Profesor: Héctor Bahamonde, PhD.

e:hector.bahamonde@uoh.cl

w:www.hectorbahamonde.com

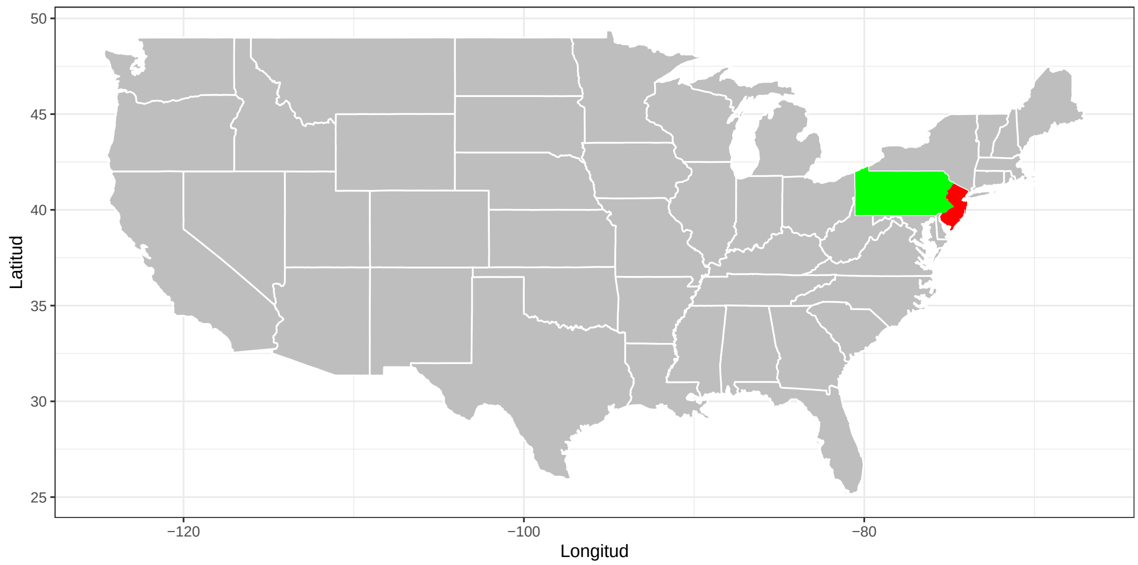
Curso: MLE.

TA: Gonzalo Barría.

I. Dif in Dif

Card and Krueger (1994) tratan de usar la geografía como *as-if random* factor (?). La pregunta que ellos tenían era cuál es el efecto (causal?) en el empleo ante un incremento en el salario en el sector de comida rápida. En otras palabras, *qué ocurre con el empleo en este sector cuando aumenta el salario, sube o baja?*

```r
# instalar librerias
p_load(ggplot2,dplyr,maps)
# declarar datos
all_states <- map_data("state")
# plot
ggplot(all_states,
       aes(x=long, y=lat, group = group)) +
  geom_polygon(fill="grey", colour = "white") +
  geom_polygon(fill="green", data = filter(all_states, region %in% c("pennsylvania"))) +
  geom_polygon(fill="red", data = filter(all_states, region %in% c("new jersey"))
               ) +
  theme_bw() +
  xlab("Longitud") +
  ylab("Latitud")
```

