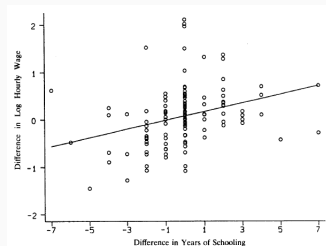


FIGURE 1. INTRAPAIR RETURNS TO SCHOOLING, IDENTICAL TWINS

TABLE 3—ORDINARY LEAST-SQUARES (OLS), GENERALIZED LEAST-SQUARES (GLS), INSTRUMENTAL-VARIABLES (IV), AND FIXED-EFFECTS ESTIMATES OF LOG WAGE EQUATIONS FOR IDENTICAL TWINS^a

Variable	OLS (i)	GLS (ii)	GLS (iii)	IV ^a (iv)	First difference (v)	First difference by IV (vi)
Own education	0.084 (0.014)	0.087 (0.015)	0.088 (0.015)	0.116 (0.030)	0.092 (0.024)	0.167 (0.043)
Sibling's education	—	—	-0.007 (0.015)	-0.037 (0.029)	—	—
Age	0.088 (0.019)	0.090 (0.023)	0.090 (0.023)	0.088 (0.019)	—	—
Age squared (÷ 100)	-0.087 (0.023)	-0.089 (0.028)	-0.090 (0.029)	-0.087 (0.024)	—	—
Male	0.204 (0.063)	0.204 (0.077)	0.206 (0.077)	0.206 (0.064)	—	—
White	-0.410 (0.127)	-0.417 (0.143)	-0.424 (0.144)	-0.428 (0.128)	—	—
Sample size:	298	298	298	298	149	149
R ² :	0.260	0.219	0.219	—	0.092	—

Notes: Each equation also includes an intercept term. Numbers in parentheses are estimated standard errors.

^aOwn education and sibling's education are instrumented for using each sibling's report of the other sibling's education as instruments.

Experimentos naturales

TABLE 3—ORDINARY LEAST-SQUARES (OLS), GENERALIZED LEAST-SQUARES (GLS), INSTRUMENTAL-VARIABLES (IV), AND FIXED-EFFECTS ESTIMATES OF LOG WAGE EQUATIONS FOR IDENTICAL TWINS^a

Variable	OLS (i)	GLS (ii)	GLS (iii)	IV ^a (iv)	First difference (v)	First difference by IV (vi)
Own education	0.084 (0.014)	0.087 (0.015)	0.088 (0.015)	0.116 (0.030)	0.092 (0.024)	0.167 (0.043)
Sibling's education	—	—	-0.007 (0.015)	-0.037 (0.029)	—	—
Age	0.088 (0.019)	0.090 (0.023)	0.090 (0.023)	0.088 (0.019)	—	—
Age squared (÷ 100)	-0.087 (0.023)	-0.089 (0.028)	-0.090 (0.029)	-0.087 (0.024)	—	—
Male	0.204 (0.063)	0.204 (0.077)	0.206 (0.077)	0.206 (0.064)	—	—
White	-0.410 (0.127)	-0.417 (0.143)	-0.424 (0.144)	-0.428 (0.128)	—	—
Sample size:	298	298	298	298	149	149
R ² :	0.260	0.219	0.219	—	0.092	—

Notes: Each equation also includes an intercept term. Numbers in parentheses are estimated standard errors.

^aOwn education and sibling's education are instrumented for using each sibling's report of the other sibling's education as instruments.

¿Qué es una variable instrumental?

