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

Teoría

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

Índice

1. De qué trata esta clase

En la Clase 9 aprendimos la *mecánica* del modelo lineal: cómo estimar $\mathbb{E}[Y \mid X] = \beta_0 + \beta_1 X$ por mínimos cuadrados (OLS) y cómo leer los coeficientes. Esta clase es la del **oficio**: ¿qué decisiones tomamos al *correr una regresión en la vida real*? ¿Qué variables incluir? ¿Y si la relación no es una recta? ¿Qué pasa cuando *falta* una variable importante?

La gran lección de la Clase 9 fue que $\hat{\beta}_j$ se lee *ceteris paribus* (manteniendo lo demás constante). El reverso de esa moneda es el tema central de hoy: **si dejas fuera una variable relevante, el coeficiente que sí estimas sale "contaminado"**. Ese es el **sesgo de variable omitida (omitted variable bias, OVB)**, el concepto más importante para interpretar regresiones con datos observacionales.

Mapa de la clase

Cuatro grandes decisiones prácticas al correr regresiones:

1. **¿Qué controlar?** Variables de control (covariables): cuándo ayudan y cuándo estorban (*bad controls*).
2. **¿Qué forma funcional?** Logaritmos, polinomios (término cuadrático), interacciones: cuando la recta no basta.
3. **¿Qué pasa si falta una variable?** El **OVB**: su fórmula, su *signo* y cómo los **efectos fijos (fixed effects)** ayudan a controlarlo.
4. **¿Cuándo el diseño da causalidad?** Introducción a la **regresión discontinua (regression discontinuity design, RDD)**.

Cerramos con los errores estándar en la práctica (heterocedasticidad, robustos) y cómo hacer todo esto en R.

2. Variables de control (covariables)

Definición 2.1 (Variable de control (control variable / covariate))**.** Una **variable de control** es un regresor que añadimos al modelo *no* porque nos interese su coeficiente, sino para **limpiar** el coeficiente de la variable que sí nos importa. En

$$Y_i = \beta_0 + \beta_1 X_i + \underbrace{\gamma_1 W_{1i} + \cdots + \gamma_p W_{pi}}_{\text{controles}} + \varepsilon_i,$$

X es la variable de interés y las W son los controles. $\hat{\beta}_1$ pasa a ser el efecto de X *manteniendo fijos* los controles.

Si queremos el efecto de la educación sobre el salario pero las personas más educadas también tienen (digamos) más experiencia, comparar sus salarios mezcla las dos cosas. *Controlar* por experiencia es comparar personas con la **misma** experiencia: así aislamos el papel de la educación. Cada control que agregas "cierra una puerta" por la que se colaba una explicación alternativa.

La tentación es "controlar por todo". Pero un **mal control (bad control)** es una variable que es ella misma un *resultado* del tratamiento X (un *outcome*, no una causa previa). Controlar por algo posterior a X puede **introducir** sesgo en vez de quitarlo. Ejemplo: para el efecto de la educación (X) sobre el salario (Y), **no** controles por "tipo de ocupación": la ocupación es un *canal* por el que la educación sube el salario; al fijarla, borras parte del efecto que quieres medir. Regla práctica: controla por variables **predeterminadas** (anteriores a X: edad, sexo, región de nacimiento), no por consecuencias de X.

Ejemplo 2.1 (Un buen control cambia la conclusión)**.** Regresión del *gasto en salud* (Y) sobre *si el hogar recibe un programa social* (X). Sin controlar, los hogares del programa gastan *menos* en salud —pero son más pobres de entrada. Al controlar por **ingreso** (una variable *predeterminada*), el signo se invierte: a igual ingreso, el programa se asocia con *más* gasto en salud. El ingreso era un confusor; controlarlo era lo correcto. (Controlar por "si tiene seguro", que el programa mismo otorga, sería un *bad control*.)

3. Formas funcionales: cuando la recta no basta

Hasta ahora el modelo era lineal en X. Pero muchas relaciones reales *se curvan*: el salario sube con la experiencia pero cada vez menos; un ingreso que se duplica no dobla el gasto en un bien. La buena noticia (Duflo, L20): **seguimos usando OLS**; solo cambiamos *en qué forma* entran las variables.

