14.310x – Data Analysis for Social Scientists

MicroMasters en Statistics and Data Science (SDS) – MITx

BREIT (Brescia Institute of Technology) · Aporta — Grupo Breca

Clase 10

Regresión en la Práctica y Sesgo de Variable Omitida

Soluciones

Módulos oficiales del MicroMaster:
M9 – Practical Issues in Running Regressions and Omitted Variable Bias

B.Sc. Cristian Lazo Quispe
clazoq@uni.pe
linkedin.com/in/cristian-lazo-quispe · github.com/CristianLazoQuispe

Material del *Advanced Program in Data Science & Global Skills*, diseñado y desarrollado por **BREIT (Brescia Institute of Technology)**, iniciativa de **Aporta** —plataforma de innovación e impacto social del **Grupo Breca**— en alianza con el **MIT IDSS** (Institute for Data, Systems, and Society).

27 de julio de 2026

Niveles de dificultad *(difficulty levels)*

Cada problema está rotulado por color según su nivel, de lo más accesible (Nivel 1) a lo más difícil (Nivel 5). Empieza por donde te sientas cómodo y avanza a tu ritmo; los niveles 4–5 son extensiones para profundizar, no un requisito.

- **Nivel 1 – Fundamentos**: calentamiento muy guiado; para quien parte de casi cero.
- **Nivel 2 – Núcleo**: el nivel objetivo del curso/examen.
- **Nivel 3 – Aplicación**: combina varias ideas en contexto realista.
- **Nivel 4 – Reto**: difícil, integra varios temas.
- **Nivel 5 – Reto+**: muy difícil / fuera del scope típico.

1. Problemas y soluciones

Nivel 1 – Fundamentos (warm-up)

Para quien parte de casi cero: cálculos directos y guiados.

Ejercicio 1.1 (Leer un coeficiente log-lin (en por ciento))**.** Una regresión del *logaritmo* del salario sobre los años de educación da

$$\widehat{\ln(\text{salario})} = 6.2 + 0.09\,\text{educación}.$$

(a) Como Y está en logaritmo, ¿el 0.09 se lee en soles o en por ciento?

(b) Interpreta el 0.09 en una frase.

Solución

Qué nos piden. Traducir un número de la computadora (0.09) a una frase que un ser humano entienda. Todo el chiste está en *dónde* vive ese coeficiente.
La idea (el paso mental clave). Antes de calcular nada, mira **qué está en logaritmo**. Aquí el lado izquierdo es ln(salario), no el salario "pelado". Y hay una regla de oro que conviene grabarse: cuando la variable dependiente entra en **log**, un cambio en ella se lee como **cambio porcentual**, no en unidades. La razón intuitiva: un cambio pequeño en $\ln Y$ es (aprox.) el cambio *relativo* de Y, es decir $\Delta \ln Y \approx \Delta Y/Y$, y "sobre Y" es justamente "por ciento".
(a) Entonces el 0.09 se lee en **por ciento**. Técnicamente es una *semi-elasticidad* (modelo log-lin: Y en log, X en nivel).
(b) Para pasarlo a por ciento multiplicamos por 100:

$$100 \times 0.09 = 9.$$

En palabras, hablando como a un estudiante: *"cada año adicional de educación se asocia, en promedio y a igualdad de lo demás, con un salario aproximadamente un* 9 *por ciento*

más alto".
Error común. Decir "0.09 soles". Eso sería correcto solo si el lado izquierdo fuera el salario en nivel (lin-lin). Al estar en log, olvidarse del logaritmo cambia por completo la interpretación (¡de céntimos a un 9 por ciento!).

Ejercicio 1.2 (Signo del sesgo con la regla del producto)**.** El sesgo de variable omitida es $\beta_2 \cdot \delta$, donde β_2 es el efecto de la omitida sobre Y y $\delta = \mathrm{Cov}(X, W)/\mathrm{Var}(X)$ es la relación entre la omitida W y el regresor X.

(a) Si $\beta_2 > 0$ y $\delta > 0$, ¿el sesgo es hacia arriba o hacia abajo?

(b) Si $\beta_2 > 0$ y $\delta < 0$, ¿y ahora?

Solución

Qué nos piden. No calcular un número, sino la *dirección* del error: ¿el coeficiente que estimamos sale *de más* o *de menos*?
La idea. El sesgo es un **producto** de dos cosas, $\beta_2 \cdot \delta$. Y la regla de los signos de la multiplicación que aprendimos en el colegio manda: $(+)(+) = +$, $(+)(-) = -$. Así que basta multiplicar los signos. Fijemos el diccionario: *signo positivo del sesgo = "hacia arriba" = el coeficiente estimado* ***sobrestima*** *al verdadero*; *negativo = "hacia abajo" = lo* ***subestima***.
(a) $\beta_2 > 0$ y $\delta > 0$: el producto $\beta_2 \cdot \delta$ es

$$(+) \times (+) = (+) \;\Rightarrow\; \text{sesgo } \textbf{hacia arriba}.$$

El coeficiente estimado queda *por encima* del verdadero: lo sobrestima.
(b) $\beta_2 > 0$ y $\delta < 0$: ahora

$$(+) \times (-) = (-) \;\Rightarrow\; \text{sesgo } \textbf{hacia abajo}.$$

Lo subestima.
En palabras. Lo bonito de esto es que *no necesitamos tener* la variable omitida para razonar sobre el sesgo: si sabemos (o sospechamos) el signo con que la omitida afecta a Y (β_2) y el signo con que se relaciona con X (δ), ya sabemos hacia dónde nos engaña el coeficiente. Guarda esta idea: la usaremos una y otra vez.

Ejercicio 1.3 (¿Control bueno o *bad control*?)**.** Queremos el efecto de *años de educación* (X) sobre el *salario* (Y). Para cada candidato di si es un buen control (predeterminado) o un *bad control* (consecuencia de X):

(a) la *región de nacimiento* de la persona;

(b) la *ocupación* actual (que la educación ayuda a conseguir).

Solución

Qué nos piden. Decidir a qué "ajustar" el modelo. La pregunta clave para cada candidato es una sola: *¿esta variable ocurrió ANTES o DESPUÉS de la educación?*
La idea (el test del tiempo). Un buen control es una variable **predeterminada**: algo que ya estaba fijado *antes* de que la persona decidiera cuánto estudiar, y que puede confundir la comparación. Un *bad control* es una variable **posterior** a X, es decir, una *consecuencia* del propio tratamiento. Controlar por una consecuencia es como tapar el camino por el que el efecto viaja: borra parte de lo que queremos medir.
(a) Región de nacimiento. ¿Cuándo se fijó? Al nacer, muchísimo antes de estudiar. Es **predeterminada**. Además puede confundir (regiones con más escuelas y mejores salarios). Conclusión: **buen control**, incluirla ayuda.
(b) Ocupación actual. ¿Cuándo se fijó? *Después* de estudiar —de hecho, estudiar es justo lo que te ayuda a conseguir mejor ocupación. La ocupación es un **canal** por el que la educación sube el salario. Conclusión: **bad control**. Si la "mantenemos fija", comparamos personas con más y menos educación *pero con el mismo puesto*, y así escondemos justamente la parte del efecto que opera consiguiendo mejores puestos.
Regla para llevarse. Controla lo que vino *antes* de X; desconfía de lo que vino *después*.

Ejercicio 1.4 (Punto de retorno de un modelo cuadrático)**.** Un perfil de experiencia estima

$$\widehat{\ln(\text{salario})} = \cdots + 0.05\,\text{exper} - 0.001\,\text{exper}^2.$$

El punto de retorno (turning point) es $X^* = -\beta_1/(2\beta_2)$.

(a) Calcula X^*.

(b) ¿El salario sube o baja *después* de ese punto?

Solución

Qué nos piden. Encontrar la "cima" de la curva: en qué valor de experiencia el salario deja de subir.
La idea. Un modelo con X y X^2 dibuja una parábola. La cima (o el fondo) está donde la pendiente se hace cero. La fórmula del vértice ya nos la dan: $X^* = -\beta_1/(2\beta_2)$. Solo hay que identificar bien quién es β_1 y quién β_2 y sustituir con cuidado con los signos.
Paso 1: identificar los coeficientes. $\beta_1 = 0.05$ (el que acompaña a `exper`) y $\beta_2 = -0.001$ (el que acompaña a $\texttt{exper}^2$). Ojo: β_2 es *negativo*.
Paso 2: sustituir en la fórmula.

$$X^* = -\frac{\beta_1}{2\beta_2} = -\frac{0.05}{2 \times (-0.001)} = -\frac{0.05}{-0.002} = \frac{0.05}{0.002} = 25.$$

Fíjate cómo los dos signos menos se cancelan y queda positivo (¡lo cual tiene sentido: los años de experiencia no pueden ser negativos!).
(a) $X^* = 25$ años de experiencia.
(b) ¿Sube o baja después? Mira el signo de β_2. Como $\beta_2 = -0.001 < 0$, la parábola *abre*

hacia abajo (cóncava): primero sube, llega a la cima en 25, y luego **baja**. En palabras: la experiencia ayuda al salario, pero cada vez menos, y pasados los ~ 25 años el efecto se vuelve levemente negativo (rendimientos decrecientes).

Ejercicio 1.5 (Leer una interacción básica)**.** Se estima, con $D = 1$ si es mujer y 0 si es hombre,

$$\widehat{\text{salario}} = 1000 + 150\,\text{educación} - 200\,D - 40\,(\text{educación} \cdot D).$$

(a) ¿Cuánto sube el salario por un año más de educación para un **hombre**?

(b) ¿Y para una **mujer**?

Solución

Qué nos piden. El "retorno de un año de educación", pero ese retorno ahora *depende de si eres hombre o mujer* —por eso hay una interacción.

