Econometría I
Modelo Lineal de Regresión: Estimación por MCO

Ramiro de Elejalde

Facultad de Economía y Finanzas
Universidad Alberto Hurtado

Outline

- Motivación: Efecto del tamaño de clase en el rendimiento académico (California)
- Modelo de regresión lineal: Supuestos
- Identificación
- Estimación por Mínimos Cuadrados Ordinarios (OLS)
- Propiedades
- Homoscedasticidad
- Estimación de la matriz de varianzas-covarianzas
- Bondad de ajuste: R^2
- Sesgo de variable omitida
- Interpretación de efectos marginales
    - Modelo lineal, cuadrático, y con interacciones
    - Modelos con logaritmo
    - Modelos con variables binarias o dummies

Motivación: Tamaño de clase y rendimiento académico

Efecto del tamaño de clase en el rendimiento académico (California)

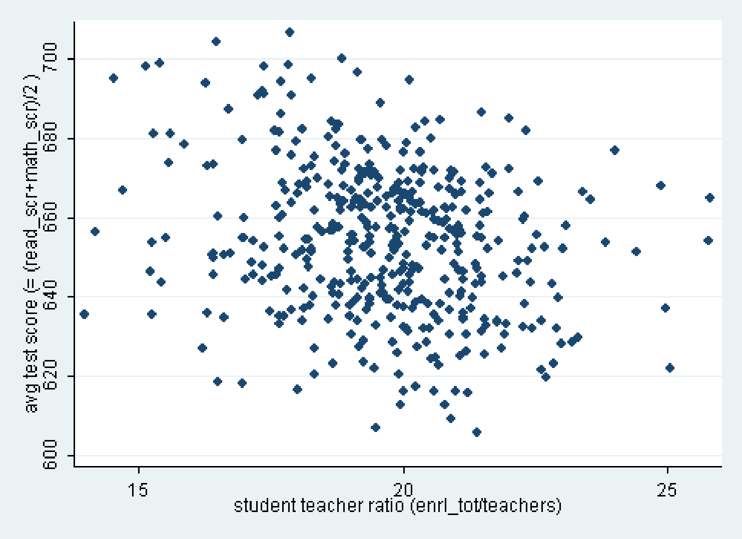
- Pregunta de política: ¿Cuál es el efecto de reducir el tamaño de clase en el rendimiento escolar?
- Datos: Todos los distritos escolares de escuelas K-6 y K-8 de California ($N = 420$)
- Variables:
  - Notas de examen de 5to grado (Stanford-9 achievement test, matemática y lectura), media del distrito (testscr)
  - Ratio estudiantes por profesor: número de estudiantes en el distrito divido por el número de profesores tiempo completo (str).

Estadísticas descriptivas

Table 1: Resumen de la distribución de los ratios de estudiantes por profesor y notas en test estandarizados para 5to grado para 420 distritos escolares en California en 1998

			Percentil				
	Media	Desviación Estándar	10%	25%	50% (mediana)	75%	90%
Estudiantes/profesor	19.6	1.9	17.3	18.6	19.7	20.9	21.9
Nota en test	654.2	19.1	630.4	640.0	654.5	666.7	679.1

Figure 1: Scatter plot de notas versus ratio estudiantes/profesor



Efecto causal

$$testscr_i = \beta_0 + \beta_1\, str_i + u_i,$$

donde u_i contiene todos los factores que influyen sobre el rendimiento escolar además del tamaño de clase.

- Efecto causal: β_1 mide el efecto causal de aumentar el tamaño de clase en rendimiento académico.
  Podemos usar β_1 para responder: ¿Cuál es el efecto de cambiar un estudiante de una clase de 22 estudiantes a una clase de 17 estudiantes?

Efecto causal

$$testscr_i = \beta_0 + \beta_1\, str_i + u_i,$$

donde u_i contiene todos los factores que influyen sobre el rendimiento escolar además del tamaño de clase.

- Efecto causal: β_1 mide el efecto causal de aumentar el tamaño de clase en rendimiento académico.
- Para poder estimar β_1 necesitamos imponer supuestos sobre u_i.

  Formalmente: $\mathbb{E}(u_i|str_i) = 0$.

Efecto causal

$$testscr_i = \beta_0 + \beta_1\, str_i + u_i,$$

- Efecto causal: Necesitamos $\mathbb{E}(u_i|str_i) = 0$.
- Intuitivamente: No debe haber diferencias sistemáticas en factores que puedan afectar el rendimiento académico en alumnos que asisten a escuelas con distintos tamaño de clase.
    1. Buscar factores que puedan afectar el rendimientos académico (además del tamaño de clase).
    2. Intuir si estos factores difieren según el tamaño de clase.

Efecto causal

$$testscr_i = \beta_0 + \beta_1\, str_i + u_i,$$

- Efecto causal: Necesitamos $\mathbb{E}(u_i | str_i) = 0$.
- Expresa la idea que es imposible observar simultáneamente un mismo estudiante en un clase de 17 alumnos y en una clase de 22 alumnos. Sin embargo, podemos comparar los resultados de estudiantes en clases de 17 alumnos y en clases de 22 alumnos. Establece, bajo que condiciones, esta comparación permite medir el efecto causal buscado.

Efecto causal

$$testscr_i = \beta_0 + \beta_1\, str_i + u_i$$

donde u_i contiene todos los factores que influyen sobre el rendimiento escolar además del tamaño de clase.

- ¿En qué situaciones el supuesto $\mathbb{E}(u_i|str_i) = 0$ puede ser válido?
- Asignación aleatoria por tamaño de clase. ¿Por qué?

  Ejemplo: El proyecto STAR (Student-Teacher Achievement Ratio) en Tennessee, USA.

Efecto causal

- Si no hay asignación aleatoria por tamaño de clase debemos buscar una estrategia alternativa para identificar el efecto causal
- Comparamos alumnos con variables observables similares (ingreso, educación de los padres, motivación, etc.) pero en clases de distinto tamaño: asignación aleatoria condicional en observables.

$$testscr_i = \beta_0 + \beta_1\, str_i + \gamma'\, x_i + v_i,$$

donde x_i incluye ingreso, educación de los padres, motivación, etc.

- Ejemplo: Efecto de veterano de guerra en salarios.

Efecto causal

$$testscr_i = \beta_0 + \beta_1\, str_i + u_i$$

donde $\mathbb{E}(u_i|str_i) = 0$

- Esperanza condicional:

$$\mathbb{E}(testscr_i|str_i) = \beta_0 + \beta_1\, str_i$$

Importante: El supuesto $\mathbb{E}(u_i|str_i) = 0$ nos permite dar una interpretación causal a la esperanza condicional.

Efecto causal

$$testscr_i = \beta_0 + \beta_1 \, str_i + u_i$$

donde $\mathbb{E}(u_i | str_i) = 0$

- Modelo de regresión lineal:

$$testscr_i = \beta_0 + \beta_1 \, str_i + u_i,$$

donde $\beta_1 = \frac{\text{Cov}(str_i, testscr_i)}{\text{Var}(str_i)}$, $\beta_0 = \mathbb{E}(testscr_i) - \beta_1 \, \mathbb{E}(str_i)$.

Importante: Inclusive si creemos que la esperanza condicional no es lineal $y_i = m(x_i) + u_i$, el modelo de regresión nos da la mejor aproximación lineal.

Estimación por Mínimos Cuadrados Ordinarios

$$testscr_i = \beta_0 + \beta_1\, str_i + u_i,$$

donde $\mathbb{E}(u_i|str_i) = 0$.

- Estimación por MCO

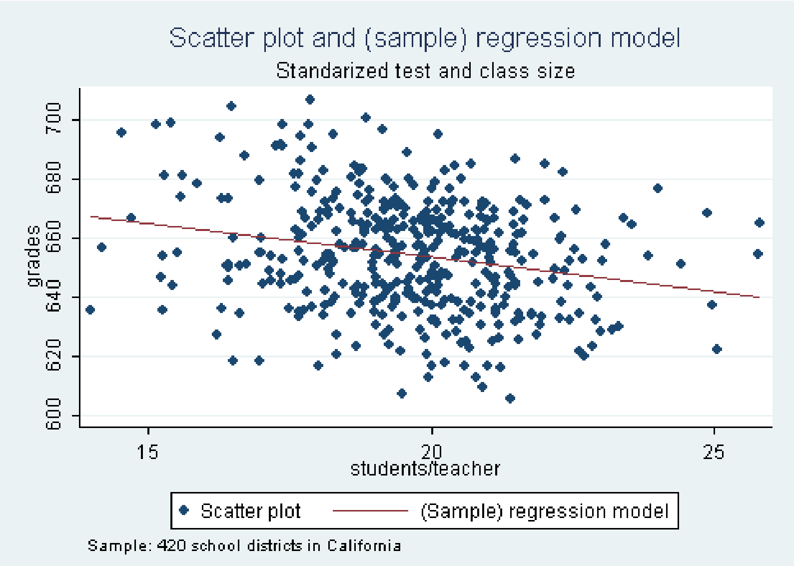
  Utilizamos el principio de analogía, es decir reemplazar momentos poblacionales por momentos muestrales.