3.1. Logaritmos: interpretar en porcentajes y elasticidades

Las tres combinaciones con logaritmos

Según quién va en logaritmo, el coeficiente β_1 se lee distinto:

Modelo	**Ecuación**	**Lectura de β_1**
lin-lin (nivel-nivel)	$Y = \beta_0 + \beta_1 X$	$+1$ en $X \Rightarrow +\beta_1$ en Y (unidades)
log-lin (semi-elasticidad)	$\ln Y = \beta_0 + \beta_1 X$	$+1$ en $X \Rightarrow \approx 100\,\beta_1$ por ciento en Y
lin-log	$Y = \beta_0 + \beta_1 \ln X$	$+1$ por ciento en $X \Rightarrow \approx \beta_1/100$ en Y
log-log (elasticidad)	$\ln Y = \beta_0 + \beta_1 \ln X$	$+1$ por ciento en $X \Rightarrow \approx \beta_1$ por ciento en Y

El caso **log-log** da directamente la **elasticidad (elasticity)**, un número sin unidades muy usado en economía.

Un cambio pequeño en $\ln Y$ es (aproximadamente) el *cambio relativo* de Y: $\Delta \ln Y \approx \Delta Y/Y$. Por eso, cuando Y entra en log, el coeficiente mide cambios *porcentuales* en lugar de cambios en soles. Es ideal cuando lo que importa es "cuánto por ciento", no "cuántos soles", y cuando Y es siempre positivo y muy asimétrico (salarios, precios, población).

Ejemplo 3.1 (Semi-elasticidad de la educación (datos de la práctica))**.** Con `mincer.csv` estimamos $\ln(\text{salario}) = \beta_0 + 0.085 \cdot \text{educación} + \dots$ El coeficiente 0.085 es una **semi-elasticidad**: cada año adicional de educación se asocia con un salario 8.5 **por ciento** más alto (aprox.), a igualdad del resto. En unidades serían "soles", pero en log leemos "por ciento": mucho más natural para comparar entre personas de distinto nivel salarial.

3.2. Polinomios: el término cuadrático y el punto de retorno

Modelo cuadrático (quadratic / polynomial model)

Para capturar una relación *curva* agregamos X^2 como un regresor más:

$$Y = \beta_0 + \beta_1 X + \beta_2 X^2 + \varepsilon.$$

Sigue siendo OLS (lineal *en los parámetros*). El *efecto* de X ya no es constante: $\dfrac{\partial Y}{\partial X} = \beta_1 + 2\beta_2 X$ depende del nivel de X. Si $\beta_2 < 0$ la curva es **cóncava** (sube y luego baja); el máximo —el **punto de retorno (turning point)**— está en

$$\boxed{X^* = -\frac{\beta_1}{2\beta_2}\,.}$$

Ejemplo 3.2 (Perfil de experiencia (Mincer))**.** Con `mincer.csv`, $\ln(\text{salario}) = \cdots + 0.041 \cdot \text{exper} - 0.0007 \cdot \text{exper}^2$. Como $\beta_2 < 0$, el salario sube con la experiencia pero *a tasa decreciente*. El punto de retorno es $X^* = -0.041/(2 \cdot (-0.0007)) \approx 29$ años: hasta los ~ 29 años de experiencia el salario crece; después se aplana y cae levemente. Sin el término cuadrático, una recta forzaría una pendiente constante y malinterpretaría todo el perfil.