Es una variable que ayuda a detectar causalidad en vez de correlación

En $y = \beta x + \text{error}$ sería añadir un predictor que,

1. Se correlaciona con x
2. NO se correlaciona con error
3. Se correlaciona con y por medio de x

Otro ejemplo de variable instrumental (tomado de Wikipedia)

$$\text{Salud} = \beta \text{ Fumar} + \text{error}$$

Fumar y salud pueden ser endógenas e.g. estrés lleva a fumar; no fumar causa estrés.

$$\text{Salud} = \beta \text{Impuesto}_{\text{tabaco}} + \text{error}$$

Impuesto es una variable instrumental pues se correlaciona con fumar, afecta salud via decisión de fumar, y no forma parte del error (i.e. no afecta salud de forma obvia)

Experimentos naturales

TABLE 3—ORDINARY LEAST-SQUARES (OLS), GENERALIZED LEAST-SQUARES (GLS), INSTRUMENTAL-VARIABLES (IV), AND FIXED-EFFECTS ESTIMATES OF LOG WAGE EQUATIONS FOR IDENTICAL TWINS^a

Variable	OLS (i)	GLS (ii)	GLS (iii)	IV ^a (iv)	First difference (v)	First difference by IV (vi)
Own education	0.084 (0.014)	0.087 (0.015)	0.088 (0.015)	0.116 (0.030)	0.092 (0.024)	0.167 (0.043)
Sibling's education	—	—	-0.007 (0.015)	-0.037 (0.029)	—	—
Age	0.088 (0.019)	0.090 (0.023)	0.090 (0.023)	0.088 (0.019)	—	—
Age squared (÷ 100)	-0.087 (0.023)	-0.089 (0.028)	-0.090 (0.029)	-0.087 (0.024)	—	—
Male	0.204 (0.063)	0.204 (0.077)	0.206 (0.077)	0.206 (0.064)	—	—
White	-0.410 (0.127)	-0.417 (0.143)	-0.424 (0.144)	-0.428 (0.128)	—	—
Sample size:	298	298	298	298	149	149
R ² :	0.260	0.219	0.219	—	0.092	—

Notes: Each equation also includes an intercept term. Numbers in parentheses are estimated standard errors.

^aOwn education and sibling's education are instrumented for using each sibling's report of the other sibling's education as instruments.

Instrumentos son debatibles
DISCUTIR: ¿es el reporte que cada gemelo da sobre la educación de su hermano un instrumento valido?

Datos: algunas fuentes y consideraciones

Algunas fuentes de datos

- ▶ Panel Study in Income Dynamics (PSID) (Brown, Duncan, Stafford (1996))
- ▶ Current Population Survey (CPS) (Polivka, 1996)
- ▶ National Longitudinal Survey (NLS)
- ▶ National Longitudinal Surveys of Youth (NLSY)
- ▶ Survey of Income and Program Participation (SIPP)
- ▶ Health and Retirement Study (HRS)
- ▶ World Bank's Living Standards Measurement Study (LSMS)
- ▶ U.S. National Center for Health Statistics
- ▶ Journals e.g. Journal of Applied Econometrics
- ▶ Kaggle (varias bases de datos)

Revisar/arreglar la data (NO MANIPULAR)

- ▶ Celdas sin información
- ▶ Errores al teclear la información
- ▶ Remover códigos (eg. 999 para información perdida)
- ▶ Revisar escalas (e.g. z-transform)
- ▶ CONOZCA SUS DATOS; haga tablas y gráficas con estadísticas descriptivas

Intermedio: Tutorial R

PRIMERO UN TUTORIAL DE R

(<https://www.statmethods.net/r-tutorial/index.html>)



