

# Diferencias en diferencias

Junio, 2017

Oficina de Planificación Estratégica y Efectividad en el Desarrollo



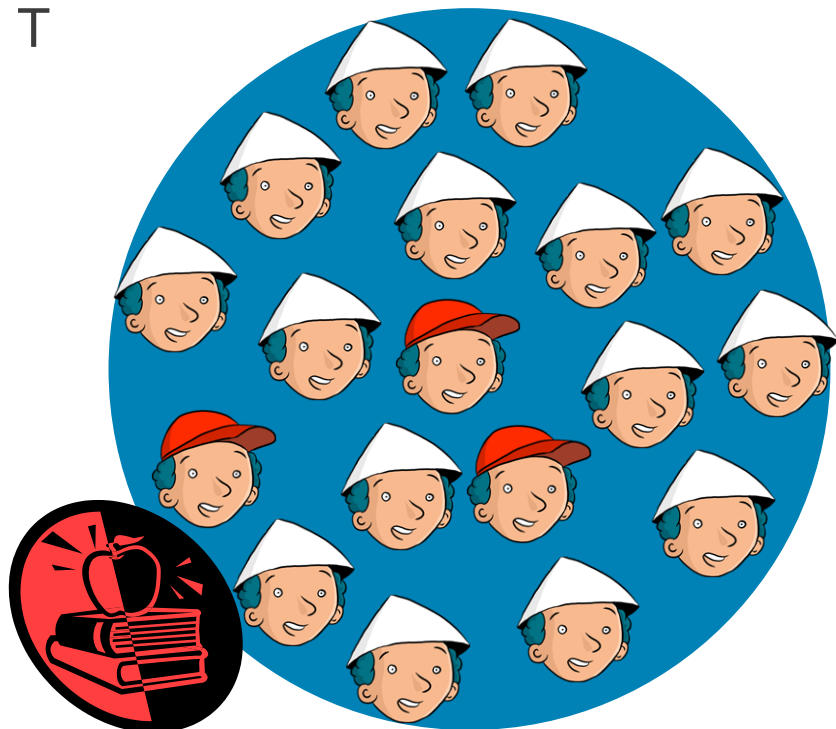
# Organización

- I. Intuición
- II. Definición
- III. Ejercicio
- IV. Cohortes sintéticos

# **I. Intuición**

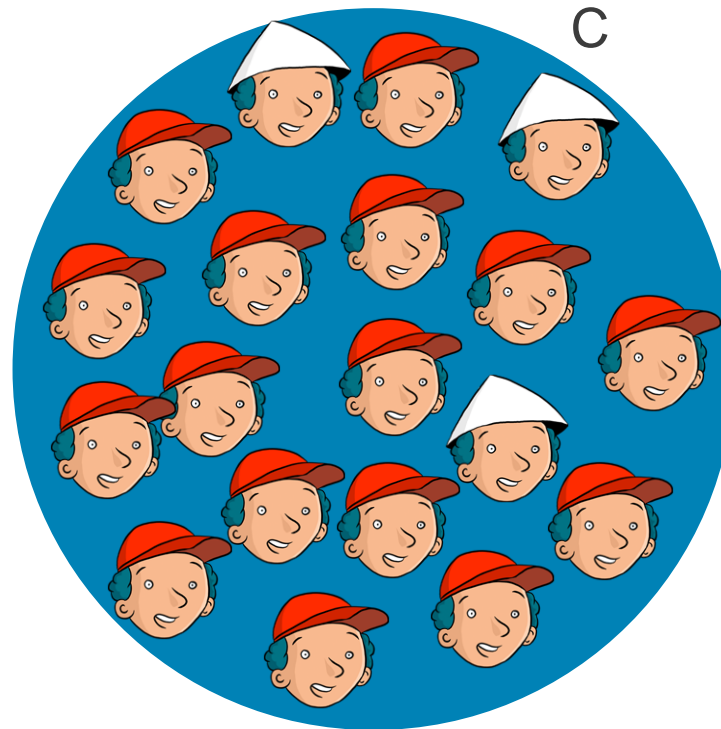