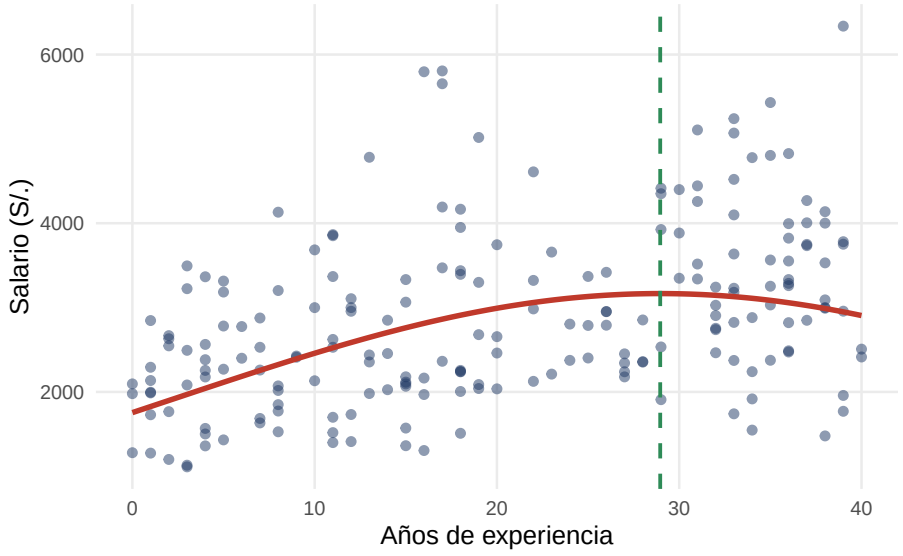

Figura 1: Perfil de experiencia con **término cuadrático**. La curva roja ($\beta_2 < 0$) sube y luego se aplana; la línea verde punteada marca el **punto de retorno** $X^* \approx 29$ años. Una sola recta no podría capturar esta forma.

3.3. Interacciones y dummies por tramos

Interacción (interaction): cuando un efecto depende de otra variable

Una **interacción** es el producto de dos regresores: $X \cdot W$. En $Y = \beta_0 + \beta_1 X + \beta_2 W + \beta_3 (X \cdot W) + \varepsilon$, el efecto de X es $\beta_1 + \beta_3 W$: *depende del valor de* W. Ejemplo: el retorno de la educación (X) podría ser distinto para hombres y mujeres (W). El DiD de la Clase 9 (`tratado`×`post`) era exactamente una interacción.

Otra forma de "doblar" la recta sin comprometerse a una fórmula: **partir el rango de X en intervalos** y meter una dummy por tramo. Así cada tramo tiene su propio nivel (o su propia pendiente, con las variables *piecewise linear*). Es una manera flexible y sencilla de dejar que los datos elijan la forma —la antesala de la regresión no paramétrica.

4. Sesgo de variable omitida (Omitted Variable Bias)

Este es el corazón de la clase. La pregunta: si el modelo "verdadero" incluye una variable pero yo la **omito**, ¿qué le pasa al coeficiente que sí estimo?

La fórmula del OVB: regresión "corta" vs "larga"

Supongamos que el modelo correcto (**long regression**) es

$$Y = \beta_0 + \beta_1 X + \beta_2 W + \varepsilon,$$

pero corremos la **short regression** $Y = \tilde{\beta}_0 + \tilde{\beta}_1 X + u$ (omitimos W). Entonces

$$\boxed{\mathbb{E}[\tilde{\beta}_1] = \beta_1 + \beta_2\,\delta\,,} \qquad \text{donde } \delta = \frac{\widehat{\mathrm{Cov}}(X,W)}{\widehat{\mathrm{Var}}(X)} \text{ es la pendiente de } W \sim X.$$

El **sesgo** es $\beta_2\,\delta$: el efecto de la omitida sobre Y (β_2) por la relación entre la omitida y el regresor incluido (δ).

El sesgo $\beta_2\,\delta$ es cero si *cualquiera* de los dos factores es cero:

- Si $\beta_2 = 0$ (la omitida *no* afecta a Y): no importa omitirla.
- Si $\delta = 0$ (la omitida *no* se correlaciona con X): tampoco sesga —por eso la aleatorización de un RCT (Clase 8) "salva" a OLS: hace a X independiente de todo lo demás, así que $\delta = 0$ para cualquier omitida.

Solo una variable que afecta a Y *y* se correlaciona con X (un confusor) genera OVB.

La tabla de signos: ¿hacia arriba o hacia abajo?

El signo del sesgo es el signo del producto $\beta_2 \cdot \delta$:

	$\mathrm{Corr}(X,W) > 0$ $(\delta > 0)$	$\mathrm{Corr}(X,W) < 0$ $(\delta < 0)$
$\beta_2 > 0$ (la omitida sube Y)	sesgo **hacia arriba** $(+)$	sesgo **hacia abajo** $(-)$
$\beta_2 < 0$ (la omitida baja Y)	sesgo **hacia abajo** $(-)$	sesgo **hacia arriba** $(+)$

"Hacia arriba" significa que $\tilde{\beta}_1$ *sobrestima* a β_1; "hacia abajo", que lo *subestima*. Saber el signo permite razonar sobre el sesgo *aunque no tengamos* la variable omitida.

