Econometría aplicada con R para la investigación

Guía práctica paso a paso para tesis, artículos científicos y análisis de datos

Willian Sebastian Flores Sotelo

Lima, Perú; 2026

Econometría aplicada con R para la investigación
Guía práctica paso a paso para tesis, artículos científicos y análisis de datos
Serie Econometría sin miedo, n.º 4
Autor-editor: © Willian Sebastian Flores Sotelo. Todos los derechos reservados.
Jr. Piura 282, piso 2, Urb. La Libertad, Comas – Lima – Perú
Primera edición digital: octubre de 2026.
Hecho el Depósito Legal en la Biblioteca Nacional del Perú N.° 2026-13818
Libro digital disponible en: Amazon Kindle, Apple Books y Google Play Books.
Ninguna parte de esta obra puede reproducirse sin autorización escrita del autor, salvo citas breves con fines de reseña o docencia.
Los scripts .R del libro se ofrecen para uso educativo. Las bases de datos citadas pertenecen a sus instituciones productoras.
Edición digital. Distribuido por Amazon Kindle Direct Publishing, Apple Books y Google Play Books.
R es software libre de The R Foundation for Statistical Computing. RStudio es una marca de Posit Software, PBC.
Este libro es una obra independiente: no está patrocinado, autorizado ni avalado por The R Foundation ni por Posit.
Stata, EViews, SPSS y Excel son marcas de sus respectivos titulares.

Contenido breve

ACERCA DEL AUTOR iv
PREFACIO: CÓMO USAR ESTE LIBRO xi
Parte Uno. Pensar como econometrista 1
Capítulo 1 ¿Por qué tu regresión miente? Correlación y causalidad 2
Capítulo 2 Tus datos listos en R en una tarde: instalar R y RStudio, de la ENAHO y el Banco Mundial a una base limpia 14
Capítulo 3 Prueba de hipótesis e intervalos de confianza: la estadística que tu investigación necesita 35
Parte Dos. La regresión, bien hecha 48
Capítulo 4 Regresión lineal simple paso a paso 49
Capítulo 5 Regresión lineal múltiple y el sesgo de variable omitida 61
Capítulo 6 Cómo interpretar una regresión sin pánico 74
Capítulo 7 Supuestos de la regresión: heterocedasticidad, multicolinealidad y autocorrelación 88
Capítulo 8 Regresión logística: modelos logit y probit 102
Parte Tres. Datos que se mueven en el tiempo 116
Capítulo 9 Datos de panel: efectos fijos vs. aleatorios y test de Hausman 117
Capítulo 10 Series de tiempo y modelo ARIMA paso a paso 133
Capítulo 11 Pronósticos con modelos VAR y cointegración 149
Parte Cuatro. Evaluación de impacto 163
Capítulo 12 Evaluación de impacto I: diferencias en diferencias 164
Capítulo 13 Evaluación de impacto II: variables instrumentales 178
Capítulo 14 Evaluación de impacto III: regresión discontinua y propensity score matching 190
Parte Cinco. Del resultado a la tesis y al artículo 205
Capítulo 15 Cómo redactar los resultados de tu tesis o artículo científico 206
APÉNDICE A. EQUIVALENCIAS R · STATA · EVIEWS · SPSS 218
APÉNDICE B. BASES DE DATOS DEL LIBRO 223

Contenido

ACERCA DEL AUTOR iv

PREFACIO: CÓMO USAR ESTE LIBRO..... xi

Parte Uno. Pensar como econometrista 1

Capítulo 1 ¿Por qué tu regresión miente? Correlación y causalidad 2

Lo que vas a lograr 2

1.1 Las que pidieron crédito no eran iguales a las que no pidieron 3

1.2 Dos negocios que nunca verás a la vez..... 3

1.3 Tres piezas dentro de una diferencia 3

1.4 Un dibujo de tres flechas 4

1.5 Cuando una moneda decide 4

1.6 Tu tesis sin sorteo 4

1.7 Manos a la obra..... 5

1.8 El error del tesista apurado: confundir significativo con causal 10

1.9 El error que cometen los investigadores en la redacción de artículos 11

RESUMEN Y CONCLUSIONES..... 11

EJERCICIOS RESUELTOS..... 11

Para tu tesis o artículo 13

Referencias 13

Capítulo 2 Tus datos listos en R en una tarde: instalar R y RStudio, de la ENAHO y el Banco Mundial a una base limpia 14

Lo que vas a lograr 14

2.1 R es el motor y RStudio, el tablero..... 15

2.2 Un proyecto por investigación..... 17

2.3 Paquetes: se instalan una vez y se cargan en cada sesión 18

2.4 La ENAHO viene en piezas..... 19

2.5 Cada encuestado vale por cientos 19

2.6 Manos a la obra..... 20

2.7 El error que no da ningún mensaje: la llave convertida en número 31

2.8 El error que cometen los investigadores en la redacción de artículos 31

RESUMEN Y CONCLUSIONES..... 32

EJERCICIOS RESUELTOS..... 32

Para tu tesis o artículo 33

Referencias 34

Capítulo 3 Prueba de hipótesis e intervalos de confianza: la estadística que tu investigación necesita 35

