## 14.310x – Data Analysis for Social Scientists

MicroMasters en Statistics and Data Science (SDS) – MITx

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

# Clase 9

## Modelo Lineal Simple y Multivariado

*Soluciones*

**Módulos oficiales del MicroMaster:**
M8 – Single and Multivariate Linear Models

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).



11 de agosto 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 una recta de regresión dada)**.** Un analista ajusta la relación entre las horas de estudio semanal ($X$) y la nota final (sobre 20) y obtiene la *recta de regresión* (regression line)

$$\hat{Y} = 8 + 1.5\,X.$$

(a) ¿Cuál es el *intercepto* (intercept) $\hat{\beta}_0$ y cuál la *pendiente* (slope) $\hat{\beta}_1$?

(b) Predice la nota de alguien que estudia $X = 4$ horas.

(c) Predice la nota de alguien que estudia $X = 6$ horas.

**Solución**

(a) $\hat{\beta}_0 = 8$ (intercepto) y $\hat{\beta}_1 = 1.5$ (pendiente).
(b) $\hat{Y} = 8 + 1.5(4) = 8 + 6 = 14$.
(c) $\hat{Y} = 8 + 1.5(6) = 8 + 9 = 17$. Cada hora extra de estudio sube la nota predicha en 1.5 puntos (la pendiente).

**Ejercicio 1.2** (Interpretar la pendiente en palabras)**.** Para una muestra de distritos, una regresión del gasto mensual en libros $Y$ (en soles) sobre el ingreso familiar $X$ (en cientos de soles) da

$$\hat{Y} = 20 + 3\,X.$$

(a) Interpreta el valor $\hat{\beta}_1 = 3$ *en palabras* (incluye las unidades).

(b) Interpreta el intercepto $\hat{\beta}_0 = 20$: ¿qué representa cuando $X = 0$?

**Solución**

(a) Por cada 100 soles más de ingreso familiar (un aumento de 1 en $X$), el gasto en libros predicho sube en $\hat{\beta}_1 = 3$ **soles**. Es el efecto (asociación) de una unidad más de $X$ sobre $Y$.
(b) El intercepto $\hat{\beta}_0 = 20$ es el gasto predicho cuando el ingreso es $X = 0$: 20 soles. (A menudo el intercepto no tiene interpretación práctica si $X = 0$ cae fuera del rango de los datos; aquí es solo el punto donde la recta corta el eje $Y$.)

**Ejercicio 1.3** (Valor ajustado y residuo)**.** Con la recta $\hat{Y} = 8 + 1.5\,X$ (nota vs. horas de estudio), una alumna estudió $X = 6$ horas y sacó una nota *real* de $Y = 15$.

(a) Calcula el *valor ajustado* (fitted value) $\hat{Y}$ para esta alumna.

(b) Calcula el *residuo* (residual) $\hat{\varepsilon} = Y - \hat{Y}$.

(c) ¿La recta *sobrestima* o *subestima* su nota?

**Solución**

(a) $\hat{Y} = 8 + 1.5(6) = 17$.
(b) $\hat{\varepsilon} = Y - \hat{Y} = 15 - 17 = -2$.
(c) El residuo es **negativo**, así que la recta **sobrestima**: predijo 17 pero la alumna sacó 15 (quedó 2 puntos por debajo de la recta).

**Ejercicio 1.4** (Promedios, $\widehat{\text{Cov}}$ y $\widehat{\text{Var}}$ muestrales)**.** Para 5 familias se registra el ingreso $X$ (en cientos de soles) y el ahorro $Y$ (en cientos de soles):

$$(X, Y): \ (2,5),\ (4,8),\ (6,9),\ (8,13),\ (10,15).$$

(a) Calcula los promedios $\bar{X}$ y $\bar{Y}$.

(b) Calcula $\sum_i (X_i - \bar{X})(Y_i - \bar{Y})$ y $\sum_i (X_i - \bar{X})^2$.

**Solución**

(a) $\bar{X} = \dfrac{2+4+6+8+10}{5} = \dfrac{30}{5} = 6$; $\bar{Y} = \dfrac{5+8+9+13+15}{5} = \dfrac{50}{5} = 10$.
(b) Desvíos $X - \bar{X} = (-4, -2, 0, 2, 4)$ y $Y - \bar{Y} = (-5, -2, -1, 3, 5)$. Productos: $20, 4, 0, 6, 20$, suma $\sum (X - \bar{X})(Y - \bar{Y}) = 50$. Cuadrados de $X$: $16, 4, 0, 4, 16$, suma $\sum (X - \bar{X})^2 = 40$.

**Ejercicio 1.5** (Calcular $\hat{\beta}_1$ y $\hat{\beta}_0$ a mano (guiado))**.** Usa los mismos datos del problema anterior: $(X, Y): \ (2,5),\ (4,8),\ (6,9),\ (8,13),\ (10,15)$, con $\bar{X} = 6$, $\bar{Y} = 10$, $\sum (X - \bar{X})(Y - \bar{Y}) = 50$ y $\sum (X - \bar{X})^2 = 40$.

(a) Calcula la pendiente $\hat{\beta}_1 = \dfrac{\sum(X-\bar{X})(Y-\bar{Y})}{\sum(X-\bar{X})^2}$.

(b) Calcula el intercepto $\hat{\beta}_0 = \bar{Y} - \hat{\beta}_1 \bar{X}$.

(c) Escribe la recta de regresión $\hat{Y} = \hat{\beta}_0 + \hat{\beta}_1 X$.

**Solución**

(a) $\hat{\beta}_1 = \dfrac{50}{40} = 1.25$.
(b) $\hat{\beta}_0 = 10 - 1.25 \cdot 6 = 10 - 7.5 = 2.5$.
(c) $\hat{Y} = 2.5 + 1.25\,X$. (Por cada 100 soles más de ingreso, el ahorro predicho sube 1.25 cientos = 125 soles.)

**Ejercicio 1.6** (Calcular un valor ajustado y un residuo dentro del dataset)**.** Con la recta recién obtenida $\hat{Y} = 2.5 + 1.25\,X$ y la familia con $X = 6$, $Y = 9$:

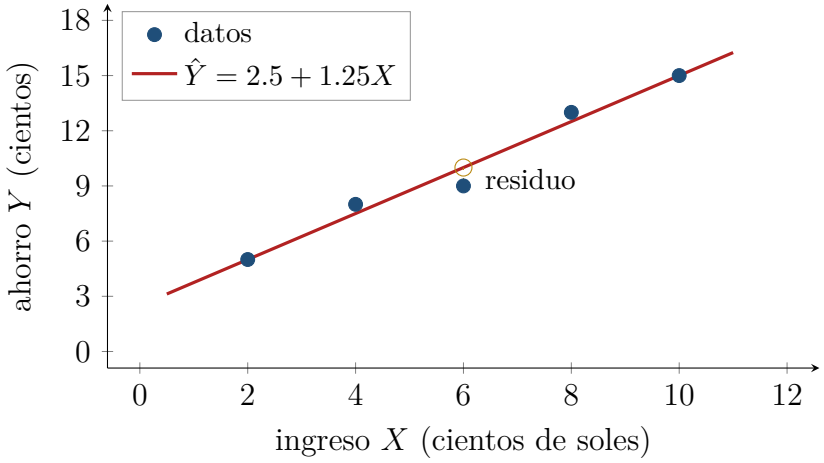
(a) Calcula el valor ajustado $\hat{Y}$ para $X = 6$.

(b) Calcula el residuo $\hat{\varepsilon} = Y - \hat{Y}$.

**Solución**

(a) $\hat{Y} = 2.5 + 1.25(6) = 2.5 + 7.5 = 10$.
(b) $\hat{\varepsilon} = 9 - 10 = -1$. La familia ahorra 1 (cien soles) *menos* de lo que predice la recta para su ingreso.

**Ejercicio 1.7** (Nube de puntos con recta ajustada: lectura visual)**.** La figura muestra 5 puntos $(X, Y)$ y su recta de regresión $\hat{Y} = 2.5 + 1.25\,X$ (la del problema anterior).




(a) ¿La pendiente de la recta es positiva o negativa? ¿Qué significa eso para la relación ingreso–ahorro?

(b) El punto en $X = 6$ está *debajo* de la recta. ¿Su residuo es positivo o negativo?

**Solución**

(a) La pendiente es **positiva** ($\hat{\beta}_1 = 1.25 > 0$): a mayor ingreso, mayor ahorro predicho (relación creciente).
(b) El punto observado $(6, 9)$ está debajo de su ajustado $(6, 10)$, así que el residuo es **negativo** ($\hat{\varepsilon} = 9 - 10 = -1$). Los puntos por encima de la recta tienen residuo positivo; los de abajo, negativo.

## Nivel 2 – Núcleo (core)

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

**Ejercicio 1.8** (OLS completo desde cero (pendiente, intercepto, ajustados, residuos))**.** Para 5 trabajadores se mide la educación $X$ (años por encima de primaria) y el salario por hora $Y$ (en soles). Datos:

$$(X, Y): \ (1, 3),\ (2, 9),\ (3, 10),\ (4, 11),\ (5, 12).$$

(a) Calcula $\bar{X}$, $\bar{Y}$, $\sum(X - \bar{X})(Y - \bar{Y})$ y $\sum(X - \bar{X})^2$.

(b) Obtén $\hat{\beta}_1$ y $\hat{\beta}_0$.

(c) Calcula los valores ajustados $\hat{Y}_i$ y los residuos $\hat{\varepsilon}_i$, y verifica que $\sum_i \hat{\varepsilon}_i = 0$.

**Solución**

(a) $\bar{X} = \dfrac{1+2+3+4+5}{5} = 3$; $\bar{Y} = \dfrac{3+9+10+11+12}{5} = \dfrac{45}{5} = 9$. Desvíos $X - \bar{X} = (-2, -1, 0, 1, 2)$, $Y - \bar{Y} = (-6, 0, 1, 2, 3)$. Productos: $12, 0, 0, 2, 6$, suma $= 20$. Cuadrados de $X$: $4, 1, 0, 1, 4$, suma $= 10$.
(b) $\hat{\beta}_1 = \dfrac{20}{10} = 2$; $\hat{\beta}_0 = \bar{Y} - \hat{\beta}_1 \bar{X} = 9 - 2 \cdot 3 = 3$. Recta: $\hat{Y} = 3 + 2X$.
(c) Ajustados $\hat{Y}_i = 3 + 2X_i = (5, 7, 9, 11, 13)$. Residuos $\hat{\varepsilon}_i = Y_i - \hat{Y}_i = (3 - 5,\ 9 - 7,\ 10 - 9,\ 11 - 11,\ 12 - 13) = (-2, 2, 1, 0, -1)$. Suma $= -2 + 2 + 1 + 0 - 1 = 0$. ✓ (OLS siempre da $\sum_i \hat{\varepsilon}_i = 0$ cuando hay intercepto.)