La idea (aislar el efecto de educación). El *efecto* de subir la educación en 1 es cuánto cambia $\widehat{\text{salario}}$ cuando `educación` sube en una unidad. Junta los términos que llevan `educación`: son $150\,\text{educación}$ y $-40\,(\text{educación} \cdot D)$. Factorizando la educación, el efecto marginal es

$$\frac{\partial \widehat{\text{salario}}}{\partial\,\text{educación}} = 150 - 40\,D.$$

Fíjate: *depende de* D. Eso es lo que hace una interacción. (El $-200\,D$ no aparece aquí porque no multiplica a la educación: solo mueve el *intercepto*, no la pendiente.)

(a) Hombre $\Rightarrow D = 0$:

$$150 - 40 \times 0 = 150.$$

Cada año extra de educación se asocia con +150 soles.

(b) Mujer $\Rightarrow D = 1$:

$$150 - 40 \times 1 = 110.$$

Cada año extra se asocia con +110 soles.

En palabras. El -40 de la interacción dice que el *retorno* de la educación es 40 soles menor para las mujeres que para los hombres en estos datos. La interacción cambia la *pendiente* (cuánto rinde estudiar), no solo el nivel.

Ejercicio 1.6 (Identificar la running variable y el umbral)**.** Un programa otorga una beca a todo estudiante que obtenga **al menos** 14 en un examen de admisión (nota de 0 a 20). Queremos el efecto de la beca sobre el ingreso futuro con un diseño de regresión discontinua (RDD).

(a) ¿Cuál es la *running variable* y cuál el *umbral* a_0?

(b) Escribe la regla del tratamiento D en términos de la nota a.

Solución

Qué nos piden. Traducir un programa de la vida real al lenguaje del RDD. Dos piezas: la variable que decide quién entra (*running variable*) y el corte donde se decide (*umbral*).
La idea. En un RDD siempre hay una variable continua que "manda", y una línea imaginaria: si la cruzas, recibes el tratamiento; si no, no. Pregúntate: *¿qué número mira el programa para darte la beca, y a partir de qué valor te la da?*
(a) El programa mira la **nota del examen**: esa es la *running variable* a. El corte es $a_0 = 14$ ("al menos 14").
(b) La regla del tratamiento se escribe con una función indicadora (que vale 1 si la condición se cumple y 0 si no):

$$D = \mathbf{1}[a \geq 14] = \begin{cases} 1 & \text{si } a \geq 14 \quad \text{(recibe beca)}, \\ 0 & \text{si } a < 14 \quad \text{(no recibe)}. \end{cases}$$

En palabras. La beca "se prende" exactamente en 14. Toda la magia del RDD vivirá en comparar a los que sacaron 13.9 con los que sacaron 14.1: casi idénticos, salvo por la beca.

Ejercicio 1.7 (¿Qué hace `factor(entidad)`?)**.** En R corremos `lm(y ~ x + factor(escuela))` sobre datos de alumnos de varias escuelas.

(a) ¿Qué agrega `factor(escuela)` al modelo?

(b) En una frase, ¿qué "controla" esa parte del modelo?

Solución

Qué nos piden. Entender qué hace, por dentro, ese `factor(escuela)` que escribimos casi de memoria.
La idea. `factor()` le dice a R: "esta variable es *categórica*, no un número". Con eso, R crea **una variable dummy por cada escuela** (dejando una como categoría base para no caer en la trampa de las dummies). En la práctica, cada escuela recibe su propio **intercepto**. A esas dummies de entidad las llamamos **efectos fijos**.
(a) Agrega **una dummy (un intercepto propio) por escuela**: efectos fijos de escuela.
(b) ¿Qué controla? Todo lo que es **constante dentro de cada escuela**: su calidad de gestión, su infraestructura, el barrio, la cultura del lugar... ¡esté medido o no! Es como decir "compara a los alumnos *dentro* de la misma escuela", de modo que las diferencias *entre* escuelas dejan de contaminar el coeficiente de `x`. Por eso el $\hat{\beta}_x$ que sale usa solo la variación *within* (dentro) de cada escuela.

Nivel 2 – Núcleo (core)

El nivel objetivo del curso/examen; todos deberían llegar aquí.

Ejercicio 1.8 (Aplicar la fórmula del OVB)**.** El modelo verdadero es $Y = \beta_0 + \beta_1 X + \beta_2 W + \varepsilon$ con $\beta_1 = 100$ y $\beta_2 = 300$. La pendiente de W sobre X es $\delta = 0.4$. Corremos la regresión *corta* (omitiendo W).

(a) ¿Cuál es el valor esperado de $\tilde{\beta}_1$ (el coeficiente de la corta)?

(b) ¿Cuánto es el sesgo y en qué dirección?

Solución

Qué nos piden. Predecir qué coeficiente esperaría darnos la regresión "mocha" (la que omite W), y compararlo con el verdadero.
La idea (la fórmula estrella de la clase). Cuando el modelo correcto lleva X y W pero corremos solo Y sobre X, no obtenemos β_1 limpio: obtenemos

$$\mathbb{E}[\tilde{\beta}_1] = \underbrace{\beta_1}_{\text{lo verdadero}} + \underbrace{\beta_2\,\delta}_{\text{el sesgo}}.$$

Así que el plan es: (1) escribir la fórmula, (2) enchufar los tres números, (3) leer el sesgo como el "pedazo de más".
Paso 1: la fórmula. $\mathbb{E}[\tilde{\beta}_1] = \beta_1 + \beta_2\,\delta$.
Paso 2: sustituir. Con $\beta_1 = 100$, $\beta_2 = 300$, $\delta = 0.4$:

$$\mathbb{E}[\tilde{\beta}_1] = 100 + 300 \times 0.4 = 100 + 120 = 220.$$

(a) La corta esperaría darnos $\tilde{\beta}_1 = \mathbf{220}$.
(b) El **sesgo** es el "extra" que la fórmula le pega a β_1:

$$\text{sesgo} = \beta_2\,\delta = 120.$$

Es *positivo*, así que va **hacia arriba**. En palabras, y para que duela un poco: el efecto real de X es 100, pero al omitir W la regresión nos entregaría 220, ¡más del doble! Todo ese 120 de más no es efecto de X: es efecto de W colado a través de X.

Ejercicio 1.9 (Elasticidad log-log)**.** Una regresión log-log de la cantidad demandada sobre el precio da

$$\widehat{\ln(\text{cantidad})} = 8.0 - 1.3\,\ln(\text{precio}).$$

(a) ¿Qué nombre recibe el coeficiente -1.3?

(b) Si el precio sube 10 por ciento, ¿cuánto cambia (aprox.) la cantidad?

Solución

Qué nos piden. Nombrar y usar el coeficiente de un modelo donde *ambos* lados están en logaritmo.
La idea. Recuerda la regla: cada lado en log ⇒ ese lado se lee en por ciento. Aquí *los*

dos están en log (log-log), así que el coeficiente relaciona "por ciento de X" con "por ciento de Y". Eso tiene nombre propio en economía: **elasticidad**.
(a) El -1.3 es la **elasticidad** (precio) de la demanda. En log-log, el coeficiente *es* directamente la elasticidad —no hay que multiplicar por nada.
(b) La elasticidad se lee así: "un 1 por ciento más de X mueve a Y en (elasticidad) por ciento". Como nos suben el precio 10 por ciento, multiplicamos:

$$\Delta\,\%\,\text{cantidad} \approx (-1.3) \times 10 = -13.$$

La cantidad demandada cae aproximadamente 13 **por ciento**.
En palabras. El signo negativo es la ley de demanda (sube el precio, baja la cantidad). Y como $|-1.3| > 1$, decimos que la demanda es *elástica*: reacciona *más que proporcionalmente* al precio (un 10 por ciento de subida provoca un 13 por ciento de caída).

Ejercicio 1.10 (Efecto no constante en un modelo cuadrático)**.** Con $\widehat{Y} = 20 + 3X - 0.2X^2$ (Y = rendimiento, X = horas de estudio):

(a) El efecto marginal es $\partial Y/\partial X = 3 - 0.4X$. Calcúlalo en $X = 2$ y en $X = 6$.

(b) ¿A partir de cuántas horas estudiar de más *empeora* el rendimiento predicho?

Solución

Qué nos piden. Mostrar que en un modelo cuadrático el "efecto de una hora más" *no es siempre el mismo*: cambia según cuántas horas ya llevas.
La idea. En una recta, la pendiente es fija. En una parábola, la pendiente cambia punto a punto; se obtiene derivando: $\partial Y/\partial X = 3 - 0.4X$. Esta expresión es una "máquina": le metes un valor de X y te dice cuánto rinde la siguiente hora ahí.
(a) Evaluamos la máquina.

$$X = 2: \quad 3 - 0.4(2) = 3 - 0.8 = 2.2; \qquad X = 6: \quad 3 - 0.4(6) = 3 - 2.4 = 0.6.$$

Interpretación: cuando recién llevas 2 horas, una hora más suma ~ 2.2 puntos; cuando ya llevas 6, esa hora extra apenas suma ~ 0.6. El efecto *se desinfla* —rendimientos decrecientes.
(b) "Empeora" significa que la siguiente hora *resta*, es decir, que el efecto marginal se vuelve negativo. Buscamos dónde cruza cero:

$$3 - 0.4X = 0 \;\Rightarrow\; X = \frac{3}{0.4} = 7.5.$$

A partir de $X^* = 7.5$ horas, $\partial Y/\partial X < 0$: estudiar de más *baja* el rendimiento predicho (piensa en fatiga, saturación). Y comprobemos con la fórmula del vértice, que debe dar lo mismo: $X^* = -\beta_1/(2\beta_2) = -3/(2 \cdot -0.2) = 7.5$. Coincide.

Ejercicio 1.11 (El coeficiente se mueve al controlar: ¿por qué?)**.** Regresión del salario sobre la educación. Corta (solo educación): coef = 302. Larga (añadiendo la habilidad, normalmente no observada): coef de educación = 181.

(a) ¿Cómo se llama el fenómeno que hace que el coeficiente cambie de 302 a 181?

(b) Sabiendo que gente más hábil estudia más y gana más, explica por qué la corta *sobrestimaba*.

