

Causalidad, Experimentos Aleatorizados y Regresión No Paramétrica

Clase 8 — 14.310x: Data Analysis for Social Scientists

B.Sc. Cristian Lazo Quispe

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

Programa SDS (Statistics and Data Science, MITx) · BREIT (Brescia Institute of Technology), iniciativa de Aporta —Grupo Breca—
en alianza con MIT IDSS.

Basado en MIT 14.310x (CC BY-NC-SA 4.0).

✉ clazoq@uni.pe · [in linkedin.com/in/cristian-lazo-quispe](https://www.linkedin.com/in/cristian-lazo-quispe) · github.com/CristianLazoQuispe

¡Arrancamos con un Kahoot! (repaso Clases 5–8)

Únete ahora

Entra a **kahoot.it** con el **PIN** en pantalla. Son **32 preguntas** rápidas de repaso: distribuciones y TLC, estimación e intervalos, pruebas de hipótesis y causalidad.

Reglas del juego

- Responde rápido: la velocidad da más puntos.
- Cuidado con las **trampas** (el error clásico).
- Al final comentamos 2–3 preguntas clave.



¿Qué veremos hoy?

- 1 Causalidad y resultados potenciales
- 2 Sesgo de selección y confusión
- 3 Experimentos aleatorizados (RCT)
- 4 Analizar un RCT
- 5 Regresión no paramétrica
- 6 Simulación en R
- 7 Manos a la obra

Contenido

- 1 Causalidad y resultados potenciales
- 2 Sesgo de selección y confusión
- 3 Experimentos aleatorizados (RCT)
- 4 Analizar un RCT
- 5 Regresión no paramétrica
- 6 Simulación en R
- 7 Manos a la obra

Repaso visual: el contrafactual en 3 minutos (video)

Mira el video para calentar

Un recordatorio en imágenes de la idea central de hoy: el **contrafactual (counterfactual)** — qué le habría pasado al *mismo* paciente si *no* hubiera recibido el tratamiento.

Enlace

▶ youtube.com/watch?v=QJI0c1Z3hyA

(empieza en el minuto 3:06).

Pregunta para el chat / breakout

¿Por qué es *imposible* observar los dos mundos —con y sin tratamiento— para la misma persona al mismo tiempo?



Caso observado vs. caso hipotético para el mismo paciente X.

¿Qué significa que “X causa Y”?

La idea central (Rubin, resultados potenciales)

Decir que una **causa (cause)** produce un **efecto (effect)** es comparar lo que pasó con lo que *habría pasado* en un mundo **contrafactual (counterfactual)**: el mismo individuo, al mismo tiempo, con y sin el tratamiento.

Definición 1 (Resultados potenciales y efecto causal)

Para cada unidad i con tratamiento $W_i \in \{0, 1\}$ definimos dos resultados:

$$Y_i(1) \text{ (con tratamiento),} \quad Y_i(0) \text{ (sin tratamiento, control).}$$

El **efecto causal individual** es $\tau_i = Y_i(1) - Y_i(0)$.

El problema fundamental de la inferencia causal (Holland, 1986)

Solo observamos *uno* de los dos: $Y_i^{\text{obs}} = Y_i(W_i)$. El otro (Y_i^{miss}) nunca se ve. ¡Es un problema de **datos faltantes!**

Ejemplo clásico: aspirina y dolor de cabeza (potential outcomes)

Cuatro historias causales posibles

Para *una* persona con dolor de cabeza, comparar $Y(\text{aspirina})$ con $Y(\text{no aspirina})$ da **cuatro** escenarios: mejora solo con aspirina, sin efecto (con o sin dolor en ambos), o mejora solo sin aspirina.

La casilla que siempre falta

En la vida real solo llenamos *una* columna: si tomaste aspirina ves $Y(\text{aspirina})$; la otra queda **en blanco** (el contrafactual). Por eso el efecto causal individual no se *observa*, se *infiere*.

1. Headache gone only with aspirin:
 $Y(\text{Aspirin}) = \text{No Headache}, Y(\text{No Aspirin}) = \text{Headache}$
2. No effect of aspirin, with a headache in both cases:
 $Y(\text{Aspirin}) = \text{Headache}, Y(\text{No Aspirin}) = \text{Headache}$
3. No effect of aspirin, with the headache gone in both cases:
 $Y(\text{Aspirin}) = \text{No Headache}, Y(\text{No Aspirin}) = \text{No Headache}$
4. Headache gone only without aspirin:
 $Y(\text{Aspirin}) = \text{Headache}, Y(\text{No Aspirin}) = \text{No Headache}$

Unit	Potential Outcomes		Causal Effect
	$Y(\text{Aspirin})$	$Y(\text{No Aspirin})$	
You	No Headache	Headache	Improvement due to Aspirin

Resultados potenciales y efecto causal — ejemplo tomado de tu material de clase.

Correlación $\neq$ causalidad: una correlación espuria

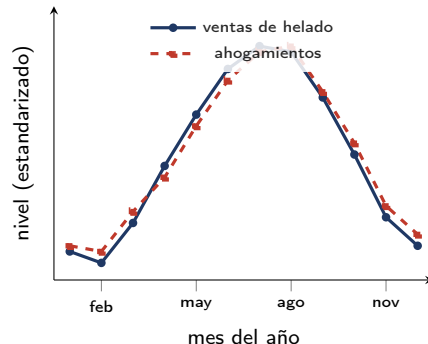
El verano lo explica todo

En una ciudad, las ventas de helado y los ahogamientos suben y bajan *juntos* a lo largo del año. ¿El helado causa ahogamientos?