  $\beta_1 = \frac{\text{Cov}(str_i, testscr_i)}{\text{Var}(str_i)}$ se estima con

  $$\widehat{\beta}_1 = \frac{\widehat{\text{Cov}}(str_i, testscr_i)}{\widehat{\text{Var}}(str_i)},$$

  y $\beta_0 = \mathbb{E}(testscr_i) - \beta_1\, \mathbb{E}(str_i)$ se estima con $\widehat{\beta}_0 = \overline{testscr} - \widehat{\beta}_1 \overline{str}$.
- Consistencia y distribución asintótica normal.

Estimación MCO: Tamaño de clase y rendimiento académico



Estimación MCO: Tamaño de clase y rendimiento académico

Salida de Stata

```
reg testscr str, robust

Linear regression                                      Number of obs =     420
                                                       F(  1,   418) =   19.26
                                                       Prob > F      =  0.0000
                                                       R-squared     =  0.0512
                                                       Root MSE      =  18.581

------------------------------------------------------------------------------
             |               Robust
     testscr |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
     str     |  -2.279808   .5194892    -4.39   0.000    -3.300945   -1.258671
     _cons   |    698.933   10.36436    67.44   0.000     678.5602    719.3057
------------------------------------------------------------------------------
```

Estimación MCO: Tamaño de clase y rendimiento académico

Salida de Stata

```
.     reg testscr str  el_pct, robust

Linear regression                                      Number of obs =     420
                                                       F(  2,   417) =  223.82
                                                       Prob > F      =  0.0000
                                                       R-squared     =  0.4264
                                                       Root MSE      =  14.464

------------------------------------------------------------------------------
             |               Robust
     testscr |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
         str |  -1.101296   .4328472    -2.54   0.011     -1.95213   -.2504616
      el_pct |  -.6497768   .0310318   -20.94   0.000     -.710775   -.5887786
       _cons |   686.0322   8.728224    78.60   0.000     668.8754     703.189
------------------------------------------------------------------------------
```

- En esta unidad discutimos estimación puntual, se (robustos!) y R^2.
- En la siguiente unidad contraste de hipótesis (t, F) e intervalos de confianza.

Modelo de regresión lineal: Supuestos

Modelo de regresión lineal

$$
\begin{aligned}
y_i &= \beta_0 + \beta_1 x_{1i} + ... + \beta_K x_{Ki} + u_i \\
&= x_i'\beta + u_i,
\end{aligned}
$$

donde y_i es una variable aleatoria (escalar), x_i es un vector aleatorio de $K \times 1$ (para no tener que trabajar con $K+1$) y u_i es una variable aleatoria (escalar).

Modelo de regresión lineal

$$y_i = x_i'\beta + u_i$$

- Supuestos:

OLS1 Muestra aleatoria de tamaño $N \implies \{y_i, x_i\}_{i=1}^N$ i.i.d.

OLS2 No correlación entre las variables explicativas y el inobservado: $\mathbb{E}(x_i u_i) = 0$.

OLS3 No colinealidad perfecta entre las variables explicativas: $\mathbb{E}(x_i x_i')$ es invertible o rango$(\mathbb{E}(x_i x_i')) = K$.

OLS4 Condiciones de regularidad ("No fat tails"): $\mathbb{E}(||x_i||^4) < \infty$ y $\mathbb{E}(y_i^4) < \infty$.

Identificación

Identificación

- Escribir los parámetros como una función de momentos poblacionales: ¿Qué momentos poblacionales permiten identificar los parámetros de interés?

Identificación

- Escribir los parámetros como una función de momentos poblacionales: ¿Qué momentos poblacionales permiten identificar los parámetros de interés?

$$
\begin{aligned}
& \mathbb{E}(x_i u_i) = 0 \quad \text{usando OLS2,} \\
\iff & \mathbb{E}[x_i(y_i - x_i'\beta)] = 0, \\
\iff & \mathbb{E}(x_i y_i) - \mathbb{E}(x_i x_i')\beta = 0, \\
\iff & \beta = \mathbb{E}(x_i x_i')^{-1}\,\mathbb{E}(x_i y_i) \quad \text{usando OLS3.}
\end{aligned}
$$

- OLS2 y OLS3 son los supuestos que identifican β.

Estimación por Mínimos Cuadrados Ordinarios (OLS)

Mínimos Cuadrados Ordinarios (OLS)

- Principio de analogía: reemplazar momentos poblacionales por momentos muestrales.

Mínimos Cuadrados Ordinarios (OLS)

- Principio de analogía: reemplazar momentos poblacionales por momentos muestrales.

  Estamos interesados en:

  $$\beta = \mathbb{E}(x_i x_i')^{-1}\, \mathbb{E}(x_i y_i).$$

  Usando el principio de analogía, lo estimamos con

  $$\hat{\beta} = \left(\frac{1}{N} \sum_{i=1}^{N} x_i x_i' \right)^{-1} \frac{1}{N} \sum_{i=1}^{N} x_i y_i.$$

Mínimos Cuadrados Ordinarios (OLS)

- Método de momentos:

Mínimos Cuadrados Ordinarios (OLS)

- Método de momentos:

  El modelo poblacional cumple

$$\mathbb{E}[x_i(y_i - x_i'\beta)] = 0.$$

  El estimador satisface una condición similar en la muestra

$$\frac{1}{N}\sum_{i=1}^{N} x_i(y_i - x_i'\hat{\beta}) = 0.$$

- Note que

$$\text{Cov}(x_i, u_i) = \text{Cov}(x_i, y_i - x_i'\beta) = 0,$$
$$\widehat{\text{Cov}}(x_i, \hat{u}_i) = \widehat{\text{Cov}}(x_i, y_i - x_i'\hat{\beta}) = 0.$$

Mínimos Cuadrados Ordinarios (OLS)

$$\hat{\beta} = \left(\frac{1}{N} \sum_{i=1}^{N} x_i x_i' \right)^{-1} \frac{1}{N} \sum_{i=1}^{N} x_i y_i.$$

- En forma matricial

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

donde X es una matriz de $N \times K$ e y es un vector de $N \times 1$ dados por

$$X = \begin{pmatrix} x_1' \\ \vdots \\ x_N' \end{pmatrix} \quad \text{e} \quad y = \begin{pmatrix} y_1 \\ \vdots \\ y_N \end{pmatrix}$$

Modelo de regresión simple

$$y_i = \beta_0 + \beta_1 x_{1i} + u_i,$$

donde

$$\beta_1 = \frac{\mathrm{Cov}(x_{1i}, y_i)}{\mathrm{Var}(x_{1i})},$$
$$\beta_0 = \mathbb{E}(y_i) - \beta_1\,\mathbb{E}(x_{1i}).$$

Modelo de regresión simple

$$y_i = \beta_0 + \beta_1 x_{1i} + u_i,$$

donde

$$\beta_1 = \frac{\mathrm{Cov}(x_{1i}, y_i)}{\mathrm{Var}(x_{1i})},$$
$$\beta_0 = \mathbb{E}(y_i) - \beta_1\, \mathbb{E}(x_{1i}).$$

El estimador OLS es,

$$\hat{\beta}_1 = \frac{\frac{1}{N}\sum_{i=1}^{N} x_{1i}(y_i - \bar{y})}{\frac{1}{N}\sum_{i=1}^{N} x_{1i}(x_{1i} - \bar{x}_1)} = \frac{\frac{1}{N}\sum_{i=1}^{N} y_i(x_{1i} - \bar{x}_1)}{\frac{1}{N}\sum_{i=1}^{N} x_{1i}(x_{1i} - \bar{x}_1)} = \frac{\widehat{\mathrm{Cov}}(x_{1i}, y_i)}{\widehat{\mathrm{Var}}(x_{1i})},$$
$$\hat{\beta}_0 = \bar{y} - \hat{\beta}_1 \bar{x}_1.$$

Regresión particionada

Dado el modelo de regresión múltiple

$$y_i = \beta_0 + \beta_1 x_{1i} + ... + \beta_k x_{ki} + ... + \beta_K x_{Ki} + u_i.$$

Podemos estimar β_1 a partir de:

$$\hat{\beta}_1 = \frac{\widehat{\text{Cov}}(\tilde{x}_{1i}, y_i)}{\widehat{\text{Var}}(\tilde{x}_{1i})},$$

donde

$$x_{1i} = \sum_{j \neq 1} \hat{\gamma}_j \, x_{ji} + \tilde{x}_{1i} = \hat{x}_{1i} + \tilde{x}_{1i}.$$

Regresión particionada

- Demostración:

$$\hat{\beta}_1 = \frac{\widehat{\text{Cov}}(\tilde{x}_{1i}, y_i)}{\widehat{\text{Var}}(\tilde{x}_{1i})},$$
$$= \frac{\widehat{\text{Cov}}(\tilde{x}_{1i}, \hat{\beta}_0 + \hat{\beta}_1 x_{1i} + \ldots + \hat{\beta}_K x_{Ki} + \hat{u}_i)}{\widehat{\text{Var}}(\tilde{x}_{1i})}.$$

(1) $\widehat{\text{Cov}}(\tilde{x}_{1i}, x_{ji}) = 0$ para todo $j \neq 1$, por construcción,

(2) $\widehat{\text{Cov}}(\tilde{x}_{1i}, \hat{u}_i) = 0$ porque $\hat{u}_i$ no está correlado con x_i's y $\tilde{x}_{1i}$ es una función lineal de x_i's.