Lo que vas a lograr 35

3.1 Si la ENAHO hubiera sorteado otras casas . 36

3.2 El intervalo de confianza: tu cifra con su margen 36

3.3 La prueba de hipótesis: supón lo aburrido y mide tu sorpresa..... 36

3.4 Veinte pruebas, un hallazgo seguro 37

3.5 Manos a la obra..... 37

3.6 El error que convierte el azar en hallazgo: probar hasta que algo salga..... 44

3.7 El error que cometen los investigadores en la redacción de artículos 44

RESUMEN Y CONCLUSIONES..... 45

EJERCICIOS RESUELTOS..... 45

Para tu tesis o artículo 46

Referencias 47

Parte Dos. La regresión, bien hecha 48

Capítulo 4 Regresión lineal simple paso a paso 49

Lo que vas a lograr 49

4.1 Una recta para treinta y tres años..... 50

4.2 ¿Qué recta elige R? La de los cuadrados más pequeños 50

4.3 Cómo leer la pendiente y la constante 50

4.4 El R^2: cuánto de la variación explica la recta 51
4.5 Manos a la obra 51
4.6 El error que más se ve en las tesis de macroeconomía: no mirar quién dibuja la recta 57
4.7 El error que cometen los investigadores en la redacción de artículos 57
RESUMEN Y CONCLUSIONES 58
EJERCICIOS RESUELTOS 58
Para tu tesis o artículo 59
Referencias 60
Capítulo 5 Regresión lineal múltiple y el sesgo de variable omitida 61
Lo que vas a lograr 61
5.1 Qué significa «manteniendo lo demás constante» 62
5.2 La variable que falta deja una huella medible 62
5.3 ¿Hacia dónde empuja el sesgo? 62
5.4 No todo control es un buen control 63
5.5 Variables dicotómicas y categorías de referencia 63
5.6 El R^2 siempre sube; el ajustado, no 63
5.7 Manos a la obra 63
5.8 El error que comete casi todo tesista: creer que más controles siempre acercan a la verdad 70
5.9 El error que cometen los investigadores en la redacción de artículos 70
RESUMEN Y CONCLUSIONES 70
EJERCICIOS RESUELTOS 71
Para tu tesis o artículo 72
Referencias 72
Capítulo 6 Cómo interpretar una regresión sin pánico 74
Lo que vas a lograr 74
6.1 Cuatro pisos en una sola salida 75
6.2 Cada coeficiente responde a una comparación 75
6.3 Niveles y logaritmos: cuatro maneras de leer una pendiente 75
6.4 Ficticias, categorías e interacciones 76
6.5 Cuando la pendiente cambia: efectos marginales 76
6.6 Manos a la obra 76
6.7 El error que nace de leer una fila sin mirar las demás 84
6.8 El error que cometen los investigadores en la redacción de artículos 85
RESUMEN Y CONCLUSIONES 85
EJERCICIOS RESUELTOS 85
Para tu tesis o artículo 86
Referencias 87
Capítulo 7 Supuestos de la regresión: heterocedasticidad, multicolinealidad y autocorrelación 88
Lo que vas a lograr 88
7.1 Supuestos que protegen al coeficiente y supuestos que protegen al error estándar 89
7.2 Heterocedasticidad: el ruido no es igual para todos 89
7.3 Errores agrupados: vecinos que se parecen 89
7.4 Multicolinealidad: dos variables que se pisan 89
7.5 Autocorrelación: el error de este mes recuerda al anterior 90
7.6 Manos a la obra 90
7.7 El error que convierte las pruebas de supuestos en un trámite 98
7.8 El error que cometen los investigadores en la redacción de artículos 99
RESUMEN Y CONCLUSIONES 99
EJERCICIOS RESUELTOS 100
Para tu tesis o artículo 101
Referencias 101
Capítulo 8 Regresión logística: modelos logit y probit 102

Lo que vas a lograr 102
8.1 Cuando la respuesta es sí o no 103
8.2 La curva en forma de S 103
8.3 Tres maneras de traducir un logit 103
8.4 Cómo saber si el modelo ajusta bien 103
8.5 Manos a la obra 104
8.6 El error que convierte un odds ratio en una probabilidad 112
8.7 El error que cometen los investigadores en la redacción de artículos 112
RESUMEN Y CONCLUSIONES 113
EJERCICIOS RESUELTOS 113
Para tu tesis o artículo 114
Referencias 115
Parte Tres. Datos que se mueven en el tiempo 116
Capítulo 9 Datos de panel: efectos fijos vs. aleatorios y test de Hausman 117
Lo que vas a lograr 117
9.1 Un panel: los mismos países, muchos años 118
9.2 Lo que no se mide y no cambia 118
9.3 Efectos fijos: cada país contra sí mismo 118
9.4 Los efectos de año: el viento que sopla para todos 118
9.5 Efectos aleatorios: restar solo una parte 119
9.6 El test de Hausman, sin mitos 119
9.7 Errores agrupados: los años de un país no son independientes 119
9.8 Manos a la obra 119
9.9 El error de elegir el modelo por un solo valor *p* 129
9.10 El error que cometen los investigadores en la redacción de artículos 129
RESUMEN Y CONCLUSIONES 130
EJERCICIOS RESUELTOS 130
Para tu tesis o artículo 131
Referencias 132
Capítulo 10 Series de tiempo y modelo ARIMA paso a paso 133
Lo que vas a lograr 133
10.1 En una serie, el orden también es información 134
10.2 Estacionariedad: que el pasado sirva de guía 134
10.3 Dos pruebas con hipótesis opuestas 134
10.4 Las tres letras del ARIMA 134
10.5 El correlograma: la huella de cada modelo 135
10.6 Manos a la obra 135
10.7 El error de juzgar un ARIMA por su ajuste 145
10.8 El error que cometen los investigadores en la redacción de artículos 145
RESUMEN Y CONCLUSIONES 146
EJERCICIOS RESUELTOS 146
Para tu tesis o artículo 147
Referencias 148
Capítulo 11 Pronósticos con modelos VAR y cointegración 149
Lo que vas a lograr 149
11.1 Un modelo donde todas las series se explican entre sí 150
11.2 Tres preguntas que responde un VAR 150
11.3 Cointegración: dos caminantes atados por una cuerda 150
11.4 Manos a la obra 151
11.5 El error que arruina más análisis de series: estimar en niveles sin probar la cointegración ... 159
11.6 El error que cometen los investigadores en la redacción de artículos 160
RESUMEN Y CONCLUSIONES 160
EJERCICIOS RESUELTOS 160
Para tu tesis o artículo 162
Referencias 162
Parte Cuatro. Evaluación de impacto 163

Capítulo 12 Evaluación de impacto I: diferencias en diferencias 164
Lo que vas a lograr 164
12.1 Antes y después: la resta que engaña 165
12.2 Cuatro promedios y una resta doble 165
12.3 El supuesto del que depende todo 165
12.4 De dos periodos a muchos: efectos fijos y estudio de eventos 166
12.5 Cuando la política llega en años distintos 166
12.6 Manos a la obra 166
12.7 El error que convierte una tendencia en un efecto 174
12.8 El error que cometen los investigadores en la redacción de artículos 175
RESUMEN Y CONCLUSIONES 175
EJERCICIOS RESUELTOS 175
Para tu tesis o artículo 176
Referencias 177
Capítulo 13 Evaluación de impacto II: variables instrumentales 178
Lo que vas a lograr 178
13.1 La división del capítulo 1, con nombre propio 179
13.2 Tres condiciones para un buen instrumento 179
13.3 La forma reducida, la primera etapa y su cociente 179
13.4 ¿De quién es el efecto? 179
13.5 Cuando el instrumento casi no mueve nada 180
13.6 Manos a la obra 180
13.7 El error que se esconde detrás de un instrumento ingenioso 186
13.8 El error que cometen los investigadores en la redacción de artículos 186
RESUMEN Y CONCLUSIONES 186
EJERCICIOS RESUELTOS 187
Para tu tesis o artículo 188
Referencias 188
Capítulo 14 Evaluación de impacto III: regresión discontinua y propensity score matching 190
Lo que vas a lograr 190
14.1 Una frontera que reparte al azar 191
14.2 Lo que no debe saltar 191
14.3 Nítida o difusa 191
14.4 Cuán cerca mirar 191
14.5 Manos a la obra: el corte del SISBEN 192
14.6 Cuando no hay corte: emparejar a los parecidos 198
14.7 Manos a la obra: el emparejamiento frente a un experimento 198
14.8 ¿Cuál método elijo? 201
14.9 El error que convierte un emparejamiento en un experimento 202
14.10 El error que cometen los investigadores en la redacción de artículos 202
RESUMEN Y CONCLUSIONES 202
EJERCICIOS RESUELTOS 203
Para tu tesis o artículo 203
Referencias 204
Parte Cinco. Del resultado a la tesis y al artículo .. 205
Capítulo 15 Cómo redactar los resultados de tu tesis o artículo científico 206
Lo que vas a lograr 206
15.1 El capítulo de resultados responde preguntas 207
15.2 Una tabla que se entiende sola 207
15.3 Cómo se informa cada número 207
15.4 Tu análisis debe poder correr otra vez 208
15.5 Manos a la obra 208
15.6 El error que comete casi todo tesista: el número que cambió en el camino 214
15.7 El error que cometen los investigadores en la redacción de artículos 215
RESUMEN Y CONCLUSIONES 215