- No: una **tercera variable**, el *calor del verano*, mueve a ambos a la vez.
- La correlación es **real** pero el vínculo causal es **falso** (*spurious*).

Lección

Ver que dos cosas se mueven juntas no dice *qué empuja a qué*, ni si hay un **confusor** (**confounder**) detrás.



Ambas series suben en verano: alta correlación, pero el motor común es el calor, no el helado.

Demo resuelto: efecto individual y ATE de una tabla

Problema

En un mundo “de juguete” *conocemos los dos* resultados potenciales de $N = 4$ personas (analgésico para el dolor, escala 0–10, menos es mejor). Se pide: el efecto individual τ_i y el efecto promedio (ATE).

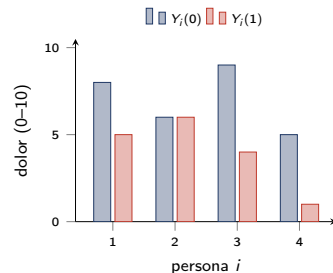
i	$Y_i(0)$	$Y_i(1)$	$\tau_i = Y_i(1) - Y_i(0)$
1	8	5	-3
2	6	6	0
3	9	4	-5
4	5	1	-4

Solución

El ATE (average treatment effect) es el promedio de los τ_i :

$$\text{ATE} = \frac{1}{N} \sum_{i=1}^N (Y_i(1) - Y_i(0)) = \frac{-3 + 0 - 5 - 4}{4} = \boxed{-3}.$$

En promedio el analgésico *baja* el dolor en 3 puntos. Persona 2: sin efecto.



La barra roja casi siempre es más baja:
el tratamiento reduce el dolor (excepto $i = 2$).

Práctica A — Causalidad: efecto individual y ATE

(¡inténtalo!)

Problema

Conocemos *ambos* resultados potenciales (dolor 0–10, menos es mejor) de 3 personas:

i	$Y_i(0)$ (sin analgésico)	$Y_i(1)$ (con analgésico)
1	7	4
2	5	5
3	9	3

Tu turno (chat / breakout)

Calcula el **efecto individual** $\tau_i = Y_i(1) - Y_i(0)$ de cada persona y el **ATE** (promedio). ¿Alguien no responde al analgésico?

Práctica A — Solución paso a paso

Solución

Paso 1 — efecto individual $\tau_i = Y_i(1) - Y_i(0)$:

$$\tau_1 = 4 - 7 = -3, \quad \tau_2 = 5 - 5 = 0, \quad \tau_3 = 3 - 9 = -6.$$

Paso 2 — ATE (promedio de los τ_i):

$$\text{ATE} = \frac{-3 + 0 - 6}{3} = \frac{-9}{3} = \boxed{-3}.$$

Interpretación: en promedio el analgésico *baja* el dolor en 3 puntos. La persona 2 **no responde** ($\tau_2 = 0$): mismo dolor con y sin pastilla.

Práctica B — Causalidad: lo observado y lo que falta

Problema

Usa la misma tabla. La **persona 1** **sí** tomó el analgésico ($W_1 = 1$).

- ¿Qué valor *observamos*, Y_1^{obs} ?
- ¿Cuál es el contrafactual que *falta*, Y_1^{miss} ?
- ¿Por qué, en la vida real, no podríamos calcular τ_1 ?

Tu turno

Piénsalo con la notación $Y^{\text{obs}} = Y(W)$.

Práctica B — Solución paso a paso

Solución

Paso 1 — qué se observa. Como $W_1 = 1$, observamos el potencial *con* tratamiento:

$$Y_1^{\text{obs}} = Y_1(1) = 4.$$

Paso 2 — qué falta. El contrafactual es “cómo le habría ido *sin* pastilla”:

$$Y_1^{\text{miss}} = Y_1(0) = 7 \quad (\text{nunca se observa en la realidad}).$$

Paso 3 — por qué no vemos τ_1 . El efecto individual $\tau_1 = Y_1(1) - Y_1(0)$ necesita **los dos**; solo tenemos uno (**problema fundamental de la inferencia causal**). Aquí conocemos ambos únicamente porque es un “mundo de juguete”.

Contenido

- 1 Causalidad y resultados potenciales
- 2 Sesgo de selección y confusión
- 3 Experimentos aleatorizados (RCT)
- 4 Analizar un RCT
- 5 Regresión no paramétrica
- 6 Simulación en R
- 7 Manos a la obra

Cómo leer la notación: W , $Y(1)$, $Y(0)$ y la barra “|”

El diccionario (ejemplo: pastilla para el dolor)

- $W = 1$: la persona **sí** recibió el tratamiento (*treated*).
- $W = 0$: la persona **no** lo recibió (control).
- $Y(1)$: resultado *si recibe* el tratamiento.
- $Y(0)$: resultado *si NO recibe* el tratamiento (nivel base).
- $\mathbb{E}[\cdot]$: **promedio** (esperanza) sobre un grupo.
- La barra “|” se lee “**entre los que...**” (o “**dado que**”).