**Ejercicio 1.9** ($R^2$ vía SST, SSR y SSM)**.** Sigue con la regresión anterior ($\hat{Y} = 3 + 2X$, residuos $(-2, 2, 1, 0, -1)$, $\bar{Y} = 9$, $Y = (3, 9, 10, 11, 12)$).

(a) Calcula $\text{SST} = \sum_i (Y_i - \bar{Y})^2$ y $\text{SSR} = \sum_i \hat{\varepsilon}_i^2$.

(b) Calcula $R^2 = 1 - \text{SSR/SST}$ e interprétalo.

**Solución**

(a) $Y - \bar{Y} = (-6, 0, 1, 2, 3)$, cuadrados $36, 0, 1, 4, 9$: SST $= 36 + 0 + 1 + 4 + 9 = 50$. Residuos al cuadrado: $4, 4, 1, 0, 1$: SSR $= 4 + 4 + 1 + 0 + 1 = 10$.
(b) $R^2 = 1 - \dfrac{10}{50} = 1 - 0.2 = 0.8$. La educación explica el 80 % de la variación del salario en esta muestra (el 20 % restante queda en los residuos).

**Ejercicio 1.10** (Estimar $\sigma^2$ con $s^2 = \text{SSR}/(n-2)$)**.** En la misma regresión, SSR $= 10$ y $n = 5$.

(a) ¿Por qué el denominador es $n - 2$ y no $n$?

(b) Calcula la estimación insesgada de la varianza del error $s^2 = \dfrac{\text{SSR}}{n-2}$ y la desviación estándar de la regresión $s = \sqrt{s^2}$.

**Solución**

(a) Porque al estimar la recta "gastamos" **dos** grados de libertad: uno por $\hat{\beta}_0$ y otro por $\hat{\beta}_1$. Dividir por $n - 2$ (y no por $n$) hace que $s^2$ sea **insesgado** para $\sigma^2 = \text{Var}(\varepsilon)$ (igual que el $-1$ en la varianza muestral, pero aquí se estiman dos parámetros).
(b) $s^2 = \dfrac{10}{5-2} = \dfrac{10}{3} \approx 3.33$ y $s = \sqrt{10/3} \approx 1.83$ soles. (Esa $s$ aparecerá en los errores estándar y las pruebas de la Clase 10.)

**Ejercicio 1.11** (La relación $\hat{\beta}_1 = \widehat{\text{Cov}}(X,Y)/\widehat{\text{Var}}(X)$)**.** La pendiente de OLS se puede escribir como $\hat{\beta}_1 = \dfrac{\widehat{\text{Cov}}(X,Y)}{\widehat{\text{Var}}(X)}$, donde $\widehat{\text{Cov}}(X,Y) = \dfrac{1}{n}\sum_i (X_i - \bar{X})(Y_i - \bar{Y})$ y $\widehat{\text{Var}}(X) = \dfrac{1}{n}\sum_i (X_i - \bar{X})^2$ (mismo $n$ en ambas).

(a) Con el dataset $(X,Y)$ : $(1,3), (2,9), (3,10), (4,11), (5,12)$ ya sabes que $\sum(X - \bar{X})(Y - \bar{Y}) = 20$ y $\sum(X - \bar{X})^2 = 10$ ($n = 5$). Calcula $\widehat{\text{Cov}}(X,Y)$ y $\widehat{\text{Var}}(X)$.

(b) Verifica que $\hat{\beta}_1 = \widehat{\text{Cov}}(X,Y)/\widehat{\text{Var}}(X)$ coincide con el $\hat{\beta}_1 = 2$ de antes. ¿Por qué el $1/n$ se cancela?

**Solución**

(a) $\widehat{\text{Cov}}(X,Y) = \dfrac{20}{5} = 4$; $\widehat{\text{Var}}(X) = \dfrac{10}{5} = 2$.
(b) $\hat{\beta}_1 = \dfrac{4}{2} = 2$. ✓ El factor $1/n$ aparece en el numerador y en el denominador, así que se cancela: $\dfrac{(1/n)\sum(X - \bar{X})(Y - \bar{Y})}{(1/n)\sum(X - \bar{X})^2} = \dfrac{\sum(X - \bar{X})(Y - \bar{Y})}{\sum(X - \bar{X})^2}$. Por eso da igual usar

covarianza/varianza muestrales o las sumas crudas.

**Ejercicio 1.12** (La recta de regresión pasa por $(\bar{X}, \bar{Y})$)**.** Una propiedad clave de OLS (con intercepto) es que la recta ajustada *siempre* pasa por el punto de promedios $(\bar{X}, \bar{Y})$.

(a) Usando $\hat{\beta}_0 = \bar{Y} - \hat{\beta}_1 \bar{X}$, demuestra algebraicamente que al evaluar la recta en $X = \bar{X}$ se obtiene $\hat{Y} = \bar{Y}$.

(b) Verifícalo con la recta $\hat{Y} = 3 + 2X$ y los promedios $\bar{X} = 3$, $\bar{Y} = 9$.

**Solución**

(a) Evaluamos la recta en $X = \bar{X}$:

$$\hat{Y}(\bar{X}) = \hat{\beta}_0 + \hat{\beta}_1 \bar{X} = (\bar{Y} - \hat{\beta}_1 \bar{X}) + \hat{\beta}_1 \bar{X} = \bar{Y}.$$

Los términos $\hat{\beta}_1 \bar{X}$ se cancelan, así que $\hat{Y}(\bar{X}) = \bar{Y}$ exactamente: la recta pasa por $(\bar{X}, \bar{Y})$.
(b) $\hat{Y}(3) = 3 + 2 \cdot 3 = 9 = \bar{Y}$. ✓

**Ejercicio 1.13** (Predicción dentro y fuera del rango (extrapolación))**.** Con la regresión salario–educación $\hat{Y} = 3 + 2X$ ($X$ = años sobre primaria, $Y$ = soles/hora; los datos van de $X = 1$ a $X = 5$):

(a) Predice el salario para $X = 4$ (dentro del rango).

(b) Predice el salario para $X = 20$. ¿Por qué hay que tomar esta predicción con cautela?

**Solución**

(a) $\hat{Y} = 3 + 2(4) = 11$ soles/hora.
(b) $\hat{Y} = 3 + 2(20) = 43$ soles/hora. **Cautela:** $X = 20$ está muy lejos del rango observado (1 a 5); predecir allí es *extrapolar* y supone que la relación lineal sigue valiendo fuera de los datos, lo cual no está garantizado (el salario podría aplanarse). La regresión solo es confiable dentro (o cerca) del rango de $X$ observado.

**Ejercicio 1.14** (Interpretar un coeficiente con signo y unidades)**.** Un estudio sobre 200 viviendas regresa el precio de alquiler $Y$ (en soles/mes) sobre la distancia al centro $X$ (en km) y obtiene

$$\hat{Y} = 1\,800 - 90\,X, \qquad R^2 = 0.42.$$

(a) Interpreta $\hat{\beta}_1 = -90$ en palabras (signo y unidades).

(b) ¿Qué dice $R^2 = 0.42$ sobre el ajuste?

**Solución**

(a) Por cada km adicional de distancia al centro, el alquiler predicho **baja** en 90 soles/mes (el signo negativo indica relación decreciente: más lejos, más barato).
(b) $R^2 = 0.42$: la distancia al centro explica el 42 % de la variación del alquiler; el 58 % restante se debe a otros factores (tamaño, antigüedad, seguridad, etc.) no incluidos en este modelo simple.

**Ejercicio 1.15** (Efecto de cambiar las unidades de $X$)**.** La regresión $\hat{Y} = 3 + 2X$ tiene $X$ en *años* de educación e $Y$ en soles/hora. Un colega quiere medir la educación en *meses* ($X' = 12X$).

(a) Si $X' = 12X$, ¿cuál es la nueva pendiente $\hat{\beta}_1'$ de la regresión de $Y$ sobre $X'$? (Pista: el efecto de *un mes* debe ser 1/12 del efecto de un año.)

(b) ¿Cambia el intercepto $\hat{\beta}_0$? ¿Cambia el $R^2$?

**Solución**

(a) Como $X = X'/12$, sustituyendo: $\hat{Y} = 3 + 2(X'/12) = 3 + \frac{2}{12}X' = 3 + \frac{1}{6}X'$. La nueva pendiente es $\hat{\beta}_1' = \frac{2}{12} = \frac{1}{6} \approx 0.167$ soles por *mes* de educación (consistente: 12 meses $\times \frac{1}{6} = 2$ soles, el efecto de un año).
(b) El **intercepto no cambia** ($\hat{\beta}_0 = 3$): cuando $X = 0$ también $X' = 0$. El $R^2$ **no cambia**: reescalar linealmente un regresor no altera la bondad de ajuste, solo las unidades del coeficiente.

**Ejercicio 1.16** (Variable dummy: la pendiente es una diferencia de medias)**.** Sea $D$ una variable *dummy* (indicadora): $D = 1$ si la persona vive en zona urbana y $D = 0$ si vive en zona rural. Se regresa el ingreso $Y$ (soles) sobre $D$:

$$\hat{Y} = 900 + 400\, D.$$

(a) ¿Cuál es el ingreso promedio predicho de los **rurales** ($D = 0$) y de los **urbanos** ($D = 1$)?

(b) ¿Qué representa $\hat{\beta}_1 = 400$?