Entonces,

$$\frac{\widehat{\text{Cov}}(\tilde{x}_{1i}, y_i)}{\widehat{\text{Var}}(\tilde{x}_{1i})} = \hat{\beta}_1 \frac{\widehat{\text{Cov}}(\tilde{x}_{1i}, x_{1i})}{\widehat{\text{Var}}(\tilde{x}_{1i})} = \hat{\beta}_1 \frac{\widehat{\text{Cov}}(\tilde{x}_{1i}, \hat{x}_{1i} + \tilde{x}_{1i})}{\widehat{\text{Var}}(\tilde{x}_{1i})} = \hat{\beta}_1 \frac{\widehat{\text{Cov}}(\tilde{x}_{1i}, \tilde{x}_{1i})}{\widehat{\text{Var}}(\tilde{x}_{1i})} = \hat{\beta}_1.$$

Predicción

- Los coeficientes β de la regresión poblacional se pueden definir como la solución del problema de mínimos cuadrados en la población:

$$\beta = \arg\min_b \mathbb{E}[(y_i - x_i'b)^2].$$

- Los estimadores por OLS $\hat{\beta}$ de β se pueden definir como la solución del problema de mínimos cuadrados en la muestra:

$$\hat{\beta} = \arg\min_b \frac{1}{N}\sum_{i=1}^{N}(y_i - x_i'b)^2.$$

  Utilizando las condiciones de primer orden

$$\frac{1}{N}\sum_{i=1}^{N} x_i(y_i - x_i'\hat{\beta}) = 0 \implies \hat{\beta} = \left(\frac{1}{N}\sum_{i=1}^{N} x_i x_i'\right)^{-1} \frac{1}{N}\sum_{i=1}^{N} x_i y_i.$$

Propiedades

Propiedades: Consistencia $\text{plim}\,\hat{\beta} = \beta$

- Paso intermedio

$$\hat{\beta} = \left(\frac{1}{N}\sum_{i=1}^{N} x_i x_i'\right)^{-1}\left(\frac{1}{N}\sum_{i=1}^{N} x_i y_i\right) = \beta + \left(\frac{1}{N}\sum_{i=1}^{N} x_i x_i'\right)^{-1}\left(\frac{1}{N}\sum_{i=1}^{N} x_i u_i\right)$$

- Demostración

$$\text{plim}\,\hat{\beta} = \text{plim}\,\beta + \left(\text{plim}\,\frac{1}{N}\sum_{i=1}^{N} x_i x_i'\right)^{-1} \text{plim}\left(\frac{1}{N}\sum_{i=1}^{N} x_i u_i\right) = \beta,$$

usando LGN y TFC en

$$\text{plim}\left(\frac{1}{N}\sum_{i=1}^{N} x_i x_i'\right) = \mathbb{E}(x_i x_i'),$$

$$\text{plim}\,\frac{1}{N}\sum_{i=1}^{N} x_i u_i = \mathbb{E}(x_i u_i) = 0,$$

Repaso: Matriz de varianzas-covarianzas

- Dada una variable aleatoria y su varianza es $\text{Var}(y) = \mathbb{E}[(y - \mathbb{E}\, y)^2] = \mathbb{E}(y^2) - (\mathbb{E}\, y)^2$.
- Dada un vector de variables aleatorias $\mathbf{y}$ su matrix de varianzas-covarianzas es

$$\text{Var}(\mathbf{y}) = \mathbb{E}[(\mathbf{y} - \mathbb{E}\,\mathbf{y})(\mathbf{y} - \mathbb{E}\,\mathbf{y})'] = \mathbb{E}(\mathbf{y}\mathbf{y}') - (\mathbb{E}\,\mathbf{y})(\mathbb{E}\,\mathbf{y})'.$$

- Ejemplo: $\mathbf{y} = (y_1 \quad y_2)'$.

Propiedades: Distribución asintótica normal

- Demostración

$$\sqrt{N}(\hat{\beta} - \beta) = \left(\frac{1}{N}\sum_{i=1}^{N} x_i x_i'\right)^{-1} \sqrt{N}\frac{1}{N}\sum_{i=1}^{N} x_i u_i.$$

Usando LGN+TFC y TCL en,

$$\text{plim}\left(\frac{1}{N}\sum_{i=1}^{N} x_i x_i'\right)^{-1} = \mathbb{E}(x_i x_i')^{-1},$$

$$\sqrt{N}\frac{1}{N}\sum_{i=1}^{N} x_i u_i \xrightarrow{d} \text{Normal}(0, \mathbb{E}(u_i^2 x_i x_i')).$$

Usando el teorema de Slutsky

$$\sqrt{N}(\hat{\beta} - \beta) \xrightarrow{d} \text{Normal}(0, \mathbb{E}(x_i x_i')^{-1}\, \mathbb{E}(u_i^2 x_i x_i')\, \mathbb{E}(x_i x_i')^{-1}).$$

Propiedades: Distribución asintótica normal

$$\sqrt{N}(\hat{\beta} - \beta) \xrightarrow{d} \text{Normal}(0, \mathbb{E}(x_i x_i')^{-1}\, \mathbb{E}(u_i^2 x_i x_i')\, \mathbb{E}(x_i x_i')^{-1}),$$

donde $\mathbb{E}(u_i^2 x_i x_i') < \infty$ bajo el supuesto $\mathbb{E}(||x_i||^4) < \infty$ y $\mathbb{E}(y_i^4) < \infty$.

- Decimos

$$\hat{\beta} \overset{a}{\sim} \text{Normal}(\beta, \text{AVar}(\hat{\beta})), \text{ donde}$$
$$\text{AVar}(\hat{\beta}) = \frac{1}{N}\, \mathbb{E}(x_i x_i')^{-1}\, \mathbb{E}(u_i^2 x_i x_i')\, \mathbb{E}(x_i x_i')^{-1}.$$

Ejemplo: Modelo de regresión simple

$$y_i = \beta_0 + \beta_1 x_{1i} + u_i,$$

donde $\mathbb{E}(u_i) = \mathbb{E}(x_{1i} u_i) = 0$

El estimador OLS es,

$$\hat{\beta}_1 = \frac{\frac{1}{N}\sum_{i=1}^{N} y_i (x_{1i} - \bar{x}_1)}{\frac{1}{N}\sum_{i=1}^{N} x_{1i}(x_{1i} - \bar{x}_1)} = \frac{\widehat{\text{Cov}}(x_{1i}, y_i)}{\widehat{\text{Var}}(x_{1i})}.$$

Ejemplo: Modelo de regresión simple

$$y_i = \beta_0 + \beta_1 x_{1i} + u_i,$$

donde $\mathbb{E}(u_i) = \mathbb{E}(x_{1i}u_i) = 0$

El estimador OLS es,

$$\hat{\beta}_1 = \frac{\frac{1}{N}\sum_{i=1}^{N} y_i(x_{1i} - \bar{x}_1)}{\frac{1}{N}\sum_{i=1}^{N} x_{1i}(x_{1i} - \bar{x}_1)} = \frac{\widehat{\text{Cov}}(x_{1i}, y_i)}{\widehat{\text{Var}}(x_{1i})}.$$

- La distribución asintótica es

$$\sqrt{N}(\hat{\beta}_1 - \beta_1) \xrightarrow{d} \text{Normal}(0, \frac{\mathbb{E}[u^2(x_1 - \mathbb{E}(x_1))^2]}{\text{Var}(x_1)^2}),$$

- Decimos

$$\hat{\beta}_1 \overset{a}{\sim} \text{Normal}(\beta_1, \text{AVar}(\hat{\beta}_1)), \text{ donde}$$

$$\text{AVar}(\hat{\beta}_1) = \frac{1}{N}\frac{\mathbb{E}[u^2(x_1 - \mathbb{E}(x_1))^2]}{\text{Var}(x_1)^2}.$$

Homoscedasticidad

Homoscedasticidad: $\mathbb{E}(u_i^2|x_i) = \mathbb{E}(u_i^2) = \sigma^2$

- Entonces

$$\mathbb{E}(u_i^2 x_i x_i') = \mathbb{E}[\mathbb{E}(u_i^2|x_i) x_i x_i'] = \sigma^2\,\mathbb{E}(x_i x_i'),$$

Por lo tanto

$$\begin{aligned}
\text{AVar}(\hat{\beta}) &= \frac{1}{N}\,\mathbb{E}(x_i x_i')^{-1}\,\mathbb{E}(u_i^2 x_i x_i')\,\mathbb{E}(x_i x_i')^{-1}, \\
&= \frac{1}{N}\,\mathbb{E}(x_i x_i')^{-1}\sigma^2\,\mathbb{E}(x_i x_i')\,\mathbb{E}(x_i x_i')^{-1}, \\
&= \frac{\sigma^2}{N}\,\mathbb{E}(x_i x_i')^{-1}.
\end{aligned}$$

Homoscedasticidad

- En el modelo de regresión simple $y_i = \beta_0 + \beta_1 x_{1i} + u_i$:

$$\text{AVar}(\hat{\beta}_1) = \frac{\mathbb{E}(u_i^2)}{N} \frac{1}{\text{Var}(x_{1i})}.$$

