

Potential Outcomes

Rodrigo Negrete Pérez

February 2, 2022

- 1 Introducción
- 2 Potential Outcomes
- 3 Efectos Causales
- 4 Simulaciones
- 5 Sesgo de selección
- 6 SUTVA

Section 1

Introducción

Introducción

El hilo conductor del curso es la pregunta de la causalidad

- No cualquier tipo de pregunta causal
 - ¿Cuál es el efecto de A sobre B?
- A manera de ejemplo: efecto de la educación superior sobre el salario

Educación y salario

- Parece intuitivo, pero pueden pasar muchas cosas en medio de la relación
- ¿Cuál es el efecto. . . de la educación?
- Podemos plantear un modelo: con datos simulados podemos saber qué estima nuestro modelo
 - Si estimamos con sesgo
 - ventajas y desventajas

Section 2

Potential Outcomes

Potential Outcomes

Suponemos que la entidad tiene dos posibles resultados: bajo tratamiento y sin tratamiento (control)

- Si Rodrigo va a la universidad, su salario sería de 80k mensuales
- Si no va a la universidad, su salario sería de 25k

Entonces, denotamos $Y_{0R} = 25$ al outcome bajo control y $Y_{1R} = 80$ al outcome bajo tratamiento

Problema fundamental de la inferencia causal

El problema es que solo observamos uno de los dos posibles potential outcomes:

- Si Rodrigo va a la uni, observaré Y_{1R}
- Si no, Y_{0R}

$$Y_i = \begin{cases} Y_{1i}, & \text{si } t = 1 \\ Y_{0i}, & \text{si } t = 0 \end{cases} \quad (1)$$

Entonces, podemos escribir la Y observada como

$$Y_i = Y_{0i} + (Y_{1i} - Y_{0i})T_i$$

Generación de P.O.

Vamos a generar los Potential Outcomes

Supongamos que:

- $Y_{0i} \sim N(30, 10)$
- $Y_{1i} \sim N(50, 35)$

```
set.seed(2022)
n<-10000
i<-1:n
y0<-rnorm(n, 30, 10)
y1<-rnorm(n, 50, 35)
df<-data.frame(i,y0,y1)
```

##	i	y0	y1
## 1	1	39.0014199	88.53769
## 2	2	18.2665423	58.19311
## 3	3	21.0251464	31.82775
## 4	4	15.5549860	33.65446
## 5	5	26.6898642	68.54422
## 6	6	0.9937101	61.58160

Asignación del tratamiento

Tenemos que asignar el tratamiento

- Podríamos hacerlo con un vector de 1 y 0, y que t sea equiprobable: simple random assignment
- Podríamos asignar t a mitad de la muestra: complete random assignment
- Para complete ra es más práctico usar el paquete randomizr: `complete_ra()`

```
library(randomizr)
```

	i	y0	y1	t
1	1	39.00	88.54	1
2	2	18.27	58.19	1
3	3	21.03	31.83	0
4	4	15.55	33.65	1
5	5	26.69	68.54	0
6	6	0.99	61.58	0

Así, podemos añadir la Y observada

```
df<-mutate(df,y_observada=y0+(y1-y0)*t)
```

	i	y0	y1	t	y_observada
1	1	39.00	88.54	1	88.54
2	2	18.27	58.19	1	58.19
3	3	21.03	31.83	0	21.03
4	4	15.55	33.65	1	33.65
5	5	26.69	68.54	0	26.69
6	6	0.99	61.58	0	0.99

Section 3

Efectos Causales

Efecto Causal Individual

Cuando tenemos ambos P.O. podemos calcular el **Efecto Causal Individual** (δ_i) tomando la diferencia entre Y_{1i} y Y_{0i}

$$\delta_i = Y_{1i} - Y_{0i}$$

- Por ejemplo, decimos que el efecto de la uni en Rodrigo es de $\delta_R = 80 - 25$ pesos mensuales
- Hay una gran heterogeneidad en los efectos causales individuales
- Hay 5 individuos cuyo salario es mayor sin universidad

```
df<-mutate(df, delta=y1-y0)
```

	i	y0	y1	t	y_observada	delta
1	1	39.00	88.54	1	88.54	49.54
2	2	18.27	58.19	1	58.19	39.93
3	3	21.03	31.83	0	21.03	10.80
4	4	15.55	33.65	1	33.65	18.10
5	5	26.69	68.54	0	26.69	41.85
6	6	0.99	61.58	0	0.99	60.59

Efecto Causal Promedio

- Nos importa el efecto que tiene la uni en general: el efecto promedio

$$ATE = E[Y_{1i} - Y_{0i}] = E[\delta_i]$$

- Basta con tomar el promedio de los efectos causales individuales, pero para toda la muestra.