Léelo en voz alta, de derecha a izquierda

$\mathbb{E}[Y(0) | W = 1]$ se lee:

“el dolor **sin** pastilla ($Y(0)$), en promedio, **entre los que sí** tomaron la pastilla ($W = 1$)”.

Observado vs. contrafactual

- $\mathbb{E}[Y(0) | W = 0]$: sin pastilla, entre los *controles* $\Rightarrow$ **se observa** (ellos no tomaron nada).
- $\mathbb{E}[Y(0) | W = 1]$: sin pastilla, entre los *tratados* $\Rightarrow$ **NO se observa** (ellos sí tomaron). Es el **contrafactual de los tratados**.

Descomposición paso a paso: de lo observado al efecto

Punto de partida: lo que sí medimos

La diferencia de promedios *observados* entre tratados y controles:

$$D = \mathbb{E}[Y \mid W = 1] - \mathbb{E}[Y \mid W = 0].$$

Los tres pasos

Paso 1. Cada grupo solo muestra el potencial que le tocó: el tratado muestra $Y(1)$, el control muestra $Y(0)$:

$$D = \mathbb{E}[Y(1) \mid W = 1] - \mathbb{E}[Y(0) \mid W = 0].$$

Paso 2 (el truco). Sumamos y restamos $\mathbb{E}[Y(0) \mid W = 1]$ (sumar 0 no cambia nada):

$$D = \underbrace{\mathbb{E}[Y(1) \mid W = 1] - \mathbb{E}[Y(0) \mid W = 1]}_{\text{comparo dentro de los tratados}} + \underbrace{\mathbb{E}[Y(0) \mid W = 1] - \mathbb{E}[Y(0) \mid W = 0]}_{\text{comparo el nivel base } Y(0)}.$$

Paso 3. Le ponemos nombre a cada corchete:

$$D = \underbrace{\mathbb{E}[Y(1) - Y(0) \mid W = 1]}_{\text{ATT: efecto en los tratados}} + \underbrace{\mathbb{E}[Y(0) \mid W = 1] - \mathbb{E}[Y(0) \mid W = 0]}_{\text{sesgo de selección}}.$$

Por qué el truco funciona

El 1.^{er} corchete resta dos potenciales del *mismo* grupo (tratados) $\Rightarrow$ es un **efecto causal**. El 2.^{do} compara el nivel base $Y(0)$ de *grupos distintos* $\Rightarrow$ mide qué tan distintos eran *antes* de tratar.

¿Por qué la diferencia observada $\neq$ ATE?

Descomposición de la diferencia de medias

Sin aleatorizar, comparamos a los tratados con los no tratados. Esa diferencia se parte en dos:

$$\underbrace{\mathbb{E}[Y \mid W = 1] - \mathbb{E}[Y \mid W = 0]}_{\text{diferencia observada}} = \underbrace{\mathbb{E}[Y(1) - Y(0) \mid W = 1]}_{\text{efecto sobre los tratados (ATT)}} + \underbrace{\mathbb{E}[Y(0) \mid W = 1] - \mathbb{E}[Y(0) \mid W = 0]}_{\text{sesgo de selección}}.$$

¿De dónde sale el sesgo de selección (selection bias)?

De que tratados y controles *ya eran distintos* antes del tratamiento (en $Y(0)$):

- Quien toma una pastilla para el dolor suele tener un dolor *peor* de base.
- Quien va a la universidad tiene atributos que difieren de quien no va.

¿Cuándo desaparece el sesgo?

Cuando $\mathbb{E}[Y(0) \mid W = 1] = \mathbb{E}[Y(0) \mid W = 0]$, es decir cuando el grupo no influye en el resultado base. La **aleatorización** garantiza justo eso.

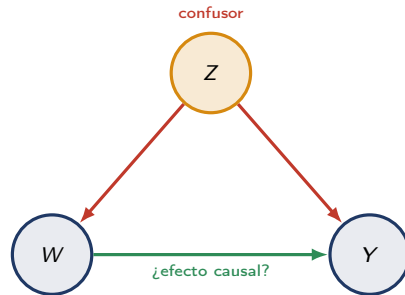
La anatomía de un confusor

Un confusor abre una “puerta trasera”

Un **confusor (confounder)** Z afecta tanto al tratamiento W como al resultado Y . Entonces W e Y se mueven juntos *sin* que W cause Y .

- Flecha causal real: línea **verde** (puede ser 0).
- Camino *espurio*: las dos flechas **rojas** vía Z .

Controlar/balancear Z “cierra” la puerta trasera.



Z (p. ej. “gravedad del caso” o “motivación”) empuja a la vez a W y a Y , creando correlación sin causalidad.

Demo resuelto: confusión numérica (Simpson en miniatura)

Problema

Un “tratamiento” (programa de tutoría) se ofrece, pero los alumnos *motivados* lo eligen más (confusor). En verdad $\tau_i = +1$ punto para todos. Promedios *observados* (nota 0–20):

	$\bar{Y}$ tratados	$\bar{Y}$ control
nivel base $Y(0)$	16	12
efecto real τ	+1	—
nota observada	17	12

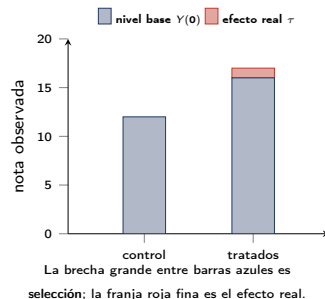
Se pide: la diferencia observada, el ATE real y el sesgo.