  Interpretación

Homoscedasticidad

- En el modelo de regresión múltiple

$$y_i = \beta_0 + \beta_1 x_{1i} + ... + \beta_k x_{ki} + ... + \beta_K x_{Ki} + v_i.$$

Usando la regresión particionada,

$$\mathrm{AVar}(\hat{\beta}_1) = \frac{\mathbb{E}(v_i^2)}{N} \frac{1}{\mathrm{Var}(\tilde{x}_{1i})}.$$

Homoscedasticidad

- En el modelo de regresión simple $y_i = \beta_0 + \beta_1 x_{1i} + u_i$:

$$\text{AVar}(\hat{\beta}_1) = \frac{\mathbb{E}(u_i^2)}{N} \frac{1}{\text{Var}(x_{1i})}.$$

- En el modelo de regresión múltiple

$$y_i = \beta_0 + \beta_1 x_{1i} + ... + \beta_k x_{ki} + ... + \beta_K x_{Ki} + v_i.$$

$$\text{AVar}(\hat{\beta}_1) = \frac{\mathbb{E}(v_i^2)}{N} \frac{1}{\text{Var}(\tilde{x}_{1i})}.$$

- Comparar

Homoscedasticidad

Si la esperanza condicional no es lineal, entonces el modelo de regresión es heteroscedástico. (Incluso si $\mathrm{Var}(y_i|x_i)$ es constante.)

Homoscedasticidad

Si la esperanza condicional no es lineal, entonces el modelo de regresión es heteroscedástico. (Incluso si $\text{Var}(y_i|x_i)$ es constante.)

- Demostración:

$$\begin{aligned}\mathbb{E}(u_i^2|x_i) &= \mathbb{E}((y_i - x_i'\beta)^2|x_i),\\ &= \mathbb{E}\{[(y_i - \mathbb{E}(y_i|x_i)) + (\mathbb{E}(y_i|x_i) - x_i'\beta)]^2|x_i\},\\ &= \mathbb{E}[(y_i - \mathbb{E}(y_i|x_i))^2|x_i] + \mathbb{E}[(\mathbb{E}(y_i|x_i) - x_i'\beta)^2|x_i],\\ &= \text{Var}(y_i|x_i) + \mathbb{E}[(\mathbb{E}(y_i|x_i) - x_i'\beta)^2|x_i].\end{aligned}$$

  Inclusive si $\text{Var}(y_i|x_i)$ es constante, $\mathbb{E}(u_i^2|x_i)$ depende de x_i según el modelo de regresión sea una mejor o peor aproximación a la esperanza condicional para distintos valores de x_i.

- Concluimos que el supuesto de homoscedasticidad es poco probable que se cumpla en la práctica.

Ejemplo: Educación y salarios en US

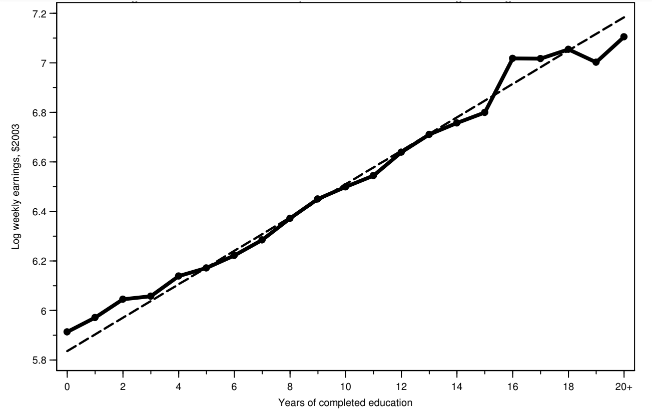


Figure 2: Regresión y esperanza condicional del log salarios semanales en el nivel educativo. La muestra incluye hombres blancos entre 40 y 49 años de la muestra del 5% del Censo 1980 (IPUMS)

Propiedades en muestras pequeñas: Insesgadez

Si suponemos $\mathbb{E}(u_i|x_i) = 0$ entonces $\mathbb{E}(\hat{\beta}) = \beta$ donde $X = (x_1, ..., x_N)$.

Propiedades en muestras pequeñas: Insesgadez

Si suponemos $\mathbb{E}(u_i|x_i) = 0$ entonces $\mathbb{E}(\hat{\beta}) = \beta$ donde $X = (x_1, ..., x_N)$.

- Demostración:

$$\mathbb{E}(\hat{\beta}|X) = \beta + \left(\frac{1}{N}\sum_{i=1}^{N} x_i x_i'\right)^{-1} \frac{1}{N}\sum_{i=1}^{N} x_i\, \mathbb{E}(u_i|x_i) = \beta,$$

  usando $\mathbb{E}(u_i|X) = \mathbb{E}(u_i|x_i)$ por el supuesto iid y $\mathbb{E}(u_i|x_i) = 0$.
  Luego por la LEI

$$\mathbb{E}(\hat{\beta}) = \mathbb{E}[\mathbb{E}(\hat{\beta}|X)] = \mathbb{E}(\beta) = \beta.$$

Teorema de Gauss-Markov

Si $\mathbb{E}(u_i|x_i) = 0$ y $\mathbb{E}(u_i^2|x_i) = \sigma^2$ (homoscedasticidad), entonces $\hat{\beta}$ es el estimador de menor varianza (condicional en X) entre los estimadores insesgados y lineales en y.

Teorema de Gauss-Markov

Si $\mathbb{E}(u_i|x_i) = 0$ y $\mathbb{E}(u_i^2|x_i) = \sigma^2$ (homoscedasticidad), entonces $\hat{\beta}$ es el estimador de menor varianza (condicional en X) entre los estimadores insesgados y lineales en y.

- Demostración (sketch): Calculamos la varianza condicional

$$\begin{aligned}
\mathrm{Var}(\hat{\beta}|X) &= \mathbb{E}[(\hat{\beta}-\beta)(\hat{\beta}-\beta)'|X], \\
&= \left(\frac{1}{N}\sum_{i=1}^N x_i x_i'\right)^{-1} \mathbb{E}\left[\left(\frac{1}{N}\sum_{i=1}^N x_i u_i\right)\left(\frac{1}{N}\sum_{i=1}^N x_i' u_i\right)|X\right]\left(\frac{1}{N}\sum_{i=1}^N x_i x_i'\right)^{-1}, \\
&= \left(\frac{1}{N}\sum_{i=1}^N x_i x_i'\right)^{-1} \mathbb{E}\left[\left(\frac{1}{N^2}\sum_{i=1}^N x_i x_i' u_i^2\right)|X\right]\left(\frac{1}{N}\sum_{i=1}^N x_i x_i'\right)^{-1}, \\
&= \left(\frac{1}{N}\sum_{i=1}^N x_i x_i'\right)^{-1} \left(\frac{1}{N^2}\sum_{i=1}^N x_i x_i' \,\mathbb{E}(u_i^2|x_i)\right)\left(\frac{1}{N}\sum_{i=1}^N x_i x_i'\right)^{-1}, \\
&= \frac{\sigma^2}{N}\left(\frac{1}{N}\sum_{i=1}^N x_i x_i'\right)^{-1}.
\end{aligned}$$

Teorema de Gauss-Markov

Si $\hat{\beta}$ minimiza la varianza condicional $\text{Var}(b|X)$ entonces minimiza la varianza no condicional $\text{Var}(b)$, donde b son estimadores insesgados de β.

Teorema de Gauss-Markov

Si $\hat{\beta}$ minimiza la varianza condicional $\text{Var}(b|X)$ entonces minimiza la varianza no condicional $\text{Var}(b)$, donde b son estimadores insesgados de β.

- Demostración: Sabemos que

$$\text{Var}(\hat{\beta}|X) \leq \text{Var}(b|X)$$

donde b es un estimador insesgado de β. Utilizando la descomposición de varianza

$$\text{Var}(b) = \mathbb{E}[\text{Var}(b|X)] + \text{Var}[\mathbb{E}(b|X)].$$

Pero hemos demostrado anteriormente que:

$$\text{Var}[\mathbb{E}(\hat{\beta}|X)] = \text{Var}(\beta) = 0.$$

Entonces $\text{Var}(\hat{\beta}|X) \leq \text{Var}(b|X) \implies \text{Var}(\hat{\beta}) \leq \text{Var}(b)$ donde b es un estimador insesgado de β.

Estimación de la matriz de varianzas-covarianzas

Estimación de la matriz de varianzas-covarianzas

- Supongamos homoscedasticidad $\mathbb{E}(u_i^2|x_i) = \sigma^2$.
  Estamos interesados en un estimador consistente de

$$\text{AVar}(\sqrt{N}(\hat{\beta} - \beta)) = \sigma^2 \, \mathbb{E}(x_i x_i')^{-1}.$$

Estimación de la matriz de varianzas-covarianzas

- Supongamos homoscedasticidad $\mathbb{E}(u_i^2|x_i) = \sigma^2$.
  Estamos interesados en un estimador consistente de

$$\text{AVar}(\sqrt{N}(\hat{\beta} - \beta)) = \sigma^2 \, \mathbb{E}(x_i x_i')^{-1}.$$

  Puede ser una buena idea usar:

$$\widehat{\text{AVar}}(\sqrt{N}(\hat{\beta} - \beta)) = \left(\frac{1}{N}\sum_{i=1}^{N} \hat{u}_i^2\right)\left(\frac{1}{N}\sum_{i=1}^{N} x_i x_i'\right)^{-1} = \hat{\sigma}^2 \left(\frac{1}{N}\sum_{i=1}^{N} x_i x_i'\right)^{-1},$$

  donde $\hat{u}_i = y_i - x_i'\hat{\beta}$.