Ejemplo 4.1 (OVB con la "habilidad" omitida (datos verificados de la práctica))**.** El caso clásico: el retorno de la educación sobre el salario, con la **habilidad (ability)** omitida —no la observamos, pero mueve tanto la educación (gente más hábil estudia más: $\delta > 0$) como el salario ($\beta_2 > 0$). Con `ovb.csv` (donde *sí* guardamos la habilidad para poder *ver* el sesgo):

corta: $\widehat{\text{salario}} = \cdots + \mathbf{302}\cdot\text{educación}$, larga: $\widehat{\text{salario}} = \cdots + \mathbf{181}\cdot\text{educación} + 331\cdot\text{habilidad}$.

El coeficiente de educación *cae* de 302 a 181 al controlar la habilidad. La fórmula lo predice *exactamente*: con Corr(educ,habil) = 0.83, la pendiente $\delta = 0.366$ y $\beta_2 = 331$, el sesgo es

$$\beta_2\,\delta = 331 \times 0.366 = \mathbf{121} = 302 - 181.$$

Como $\beta_2 > 0$ y $\delta > 0$, el sesgo es **hacia arriba**: sin controlar la habilidad, ¡*sobrestimamos* el retorno de la educación en 121 soles por año!

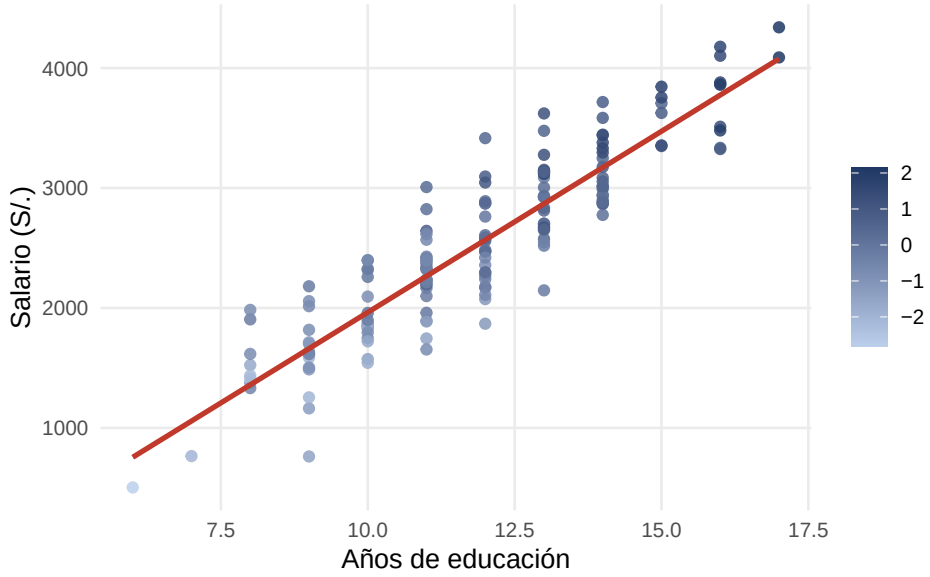

Figura 2: Salario vs. educación, con cada punto coloreado por su **habilidad** (más oscuro = más hábil). Los más educados tienden a ser más hábiles (color más oscuro a la derecha): parte de la pendiente cruda (recta roja, 302) es en realidad el efecto de la habilidad. Al controlarla, el retorno "limpio" de la educación baja a ~ 181.

El OVB es exactamente lo que ocurre cuando falla el supuesto de **exogeneidad** del modelo clásico: al meter la omitida dentro del error, X queda correlacionado con el error ($\text{Cov}(X, \varepsilon) \neq 0$) y $\hat{\beta}_1$ deja de ser insesgado. Por eso los coeficientes de una regresión observacional miden *asociación*, no *causa*, salvo que hayamos controlado todos los confusores relevantes.

5. Efectos fijos (fixed effects)

Muchas variables omitidas son **características fijas de un grupo**: la "cultura" de un distrito, la calidad de gestión de una escuela, factores no medidos de un individuo seguido en el tiempo. Los **efectos fijos** son un truco elegante para controlarlas *todas a la vez*, aunque no las midamos.

Efectos fijos (fixed effects) via dummies de entidad

Si los datos se agrupan en entidades j (distritos, escuelas, personas, años), añadimos una **dummy por entidad** α_j:

$$Y_{ij} = \alpha_j + \beta X_{ij} + \varepsilon_{ij}.$$

Cada α_j absorbe *todo* lo que es constante dentro de la entidad j (observado o no). El $\hat{\beta}$ resultante usa solo la variación **dentro de (within)** cada entidad: compara a j consigo mismo, no a un j con otro.