Solución

Diferencia observada = $17 - 12 = 5$. Efecto real ATE = +1.

$$\text{sesgo} = \underbrace{16}_{\mathbb{E}[Y(0)|W=1]} - \underbrace{12}_{\mathbb{E}[Y(0)|W=0]} = +4.$$

$5 = 1 + 4$: ¡el 80 % de la “mejora” es **selección**, no efecto!



Práctica A — Sesgo de selección

(¡inténtalo!)

Problema

Un programa de tutoría se ofrece *sin* aleatorizar: los alumnos **motivados** se apuntan más. Notas 0–20:

- Nota media *observada*: tratados $\bar{Y}_t = 17$, control $\bar{Y}_c = 12$.
- Nivel base $Y(0)$: tratados = 15, control = 12. Efecto real $ATT = +2$.

Tu turno

Descompón la **diferencia observada** en efecto + **sesgo de selección**. ¿Qué fracción de la “mejora” es sesgo?

Práctica A — Solución paso a paso

Solución

Paso 1 — diferencia observada:

$$\bar{Y}_t - \bar{Y}_c = 17 - 12 = 5.$$

Paso 2 — sesgo de selección (diferencia en el nivel base $Y(0)$):

$$\text{sesgo} = \mathbb{E}[Y(0) \mid W=1] - \mathbb{E}[Y(0) \mid W=0] = 15 - 12 = 3.$$

Chequeo: nota tratados = base 15 + ATT 2 = 17. ✓

$$\underbrace{5}_{\text{observada}} = \underbrace{2}_{\text{ATT}} + \underbrace{3}_{\text{sesgo}}.$$

Interpretación: $3/5 = 60\%$ de la “mejora” es **selección**, no efecto.

Práctica B — Caza el confusor

Problema

Un titular dice: *“Quienes van al gimnasio están más sanos”*. ¿Por qué esto **no prueba** que el gimnasio *cause* salud?

Tu turno

Nombra un **confusor** Z y explica cómo abre la “puerta trasera” entre ir al gimnasio (W) y la salud (Y).

Práctica B — Solución paso a paso

Solución

Paso 1 — un confusor plausible: Z = hábitos/motivación (o el ingreso).

Paso 2 — cómo confunde: Z afecta a *ambos*:

- $Z \rightarrow W$: los más motivados *van más* al gimnasio.
- $Z \rightarrow Y$: los más motivados también *comen y duermen mejor* $\Rightarrow$ más salud.

Paso 3 — la puerta trasera: gimnasio y salud suben juntos por el camino espurio $W \leftarrow Z \rightarrow Y$, *sin* que el gimnasio sea la causa.

Solución: aleatorizar (o balancear/controlar Z) cierra esa puerta.

Contenido

- 1 Causalidad y resultados potenciales
- 2 Sesgo de selección y confusión
- 3 Experimentos aleatorizados (RCT)**
- 4 Analizar un RCT
- 5 Regresión no paramétrica
- 6 Simulación en R
- 7 Manos a la obra

Repaso visual: ¿qué es un ensayo aleatorizado? (video)

Video (Cancer Research UK, ~ 2 min)

Cómo y por qué se **aleatoriza (randomise)** a los pacientes en dos grupos —A y B— para probar un tratamiento sin sesgo.

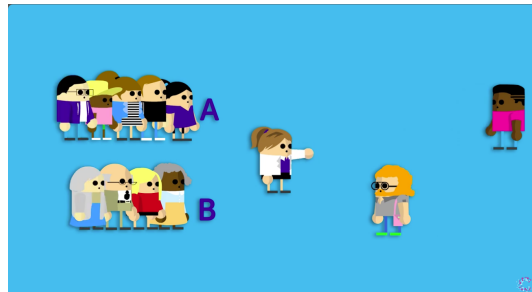
Enlace

▶ youtube.com/watch?v=fkOCYov1p-o

"What is a randomised trial?" (subtítulos en inglés).

Pregunta para el chat / breakout

Si el médico eligiera a *dedo* quién va al grupo A y quién al B, ¿qué problema aparecería? (pista: sesgo de selección).



La lotería reparte a las personas en los grupos A y B.

La aleatorización resuelve el problema de selección

Experimento completamente aleatorizado (completely randomized)

Asignamos N_t unidades al **tratamiento** y N_c al **control** *al azar*. Entonces la asignación W **no depende** de los resultados potenciales: $W \perp\!\!\!\perp (Y(0), Y(1))$.

Consecuencia clave

El sesgo de selección se *anula*: $\mathbb{E}[Y(0) \mid W = 1] = \mathbb{E}[Y(0) \mid W = 0]$, y por tanto

$$\mathbb{E}[Y \mid W = 1] - \mathbb{E}[Y \mid W = 0] = \mathbb{E}[Y(1) - Y(0)] = \text{ATE}.$$

¡La simple **diferencia de medias** estima el efecto causal promedio!

A/B testing

Un test A/B (web, app, marketing) es un RCT: la mitad de usuarios ve la versión *A* (control) y la otra mitad la *B* (tratamiento), asignados al azar. Comparar medias da el efecto causal del cambio.