Estimación de la matriz de varianzas-covarianzas

- Supongamos homoscedasticidad $\mathbb{E}(u_i^2|x_i) = \sigma^2$.
  Estamos interesados en un estimador consistente de

$$\text{AVar}(\sqrt{N}(\hat{\beta} - \beta)) = \sigma^2\, \mathbb{E}(x_i x_i')^{-1}.$$

  Puede ser una buena idea usar:

$$\widehat{\text{AVar}}(\sqrt{N}(\hat{\beta} - \beta)) = \left(\frac{1}{N}\sum_{i=1}^{N} \hat{u}_i^2\right)\left(\frac{1}{N}\sum_{i=1}^{N} x_i x_i'\right)^{-1} = \hat{\sigma}^2 \left(\frac{1}{N}\sum_{i=1}^{N} x_i x_i'\right)^{-1},$$

  donde $\hat{u}_i = y_i - x_i'\hat{\beta}$.

  Demostración: Anteriormente demostramos que

$$\text{plim}\left(\frac{1}{N}\sum_{i=1}^{N} x_i x_i'\right)^{-1} = \mathbb{E}(x_i x_i')^{-1},$$

  nos queda demostrar que $\text{plim}\,\hat{\sigma}^2 = \sigma^2$.

Homoscedasticidad

- Nos queda demostrar que $\operatorname{plim} \hat{\sigma}^2 = \sigma^2$:

Homoscedasticidad

- Nos queda demostrar que $\text{plim}\,\hat{\sigma}^2 = \sigma^2$:

$$\begin{aligned}
\frac{1}{N}\sum_{i=1}^{N}\hat{u}_i^2 &= \frac{1}{N}\sum_{i=1}^{N}(y_i - x_i'\hat{\beta})^2, \\
&= \frac{1}{N}\sum_{i=1}^{N}[(y_i - x_i'\beta) - (x_i'\hat{\beta} - x_i'\beta)]^2, \\
&= \frac{1}{N}\sum_{i=1}^{N}[(y_i - x_i'\beta)^2 + (x_i'\hat{\beta} - x_i'\beta)^2 - 2(x_i'\hat{\beta} - x_i'\beta)(y_i - x_i'\beta)], \\
&= \frac{1}{N}\sum_{i=1}^{N}[u_i^2 + (\hat{\beta} - \beta)'x_i x_i'(\hat{\beta} - \beta) - 2(\hat{\beta} - \beta)'x_i u_i], \\
&= \frac{1}{N}\sum_{i=1}^{N}u_i^2 + (\hat{\beta} - \beta)'\left(\frac{1}{N}\sum_{i=1}^{N}x_i x_i'\right)(\hat{\beta} - \beta) - 2(\hat{\beta} - \beta)'\frac{1}{N}\sum_{i=1}^{N}x_i u_i.
\end{aligned}$$

Homoscedasticidad

- Nos queda demostrar $\text{plim}\,\frac{1}{N}\sum_{i=1}^{N}\hat{u}_i^2 = \sigma^2$:

$$\frac{1}{N}\sum_{i=1}^{N}\hat{u}_i^2 = \frac{1}{N}\sum_{i=1}^{N}u_i^2 + (\hat{\beta}-\beta)'\left(\frac{1}{N}\sum_{i=1}^{N}x_i x_i'\right)(\hat{\beta}-\beta) - 2(\hat{\beta}-\beta)'\frac{1}{N}\sum_{i=1}^{N}x_i u_i.$$

Utilizando LGN y TFC

$$\text{plim}\,\frac{1}{N}\sum_{i=1}^{N}u_i^2 = \mathbb{E}(u_i^2) = \sigma^2,$$

$$\text{plim}(\hat{\beta}-\beta)'\,\text{plim}\left(\frac{1}{N}\sum_{i=1}^{N}x_i x_i'\right)\text{plim}(\hat{\beta}-\beta) = 0\times\mathbb{E}(x_i x_i')\times 0 = 0,$$

$$\text{plim}(\hat{\beta}-\beta)'\,\text{plim}\left(\frac{1}{N}\sum_{i=1}^{N}x_i u_i\right) = 0\times\mathbb{E}(x_i u_i) = 0.$$

Concluimos que $\text{plim}\,\hat{\sigma}^2 = \text{plim}\,\frac{1}{N}\sum_{i=1}^{N}\hat{u}_i^2 = \mathbb{E}(u_i^2) = \sigma^2$.

Homoscedasticidad

- Los programas estadísticos como Stata utilizan como estimador de σ^2 a

$$\tilde{\sigma}^2 = \frac{1}{N-K}\sum_{i=1}^{N}\hat{u}_i^2 = \frac{N}{N-K}\frac{1}{N}\sum_{i=1}^{N}\hat{u}_i^2 \xrightarrow{p} \sigma^2.$$

  La justificación es que $\tilde{\sigma}^2$ es un estimador insesgado de σ^2 cuando asumimos $\mathbb{E}(u_i|x_i) = 0$ y $\mathbb{E}(u_i^2|x_i) = \sigma^2$ (homoscedasticidad).
- Desde el punto de visto asintótico este ajuste de grados de libertad es irrelevante.

Homoscedasticidad

- Resumen: Bajo homoscedasticidad

  Decimos $\hat{\beta} \overset{a}{\sim} \text{Normal}(\beta, \text{AVar}(\hat{\beta}))$, donde

$$\text{AVar}(\hat{\beta}) = \frac{1}{N}\,\mathbb{E}(u_i^2)\,\mathbb{E}(x_i x_i')^{-1} = \frac{\sigma^2}{N}\,\mathbb{E}(x_i x_i')^{-1}.$$

  Un estimador consistente de $\text{AVar}(\hat{\beta})$ viene dado por

$$\widehat{\text{AVar}}(\hat{\beta}) = \frac{1}{N}\left(\frac{1}{N}\sum_{i=1}^{N}\hat{u}_i^2\right)\left(\frac{1}{N}\sum_{i=1}^{N}x_i x_i'\right)^{-1} = \frac{\hat{\sigma}^2}{N}\left(\frac{1}{N}\sum_{i=1}^{N}x_i x_i'\right)^{-1}.$$

  Por lo tanto $s.e.(\hat{\beta}_k) = \sqrt{\widehat{\text{AVar}}(\hat{\beta})_{kk}}$

Heteroscedasticidad

- Bajo heteroscedasticidad.

  Estamos interesados en un estimador consistente de

$$\mathrm{AVar}(\sqrt{N}(\hat{\beta}-\beta)) = \mathbb{E}(x_i x_i')^{-1}\,\mathbb{E}(u_i^2 x_i x_i')\,\mathbb{E}(x_i x_i')^{-1}.$$

  Puede ser buena idea usar:

$$\widehat{\mathrm{AVar}}(\sqrt{N}(\hat{\beta}-\beta)) = \left(\frac{1}{N}\sum_{i=1}^{N} x_i x_i'\right)^{-1}\left(\frac{1}{N}\sum_{i=1}^{N} \hat{u}_i^2 x_i x_i'\right)\left(\frac{1}{N}\sum_{i=1}^{N} x_i x_i'\right)^{-1},$$

  donde $\hat{u}_i = y_i - x_i'\hat{\beta}$.

- Usando argumentos similares al caso homoscedástico pero más engorrosos se puede demostrar que

$$\operatorname{plim} \frac{1}{N}\sum_{i=1}^{N} \hat{u}_i^2 x_i x_i' = \mathbb{E}(u_i^2 x_i x_i').$$

Heteroscedasticidad

- Resumen: Bajo heteroscedasticidad

Decimos $\hat{\beta} \overset{a}{\sim} \text{Normal}(\beta, \text{AVar}(\hat{\beta}))$, donde

$$\text{AVar}(\hat{\beta}) = \frac{1}{N}\,\mathbb{E}(x_i x_i')^{-1}\,\mathbb{E}(u_i^2 x_i x_i')\,\mathbb{E}(x_i x_i')^{-1}.$$

Un estimador consistente de $\text{AVar}(\hat{\beta})$ viene dado por

$$\widehat{\text{AVar}}(\hat{\beta}) = \frac{1}{N}\left(\frac{1}{N}\sum_{i=1}^{N} x_i x_i'\right)^{-1}\left(\frac{1}{N}\sum_{i=1}^{N} \hat{u}_i^2 x_i x_i'\right)\left(\frac{1}{N}\sum_{i=1}^{N} x_i x_i'\right)^{-1}.$$

Por lo tanto $s.e.(\hat{\beta}_k) = \sqrt{\widehat{\text{AVar}}(\hat{\beta})_{kk}}$

Forma matricial

- Bajo homoscedasticidad

$$\widehat{\text{AVar}}(\hat{\beta}) = \frac{\hat{\sigma}^2}{N}\left(\frac{1}{N}\sum_{i=1}^{N} x_i x_i'\right)^{-1}.$$

Definimos donde X es una matriz de $N \times K$ e y es un vector de $N \times 1$ dados por

$$X = \begin{pmatrix} x_1' \\ \vdots \\ x_N' \end{pmatrix} \quad \text{e} \quad y = \begin{pmatrix} y_1 \\ \vdots \\ y_N \end{pmatrix}$$

Forma matricial

- Bajo homoscedasticidad

$$\widehat{\text{AVar}}(\hat{\beta}) = \frac{\hat{\sigma}^2}{N}\left(\frac{1}{N}\sum_{i=1}^{N} x_i x_i'\right)^{-1}.$$

En forma matricial

$$\widehat{\text{AVar}}(\hat{\beta}) = \hat{\sigma}^2 \left(X'X\right)^{-1},$$

donde X es una matriz de $N \times K$, y es un vector de $N \times 1$ y

$$\hat{\sigma}^2 = \frac{1}{N}\hat{u}'\hat{u},$$

donde $\hat{u} = y - X\hat{\beta}$ es un es un vector de $N \times 1$.

