RCTs

Oscar Gálvez-Soriano

University of Houston
Department of Economics

November, 2021

Tipos de datos

- Corte transversal, series de tiempo y datos panel.
- Datos experimentales y no-experimentales (observables).

Muestras aleatorias

- Población.
- Muestra.
- Muestra aleatoria. Cada observación se obtiene aleatoriamente de la población.
- Muestreo estratificado.

Evaluación de programas

- Efecto de intervenciones y tratamientos en un resultado.
- Evaluación de programas sociales.
  - Proceso de evaluación vs. evaluación de impacto.
- Validez interna vs. validez externa.

¿Los hospitales hacen a las personas más sanas?

Grupo	Tamaño de muestra	Salud (promedio)	SE
Hospital	7,774	3.21	0.01423
No hospital	90,049	3.93	0.00335

- La differencia de medias es -0.72, con un SE 0.0146.
- ¿Este resultado tiene una interpretación causal?

El problema de selección

Definamos a la variable de tratamiento como

$$D_i = \{0, 1\}$$

La variable de interés la denotaremos como y_i.

Los resultados potenciales son

$$y_i = \begin{cases} Y_{1i} & D_i = 1 \\ Y_{0i} & D_i = 0 \end{cases}$$

$$y_i = Y_{0i} + (Y_{1i} - Y_{0i})D_i$$

El problema es que sólo observamos Y_0 o Y_1 para un individuo en particular.

El problema de selección

Podemos escribir la diferencia de y_i entre tratados y no-tratados como la suma de dos términos:

$$E\left[y_i|D_i=1\right]-E\left[y_i|D_i=0\right]=E\left[Y_{1i}|D_i=1\right]-E\left[Y_{0i}|D_i=0\right]$$

$$=E\left[Y_{1i}|D_i=1\right]-E\left[Y_{0i}|D_i=1\right]+E\left[Y_{0i}|D_i=1\right]-E\left[Y_{0i}|D_i=0\right]$$

Los dos primeros términos de la última igualdad se conoce como "Average Treatment Effect on the Treated (ATE)". Los últimos dos términos se conocen como el "Sesgo de Selección" (selection bias).

El problema de selección

Una **asignación aleatoria** resuelve el **problema de selección**.

$$E\left[y_i|D_i=1\right]-E\left[y_i|D_i=0\right]$$

$$=E\left[Y_{1i}|D_i=1\right]-E\left[Y_{0i}|D_i=1\right]+E\left[Y_{0i}|D_i=1\right]-E\left[Y_{0i}|D_i=0\right]$$

Ejemplo: Programa de entrenamiento laboral

Lalonde, Robert (1986). "Evaluating the Econometric Evaluations of Training Programs Using Experimental Data." American Economic Review, 76(4): 604-620.

- RQ: Los trabajadores poco calificados tienen problemas para encontrar trabajo. ¿Capacitarlos incrementa su salario?
- ¿Podemos comparar participantes y no participantes del programa?

Ejemplo: Programa de entrenamiento laboral

- Tenemos dos muestras independientes: grupo de tratamiento y grupo de control.
- El grupo de tratamiento tiene: 297 observaciones, con ingreso promedio $5,976 (SE $402).
- El grupo de control tiene: 425 observaciones, con ingreso promedio $5,090 (SE $277).
- Efecto promedio $(y_1 - y_0)$: $5,976-$5,090=$886.
- SE de la diferencia:

$$SE(\bar{y}_1 - \bar{y}_0) = [var(\bar{y}_1) + var(\bar{y}_0)]^{1/2}$$

$$SE(\bar{y}_1 - \bar{y}_0) = \left[\frac{var(y_1)}{n_1} + \frac{var(y_0)}{n_0}\right]^{1/2} = 488$$

Análisis de Regresión con experimentos

Podemos estimar el "Average Treatment Effect" usando una regresión. Recordemos que:

$$y_i = Y_{0i} + (Y_{1i} - Y_{0i})D_i$$

Podríamos reescribir esta ecuación como:

$$y_i = \beta_0 + \beta_1 D_i + u_i$$

Al estimar esta ecuación con MCO, el coeficiente β_1 puede ser interpretado como el "Average Treatment Effect".

Veamos ejemplos en STATA.