Efecto promedio sobre los tratados

- De igual manera, podemos calcular el efecto promedio, pero sobre los tratados

$$ATT = E[Y_{1i} - Y_{0i} | T_i = 1]$$

- Es calcular el ATE, pero solo para aquellos que fueron tratados.

Efecto promedio sobre los controles

- Análogamente, el ATC es el ATE, pero consideramos solo a aquellos que fueron controles

$$ATC = E[Y_{1i} - Y_{0i} | T_i = 0]$$

- Con datos generados es muy fácil calcular los efectos causales
- Tomamos los efectos causales individuales, los dividimos en subconjuntos, y luego calculamos promedios
- En R, las funciones **filter()** y **with()** nos van a ser muy útiles

```
ate<-with(df, mean(delta))  
att<-with(filter(df, t==1), mean(delta))  
atc<-with(filter(df, t==0), mean(delta))
```

- Obtenemos que el $ATE=20.5874983$
 - $ATT=20.3810284$
 - $ATC=20.7939683$

Contraste Ingenuo

- Sin embargo, el investigador no puede observar estos efectos. Ninguno es observable
- Lo que el investigador puede observar es el Contraste ingenuo: el promedio del outcome para los tratados menos el promedio del outcome para los no tratados

$$CI = E[Y_{1i} | T_i = 1] - E[Y_{0i} | T_i = 0]$$

- En nuestro df definimos las `y_observada(s)`
- Para calcular el CI, basta calcular el promedio de las `y_observadas` de los que fueron tratados, y restarle el promedio de las `y_observadas` de los controles.

```
ci<-with(filter(df, t==1),  
          mean(y_observada))-  
with(filter(df, t==0),  
      mean(y_observada))
```

- Obtenemos un $CI=20.3875545$, el cual es muy parecido al $ATE=20.5874983$
- Entonces, ¿por qué tanto problema?

Section 4

Simulaciones

Virtudes de las simulaciones

- Los datos que generamos provienen de una distribución normal.
- Podemos calcular muchas cosas usando cálculo diferencial.
 - Por ejemplo, ya sabíamos que el ATE iba a ser de 20
 - Sin embargo, por el elemento aleatorio no dio 20, sino 20.5874983
 - Pero no siempre será tan fácil como aplicar las propiedades de la normal.
- Alternativamente al cálculo diferencial, podemos hacer el cálculo muchas muchas veces y ver hacia dónde tiende

- Por ejemplo, ¿nuestro CI coincidió con el ATE por el elemento aleatorio? ¿Fue una casualidad?
- ¡Simulemos! Hagamos el mismo cálculo, pero varias veces: para cada df tomemos el ATE y el CI.
- En R, podemos hacerlo fácilmente con un FOR loop.
- Hagámoslo unas 5000 veces

```
numrep<-5000
```