# Problema de evaluación

T



$$E[Y_i|T] = 630$$

C



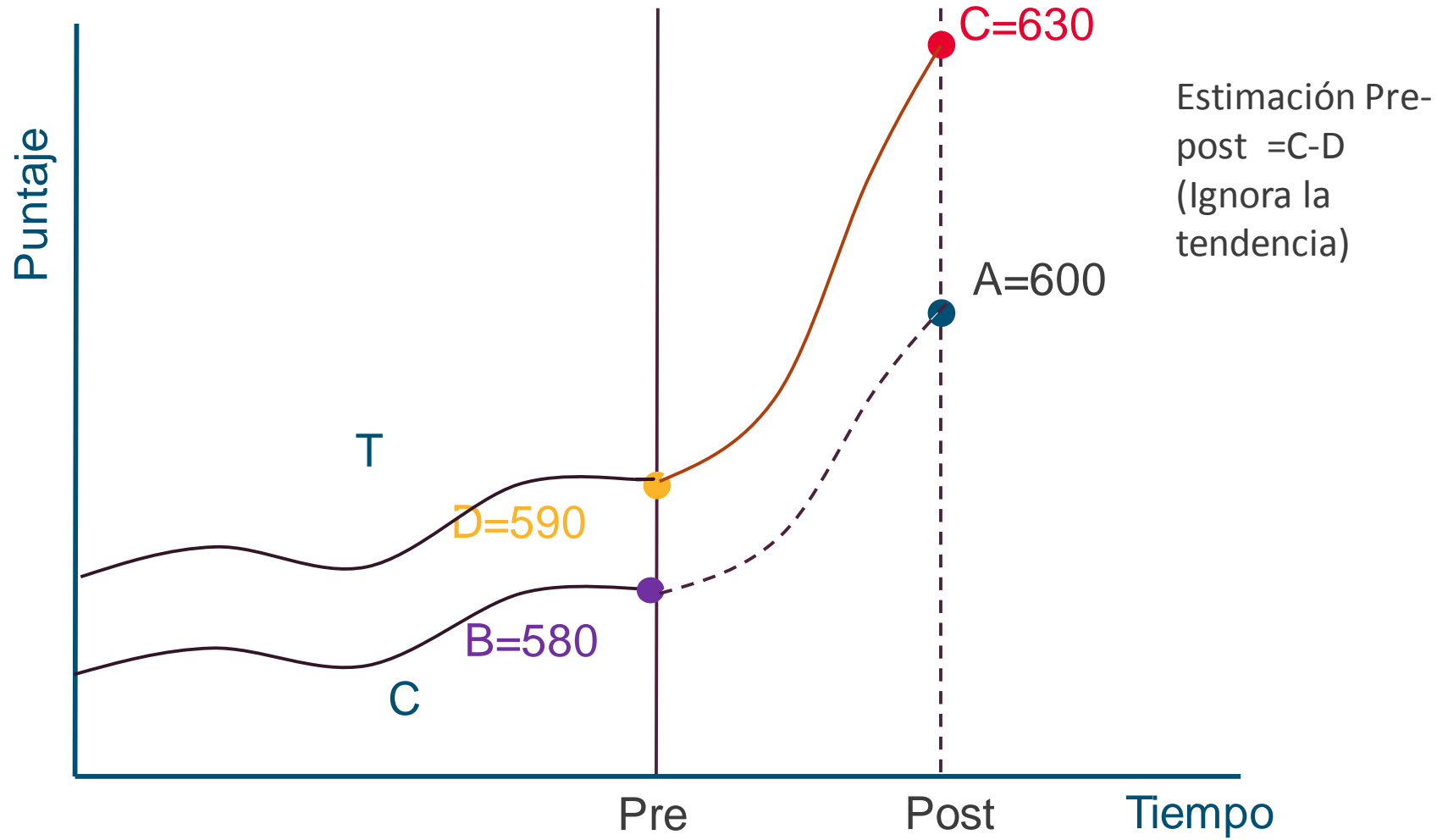
$$\hat{E}[Y_i|C] = 600$$

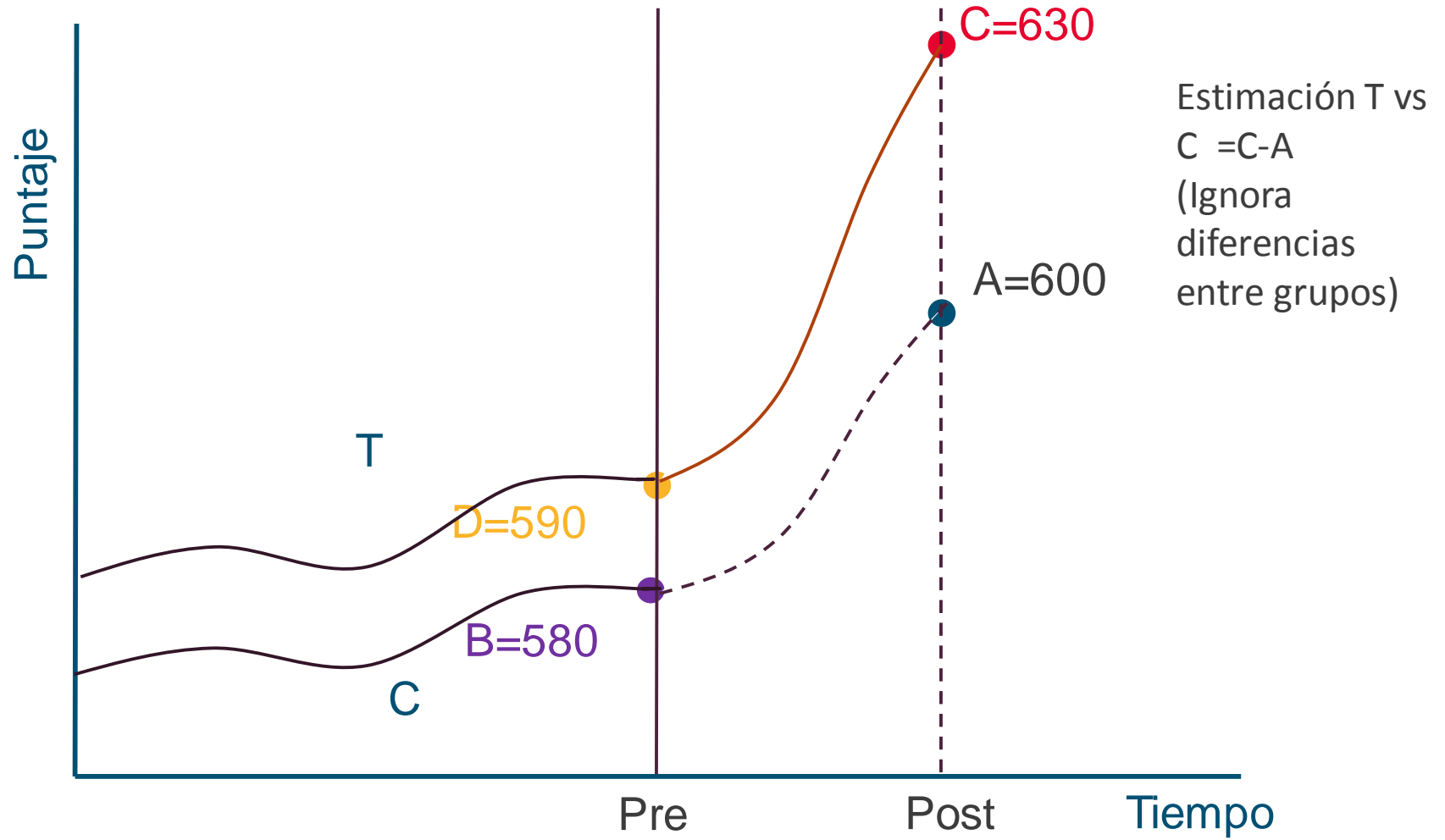
IMPACTO=630-600=30 puntos

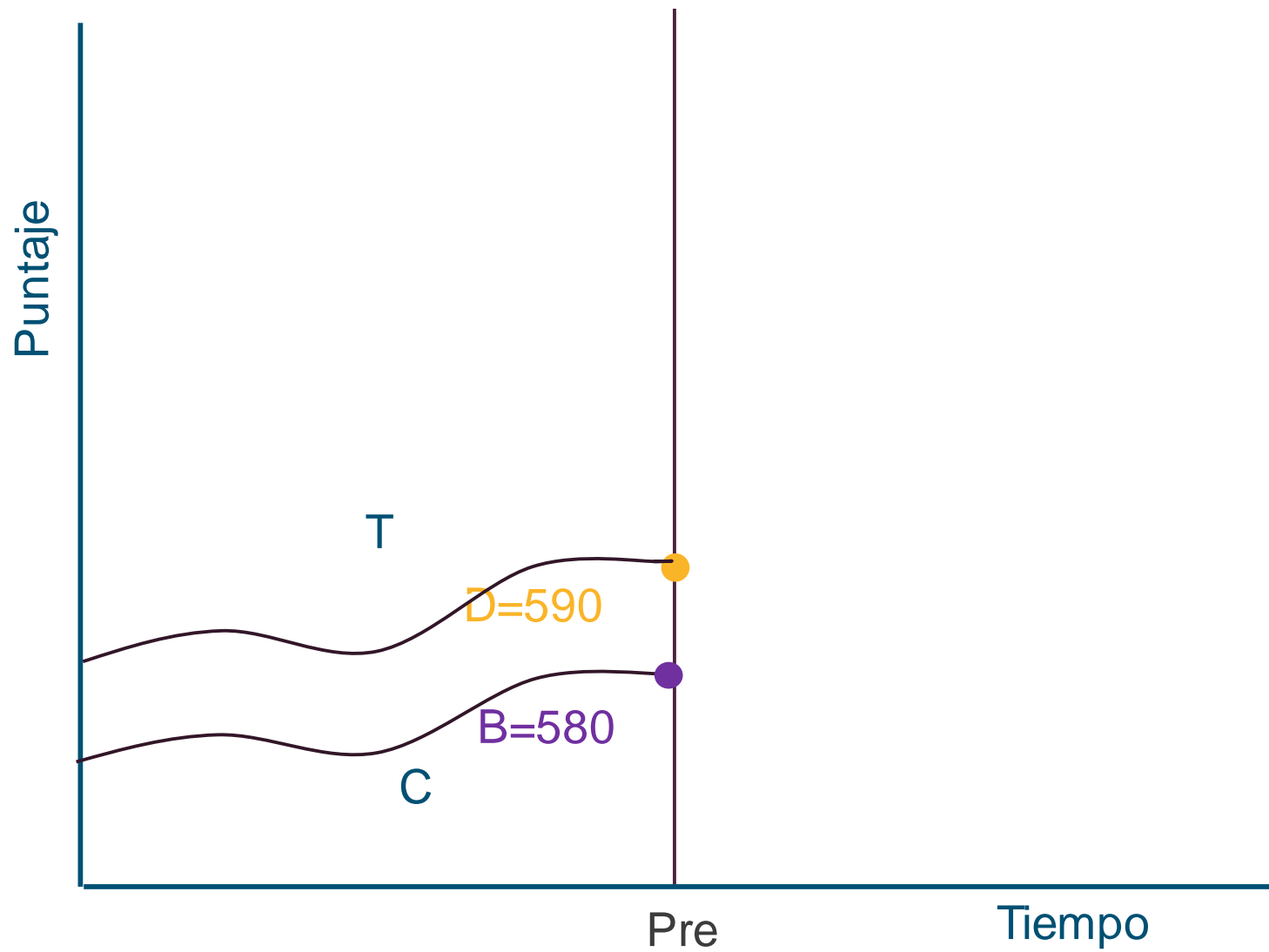
# Reto de evaluación

$$\begin{aligned}\beta &= E[Y_i^T | T] - E[Y_i^C | C] \\ &= E[Y_i^T | T] - E[Y_i^C | T] - E[Y_i^C | C] + E[Y_i^C | T] \\ &= \underbrace{E[(Y_i^T - Y_i^C) | T]}_{\text{Efecto por tratamiento}} + \underbrace{E[Y_i^C | T] - E[Y_i^C | C]}_{\text{Sesgo de selección}}\end{aligned}$$

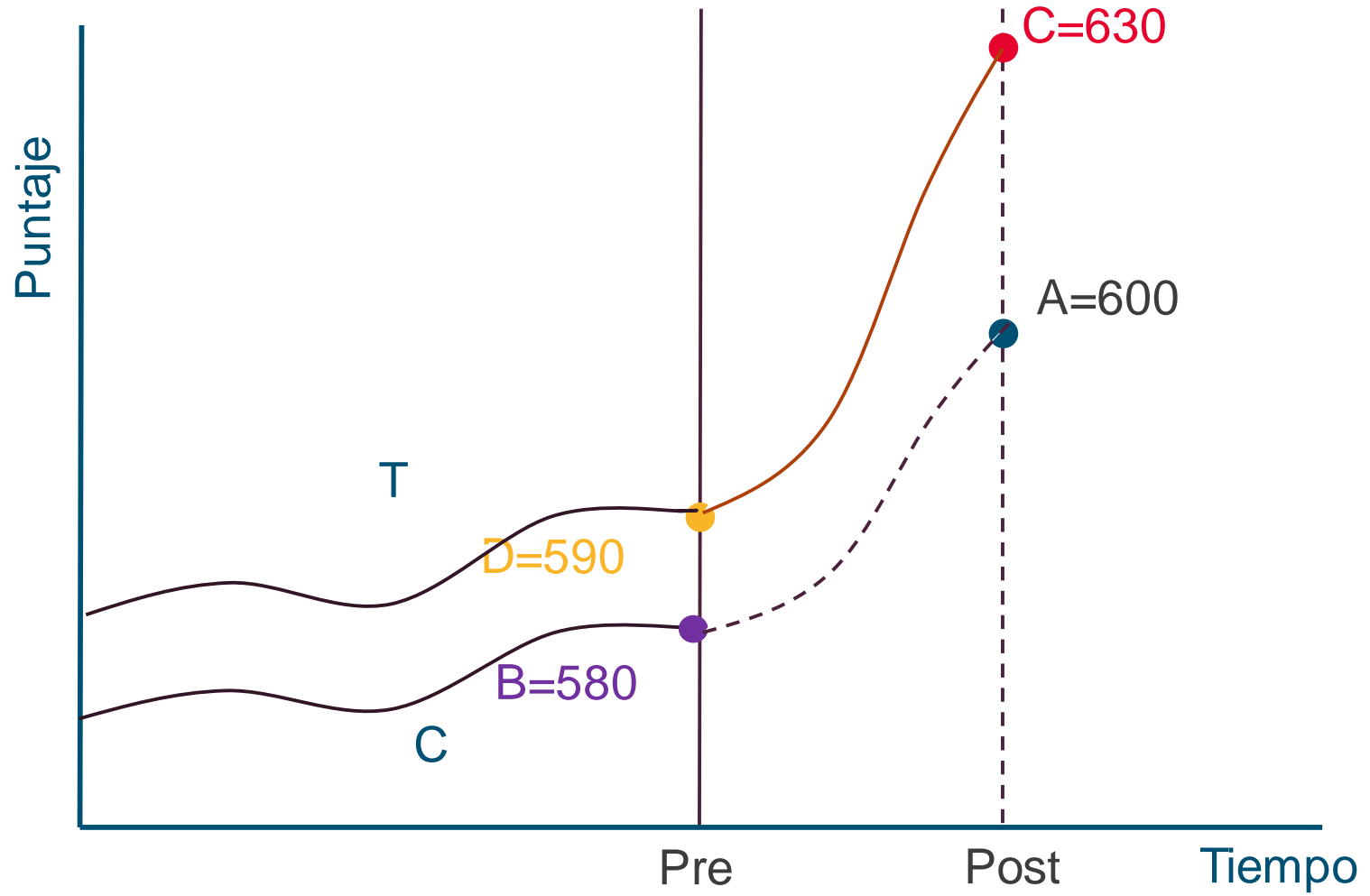
Supongamos que hay datos un periodo ...Cómo nos puede ayudar esta información?