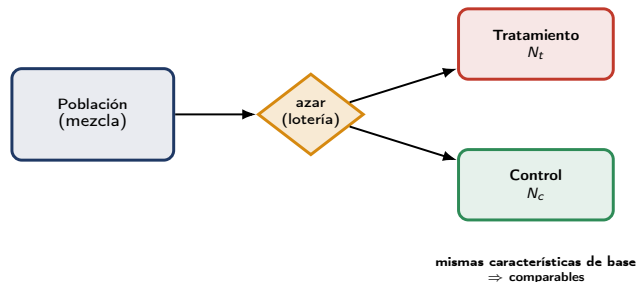
Aleatorizar $\Rightarrow$ grupos comparables

La intuición visual

Antes de tratar, la lotería reparte por igual a motivados y no motivados, jóvenes y mayores. . . Los dos grupos quedan **balanceados** “en promedio”.

- Mismas características de base.
- La *única* diferencia sistemática es el tratamiento.

Por eso la diferencia de medias se vuelve *causal*.



Demo resuelto: estimar el efecto en un RCT

Problema

Un RCT prueba un fertilizante. Rendimiento (qq/ha) en parcelas asignadas al azar:

Tratamiento ($N_t = 5$)	22, 25, 24, 27, 22
Control ($N_c = 5$)	19, 21, 18, 20, 22

Se pide: el efecto estimado $\hat{\tau} = \bar{Y}_t - \bar{Y}_c$.

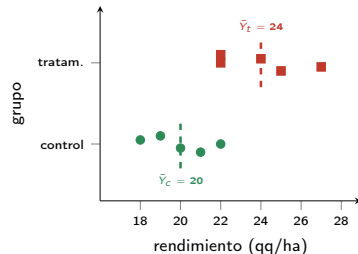
Solución

$$\bar{Y}_t = \frac{22 + 25 + 24 + 27 + 22}{5} = \frac{120}{5} = 24,$$

$$\bar{Y}_c = \frac{19 + 21 + 18 + 20 + 22}{5} = \frac{100}{5} = 20.$$

$$\hat{\tau} = \bar{Y}_t - \bar{Y}_c = 24 - 20 = +4 \text{ qq/ha.}$$

Como hubo aleatorización, interpretamos +4 como efecto **causal** promedio.



Las líneas punteadas son las medias;
su separación es $\hat{\tau} = +4$.

Práctica A — RCT: estima el efecto

(¡inténtalo!)

Problema

Un RCT prueba un fertilizante (rendimiento en qq/ha), parcelas asignadas al azar:

Tratamiento ($n = 4$)	23, 26, 25, 26
Control ($n = 4$)	20, 19, 22, 19

Tu turno

Estima el efecto $\hat{\tau} = \bar{Y}_t - \bar{Y}_c$. ¿Por qué lo llamamos *causal*?

Práctica A — Solución paso a paso

Solución

Paso 1 — media del tratamiento:

$$\bar{Y}_t = \frac{23 + 26 + 25 + 26}{4} = \frac{100}{4} = 25.$$

Paso 2 — media del control:

$$\bar{Y}_c = \frac{20 + 19 + 22 + 19}{4} = \frac{80}{4} = 20.$$

Paso 3 — efecto estimado:

$$\hat{\tau} = \bar{Y}_t - \bar{Y}_c = 25 - 20 = +5 \text{ qq/ha.}$$

Como hubo **aleatorización**, los grupos son comparables y +5 es el efecto **causal** promedio (no solo una asociación).

Práctica B — ¿Por qué la aleatorización lo vuelve causal?

Problema

En ese mismo RCT, la simple diferencia de medias *sí* estima el efecto causal. En cambio, con datos **observacionales** (sin aleatorizar) no. **¿Por qué?**

Tu turno

Usa la idea de **sesgo de selección** y de $W \perp\!\!\!\perp (Y(0), Y(1))$.

Práctica B — Solución paso a paso

Solución

Paso 1 — qué logra el azar. Al asignar W al azar, $W \perp\!\!\!\perp (Y(0), Y(1))$: los grupos son *iguales en promedio* en su nivel base.

$$\mathbb{E}[Y(0) \mid W=1] = \mathbb{E}[Y(0) \mid W=0] \Rightarrow \text{sesgo} = 0.$$

Paso 2 — consecuencia. Sin sesgo, la diferencia observada es el ATE:

$$\mathbb{E}[Y \mid W=1] - \mathbb{E}[Y \mid W=0] = \mathbb{E}[Y(1) - Y(0)] = \text{ATE}.$$

Paso 3 — el caso observacional. Ahí los grupos difieren *desde antes* (sesgo $\neq 0$), y la diferencia **mezcla** efecto + selección $\Rightarrow$ no es causal.

Contenido

- 1 Causalidad y resultados potenciales
- 2 Sesgo de selección y confusión
- 3 Experimentos aleatorizados (RCT)
- 4 Analizar un RCT**
- 5 Regresión no paramétrica
- 6 Simulación en R
- 7 Manos a la obra

Efecto, error estándar e intervalo de confianza

Estimador del efecto y su variabilidad (conecta con la Clase 6)

El estimador del ATE es la diferencia de medias, y es **insesgado**:

$$\hat{\tau} = \bar{Y}_t - \bar{Y}_c.$$

Como los grupos son independientes, las varianzas se *suman*:

$$\widehat{\text{Var}}(\hat{\tau}) = \frac{s_t^2}{N_t} + \frac{s_c^2}{N_c}, \quad \text{SE}(\hat{\tau}) = \sqrt{\widehat{\text{Var}}(\hat{\tau})}.$$

Intervalo de confianza y prueba

$$\text{IC}_{1-\alpha} = \hat{\tau} \pm t_{\text{crit}} \cdot \text{SE}(\hat{\tau}), \quad t = \frac{\hat{\tau}}{\text{SE}(\hat{\tau})}.$$

Probamos $H_0 : \text{ATE} = 0$ vs. $H_1 : \text{ATE} \neq 0$. Con N grande, $t_{\text{crit}} \approx 1.96$ al 95 % (aprox. normal); con N chico, usamos la t con $N_t + N_c - 2$ gl.