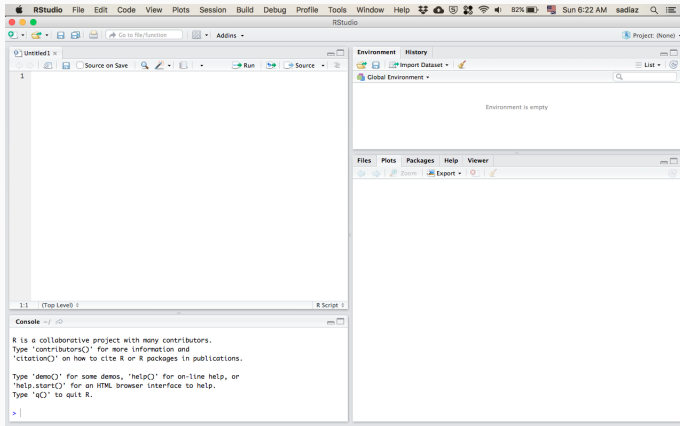
Bajar R

<https://cran.r-project.org/>

Bajar R-Studio

<https://www.rstudio.com/products/rstudio/download/>

Ambiente Rstudio



Cuadro 1: Operadores

Operador	Descripción
+	Suma
-	Resta
*	Multiplicación
÷	División
% %	Modulo
^, **	Exponenciación

Cuadro 2: Operadores Lógicos

Operador	Descripción
>	Mayor
>=	Mayor o igual
<	Menor
<=	Menor o igual
==	Igual
!=	Diferente

Tipos de datos

Character, Numeric, Integer, Complex, Logical, Factors

Estructuras de datos

Vector, List, Matrix, Data frame

Condicional

Loops (For y While)

Funciones

Datos (Importar, Indexar, y Formatear)

Graficar (ir a Tutorial.R)

Capítulo 3 (Stata book) & 4 (Other book): Modelos Lineales

Ejemplo: retornos de educación

$$\ln(\text{salario}_i) = \alpha \text{educacion}_i + \beta \text{controles}_i + u_i$$

- ▶ La transformación a logaritmo es opcional, pero puede ayudar a producir errores homoskedastic (i.e. que no depende de nivel de educación)
- ▶ α puede interpretarse directamente en ESTA regresión e.g. si $\hat{\alpha} = 0,1$, un año de educación aumenta 10% el salario
- ▶ En ESTA regresión educación es endógena, no es aleatoria, no se puede concluir causalidad.

Objetivo: estimar parámetros α y β

$$\ln(\text{salario}_i) = \alpha \text{educacion}_i + \beta \text{controles}_i + u_i$$

Método: minimizar perdida (L) esperada

$$\min_{\hat{y}} E[L(y - \hat{y})]$$

L puede cambiar pero la más conocida es $L = \sum_{i=1}^N (y_i - \beta x)$

Estimador OLS (ordinary least squares)

$$\sum_{i=1}^N u_i^2 = (y - \beta X)'(y - \beta X)$$

Ponemos la derivada con respecto a β igual a zero, obtenemos el estimador OLS,

$$\hat{\beta}_{OLS} = (X'X)^{-1}X'y$$

El estimador puede ser robusto ante heteroscedasticidad i.e. no es crítico que el error dependa de x . Pero hay que asegurarse que el programa (e.g. R) haga los supuestos correctos.

Estimadores que incluyen la varianza del error (ir a LinearRegression.R).