- Preparemos el loop
- Luego podremos mover el número de repeticiones y la cantidad de individuos

```
set.seed(2022)
#semilla
numrep<-5000
n<-10000

ate_estimates<-NULL #vectores vacios
ci_estimates<-NULL  # para guardar datos
```

```

for (i in 1:numrep) {
  # Creacion df
  df<-data.frame(i=1:n,
                 y0=rnorm(n, 30, 10),
                 y1=rnorm(n, 50, 35),
                 t=complete_ra(n)) %>%
  mutate(y_observada=y0+(y1-y0)*t,
         delta=y1-y0)

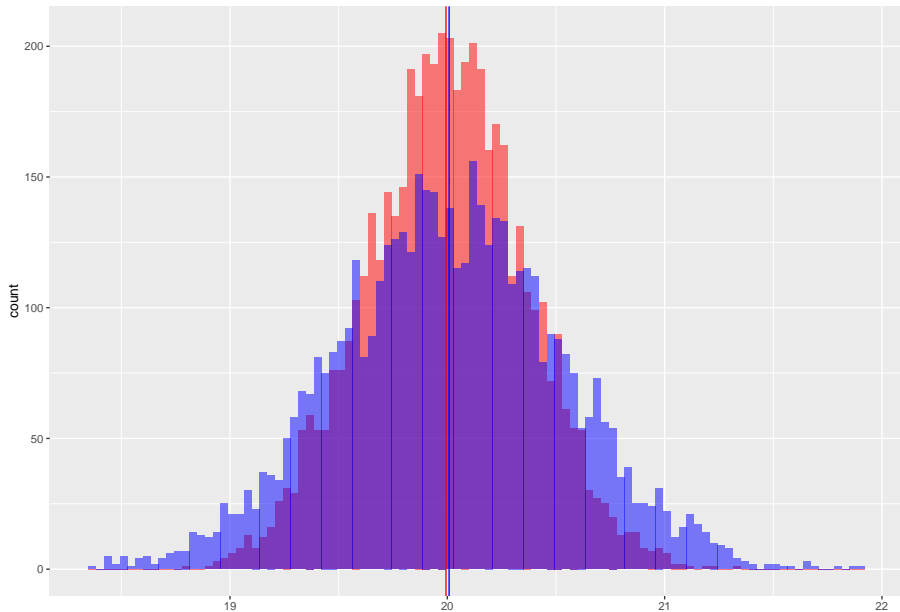
  # Calculo efectos causales
  # Debemos guardarlos como entrada en vectores
  ate_estimates[i]<- with(df, mean(delta))

  ci_estimates[i]<- with(filter(df, t==1),
                          mean(y_observada))-
    with(filter(df, t==0),
          mean(y_observada))
}

```

- Hagamos un análisis visual
- Hagamos un histograma para los ATE's y otro para los CI's

```
ggplot()+  
  geom_histogram(aes(ate_estimates),  
                 fill='red', alpha=.5,  
                 bins = 100)+  
  geom_histogram(aes(ci_estimates),  
                 fill='blue', alpha=.5,  
                 bins=100)+  
  geom_vline(xintercept = mean(ate_estimates),  
             color='red')+  
  geom_vline(xintercept = mean(ci_estimates),  
             color='blue')+  
  xlab('')
```

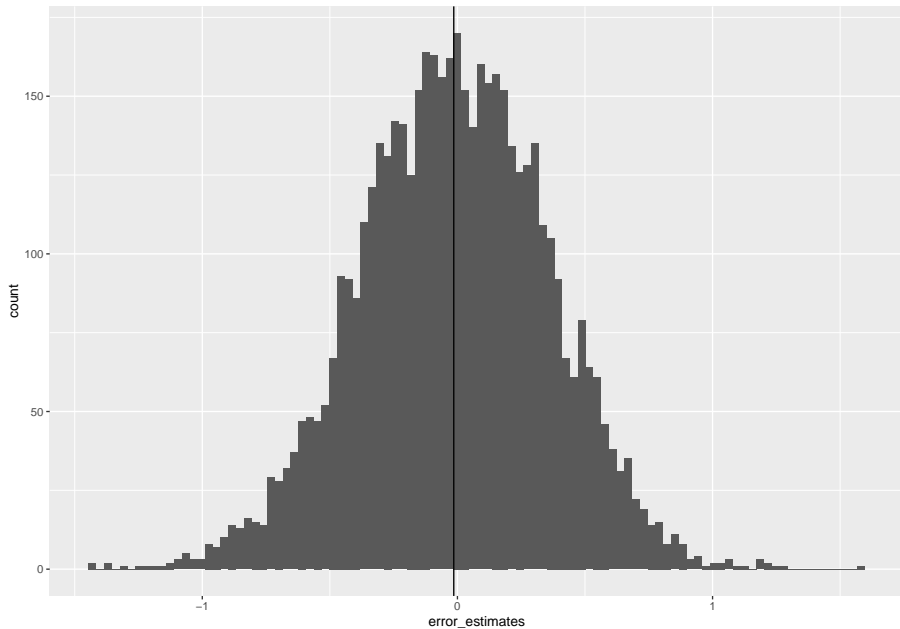


- Tratemus de graficar las diferencias

```
error_estimates<-ate_estimates-  
ci_estimates
```