**Solución**

(a) Rurales ($D = 0$): $\hat{Y} = 900 + 400(0) = 900$ soles. Urbanos ($D = 1$): $\hat{Y} = 900 + 400(1) = 1\,300$ soles. Es decir, $\hat{\beta}_0 = 900$ es la media del grupo base (rural) y $\hat{\beta}_0 + \hat{\beta}_1 = 1\,300$ la del grupo urbano.
(b) $\hat{\beta}_1 = 400$ es la **diferencia de medias** entre urbanos y rurales: los urbanos ganan en promedio 400 soles más. Con un solo regresor dummy, la pendiente OLS coincide con

$\bar{Y}_{D=1} - \bar{Y}_{D=0}$.

**Ejercicio 1.17** (De $R^2$ a SSR y SSM)**.** Una regresión con $n = 30$ observaciones reporta $\text{SST} = 500$ y $R^2 = 0.64$.

(a) Calcula SSR (suma de cuadrados de residuos) y $\text{SSM} = \text{SST} - \text{SSR}$ (suma de cuadrados del modelo).

(b) Estima $\sigma^2$ con $s^2 = \text{SSR}/(n-2)$.

**Solución**

(a) Como $R^2 = 1 - \text{SSR}/\text{SST}$, se tiene $\text{SSR} = (1-R^2)\,\text{SST} = (1-0.64)(500) = 0.36 \cdot 500 = 180$. $\text{SSM} = \text{SST} - \text{SSR} = 500 - 180 = 320$ (lo "explicado" por el modelo); en efecto $\text{SSM}/\text{SST} = 320/500 = 0.64 = R^2$.
(b) $s^2 = \dfrac{180}{30-2} = \dfrac{180}{28} \approx 6.43$.

**Ejercicio 1.18** ($R^2$ en regresión bivariada y el coeficiente de correlación)**.** El material de clase (MIT 14.310x, L17) recuerda que, en regresión *bivariada* (un solo regresor), $R^2 = r_{XY}^2$, el cuadrado del coeficiente de correlación muestral entre $X$ e $Y$.

(a) Si la correlación muestral es $r_{XY} = 0.9$, ¿cuánto vale $R^2$?

(b) Si en otra regresión $R^2 = 0.49$ y la pendiente es *negativa*, ¿cuál es $r_{XY}$ (con su signo)?

**Solución**

(a) $R^2 = r_{XY}^2 = (0.9)^2 = 0.81$.
(b) $r_{XY} = \pm\sqrt{0.49} = \pm 0.7$. Como la pendiente es negativa y $r_{XY}$ tiene el *mismo signo* que $\hat{\beta}_1$ (ambos comparten el signo de la covarianza), $r_{XY} = -0.7$.

## Nivel 3 – Aplicación

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

**Ejercicio 1.19** (Regresión múltiple: "manteniendo constantes los demás")**.** Con datos de un mercado laboral peruano se estima el salario mensual $Y$ (en soles) según educación ed (años), experiencia exp (años) y un dummy mujer (= 1 si es mujer):

$$\widehat{Y} = 800 + 250\,\text{ed} + 60\,\text{exp} - 300\,\text{mujer}.$$

(1) Interpreta el coeficiente 250 de educación *manteniendo constantes* la experiencia y el sexo.

(2) Interpreta el coeficiente $-300$ del dummy mujer.

(3) Predice el salario de un **hombre** con ed $= 5$, exp $= 10$ y de una **mujer** con los mismos valores. ¿Cuál es la brecha?

**Solución**

(1) A igual experiencia y mismo sexo, cada año adicional de educación se asocia con 250 soles más de salario mensual (efecto *parcial* de la educación, "manteniendo constantes los demás").
(2) A igual educación y experiencia, una mujer gana en promedio 300 soles **menos** que un hombre (la categoría base es "hombre", con mujer $= 0$). Es una brecha salarial condicional, no necesariamente causal.
(3) Hombre: $\hat{Y} = 800 + 250(5) + 60(10) - 300(0) = 800 + 1\,250 + 600 = 2\,650$. Mujer: $\hat{Y} = 800 + 1\,250 + 600 - 300 = 2\,350$. Brecha $= 2\,650 - 2\,350 = 300$ soles, igual al coeficiente del dummy.
*(Adaptado de MIT 14.310x, L17–L18: multiple regression e interpretación de coeficientes.)*

**Ejercicio 1.20** (Variables dummy con varias categorías y la trampa de las dummies)**.** Se quiere incluir la *región* (Costa, Sierra, Selva) en una regresión de salarios. Un analista crea tres dummies: Costa, Sierra, Selva (cada una 0/1) y las mete *todas* junto con el intercepto.

(1) Explica por qué esto provoca *colinealidad perfecta* (la "trampa de las dummies", *dummy variable trap*).

(2) Propon la solución correcta. Si se omite "Costa" como **categoría base**, ¿cómo se interpreta el coeficiente de Sierra?

**Solución**

(1) Como toda persona pertenece a exactamente una región, Costa + Sierra + Selva $= 1$ para todos, que es justo la columna del intercepto (una columna de unos). Las cuatro columnas son linealmente dependientes, así que $\mathbf{X}^\top\mathbf{X}$ es **singular** (no invertible) y OLS no tiene solución única.
(2) Solución: incluir el intercepto y *solo dos* dummies, dejando una región como **categoría base** (*reference category*). Con "Costa" como base, el coeficiente de Sierra es la diferencia salarial promedio **Sierra vs. Costa**, manteniendo lo demás constante (análogo con Selva). En general, con $k$ categorías se usan $k - 1$ dummies.
*(Adaptado de MIT 14.310x, L17: dummy variables.)*

**Ejercicio 1.21** (Regresión simple vs. múltiple: el coeficiente cambia)**.** Para estudiar el efecto de las *horas de sueño* sobre la nota, se corren dos regresiones sobre los mismos estudiantes:

$$\text{(simple) } \widehat{\text{nota}} = 8 + 1.2\,\text{sueño}; \qquad \text{(múltiple) } \widehat{\text{nota}} = 6 + 0.4\,\text{sueño} + 0.9\,\text{estudio}.$$

(1) ¿Por qué el coeficiente del sueño cae de 1.2 a 0.4 al añadir las horas de estudio?

(2) ¿Cuál de los dos coeficientes (1.2 o 0.4) se acerca más al efecto *propio* del sueño? Explica.

**Solución**

(1) En la regresión simple, el sueño "carga" parte del efecto de una variable correlacionada y omitida: quienes duermen más suelen también *estudiar más* y esas horas de estudio elevan la nota. Al incluir estudio en la múltiple, ese canal se controla y el coeficiente del sueño queda *neto* (0.4): el efecto que antes era 1.2 incluía la contribución del estudio.
(2) El de la regresión **múltiple** (0.4) se acerca más al efecto propio del sueño, porque compara estudiantes *con las mismas horas de estudio* (mantiene constante el estudio). El 1.2 confunde sueño con estudio (sesgo por variable omitida; ver Nivel 4).
*(Adaptado de MIT 14.310x, L17–L18.)*

**Ejercicio 1.22** (Predecir con varios regresores e interpretar el ajuste)**.** Una municipalidad modela el gasto mensual en transporte $Y$ (soles) de los hogares:

$$\widehat{Y} = 50 + 0.05\,\text{ingreso} + 120\,\text{auto} - 15\,\text{distancia}, \qquad R^2 = 0.55,$$

donde ingreso está en soles, auto $= 1$ si el hogar tiene auto y distancia en km al trabajo.

(1) Predice el gasto de un hogar con ingreso 3 000, auto $= 1$ y distancia $= 8$.

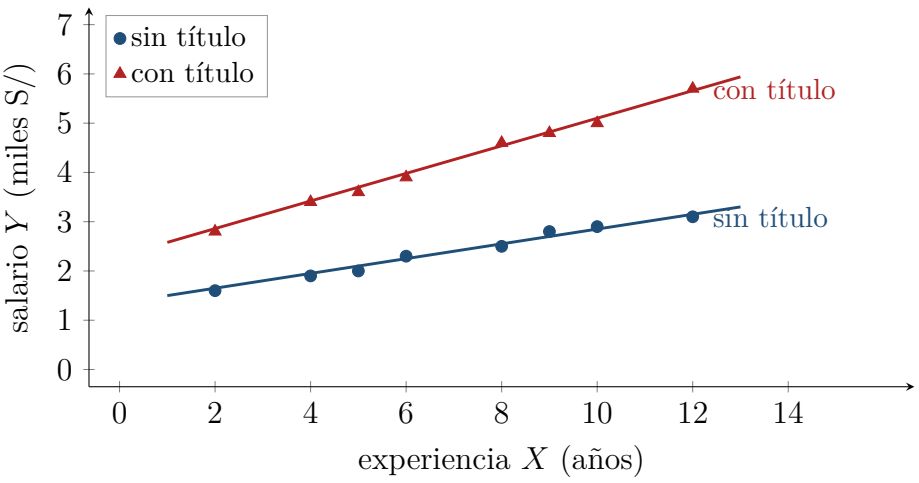
(2) Interpreta el coeficiente 120 del dummy auto y el $-15$ de distancia, ambos "manteniendo constantes los demás".

(3) ¿Qué informa $R^2 = 0.55$?

**Solución**

(1) $\hat{Y} = 50 + 0.05(3\,000) + 120(1) - 15(8) = 50 + 150 + 120 - 120 = 200$ soles.
(2) 120: a igual ingreso y distancia, tener auto se asocia con 120 soles más de gasto en transporte. $-15$: a igual ingreso y tenencia de auto, cada km adicional al trabajo *reduce* el gasto predicho en 15 soles (quizá porque vivir más lejos va con otras decisiones; es asociación, no causa).
(3) $R^2 = 0.55$: el modelo explica el 55 % de la variación del gasto en transporte entre hogares; el 45 % restante depende de factores no incluidos.
*(Adaptado de MIT 14.310x, L18: multivariate linear model.)*

**Ejercicio 1.23** (Comparar dos rectas en una nube de puntos (dummy de grupo))**.** La figura muestra el salario $Y$ (miles de soles) vs. años de experiencia $X$ para dos grupos: con título universitario (rojo) y sin título (azul), junto con la recta ajustada de cada grupo.




(1) ¿Qué grupo tiene mayor intercepto y cuál mayor pendiente? Interpreta ambas diferencias.