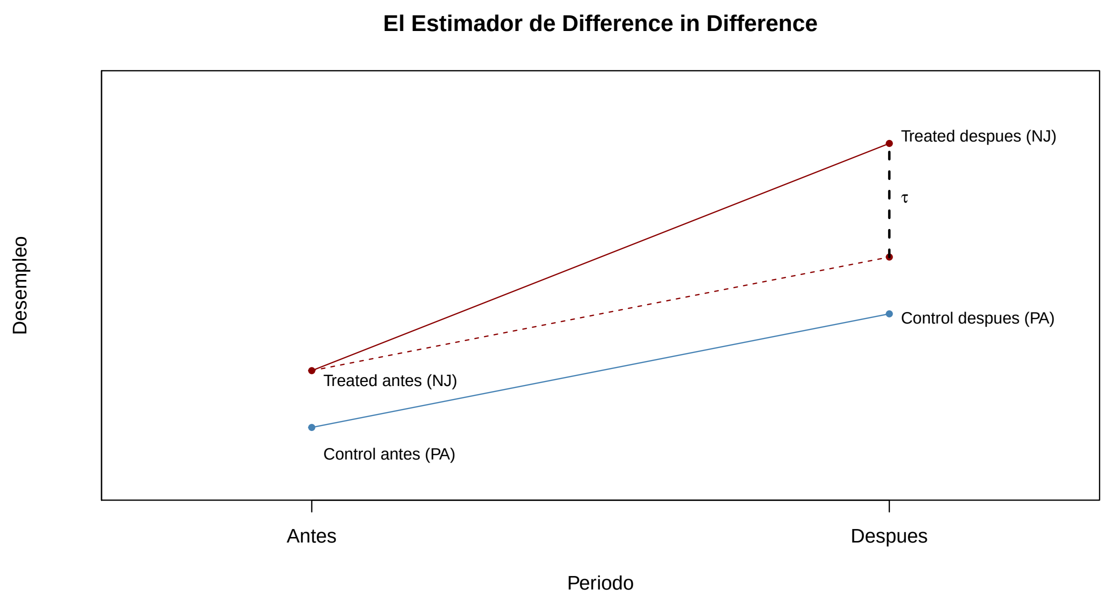


Sin embargo, PA y NJ puede que hayan sido distintos en varias características (que es lo más probable, no?). El problema es que si quisiéramos calcular el efecto causal $\text{Desempleo}_{\text{PA}}$-$\text{Desempleo}_{\text{NJ}}$ cuando subimos el salario en NJ, ya PA podría contar con un piso que no estemos tomando en consideración. Cómo podríamos calcular el efecto causal τ, pero *tomando en cuenta las características de base de PA*?

Veámos un gráfico que podría aclarar lo que queremos.

```
plot(c(0, 1), c(6, 8),
     type = "p",
     ylim = c(5, 12),
     xlim = c(-0.3, 1.3),
     main = "El Estimador de Difference in Difference",
     xlab = "Periodo",
     ylab = "Desempleo",
     col = "steelblue",
     pch = 20,
     xaxt = "n",
     yaxt = "n")
#
```

```r
axis(1, at = c(0, 1), labels = c("Antes", "Despues"))
axis(2, at = c(0, 13))
# treatment
points(c(0, 1, 1), c(7, 9, 11),
       col = "darkred",
       pch = 20)
#
lines(c(0, 1), c(7, 11), col = "darkred")
lines(c(0, 1), c(6, 8), col = "steelblue")
lines(c(0, 1), c(7, 9), col = "darkred", lty = 2)
lines(c(1, 1), c(9, 11), col = "black", lty = 2, lwd = 2)
#
text(1, 10, expression(tau), cex = 0.8, pos = 4)
text(0, 5.5, "Control antes (PA)", cex = 0.8 , pos = 4)
text(0, 6.8, "Treated antes (NJ)", cex = 0.8 , pos = 4)
text(1, 7.9, "Control despues (PA)", cex = 0.8 , pos = 4)
text(1, 11.1, "Treated despues (NJ)", cex = 0.8 , pos = 4)
```



- Cuál es el contrafactual?
- Qué significa o cómo leemos τ?

De manera más formal, $\tau = \beta_1$ en Equation 1:

$$\Delta y_i = \beta_0 + \beta_1 z_i + \epsilon_i \tag{1}$$

donde z es el *estado* del tratamiento, basicamente un vector x de 0's y 1's, Δy_i es la diferencia (el "delta") en y cuando $z(0)$ cambia a $z(1)$, y $\tau = \beta_1$ es el *DID estimator* especificado en Equation 2:

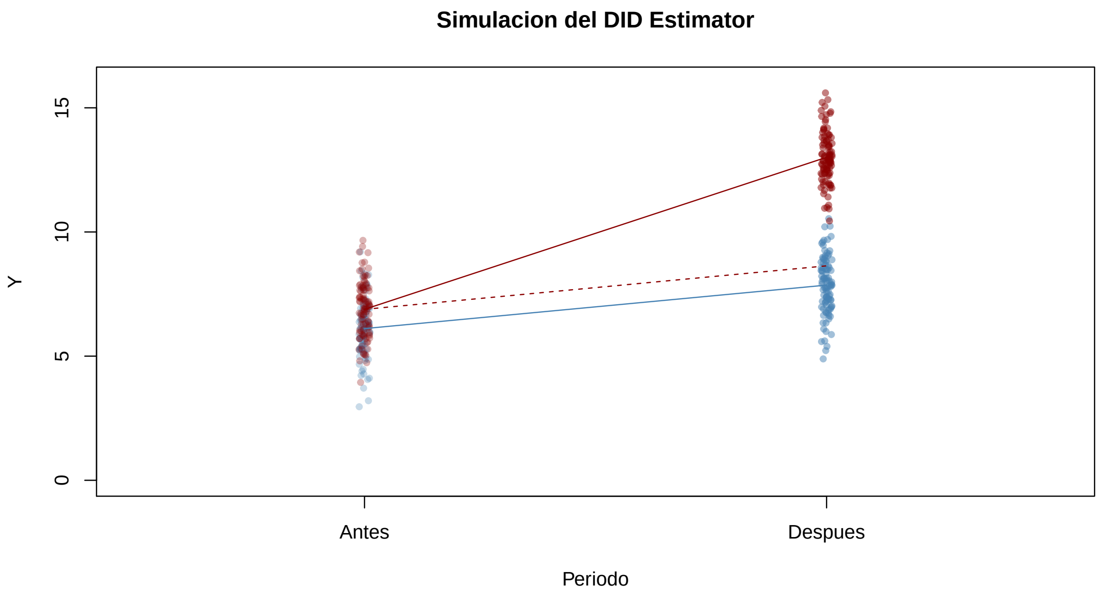
$$\tau = (y_{\text{Control Antes}} - y_{\text{Control Después}}) - (y_{\text{Treated Antes}} - y_{\text{Treated Después}}) \tag{2}$$