Si la variable omitida es *fija dentro de la entidad* (p. ej. la calidad de gestión de una escuela, que no cambia entre sus alumnos), el efecto fijo α_j la captura por completo: ya no está en el error, ya no sesga. Es como controlar por "todo lo que hace especial a esa entidad" sin tener que medirlo. **Limitación:** solo controla lo *invariante* dentro de la entidad; algo que varíe dentro (y sea confusor) sigue sesgando.

Ejemplo 5.1 (La paradoja de Simpson: pooled vs. within)**.** Con `panel.csv` (6 distritos; X = gasto por alumno, Y = rendimiento):

- **Pooled** (ignora el distrito): $\hat{\beta} = -0.38$ —¡pendiente *negativa*! Parece que gastar más *empeora* el rendimiento.
- **Con efectos fijos** `+ factor(distrito)`: $\hat{\beta} = +1.35$ —*positiva*, la verdadera relación dentro de cada distrito.

¿Qué pasó? Los distritos con mayor rendimiento "de base" (alto α_j) gastan *menos* por alumno: un confusor de entidad que correlaciona negativamente X con α. Al ignorarlo (pooled) el signo se invierte —la **paradoja de Simpson**. Los efectos fijos lo arreglan. Es OVB en acción: la omitida era "el distrito".

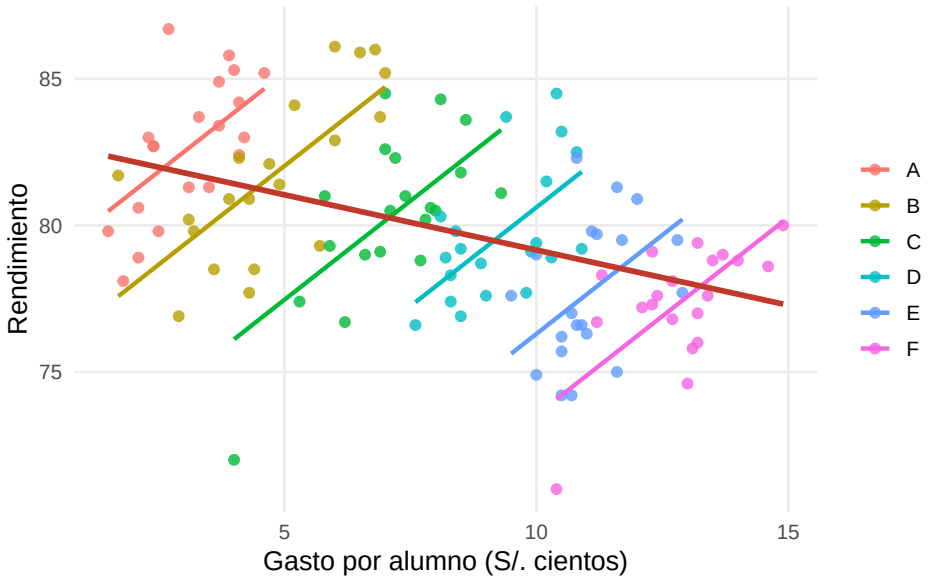

Figura 3: **Paradoja de Simpson.** Dentro de cada distrito (cada color) más gasto se asocia con *mejor* rendimiento (rectas *within* con pendiente positiva). Pero la recta **pooled** (roja, que ignora el distrito) tiene pendiente *negativa*, porque los distritos con más rendimiento de base gastan menos. Controlar el distrito con efectos fijos recupera el signo correcto.

6. Errores estándar en la práctica

Elegir bien los regresores da coeficientes correctos; pero para *la inferencia* (errores estándar, pruebas t, IC de la Clase 9) importa también un supuesto que se viola con frecuencia: la **homocedasticidad**.

Heterocedasticidad y errores estándar robustos

La **heterocedasticidad (heteroskedasticity)** es cuando la varianza del error *no* es constante: $\text{Var}(\varepsilon_i) = \sigma_i^2$ cambia con X (la nube "abre como un embudo"). No sesga los $\hat{\beta}$, pero **sí** sesga sus errores estándar clásicos —y con ello los t, p-valores e intervalos. Solución estándar: **errores estándar robustos (robust / heteroskedasticity-consistent, tipo White)**, que no asumen varianza constante.