(2) Si quisieras capturar ambas diferencias (nivel y pendiente) en *una sola* regresión, ¿qué términos incluirías? (idea)

**Solución**

(1) El grupo **con título** (rojo) tiene mayor intercepto ($\approx 2.3$ vs. $\approx 1.35$): a 0 años de experiencia ya parte de un salario más alto. También tiene mayor **pendiente** ($\approx 0.28$ vs. $\approx 0.15$): cada año de experiencia "rinde" más para los titulados, así que la brecha *crece* con la experiencia.
(2) Incluir un dummy titulo (diferencia de *nivel*/intercepto) y un **término de interacción** titulo$\times$experiencia (diferencia de *pendiente*): $\widehat{Y} = \beta_0 + \beta_1 \text{exp} + \beta_2 \text{titulo} + \beta_3(\text{titulo} \times \text{exp})$. Así $\beta_2$ desplaza el intercepto y $\beta_3$ cambia la pendiente para los titulados.
*(Adaptado de MIT 14.310x, L17–L18: dummies e interacciones.)*

**Ejercicio 1.24** (Construir la recta y el $R^2$ en contexto (gasto vs. ingreso))**.** Una encuesta a 5 hogares registra el ingreso $X$ (en cientos de soles) y el gasto en alimentos $Y$ (en cientos de soles):

$$(X, Y): \ (1,3),\ (2,9),\ (3,10),\ (4,11),\ (5,12).$$

(Es el mismo patrón numérico del Núcleo, ahora en contexto.)

(1) Halla la recta de regresión $\hat{Y} = \hat{\beta}_0 + \hat{\beta}_1 X$.

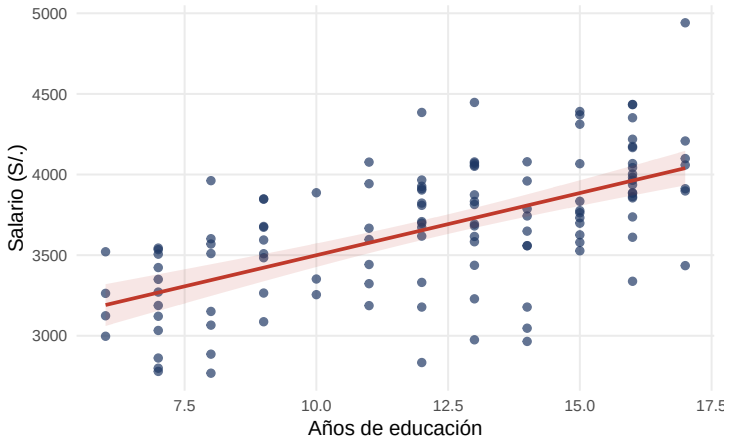
(2) Calcula $R^2$ e interprétalo.

(3) Estima la *propensión marginal a gastar* en alimentos por cada 100 soles de ingreso e interpreta el intercepto en este contexto.

**Solución**

(1) $\bar{X} = 3$, $\bar{Y} = 9$, $\sum(X - \bar{X})(Y - \bar{Y}) = 20$, $\sum(X - \bar{X})^2 = 10$, así $\hat{\beta}_1 = 20/10 = 2$ y $\hat{\beta}_0 = 9 - 2 \cdot 3 = 3$: $\hat{Y} = 3 + 2X$.
(2) SST $= 50$, SSR $= 10$ (residuos $-2, 2, 1, 0, -1$), así $R^2 = 1 - 10/50 = 0.8$: el ingreso explica el 80 % de la variación del gasto en alimentos.
(3) La pendiente $\hat{\beta}_1 = 2$ es la *propensión marginal a gastar*: por cada 100 soles más de ingreso, el gasto en alimentos sube 200 soles. El intercepto $\hat{\beta}_0 = 3$ (es decir 300 soles) sería el gasto en alimentos de un hogar con ingreso "0"; es una extrapolación fuera del rango y conviene leerlo solo como el corte de la recta.
*(Adaptado de MIT 14.310x, L17.)*

**Ejercicio 1.25** (Interpretar coeficientes en R: de una a dos variables)**.** El archivo `salarios.csv` tiene 120 trabajadores: `salario` (S/.), `educacion`, `experiencia` y `region`. Estimaremos el salario **añadiendo regresores de a poco** para leer *qué representa cada coeficiente*. La nube salario vs. educación con la recta simple (M1):




**Completa el código**

```
datos <- read.csv("salarios.csv")
m1 <- lm(salario ~ ____, data = datos)              # una variable (solo
    educacion)
summary(m1)                  # coeficiente, p-value, R^2
m2 <- lm(salario ~ educacion + ____, data = datos)   # dos variables (+
    experiencia)
summary(m2)
```

(a) Completa las dos casillas.

(b) ¿Qué representa el coeficiente de `educacion` en **M1** y en **M2**? ¿Cuál es *ceteris paribus*? ¿Son significativos (mira el $p$-value)?

(c) El coeficiente de `educacion` pasa de $\approx 77$ (M1) a $\approx 197$ (M2). ¿Por qué cambia tanto al agregar la experiencia?

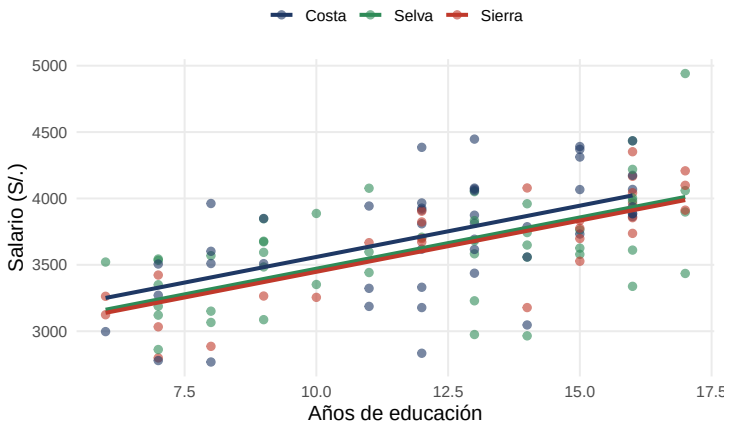
**Solución**

**Código completo**

```
datos <- read.csv("salarios.csv")
m1 <- lm(salario ~ educacion, data = datos)
summary(m1)   # educacion 77.1  (p<0.001)   R^2 = 0.37
m2 <- lm(salario ~ educacion + experiencia, data = datos)
summary(m2)   # educacion 197.2   experiencia 67.8  (p<0.001)   R^2 = 0.76
cor(datos$educacion, datos$experiencia)   # -0.83
```

**(a)** Las casillas son `educacion` y `experiencia`.
**(b) M1** ($\hat{\beta}_{educ} \approx 77$): la asociación *cruda* — cada año de educación va con ≈S/.77 más, *sin controlar nada*. **M2** ($\hat{\beta}_{educ} \approx 197$): el efecto *ceteris paribus* — ≈S/.197 por año de educación **manteniendo la experiencia constante** (y ≈S/.68 por año de experiencia). Ambos con $p < 0.001$; el $R^2$ sube de 0.37 a 0.76.
**(c)** Es **sesgo por variable omitida** (*omitted variable bias*): educación y experiencia están muy correlacionadas ($r = -0.83$: quien estudia más empieza a trabajar más tarde). En M1, al *omitir* la experiencia, el coeficiente de educación cargaba con parte del efecto (a la baja) de la experiencia; al controlarla en M2 queda limpio y sube a 197. Moraleja: *el coeficiente de un regresor depende de qué más haya en el modelo.*
*(Adaptado de MIT 14.310x, L17–L18: regresión múltiple, ceteris paribus, OVB.)*

**Ejercicio 1.26** (Dummy y modelo mixto en R: cada coeficiente y el $F$-test)**.** Seguimos con `salarios.csv`. Ahora usamos la variable *categórica* `region` (Costa, Sierra, Selva) como **dummy**, y luego mezclamos todo. Con dummies las rectas por región son **paralelas** (mismo slope, distinto intercepto):



**Completa el código**

```
datos <- read.csv("salarios.csv")
m3 <- lm(salario ~ educacion + ____(region), data = datos)   # dummy de region
summary(m3)                 # Costa = base; cada coef = diferencia vs Costa
m4 <- lm(salario ~ educacion + experiencia + factor(region), data = datos)  #
   mixto
summary(m4)
# F-test: las dummies de region, juntas, agregan algo?
m_sin <- lm(salario ~ educacion + experiencia, data = datos)
anova(m_sin, ____)          # compara sin region vs con region
```

(a) Completa las dos casillas.

(b) ¿Qué representa el coeficiente de `Sierra`? ¿Respecto de quién se mide?

(c) ¿Las dummies de región son **conjuntamente** significativas? ¿Qué prueba usas y en qué se diferencia de mirar cada $p$-value por separado?

## Solución

**Código completo**

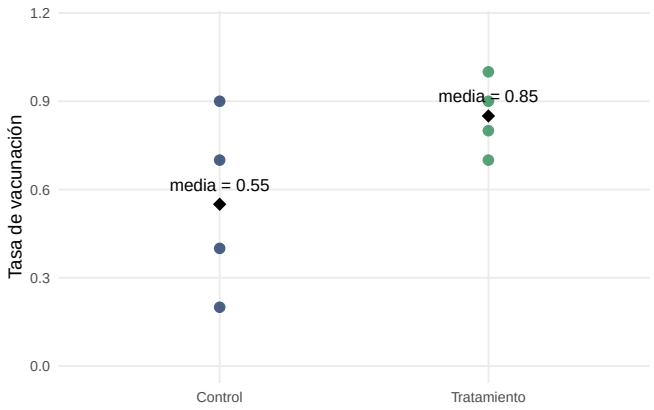
```
m3 <- lm(salario ~ educacion + factor(region), data = datos)
summary(m3)   # Selva -88.7  Sierra -112.1   educacion 77.2   R^2 = 0.383
m4 <- lm(salario ~ educacion + experiencia + factor(region), data = datos)
summary(m4)   # educacion 197.5  experiencia 67.9  Selva -85.3  Sierra
   -119.9  R^2=0.773
m_sin <- lm(salario ~ educacion + experiencia, data = datos)
anova(m_sin, m4)   # F = 3.55,  p = 0.032
```

