

Pruebas de hipótesis: Resumiendo información acerca de los efectos causales

Lily Medina

28 November 2022

Pruebas de hipótesis

Panorama general

- ▶ Estamos interesados en cuáles es el efecto causal de una intervención; Entonces:
 1. Formulamos una hipótesis acerca de ese efecto causal. Por ejemplo: X tiene un efecto positivo sobre Y.
 2. Probamos la hipótesis
 3. Podemos decir si nuestra hipótesis puede ser o no cierta a la luz de la evidencia recolectada.

Qué veremos en esta sesión?

- ▶ Problema fundamental de la inferencia causal: para una unidad sólo podemos observar el valor que toma la variable de interés Y si esa unidad fue asignada a control o al tratamiento, pero no a los dos.
- ▶ Estimación del promedio de los efectos causales
- ▶ Resultados potenciales
- ▶ Valores p

Tres escenarios hipotéticos:

1. El mejor escenario: Podemos observar simultáneamente para cada unidad el valor de la variable de interés (Y) cuando estas son asignadas al grupo de control ($T = 0$) y cuando son asignadas al grupo de tratamiento ($T = 1$.)
2. El segundo mejor escenario: Podemos realizar un número infinito de experimentos
3. El tercer mejor escenario: Podemos realizar SÓLO UN experimento

Primer escenario: podemos observar el efecto
causal para cada unidad

Primer escenario: El mejor escenario

Podemos observar el valor de la variable de interés (Y) cuando las unidades son asignadas al control ($T = 0$) y cuando son asignadas a recibir el tratamiento ($T = 1$.)

participantes	control	tratamiento
Noelia	0	1
Olivia	2	2
Simón	3	1
Jose	1	4
Santiago	3	5
Alejandra	5	5
Juan	6	7
Antonio	2	4
Sofia	3	5
Emiliano	4	6

Preguntas:

- ▶ ¿Cuál es el efecto del tratamiento para cada participante?
- ▶ ¿Cuál es el promedio de los efectos del tratamiento? O cómo calcularlo?

Efecto del tratamiento para cada participante

```
RP$diferencia <- RP$tratamiento - RP$control
```

participantes	control	tratamiento	diferencia
Noelia	0	1	1
Olivia	2	2	0
Simón	3	1	-2
Jose	1	4	3
Santiago	3	5	2
Alejandra	5	5	0
Juan	6	7	1
Antonio	2	4	2
Sofia	3	5	2
Emiliano	4	6	2

Promedio de los efectos del tratamiento

participantes	control	tratamiento	diferencia
Noelia	0	1	1
Olivia	2	2	0
Simón	3	1	-2
Jose	1	4	3
Santiago	3	5	2
Alejandra	5	5	0
Juan	6	7	1
Antonio	2	4	2
Sofia	3	5	2
Emiliano	4	6	2

```
mean(RP$diferencia)
```

```
[1] 1.1
```

```
mean(RP$tratamiento- RP$control )
```

```
[1] 1.1
```

Recapitulando

- ▶ ¿Cómo definimos nuestro primer escenario?

Recapitulando

- ▶ ¿Cómo definimos nuestro primer escenario?

Como un escenario en el que podemos observar simultáneamente para cada unidad el valor de la variable de interés (Y) cuando estas son asignadas al grupo de control ($T = 0$) y cuando son asignadas al grupo de tratamiento ($T = 1$.)

Problemas con el primer escenario

- ▶ ¿Cuál es el problema con este escenario?

Problemas con el primer escenario

- ▶ ¿Cuál es el problema con este escenario?

Una vez conducimos nuestro experimento, sólo observamos un resultado potencial para cada unidad. Si Noelia es asignada a recibir el tratamiento, ya no podemos observar qué habría pasado si Noelia hubiera sido asignada a control y viceversa.

Problemas con el primer escenario

► ¿Cuál es el problema con este escenario?

Una vez conducimos nuestro experimento, sólo observamos un resultado potencial para cada unidad. Si Noelia es asignada a recibir el tratamiento, ya no podemos observar qué habría pasado si Noelia hubiera sido asignada a control y viceversa.

No podemos calcular el promedio del efecto del tratamiento como vimos en la diapositiva anterior. No importa cuántos datos tengamos. Nunca vamos a poder observar los dos resultados potenciales al mismo tiempo.

Problemas con el primer escenario

- ▶ ¿Cuál es el problema con este escenario?

Una vez conducimos nuestro experimento, sólo observamos un resultado potencial para cada unidad. Si Noelia es asignada a recibir el tratamiento, ya no podemos observar qué habría pasado si Noelia hubiera sido asignada a control y viceversa.

No podemos calcular el promedio del efecto del tratamiento como vimos en la diapositiva anterior. No importa cuántos datos tengamos. Nunca vamos a poder observar los dos resultados potenciales al mismo tiempo.

Esto es a lo que llamamos **el problema fundamental de la inferencia causal**.