Si los datos están agrupados (alumnos dentro de aulas, hogares dentro de distritos) y los errores se parecen dentro del grupo, hay que usar **errores estándar agrupados (clustered standard errors)** al nivel del grupo. Ignorarlo suele *subestimar* la incertidumbre (IC demasiado angostos, falsas significancias). Regla práctica: agrupa al nivel donde se asignó el tratamiento o donde hay dependencia.

7. Introducción a la regresión discontinua (RDD)

Cerramos con una joya del diseño: una situación en la que una simple regresión *sí* recupera un efecto causal, sin necesidad de un experimento aleatorizado.

Regression Discontinuity Design (RDD)

El **RDD** sirve cuando un tratamiento se asigna según si una variable continua —la **running variable** a— cruza un **umbral (threshold, cutoff)** a_0. Formalmente $D = \mathbf{1}[a \geq a_0]$. Ejemplos (Duflo, L20): una beca para quien saca al menos P puntos; la edad mínima legal para beber ($a_0 = 21$); una elección ganada o perdida en el 50 por ciento. El análisis más simple usa la dummy para *desplazar el intercepto* en el corte:

$$Y_i = \beta_0 + \tau\, D_i + \gamma\,(a_i - a_0) + \varepsilon_i,$$

y τ (el **salto** en a_0) es el efecto causal local.

Alguien con 59.9 puntos y alguien con 60.1 son **casi idénticos** en todo —esfuerzo, talento, contexto— salvo en una cosa: uno recibió la beca y el otro no. El umbral crea una "mini-aleatorización" local. Si el resultado Y *saltara* en a_0, ese salto no puede deberse a nada que cambie *suavemente* con el puntaje: solo al tratamiento. El **supuesto clave**: en ausencia del programa, Y habría seguido *sin saltos* a través del umbral (continuidad).

Ejemplo 7.1 (Beca por puntaje (datos de la práctica))**.** Con `rdd.csv` (beca si el puntaje ≥ 60), estimando una recta local a cada lado del corte en una ventana de ± 15 puntos, el salto estimado es $\hat{\tau} \approx 9.1$ (el verdadero es 8). Interpretación: cruzar el umbral y recibir la beca eleva el resultado (ingreso futuro) en ~ 9 unidades, *para los que están cerca del corte*. Es un efecto **local** (LATE): vale cerca de a_0, no necesariamente para un alumno con 30 o con 90 puntos.

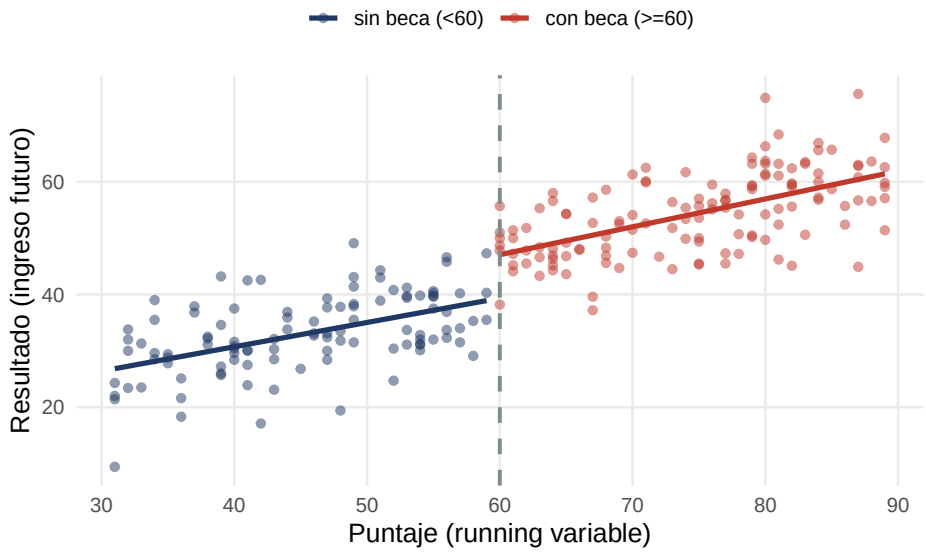

Figura 4: **RDD:** el resultado sube suave con el puntaje, pero **salta** en el umbral (línea gris, $a_0 = 60$) por efecto de la beca. La distancia vertical entre las dos rectas *en el corte* es el efecto causal local $\hat{\tau} \approx 9$. Lejos del umbral no comparamos "peras con peras"; solo cerca de a_0 los dos grupos son comparables.