EJERCICIOS RESUELTOS.....216
Para tu tesis o artículo 217
Referencias 217
APÉNDICE A. EQUIVALENCIAS R · STATA · EVIEWS · SPSS.....218
APÉNDICE B. BASES DE DATOS DEL LIBRO223

Capítulo 1

¿Por qué tu regresión miente? Correlación y causalidad

El 13 de octubre de 2006, el Comité Noruego del Nobel anunció el Premio de la Paz. No fue para un presidente ni para un tratado. Fue para Muhammad Yunus y el Grameen Bank, un banco que prestaba pequeñas sumas a personas pobres de Bangladesh. El comité premió sus esfuerzos por crear desarrollo económico y social «desde abajo» y escribió que el microcrédito era uno de los medios para salir de la pobreza (Norwegian Nobel Committee, 2006).

¿De dónde venía tanta confianza? En parte, de los datos. Un estudio muy citado, con hogares de Bangladesh, estimó que el consumo anual del hogar subía 18 taka por cada 100 taka adicionales prestados a las mujeres. Cuando el préstamo iba a los hombres, subía 11 (Pitt & Khandker, 1998). Era un efecto grande y tenía una explicación hermosa.

Años después, dos economistas rehicieron ese estudio con los mismos datos. Encontraron que el resultado dependía de supuestos frágiles y de unos pocos hogares con consumos extremos. Con esos datos, concluyeron, no se podía afirmar ni negar que el crédito causara el aumento (Roodman & Morduch, 2014).

Entonces alguien hizo lo que el estudio original no podía hacer. Entre 2003 y 2012, seis equipos sortearon quién o dónde recibía el microcrédito en Bosnia y Herzegovina, Etiopía, India, México, Mongolia y Marruecos. En México, Compartamos Banco ofreció crédito puerta a puerta en 238 conglomerados de Sonora, y el sorteo decidió cuáles recibían la promoción (Angelucci et al., 2015). La conclusión conjunta cabe en una línea: efectos «modestamente positivos, pero no transformadores» (Banerjee et al., 2015, p. 1).

El microcrédito no resultó inútil. Resultó **mucho menos** milagroso de lo que decían las correlaciones. Este capítulo te enseña por qué pasa eso y cómo detectarlo antes de escribir tu tesis.

Lo que vas a lograr

Al terminar este capítulo podrás:

- Explicar con un ejemplo por qué una correlación, aunque sea fuerte y significativa, no te dice cuánto **causa** una variable sobre otra.
- Escribir la pregunta contrafactual de tu investigación: ¿qué habría pasado con las mismas personas sin el tratamiento?
- Descomponer una diferencia de promedios en efecto verdadero más sesgo de selección.
- Simular datos en R con un efecto conocido y comprobar cuál método lo recupera.
- Distinguir un resultado sesgado de uno impreciso, que son dos problemas distintos.

No necesitas saber R todavía. Copiarás bloques cortos y verás qué devuelven. En el capítulo 2 aprenderás a instalar R y RStudio con calma; si aún no los tienes, puedes leer este capítulo sin computadora y volver después.

1.1 Las que pidieron crédito no eran iguales a las que no pidieron

Imagina que trabajas en una financiera y te encargan medir si el crédito mejora los negocios de tus clientas. Tienes una encuesta: 6000 emprendedoras, cuánto gana cada negocio al mes y si tomó un crédito. Comparas promedios. Las que tomaron crédito ganan 191 soles más al mes.

¿Ese es el efecto del crédito? Piensa en quién pide un préstamo. Pide más la que tiene un plan, la que ya vende bien, la que sale a buscar clientes. Esa misma iniciativa también hace que su negocio gane más, con crédito o sin él.

Por eso, la diferencia de 191 soles mezcla dos cosas. Una parte es lo que el crédito agrega. La otra es lo que esas mujeres ya traían. La encuesta no mide la iniciativa, así que no puedes separar una parte de la otra solo mirando la diferencia.

Este problema tiene nombre: **sesgo de selección**. Aparece cuando las personas eligen el tratamiento por razones que también mueven el resultado. Es el mismo problema que tuvo el estudio de Bangladesh: los hogares que tomaron crédito no eran un grupo cualquiera.

1.2 Dos negocios que nunca verás a la vez

Para saber cuánto agrega el crédito a una clienta, tendrías que ver dos versiones de su negocio en el mismo mes: una con crédito y otra sin él. La diferencia entre ambas sería el efecto para ella.

Nunca ves las dos. Ves la versión que ocurrió. La otra, la que no ocurrió, se llama **contrafactual**. Toda pregunta causal es, en el fondo, una pregunta sobre un contrafactual: ¿cuánto habría ganado esta misma mujer si no hubiera tomado el crédito?

En la práctica, como el contrafactual no se observa, se aproxima con otro grupo. Comparas a las que tomaron crédito con las que no. Esa comparación solo funciona si ambos grupos se parecen en todo lo demás. Si las que no tomaron crédito tienen menos iniciativa, son un mal reemplazo del contrafactual.

¿Qué tan malo? Eso se puede calcular.

1.3 Tres piezas dentro de una diferencia

Cunningham (2021) escribe la diferencia simple de promedios como la suma de tres piezas. La primera es el efecto promedio del tratamiento, lo que buscas. La segunda es el sesgo de selección: cuánto habrían diferido los dos grupos aunque nadie hubiera recibido el tratamiento. La tercera aparece cuando el efecto no es igual para todos.

En palabras:

diferencia de promedios = efecto verdadero + sesgo de selección + sesgo por efectos distintos

Si el efecto del crédito fuera igual para todas, la tercera pieza vale cero. Queda la segunda. En tu ejemplo, el sesgo de selección es lo que habrían ganado de más las que pidieron crédito **sin haberlo pedido**, solo por su iniciativa.

La ecuación no se puede usar con datos reales, porque el sesgo depende del contrafactual que no ves. Pero con datos simulados sí: tú decides el efecto verdadero y luego mides cuánto se equivoca cada método. Eso harás en la sección *Manos a la obra*.