Solución

Qué nos piden. Ponerle nombre y explicación a algo que verás todo el tiempo en papers: agregas un control y el coeficiente "salta".
(a) El nombre. Es el **sesgo de variable omitida (omitted variable bias, OVB)**. En la regresión corta faltaba la habilidad; al meterla (regresión larga), el coeficiente de educación cambió. Ese movimiento *es* el OVB haciéndose visible.
(b) El porqué, paso a paso. Sigamos la cadena de razonamiento:

1. La habilidad sube el salario: es su efecto directo, $\beta_2 > 0$.
2. La habilidad se relaciona positivamente con la educación (gente más hábil estudia más): $\delta > 0$.
3. En la corta, la educación *carga* con el efecto de la habilidad, porque van de la mano. La fórmula lo dice: $\mathbb{E}[\tilde{\beta}_1] = \beta_1 + \beta_2\,\delta$, y aquí $\beta_2\,\delta = (+)(+) > 0$: sesgo **hacia arriba**.
4. Por eso la corta (302) *sobrestima*: mezcla el efecto de estudiar con el de ser más hábil. Al controlar la habilidad, le "devolvemos" a la habilidad lo que era suyo y el retorno limpio de la educación baja a 181.

En palabras. 302 no era el efecto de la educación; era "educación + un poco de habilidad disfrazada de educación". El número honesto (a igual habilidad) es 181.

Ejercicio 1.12 (Completar la tabla de signos)**.** Para cada escenario, di la *dirección* del sesgo del coeficiente de X cuando se omite W.

(a) X = cigarrillos de la madre, Y = peso del bebé; W = pobreza (baja Y: $\beta_2 < 0$; más pobreza $\Rightarrow$ más fumar: $\delta > 0$).

(b) X = policías en una ciudad, Y = crimen; W = tamaño de la ciudad (sube el crimen: $\beta_2 > 0$; ciudades grandes tienen más policías: $\delta > 0$).

Solución

Qué nos piden. Aplicar la regla del producto de signos en dos casos famosos, ya con los signos "servidos" en el enunciado.
La receta. Siempre igual: sesgo $= \beta_2 \cdot \delta$; multiplica los dos signos y traduce (+ hacia

arriba/sobrestima, − hacia abajo/subestima).
(a) Cigarrillos y peso del bebé. Nos dan $\beta_2 < 0$ y $\delta > 0$:

$$\beta_2 \cdot \delta = (-)(+) = (-) \;\Rightarrow\; \text{sesgo } \textbf{hacia abajo}.$$

Interpretación: al no controlar la pobreza, el efecto de fumar sobre el peso parece *más negativo* de lo que realmente es. Parte de ese peso bajo se debe a la pobreza (que también empuja a fumar), no solo al cigarrillo.
(b) Policías y crimen. Nos dan $\beta_2 > 0$ y $\delta > 0$:

$$\beta_2 \cdot \delta = (+)(+) = (+) \;\Rightarrow\; \text{sesgo } \textbf{hacia arriba}.$$

Y aquí viene lo espectacular: el efecto *verdadero* de más policías sobre el crimen suele ser negativo (más policías, menos crimen), pero el sesgo hacia arriba puede ser tan fuerte que el coeficiente salga ¡**positivo**! Un ingenuo concluiría "más policías causan más crimen". La culpa es del tamaño de la ciudad: las grandes tienen a la vez más crimen y más policías. Es el ejemplo clásico de por qué omitir un confusor arruina todo.

Ejercicio 1.13 (Interacción: efecto para dos grupos)**.** $\widehat{Y} = 50 + 8\,X + 12\,D + 2\,(X \cdot D)$, con D una dummy de zona urbana.

(a) Escribe el efecto marginal de X como función de D.

(b) ¿Cuánto vale en zona rural ($D = 0$) y urbana ($D = 1$)?

Solución

Qué nos piden. Separar cuánto "pesa" X según la zona, otra vez con una interacción.
Paso 1: aislar los términos con X. En $\widehat{Y} = 50 + 8X + 12D + 2(X \cdot D)$, los que llevan X son $8X$ y $2(X \cdot D)$. Deriva respecto de X (o factoriza X):

$$\frac{\partial \widehat{Y}}{\partial X} = 8 + 2D.$$

El $12D$ no entra: no multiplica a X, solo sube el intercepto de la zona urbana.
(a) Efecto marginal de X: $8 + 2D$.
(b) Enchufamos cada valor de D.

$$D = 0 \text{ (rural)}: \; 8 + 2(0) = 8; \qquad D = 1 \text{ (urbana)}: \; 8 + 2(1) = 10.$$

En palabras: cada unidad más de X vale 8 en zona rural y 10 en zona urbana. La interacción (+2) es *cuánto más rinde X por ser urbano*. El 12 solo desplaza el punto de partida (el nivel), no la pendiente —no lo confundas con el efecto de X.

Ejercicio 1.14 (Pooled vs. within: ¿a cuál creer?)**.** Con datos de 6 distritos, la regresión *pooled* de rendimiento sobre gasto da pendiente −0.38; al agregar `factor(distrito)` (efectos fijos), la pendiente pasa a +1.35.

(a) ¿Cuál de las dos estima el efecto *dentro* de cada distrito?

(b) Da una razón por la que la pooled sale negativa.

Solución

Qué nos piden. Elegir entre dos números que se contradicen (uno negativo, otro positivo) y explicar el desacuerdo.
La idea (dos tipos de variación). Hay dos maneras de comparar: *entre* distritos (¿los distritos que gastan más rinden más?) y *dentro* de cada distrito (si un mismo distrito gasta más, ¿rinde más?). La *pooled* mezcla ambas; los efectos fijos (`factor(distrito)`) se quedan solo con la de *dentro* (within), porque comparan cada distrito consigo mismo.
(a) La de **efectos fijos** (+1.35) estima el efecto *dentro* de cada distrito. Es la que aisla el efecto "limpio" del gasto.
(b) Por qué la pooled sale negativa. Imagina que los distritos con mejor rendimiento "de base" (alto efecto fijo: buena gestión, alumnos favorecidos) resultan ser los que gastan *menos* por alumno. Entonces, al comparar *entre* distritos, ves "menos gasto junto con más rendimiento" —una correlación negativa que **no** es causal: la produce el distrito, no el gasto. Esa señal falsa arrastra la pooled hacia abajo hasta volverla negativa. Es OVB con nombre y apellido: la variable omitida es "el distrito". Al controlarlo con efectos fijos, esa confusión desaparece y aflora el efecto real, positivo (+1.35). *Créele al within.*

Ejercicio 1.15 (Estimar el salto en un RDD)**.** Beca si el puntaje $a \geq 60$. Se estima, en una ventana alrededor del corte,

$$\widehat{Y} = 41 + 9\,D + 0.5\,(a - 60), \qquad D = \mathbf{1}[a \geq 60].$$

(a) Predice Y justo por debajo del corte ($a = 60^{-}$, $D = 0$) y justo por encima ($a = 60^{+}$, $D = 1$).

(b) ¿Cuál es el efecto causal local de la beca?

Solución

Qué nos piden. Medir el "escalón" que la beca produce justo en el corte.
La idea. Fíjate que la running variable entra *centrada*: $(a - 60)$. Eso es a propósito: en el corte $a = 60$, ese término vale 0 y se apaga, dejando ver limpiamente el efecto de la dummy D. Así que evaluemos la recta "pegaditos" al corte por ambos lados.
(a) Justo debajo ($a = 60$, $D = 0$):

$$\widehat{Y} = 41 + 9(0) + 0.5(60 - 60) = 41 + 0 + 0 = 41.$$

Justo encima ($a = 60$, $D = 1$):

$$\widehat{Y} = 41 + 9(1) + 0.5(60 - 60) = 41 + 9 + 0 = 50.$$

La misma nota (60), pero uno recibe beca y el otro no: 41 contra 50.

(b) El efecto. El **salto** en el corte es la diferencia entre ambos:

$$50 - 41 = 9.$$

Es decir, $\hat{\tau} = 9$ (¡que es justo el coeficiente de D: por eso centramos!). En palabras: cruzar el umbral y recibir la beca eleva Y en 9 unidades *para quienes están cerca de* 60. Es un efecto **local**: vale en el corte, donde los de un lado y del otro son casi gemelos; no lo extrapoles a alguien con 30 o 90.

Ejercicio 1.16 (Detectar el *bad control* en un estudio)**.** Un estudio del efecto de un *programa de capacitación* (X) sobre el *salario* (Y) controla por: (i) edad, (ii) educación previa, (iii) *si la persona fue ascendida* después de la capacitación.

(a) ¿Cuál de los tres es un *bad control*?

(b) ¿Por qué sesga incluirlo?

Solución

Qué nos piden. Auditar una lista de controles y detectar al intruso.
La idea (otra vez el test del tiempo). Recorre la lista preguntando: *¿esto pasó antes o después de la capacitación?* Lo que pasó después y es *consecuencia* del tratamiento es sospechoso.

- (i) Edad: predeterminada (la traes de antes). [**OK**] buen control.
- (ii) Educación previa: por definición *previa*. [**OK**] buen control.
- (iii) Ser ascendido *después* de capacitarse: ocurrió *después* y muy probablemente *gracias* a la capacitación. [**!**] sospechoso.

(a) El *bad control* es (iii), **ser ascendido**.
(b) Por qué sesga. El ascenso es un **canal** por el que la capacitación sube el salario: capacitas → te ascienden → ganas más. Si "mantienes fijo" el ascenso, comparas capacitados y no capacitados *que fueron ascendidos por igual*, y así borras justamente el pedazo del efecto que viaja por los ascensos. Peor aún, condicionar en una variable posterior puede *abrir* correlaciones falsas (sesgo de selección/collider). En cambio, edad y educación previa son anteriores a X: controlarlas es legítimo y ayuda.

Ejercicio 1.17 (Interpretación lin-log)**.** Se estima $\widehat{\text{gasto en salud}} = 200 + 150 \ln(\text{ingreso})$ (gasto en soles).

(a) Si el ingreso sube 1 por ciento, ¿cuánto sube (aprox.) el gasto en salud?