Table 4.2. *Least-Squares Estimators and Their Asymptotic Variance*

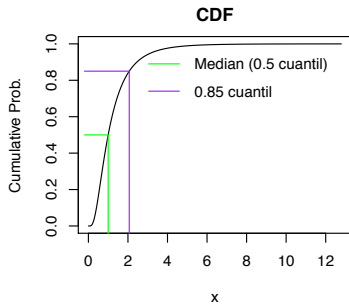
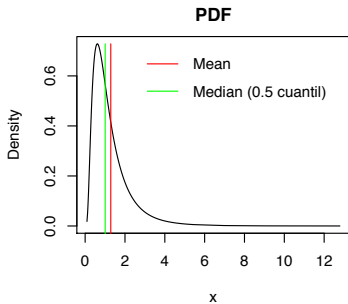
Estimator ^a	Definition	Estimated Asymptotic Variance
OLS	$\hat{\beta} = (\mathbf{X}'\mathbf{X})^{-1} \mathbf{X}'\mathbf{y}$	$(\mathbf{X}'\mathbf{X})^{-1} \mathbf{X}'\hat{\Omega}\mathbf{X} (\mathbf{X}'\mathbf{X})^{-1}$
FGLS	$\hat{\beta} = (\mathbf{X}'\hat{\Omega}^{-1}\mathbf{X})^{-1} \mathbf{X}'\hat{\Omega}^{-1}\mathbf{y}$	$(\mathbf{X}'\hat{\Omega}^{-1}\mathbf{X})^{-1}$
WLS	$\hat{\beta} = (\mathbf{X}'\hat{\Sigma}^{-1}\mathbf{X})^{-1} \mathbf{X}'\hat{\Sigma}^{-1}\mathbf{y}$	$(\mathbf{X}'\hat{\Sigma}^{-1}\mathbf{X})^{-1} \mathbf{X}'\hat{\Sigma}^{-1}\hat{\Omega}\hat{\Sigma}^{-1}\mathbf{X}(\mathbf{X}'\hat{\Sigma}^{-1}\mathbf{X})^{-1}$.

^a Estimators are for linear regression model with error conditional variance matrix Ω . For FGLS it is assumed that $\hat{\Omega}$ is consistent for Ω . For OLS and WLS the heteroskedastic robust variance matrix of $\hat{\beta}$ uses $\hat{\Omega}$ equal to a diagonal matrix with squared residuals on the diagonals.

LinearRegression.R; Sección Lineal: Ejemplo problemas con heteroscedasticidad



NUEVO TEMA: Modelos de medianas y cuantiles



Nos interesa el cuantil q de $y|x$ e.g. el cuantil 0.95 de ingresos dado un nivel de educación.

$$\mu_q(x) = F_{y|x}^{-1}(q)$$

$F_{y|x}(q)$ es la CDF de $y|x$. La CDF nos da una probabilidad pero nos interesa la inversa i.e. el valor de x que corresponde a esa probabilidad-cuantil.

Por ejemplo

$$\mu_{0,95}(\text{pregrado}) = F_{\text{ingresos}|\text{pregrado}}^{-1}(0,95)$$

Ahora definimos un modelo con heteroscedasticidad (i.e. con error u que aumenta con x),

$$y = x'\beta + u$$

$$u = x'\alpha \times \epsilon$$

$$\epsilon \sim iid[0, \sigma^2]$$

Ahora obtenemos $\mu_q(x, \beta\alpha)$. El cuantil q por definición es,

$$q = Pr[y \leq \mu_q(x, \beta\alpha)]$$

Reemplazando y obtenemos

$$\begin{aligned} q &= Pr[y \leq \mu_q(x, \beta\alpha)] \\ &= Pr[u \leq \mu_q(x, \beta\alpha) - x'\beta] \end{aligned}$$

Reemplazando u obtenemos

$$\begin{aligned} q &= Pr[y \leq \mu_q(x, \beta\alpha)] \\ &= Pr[u \leq \mu_q(x, \beta\alpha) - x'\beta] \\ &= Pr[\epsilon \leq (\mu_q(x, \beta\alpha) - x'\beta)/x'\alpha] \end{aligned}$$

Es decir, encontramos que q es la CDF del error ϵ evaluada en una función de α & β

$$\begin{aligned} q &= Pr[y \leq \mu_q(x, \beta\alpha)] \\ &= Pr[u \leq \mu_q(x, \beta\alpha) - x'\beta] \\ &= Pr[\epsilon \leq (\mu_q(x, \beta\alpha) - x'\beta)/x'\alpha] \\ &= F_\epsilon[(\mu_q(x, \beta\alpha) - x'\beta)/x'\alpha] \end{aligned}$$