1.4 Un dibujo de tres flechas

Antes de estimar cualquier cosa, dibuja en un papel qué causa qué. Para tu problema bastan tres flechas:

- La iniciativa lleva a pedir crédito.
- La iniciativa sube la utilidad del negocio.
- El crédito sube la utilidad del negocio, y esa es la flecha que quieres medir.

La iniciativa abre un camino alternativo entre crédito y utilidad, que pasa por fuera de la flecha que te interesa. Huntington-Klein (2021) llama «puertas traseras» a esos caminos: mientras uno quede abierto, la correlación entre crédito y utilidad mezcla el efecto con el sesgo.

En consecuencia, hay dos formas de cerrar esa puerta. La primera es medir la iniciativa y controlarla en la regresión. La segunda es cortar la flecha que va de la iniciativa al crédito, de modo que el crédito ya no dependa de quién es cada mujer. El dibujo te dice cuál de las dos tienes a mano.

1.5 Cuando una moneda decide

Un sorteo corta la flecha que va de la iniciativa al tratamiento. Si una moneda decide qué barrios reciben la promoción, la promoción no puede depender de la iniciativa de nadie. En promedio, los barrios sorteados y los no sorteados tienen la misma mezcla de mujeres con mucha y con poca iniciativa.

Por esa razón, los autores de los seis estudios dicen que la asignación aleatoria minimiza los sesgos de selección que afectan a los estudios observacionales (Banerjee et al., 2015). No hace falta medir la iniciativa. El sorteo la reparte por igual.

Fíjate en un detalle de Compartamos. El sorteo no decidió quién tomaba el crédito, decidió quién recibía la **promoción**. En México tomó crédito el 19 % de las personas del grupo con promoción (Banerjee et al., 2015, Tabla 2). La mayoría no lo tomó, aunque se lo ofrecieron. Ese detalle cambia cómo se lee el resultado, y lo verás en R.

1.6 Tu tesis sin sorteo

Tu tesis casi seguro usará datos de una encuesta, no de un sorteo. ¿Estás perdido? No. Estás obligado a ser honesto sobre lo que tu diseño permite afirmar.

Este libro te enseña, capítulo a capítulo, las herramientas para acercarte a la causalidad sin sorteo. Primero, controlar variables (capítulos 4 a 7) y aprovechar datos de panel (capítulo 9). Después, comparar antes y después entre grupos (capítulo 12), usar variables instrumentales (capítulo 13) y aprovechar umbrales o parejas comparables (capítulo 14). Cada una tiene un supuesto, y tu trabajo es defenderlo.

Cuando no puedas defenderlo, escribe «asociación» y no «efecto». Es menos vistoso. También es correcto.

1.7 Manos a la obra

En esta sección juegas con ventaja: **tú fabricas los datos, así que conoces la respuesta correcta antes de estimar**. Simularás 6000 emprendedoras repartidas en 240 conglomerados. Tú decidirás que el crédito sube la utilidad del negocio exactamente en 60 soles al mes. Después pondrás a prueba cuatro métodos en R y verás cuál se acerca a 60.

Aquí los datos son **simulados**: ninguna cifra de esta sección describe negocios reales ni clientas de Compartamos. El diseño imita el de Sonora (conglomerados sorteados, promoción puerta a puerta, adopción parcial) solo para que reconozcas el problema cuando lo veas en un artículo.

Paso 0. Prepara R para que veas lo mismo que el libro

Abre RStudio, crea un script nuevo con *File › New File › R Script* y escribe estas líneas antes que nada. Ejecuta cada línea con Ctrl + Enter (Cmd + Enter en Mac).

```
Sys.setenv(LANGUAGE = "es")
options(OutDec = ",", width = 80, scipen = 6, digits = 4)
set.seed(2026)
dir.create("figuras", showWarnings = FALSE)
```

`Sys.setenv(LANGUAGE = "es")` pide los mensajes de R en español. `options(OutDec = ",")` hace que R imprima los decimales con coma, como se escriben en español; `digits = 4` limita las cifras que muestra y `scipen = 6` evita la notación científica. `set.seed(2026)` fija la semilla del azar: si usas la misma semilla y corres los bloques en el mismo orden, obtendrás exactamente las cifras de este libro (R Core Team, 2026). La última línea crea la carpeta donde se guardarán los gráficos.

Si prefieres no copiar, descarga el archivo `cap01.R` del repositorio del libro: trae estas líneas y todos los bloques en orden.

Paso 1. Crea 6000 emprendedoras con una iniciativa oculta

```
# Bloque 1: 6000 emprendedoras simuladas en 240 conglomerados, con una iniciativa que nadie mide.
n_cong <- 240
por_cong <- 25
datos <- data.frame(conglomerado = rep(1:n_cong, each = por_cong))
datos$iniciativa <- rnorm(nrow(datos), mean = 0, sd = 1)
# El sorteo: la mitad de los conglomerados recibe la promoción puerta a puerta.
sorteados <- sample(1:n_cong, n_cong / 2)
datos$promocion <- as.integer(datos$conglomerado %in% sorteados)
head(datos)
nrow(datos)
```

```
  conglomerado iniciativa promocion
1            1    0,52059         0
2            1   -1,07969         0
3            1    0,13924         0
4            1   -0,08475         0
5            1   -0,66664         0
6            1   -2,51609         0
[1] 6000
```

El signo `<-` guarda un resultado con un nombre. `data.frame()` crea una base de datos, y `rep(1:n_cong, each = por_cong)` repite cada número de conglomerado 25 veces: 240 × 25 = 6000 filas. El signo `$` señala una columna dentro de la base: `datos$iniciativa` es la columna iniciativa.

`rnorm()` genera números al azar con forma de campana, centrados en 0 y con desviación estándar 1. Así nace la iniciativa de cada mujer, una variable que ninguna encuesta registra. `sample(1:n_cong, n_cong / 2)` elige al azar 120 de los 240 conglomerados, y `%in%` marca con `TRUE` a las mujeres que viven en uno de ellos. `as.integer()` convierte `TRUE` y `FALSE` en 1 y 0.

`head()` muestra las seis primeras filas y `nrow()` cuenta las filas. Las seis primeras mujeres viven en el conglomerado 1, que no salió sorteado.