(b) ¿Y si el ingreso se *duplica* (sube 100 por ciento)?

Solución

Qué nos piden. Usar el tercer caso de logaritmos: solo X en log (lin-log), Y en nivel (soles).
La idea. Aquí el que está en log es X, así que ahora el *cambio porcentual* lo pones *tú en* X y el modelo te devuelve un cambio en *unidades de* Y. La regla práctica: un 1 por ciento más de X sube Y en $\beta_1/100$. (Sale de $\Delta Y \approx \beta_1 \, \Delta \ln X$, y un 1 por ciento es $\Delta \ln X \approx 0.01$.)
(a) Subida de 1 por ciento del ingreso.

$$\Delta Y \approx \frac{\beta_1}{100} \times 1 = \frac{150}{100} = 1.5 \text{ soles.}$$

Cada 1 por ciento más de ingreso $\approx +1.5$ soles de gasto en salud.
(b) Duplicar el ingreso. Duplicar es un cambio grande, así que conviene usar el log directo: $\Delta Y \approx \beta_1 \, \Delta \ln(\text{ingreso})$, con $\Delta \ln = \ln(2) \approx 0.693$:

$$\Delta Y \approx 150 \times 0.693 \approx 104 \text{ soles.}$$

En palabras: pasar de, digamos, 1000 a 2000 de ingreso sube el gasto en salud unos 104 soles. Nota lo elegante del lin-log: el efecto de un *cambio porcentual dado* es constante en soles, sin importar de qué nivel de ingreso partas.

Nivel 3 – Aplicación

Combina dos o más ideas en contexto realista; consolida lo aprendido.

Ejercicio 1.18 (Predecir la dirección del sesgo *sin* los datos)**.** Estimamos el efecto del *tamaño de la clase* (X, n.ó de alumnos) sobre el *rendimiento* (Y), pero omitimos el *nivel socioeconómico* W del colegio. Se sabe que: colegios con mejor nivel socioeconómico tienen clases *más pequeñas* ($\text{Corr}(X, W) < 0$) y mejor rendimiento ($\beta_2 > 0$).

(a) ¿Hacia dónde se sesga el coeficiente de X?

(b) Si el efecto verdadero de clases más grandes es levemente negativo, ¿la regresión corta podría *exagerar* ese efecto negativo?

Solución

Qué nos piden. Razonar el sesgo *sin* calcular, solo con la lógica de los signos —la habilidad más útil del capítulo.
Paso 1: sacar los signos del enunciado. "Mejor nivel socioeconómico ⇒ clases más pequeñas" significa que W y X se mueven en sentidos opuestos: $\delta < 0$. Y "mejor nivel ⇒ mejor rendimiento" es $\beta_2 > 0$.

Paso 2: la regla del producto.

$$\text{sesgo} = \beta_2 \cdot \delta = (+)(-) = (-) \;\Rightarrow\; \textbf{hacia abajo}.$$

(a) El coeficiente de X (tamaño de clase) se sesga **hacia abajo**: sale *más negativo* que el verdadero.
(b) Sí, exactamente. Si el efecto real de agrandar la clase es *levemente* negativo, el sesgo hacia abajo lo empuja a parecer *muy* negativo. Parte de ese "daño de las clases grandes" es en realidad el efecto del nivel socioeconómico bajo (colegios pobres tienen clases grandes *y* peor rendimiento por muchas otras razones). En palabras: la corta *exagera* el efecto negativo. Si controlaras el nivel socioeconómico (o pusieras efectos fijos de colegio), el coeficiente subiría hacia su valor verdadero, más modesto.

Ejercicio 1.19 (Elegir la forma funcional)**.** Un análisis exploratorio (kernel) muestra que el *ingreso* sube con la *edad*, alcanza un máximo cerca de los 50 años y luego baja.

(a) ¿Basta una regresión lineal $Y = \beta_0 + \beta_1\,\text{edad}$? ¿Por qué no?

(b) Propon una especificación OLS que capture esa forma y di el signo esperado del término nuevo.

Solución

Qué nos piden. Elegir la forma del modelo a partir de la *forma del dibujo* que mostraron los datos.
La idea. Primero visualiza: "sube, llega a un máximo cerca de 50, y baja" es una **joroba** (una $\cap$). Ahora pregúntate qué curvas puede dibujar cada modelo.
(a) ¿Basta la recta? No. Una recta tiene pendiente *constante*: solo puede subir siempre o bajar siempre, jamás "subir y luego bajar". Si forzaras una recta a estos datos, te daría una pendiente "promedio" engañosa que no describe bien ni la subida de los jóvenes ni la caída de los mayores. La recta es la herramienta equivocada para una joroba.
(b) La propuesta. Agrega un término cuadrático:

$$Y = \beta_0 + \beta_1\,\text{edad} + \beta_2\,\text{edad}^2 + \varepsilon.$$

Para que sea una joroba ($\cap$, sube-y-baja) necesitamos $\beta_2 < 0$ (parábola cóncava). Y como el máximo debe caer cerca de 50, la fórmula del vértice nos da una pista de consistencia: $-\beta_1/(2\beta_2) \approx 50$. Sigue siendo OLS (lineal en los parámetros): solo añadimos una columna edad^2. (Alternativa válida, más flexible: dummies por tramos de edad.)

Ejercicio 1.20 (Anatomía de la paradoja de Simpson)**.** En cada uno de 3 hospitales, un tratamiento nuevo tiene *mayor* tasa de recuperación que el viejo. Pero juntando todos los datos (pooled), el tratamiento nuevo tiene *menor* tasa de recuperación.

(a) Explica cómo es posible en términos de una variable omitida (el hospital).

(b) ¿Qué regresión da la respuesta correcta?

Solución

Qué nos piden. Explicar una aparente contradicción: bueno en cada hospital, pero malo en total. Suena a brujería; es estadística.
La idea (busca la variable escondida). Cuando "dentro" dice una cosa y "en total" dice la contraria, casi siempre hay una variable de agrupación que se cuela. Aquí es el **hospital**. Piensa quién usa el tratamiento nuevo y en qué hospitales.
(a) El mecanismo. Supongamos que los hospitales que más adoptan el tratamiento nuevo son también los que reciben a los pacientes *más graves* (hospitales de referencia). Esos pacientes se recuperan menos *por su gravedad*, no por el tratamiento. Entonces, al juntar todo (pooled), el tratamiento nuevo aparece "asociado" a peor recuperación, pero es un espejismo: lo que arrastra el resultado es la gravedad, correlacionada con el hospital y con el uso del tratamiento. El hospital (y la gravedad que trae) es la **variable omitida**: afecta la recuperación y se relaciona con qué tratamiento se usa. Es OVB puro.
(b) La regresión correcta. La que **controla el hospital**, con efectos fijos:

```
lm(recuperacion ~ tratamiento + factor(hospital)).
```

Al comparar tratamientos *dentro* del mismo hospital (misma mezcla de gravedad), recuperamos la verdad: el tratamiento nuevo es mejor. Moraleja: cuando "dentro" y "en total" pelean, casi siempre gana "dentro" —y los efectos fijos son la forma de quedarse con "dentro".

Ejercicio 1.21 (Interpretar τ y su carácter *local*)**.** Un RDD sobre una beca por puntaje ($a_0 = 60$) estima $\hat{\tau} = 9$ con una ventana de ± 15 puntos.

(a) Interpreta $\hat{\tau} = 9$ con precisión (¿para quiénes vale?).

(b) ¿Por qué *no* podemos concluir que la beca subiría 9 el resultado de un alumno con 30 puntos?

Solución

Qué nos piden. Interpretar un efecto de RDD con la palabra clave que todos olvidan: *local*.
(a) Interpretación precisa. $\hat{\tau} = 9$ es el efecto causal de la beca sobre el resultado, **para los estudiantes cuyo puntaje está cerca de** 60 —los que "apenas" la ganan o la pierden. En la jerga, es un *LATE* (local average treatment effect): un promedio del efecto *en el corte*. No es "el efecto de la beca en general".
(b) Por qué no extrapolar a 30 **puntos.** Aquí está el corazón del RDD. Su credibilidad viene de que, *justo* en el umbral, el de 59.9 y el de 60.1 son casi gemelos: la única diferencia relevante es la beca, así que el salto es causal. Pero un alumno con 30 puntos **no tiene un gemelo del otro lado**: no hay, cerca de él, alguien prácticamente idéntico que difiera solo en la beca. El diseño no nos dice nada sobre él; extrapolar el 9 hasta 30 sería inventar, apoyándose en supuestos que el RDD *no* garantiza (que el efecto es el mismo lejos del corte). En palabras: el RDD ilumina un punto (el umbral) con mucha fuerza, pero deja

el resto a oscuras.

Ejercicio 1.22 (Decodificar interacción con dummies (estilo DiD))**.** Con *tratado* y *post* (dummies) y su interacción,

$$\widehat{Y} = 10 + 2\,\text{tratado} + 3\,\text{post} + 5\,(\text{tratado} \cdot \text{post}).$$

(a) Calcula las cuatro medias de celda (control/tratado × antes/después).

(b) ¿Cuál de los coeficientes es el efecto del programa (DiD)?

Solución

Qué nos piden. "Desarmar" una regresión con dos dummies y su interacción para recuperar las cuatro medias de una tabla 2×2, y ver dónde vive el efecto.
La idea. Cada combinación de (tratado, post) enciende (= 1) o apaga (= 0) ciertos términos. Vamos celda por celda, con calma.
(a) Las cuatro celdas.

- **Control-antes** (tratado= 0, post= 0): se apaga todo menos el intercepto: $\widehat{Y} = 10$.
- **Control-después** (tratado= 0, post= 1): se enciende `post`: $10 + 3 = 13$.
- **Tratado-antes** (tratado= 1, post= 0): se enciende `tratado`: $10 + 2 = 12$.
- **Tratado-después** (tratado= 1, post= 1): se encienden los cuatro: $10+2+3+5 = 20$.