- Hacemos un vector porque ggplot requiere un df


```
ggplot()+  
  geom_histogram(aes(x=error_estimates),  
                 bins = 100)+  
  geom_vline(xintercept =  
             mean(error_estimates))
```



- Podríamos concluir que el CI estima insesgadamente el ATE...
peeeroo
- Nos apoyamos en algunos SUPUESTOS

Section 5

Sesgo de selección

Unconfoundedness

En el centro de todas las estrategias de identificación yace el supuesto de **Unconfoundedness**

- Una vez condicionemos sobre todo lo que tenemos que controlar, la asignación del tratamiento es cuasialeatoria
- La asignación del tratamiento no puede depender de los Potential Outcomes
- Si los individuos se tratan porque creen que el tratamiento les va a funcionar, se están **autoseleccionando**.

Asignación aleatoria del tratamiento

- La asignación aleatoria del tratamiento (junto con la Ley de los Grandes Números) garantiza que la asignación aleatoria del tratamiento no depende de los P.O.
- En nuestras simulaciones, la asignación de t siempre fue aleatoria

Autoselección

- Repitamos las simulaciones, pero cambiemos la asignación de t
- Hagamos que t dependa de los P.O.

- La Educación universitaria es costosa
 - Costos monetarios
 - Costos de oportunidad
 - Tiempo
- Supongamos que solo se tratan aquellos para los que la uni les mejora considerablemente el salario mensual
- Supongamos que:

$$T_i = \begin{cases} 1, & \text{si } Y_{1i} - Y_{0i} > 10 \\ 0, & \text{e.o.c.} \end{cases} \quad (2)$$


```
set.seed(2022)
#semilla
numrep<-5000
n<-10000

ate_estimates<-NULL #vectores vacios
ci_estimates<-NULL # para guardar datos
```

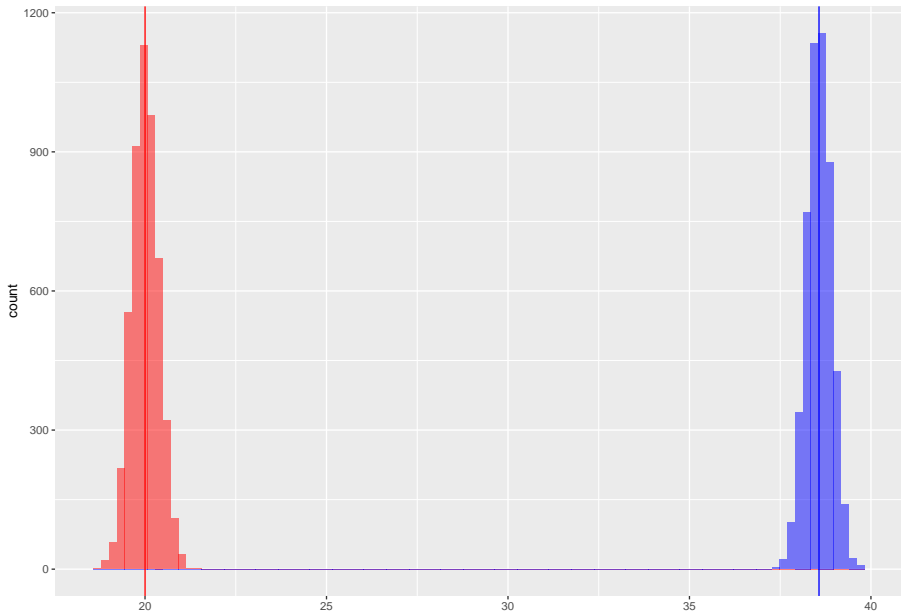
```

for (i in 1:numrep) {
  # Creacion df
  df<-data.frame(i=1:n,
                 y0=rnorm(n, 30, 10),
                 y1=rnorm(n, 50, 35)) %>%
  mutate(t=ifelse(y1-y0>10,1,0), #Cambiamos t
         y_observada=y0+(y1-y0)*t,
         delta=y1-y0)

  # Calculo efectos causales
  # Debemos guardarlos como entrada en vectores
  ate_estimates[i]<- with(df, mean(delta))

  ci_estimates[i]<- with(filter(df, t==1),
                          mean(y_observada))-
    with(filter(df, t==0),
          mean(y_observada))
}