Forma matricial

- Bajo heteroscedasticidad

$$\widehat{\text{AVar}}(\hat{\beta}) = \frac{1}{N}\left(\frac{1}{N}\sum_{i=1}^{N} x_i x_i'\right)^{-1}\left(\frac{1}{N}\sum_{i=1}^{N} \hat{u}_i^2 x_i x_i'\right)\left(\frac{1}{N}\sum_{i=1}^{N} x_i x_i'\right)^{-1}.$$

En forma matricial

$$\widehat{\text{AVar}}(\hat{\beta}) = N\left(X'X\right)^{-1} S \left(X'X\right)^{-1},$$

donde X es una matriz de $N \times K$, y es un vector de $N \times 1$ y

$$S = \frac{1}{N} X' diag(\hat{u}_1^2, \cdots, \hat{u}_N^2) X,$$

donde $diag(.)$ es una matriz diagonal de $N \times N$ con $\hat{u}_1^2, \cdots, \hat{u}_N^2$ en la diagonal principal.

Bondad de ajuste: R^2

Bondad de ajuste: R^2

- R^2 poblacional

$$R^2 = 1 - \frac{\mathbb{E}(u_i^2)}{\text{Var}(y_i)}.$$

- $\widehat{R}^2$ muestral

$$\widehat{R}^2 = 1 - \frac{\frac{1}{N}\sum_{i=1}^{N} \hat{u}_i^2}{\widehat{\text{Var}}(y_i)}.$$

Interpretación

Sesgo de variable omitida

Efecto del tamaño de clase en el rendimiento académico (California)

- Pregunta de política: ¿Cuál es el efecto de reducir el tamaño de clase en el rendimiento escolar?

Estimación MCO: Tamaño de clase y rendimiento académico

Salida de Stata

```
reg testscr str, robust

Linear regression                                      Number of obs =     420
                                                       F(  1,   418) =   19.26
                                                       Prob > F      =  0.0000
                                                       R-squared     =  0.0512
                                                       Root MSE      =  18.581

------------------------------------------------------------------------------
             |               Robust
testscr      |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
str          |  -2.279808   .5194892    -4.39   0.000    -3.300945   -1.258671
_cons        |    698.933   10.36436    67.44   0.000     678.5602    719.3057
------------------------------------------------------------------------------
```

Efecto del tamaño de clase en el rendimiento académico (California)

- Pregunta de política: ¿Cuál es el efecto de reducir el tamaño de clase en el rendimiento escolar?
- Respuesta: Si reducimos el tamaño de clase en un alumno por profesor, el rendimiento académico aumentaría, en media, en 2.28 puntos.
- ¿Es nuestra respuesta convincente? Depende del supuesto $\text{Cov}(u_i, str_i) = 0$.
- Dado que no tenemos asignación aleatoria, en este caso el supuesto no parece válido.
- Sesgo de variable omitida: Hay variables no observadas que afectan al tamaño de clase y el rendimiento escolar

Sesgo de variable omitida

- Solución: Incluir la variable relevante omitida como variable explicativa.

Sesgo de variable omitida

- Solución: Incluir la variable relevante omitida como variable explicativa.
- Intuición: Controlar por variables observables para hacer estudiantes en distintos tamaños de clase comparables. Buscamos generar asignación aleatoria condicional en observables.

Sesgo de variable omitida

- Variables que nos gustaría observar
  1. Características de las escuelas: ratio estudiantes por profesor, calidad de los profesores, computadoras (recursos materiales) por estudiante, medidas del diseño del plan de estudios, ...
  2. Características de los estudiantes: manejo del inglés, actividades extra-curriculares, ambiente de aprendizaje en la casa, nivel educativo de los padres,...
- Variables que observamos:
  1. ratio estudiantes por profesor (*str*),
  2. porcentaje de personas cuya lengua madre no es inglés en el distrito (*el_pct*),
  3. porcentaje elegible para un subsidio para el almuerzo (*meal_pct*),
  4. porcentaje que recibe un subsidio público (*calw_pct*), y
  5. ingreso medio en el distrito (*avginc*).

Estimación MCO: Tamaño de clase y rendimiento académico

Salida de Stata

	Model 1	Model 2	Model 3	Model 4	Model 5
Student/teacher	-2.28***	-1.10*	-1.00***	-1.31***	-1.01***
	(0.52)	(0.43)	(0.27)	(0.34)	(0.27)
English learner		-0.65***	-0.12***	-0.49***	-0.13***
		(0.03)	(0.03)	(0.03)	(0.04)
Lunch subsidy			-0.55***		-0.53***
			(0.02)		(0.04)
Income subsidy				-0.79***	-0.05
				(0.07)	(0.06)
Constant	698.93***	686.03***	700.15***	698.00***	700.39***
	(10.36)	(8.73)	(5.57)	(6.92)	(5.54)
R-squared	0.05	0.43	0.77	0.63	0.77
N	420.00	420.00	420.00	420.00	420.00

* p<0.05, ** p<0.01, *** p<0.001

- Signo del sesgo de variable omitida.

Sesgo de variable omitida

- El siguiente modelo de regresión mide un efecto causal

$$y_i = \beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} + u_i.$$

Sin embargo, no podemos observar x_{2i} y estimamos el modelo

$$y_i = \gamma_0 + \gamma_1 x_{1i} + v_i.$$

$$\text{plim}\, \hat{\gamma}_1 = \frac{\text{plim}\, \widehat{\text{Cov}}(x_{1i}, y_i)}{\text{plim}\, \widehat{\text{Var}}(x_{1i})} = \frac{\text{Cov}(x_{1i}, y_i)}{\text{Var}(x_{1i})} = \beta_1 + \beta_2 \frac{\text{Cov}(x_{1i}, x_{2i})}{\text{Var}(x_{1i})}.$$

$$\text{Sesgo Asintótico}(\hat{\gamma}_1) = \text{plim}\, \hat{\gamma}_1 - \beta_1 = \beta_2 \frac{\text{Cov}(x_{1i}, x_{2i})}{\text{Var}(x_{1i})}.$$

El signo del sesgo asintótico depende de β_2 y $\text{Cov}(x_{1i}, x_{2i})$

Ejemplo: Efecto de educación en salarios

- $lwage_i$ = log(salarios), $educ_i$ = años de educación, y $abil_i$ = habilidad innata.
- Tenemos los modelos

$$lwage_i = \beta_0 + \beta_{educ} educ_i + \beta_{abil} abil_i + u_i.$$

$$lwage_i = \gamma_0 + \gamma_{educ} educ_i + v_i.$$

- Entonces

$$\begin{aligned} \text{Sesgo Asintótico}(\hat{\gamma}_{educ}) &= \operatorname{plim} \hat{\gamma}_{educ} - \beta_{educ}, \\ &= \beta_{abil} \frac{\text{Cov}(educ_i, abil_i)}{\text{Var}(educ_i)}. \end{aligned}$$

$\beta_{abil} > 0$, $\text{Cov}(educ_i, abil_i) > 0 \implies \text{Sesgo Asintótico}(\hat{\gamma}_{educ}) > 0$

Interpretación de efectos marginales

Efectos marginales

- Dado una forma funcional para $\mathbb{E}(y|x_1, ..., x_K)$, el efecto marginal de cambios en x_j manteniendo constantes x_k para todo $k \neq j$ se puede aproximar por

$$\Delta\,\mathbb{E}(y|x) \approx \frac{\partial\,\mathbb{E}(y|x)}{\partial x_j}\Delta x_j,$$

para x_j continua, y

$$\Delta\,\mathbb{E}(y|x) = \mathbb{E}(y|x_j\!=\!1, x_k \text{ fijas } \forall k \neq j) - \mathbb{E}(y|x_j\!=\!0, x_k \text{ fijas } \forall k \neq j),$$

para x_j binaria.