**(a)** Las casillas son `factor` y `m4`.
**(b)** `Costa` es la **categoría base (baseline)**, omitida para evitar la *dummy variable trap*. El coef de `Sierra` ($\approx -120$ en M4) es *cuánto menos gana, en promedio, un trabajador de Sierra que uno de* ***Costa*** *con la misma educación y experiencia* (ceteris paribus): solo desplaza el intercepto de su recta.
**(c)** Cada $p$-value de la tabla prueba si *esa dummy sola* importa; para la significancia **conjunta** de la región (*joint significance*) se usa un $F$**-test** (`anova`: la prueba $F$ de Fisher que compara el modelo con y sin las dummies). Da $F = 3.55$, $p = 0.032 < 0.05$: la región **sí** es significativa en conjunto, aunque en M3 sus dummies por separado no lo fueran.
*(Adaptado de MIT 14.310x, L18–L19: dummies, categoría base y prueba F conjunta.)*

## Nivel 4 – Reto

*Difícil: integra varios temas en un mismo problema.*

**Ejercicio 1.27** (Prueba exacta de Fisher: enumerar todas las combinaciones (método MIT))**.** Un programa de salud se aplicó a 4 de 8 regiones (regiones $1, 3, 4, 8$); se mide la **tasa de vacunación** (`programa.csv`). La media es 0.85 en tratamiento y 0.55 en control, así que el **estimador del ATE** (*average treatment effect*) — la *difference in means* — vale 0.30. Ese 0.30 es el **observed test statistic** (`observed_test`); **no** es el $p$-valor:




¿Es 0.30 real o pudo salir por azar al elegir qué regiones tratar? Bajo la **sharp null** de Fisher (el programa no tiene efecto en *nadie*) cualquiera de las $\binom{8}{4} = 70$ asignaciones es igual de probable. Las **enumeramos todas** en una matriz de 0 y 1 y la **multiplicamos** por los outcomes (así, sin barajar al azar, obtenemos la *null distribution* exacta):

**Completa el código**

```
library(perm)
datos <- read.csv("programa.csv")     # region, grupo (T/C), tasa
A <- datos$tasa                       # outcome (vaccination rate)
n_treat <- sum(datos$grupo == "T")    # treated units = 4
# observed test statistic = |difference in means| (estima el ATE)
observed_test <- abs(mean(A[datos$grupo=="T"]) - mean(A[datos$grupo=="C"]))
    # 0.30

# todas las C(8,4)=70 asignaciones -> matriz de 0 y 1 (70 x 8)
perms <- chooseMatrix(8, ____)        # treated units
treatment_avg  <- perms %*% ____ / n_treat        # multiplica por A (los
    outcomes)
control_avg    <- (1 - perms) %*% A / n_treat
test_statistic <- abs(treatment_avg - control_avg)   # una por asignacion (
    sharp null)
pvalue <- mean(test_statistic >= ____)   # fraccion tan o mas extrema que
    observed_test
pvalue
```

(a) Completa las tres casillas.

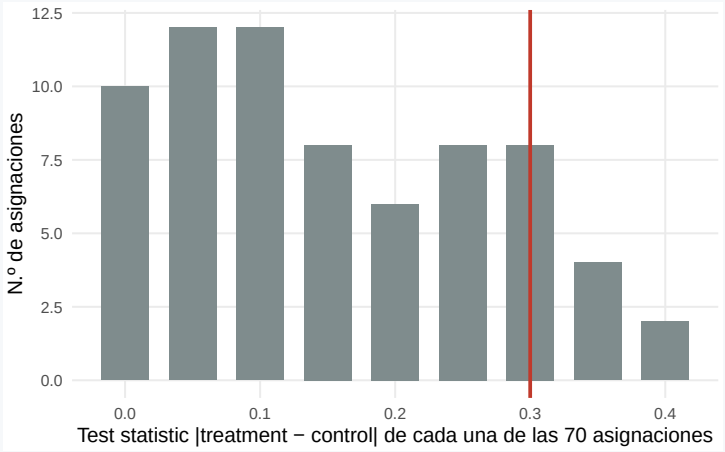
(b) ¿Qué es cada fila de `perms`, y qué calcula `perms %*% A`?

(c) ¿`observed_test` es el ATE o el $p$-valor? Con $\alpha = 0.05$, ¿rechazas la *sharp null*? Concluye.

## Solución

### Código completo

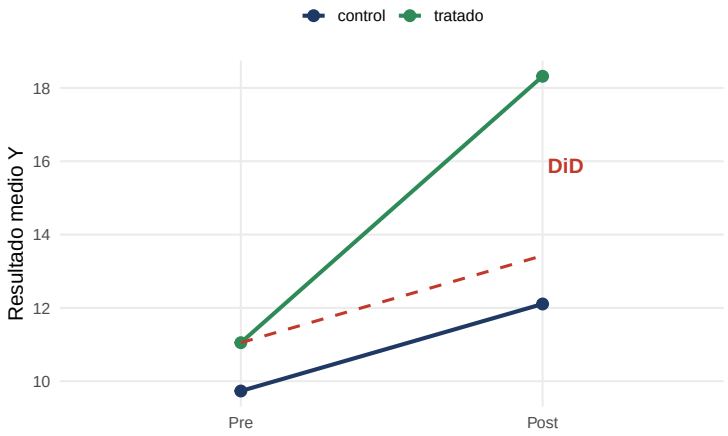
```
library(perm)
datos <- read.csv("programa.csv");  A <- datos$tasa
n_treat <- sum(datos$grupo == "T")                          # 4
observed_test <- abs(mean(A[datos$grupo=="T"]) - mean(A[datos$grupo=="C"]))
      # 0.30
perms <- chooseMatrix(8, n_treat)     # 70 x 8, cada fila una asignacion
    (0/1)
treatment_avg  <- perms %*% A / n_treat
control_avg    <- (1 - perms) %*% A / n_treat
test_statistic <- abs(treatment_avg - control_avg)
pvalue <- mean(test_statistic >= observed_test)
pvalue                                  # 14/70 = 0.20  ->  no se rechaza
```




*Null distribution* exacta: las 70 asignaciones agrupadas por su `test_statistic`; la línea roja es el `observed_test` (0.30). Solo 14 de 70 lo alcanzan.
**(a)** Las casillas son `n_treat`, `A` y `observed_test`.
**(b)** Cada *fila* de `perms` es una asignación: qué 4 regiones son tratadas (1) y cuáles control (0). `perms %*% A` suma los outcomes de las tratadas en cada asignación; al dividir entre `n_treat` da el `treatment_avg`. Recorriendo las 70 filas se obtiene el `test_statistic` en *todas* las asignaciones posibles: la *null distribution*.
**(c)** `observed_test`= 0.30 es el **test statistic observado** (la *difference in means*, magnitud del **ATE estimado**), **no** el $p$-valor. El $p$-valor es `pvalue`= $14/70 = 0.20 > 0.05$, así

que **no se rechaza** la sharp null. Aunque el ATE estimado parece grande (0.85 vs 0.55), con solo 8 regiones (70 asignaciones) no es raro bajo $H_0$: falta evidencia (y potencia). *(Del examen final de MIT 14.310x: Fisher exact test por enumeración, L15.)*

**Ejercicio 1.28** (Diferencias en diferencias (diff-in-diff) en R)**.** Un programa social se aplica a un grupo *tratado* y no a un *control*; se observa a cada uno **antes** (`post=0`) y **después** (`post=1`). El archivo `did_panel.csv` tiene columnas `tratado` (0/1), `post` (0/1) y el resultado `y`. El estimador *diff-in-diff* resta la tendencia común (la línea roja punteada es el contrafactual del tratado):




**Completa el código**

```
datos <- read.csv("did_panel.csv")
# interaccion explicita (estilo Qian): 1 solo para tratados en el post
datos$tratadopost <- datos$tratado * ____   # multiplica por post
# DiD por regresion: el ATE es el coeficiente de la interaccion
m <- lm(y ~ tratado + post + ____, data = datos)   # agrega tratadopost
summary(m)
coef(m)["tratadopost"]      # estimador DiD
```

(a) Completa las dos casillas.

(b) Calcula el DiD "a mano" con las cuatro medias de celda y verifica que coincide con la interacción.

(c) ¿El efecto es significativo? ¿Qué supuesto clave necesita el DiD?

**Solución**

**Código completo**

```
datos$tratadopost <- datos$tratado * datos$post
m <- lm(y ~ tratado + post + tratadopost, data = datos)
summary(m)
coef(m)["tratadopost"]      # ~ 4.89   (p < 0.001)

# A mano: (cambio en tratados) - (cambio en control)
with(datos, {
  dT <- mean(y[tratado==1 & post==1]) - mean(y[tratado==1 & post==0])
  dC <- mean(y[tratado==0 & post==1]) - mean(y[tratado==0 & post==0])
  dT - dC                   # ~ 4.89  (identico a la interaccion)
})
```

**(a)** Las casillas son `datos$post` y `tratadopost`.
**(b)** El DiD es (cambio del tratado) menos (cambio del control): $(\bar{y}_{T,post} - \bar{y}_{T,pre}) - (\bar{y}_{C,post} - \bar{y}_{C,pre}) \approx 4.89$, *exactamente* el coeficiente de `tratadopost`.
**(c)** Sí: $p < 0.001$. El programa sube $Y$ en $\approx 4.9$ unidades una vez que descontamos la tendencia común. Supuesto clave: **tendencias paralelas (parallel trends)** — sin el programa, tratado y control habrían evolucionado igual (la línea punteada).
*(Método del examen final de MIT 14.310x, Qian 2008: DiD con `tratado*post`; en el examen se añaden efectos fijos con `felm(...+ G(factor(admin)))`.)*

**Ejercicio 1.29** (OLS "a mano" en R: $\hat{\boldsymbol{\beta}} = (\mathbf{X}^\top\mathbf{X})^{-1}\mathbf{X}^\top\mathbf{Y}$ y verificar con `lm`)**.** La idea es *ver que `lm` no es magia*: calculamos los coeficientes OLS con álgebra de matrices y confirmamos que dan lo mismo. Usamos `salarios.csv` (`salario`, `educacion`, `experiencia`).