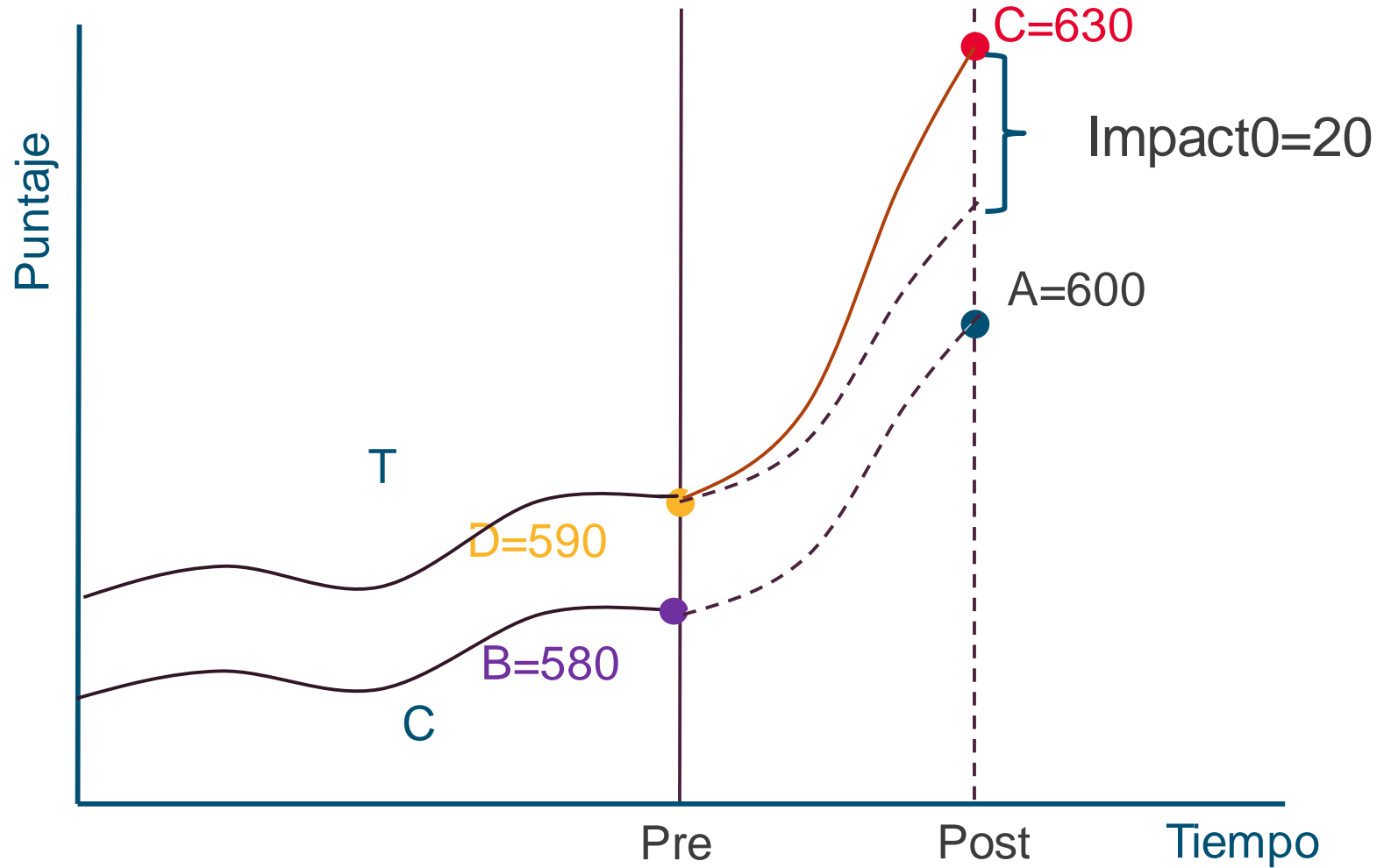


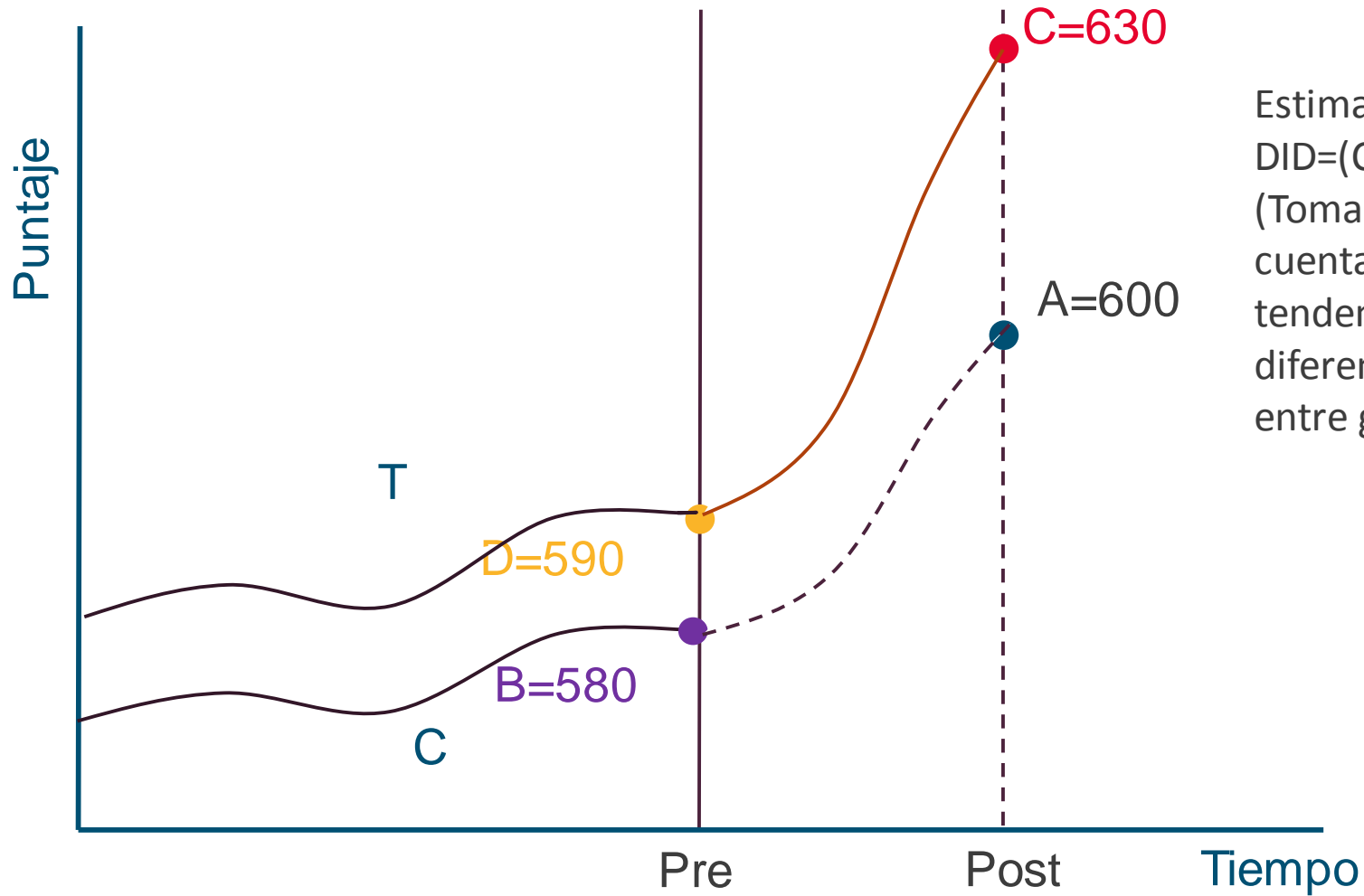






$$\text{Impacto} = (C - A) - (D - B)$$





Estimación  
 $DID = (C - A) - (D - B)$   
(Toma en cuenta las tendencias y las diferencias entre grupos)

$$Y_i = a + bT_i + cX_i + e_i$$

$$ATE = E[Y_i|T_i = 1] - E[Y_i|T_i = 0]$$

$$C = E[Y_i|T_i = 1] = a + b + cE[X_i|T_i = 1] + E[e_i|T_i = 1]$$

$$A = E[Y_i|T_i = 0] = a + 0 + cE[X_i|T_i = 0] + E[e_i|T_i = 1]$$

$$ATE = b + c(E[X_i|T_i = 1] - E[X_i|T_i = 0])$$

Supongamos tenemos un periodo antes de la intervención

$$D = E[Y_{i0}|T_i = 1] = a + b(0) + cE[X_{i0}|T_{i1} = 1]$$

$$B = E[Y_{i0}|T_i = 0] = a + 0 + cE[X_{i0}|T_{i1} = 0]$$

$$E[Y_{i0}|T_i = 1] - E[Y_{i0}|T_i = 0] = c(E[X_{i0}|T_{i1} = 1] - E[X_{i0}|T_{i1} = 0])$$