Paso 2. Decide quién toma el crédito y cuánto gana su negocio

```
# Bloque 2: quién toma el crédito y cuánto gana su negocio.
# Toman más crédito las que tienen más iniciativa y las que recibieron la promoción.
indice <- -2.1 + 0.6 * datos$iniciativa + 1.3 * datos$promocion
datos$credito <- rbinom(nrow(datos), size = 1, prob = plogis(indice))
# Efecto verdadero del crédito: 60 soles más de utilidad al mes.
datos$utilidad <- 900 + 60 * datos$credito + 250 * datos$iniciativa +
  rnorm(nrow(datos), mean = 0, sd = 400)
summary(datos[, c("credito", "utilidad")])
tapply(datos$credito, datos$promocion, mean)
```

```
    credito         utilidad
 Min.   :0,000   Min.   :-712
 1st Qu.:0,000   1st Qu.: 595
 Median :0,000   Median : 912
 Mean   :0,228   Mean   : 915
 3rd Qu.:0,000   3rd Qu.:1240
 Max.   :1,000   Max.   :3007
     0      1
0,1200 0,3367
```

Lee las dos fórmulas con calma, porque son el corazón del capítulo. La primera dice que la probabilidad de tomar crédito sube con la iniciativa (0.6) y con la promoción (1.3). `plogis()` convierte ese índice en una probabilidad entre 0 y 1, y `rbinom()` lanza una moneda cargada con esa probabilidad para cada mujer. Fíjate en que dentro del código los decimales se escriben con punto: la coma solo cambia lo que R **imprime**.

La segunda fórmula fabrica la utilidad mensual. Parte de 900 soles, suma **60 soles si tomó crédito**, suma 250 soles por cada desviación estándar de iniciativa y agrega un ruido con desviación de 400 soles. El 60 es el efecto verdadero. Desde ahora, cada método se juzga por cuánto se acerca a 60.

`summary()` resume las columnas: tomó crédito el 22,8 % y la utilidad media es de 915 soles. Algunas utilidades son negativas, como pasa con los negocios que pierden en un mes malo. `tapply()` calcula la media de la primera variable para cada valor de la segunda: tomó crédito el 12,0 % de las mujeres sin promoción y el 33,7 % de las que la recibieron.

Paso 3. Primer intento: comparar promedios

```
# Bloque 3: primer intento, comparar a las que tomaron crédito con las que no.
tapply(datos$utilidad, datos$credito, mean)
diff(tapply(datos$utilidad, datos$credito, mean))
```

```
     0      1
 871,4 1062,7
    1
191,3
```

Las que no tomaron crédito ganan 871,4 soles al mes; las que lo tomaron, 1062,7. `diff()` resta el primer valor del segundo: 191,3 soles de diferencia.

El efecto verdadero es 60. La comparación simple lo **triplica**. No es mala suerte, es el sesgo de selección: las que tomaron crédito traían más iniciativa, y la iniciativa también sube la utilidad.

Paso 4. Segundo intento: la regresión

Muchos tesistas creen que la regresión corrige este problema porque es «más avanzada». Compruébalo.

```
# Bloque 4: segundo intento, la regresión.
m_ingenua <- lm(utilidad ~ credito, data = datos)
summary(m_ingenua)
```

```
Call:
lm(formula = utilidad ~ credito, data = datos)

Residuals:
    Min      1Q  Median      3Q     Max
-1583,7  -314,0    -2,9   323,8  1944,1

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)   871,40       6,91   126,0   <2e-16 ***
credito       191,28      14,47    13,2   <2e-16 ***
---
Signif. codes:  0 ‘***’ 0,001 ‘**’ 0,01 ‘*’ 0,05 ‘.’ 0,1 ‘ ’ 1

Residual standard error: 471 on 5998 degrees of freedom
Multiple R-squared:  0,0283,    Adjusted R-squared:  0,0281
F-statistic:  175 on 1 and 5998 DF,  p-value: <2e-16
```

`lm()` estima un modelo lineal por mínimos cuadrados. La fórmula `utilidad ~ credito` se lee «utilidad explicada por crédito», y `data = datos` dice en qué base buscar las columnas. `summary()` muestra el resultado completo; en el capítulo 6 aprenderás a leer cada línea.

Por ahora mira la fila `credito`. El coeficiente es 191,28: el mismo número del paso 3. Con una sola variable binaria, la regresión es exactamente la diferencia de promedios. El estadístico $t = 13{,}2$ y el valor $p < 2e\text{-}16$ dicen que la diferencia es muy significativa.

Y es falsa como estimación del efecto. Esa es la lección más importante del capítulo: **la significancia no corrige el sesgo**. Un número puede estar estimado con mucha precisión y apuntar al lugar equivocado.

Paso 5. Tercer intento: si pudieras medir la iniciativa

En tu simulación tienes algo que ninguna encuesta tiene: la columna `iniciativa`. Agrégala a la regresión.

```
# Bloque 5: tercer intento, si pudieras medir la iniciativa.
m_oraculo <- lm(utilidad ~ credito + iniciativa, data = datos)
coef(summary(m_oraculo))
```

```
            Estimate Std. Error t value  Pr(>|t|)
(Intercept)   903,52      5,909 152,911 0,0000000
credito        53,96     12,613   4,279 0,0000191
iniciativa    255,16      5,294  48,201 0,0000000
```

`coef(summary())` muestra solo la tabla de coeficientes. Ahora el crédito vale 53,96 soles, con un error estándar de 12,613. Si sumas y restas dos errores estándar, obtienes un intervalo de 28,7 a 79,2 soles (cuenta hecha a mano), que contiene el 60 verdadero.

Controlar la iniciativa cerró la puerta trasera. El problema es que esta regresión es un **oráculo**: en la vida real nadie tiene una columna con la iniciativa exacta de cada persona. Por eso la llamamos `m_oraculo`.

Paso 6. Cuarto intento: el sorteo

No puedes medir la iniciativa. Pero tienes algo mejor que la encuesta del paso 3: un sorteo. Compara a los conglomerados con promoción y sin ella.

```
# Bloque 6: cuarto intento, usar el sorteo.
toma <- tapply(datos$credito, datos$promocion, mean)
util <- tapply(datos$utilidad, datos$promocion, mean)
efecto_sorteo <- util["1"] - util["0"]
efecto_sorteo
efecto_sorteo / (toma["1"] - toma["0"])
```

```
    1
8,946
    1
41,29
```

La primera cifra sorprende: la utilidad media sube solo 8,946 soles en los conglomerados con promoción. ¿El crédito casi no sirve? No. Recuerda el paso 2: la promoción solo subió la adopción del 12,0 % al 33,7 %. La mayoría de las mujeres con promoción no tomó crédito, así que el efecto se diluye entre todas.

Esa primera cifra es el efecto de **ofrecer** el crédito, no de tomarlo. Para pasar de uno a otro se divide entre el aumento de la adopción: 8,946 / (0,3367 − 0,1200). R hace la cuenta en la última línea y da 41,29 soles por mujer que tomó el crédito gracias a la promoción. En el capítulo 13 verás que esta división es un estimador de variables instrumentales.

¿41 en lugar de 60? El sorteo no tiene sesgo de selección, pero eso no lo hace exacto. El paso 8 te muestra cuánto puede variar.