**Completa el código**

```
datos <- read.csv("salarios.csv")
Y <- datos$salario
# matriz de diseno X: columna de 1s (intercepto) + regresores
X <- cbind(1, datos$educacion, ____)     # agrega experiencia
XtX <- t(X) %*% X
XtY <- t(X) %*% Y
beta_hat <- ____(XtX) %*% XtY            # invierte X'X con solve()
beta_hat
# verificar contra lm (mismos numeros):
coef(lm(salario ~ educacion + experiencia, data = datos))
```

(a) Completa las dos casillas.

(b) ¿Por qué la primera columna de $\mathbf{X}$ es de unos?

(c) ¿Coinciden `beta_hat` y `coef(lm(...))`? ¿Qué hace `lm` por dentro?

**Solución**

**Código completo**

```
X <- cbind(1, datos$educacion, datos$experiencia)
beta_hat <- solve(t(X) %*% X) %*% (t(X) %*% Y)
beta_hat        # 45.60 (intercepto)  197.18 (educacion)  67.80 (
    experiencia)
coef(lm(salario ~ educacion + experiencia, data = datos))   # identico
```

**(a)** Las casillas son `datos$experiencia` y `solve`.
**(b)** La columna de 1s estima el **intercepto** $\beta_0$: en $\hat{Y} = \beta_0 \cdot 1 + \beta_1 \,\text{educ} + \beta_2 \,\text{exp}$, esa columna es la que multiplica a $\beta_0$. Sin ella, la recta pasaría *forzada por el origen* ($\beta_0 = 0$), lo que casi nunca es correcto.
**(c)** Sí, coinciden hasta el último decimal (la diferencia es $\sim 10^{-10}$, solo redondeo). La fórmula $\hat{\boldsymbol{\beta}} = (\mathbf{X}^\top\mathbf{X})^{-1}\mathbf{X}^\top\mathbf{Y}$ resuelve las *normal equations* y es *exactamente* lo que `lm` calcula por dentro (con un método numérico más estable, la descomposición **QR**, en vez de invertir $\mathbf{X}^\top\mathbf{X}$ a mano). Por eso "a mano" y con `lm` sale lo mismo: `lm` *es* esta cuenta de matrices.
*(Adaptado de MIT 14.310x, L18: OLS en notación matricial.)*

**Ejercicio 1.30** (Derivar $\hat{\beta}_0, \hat{\beta}_1$ por condiciones de primer orden)**.** OLS elige $\hat{\beta}_0, \hat{\beta}_1$ minimizando la suma de cuadrados $S(\beta_0, \beta_1) = \sum_{i=1}^{n}(Y_i - \beta_0 - \beta_1 X_i)^2$.

(1) Deriva las dos *condiciones de primer orden* $\partial S/\partial\beta_0 = 0$ y $\partial S/\partial\beta_1 = 0$ (las *ecuaciones normales*).

(2) De la primera ecuación obtén $\hat{\beta}_0 = \bar{Y} - \hat{\beta}_1\bar{X}$; de la segunda, sustituyendo $\hat{\beta}_0$, obtén $\hat{\beta}_1 = \dfrac{\sum(X_i - \bar{X})(Y_i - \bar{Y})}{\sum(X_i - \bar{X})^2}$.

**Solución**

(1) Derivando:

$$\frac{\partial S}{\partial \beta_0} = \sum_i 2(Y_i - \beta_0 - \beta_1 X_i)(-1) = 0 \;\Rightarrow\; \sum_i (Y_i - \hat{\beta}_0 - \hat{\beta}_1 X_i) = 0,$$

$$\frac{\partial S}{\partial \beta_1} = \sum_i 2(Y_i - \beta_0 - \beta_1 X_i)(-X_i) = 0 \;\Rightarrow\; \sum_i X_i(Y_i - \hat{\beta}_0 - \hat{\beta}_1 X_i) = 0.$$

Estas son las *ecuaciones normales*. La primera dice $\sum_i \hat{\varepsilon}_i = 0$; la segunda, $\sum_i X_i\hat{\varepsilon}_i = 0$.
(2) De la primera: $\sum Y_i = n\hat{\beta}_0 + \hat{\beta}_1 \sum X_i$, dividiendo entre $n$ da $\bar{Y} = \hat{\beta}_0 + \hat{\beta}_1\bar{X}$, es decir

$\boxed{\hat{\beta}_0 = \bar{Y} - \hat{\beta}_1 \bar{X}}$. Sustituyendo en la segunda y usando $\sum X_i(Y_i - \bar{Y}) - \hat{\beta}_1 \sum X_i(X_i - \bar{X}) = 0$ (porque $\sum X_i(\bar{Y} - \hat{\beta}_1 \bar{X} + \hat{\beta}_1 X_i)$ reordena a esto), y como $\sum X_i(X_i - \bar{X}) = \sum (X_i - \bar{X})^2$ y $\sum X_i(Y_i - \bar{Y}) = \sum (X_i - \bar{X})(Y_i - \bar{Y})$ (los $\bar{X} \sum(\cdot)$ se anulan):

$$\boxed{\hat{\beta}_1 = \frac{\sum_i (X_i - \bar{X})(Y_i - \bar{Y})}{\sum_i (X_i - \bar{X})^2} = \frac{\widehat{\mathrm{Cov}}(X, Y)}{\widehat{\mathrm{Var}}(X)}}.$$

*(Adaptado de MIT 14.310x, L17: derivación de OLS; MIT 18.05, mínimos cuadrados.)*

**Ejercicio 1.31** (Propiedades de los residuos: $\sum \hat{\varepsilon}_i = 0$ y $\sum X_i \hat{\varepsilon}_i = 0$)**.** A partir de las ecuaciones normales del problema anterior, demuestra dos propiedades de los residuos OLS y una consecuencia geométrica.

(1) Muestra que $\sum_i \hat{\varepsilon}_i = 0$ y que $\sum_i X_i \hat{\varepsilon}_i = 0$.

(2) Deduce que $\sum_i \hat{Y}_i \hat{\varepsilon}_i = 0$ (los ajustados son ortogonales a los residuos) y que el punto $(\bar{X}, \bar{Y})$ está en la recta.

**Solución**

(1) La primera ecuación normal es exactamente $\sum_i (Y_i - \hat{\beta}_0 - \hat{\beta}_1 X_i) = \sum_i \hat{\varepsilon}_i = 0$. La segunda es $\sum_i X_i (Y_i - \hat{\beta}_0 - \hat{\beta}_1 X_i) = \sum_i X_i \hat{\varepsilon}_i = 0$. Así, con intercepto, los residuos suman cero y están *no correlacionados* con $X$.
(2) Como $\hat{Y}_i = \hat{\beta}_0 + \hat{\beta}_1 X_i$,

$$\sum_i \hat{Y}_i \hat{\varepsilon}_i = \hat{\beta}_0 \underbrace{\sum_i \hat{\varepsilon}_i}_{0} + \hat{\beta}_1 \underbrace{\sum_i X_i \hat{\varepsilon}_i}_{0} = 0.$$

Y de $\sum_i \hat{\varepsilon}_i = 0$ se sigue $\bar{Y} = \hat{\beta}_0 + \hat{\beta}_1 \bar{X}$, o sea la recta pasa por $(\bar{X}, \bar{Y})$. (Estas ortogonalidades son la razón de que el "término cruzado" desaparezca en SST = SSM + SSR.)
*(Adaptado de MIT 14.310x, L17.)*

**Ejercicio 1.32** (Sesgo por variable omitida (OVB): fórmula y signo)**.** El modelo "largo" (correcto) es $Y = \beta_0 + \beta_1 X + \beta_2 Z + \varepsilon$, pero por no observar $Z$ (habilidad) se corre el modelo "corto" $Y = \alpha_0 + \alpha_1 X + u$. Sea $\delta$ la pendiente de regresar $Z$ sobre $X$ ($\delta = \widehat{\mathrm{Cov}}(X, Z)/\widehat{\mathrm{Var}}(X)$). Se puede mostrar que $\hat{\alpha}_1 \xrightarrow{p} \beta_1 + \beta_2 \, \delta$.

(1) Identifica el *sesgo por variable omitida* y su signo si $\beta_2 > 0$ (la habilidad sube el salario) y $\delta > 0$ ($X$ y $Z$ covarían positivamente).

(2) Con $\beta_1 = 3$ (retorno causal real a la educación), $\beta_2 = 2$ y $\delta = 1.5$, calcula el coeficiente que reportaría la regresión corta. ¿Sobrestima o subestima?

(3) ¿En qué caso el sesgo es *cero* aunque $Z$ importe ($\beta_2 \neq 0$)?

**Solución**

(1) El sesgo es sesgo $= \hat{\alpha}_1 - \beta_1 = \beta_2\,\delta$. Con $\beta_2 > 0$ y $\delta > 0$, el sesgo es **positivo**: la regresión corta atribuye a $X$ parte del efecto que es de $Z$.
(2) $\hat{\alpha}_1 \to \beta_1 + \beta_2\delta = 3 + 2(1.5) = 3 + 3 = \boxed{6}$. El retorno verdadero es 3, pero la regresión corta reporta 6: **sobrestima** (lo duplica). Quienes tienen más educación ganan más en parte por su mayor habilidad, no solo por estudiar.
(3) El sesgo $\beta_2\delta = 0$ si $\delta = 0$, es decir si $Z$ *no* se correlaciona con $X$ ($\widehat{\mathrm{Cov}}(X, Z) = 0$): entonces omitir $Z$ no sesga $\hat{\alpha}_1$ (aunque $Z$ afecte a $Y$). También si $\beta_2 = 0$ ($Z$ irrelevante).
*(Adaptado de MIT 14.310x, L17–L18; preludio a variables instrumentales.)*

**Ejercicio 1.33** (Invertir $\mathbf{X}^\top\mathbf{X}$ $(2 \times 2)$ y obtener $\hat{\boldsymbol{\beta}}$)**.** Para el modelo con intercepto $Y_i = \beta_0 + \beta_1 X_i + \varepsilon_i$, la matriz de diseño es $\mathbf{X}$ con primera columna de unos y segunda columna $X_i$. Con 4 datos $X = (1, 2, 3, 4)$, $Y = (3, 8, 11, 12)$:

(1) Calcula $\mathbf{X}^\top\mathbf{X}$ y $\mathbf{X}^\top\mathbf{Y}$.