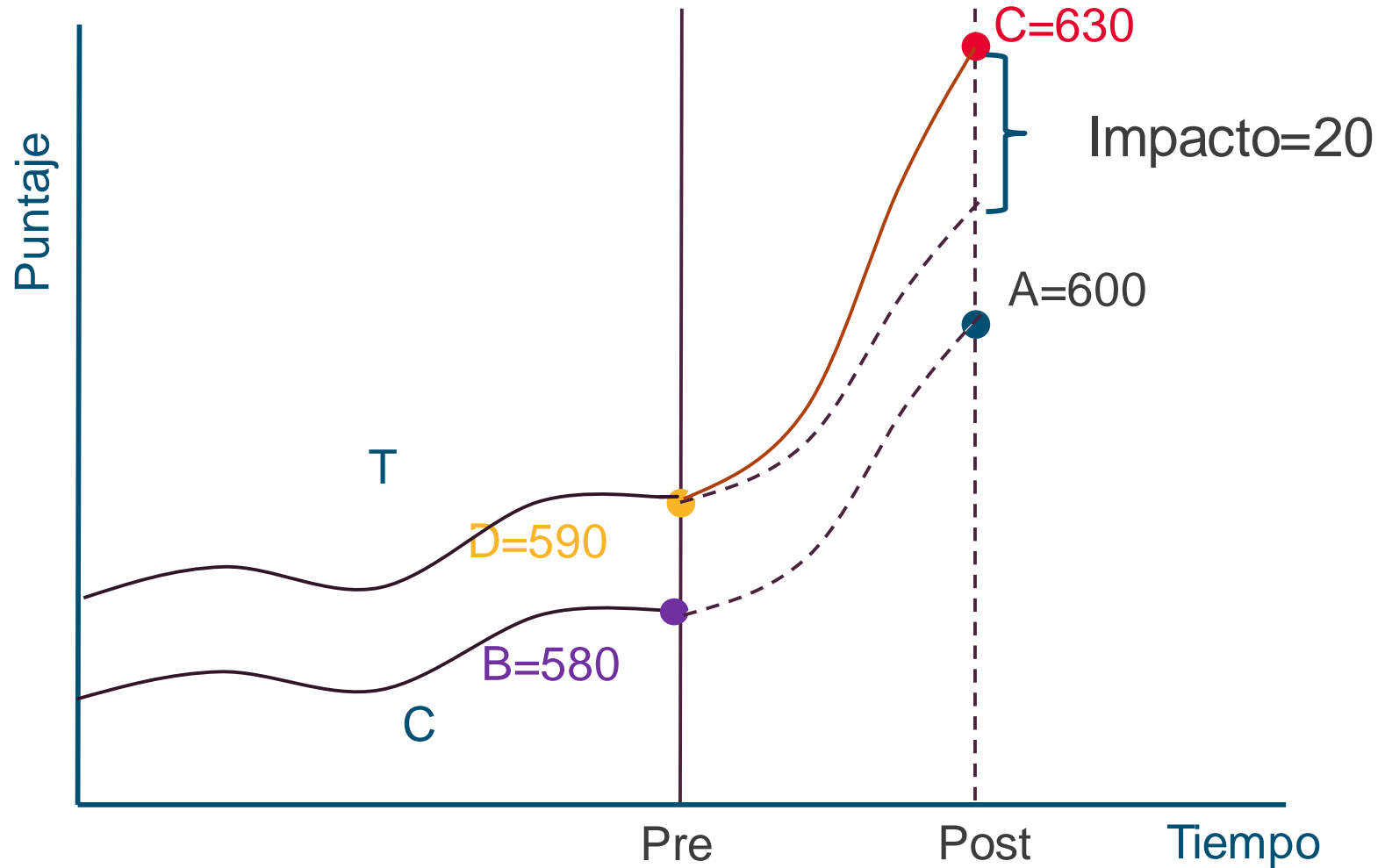
Y ahora, DID:

$$E[Y_i|T_i = 1] - E[Y_i|T_i = 0] - (E[Y_{i0}|T_i = 1] - E[Y_{i0}|T_i = 0])$$

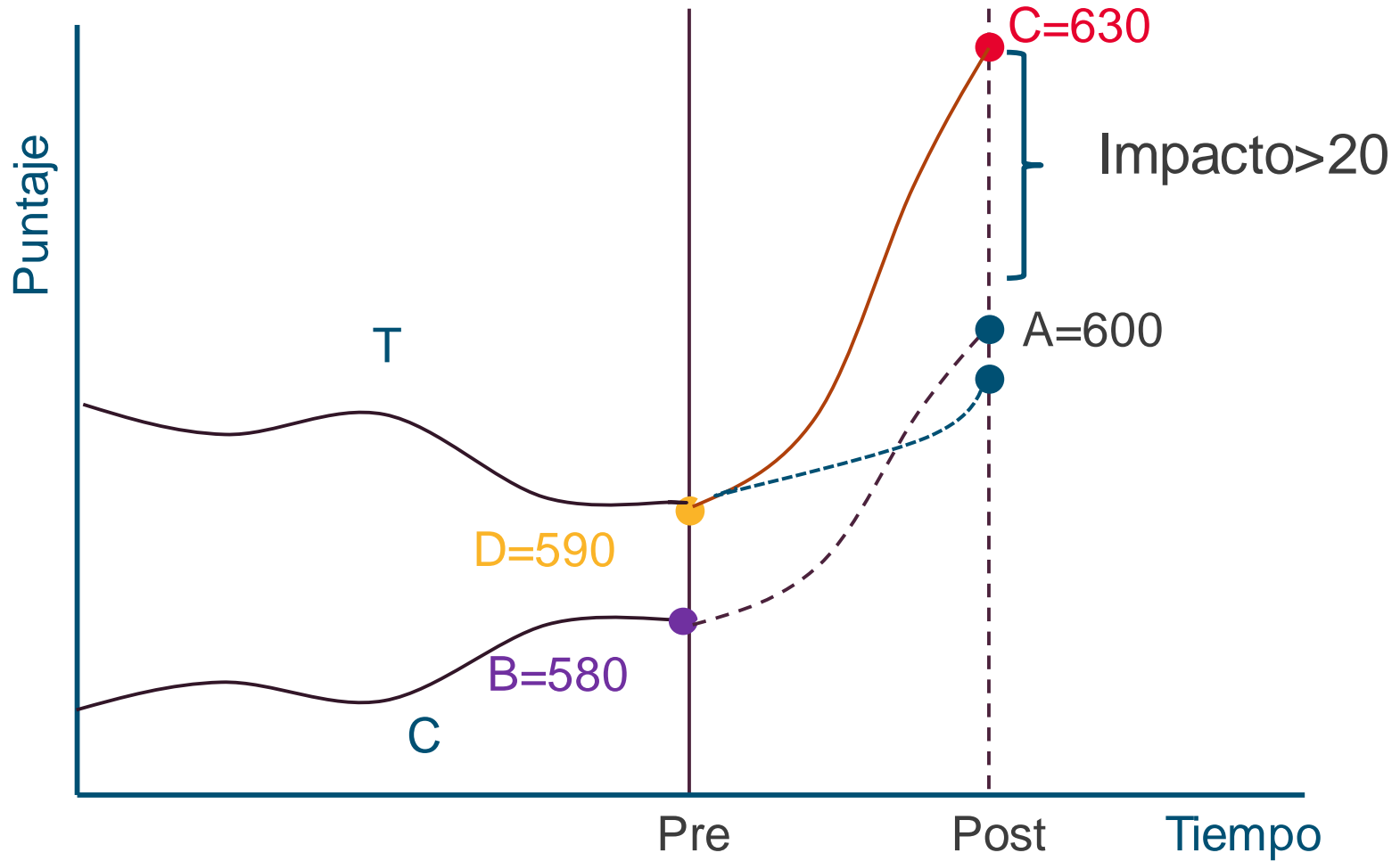
$$= b + c(E[X_i|T_i = 1] - E[X_i|T_i = 0]) - c(E[X_{i0}|T_{i1} = 1] - E[X_{i0}|T_{i1} = 0])$$