(b) ¿Dónde está el efecto del programa? La **interacción**, 5. Comprobémoslo con la lógica del DiD ("resta de restas"):

$$\underbrace{(20-12)}_{\text{cambio del tratado}} - \underbrace{(13-10)}_{\text{cambio del control}} = 8 - 3 = 5.$$

El grupo tratado subió 8; pero *aun sin programa* habría subido lo mismo que el control (3, la "marea" común del tiempo). Lo que el programa añadió *por encima* de esa marea es 5. Los otros coeficientes son la brecha inicial entre grupos (2) y la tendencia temporal (3); ninguno es el efecto causal. *El efecto siempre está en la interacción.*

Ejercicio 1.23 (Log y cuadrático juntos)**.** $\widehat{\ln(\text{salario})} = 6.0 + 0.08\,\text{educación} + 0.04\,\text{exper} - 0.0008\,\text{exper}^2$.

(a) Interpreta el 0.08 de educación (¿en qué unidades?).

(b) ¿A cuántos años de experiencia el salario predicho deja de crecer?

Solución

Qué nos piden. Leer dos ideas a la vez en un mismo modelo: un logaritmo (para la educación) y un cuadrático (para la experiencia). Se atacan por separado.
(a) El 0.08 **de educación.** El lado izquierdo está en **log**, y la educación entra en *nivel*: es un caso log-lin. Regla: coeficiente en por ciento. Entonces $100 \times 0.08 = 8$: cada año adicional de educación se asocia con un salario aproximadamente 8 **por ciento** más alto, manteniendo la experiencia constante. (El log convierte el efecto en porcentaje; la unidad "año de educación" se lee contra "por ciento de salario".)
(b) ¿Cuándo deja de crecer con la experiencia? La experiencia entra con un término cuadrático, así que "deja de crecer" es su punto de retorno. Identifica: $\beta_1 = 0.04$ (coef de `exper`) y $\beta_2 = -0.0008$ (coef de `exper`^{2}). Sustituye en el vértice:

$$X^* = -\frac{\beta_1}{2\beta_2} = -\frac{0.04}{2 \times (-0.0008)} = \frac{0.04}{0.0016} = 25.$$

A los **25** años de experiencia el salario predicho toca su máximo; después se aplana y cae levemente (porque $\beta_2 < 0$). Observa cómo cada parte del modelo cuenta su propia historia: el log para la educación, la parábola para la experiencia.

Ejercicio 1.24 (Calcular el sesgo desde una tabla de regresiones)**.** Dos modelos sobre los mismos datos:

$$\text{(corta) } \widehat{Y} = \cdots + 250\,X, \qquad \text{(larga) } \widehat{Y} = \cdots + 180\,X + 300\,W.$$

Además, una regresión auxiliar da $\widehat{W} = \cdots + \delta\,X$.

(a) Usando $\tilde{\beta}_1 = \beta_1 + \beta_2\delta$, despeja δ.

(b) Interpreta el signo de δ que obtuviste.

Solución

Qué nos piden. Usar la fórmula del OVB "al revés": en vez de predecir el sesgo, recuperar la relación oculta δ entre la omitida y el regresor, usando lo que sí vemos (los dos coeficientes de X y el efecto de W).
Paso 1: identificar cada pieza en la fórmula. La corta da $\tilde{\beta}_1 = 250$; la larga da el β_1 limpio $= 180$ y el efecto de la omitida $\beta_2 = 300$. La incógnita es δ. La fórmula:

$$\tilde{\beta}_1 = \beta_1 + \beta_2\,\delta \implies 250 = 180 + 300\,\delta.$$

Paso 2: despejar δ**.** Resta 180 a ambos lados y divide entre 300:

$$300\,\delta = 250 - 180 = 70 \implies \delta = \frac{70}{300} \approx 0.233.$$

(a) $\delta \approx 0.233$.
(b) Interpretación del signo. $\delta > 0$: la omitida W y el regresor X están *positivamente* relacionados (cuando sube X, tiende a subir W). Y como $\beta_2 = 300 > 0$ también, el sesgo

$\beta_2\,\delta = 70$ era **positivo** (hacia arriba): por eso la corta (250) salía por encima de la larga (180), justo 70 de más. Todo cuadra: la brecha entre los dos coeficientes *es* el sesgo, y el sesgo *es* $\beta_2\,\delta$.

Ejercicio 1.25 (Diseñar la lista de controles)**.** Quieres el efecto causal de *tener internet en casa* (X) sobre el *rendimiento escolar* (Y). Candidatos a control: (i) ingreso del hogar, (ii) educación de los padres, (iii) horas diarias en redes sociales, (iv) región.

(a) ¿Cuáles incluirías como *buenos* controles?

(b) ¿Cuál es un *bad control* y por qué?

Solución

Qué nos piden. Armar tú mismo la lista de controles y justificar cada decisión —la parte de "arte" del oficio.
La idea. Dos preguntas por candidato: (1) ¿es un *confusor* (afecta a Y y se relaciona con X)? Si sí, quiero controlarlo. (2) ¿es *predeterminado* o es una *consecuencia* de X? Si es consecuencia, es un bad control y lo dejo fuera.
(a) Buenos controles. (i) **ingreso del hogar**, (ii) **educación de los padres** y (iv) **región**. Los tres son *predeterminados* (no los causa tener internet) y son plausibles confusores: los hogares con más recursos tienen internet *y* mejor rendimiento por muchas razones (libros, tiempo, apoyo). Controlarlos compara "hogares parecidos" y limpia el efecto del internet.
(b) El bad control. (iii) **horas en redes sociales**. Es una *consecuencia* de tener internet: sin internet no hay redes. Es un canal (posiblemente negativo) por el que el internet afecta al rendimiento. Si lo controlas, "congelas" las horas de redes y borras esa parte del efecto —además de arriesgar sesgo de collider. En palabras: no controles el efecto secundario de tu propio tratamiento; deja que forme parte del efecto que mides.

Nivel 4 – Reto

Difícil: integra varios temas en un mismo problema.

Ejercicio 1.26 (Derivar la fórmula del OVB)**.** Modelo verdadero $Y = \beta_0 + \beta_1 X + \beta_2 W + \varepsilon$ con $\mathbb{E}[\varepsilon \mid X, W] = 0$. Corremos la regresión corta de Y sobre X y obtenemos $\tilde{\beta}_1 = \widehat{\mathrm{Cov}}(X, Y)/\widehat{\mathrm{Var}}(X)$.

(a) Sustituye el modelo verdadero en $\mathrm{Cov}(X, Y)$ y simplifica.

(b) Deduce que $\mathbb{E}[\tilde{\beta}_1] = \beta_1 + \beta_2\,\delta$ con $\delta = \mathrm{Cov}(X, W)/\,\mathrm{Var}(X)$.

Solución

Qué nos piden. *Demostrar* la fórmula que venimos usando de memoria. Vas a ver que no es magia: sale de las propiedades de la covarianza que ya conoces.
Herramientas que usaremos. La covarianza es *lineal*: $\mathrm{Cov}(X,\, aU + bV + c) = a\,\mathrm{Cov}(X,U) + b\,\mathrm{Cov}(X,V)$ (las constantes no covarían). Y el estimador OLS simple es $\tilde{\beta}_1 = \mathrm{Cov}(X,Y)/\,\mathrm{Var}(X)$.
(a) Sustituir el modelo verdadero dentro de $\mathrm{Cov}(X,Y)$**.** Como $Y = \beta_0 + \beta_1 X + \beta_2 W + \varepsilon$:

$$\mathrm{Cov}(X,Y) = \mathrm{Cov}\Big(X,\ \beta_0 + \beta_1 X + \beta_2 W + \varepsilon\Big).$$

Reparte la covarianza término a término:

$$= \underbrace{\mathrm{Cov}(X,\beta_0)}_{=0} + \beta_1 \underbrace{\mathrm{Cov}(X,X)}_{=\mathrm{Var}(X)} + \beta_2\,\mathrm{Cov}(X,W) + \underbrace{\mathrm{Cov}(X,\varepsilon)}_{=0}.$$

Dos términos mueren: $\mathrm{Cov}(X,\beta_0) = 0$ (con una constante no hay covarianza) y $\mathrm{Cov}(X,\varepsilon) = 0$ (por la exogeneidad $\mathbb{E}[\varepsilon \mid X, W] = 0$, X no lleva información del error). Queda:

$$\mathrm{Cov}(X,Y) = \beta_1\,\mathrm{Var}(X) + \beta_2\,\mathrm{Cov}(X,W).$$

(b) Dividir entre $\mathrm{Var}(X)$**.**

$$\tilde{\beta}_1 = \frac{\mathrm{Cov}(X,Y)}{\mathrm{Var}(X)} = \beta_1 + \beta_2 \underbrace{\frac{\mathrm{Cov}(X,W)}{\mathrm{Var}(X)}}_{\delta} = \beta_1 + \beta_2\,\delta.$$

Y aquí está el reconocimiento clave: ese cociente $\mathrm{Cov}(X,W)/\,\mathrm{Var}(X)$ es *exactamente* la fórmula de una pendiente OLS —la de regresar la omitida W sobre el incluido X. Por eso lo llamamos δ. Tomando esperanza, $\mathbb{E}[\tilde{\beta}_1] = \beta_1 + \beta_2\delta$: la corta no estima β_1, sino β_1 *más* el sesgo $\beta_2\delta$.

Ejercicio 1.27 (Punto de retorno: derivada y segundo orden)**.** Sea $Y = \beta_0 + \beta_1 X + \beta_2 X^2$ con $\beta_2 \neq 0$.

(a) Deriva respecto de X, iguala a cero y obtén $X^* = -\beta_1/(2\beta_2)$.

(b) Usa la segunda derivada para decir cuándo X^* es un *máximo* y cuándo un *mínimo*.