Ahora estamos en posición de encontrar $\mu_q(x, \beta\alpha)$ invirtiendo F_e

$$\begin{aligned} q &= Pr[y \leq \mu_q(x, \beta\alpha)] \\ &= Pr[u \leq \mu_q(x, \beta\alpha) - x'\beta] \\ &= Pr[\epsilon \leq (\mu_q(x, \beta\alpha) - x'\beta)/x'\alpha] \\ &= F_e[(\mu_q(x, \beta\alpha) - x'\beta)/x'\alpha] \end{aligned}$$

$$F_e^{-1}(q) = (\mu_q(x, \beta\alpha) - x'\beta)/x'\alpha$$

$$\mu_q(x, \beta\alpha) = x'(\beta + \alpha \times F_e^{-1}(q))$$

Es decir, los cuantiles son lineales en x con la heteroscedasticidad que definimos i.e. $u = x' \alpha \times \epsilon$

$$\mu_q(x, \beta \alpha) = x' (\beta + \alpha \times F_e^{-1}(q))$$

Puede haber heteroscedasticidad no lineal pero es usual usar la aproximación lineal

La función objetivo a minimizar en β_q

$$Q_N(\beta_q) = \sum_{i: y_i \geq x_i' \beta} q |y_i - x_i' \beta_q| + \sum_{i: y_i < x_i' \beta} (1 - q) |y_i - x_i' \beta_q|$$

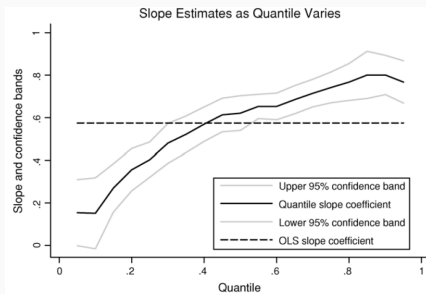
Esta función no es intuitiva, hay que obtenerla (ver Koenker and Bassett (1978)). Lo importante es que este es el costo entre data (y) y modelo lineal ($x_i' \beta$) a minimizar.

IMPORTANTE: no es diferenciable (valores absolutos), hay que usar métodos numéricos.

LinearRegression.R; Sección lineal cuantiles: ejemplo



$Gasto_{total} \sim Intercepto + \beta Gasto_{salud} + \epsilon$ en Vietnam



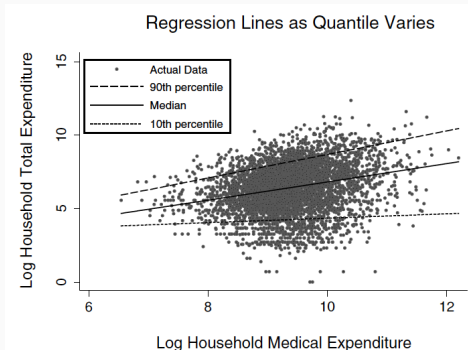
¿Qué quiere decir una pendiente = 0.57? ¿Elástico o inelástico?

¿Una pendiente menor en $q = 0.1$ y mayor en $q = 0.9$?

Modelos Lineales

El gasto en salud es inelástico pero en menor medida en cuantiles con ingreso mayores.

Pendientes diferentes por cuantil es evidencia de heteroscedasticidad o no-linealidad



Variables instrumentales

Aplicaciones:

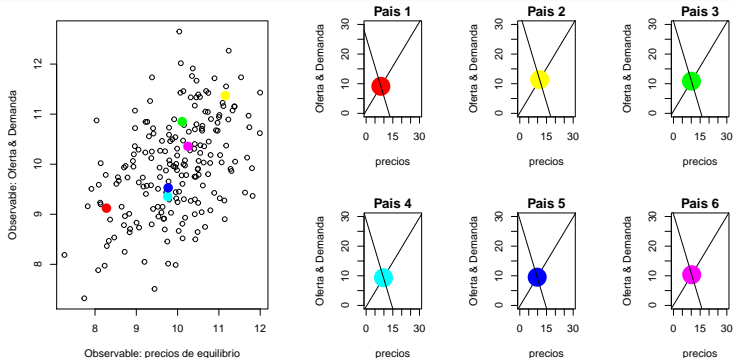
- ▶ Simultaneidad
- ▶ Errores de medición
- ▶ Variables omitidas (la aplicación más importante ... para economistas por lo menos).