$$= b$$

$$\text{Impacto} = (C-A)-(D-B)=(C-D)-(A-B)$$

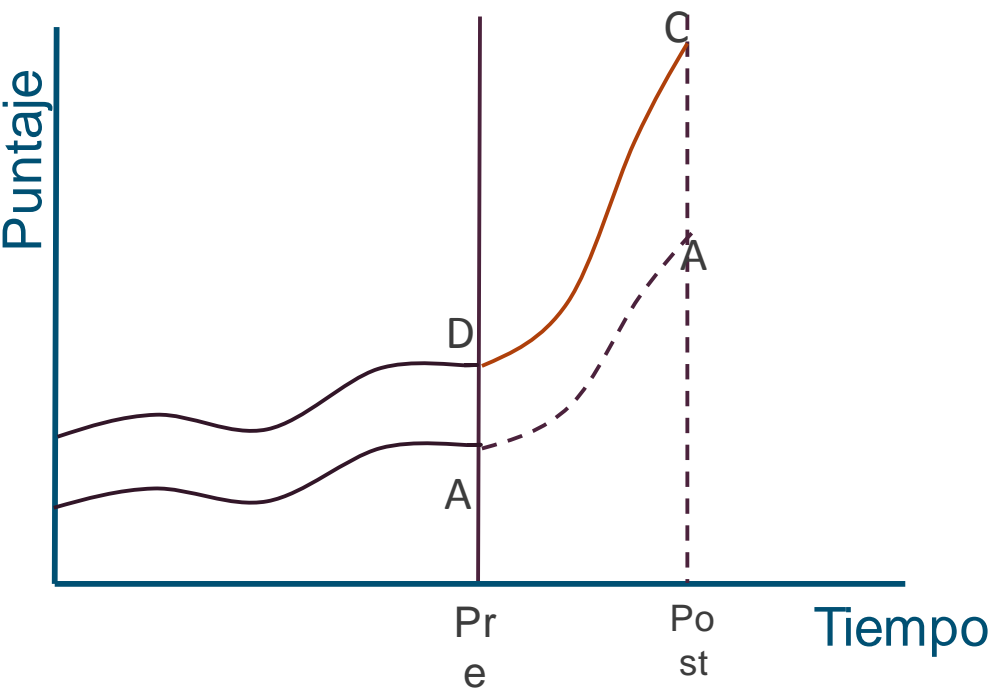


# ***Supuesto clave: Tendencias similares antes de la intervención***



# Notación alternativa

$$Y_i = \alpha + \phi T_i + \eta D2_i + \beta T_i D2_i + \epsilon_i$$



$$C = E[Y|T = 1, D2 = 1] = \alpha + \phi + \eta + \beta$$

$$A = E[Y|T = 0, D2 = 1] = \alpha + \eta$$

$$D = E[Y|T = 1, D2 = 0] = \alpha + \phi$$

$$B = E[Y|T = 0, D2 = 0] = \alpha$$

Donde  $D2 = \begin{cases} 1 & \text{si post} \\ 0 & \text{si no} \end{cases}$

$T = \begin{cases} 1 & \text{si tratamiento} \\ 0 & \text{if no} \end{cases}$

## **II. Definición**



Sea  $Z$  notación para controles

$$Y_{it} = \alpha + \phi T_{it} + \eta D2_{it} + \beta T_{it} D2_{it} + \gamma Z_{it} + \epsilon_{it}$$

$$Y_{it-1} = \alpha + \phi T_{it-1} + \eta D2_{it-1} + \beta T_{it-1} D2_{it-1} + \gamma Z_{it-1} + \epsilon_{it-1}$$

---

$$\Delta Y_{it} = \eta + \beta T_{it} + \Delta Z_{it} + \Delta \epsilon_{it}$$

Recuerda  $\Delta T_{it} = T_{i1} - T_{i0} = T_{i1} - 0 = T_{i1}$

Estimado insesgado si

$$T_i, \Delta Z_{it} \perp \Delta \epsilon_{it}$$

Estimador eficiente si

$$V[\varepsilon_{it}] = \sigma^2$$

Pero lo mas probable es que

$$\varepsilon_{it} = \rho \varepsilon_{it-1} + u_{it}$$

Así que supongamos

$$COV[\varepsilon_{it}, \varepsilon_{is}] \neq 0$$

$$COV[\varepsilon_{it}, \varepsilon_{js}] = 0$$

Sea  $\mathbf{x}=[T, D2, TD2]$  y  $\mathbf{x}$  denota controles, entonces

Se necesita un número grande de “clusters” (>50) para estimar la matriz de varianza-covarianza

$$\hat{V}[\hat{\beta}_{OLS}] = (\mathbf{x}'\mathbf{x})^{-1}(\sum_{i=1}^N \mathbf{u}'_i \mathbf{u}_i)(\mathbf{x}'\mathbf{x})^{-1}$$

$$\mathbf{u}_i = \sum_{t=1}^T \varepsilon_{it} \mathbf{x}_{it}$$

# Tips prácticos para DID

- Muestra evidencia de que las tendencias no son diferentes antes del tratamiento (obligado!)
  - Prueba de diferencias en tendencias antes del T
  - Cuando hay más de dos periodos, haz pruebas para ver si se puede predecir el tratamiento
  - Encuentra evidencia de que los grupos tampoco difieren en niveles (útil, pero no necesario)
- Restrinje la muestra a grupos que sean similares y donde te preocupa que hay diferencias en tendencias.
  - Combina con pareamiento
  - Combina con DID en otras dimensiones
  - Combina con otros métodos
- Cuando hay más de dos periodos haz pruebas placebo
  - Pretende que los datos post-intervención no existen y revisa si encuentras efectos en periodos antes de la intervención

### **III. Ejercicio**

# Resumen

El supuesto clave para DID

Las tendencias de los grupos de T y C en ausencia del tratamiento hubieran sido iguales en ausencia del tratamiento.

$$T_i, \Delta Z_{it} \perp \Delta \epsilon_{it}$$

Requiere muchas observaciones para poder estimar correctamente la matriz de varianza-copvarianza (conglomerados)

# Resumen

- **DID:** cambio pre-post entre un grupo participante y otro no-participante
- Asume que en la ausencia de programa, **tendencias** entre participantes y no-participantes serian **iguales**

## **IV. Synthetic cohorts**

# Los ingredientes

- Muestra: Datos de panel
- Sección transversal:  $J$  unidades
  - $j=1$  es la tratada
  - $j=\{2, \dots, J\}$  son las no tratadas (donantes)
- Serie de tiempo:  $T$  periodos
  - $t=\{1, \dots, T\}$
  - $T_0$  periodos antes de la intervención
  - $T_1$  periodos después de la intervención
  - $T=T_0+T_1$
- Características  $Z$  en el periodo pre-intervención