Solución

Qué nos piden. Deducir la fórmula del vértice (que hemos usado "de receta") con cálculo básico, y —lo importante— entender *por qué* el signo de β_2 decide si es cima o fondo.
(a) Primer orden (dónde la pendiente es cero). El máximo o mínimo de una

función suave ocurre donde su pendiente se anula. Deriva:

$$\frac{dY}{dX} = \beta_1 + 2\beta_2 X.$$

Iguálala a cero y despeja X:

$$\beta_1 + 2\beta_2 X = 0 \implies X^* = -\frac{\beta_1}{2\beta_2}.$$

Ese es el único punto donde la curva "se aplana".
(b) Segundo orden (¿cima o fondo?). Deriva otra vez:

$$\frac{d^2Y}{dX^2} = 2\beta_2.$$

El criterio de la segunda derivada dice: si es negativa, el punto es un **máximo**; si es positiva, un **mínimo**. Como el signo de $2\beta_2$ es el signo de β_2:

- $\beta_2 < 0 \Rightarrow$ segunda derivada $< 0 \Rightarrow$ **máximo** (la parábola abre hacia abajo, $\cap$: sube y baja). *Es el caso típico de salarios/experiencia.*
- $\beta_2 > 0 \Rightarrow$ segunda derivada $> 0 \Rightarrow$ **mínimo** (abre hacia arriba, $\cup$: baja y sube).

En palabras. β_1 ubica *dónde* está el vértice; β_2 decide si es una loma o un valle. Por eso, apenas ves el signo de β_2, ya sabes la forma de la historia.

Ejercicio 1.28 (El supuesto de identificación del RDD y una amenaza)**.** En un RDD, el tratamiento es $D = \mathbf{1}[a \geq a_0]$.

(a) Enuncia el supuesto de identificación (¿qué debe ser *continuo* en a_0?).

(b) Un examen de admisión con corte en 14 permite que el profesor "suba" notas de 13.5 a 14 a sus favoritos. ¿Por qué amenaza esto al RDD? ¿Cómo se detecta?

Solución

Qué nos piden. Nombrar el supuesto que hace *creíble* un RDD y reconocer cuándo se rompe —la diferencia entre un estudio sólido y uno que engaña.
(a) El supuesto de continuidad. La identificación del RDD descansa en esto: *en ausencia del tratamiento*, todo —el resultado esperado $\mathbb{E}[Y \mid a]$ y cualquier característica de los individuos— variaría de forma **continua** (sin saltos) al cruzar el umbral a_0. Si eso se cumple, entonces *cualquier* salto que veamos en Y justo en a_0 no puede venir de un cambio suave del puntaje: solo puede venir del tratamiento. Ese es el argumento causal.
(b) La amenaza: manipulación de la running variable (sorting). Si el profesor sube selectivamente de 13.5 a 14 a sus favoritos, entonces quienes quedan *justo encima* del corte ya no son comparables con los de *justo debajo*: los de arriba incluyen a los "favoritos", que probablemente difieren en cosas que también afectan el ingreso futuro (contactos, apoyo, motivación). Se rompe la continuidad de las características en el corte,

y el salto en Y ya no es limpio: mezcla el efecto de la beca con el de "ser favorito".
Cómo se detecta. Con un **test de densidad de McCrary**: mira cuántas observaciones hay a cada lado del corte. Si hubo manipulación, verás un *amontonamiento* sospechoso *justo encima* de 14 (mucha gente "empujada" a 14) y un hueco justo debajo. Un salto en la *cantidad* de gente (no en el resultado) alrededor de a_0 es la señal de alarma. Sin manipulación, la densidad debería cruzar el umbral suavemente.

Ejercicio 1.29 (Efectos fijos eliminan la omitida invariante)**.** Panel con entidades j: $Y_{ij} = \beta X_{ij} + \alpha_j + \varepsilon_{ij}$, donde α_j es un efecto de entidad *no observado* y posiblemente correlacionado con X.

(a) Escribe el modelo en *desviaciones respecto a la media de la entidad* (resta $\bar{Y}_{\cdot j}$, $\bar{X}_{\cdot j}$).

(b) Muestra que α_j desaparece y explica por qué eso elimina el OVB por α_j.

Solución

Qué nos piden. Demostrar, con álgebra sencilla, *por qué* los efectos fijos funcionan: no es fe, es una resta que hace desaparecer al culpable.
La idea (la transformación *within*). Si α_j es constante dentro de la entidad j, entonces al restar el *promedio de la entidad* a cada observación, ese α_j se cancela consigo mismo. Veamos.
(a) Promediar dentro de la entidad y restar. Promedia el modelo sobre las observaciones de la entidad j (el promedio de una constante es la misma constante):

$$\bar{Y}_{\cdot j} = \beta\, \bar{X}_{\cdot j} + \alpha_j + \bar{\varepsilon}_{\cdot j}.$$

Ahora resta esta ecuación de la original $Y_{ij} = \beta X_{ij} + \alpha_j + \varepsilon_{ij}$:

$$(Y_{ij} - \bar{Y}_{\cdot j}) = \beta\,(X_{ij} - \bar{X}_{\cdot j}) + (\alpha_j - \alpha_j) + (\varepsilon_{ij} - \bar{\varepsilon}_{\cdot j}).$$

(b) α_j se esfuma. El término $(\alpha_j - \alpha_j) = 0$: ¡desapareció! Queda un modelo limpio en desviaciones,

$$\tilde{Y}_{ij} = \beta\, \tilde{X}_{ij} + \tilde{\varepsilon}_{ij},$$

donde ya *no está* la variable problemática. ¿Por qué esto mata el OVB? Porque el OVB por α_j venía de que α_j estaba *escondido en el error* y correlacionado con X ($\mathrm{Cov}(X, \alpha) \neq 0$). Al eliminarlo con la resta *within*, ya no puede contaminar a $\hat{\beta}$: OLS sobre las desviaciones (que es *exactamente* lo que hace incluir `factor(j)`) es insesgado aunque α_j correlacione con X.
La letra chica (importante). Esto funciona *solo* porque α_j es **invariante** dentro de la entidad. Una variable omitida que *cambie* dentro de la entidad no se cancela con la resta y seguiría sesgando. Los efectos fijos son poderosos, pero no son magia total: controlan lo fijo del grupo, no lo que varía dentro de él.

Ejercicio 1.30 (Elasticidad log-log con un cálculo real)**.** Se estima $\ln(\text{cantidad}) = 10 - 0.8\ln(\text{precio})$. Actualmente el precio es $P = 50$ y la cantidad predicha es $Q = 200$.

(a) Usando la elasticidad, estima el cambio porcentual en Q si el precio sube a 55 (un +10 por ciento).

(b) Aproxima el cambio en *unidades* de Q.

Solución

Qué nos piden. Pasar de una elasticidad a un cambio concreto —primero en por ciento, luego en unidades— y ver qué tan buena es la aproximación lineal.
(a) Cambio porcentual vía la elasticidad. El coef -0.8 es la elasticidad. Un $+10$ por ciento en el precio se traduce en:

$$\Delta\,\%\,Q \approx (-0.8) \times 10 = -8 \text{ por ciento.}$$

(b) De por ciento a unidades. Un -8 por ciento sobre $Q = 200$ es:

$$\Delta Q \approx -0.08 \times 200 = -16 \text{ unidades } \Rightarrow\ Q \approx 200 - 16 = 184.$$

Comprobación (por qué es una *aproximación*). La elasticidad usa cambios pequeños; con +10 por ciento el cálculo exacto usa la razón de precios:

$$\left(\tfrac{55}{50}\right)^{-0.8} = 1.1^{-0.8} \approx 0.927 \ \Rightarrow\ Q = 200 \times 0.927 \approx 185.$$

O sea el cambio exacto es ≈ -7.3 por ciento (unas 185 unidades), muy cerca del -8 por ciento de la aproximación lineal. En palabras: la regla "elasticidad × por ciento" es cómoda y precisa para cambios chicos; para saltos grandes conviene la potencia exacta, pero la diferencia aquí es mínima.

Ejercicio 1.31 (Dos omitidas con sesgos de signo opuesto)**.** Estimamos el efecto de la educación (X) sobre el salario, omitiendo *dos* variables: la *habilidad* W_1 (sube el salario, correlación positiva con X) y la *mala salud crónica* W_2 (baja el salario, y las personas con más educación tienden a tener *mejor* salud, i.e. $\text{Corr}(X, W_2) < 0$).

(a) Da el signo del sesgo que aporta cada omitida.

(b) ¿Se puede firmar (determinar) el signo del sesgo total? Explica.

Solución

Qué nos piden. Extender la regla de signos a *dos* omitidas a la vez, con la sutileza de que cada una podría empujar en su propia dirección.
La idea. Con varias omitidas, el sesgo total es (aprox.) la *suma* de los sesgos individuales: $\beta_{2,1}\delta_1 + \beta_{2,2}\delta_2$. Analiza cada sumando con la regla del producto, y luego pregunta si

apuntan al mismo lado.

(a) Signo de cada sesgo.

- **Habilidad** W_1: sube el salario ($\beta_{2,1} > 0$) y correlaciona positivo con educación ($\delta_1 > 0$). Sesgo: $(+)(+) = +$, **hacia arriba**.
- **Mala salud** W_2: baja el salario ($\beta_{2,2} < 0$) y más educación trae *menos* mala salud ($\delta_2 < 0$). Sesgo: $(-)(-) = +$, ¡también **hacia arriba**! (Los dos negativos se cancelan.)

(b) ¿Se puede firmar el total? Sí, en este caso. Los dos sesgos son positivos, así que su suma es positiva pase lo que pase con las magnitudes: el sesgo total es **hacia arriba** (la corta sobrestima el retorno de la educación). *La lección general*, en cambio, es de cautela: cuando las omitidas empujan en direcciones *opuestas* (un sesgo + y otro −), el signo del total **no** se puede determinar sin conocer las magnitudes —podría quedar hacia arriba, hacia abajo o cancelarse. Aquí tuvimos suerte porque, tras aplicar la regla de signos, ambos coincidieron.

Nivel 5 – Reto+

Muy difícil / fuera del scope típico: demostraciones y casos límite.