Paso 7. Busca la iniciativa en cada grupo

¿Dónde quedó la iniciativa en cada comparación? Calcúlala por grupos.

```
# Bloque 7: la prueba del delito, ¿dónde está la iniciativa?
tapply(datos$iniciativa, datos$credito, mean)
tapply(datos$iniciativa, datos$promocion, mean)
```

```
      0       1
-0,1259  0,4122
        0         1
 0,004489 -0,010533
```

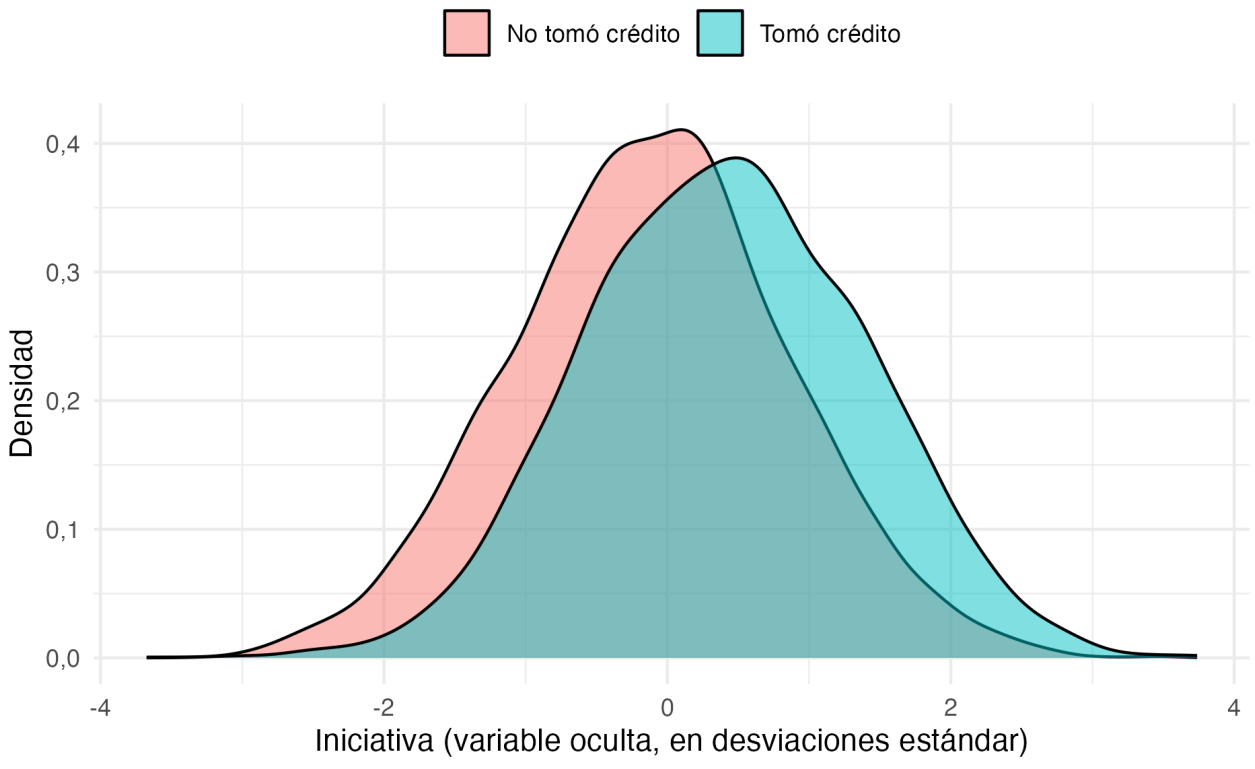
Entre las que tomaron crédito, la iniciativa media es 0,4122 desviaciones estándar; entre las que no, −0,1259. Es más de media desviación de diferencia. Ahí está el sesgo del paso 3.

Ahora mira la comparación por sorteo: 0,004489 contra −0,010533. Prácticamente cero. El sorteo repartió la iniciativa por igual entre los dos grupos, sin que nadie la midiera.

Además, con estas cifras puedes cerrar la cuenta de Cunningham (2021). El sesgo de selección esperado es 250 soles por desviación de iniciativa multiplicado por la diferencia de iniciativa: 250 × (0,4122 + 0,1259) = 134,5 soles (cuenta hecha a mano). Efecto verdadero más sesgo: 60 + 134,5 = 194,5 soles, muy cerca de los 191,3 del paso 3. Los 3 soles restantes vienen del ruido.

El gráfico muestra lo mismo de un vistazo:

```
# Bloque 8: un gráfico que muestra el sesgo.
library(ggplot2)
datos$grupo <- factor(datos$credito, labels = c("No tomó crédito", "Tomó crédito"))
g <- ggplot(datos, aes(x = iniciativa, fill = grupo)) +
  geom_density(alpha = 0.5) +
  labs(x = "Iniciativa (variable oculta, en desviaciones estándar)",
       y = "Densidad", fill = NULL) +
  scale_x_continuous(labels = scales::label_number(decimal.mark = ",")) +
  scale_y_continuous(labels = scales::label_number(decimal.mark = ",")) +
  theme_minimal(base_size = 11) + theme(legend.position = "top")
ggsave("figuras/cap01_b08_g01.png", g,
       width = 16, height = 10, units = "cm", dpi = 300)
```



`library(ggplot2)` carga el paquete de gráficos más usado en R. `ggplot()` recibe la base y, dentro de `aes()`, qué variable va en el eje horizontal y cuál define el color. `geom_density()` dibuja la distribución suavizada de cada grupo; `alpha = 0.5` la vuelve semitransparente. `labs()` pone los rótulos y `label_number(decimal.mark = ",")` escribe los números del eje con coma. `ggsave()` guarda el gráfico a 300 puntos por pulgada, la resolución que piden las revistas.

La curva de las que tomaron crédito está corrida a la derecha. No es un efecto del crédito: la iniciativa se fijó en el paso 1, **antes** de que nadie pidiera nada.

Paso 8. Sesgado o impreciso: repite el mundo 500 veces

Un solo sorteo dio 41,29. ¿Fue un mal sorteo o el método falla? Con datos reales no lo sabrías. Con una simulación puedes repetir el mundo entero muchas veces.