(2) Invierte $\mathbf{X}^\top\mathbf{X}$ (matriz $2 \times 2$) y obtén $\hat{\boldsymbol{\beta}} = (\mathbf{X}^\top\mathbf{X})^{-1}\mathbf{X}^\top\mathbf{Y}$.

(3) Verifica que $\hat{\beta}_1$ coincide con la fórmula escalar $\sum(X - \bar{X})(Y - \bar{Y}) / \sum(X - \bar{X})^2$.

**Solución**

(1) Con $n = 4$: $\sum X = 10$, $\sum X^2 = 1 + 4 + 9 + 16 = 30$, $\sum Y = 3 + 8 + 11 + 12 = 34$, $\sum XY = 1 \cdot 3 + 2 \cdot 8 + 3 \cdot 11 + 4 \cdot 12 = 3 + 16 + 33 + 48 = 100$.

$$\mathbf{X}^\top\mathbf{X} = \begin{pmatrix} n & \sum X \\ \sum X & \sum X^2 \end{pmatrix} = \begin{pmatrix} 4 & 10 \\ 10 & 30 \end{pmatrix}, \qquad \mathbf{X}^\top\mathbf{Y} = \begin{pmatrix} \sum Y \\ \sum XY \end{pmatrix} = \begin{pmatrix} 34 \\ 100 \end{pmatrix}.$$

(2) Determinante $\det = 4 \cdot 30 - 10 \cdot 10 = 120 - 100 = 20$. Para una $2 \times 2$, $\left(\begin{smallmatrix} a & b \\ c & d \end{smallmatrix}\right)^{-1} = \frac{1}{\det}\left(\begin{smallmatrix} d & -b \\ -c & a \end{smallmatrix}\right)$:

$$(\mathbf{X}^\top\mathbf{X})^{-1} = \frac{1}{20}\begin{pmatrix} 30 & -10 \\ -10 & 4 \end{pmatrix} = \begin{pmatrix} 3/2 & -1/2 \\ -1/2 & 1/5 \end{pmatrix}.$$

$$\hat{\boldsymbol{\beta}} = (\mathbf{X}^\top\mathbf{X})^{-1}\mathbf{X}^\top\mathbf{Y} = \begin{pmatrix} 3/2 & -1/2 \\ -1/2 & 1/5 \end{pmatrix}\begin{pmatrix} 34 \\ 100 \end{pmatrix} = \begin{pmatrix} \frac{3}{2} \cdot 34 - \frac{1}{2} \cdot 100 \\ -\frac{1}{2} \cdot 34 + \frac{1}{5} \cdot 100 \end{pmatrix} = \begin{pmatrix} 51 - 50 \\ -17 + 20 \end{pmatrix} = \begin{pmatrix} 1 \\ 3 \end{pmatrix}.$$

Así $\hat{\beta}_0 = 1$, $\hat{\beta}_1 = 3$: $\hat{Y} = 1 + 3X$.
(3) Escalar: $\bar{X} = 2.5$, $\bar{Y} = 8.5$, $\sum(X - \bar{X})(Y - \bar{Y})$ con desvíos $X - \bar{X} = (-1.5, -0.5, 0.5, 1.5)$, $Y - \bar{Y} = (-5.5, -0.5, 2.5, 3.5)$: productos $8.25, 0.25, 1.25, 5.25$, suma $= 15$; $\sum(X - \bar{X})^2 = 2.25 + 0.25 + 0.25 + 2.25 = 5$. $\hat{\beta}_1 = 15/5 = 3$. ✓
*(Adaptado de MIT 14.310x, L18: notación matricial.)*

**Ejercicio 1.34** (Insesgadez de OLS en el modelo simple)**.** En el modelo $Y_i = \beta_0 + \beta_1 X_i + \varepsilon_i$ con los $X_i$ fijos (o condicionando en ellos) y $\mathbb{E}[\varepsilon_i] = 0$, demuestra que $\hat{\beta}_1$ es *insesgado*: $\mathbb{E}[\hat{\beta}_1] = \beta_1$.

(1) Escribe $\hat{\beta}_1 = \sum_i w_i Y_i$ con pesos $w_i = \dfrac{X_i - \bar{X}}{\sum_j (X_j - \bar{X})^2}$, y verifica $\sum_i w_i = 0$ y $\sum_i w_i X_i = 1$.

(2) Sustituye $Y_i = \beta_0 + \beta_1 X_i + \varepsilon_i$ y toma esperanza para concluir $\mathbb{E}[\hat{\beta}_1] = \beta_1$.

**Solución**

(1) Como $\hat{\beta}_1 = \dfrac{\sum_i (X_i - \bar{X})(Y_i - \bar{Y})}{\sum_j (X_j - \bar{X})^2} = \dfrac{\sum_i (X_i - \bar{X}) Y_i}{\sum_j (X_j - \bar{X})^2}$ (el $\bar{Y}$ se cae porque $\sum_i (X_i - \bar{X}) = 0$), tenemos $\hat{\beta}_1 = \sum_i w_i Y_i$ con $w_i = (X_i - \bar{X}) / \sum_j (X_j - \bar{X})^2$. Entonces $\sum_i w_i = \dfrac{\sum_i (X_i - \bar{X})}{\sum_j (X_j - \bar{X})^2} = 0$ y

$$\sum_i w_i X_i = \frac{\sum_i (X_i - \bar{X}) X_i}{\sum_j (X_j - \bar{X})^2} = \frac{\sum_i (X_i - \bar{X})^2}{\sum_j (X_j - \bar{X})^2} = 1$$

(usando $\sum_i (X_i - \bar{X}) X_i = \sum_i (X_i - \bar{X})^2$).
(2) Sustituyendo $Y_i = \beta_0 + \beta_1 X_i + \varepsilon_i$:

$$\hat{\beta}_1 = \sum_i w_i (\beta_0 + \beta_1 X_i + \varepsilon_i) = \beta_0 \underbrace{\sum_i w_i}_{0} + \beta_1 \underbrace{\sum_i w_i X_i}_{1} + \sum_i w_i \varepsilon_i = \beta_1 + \sum_i w_i \varepsilon_i.$$

Tomando esperanza (con $X$ fijos y $\mathbb{E}[\varepsilon_i] = 0$): $\mathbb{E}[\hat{\beta}_1] = \beta_1 + \sum_i w_i \, \mathbb{E}[\varepsilon_i] = \beta_1$. ✓ OLS es insesgado.
*(Adaptado de MIT 14.310x, L17; MIT 18.05, propiedades de estimadores lineales.)*

## Nivel 5 – Reto+

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

**Ejercicio 1.35** (Derivación matricial: $\hat{\boldsymbol{\beta}} = (\mathbf{X}^\top \mathbf{X})^{-1} \mathbf{X}^\top \mathbf{Y}$)**.** En notación matricial el modelo es $\mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}$ con $\mathbf{X}$ de $n \times (k{+}1)$ y rango columna completo. OLS minimiza $S(\boldsymbol{\beta}) = \|\mathbf{Y} - \mathbf{X}\boldsymbol{\beta}\|^2 = (\mathbf{Y} - \mathbf{X}\boldsymbol{\beta})^\top (\mathbf{Y} - \mathbf{X}\boldsymbol{\beta})$.

(1) Expande $S(\boldsymbol{\beta})$ y deriva respecto a $\boldsymbol{\beta}$ para obtener las *ecuaciones normales* $\mathbf{X}^\top \mathbf{X} \hat{\boldsymbol{\beta}} = \mathbf{X}^\top \mathbf{Y}$.

(2) Despeja $\hat{\boldsymbol{\beta}}$ y di qué supuesto garantiza que la inversa exista. ¿Por qué el mínimo es global (convexidad)?

**Solución**

(1) Expandiendo (usando $\mathbf{Y}^\top\mathbf{X}\boldsymbol{\beta}$ escalar, igual a su transpuesta $\boldsymbol{\beta}^\top\mathbf{X}^\top\mathbf{Y}$):

$$S(\boldsymbol{\beta}) = \mathbf{Y}^\top\mathbf{Y} - 2\,\boldsymbol{\beta}^\top\mathbf{X}^\top\mathbf{Y} + \boldsymbol{\beta}^\top\mathbf{X}^\top\mathbf{X}\boldsymbol{\beta}.$$

Derivando con las reglas vectoriales $\partial(\mathbf{a}^\top\boldsymbol{\beta})/\partial\boldsymbol{\beta} = \mathbf{a}$ y $\partial(\boldsymbol{\beta}^\top\mathbf{A}\boldsymbol{\beta})/\partial\boldsymbol{\beta} = 2\mathbf{A}\boldsymbol{\beta}$ (con $\mathbf{A} = \mathbf{X}^\top\mathbf{X}$ simétrica):

$$\frac{\partial S}{\partial \boldsymbol{\beta}} = -2\mathbf{X}^\top\mathbf{Y} + 2\mathbf{X}^\top\mathbf{X}\boldsymbol{\beta} = \mathbf{0} \;\Rightarrow\; \boxed{\mathbf{X}^\top\mathbf{X}\hat{\boldsymbol{\beta}} = \mathbf{X}^\top\mathbf{Y}}.$$

(2) Si $\mathbf{X}$ tiene **rango columna completo** $(k+1)$, entonces $\mathbf{X}^\top\mathbf{X}$ es invertible y

$$\boxed{\hat{\boldsymbol{\beta}} = (\mathbf{X}^\top\mathbf{X})^{-1}\mathbf{X}^\top\mathbf{Y}}.$$

El mínimo es *global* porque la Hessiana es $2\mathbf{X}^\top\mathbf{X} \succeq 0$ (semidefinida positiva; definida positiva si el rango es completo), así que $S$ es convexa.
*(Adaptado de MIT 14.310x, L18: matrix notation; MIT 18.05 / álgebra lineal.)*

**Ejercicio 1.36** (Insesgadez y varianza de $\hat{\boldsymbol{\beta}}$ en forma matricial)**.** Bajo el modelo clásico, con $\mathbf{X}$ fija (o condicionando en ella), $\mathbb{E}[\boldsymbol{\varepsilon}] = \mathbf{0}$ y $\mathrm{Var}(\boldsymbol{\varepsilon}) = \mathbb{E}[\boldsymbol{\varepsilon}\boldsymbol{\varepsilon}^\top] = \sigma^2\mathbf{I}_n$.