Ejercicio 1.32 (OVB y el teorema de Frisch–Waugh–Lovell (partialling-out))**.** El teorema FWL dice que el coeficiente de X en la larga $Y = \beta_0 + \beta_1 X + \beta_2 W + \varepsilon$ se obtiene así: (1) regresa X sobre W y guarda el residuo $\tilde{X}$; (2) regresa Y sobre $\tilde{X}$.

(a) Interpreta qué es $\tilde{X}$ y por qué "limpiar X de W" da el efecto *ceteris paribus*.

(b) Conecta con el OVB: ¿por qué la regresión corta (sobre X sin limpiar) da $\beta_1 + \beta_2\delta$ en vez de β_1?

Solución

Qué nos piden. Ver el "ceteris paribus" y el OVB desde un ángulo más profundo: el de *limpiar* (partial-out) una variable de la otra. Es el mismo fenómeno, contado como proyección.

(a) Qué es $\tilde{X}$. Al regresar X sobre W, la parte de X que W "explica" queda en el ajuste, y lo que sobra —el residuo $\tilde{X}$— es **la parte de X que W *no* puede explicar**, es decir, la variación de X *ortogonal* a W. Ahora, regresar Y sobre ese $\tilde{X}$ mide cómo se mueve Y cuando cambia la parte de X que *no* acompaña a W. Y eso es precisamente "mover X manteniendo W fijo": por construcción, $\tilde{X}$ ya tiene descontado todo lo que compartía con W. Por eso el coeficiente FWL coincide *exactamente* con β_1 de la larga: es el efecto *ceteris paribus*. El teorema nos dice que "controlar por W" y "limpiar X de W antes de regresar" son la misma cosa.

(b) Conexión con el OVB. La regresión corta usa X **completo**, sin limpiar. Pero X completo se puede descomponer en dos pedazos: el pedazo alineado con W (su proyección

sobre W, cuya pendiente es δ) y el pedazo limpio $\tilde{X}$. Cuando regresas Y sobre el X sin limpiar, el pedazo alineado con W *arrastra* el efecto de W sobre Y (que es β_2): por cada unidad de X, hay δ unidades que "viajan por la vía de W", aportando $\beta_2 \cdot \delta$ al coeficiente. Por eso la corta estima

$$\beta_1 + \beta_2\,\delta,$$

la misma fórmula del OVB, pero ahora *vista* como "no limpiaste X y se te coló el efecto de W". FWL y OVB son las dos caras de la misma moneda: limpiar elimina el $\beta_2\delta$; no limpiar lo deja dentro.

Ejercicio 1.33 (Sharp vs. fuzzy RDD)**.** En el RDD *nítido (sharp)* el tratamiento salta de 0 a 1 exactamente en a_0. En el *difuso (fuzzy)*, cruzar el umbral solo *aumenta la probabilidad* de tratarse (no todos cumplen).

(a) En el sharp, el efecto es (salto en Y). En el fuzzy, ¿por qué hay que *dividir* el salto en Y por algo?

(b) Escribe el estimador fuzzy como cociente e identifica numerador y denominador.

Solución

Qué nos piden. Entender qué cambia cuando el umbral no obliga, solo *empuja* —y por qué eso obliga a "escalar" el efecto.
(a) Por qué dividir. Piensa en el sharp: en el corte, la probabilidad de tratarse pasa de 0 a 1 (un cambio de 1 unidad completa), así que el salto en Y *es* el efecto del tratamiento, directo. En el fuzzy, cruzar el umbral no lleva a todos al tratamiento: la probabilidad salta, digamos, de 0.2 a 0.7 —solo 0.5 de cambio. Entonces el salto que observamos en Y corresponde a ese empujón *parcial*, no a "tratar a todo el mundo". Para recuperar el efecto *por unidad de tratamiento*, hay que dividir el salto en Y entre el salto en la probabilidad de tratarse: le preguntamos "¿cuánto se movió Y *por cada unidad extra* de tratamiento que el umbral logró inducir?".
(b) El estimador como cociente.

$$\hat{\tau}_{\text{fuzzy}} = \frac{\text{salto en } \mathbb{E}[Y \mid a] \text{ en } a_0}{\text{salto en } \mathbb{E}[D \mid a] \text{ en } a_0} = \frac{\lim\limits_{a \downarrow a_0} \mathbb{E}[Y \mid a] - \lim\limits_{a \uparrow a_0} \mathbb{E}[Y \mid a]}{\lim\limits_{a \downarrow a_0} \mathbb{E}[D \mid a] - \lim\limits_{a \uparrow a_0} \mathbb{E}[D \mid a]}.$$

Lectura de cada pieza. El **numerador** es el salto en el *resultado* (lo que en IV se llama la *forma reducida*: el efecto "bruto" de cruzar el umbral sobre Y). El **denominador** es el salto en la *probabilidad de tratarse* (la *primera etapa*: cuánto muerde el umbral en el tratamiento). Dividir uno entre otro es, ni más ni menos, un estimador de **variables instrumentales**, donde "cruzar el umbral" instrumenta al tratamiento. Esto es exactamente el puente a la Clase 11 (IV): el fuzzy RDD es IV con un instrumento que vive en un umbral.

Ejercicio 1.34 (Bad control formal: condicionar en un *collider*)**.** Sea X (tratamiento) aleatorio, M una variable *posterior* afectada por X y por un factor no observado U que también afecta Y: $M = X + U$, $Y = U + \text{ruido}$ (así X *no* tiene efecto directo sobre Y: el efecto verdadero es 0).

(a) Sin controlar por M, ¿qué estima la regresión de Y sobre X?

(b) Al controlar por M (un post-tratamiento), argumenta por qué aparece una asociación espuria entre X e Y.

Solución

Qué nos piden. Demostrar, en un ejemplo mínimo, algo que suena paradójico: controlar por una variable puede *crear* sesgo donde no lo había. Es la versión formal del "bad control".
(a) Sin controlar por M. X es aleatorio, así que es independiente de U y del ruido. Calculemos la covarianza que gobierna la regresión de Y sobre X:

$$\text{Cov}(X, Y) = \text{Cov}(X,\ U + \text{ruido}) = \text{Cov}(X, U) + \text{Cov}(X, \text{ruido}) = 0 + 0 = 0.$$

El coeficiente es 0. **Correcto**: el efecto verdadero de X sobre Y es cero, y la regresión simple lo recupera. No había nada que controlar.
(b) Al controlar por M (el desastre). $M = X + U$ es un **collider**: dos flechas *chocan* en él ($X \to M \leftarrow U$). Cuando "fijamos" $M = m$, imponemos la restricción $X + U = m$, es decir $U = m - X$. ¡Mira lo que pasó! Dentro de un valor fijo de M, el factor U quedó *atado negativamente* a X: si X sube, U debe bajar para mantener la suma en m. Y como Y depende de U ($Y = U + \text{ruido}$), al condicionar en M inducimos una correlación entre X e Y (negativa) que **no existía**. Formalmente, $\text{Cov}(X, Y \mid M) \neq 0$ aunque $\text{Cov}(X, Y) = 0$.
La moraleja. Controlar una variable *posterior* al tratamiento (un collider) abrió un camino falso entre X e Y: creó sesgo de la nada. Por eso la regla de oro "no controles por consecuencias del tratamiento" no es una superstición: es un teorema. Un buen control cierra puertas traseras; un bad control puede *abrir* una.

Ejercicio 1.35 (OVB en forma matricial (multivariado))**.** Modelo verdadero $Y = X\beta + W\gamma + \varepsilon$, con X la matriz de regresores incluidos (incluye la constante) y W los omitidos. La regresión corta estima $\tilde{\beta} = (X^\top X)^{-1} X^\top Y$.

(a) Sustituye el modelo verdadero y muestra que $\mathbb{E}[\tilde{\beta}] = \beta + (X^\top X)^{-1} X^\top W\, \gamma$.

(b) Identifica quién juega el papel de "δ" del caso simple y da la condición para que *no* haya sesgo.

Solución

Qué nos piden. Subir la fórmula del OVB a su versión "adulta", con matrices, para ver que la misma idea vale con *muchos* regresores y *muchas* omitidas a la vez.

(a) Sustituir y simplificar. Toma el estimador corto y mete dentro el modelo verdadero $Y = X\beta + W\gamma + \varepsilon$:

$$\tilde{\beta} = (X^\top X)^{-1} X^\top Y = (X^\top X)^{-1} X^\top \Big(X\beta + W\gamma + \varepsilon \Big).$$

Distribuye el $(X^\top X)^{-1} X^\top$ sobre los tres sumandos:

$$= \underbrace{(X^\top X)^{-1}(X^\top X)}_{=I}\, \beta + (X^\top X)^{-1} X^\top W\, \gamma + (X^\top X)^{-1} X^\top \varepsilon.$$

El primer bloque colapsa a la identidad por la inversa: $(X^\top X)^{-1}(X^\top X) = I$, dejando β. Toma esperanza; con exogeneidad $\mathbb{E}[X^\top \varepsilon] = 0$ el último término muere:

$$\mathbb{E}[\tilde{\beta}] = \beta + (X^\top X)^{-1} X^\top W\, \gamma.$$

(b) Quién es "δ" aquí. Fíjate en la matriz

$$\Pi = (X^\top X)^{-1} X^\top W.$$

Esa expresión tiene la forma *exacta* de un estimador OLS: son los coeficientes de **regresar cada columna de W (cada omitida) sobre todos los X incluidos**. Es decir, Π es la versión matricial de δ. El sesgo es $\Pi\gamma$: "cómo se relacionan las omitidas con los incluidos" (Π) por "cuánto afectan las omitidas a Y" (γ) —idéntico espíritu a $\beta_2\,\delta$, ahora en bloque. **Condición de cero sesgo.** $\mathbb{E}[\tilde{\beta}] = \beta$ (insesgado) si $\Pi\gamma = 0$. Dos formas naturales de lograrlo: (i) $\gamma = 0$, las omitidas no afectan a Y (no eran relevantes); o (ii) $X^\top W = 0$, las omitidas son **ortogonales** a los incluidos —lo que ocurre, por ejemplo, cuando X se asignó al azar (un RCT), pues la aleatorización rompe toda correlación con cualquier W. Es la misma moraleja del caso simple, ahora blindada con álgebra lineal.