Primero, mira la precisión del sorteo en tu base:

```
# Bloque 9: ¿cuánta precisión tiene el sorteo?
m_sorteo <- lm(utilidad ~ promocion, data = datos)
coef(summary(m_sorteo))
```

```
            Estimate Std. Error  t value Pr(>|t|)
(Intercept)  910,600      8,714 104,4946   0,0000
promocion      8,946     12,324   0,7259   0,4679
```

El efecto de la promoción, 8,946 soles, tiene un error estándar de 12,324. El valor p = ,468: con una sola muestra, no podrías distinguirlo de cero. Ahora repite el experimento completo 500 veces:

```
# Bloque 10: repetir el mundo 500 veces.
un_mundo <- function() {
  d <- data.frame(conglomerado = rep(1:240, each = 25))
  d$iniciativa <- rnorm(nrow(d))
  d$promocion <- as.integer(d$conglomerado %in% sample(1:240, 120))
  d$credito <- rbinom(nrow(d), 1,
                      plogis(-2.1 + 0.6 * d$iniciativa + 1.3 * d$promocion))
  d$utilidad <- 900 + 60 * d$credito + 250 * d$iniciativa + rnorm(nrow(d), 0, 400)
  ingenua <- coef(lm(utilidad ~ credito, data = d))["credito"]
  toma <- tapply(d$credito, d$promocion, mean)
  util <- tapply(d$utilidad, d$promocion, mean)
  sorteo <- (util["1"] - util["0"]) / (toma["1"] - toma["0"])
  c(ingenua = unname(ingenua), sorteo = unname(sorteo))
}
mundos <- t(replicate(500, un_mundo()))
colMeans(mundos)
apply(mundos, 2, sd)
```

```
ingenua  sorteo
 192,85   63,02
ingenua  sorteo
  14,41   62,88
```

`function()` crea tu propia función: `un_mundo()` fabrica una base nueva con las mismas reglas y devuelve las dos estimaciones. `replicate(500, …)` la corre 500 veces y `t()` acomoda el resultado en dos columnas. `colMeans()` promedia cada columna y `apply(mundos, 2, sd)` calcula su desviación estándar.

Lee las dos líneas juntas. En promedio, la regresión ingenua da 192,85 soles, lejísimos de 60, y casi no varía (desviación de 14,41): es **precisa y sesgada**. El sorteo da en promedio 63,02, muy cerca de 60, pero de un mundo a otro varía mucho (desviación de 62,88): es **insesgado e impreciso**.

En suma, esa es la diferencia que debes llevarte. El sesgo no se arregla con más datos; la imprecisión, sí. Los seis estudios de microcrédito encontraron tasas de adopción bajas (19 % en México), y una adopción baja hace que un sorteo, por honesto que sea, mida con mucho ruido (Banerjee et al., 2015).

1.8 El error del tesista apurado: confundir significativo con causal

Abre cualquier repositorio de tesis y buscarás poco para encontrar esta frase: «Se concluye que el acceso al crédito influye significativamente en la utilidad de las mypes (p < ,05)». La frase mezcla dos afirmaciones que el capítulo acaba de separar.

«Significativamente» habla de la precisión: el coeficiente difícilmente es cero. «Influye» habla de causalidad: si le das crédito a una mype, su utilidad sube. Tu paso 4 tenía un p < 2e-16 y triplicaba el efecto. La significancia era real y la influencia estaba inflada.

¿Tu diseño es una encuesta sin sorteo y sin una estrategia de los capítulos 12 a 14? Entonces escribe lo que tus datos sostienen: «Las mypes con crédito reportan una utilidad mayor en 191 soles al mes (p < ,001). Esta diferencia no aísla el efecto del crédito, porque las empresas que lo solicitan pueden diferir en características no observadas, como la iniciativa de sus dueñas».

Tampoco sirve llenar la regresión de controles

Sin embargo, el remedio equivocado es meter en la regresión todas las variables disponibles, con la esperanza de que alguna «capture» la iniciativa. El ejercicio 3 te muestra que una variable parecida a la iniciativa corrige poco. Y en el capítulo 5 verás que controlar por una variable que el propio tratamiento modifica puede **crear** un sesgo que no existía. Un control solo ayuda si sabes qué puerta trasera cierra.

1.9 El error que cometen los investigadores en la redacción de artículos

En la discusión de muchos artículos se lee algo así: «Nuestros resultados demuestran que el microcrédito mejora el bienestar de los hogares». Tiene tres fallas.

1. **El verbo**. «Demuestran» y «mejora» son lenguaje causal; si el diseño es observacional, el verbo debe ser «se asocia con».

2. **La cifra ausente**. No dice cuánto mejora ni con qué precisión. Un lector no puede comparar ese hallazgo con los seis estudios aleatorizados.

3. **La amenaza callada**. No menciona la selección: quiénes piden crédito y por qué.

Cómo se escribe bien: «En la muestra simulada (n = 6000), las emprendedoras que tomaron crédito obtuvieron una utilidad mensual mayor en 191,3 soles (EE = 14,47; t(5998) = 13,2; p < ,001). Sin embargo, al usar la asignación aleatoria de la promoción, el efecto por cada mujer que tomó crédito fue de 41,3 soles. El efecto de la promoción no se distinguió de cero (b = 8,9; EE = 12,32; p = ,468). La diferencia entre ambas estimaciones es coherente con un sesgo de selección positivo».

RESUMEN Y CONCLUSIONES

1. Una correlación mide si dos variables se mueven juntas. No dice cuánto cambia una cuando tú cambias la otra.

2. Toda pregunta causal compara lo que ocurrió con un contrafactual que no se observa. El grupo de comparación es un reemplazo del contrafactual, y solo sirve si se parece al grupo tratado en todo lo demás.

3. La diferencia simple de promedios es igual al efecto verdadero más el sesgo de selección (más un sesgo adicional si el efecto varía entre personas). En tu simulación, 191,3 ≈ 60 + 134,5.

4. La regresión `lm(utilidad ~ credito)` da exactamente la misma diferencia y, además, la declara muy significativa. La significancia no corrige el sesgo.

5. Un sorteo reparte por igual lo que no mides. Es insesgado, pero puede ser impreciso: con adopción baja, como en los estudios de microcrédito, necesita muestras grandes.

En el capítulo 2 instalarás R y RStudio con calma, conocerás sus paneles y prepararás tu primera base real: la ENAHO 2024 del INEI y los indicadores del Banco Mundial, listos para analizar.

EJERCICIOS RESUELTOS

Ejercicio 1. Educación e ingreso en la ENAHO 2024 (datos reales). Descarga la base de trabajadores del libro, comprueba el archivo y calcula la correlación entre los años de educación y el logaritmo del ingreso mensual. Luego estima la regresión. ¿Cuánto más gana, en promedio, quien tiene un año más de educación? ¿Es ese el efecto de estudiar un año más?

```
# ===== EJERCICIOS RESUELTOS =====
# Ejercicio 1: educación e ingreso en la ENAHO 2024 (datos reales).
url <- "https://zenodo.org/records/23249894/files/enaho2024_trabajadores.csv?download=1"
archivo <- tempfile(fileext = ".csv")
download.file(url, archivo, mode = "wb", quiet = TRUE)
unname(tools::md5sum(archivo))
trab <- read.csv(archivo)
trab <- subset(trab, ingreso_mes > 0)
cor(trab$educ, log(trab$ingreso_mes))
coef(lm(log(ingreso_mes) ~ educ, data = trab))
```

```
[1] "96cd2764db40ff2b9a9977fe8108a6cb"
[1] 0,4136
(Intercept)        educ
     5,7956      0,1013
```

Solución. `download.file()` baja el archivo a una ruta temporal creada por `tempfile()`. `tools::md5sum()` calcula su huella digital. Si ves 96cd2764db40ff2b9a9977fe8108a6cb, tienes exactamente la misma base que el libro, un extracto de la ENAHO 2024 publicado en Zenodo (Flores Sotelo, 2026; Instituto Nacional de Estadística e Informática [INEI], 2024). `read.csv()` la lee; son ocupados de 18 a 65 años con ingreso positivo, y `subset()` lo asegura.