8. Regresión en la práctica en R

Todo lo de esta clase se hace con la misma `lm()` de la Clase 9, cambiando cómo entran las variables. Lo nuevo: `log()`, `I(...^2)` para el cuadrático, `:` o `*` para interacciones y `factor()` para efectos fijos.

Controles, logaritmos, cuadrático e interacciones

```
datos <- read.csv("mincer.csv")

# log-lin: coef de educacion = semi-elasticidad (~8.5% por anio)
m_log  <- lm(log(salario) ~ educacion + experiencia, data = datos)

# termino cuadratico: I(experiencia^2) protege el "^" dentro de la formula
m_quad <- lm(log(salario) ~ educacion + experiencia + I(experiencia^2),
             data = datos)
```

```
b <- coef(m_quad)
turning <- -b["experiencia"] / (2 * b["I(experiencia^2)"])   # ~29 anios

# interaccion: el retorno de educacion depende del sexo (educacion:sexo)
# m_int <- lm(log(salario) ~ educacion * factor(sexo), data = datos)
```

Sesgo de variable omitida y efectos fijos

```
ovb <- read.csv("ovb.csv")
m_short <- lm(salario ~ educacion, data = ovb)               # omite habilidad
    : 302
m_long  <- lm(salario ~ educacion + habilidad, data = ovb)   # controla:
     181
# el coef de educacion CAE al controlar => habia sesgo hacia arriba

panel <- read.csv("panel.csv")
m_pool <- lm(rendimiento ~ gasto, data = panel)                    # pooled:
   -0.38
m_fe   <- lm(rendimiento ~ gasto + factor(distrito), data = panel) # FE:
   +1.35
# factor(distrito) = una dummy por distrito = efectos fijos
```

Para SE robustos y agrupados se usa el paquete `sandwich` (con `lmtest`): `coeftest(m, vcov = vcovHC(m, "HC1"))`. Para un RDD básico se centra la running variable y se estima una recta local a cada lado: `lm(y ~ D + I(a-a0) + D:I(a-a0), data = ventana)`; el coeficiente de `D` es el salto $\hat{\tau}$. (Existe el paquete `rdrobust` para hacerlo bien: elige la ventana óptima y reporta el efecto.)

Resumen

Resumen de la Clase 10

- **Controles:** agregar covariables limpia el coeficiente de interés (*ceteris paribus*). Controla por variables *predeterminadas*; evita los *bad controls* (consecuencias del tratamiento).
- **Formas funcionales:** sigue siendo OLS. **Log** ⇒ leer en por ciento (semi-elasticidad) o elasticidad (log-log). **Cuadrático** ⇒ efecto no constante, punto de retorno $X^* = -\beta_1/(2\beta_2)$. **Interacción** ⇒ un efecto que depende de otra variable.
- **OVB (lo más importante):** omitir una variable relevante sesga el coeficiente: $\mathbb{E}[\tilde{\beta}_1] = \beta_1 + \beta_2\delta$. Hace falta que la omitida *afecte a* Y ($\beta_2 \neq 0$) *y se correlacione con* X ($\delta \neq 0$). El signo del sesgo = signo de $\beta_2 \cdot \delta$.

- **Efectos fijos:** dummies de entidad (`factor()`) absorben todo lo constante dentro del grupo ⇒ controlan confusores invariantes sin medirlos (usan la variación *within*). Cuidado con la paradoja de Simpson.
- **Errores estándar:** la heterocedasticidad no sesga $\hat{\beta}$ pero sí sus SE ⇒ usar SE **robustos**; con datos agrupados, SE **clustered**.
- **RDD:** tratamiento asignado por un umbral en una *running variable*; el **salto** τ en el corte identifica un efecto causal *local*, bajo el supuesto de continuidad.

Mensaje central: en la práctica, un coeficiente de regresión solo significa lo que uno cree que significa si el modelo tiene las variables correctas y la forma correcta. El **OVB** es la advertencia permanente: pregunta siempre "¿qué me falta controlar, y hacia dónde me sesgaría?". La Clase 11 llevará esto al extremo con *variables instrumentales* para cuando *no podemos* controlar el confusor.

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`.