El contrafactual:

$$Y_{it}(0) = \sum_{j \in \mathcal{C}} w_j^* Y_{jt}$$

Con ponderadores tales que:

$$\sum_{j=2}^{J+1} w_j^* \bar{Y}_j^{K_1} = \bar{Y}_1^{K_1}, \dots, \sum_{j=2}^{J+1} w_j^* \bar{Y}_j^{K_M} = \bar{Y}_1^{K_M}.$$

$$\sum_{j=2}^{J+1} w_j^* Z_j = Z_1$$

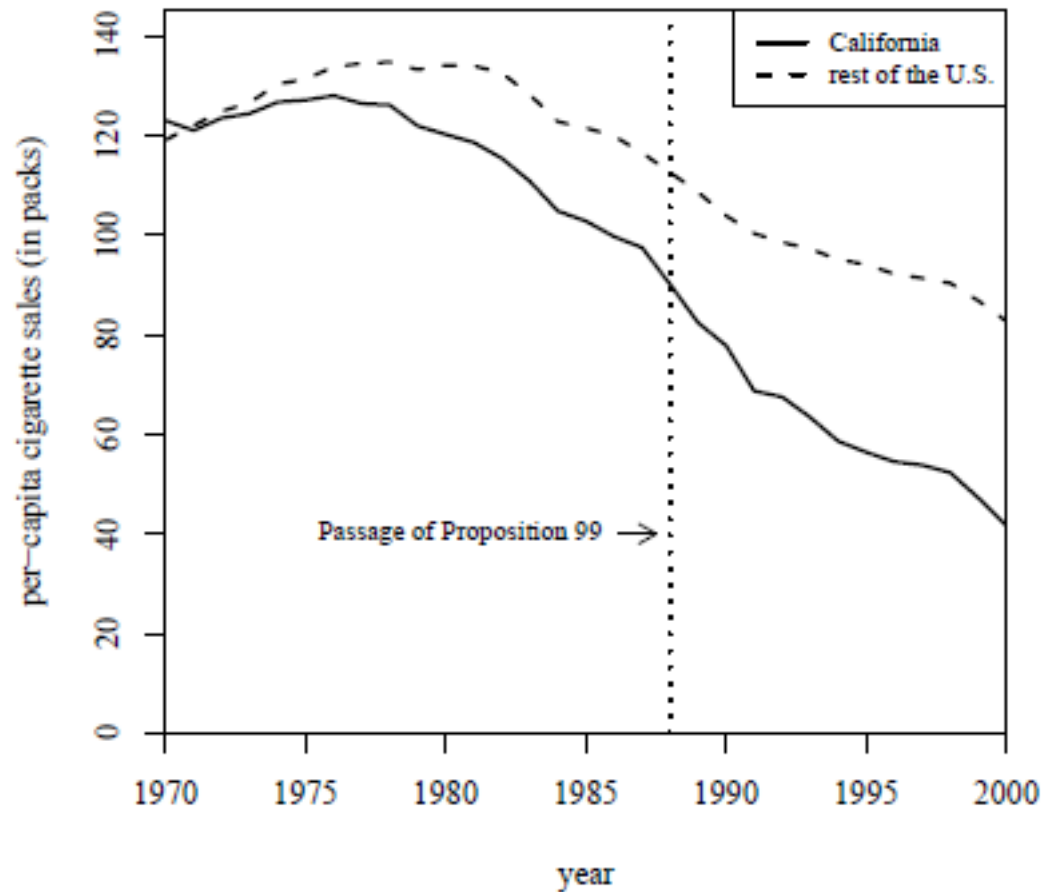
El impacto se estima como:

$$\hat{\alpha}_{1t} = Y_{1t} - \sum_{j=2}^{J+1} w_j^* Y_{jt}$$

**Abadie, A., A. Diamond, y J. Hainmueller (2010) “Synthetic Control Methods for Comparative Case Studies: Estimating the Effect of California's Tobacco Control Program,” Journal of the American Statistical Association**

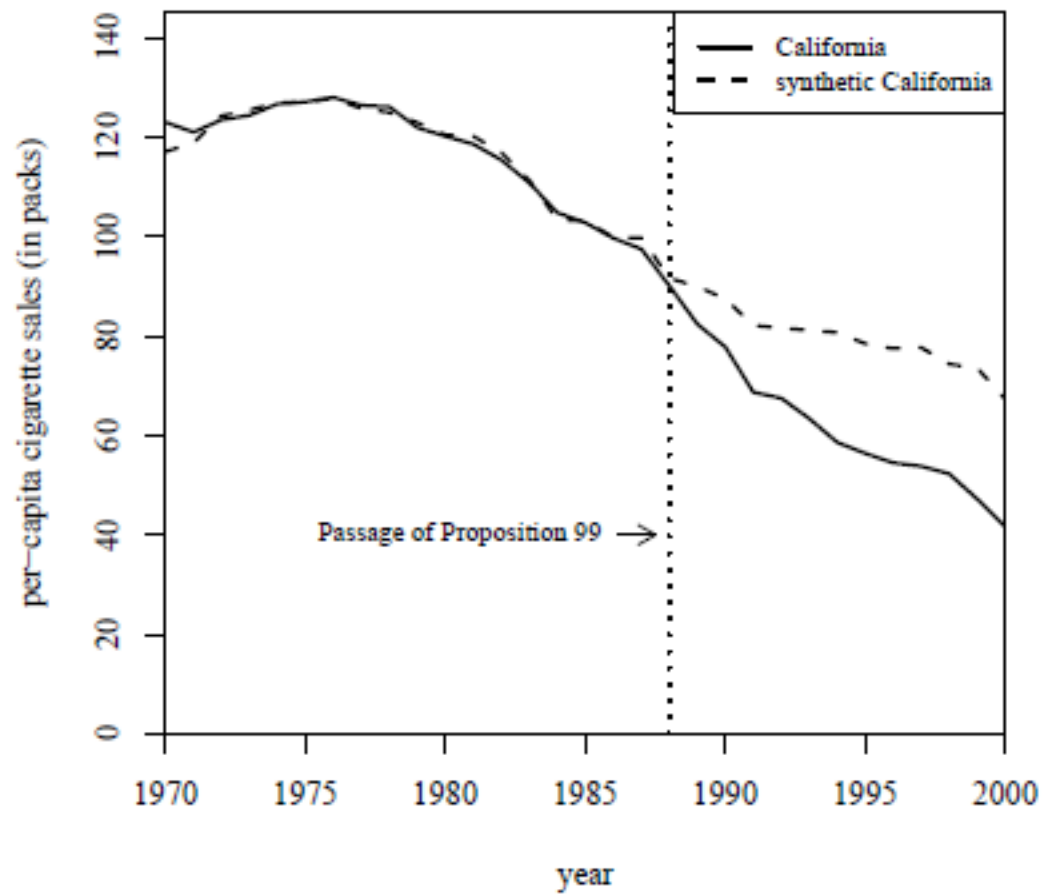
En 1988, California pasó una nueva ley sobre Tabaco:

1. Aumentó el impuesto en 25 centavos/paquete
2. Destinó los fondos recaudados por el impuesto a programas de salud y anti-Tabaco.
3. Financió campañas anti-tabaco en los medios
4. Estimuló normas jurídicas sobre aire limpio
5. Produjo mas de \$100 millones anuales en proyectos contra el tabaco