La correlación es 0,4136. La pendiente de 0,1013 en una regresión con logaritmo se lee, aproximadamente, como un ingreso 10 % mayor por cada año adicional de educación. ¿Es el efecto de estudiar? No necesariamente. Quienes estudian más pueden diferir en habilidad, en la educación de sus padres o en la región donde viven. Es la misma estructura del microcrédito: una variable oculta que mueve las dos cosas. En los capítulos 5 y 13 volverás a esta regresión.

Ejercicio 2. Una promoción que convence a más gente. Repite la simulación con una promoción mucho más eficaz (coeficiente 3 en lugar de 1.3). ¿Qué pasa con la adopción y con la estimación del sorteo?

```
# Ejercicio 2: una promoción que convence a más gente.
indice2 <- -2.1 + 0.6 * datos$iniciativa + 3 * datos$promocion
datos$credito2 <- rbinom(nrow(datos), 1, plogis(indice2))
datos$utilidad2 <- 900 + 60 * datos$credito2 + 250 * datos$iniciativa +
  rnorm(nrow(datos), 0, 400)
toma2 <- tapply(datos$credito2, datos$promocion, mean)
toma2
util2 <- tapply(datos$utilidad2, datos$promocion, mean)
(util2["1"] - util2["0"]) / (toma2["1"] - toma2["0"])
```

```
     0      1
0,1213 0,6897
    1
60,49
```

Solución. Con la nueva promoción, la adopción pasa del 12,1 % al 69,0 %: una diferencia de casi 57 puntos, frente a los 22 puntos del paso 6. La estimación del sorteo es 60,49 soles, casi exactamente el efecto verdadero. Cuando el sorteo mueve a más gente, la señal crece frente al ruido. Por eso, al diseñar o leer un experimento, la primera cifra que se mira es cuánto cambió la adopción.

Ejercicio 3. Controlar por una variable que se parece a la iniciativa. En la práctica no tienes la iniciativa, pero sí los años de experiencia del negocio, que se le parecen. Simula esa variable y agrégala a la regresión. ¿Se corrige el sesgo?

```
# Ejercicio 3: controlar por una variable que se parece a la iniciativa.
# Los años de experiencia del negocio siguen a la iniciativa, pero con mucho ruido.
datos$experiencia <- pmax(0, round(6 + 1.5 * datos$iniciativa + rnorm(nrow(datos), 0, 3)))
cor(datos$experiencia, datos$iniciativa)
coef(lm(utilidad ~ credito + experiencia, data = datos))["credito"]
```

```
[1] 0,4404
credito
    166
```

Solución. `pmax(0, …)` impide que la experiencia sea negativa y `round()` la deja en años enteros. La correlación entre experiencia e iniciativa es 0,4404: se parecen, pero no son lo mismo. Al controlar por experiencia, el coeficiente del crédito baja de 191,3 a 166 soles. Sigue a casi el triple del efecto verdadero. Un control imperfecto cierra la puerta trasera solo un poco; la parte de la iniciativa que la experiencia no recoge sigue sesgando la estimación.

Para tu tesis o artículo

1. Escribe en una oración la pregunta contrafactual de tu investigación: «¿Cuánto habría sido *Y* para las mismas unidades si no hubieran recibido *X*?».

2. Dibuja el diagrama de flechas de tu problema y nombra al menos dos variables ocultas que podrían mover a la vez tu *X* y tu *Y*.

3. Revisa tu objetivo general y tus conclusiones: si usan «influye», «efecto» o «impacto», comprueba que tu diseño lo sostenga. Si no, cambia el verbo a «se asocia».

Lo que te va a preguntar el jurado: «¿Cómo descarta que las unidades con *X* ya fueran distintas antes del tratamiento?», «¿Qué variable omitida podría explicar su resultado?» y «¿Su significancia demuestra causalidad?». La respuesta a la última es no, y conviene que lo digas tú antes que el jurado.

Si es un artículo: los revisores esperan una sección de amenazas a la identificación. Nombra la fuente de selección, di en qué dirección sesgaría tu estimación y, si puedes, compárala con evidencia experimental del mismo tema, como hacen Banerjee et al. (2015) con el microcrédito.

Referencias

Angelucci, M., Karlan, D., & Zinman, J. (2015). Microcredit impacts: Evidence from a randomized microcredit program placement experiment by Compartamos Banco. *American Economic Journal: Applied Economics*, *7*(1), 151–182. https://doi.org/10.1257/app.20130537

Banerjee, A., Karlan, D., & Zinman, J. (2015). Six randomized evaluations of microcredit: Introduction and further steps. *American Economic Journal: Applied Economics*, *7*(1), 1–21. https://doi.org/10.1257/app.20140287

Cunningham, S. (2021). *Causal inference: The mixtape*. Yale University Press. https://doi.org/10.12987/9780300255881

Flores Sotelo, W. S. (2026). *Econometría aplicada con EViews para la investigación: Datos del libro (extractos ENAHO 2024)* (Versión 1.1.0) [Conjunto de datos]. Zenodo. https://doi.org/10.5281/zenodo.23249774

Huntington-Klein, N. (2021). *The effect: An introduction to research design and causality*. Chapman & Hall/CRC. https://doi.org/10.1201/9781003226055

Instituto Nacional de Estadística e Informática. (2024). *Ficha técnica: Encuesta Nacional de Hogares 2024*. https://proyectos.inei.gob.pe/microdatos/

Norwegian Nobel Committee. (2006). *The Nobel Peace Prize 2006* [Comunicado de prensa]. https://www.nobelprize.org/prizes/peace/2006/press-release/

Pitt, M. M., & Khandker, S. R. (1998). The impact of group-based credit programs on poor households in Bangladesh: Does the gender of participants matter? *Journal of Political Economy*, *106*(5), 958–996. https://doi.org/10.1086/250037

R Core Team. (2026). *R: A language and environment for statistical computing* (Versión 4.6.1). R Foundation for Statistical Computing. https://doi.org/10.32614/R.manuals

Roodman, D., & Morduch, J. (2014). The impact of microcredit on the poor in Bangladesh: Revisiting the evidence. *The Journal of Development Studies*, *50*(4), 583–604. https://doi.org/10.1080/00220388.2013.858122