Lineal, cuadrático, y con interacciones

- Lineal

$$\mathbb{E}(y|x_1, x_2) = \beta_0 + \beta_1\, x_1 + \beta_2\, x_2,$$
$$\frac{\partial\, \mathbb{E}(y|x)}{\partial x_j} = \beta_j \Rightarrow \Delta\, \mathbb{E}(y|x) = \beta_j\, \Delta x_j.$$

Lineal, cuadrático, y con interacciones

- Cuadrático

$$\begin{aligned}\mathbb{E}(y|x_1, x_2) &= \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_2^2, \\ \frac{\partial\, \mathbb{E}(y|x)}{\partial x_2} &= \beta_2 + 2\beta_3 x_2 \Rightarrow \Delta\, \mathbb{E}(y|x) \approx (\beta_2 + 2\beta_3 x_2)\Delta x_2.\end{aligned}$$

Lineal, cuadrático, y con interacciones

- Interacciones

$$\mathbb{E}(y|x_1, x_2) = \beta_0 + \beta_1\, x_1 + \beta_2\, x_2 + \beta_3\, x_2 x_1,$$
$$\frac{\partial\, \mathbb{E}(y|x)}{\partial x_2} = \beta_2 + \beta_3\, x_1 \Rightarrow \Delta\, \mathbb{E}(y|x) \approx (\beta_2 + \beta_3\, x_1)\Delta x_2.$$

Efecto del tamaño de clase en el rendimiento educativo

- Modelo

$$\mathbb{E}(testscr|str, el\_pct) = \beta_0 + \beta_{str}\, str + \beta_{el\_pct}\, el\_pct,$$

  donde *testscr* es la nota en un test estandarizado, *str* es el ratio de estudiantes por profesor y *el_pct* es el porcentaje de estudiantes cuya lengua madre no es inglés (en %).

- Información: $\mathbb{E}(testscr) = 654$, $\mathbb{E}(str) = 20$, $\mathbb{E}(el\_pct) = 16$, $\beta_{str} = -1.10$, $\beta_{el\_pct} = -0.65$.
- Efecto marginal de *str* y *el_pct*

Efecto de la educación en salarios

- Modelo

$$\mathbb{E}(wage|educ, exper, IQ) = \beta_0 + \beta_{educ}\, educ + \beta_{exper}\, exper \\ + \beta_{exper^2}\, exper^2 + \beta_{IQ}\, IQ + \beta_{educ\_IQ}\, educ \times IQ,$$

donde *wage* es salarios mensuales para hombres entre 28 y 38 años, *educ* son años de educación formal,*exper* son años de experiencia laboral y *IQ* es el resultado en test de inteligencia.

- Información: $\mathbb{E}(wage) = 938$, $\mathbb{E}(educ) = 13$, $\mathbb{E}(exper) = 12$, $\mathbb{E}(IQ) = 101$, $sd(IQ) = 15.05$,$\beta_{educ} = 22.80$, $\beta_{exper} = 18.34$, $\beta_{exper^2} = -0.05$, $\beta_{IQ} = 0.69$ y $\beta_{educ\_IQ} = 0.33$.
- Efecto marginal de *exper* y *educ*

Nivel-Log

$$
\begin{aligned}
\mathbb{E}(y|x) &= \beta_0 + \beta_1 \log x_1 + \beta_2 x_2, \\
\frac{\partial\, \mathbb{E}(y|x)}{\partial x_1} &= \beta_1 \frac{1}{x_1} \\
\Rightarrow \Delta\, \mathbb{E}(y|x) &\approx \beta_1 \frac{\Delta x_1}{x_1}, \\
&= \frac{\beta_1}{100} \left(\frac{\Delta x_1}{x_1} \times 100 \right).
\end{aligned}
$$

Nota: Diferencias entre cambios porcentuales versus cambios en puntos porcentuales.

En la práctica, $\log x_1$ versus x_1 medido en puntos porcentuales.

Efectos de la contaminación en los precios de las viviendas

- Modelo

$$\mathbb{E}(price|nox, rooms, stratio) = \beta_0 + \beta_{lnox}\log(nox) + \beta_{rooms}\, rooms + \beta_{stratio}\, stratio,$$

  donde *price* es el precio de la vivienda, *nox* es la cantidad de óxido nitroso en el aire (contaminación), *rooms* es el número de ambientes y *stratio* es el ratio estudiantes/profesor en la comuna.
- Información: $\mathbb{E}(price) = 22,511$, $\mathbb{E}(nox) = 5.55$, $\mathbb{E}(rooms) = 6.30$, $\mathbb{E}(stratio) = 18$, $\beta_{lnox} = -9,344$, $\beta_{rooms} = 7,069$, y $\beta_{stratio} = -1,130$.
- Efecto marginal de *nox*.

Log-nivel

$$
\begin{aligned}
\log \mathbb{E}(y|x_1, x_2) &= \beta_0 + \beta_1\, x_1 + \beta_2\, x_2, \\
\iff \mathbb{E}(y|x_1, x_2) &= \exp(\beta_0 + \beta_1\, x_1 + \beta_2\, x_2), \\
\frac{\partial\, \mathbb{E}(y|x)}{\partial x_1} &= \mathbb{E}(y|x)\, \beta_1, \\
\Rightarrow \Delta\, \mathbb{E}(y|x) &\approx \mathbb{E}(y|x)\, \beta_1 \Delta x_1, \\
\Rightarrow \frac{\Delta\, \mathbb{E}(y|x)}{\mathbb{E}(y|x)} \times 100 &\approx (\beta_1 \times 100) \Delta x_1.
\end{aligned}
$$

Efecto de la educación en salarios

- Modelo

$$\mathbb{E}(\log(wage)|educ, exper) = \beta_0 + \beta_{educ}\, educ \\ + \beta_{exper}\, exper + \beta_{exper^2}\, exper^2,$$

  donde *wage* es salarios mensuales, *educ* son años de educación formal, y *exper* son años de experiencia laboral.

- Información: $\beta_{educ} = 0.065$.
- Efecto marginal de *educ*.

Log-nivel

$$\mathbb{E}(y|x) = \exp(x'\beta) \Rightarrow \log \mathbb{E}(y|x) = x'\beta.$$

Pero, en general, trabajamos con

$$\mathbb{E}(\log y|x) = x'\beta,$$

utilizando un modelo

$$\log y = x'\beta + u, \quad \text{con } u \perp\!\!\!\perp x,$$

entonces

$$\frac{\partial\, \mathbb{E}(\log y|x)}{\partial x_j} = \frac{\partial \log(\mathbb{E}(y|x))}{\partial x_j} = \beta_j.$$

Note que necesitamos imponer independencia.

Efectos de la contaminación en los precios de las viviendas

- Modelo

$$\mathbb{E}(\log(\textit{price})|\textit{nox}, \textit{rooms}, \textit{stratio}) = \beta_0 + \beta_{\textit{lnox}} \log(\textit{nox}) + \beta_{\textit{rooms}}\, \textit{rooms} + \beta_{\textit{stratio}}\, \textit{stratio},$$

  donde *price* es el precio de la vivienda, *nox* es la cantidad de óxido nitroso en el aire (contaminación), *rooms* es el número de habitaciones y *stratio* es el ratio estudiante/profesor en la comuna.

- Información: $\mathbb{E}(\textit{rooms}) = 6.30$, $\mathbb{E}(\textit{stratio}) = 18$, $\beta_{\textit{rooms}} = 0.257$, y $\beta_{\textit{stratio}} = -0.051$.
- Efecto marginal de *rooms* y *stratio*.

Efectos de la contaminación en los precios de las viviendas

- Modelo

$$\mathbb{E}(\log(price)|nox, rooms, stratio) = \beta_0 + \beta_{lnox} \log(nox) \\ + \beta_{rooms}\, rooms + \beta_{stratio}\, stratio,$$

  donde *price* es el precio de la vivienda, *nox* es la cantidad de óxido nitroso en el aire (contaminación), *rooms* es el número de habitaciones y *stratio* es el ratio estudiante/profesor en la comuna.
- Información: $\beta_{lnox} = -0.645$.
- Efecto marginal de *nox*.

Log-log

$$
\begin{aligned}
\log \mathbb{E}(y|x_1, x_2) &= \beta_0 + \beta_1 \log(x_1) + \beta_2\, x_2, \\
\iff \mathbb{E}(y|x_1, x_2) &= \exp(\beta_0 + \beta_1 \log(x_1) + \beta_2 x_2), \\
\frac{\partial\, \mathbb{E}(y|x)}{\partial x_1} &= \frac{\mathbb{E}(y|x)}{x_1}\beta_1, \\
\Rightarrow \Delta\, \mathbb{E}(y|x) &\approx \frac{\mathbb{E}(y|x)}{x_1}\beta_1 \Delta x_1, \\
\Rightarrow \frac{\Delta\, \mathbb{E}(y|x)}{\mathbb{E}(y|x)} \times 100 &\approx \beta_1 \left(\frac{\Delta x_1}{x_1} \times 100 \right).
\end{aligned}
$$

Resumen: Modelos con logaritmos

- Resumen
  1. Nivel-log

     $$y = \beta_0 + \beta_1 \log x_1 + \beta_2 x_2 + u,$$

     Un cambio de x_1 en un 1% está asociado con un cambio en la media de y de 0.01 β_1 unidades.
  2. Log-nivel

     $$\log y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + u,$$

     Un cambio de x_1 en 1 unidad está asociado con un cambio en la media de y de $100\beta_1$%.
  3. Log-log

     $$\log y = \beta_0 + \beta_1 \log(x_1) + \beta_2 x_2 + u,$$

     Un cambio de x_1 en un 1% está asociado con un cambio en la media de y de β_1% (elasticidad).