Demo resuelto: análisis completo de un RCT

Problema

Un programa de transferencias mide el gasto en alimentos (soles/semana). $N_t = N_c = 100$. Resultados: $\bar{Y}_t = 85$, $s_t = 20$; $\bar{Y}_c = 78$, $s_c = 18$. **Se pide:** $\hat{\tau}$, su SE, IC al 95 % y la prueba ($\alpha = 0.05$).

Solución

1) **Efecto:** $\hat{\tau} = 85 - 78 = +7$ soles.

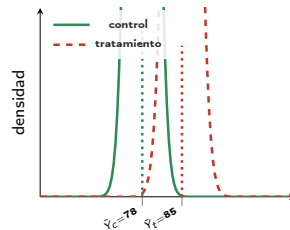
2) **SE:**

$$SE = \sqrt{\frac{20^2}{100} + \frac{18^2}{100}} = \sqrt{4 + 3.24} = \sqrt{7.24} \approx 2.69.$$

3) **IC 95 %** (N grande, $z = 1.96$):

$$7 \pm 1.96(2.69) = 7 \pm 5.27 = [1.73, 12.27].$$

4) **Prueba:** $t = 7/2.69 \approx 2.60 > 1.96 \Rightarrow$ **rechazo** H_0 ($p \approx 0.009$). El programa *aumentó* el gasto en alimentos.



gasto (soles/semana)

Distribución de $\bar{Y}$ en cada grupo.

Apenas se solapan: la diferencia es *significativa*.

Práctica A — Error estándar e IC del efecto

(¡inténtalo!)

Problema

Un RCT mide el gasto en alimentos (soles/semana), $N_t = N_c = 64$:

$$\bar{Y}_t = 90, s_t = 16; \quad \bar{Y}_c = 84, s_c = 12.$$

Tu turno

Calcula $\hat{\tau}$, su **error estándar** y el **IC al 95 %** ($z = 1.96$).

$$SE(\hat{\tau}) = \sqrt{\frac{s_t^2}{N_t} + \frac{s_c^2}{N_c}}.$$

Práctica A — Solución paso a paso

Solución

Paso 1 — efecto: $\hat{\tau} = \bar{Y}_t - \bar{Y}_c = 90 - 84 = 6$.

Paso 2 — error estándar:

$$SE = \sqrt{\frac{16^2}{64} + \frac{12^2}{64}} = \sqrt{\frac{256}{64} + \frac{144}{64}} = \sqrt{4 + 2.25} = \sqrt{6.25} = 2.5.$$

Paso 3 — IC 95 %:

$$\hat{\tau} \pm z \cdot SE = 6 \pm 1.96(2.5) = 6 \pm 4.9 = [1.1, 10.9].$$

El intervalo **no contiene el 0**: el efecto es plausible y positivo.

Práctica B — Estadístico t y decisión

Problema

Con los mismos datos ($\hat{\tau} = 6$, $SE = 2.5$), prueba $H_0 : ATE = 0$ vs. $H_1 : ATE \neq 0$ a nivel $\alpha = 0.05$ (dos colas).

Tu turno

Calcula $t = \hat{\tau} / SE$, compáralo con 1.96 y **decide**. ¿El programa tuvo efecto?

Práctica B — Solución paso a paso

Solución

Paso 1 — estadístico:

$$t = \frac{\hat{\tau}}{SE} = \frac{6}{2.5} = 2.4.$$

Paso 2 — comparar con el crítico: $|t| = 2.4 > 1.96 \Rightarrow$ **rechazo** H_0 ($p \approx 0.016$).

Paso 3 — coherencia: el IC $[1.1, 10.9]$ *no* contiene 0, igual que el test.

Conclusión: el programa **aumentó** el gasto en alimentos de forma estadísticamente significativa.

Contenido

- 1 Causalidad y resultados potenciales
- 2 Sesgo de selección y confusión
- 3 Experimentos aleatorizados (RCT)
- 4 Analizar un RCT
- 5 Regresión no paramétrica**
- 6 Simulación en R
- 7 Manos a la obra

Regresión sin imponer una fórmula

La idea

Queremos la relación $\mathbb{E}[Y \mid X = x] = g(x)$ *sin* suponer que es una recta. La regresión **no paramétrica** (**nonparametric**) deja que los *datos* dibujen la curva, suavizando localmente.

Definición 2 (Regresión por kernel (Nadaraya–Watson))

En cada punto x , promediamos los Y_i de los vecinos, pesando más a los cercanos:

$$\hat{g}(x) = \frac{\sum_{i=1}^n Y_i K\left(\frac{x-x_i}{h}\right)}{\sum_{i=1}^n K\left(\frac{x-x_i}{h}\right)},$$

con **kernel** K (peso que decae con la distancia) y **ancho de banda** (bandwidth) h .