Segundo escenario: podemos realizar un número infinito de experimentos

¿Cuál es la situación ahora?

- ▶ No podemos observar el efecto causal para cada unidad
- ▶ Intuición: recibes el tratamiento (ejemplo: un medicamento) o no lo recibes. Pero las dos cosas no pasan al mismo tiempo
- ▶ Esto es a lo que llamamos **el problema fundamental de la inferencia causal**.
- ▶ No observamos los valores contrafactuales
- ▶ Hay muchos datos faltantes

¿Cómo podemos solucionarlo?

- ▶ Aunque nunca podremos observar el efecto causal para cada unidad. Sí podemos calcular el promedio de los valores que sí observamos.
- ▶ Gracias a selección/asignación aleatoria podemos obtener estimaciones confiables.

Veámos lo que nos permite el selección/asignación aleatoria

Enfoquémonos por ahora en el promedio del resultado potencial bajo el tratamiento

participantes	control	tratamiento
Noelia	0	1
Olivia	2	2
Simón	3	1
Jose	1	4
Santiago	3	5
Alejandra	5	5
Juan	6	7
Antonio	2	4
Sofia	3	5
Emiliano	4	6

```
mean(RP$tratamiento)
```

```
[1] 4
```

Pregunta

¿Podemos obtener una estimación confiable de este promedio si no observamos el resultado potencial bajo el tratamiento para cada participante?

Por ejemplo

¿Qué pasa si hacemos solo seleccionamos 5 unidades aleatoriamente para el grupo del tratamiento?

participantes	tratamiento	seleccion
Noelia	1	si
Olivia		no
Simón		no
Jose	4	si
Santiago	5	si
Alejandra		no
Juan		no
Antonio	4	si
Sofia	5	si
Emiliano		no

```
mean(muestra$tratamiento, na.rm = TRUE)
```

```
[1] 3.8
```

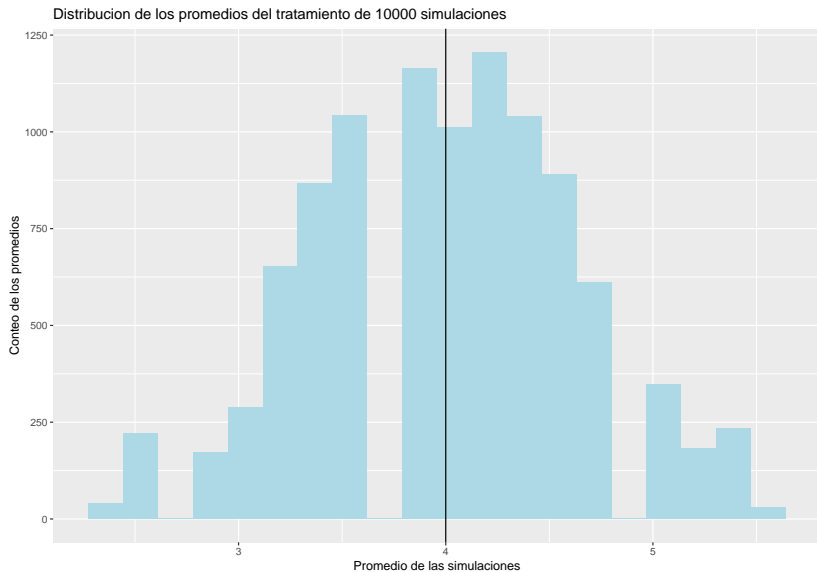
¿Qué pasa si hacemos esa misma selección aleatoria una vez más?

participantes	tratamiento	seleccion
Noelia		no
Olivia	2	si
Simón	1	si
Jose	4	si
Santiago	5	si
Alejandra		no
Juan	7	si
Antonio		no
Sofia		no
Emiliano		no

```
mean(muestra$tratamiento, na.rm = TRUE)
```

```
[1] 3.8
```

¿Qué pasa si hacemos esto **muchas** veces?



Ahora comparemos el promedio de las simulaciones con el real

Simulaciones

```
mean(promedios)
```

```
## [1] 4.00652
```

Promedio “real”

```
mean(RP$tratamiento)
```

```
## [1] 4
```


Panorama general

¿Qué hemos aprendido?

- ▶ Que podemos estimar el promedio real de una cantidad sin tener todos las observaciones.

¿Para qué nos sirve este aprendizaje?

- ▶ Vamos a utilizar esta misma idea para estimar el promedio de los efectos causales.

Asignación Aleatoria

participantes	tratamiento	control	asignacion
Noelia	1		1
Olivia		2	0
Simón		3	0
Jose		1	0
Santiago	5		1
Alejandra		5	0
Juan	7		1
Antonio	4		1
Sofia		3	0
Emiliano	6		1

Si `asignacion = 1`, el participante fue asignado al grupo de tratamiento y si `asignacion = 0`, el participante fue asignado al control