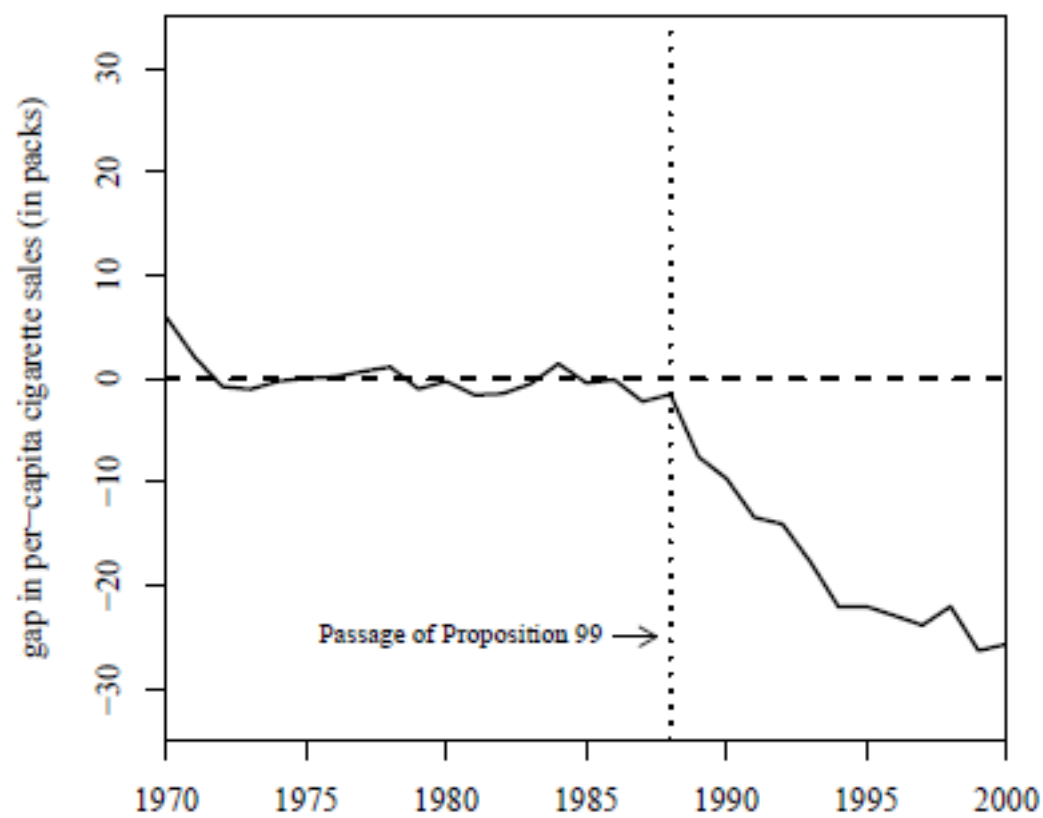
Estados con programas que habian pasado programas similares se excluyen del “resto”: AK, AZ, FL, HA, MA, MD, MI, NJ, NY, OR, WA, DC





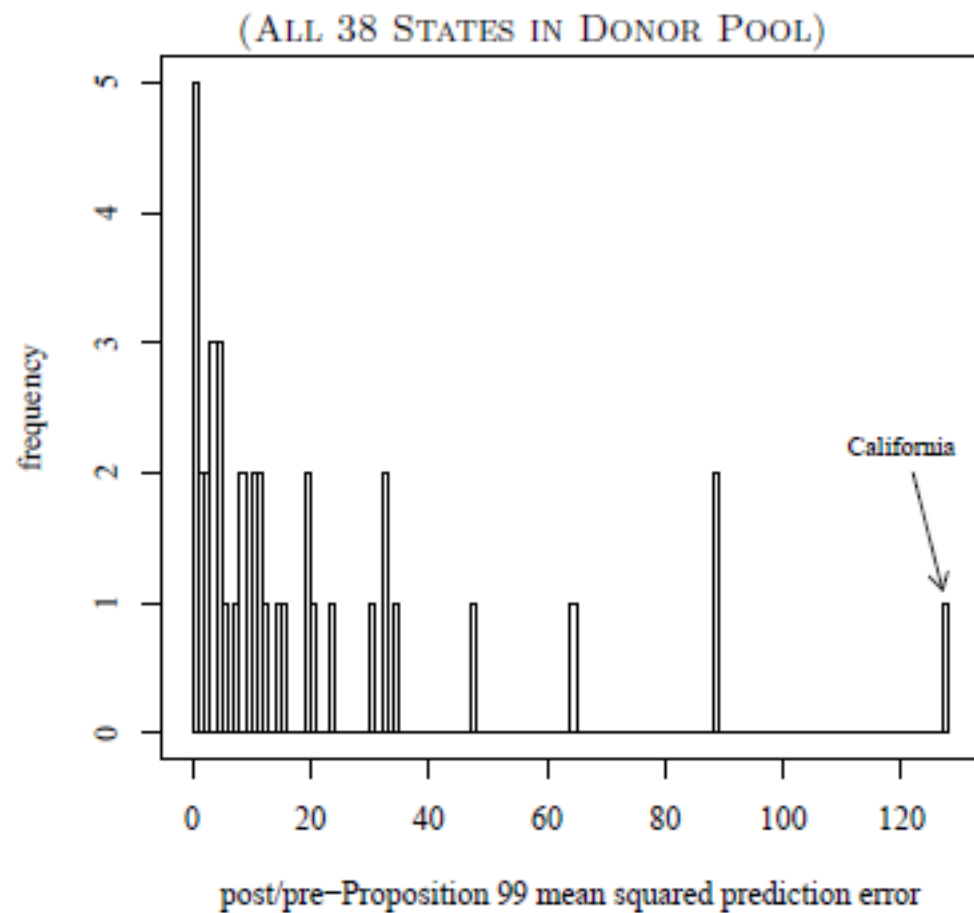
Variables	California		Average of 38 control states
	Real	Synthetic	
Ln(GDP per capita)	10.08	9.86	9.86
Percent aged 15-24	17.40	17.40	17.29
Retail price	89.42	89.41	87.27
Beer consumption per capita	24.28	24.20	23.75
Cigarette sales per capita 1988	90.10	91.62	114.20
Cigarette sales per capita 1980	120.20	120.43	136.58
Cigarette sales per capita 1975	127.10	126.99	132.81

*Note:* All variables except lagged cigarette sales are averaged for the 1980-1988 period (beer consumption is averaged 1984-1988).



# Cálculo de errores

- Permutar para obtener placebos
- El proceso permite estimar intervalos de confianza
- El valor  $p$  se calcula como el porcentaje de diferencias mayores a la estimada sobre la asignación de tratamiento real.





# Referencias

Abadie, A., A. Diamond, y J. Hainmueller (2010) "Synthetic Control Methods for Comparative Case Studies: Estimating the Effect of California's Tobacco Control Program," Journal of the American Statistical Association

Bertrand, M., Duflo, M. and Mullainathan, S. (2004) "How Much Should We Trust Differences-in-Differences Estimates?"; Quarterly Journal of Economics, 119(1), pp. 249-75.

Cameron, C. and Trivedi, P. (2005) Microeconometrics. Methods and Applications, Cambridge University Press. Chapter 21.

Duflo, E. (2001) "Schooling and Labor Market Consequences of School Construction in Indonesia: Evidence form an Unusual Policy Experiment". The American Economic Review. Vol. 91, No. 4. September, pp. 795-813.

Gertler P., Martinez, M., Premand, P., Rawlings, L. and Vermeersch, C. (2011) Impact Evaluation in Practice. Washington DC: The World Bank. Chapter 6.