El compromiso del ancho de banda h (sesgo–varianza)

h grande $\Rightarrow$ curva muy suave pero **sesgada** (pierde detalle); h chico $\Rightarrow$ sigue cada zigzag pero con mucha **varianza** (ruidosa). Se elige con *validación cruzada*.

¿Qué es un kernel? Un ejemplo mínimo

Un kernel = una regla de PESOS

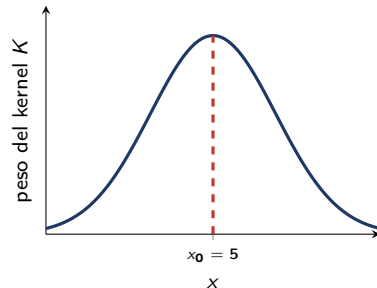
Para estimar el y “típico” en un punto x_0 no promediamos todo por igual: damos **más peso a los vecinos cercanos** a x_0 y menos a los lejanos. El **kernel K** decide cuánto pesa cada punto.

Ejemplo: estimar $\hat{g}(5)$ con 3 vecinos

x_i	y_i	peso K
4	6	0.8
5	7	1.0
7	5	0.3

$$\hat{g}(5) = \frac{0.8(6) + 1.0(7) + 0.3(5)}{0.8 + 1.0 + 0.3} = \frac{13.3}{2.1} \approx 6.3.$$

Es un **promedio ponderado**: el vecino en $x=5$ manda; el lejano $x=7$ casi no.



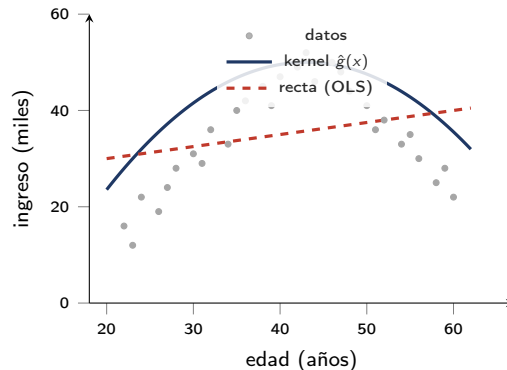
El peso (curva azul) es máximo en x_0 y cae con la distancia.

Kernel vs. recta: cuando la relación se curva

Lectura del gráfico

Los ingresos suben con la edad, llegan a un *máximo* y luego bajan (jubilación). Una **recta** (OLS) no puede capturar esa joroba.

- **Recta**: impone forma lineal, se *equivoca* en los extremos.
- **Kernel**: sigue la curva real, sin fórmula previa.



La recta sube siempre; el kernel detecta el pico hacia los ~ 43 años y la caída posterior.

Demo/intuición: el kernel como promedio local

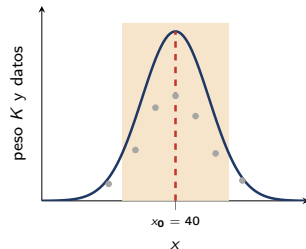
Problema (intuición)

Para estimar $\hat{g}(x_0)$ en $x_0 = 40$ con ancho de banda h , ¿qué hace el kernel? **Se pide:** describir el cálculo y el rol de h .

Solución

- Centramos una “ventana” en $x_0 = 40$ (banda **ámbar**, ancho $\propto h$).
- Cada punto dentro recibe un **peso** $K\left(\frac{x-x_0}{h}\right)$: más cerca, más peso.
- $\hat{g}(40) = \text{promedio ponderado}$ de los Y_i de esa ventana.
- Deslizamos la ventana sobre una rejilla de puntos $\Rightarrow$ curva.

h **controla el ancho de ventana**: más ancho = más suave (más sesgo); más angosto = más fiel pero ruidoso.



Banda **ámbar**: la ventana de ancho $\propto h$.
Curva azul: el peso K , máximo en x_0 y decreciente hacia los lados.

Práctica A — Kernel uniforme a mano

(¡inténtalo!)

Problema

Salario y (cientos S/. /h) vs. años de educación x . Seis datos:

$$(x, y) : (1, 2), (3, 5), (5, 7), (7, 8), (9, 8), (11, 7).$$

Estima $\hat{g}(6)$ con un **kernel uniforme**: peso 1 si $|x - 6| \leq 4$ y 0 si no (bandwidth $h = 4$).

Tu turno

¿Qué puntos *entran* en la ventana y cuánto vale $\hat{g}(6)$?

$$\hat{g}(x_0) = \frac{\sum_i K\left(\frac{x_i - x_0}{h}\right) y_i}{\sum_i K\left(\frac{x_i - x_0}{h}\right)}.$$

Práctica A — Solución paso a paso

Solución

Paso 1 — quién entra. Con $|x - 6| \leq 4$ entran $x \in \{3, 5, 7, 9\}$; *salen* $x = 1$ y $x = 11$ (distancia $5 > 4$).

Paso 2 — pesos. El kernel uniforme da peso 1 a cada punto de la ventana.

Paso 3 — promedio (ponderado) local:

$$\hat{g}(6) = \frac{5 + 7 + 8 + 8}{4} = \frac{28}{4} = \boxed{7.0}.$$

Es simplemente el **promedio de los vecinos** dentro de la ventana.