Prácticas en R — házlas *desde cero*

Estos cuatro ejercicios se resuelven **en R, desde cero**, solo con el enunciado (que ya te dice el archivo de datos, sus columnas y los puntos a encontrar). **Si te trabas**, pide el archivo `pN_*.R` con espacios `____` para completar. El **solucionario** (código + números + interpretación) se entrega al final de la clase (y está en `pN_*_sol.R`).

Ejercicio 1.36 (R desde cero — Sesgo de variable omitida (`ovb.csv`))**.** Tienes el archivo `ovb.csv` con tres columnas: `salario`, `educacion` y `habilidad`. La *habilidad* casi nunca se observa en datos reales; aquí la incluimos a propósito para poder *ver* el sesgo. **Se pide** (todo en R):

(a) Corre la regresión *corta* de `salario` sobre `educacion` y anota el coeficiente de `educacion`.

(b) Corre la *larga* agregando `habilidad`; anota el nuevo coeficiente de `educacion` y el de `habilidad`.

(c) Calcula δ como la pendiente de regresar `habilidad` sobre `educacion`.

(d) Verifica que el cambio del coeficiente de `educacion` (corta − larga) es igual a $\beta_2 \cdot \delta$. ¿El sesgo es *hacia arriba* o *hacia abajo*? Interpreta.

Solución

Plan. Dos regresiones (corta y larga) y una auxiliar para δ; luego comparar.

Solución en R

```
datos <- read.csv("ovb.csv")
m_short <- lm(salario ~ educacion, data = datos)              # (a) coef
    educacion = 302
m_long  <- lm(salario ~ educacion + habilidad, data = datos)  # (b)
    educacion = 181, habilidad = 331
delta   <- coef(lm(habilidad ~ educacion, data = datos))["educacion"]  # (c
    ) 0.366
coef(m_short)["educacion"] - coef(m_long)["educacion"]        # 121 (cambio
     real)
coef(m_long)["habilidad"] * delta                            # 121 = beta2
    * delta
```

Lectura de los números. (a) corta: educación = 302. (b) larga: educación = 181, habilidad = 331. (c) $\delta = 0.366$. (d) El cambio $302 - 181 = 121$ coincide *exactamente* con $\beta_2 \cdot \delta = 331 \times 0.366 = 121$. Como $\beta_2 > 0$ y $\delta > 0$, el sesgo es **hacia arriba**: sin controlar la habilidad, la corta *sobrestimaba* el retorno de la educación (creía 302 cuando el efecto limpio es 181). Moraleja: cuando el coeficiente *salta* al agregar un control, había sesgo de variable omitida.

Ejercicio 1.37 (R desde cero — Formas funcionales: log y cuadrático (`mincer.csv`))**.** El archivo `mincer.csv` tiene `salario`, `educacion` y `experiencia`. Queremos un modelo donde el salario entre en *logaritmo* (para leer en por ciento) y la experiencia tenga un término *cuadrático* (para capturar que sube y luego se aplana). **Se pide** (en R):

(a) Ajusta ln(`salario`) sobre `educacion` y `experiencia`, agregando el término cuadrático con `I(experiencia^2)`.

(b) Interpreta el coeficiente de `educacion` como *semi-elasticidad* (en por ciento).

(c) Calcula el *punto de retorno* de la experiencia, $-\beta_1/(2\beta_2)$.

(d) ¿Qué le pasa al salario *después* de ese punto? Justifica con el signo de β_2.

Solución

Plan. Un solo `lm` con el término cuadrático; luego leer coeficientes.

Solución en R

```
datos <- read.csv("mincer.csv")
m <- lm(log(salario) ~ educacion + experiencia + I(experiencia^2), data =
    datos)
b <- coef(m)
100 * b["educacion"]                          # (b) semi-elasticidad ~ 8.5 %
     por anio
-b["experiencia"] / (2 * b["I(experiencia^2)"])   # (c) punto de retorno ~
    29 anios
```

Lectura. (b) El coeficiente de `educacion` (≈ 0.085) se lee en por ciento porque Y está en log: cada año de educación $\approx +8.5$ por ciento de salario. (c) El punto de retorno cae en ≈ 29 años de experiencia. (d) Como el coeficiente de `experiencia`^{2} es *negativo* ($\beta_2 < 0$, parábola cóncava), el salario sube con la experiencia hasta ~ 29 años y después *se aplana y cae* levemente (rendimientos decrecientes). Una recta no podría describir ese perfil.

Ejercicio 1.38 (R desde cero — Efectos fijos y paradoja de Simpson (`panel.csv`))**.** El archivo `panel.csv` tiene `distrito`, `gasto` (por alumno) y `rendimiento`, para 6 distritos. **Se pide** (en R):

(a) Corre la regresión *pooled* de `rendimiento` sobre `gasto` y anota la pendiente.

(b) Agrega `factor(distrito)` (efectos fijos) y anota la nueva pendiente.

(c) ¿Cambia el *signo*? Explica por qué (¿qué variable omitida hay?) y a cuál de las dos regresiones le creerías.

Solución

Plan. Comparar la pendiente del gasto con y sin controlar el distrito.

Solución en R

```
datos <- read.csv("panel.csv")
coef(lm(rendimiento ~ gasto, data = datos))["gasto"]                    #
    pooled = -0.38
coef(lm(rendimiento ~ gasto + factor(distrito), data = datos))["gasto"]  #
    efectos fijos = +1.35
```

Lectura. (a) La *pooled* da pendiente **negativa** (-0.38). (b) Con efectos fijos da **positiva** ($+1.35$). (c) ¡El signo se invierte! Es la **paradoja de Simpson**: los distritos con mejor

rendimiento de base gastan *menos* por alumno, y esa correlación entre distritos (espuria) arrastra la pooled hacia abajo. La variable omitida es "el distrito". Al controlarlo con `factor(distrito)`, comparamos cada distrito consigo mismo (variación *within*) y recuperamos la relación verdadera, positiva. **Créele al de efectos fijos.**

Ejercicio 1.39 (R desde cero — Regresión discontinua (`rdd.csv`))**.** El archivo `rdd.csv` tiene `puntaje`, `beca` y `resultado` (ingreso futuro). La beca se otorga si `puntaje`≥ 60. **Se pide** (en R):

(a) Quédate con una *ventana* alrededor del corte, $|\texttt{puntaje} - 60| \leq 15$, y crea la running variable *centrada* `x = puntaje - 60`.

(b) Ajusta `resultado` sobre `beca`, `x` y su interacción `beca:x`.

(c) El coeficiente de `beca` es el *salto* en el umbral. Interprétalo: ¿para quiénes vale ese efecto?

Solución

Plan. Recortar una ventana, centrar el puntaje y estimar una recta local a cada lado.

Solución en R

```
datos <- read.csv("rdd.csv")
win <- subset(datos, abs(puntaje - 60) <= 15)   # (a) ventana
win$x <- win$puntaje - 60                        #     running variable
    centrada
m <- lm(resultado ~ beca + x + beca:x, data = win)  # (b) recta local a
    cada lado
coef(m)["beca"]                                  # (c) el salto ~ 9  (
    verdadero = 8)
```

Lectura. El coeficiente de `beca` (≈ 9) es el **salto** del resultado justo en el corte: el efecto causal de la beca. Centramos `x` para que, en el umbral (`x` $= 0$), el coeficiente de `beca` sea exactamente ese salto. Es un efecto **local** (LATE): vale para los alumnos *cerca de* 60 (los que "apenas" ganan o pierden la beca), no para uno con 30 o con 90 puntos, que no tienen un "gemelo" del otro lado del corte.

Fuentes y atribución

Material de repaso **basado en MIT 14.310x – Data Analysis for Social Scientists** (Esther Duflo & Sara Fisher Ellison), MIT OpenCourseWare, licencia **CC BY-NC-SA 4.0** (uso no comercial, con atribución, compartir bajo la misma licencia), y en las fuentes de la bibliografía. Los enunciados se basan en el material del curso y en diversas fuentes abiertas; cuando un problema se adapta de una fuente específica, se cita en el enunciado.

Para esta clase (M9 — Regresión en la práctica y OVB) se usaron, además, como *referencia/inspiración* (sin copiar texto ni código): MIT 14.310x lecciones **L19–L20** (Prof. E. Duflo) y **Homework 9** (datos `fastfood.csv`); y los repositorios abiertos de la comunidad `mynameisjanus/14310xDataAnalysis` (Module 9: OVB, formas funcionales, RDD), `Data-Analysis-Cheat-Sheet-for-Social-Scientists` (Regression Practice: OVB, DiD), `dnackat/data-analysis-for-social-scientists-mitx` (Homework 9: Card & Krueger, RDD) y el *Fall 2020* Module 9 —todos **sin licencia explícita**, usados solo como referencia, no como copia. Textos de apoyo: Wooldridge, *Introductory Econometrics*; Angrist & Pischke, *Mostly Harmless Econometrics* (OVB, efectos fijos, RDD).

Bibliografía.

- Duflo, E. & Ellison, S. F. *14.310x: Data Analysis for Social Scientists*. MIT OpenCourseWare (slides, lecture notes y problem sets). ocw.mit.edu.
- Larsen, R. J. & Marx, M. L. *An Introduction to Mathematical Statistics and Its Applications*, 6.ª ed., Pearson, 2018.
- DeGroot, M. H. & Schervish, M. J. *Probability and Statistics*, 4.ª ed., Pearson, 2012.
- Blitzstein, J. & Hwang, J. *Introduction to Probability* (Harvard Stat 110); W. Chen, *Probability Cheatsheet*.
- Diez, D., Çetinkaya-Rundel, M. & Barr, C. *OpenIntro Statistics*. openintro.org.
- Materiales abiertos de la comunidad del curso en GitHub: `gevorg-hunanyan/probability`; `hanmochen/Probability-Theory-2020`; `mynameisjanus/14310xDataAnalysis`.