Simultaneidad

Modelos Lineales

Solo observamos precios de equilibrio (panel izq.), que vienen de dos ecuaciones simultáneas (paneles der.)

¿Cómo estimar la pendiente de oferta y demanda con observables?

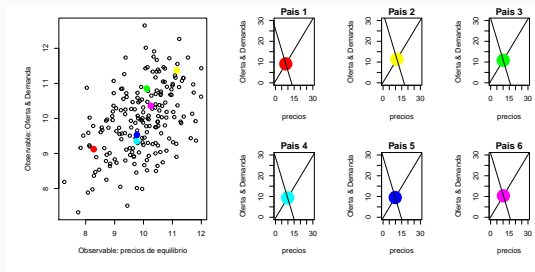


Modelos Lineales

Solución: introduzco variables instrumentales e.g. interceptos diferentes para oferta y demanda (punto-intercepto)

Por ejemplo, política industrial desplaza el intercepto para arriba.

Mientras que política de salud no lo desplaza



Variable omitida / Causalidad

Problemas de especificación i.e. el modelo no refleja el proceso causal que generó la data

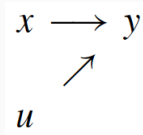
- ▶ La forma funcional de y no es lineal
- ▶ Endogeneidad (correlación no causalidad)
- ▶ Variable omitida
- ▶ Heterogeneidad de parámetros (random effects)

Modelos Lineales

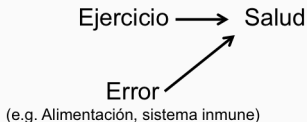
¿Qué asume este modelo? x causa y en proporción dada por β

$$y = \beta x + u$$

¿Están de acuerdo?

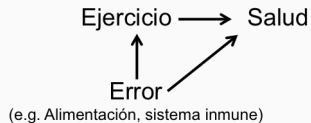
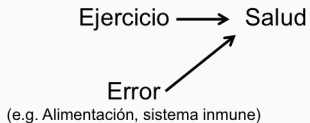


¿Están de acuerdo?



¿Cuál es la estructura causal?

¿Qué implica para la regresión la de la derecha?



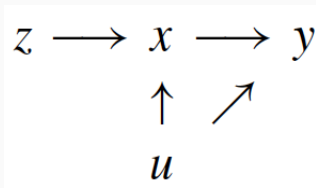
x afecta u (el ruido). El dato y cambia en función de ambos

$$y = \beta x + u(x)$$

$$\frac{dy}{dx} = \beta + \frac{du}{dx}$$

El estimador OLS de β esta sesgado, incluye información proveniente del error. No es posible interpretarlo como fuerza causal, solo como correlación.

Solución: variable instrumental Z



Características

- ▶ z no está correlacionada con u
- ▶ z se correlaciona con x

LinearRegression.R; Sección OLS is inconsistent



Ejemplos variables instrumentales (z)

¿Por qué hallazgo de pozos? Use las dos condiciones de VI

- ▶ $y = \text{precio}_{\text{petroleo}}$; $x = \text{demanda}_{\text{petroleo}}$; $z : \text{hallazgo pozos}$

QUIZ

¿Por qué cercanía a la escuela? Use las dos condiciones de VI

- ▶ $y = \text{Salario}$; $x = \text{Educacion}_{\text{nivel}}$; $z : \text{cercana escuela}$



Ashenfelter, O. and Krueger, A. (1994).

Estimates of the economic return to schooling from a new sample of twins.

The American Economic Review, 84(5):1157–1173.



Blohm, G., Schrater, P., and Kording, K. (2017).

Cosmo 2017.

Cosmo, 1(1):1.