Práctica B — Kernel triangular vs. recta

Problema

Mismos datos y $x_0 = 6$, ahora con **kernel triangular** $K(u) = \max(0, 1 - |u|)$, con $u = \frac{x - 6}{4}$ (bandwidth $h = 4$).

Tu turno

Calcula los **pesos** $K(u_i)$ y obtén $\hat{g}(6)$. Compáralo con la recta OLS $\hat{y}(6) = 7.17$: ¿qué capta el kernel que la recta no?

Práctica B — Solución paso a paso

Solución

Paso 1 — pesos $K(u_i)$ con $u_i = (x_i - 6)/4$:

x_i	1	3	5	7	9	11
u_i	-1.25	-0.75	-0.25	0.25	0.75	1.25
$K(u_i)$	0	0.25	0.75	0.75	0.25	0

Paso 2 — promedio ponderado:

$$\hat{g}(6) = \frac{0.25(5) + 0.75(7) + 0.75(8) + 0.25(8)}{0.25 + 0.75 + 0.75 + 0.25} = \frac{1.25 + 5.25 + 6 + 2}{2} = \frac{14.5}{2} = \boxed{7.25}.$$

Paso 3 — vs. la recta. El kernel (7.25) pondera *más* a los vecinos cercanos; la recta (7.17) impone forma lineal y no capta la **joroba** (el salario se aplana con mucha educación). **Bandwidth:** h grande $\Rightarrow$ muy suave (sesgo); h chico $\Rightarrow$ ruidoso (varianza).

La otra cara de lo “no paramétrico”: comparar distribuciones

¿Qué tiene que ver esto con la regresión por kernel?

“No paramétrico” significa **no imponer una forma**. Es un paraguas con **dos** herramientas:

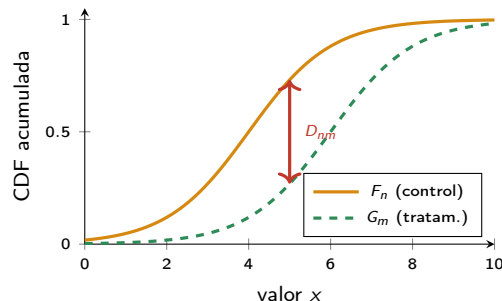
- **Regresión** por kernel: estima la curva $\mathbb{E}[Y | X]$ sin suponer una recta (*lo que acabamos de ver*).
- **Comparación** de distribuciones: ¿el tratamiento cambió *toda* la distribución de Y , no solo la media? Para eso está **Kolmogorov–Smirnov (KS)**.

Kolmogorov–Smirnov (KS)

Compara las dos **CDF empíricas** F_n (control) y G_m (tratamiento) y mide su **mayor brecha vertical**:

$$D_{nm} = \max_x |F_n(x) - G_m(x)|.$$

D_{nm} grande $\Rightarrow$ las distribuciones difieren. En R: `ks.test`.



El tratamiento corre la curva a la *derecha* (subió los valores).

Contenido

- 1 Causalidad y resultados potenciales
- 2 Sesgo de selección y confusión
- 3 Experimentos aleatorizados (RCT)
- 4 Analizar un RCT
- 5 Regresión no paramétrica
- 6 Simulación en R**
- 7 Manos a la obra

Simulación: un RCT de principio a fin

Idea

Simulamos un RCT donde el tratamiento tiene un efecto *real* $\tau = 5$. Aleatorizamos, comparamos medias y corremos un `t.test`: la diferencia estimada debe rondar 5 y resultar **significativa**.

En R

```
set.seed(2026); n <- 200; tau <- 5
W <- rbinom(n, 1, 0.5) # asignacion aleatoria 0/1
Y0 <- rnorm(n, mean = 50, sd = 10) # resultado potencial control
Y <- Y0 + tau * W # tratados reciben +tau
mean(Y[W==1]) - mean(Y[W==0]) # diferencia de medias ~ 5
t.test(Y ~ W) # IC y p-valor del efecto
```

Lectura

Como W se asigna *al azar*, los grupos son comparables y la diferencia de medias estima el ATE sin sesgo. Cambia `tau <- 0`: ahora $\sim 95\%$ de las veces el IC contendrá el 0 (no rechazas H_0).

Contenido

- 1 Causalidad y resultados potenciales
- 2 Sesgo de selección y confusión
- 3 Experimentos aleatorizados (RCT)
- 4 Analizar un RCT
- 5 Regresión no paramétrica
- 6 Simulación en R
- 7 Manos a la obra**

¿Y ahora?

- Repasa el marco de **resultados potenciales**: $Y(1)$, $Y(0)$, el efecto individual y el ATE; recuerda el *problema fundamental* (solo ves uno).
- Distingue **correlación** de **causalidad**: busca el confusor y el sesgo de selección antes de afirmar “X causa Y”.
- Para un **RCT**: la diferencia de medias es el efecto causal; reporta también su **SE**, su **IC** y la **prueba** (Clase 6).
- Cuando la relación se curva, prueba **regresión por kernel**: en R con `ksmooth`, `loess` o `npreg` (paquete `np`); para comparar distribuciones, `ks.test`.
- Simula con `rbinom`, `rnorm`, `t.test` y `replicate(...)` para *ver* cómo la aleatorización elimina el sesgo.

¡A practicar!