(1) Demuestra que $\mathbb{E}[\hat{\boldsymbol{\beta}}] = \boldsymbol{\beta}$ (insesgadez).

(2) Demuestra que $\mathrm{Var}(\hat{\boldsymbol{\beta}}) = \sigma^2(\mathbf{X}^\top\mathbf{X})^{-1}$.

(3) Aplica al dataset $X = (1,2,3,4)$, $Y = (3,8,11,12)$ (Reto): allí $(\mathbf{X}^\top\mathbf{X})^{-1} = \left(\begin{smallmatrix} 3/2 & -1/2 \\ -1/2 & 1/5 \end{smallmatrix}\right)$ y la SSR resultó 4 $(n = 4)$. Estima $\sigma^2$ con $s^2 = \mathrm{SSR}/(n-2)$ y da $\widehat{\mathrm{Var}}(\hat{\beta}_0)$, $\widehat{\mathrm{Var}}(\hat{\beta}_1)$ y $\widehat{\mathrm{Cov}}(\hat{\beta}_0, \hat{\beta}_1)$.

**Solución**

(1) Sustituyendo $\mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}$ en $\hat{\boldsymbol{\beta}} = (\mathbf{X}^\top\mathbf{X})^{-1}\mathbf{X}^\top\mathbf{Y}$:

$$\hat{\boldsymbol{\beta}} = (\mathbf{X}^\top\mathbf{X})^{-1}\mathbf{X}^\top(\mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}) = \boldsymbol{\beta} + (\mathbf{X}^\top\mathbf{X})^{-1}\mathbf{X}^\top\boldsymbol{\varepsilon}.$$

Tomando esperanza, $\mathbb{E}[\hat{\boldsymbol{\beta}}] = \boldsymbol{\beta} + (\mathbf{X}^\top\mathbf{X})^{-1}\mathbf{X}^\top\mathbb{E}[\boldsymbol{\varepsilon}] = \boldsymbol{\beta}$. ✓
(2) Sea $\mathbf{A} = (\mathbf{X}^\top\mathbf{X})^{-1}\mathbf{X}^\top$, de modo que $\hat{\boldsymbol{\beta}} - \boldsymbol{\beta} = \mathbf{A}\boldsymbol{\varepsilon}$. Entonces

$$\mathrm{Var}(\hat{\boldsymbol{\beta}}) = \mathbf{A}\;\mathrm{Var}(\boldsymbol{\varepsilon})\,\mathbf{A}^\top = \mathbf{A}\,(\sigma^2\mathbf{I}_n)\,\mathbf{A}^\top = \sigma^2\mathbf{A}\mathbf{A}^\top.$$

Con $\mathbf{A}\mathbf{A}^\top = (\mathbf{X}^\top\mathbf{X})^{-1}\mathbf{X}^\top\mathbf{X}(\mathbf{X}^\top\mathbf{X})^{-1} = (\mathbf{X}^\top\mathbf{X})^{-1}$ (la inversa es simétrica), queda $\boxed{\mathrm{Var}(\hat{\boldsymbol{\beta}}) = \sigma^2(\mathbf{X}^\top\mathbf{X})^{-1}}$.

(3) $s^2 = \dfrac{\text{SSR}}{n-2} = \dfrac{4}{4-2} = 2$. Multiplicando la inversa por $s^2 = 2$:

$$\widehat{\text{Var}}(\hat{\boldsymbol{\beta}}) = 2\begin{pmatrix} 3/2 & -1/2 \\ -1/2 & 1/5 \end{pmatrix} = \begin{pmatrix} 3 & -1 \\ -1 & 2/5 \end{pmatrix}.$$

Así $\widehat{\text{Var}}(\hat{\beta}_0) = 3$, $\widehat{\text{Var}}(\hat{\beta}_1) = 2/5 = 0.4$ y $\widehat{\text{Cov}}(\hat{\beta}_0, \hat{\beta}_1) = -1 < 0$ (consistente con el comentario de L17: si $\bar{X} > 0$, la covarianza entre los estimadores es negativa). Errores estándar: $\text{ee}(\hat{\beta}_0) = \sqrt{3} \approx 1.73$, $\text{ee}(\hat{\beta}_1) = \sqrt{0.4} \approx 0.63$.
*(Adaptado de MIT 14.310x, L17–L18; MIT 18.05, distribución de estimadores.)*

**Ejercicio 1.37** (Frisch–Waugh–Lovell: "partialling out")**.** El teorema de Frisch–Waugh–Lovell (FWL) dice que, en $Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \varepsilon$, el coeficiente $\hat{\beta}_1$ de la regresión múltiple se puede obtener así: (i) regresar $X_1$ sobre el resto $(1, X_2)$ y guardar los residuos $\tilde{X}_1$; (ii) regresar $Y$ sobre el resto y guardar $\tilde{Y}$ (o usar $Y$ directamente); (iii) regresar $\tilde{Y}$ sobre $\tilde{X}_1$.

(1) Escribe la fórmula resultante $\hat{\beta}_1 = \dfrac{\sum_i \tilde{X}_{1i} Y_i}{\sum_i \tilde{X}_{1i}^2}$ y explica en palabras qué significa "*partialling out*" (parcializar) $X_2$.

(2) Usa FWL para explicar *por qué* en una regresión múltiple el coeficiente de $X_1$ es su efecto "manteniendo constante $X_2$", y cómo se conecta con el sesgo por variable omitida del Nivel 4.

**Solución**

(1) $\tilde{X}_{1i}$ es la parte de $X_1$ que *no* explica $X_2$ (sus residuos al regresar $X_1$ sobre 1 y $X_2$). Regresar $Y$ sobre $\tilde{X}_1$ da

$$\hat{\beta}_1 = \frac{\sum_i \tilde{X}_{1i} Y_i}{\sum_i \tilde{X}_{1i}^2} = \frac{\widehat{\text{Cov}}(\tilde{X}_1, Y)}{\widehat{\text{Var}}(\tilde{X}_1)}.$$

(Como $\tilde{X}_1$ es ortogonal a 1 y a $X_2$, usar $Y$ o $\tilde{Y}$ en el numerador da lo mismo.) "*Partialling out* $X_2$" significa: quitar de $X_1$ la variación que comparte con $X_2$, y relacionar con $Y$ solo la variación *limpia* de $X_1$.
(2) Por eso $\hat{\beta}_1$ mide el efecto de $X_1$ *usando solo* la variación de $X_1$ no explicada por $X_2$, es decir comparando observaciones *con el mismo* $X_2$: ese es el sentido exacto de "manteniendo $X_2$ constante". Conexión con OVB: si *omitimos* $X_2$, no parcializamos nada y la regresión simple usa toda la variación de $X_1$ (incluida la compartida con $X_2$), por lo que su coeficiente recoge $\beta_1 + \beta_2\delta$ (sesgo $\beta_2\delta$). FWL muestra formalmente que controlar por $X_2$ = parcializarlo = eliminar ese sesgo.
*Verificación numérica (simulada en R, $n = 50$, modelo $Y = 1 + 2X_1 + 3X_2 +$ ruido pequeño):* la regresión múltiple da $\hat{\beta}_1 \approx 2.005$; el procedimiento FWL (regresar los residuos $\tilde{Y}$ sobre $\tilde{X}_1$) da el *mismo* $\hat{\beta}_1 \approx 2.005$. Coinciden, como predice el teorema.
*(Adaptado de MIT 14.310x, L18: multivariate OLS; Frisch–Waugh–Lovell.)*

**Ejercicio 1.38** (Multicolinealidad perfecta $\Rightarrow$ $\mathbf{X}^\top\mathbf{X}$ singular)**.** Suponga que un regresor es combinación lineal exacta de otros, p. ej. se incluyen $X_1$ = estatura en cm y $X_2$ = estatura en metros ($X_2 = 0.01\,X_1$), además del intercepto.

(1) Argumenta por qué la matriz $\mathbf{X}$ *no* tiene rango columna completo y, por tanto, $\mathbf{X}^\top\mathbf{X}$ es **singular** (no invertible), de modo que $\hat{\boldsymbol{\beta}}$ no es único.

(2) Da una intuición de por qué el problema es de *identificación*: ¿qué familia de coeficientes $(\beta_1, \beta_2)$ produce *exactamente* el mismo ajuste? Relaciónalo con la "trampa de las dummies" (Nivel 3) y con el supuesto de identificación de L18 ($\mathbf{X}$ de rango completo).

**Solución**

(1) Si $X_2 = 0.01\,X_1$, la columna de $X_2$ es un múltiplo de la de $X_1$: las columnas son **linealmente dependientes**, así que $\text{rango}(\mathbf{X}) < k+1$. Para una $\mathbf{X}$ sin rango completo, $\mathbf{X}^\top\mathbf{X}$ también tiene rango deficiente y $\det(\mathbf{X}^\top\mathbf{X}) = 0$: es **singular**, no admite inversa, y la fórmula $\hat{\boldsymbol{\beta}} = (\mathbf{X}^\top\mathbf{X})^{-1}\mathbf{X}^\top\mathbf{Y}$ no está definida (hay infinitas soluciones de las ecuaciones normales).
(2) El ajuste depende de $\beta_1 X_1 + \beta_2 X_2 = \beta_1 X_1 + \beta_2(0.01 X_1) = (\beta_1 + 0.01\beta_2)X_1$. Cualquier par $(\beta_1, \beta_2)$ con la *misma* suma $\beta_1 + 0.01\beta_2$ da *idéntico* $\hat{Y}$ y la misma SSR; los datos no pueden distinguir entre ellos. Es un problema de **identificación**: no se pueden separar los efectos de dos variables que se mueven juntas perfectamente. Es la misma raíz que la "trampa de las dummies" ($\sum$ dummies $= 1 =$ intercepto) y por eso L18 exige el supuesto de identificación: $n > k+1$ y $\mathbf{X}$ de *rango columna completo* (regresores linealmente independientes, $\mathbf{X}^\top\mathbf{X}$ invertible). La solución práctica: eliminar el regresor redundante (usar solo cm *o* solo metros).
*(Adaptado de MIT 14.310x, L18: identification y full column rank.)*

## Fuentes y atribución



**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`.