Si te fijas, el estimador es "la diferencia de la diferencia" (o "*difference in difference*").

Cocinemos unos datos.

```r
n <- 200
# definir tau
TEffect <- 4
# generar dummy de treatment
z <- c(rep(0, n/2), rep(1, n/2))
# simular pre y post treatment en la y
y_pre <- 7 + rnorm(n)
y_pre[1:n/2] <- y_pre[1:n/2] - 1
y_post <- 7 + 2 + TEffect * z + rnorm(n)
y_post[1:n/2] <- y_post[1:n/2] - 1
#
p_load(scales) # para usar alpha abajo (colores)
pre <- rep(0, length(y_pre[z==0]))
post <- rep(1, length(y_pre[z==0]))
# t=1
plot(jitter(pre, 0.6),
     y_pre[z == 0],
```

```
     ylim = c(0, 16),
     col = alpha("steelblue", 0.3),
     pch = 20,
     xlim = c(-0.5, 1.5),
     ylab = "Y",
     xlab = "Periodo",
     xaxt = "n",
     main = "Simulacion del DID Estimator")
axis(1, at = c(0, 1), labels = c("Antes", "Despues"))
# treatment t=1
points(jitter(pre, 0.6),
       y_pre[z == 1],
       col = alpha("darkred", 0.3),
       pch = 20)
# control t=2
points(jitter(post, 0.6),
       y_post[z == 0],
       col = alpha("steelblue", 0.5),
       pch = 20)
# treatment t=2
points(jitter(post, 0.6),
       y_post[z == 1],
       col = alpha("darkred", 0.5),
       pch = 20)
# lineas
lines(c(0, 1), c(mean(y_pre[z == 1]), mean(y_post[z == 1])), col = "darkred")
lines(c(0, 1), c(mean(y_pre[z == 0]), mean(y_post[z == 0])), col = "steelblue")
lines(c(0, 1), c(mean(y_pre[z == 1]), mean(y_post[z == 0]) +
(mean(y_pre[z == 1])-mean(y_pre[z == 0]))), col = "darkred", lty = 2)
```



Ahora calculemos τ a mano:

```
mean(y_post[z == 1]) - mean(y_pre[z == 1]) -
(mean(y_post[z == 0]) - mean(y_pre[z == 0]))

## [1] 4.371355
```

Tambien podemos usar Equation 1 para calcular τ (que recuerda es igual a β_1):

```
lm(I(y_post - y_pre) ~ z)

##
## Call:
## lm(formula = I(y_post - y_pre) ~ z)
##
## Coefficients:
## (Intercept)            z
##       1.752        4.371
```

Nota que z es el *estado* del tratamiento, basicamente un vector x de 0's y 1's, y τ es el *efecto* "causal" asociado a la administración de z. Es análogo a decir $\beta_1 x_1$.

```
knitr::purl('FE_DifDif.Rnw')

## Error in parse_block(g[-1], g[1], params.src, markdown_mode):  Duplicate chunk label
'setup', which has been used for the chunk:
## if (!require("pacman")) install.packages("pacman"); library(pacman)
## p_load(knitr)
## set.seed(2020)
## options(scipen=9999999)
## if (!require("pacman")) install.packages("pacman"); library(pacman)

Stangle('FE_DifDif.Rnw')

## Writing to file FE_DifDif.R
```

References

Card, David and Alan Krueger (1994). "Minimum Wages and Employment: A Case Study of the Fast-Food Industry in New Jersey and Pennsylvania." In: *The American Economic Review* 84.4, pp. 772–793.