Variables binarias o dummies

- Una variable dummy d_j (por ejemplo, *female*)

$$\begin{aligned}
&\mathbb{E}(y|d_1) = \beta_0 + \beta_1\, d_1, \\
&\Rightarrow \mathbb{E}(y|d_1 = 1) = \beta_0 + \beta_1 \\
&\Rightarrow \mathbb{E}(y|d_1 = 0) = \beta_0 \\
&\Rightarrow \mathbb{E}(y|d_1 = 1) - \mathbb{E}(y|d_1 = 0) = \beta_1 .
\end{aligned}$$

Efecto de ser veterano de guerra en ingresos

- Modelo

$$\mathbb{E}(earnings|dvet) = \beta_0 + \beta_{dvet}\, dvet,$$

  donde *earnings* es el ingreso en 1988-1991 de los hombres que solicitaron entrar a las fuerzas armadas entre 1979 y 1982 y *dvet* es una dummy de veterano de guerra.
- Información: $\mathbb{E}(earnings) = 13,611$, y $\beta_{dvet} = 1,718$.
- Efecto marginal de *dvet*.

Efecto de educación universitaria en salarios

- Modelo

$$\mathbb{E}(ahe|bachelor) = \beta_0 + \beta_{bachelor}\, bachelor,$$

  donde *ahe* son los ingresos por hora en 2008 de personas entre 25 y 34 años que poseen estudios secundarios o universitarios y *bachelor* es una dummy de graduado universitario.
- Información: $\mathbb{E}(ahe) = 18.97$, $\beta_0 = 15.33$ y $\beta_{bachelor} = 7.28$.
- Efecto marginal de *bachelor*.

Variables binarias o dummies

- Interacciones entre variables dummies (por ejemplo, *female*, y *married*)

$$
\begin{aligned}
&\mathbb{E}(y|d_1, d_2) = \beta_0 + \beta_1\, d_1 + \beta_2\, d_2 + \beta_3\, d_1 \times d_2, \\
&\Rightarrow \mathbb{E}(y|d_1\!=\!1, d_2\!=\!1) = \beta_0 + \beta_1 + \beta_2 + \beta_3 \\
&\Rightarrow \mathbb{E}(y|d_1\!=\!1, d_2\!=\!0) = \beta_0 + \beta_1 \\
&\Rightarrow \mathbb{E}(y|d_1\!=\!0, d_2\!=\!1) = \beta_0 + \beta_2 \\
&\Rightarrow \mathbb{E}(y|d_1\!=\!0, d_2\!=\!0) = \beta_0
\end{aligned}
$$

Variables binarias o dummies

- Interacciones entre variables dummies (por ejemplo, *female*, y *married*)

$$\begin{aligned}
&\mathbb{E}(y|d_1, d_2) = \beta_0 + \beta_1\, d_1 + \beta_2\, d_2 + \beta_3\, d_1 \times d_2,\\
&\Rightarrow \mathbb{E}(y|d_1\!=\!1, d_2\!=\!1) = \beta_0 + \beta_1 + \beta_2 + \beta_3\\
&\Rightarrow \mathbb{E}(y|d_1\!=\!1, d_2\!=\!0) = \beta_0 + \beta_1\\
&\Rightarrow \mathbb{E}(y|d_1\!=\!0, d_2\!=\!1) = \beta_0 + \beta_2\\
&\Rightarrow \mathbb{E}(y|d_1\!=\!0, d_2\!=\!0) = \beta_0
\end{aligned}$$

Dos modelos equivalentes

$$\begin{aligned}
y &= \beta_0 + \beta_1\, female + \beta_2\, married + \\
&\quad \beta_3\, female \times married + u, \qquad (1)\\
y &= \delta_0 + \delta_1\, female_married + \delta_2\, female_nonmarried + \\
&\quad \delta_3\, male_married + u. \qquad (2)
\end{aligned}$$

Efecto de ser veterano de guerra en ingresos

- Modelo

$$\begin{aligned}\mathbb{E}(earnings|dvet, dnwhite) =& \beta_0 + \beta_{dvet}\, dvet + \beta_{dnwhite}\, dnwhite \\ &+ \beta_{dvet\_dnwhite}\, dvet \times dnwhite,\end{aligned}$$

  donde *earnings* es el ingreso en 1988-1991 de los hombres que solicitaron entrar a las fuerzas armadas entre 1979 y 1982, *dvet* es una dummy de veterano de querra y *dnwhite* es una dummy NO es de raza blanca.

- Información: $\mathbb{E}(earnings) = 13,611$, $\beta_{dvet} = 1,233$, $\beta_{dnwhite} = -3,391$ y $\beta_{dvet\_dnwhite} = 1,215$.
- Efecto marginal de *dvet*.

Efecto de educación universitaria en salarios

- Modelo

$$\begin{aligned}\mathbb{E}(ahe|bachelor) =& \beta_0 + \beta_{bachelor}\, bachelor + \beta_{d2008}\, d2008 \\ &+ \beta_{bachelor\_d2008}\, bachelor \times d2008,\end{aligned}$$

  donde *ahe* es el ingresos (en \$ de 2008) por hora en 1992 y 2008 de personas entre 25 y 34 años que poseen estudios secundarios o universitarios y *bachelor* es una dummy de graduado universitario.
- Información: $\mathbb{E}(ahe) = 13.32$, $\beta_0 = 6.51$,$\beta_{bachelor} = 2.75$, $\beta_{d2008} = 8.82$ y $\beta_{bachelor\_d2008} = 4.83$.
- Efecto marginal de *bachelor* en 1992 y 2008.

Variables binarias o dummies

- Interacciones entre variables dummies y continuas

$$
\begin{aligned}
&\mathbb{E}(y|d_1, x) = \beta_0 + \beta_1\, d_1 + \beta_2\, x + \beta_3\, d_1 \times x,\\
&\Rightarrow \mathbb{E}(y|d_1 = 1, x) = \beta_0 + \beta_1 + (\beta_2 + \beta_3)x\\
&\Rightarrow \mathbb{E}(y|d_1 = 0, x) = \beta_0 + \beta_2\, x\\
&\Rightarrow \Delta\, \mathbb{E}(y|x) \approx (\beta_2 + \beta_3\, d_1)\Delta x.
\end{aligned}
$$

Efecto de la educación en salarios

- Modelo

$$\begin{aligned}\mathbb{E}(\log(\textit{wage})|\textit{educ}, \textit{female}) = & \beta_0 + \beta_{\textit{educ}}\,\textit{educ} + \beta_{\textit{female}}\,\textit{female} \\ & + \beta_{\textit{educ\_female}}\,\textit{educ} \times \textit{female},\end{aligned}$$

  donde *wage* es el ingreso por hora para datos de US en 2004 (CPS Marzo 2005), *educ* es años de educación y *female* es una dummy de mujer.

- Información: $\beta_{\textit{educ}} = 0.861$, $\beta_{\textit{female}} = -0.484$ y $\beta_{\textit{educ\_female}} = 0.0180$.
- Efecto marginal de *educ*.

Cambios grandes con logaritmo

- Cuando los efectos marginales son grandes, $\frac{\partial \log y}{\partial x}$ puede que no sea una buena aproximación a $\frac{\frac{\Delta y}{y}}{\Delta x}$. Podemos calcular el cambio exacto.

Cambios grandes con logaritmo

- Cuando los efectos marginales son grandes, $\frac{\partial \log y}{\partial x}$ puede que no sea una buena aproximación a $\frac{\frac{\Delta y}{y}}{\Delta x}$. Podemos calcular el cambio exacto.
- Dado el modelo $\log y = \beta_0 + \beta_1 x + u$, donde $u \perp\!\!\!\perp x$. Podemos escribir

$$\mathbb{E}(y|x) = \exp(\beta_0 + \beta_1 x)\,\mathbb{E}(\exp(u)),$$

de donde

$$\begin{aligned}\frac{\mathbb{E}(y|x = \bar{x} + 1) - \mathbb{E}(y|x = \bar{x})}{\mathbb{E}(y|x = \bar{x})} \times 100 &= \left(\frac{\exp(\beta_0 + \beta_1(\bar{x} + 1))}{\exp(\beta_0 + \beta_1(\bar{x}))} - 1\right) \times 100 \\ &= (\exp(\beta_1) - 1) \times 100.\end{aligned}$$

Efectos de la contaminación en los precios de las viviendas

- Modelo

$$\mathbb{E}(\log(price)|nox, rooms, stratio) = \beta_0 + \beta_{lnox} \log(nox) + \beta_{rooms}\, rooms + \beta_{stratio}\, stratio,$$

donde *price* es el precio de la vivienda, *nox* es la cantidad de óxido nitroso en el aire (contaminación), *rooms* es el número de habitaciones y *stratio* es el ratio estudiante/profesor en la comuna.

- Información: $\mathbb{E}(rooms) = 6.30$ y $\mathbb{E}(stratio) = 18$, $\beta_{rooms} = 0.257$
- Efecto marginal de *rooms*

$(\exp(\beta_{rooms}) - 1) \times 100 = 29.3\%.$