```



- La diferencia es dramática:
 - El promedio de nuestros ATE's es 20.0008954, mientras que el promedio de los CI's es 38.5690034
- Nuestro CI tiende a sobreestimar el efecto real de la uni (y por mucho)
- Extrapolando: si los individuos se autoseleccionan, el CI va a estar sesgado
- La dirección y tamaño del sesgo va a depender del proceso de autoselección y los datos

- ¿Qué está pasando?
- Regresemos a la definición de CI y manipulemos un poco la expresión

$$CI = E[Y_{1i}|T_i = 1] - E[Y_{0i}|T_i = 0]$$

- Sumemos un cero $E[Y_{0i}|T_i = 1]$

$$E[Y_{1i}|T_i = 1] - E[Y_{0i}|T_i = 0] = E[Y_{1i}|T_i = 1] - E[Y_{0i}|T_i = 1] + E[Y_{0i}|T_i = 1] - E[Y_{0i}|T_i = 0] \quad (3)$$

Si agrupamos los términos, notamos que primero aparece el ATT, seguido del Sesgo de selección

$$\text{Sesgo de Selección} = E[Y_{0i} | T_i = 1] - E[Y_{0i} | T_i = 0]$$

- El sesgo de selección es, en promedio, qué tan diferentes eran los tratados de los no tratados antes del tratamiento, de ahí que sea la diferencia en y_0

Entonces

$$CI = ATT + \text{Sesgo de Selección}$$

- Hagamos una sola corrida para ver qué está pasando
- Intenten hacer las simulaciones en casa

```
set.seed(2022)
n<-10000
i<-1:n
y0<-rnorm(n, 30, 10)
y1<-rnorm(n, 50, 35)
df<-data.frame(i,y0,y1)
# Asignamos el tratamiento
df<-mutate(df, t=ifelse(y1-y0>10,1,0) )
# Ya con el tratamiento, ponemos la y observada
df<-mutate(df, y_observada=y0+(y1-y0)*t)
# Podemos poner el efecto causal individual
df<-mutate(df, delta=y1-y0)
```

```
df<-mutate(df, delta=y1-y0)
ate<-with(df, mean(delta))
att<-with(filter(df, t==1), mean(delta))
atc<-with(filter(df, t==0), mean(delta))
ci<-with(filter(df, t==1),
          mean(y_observada))-
  with(filter(df, t==0),
        mean(y_observada))
s.bias<-with(filter(df, t==1), mean(y0))-
  with(filter(df, t==0), mean(y0))
```

```
print(ate)
```

```
## [1] 20.5875
```

```
print(ci)
```

```
## [1] 38.44198
```

```
print(att)
```

```
## [1] 42.93694
```

```
print(s.bias)
```

```
## [1] -4.494957
```

El ATE y el CI difieren por mucho. El CI es el ATT más el Sesgo de selección. El problema es que el ATT difiere enormemente del ATE porque los que se tratan son solo aquellos cuyo efecto causal es grande. Cuando hay sesgo de selección el ATE es muy diferente al CI.

Ley de los grandes números

- La asignación aleatoria no asegura unconfoundedness
- Necesitamos una muestra considerable
- También, por pura chance, puede que nuestros controles y tratamientos difieran significativamente
- Podríamos hacer simulaciones cuando veamos la regresión y podamos añadir otras variables al DGP

Section 6

SUTVA

SUTVA

A grandes rasgos, SUTVA pide:

- No Spillovers
 - Si no, confundiríamos controles con tratamiento
- Misma dosis