Selección en observables

En otras palabras: el término de error (el efecto no observable en Y) no presenta diferencias entre el grupo tratado y no tratado

El error es igual a si fue tratado o no ⇒ Media condicional → INDEPENDENCIA

$E(u_i|T_i=1) = E(u_i|T_i=0)$

$E(Y_{1i}|T_i=1) - E(Y_{0i}|T_i=1) + [E(Y_{1i}|T_i=0) - E(Y_{0i}|T_i=0)]$ (subestimar; sobre estimar)

si tengo S.S

SESGO DE SELECCION

Factores no observados no están relacionados con si se trató o no

Selección en observables

Tomemos la ecuación de la regresión

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

- Comparación de observables no ofrece buena medida pues tendré un sesgo de selección asociado

[recordemos que los valores esperados de Y_i dependerán no solo del tratamiento si no que de más cosas]

ALEATORIZACIÓN en un experimento controlado

Sin experimento → **SELECCION OBSERVABLES**

Factores observados

u correlacionado a X_i

Y_i depende de la relación de X_i y T_i

$E(Y_{1,1}|T_i=1, X_i) = E(Y_{i,1}|T_i=0, X_i)$

$E(Y_{1,1}|T_i=1, X) - E(Y_{0i}|T_i=0, X) =$ EFECTO CAUSAL PROMEDIO

- Si no controlamos por X_i, se va al término de error
- Clave: determinar cuáles X_i incorporar

Omisión de variable relevante

Poco frecuente que tengamos todas las variables que determinan un X_i tal que controlamos todos los factores contenido en los residuos.

→ Si $corr(u, X_i) \neq 0$ — sesgo de selección
→ Si $E(u|T_i) \neq 0$

Imaginemos el modelo real ---> Tenemos dos set de variables independientes (X_1 y X_2 vectores de variables)

$$y = X_1\beta_1 + X_2\beta_2 + u \quad *$$

El modelo que estimamos (excluye X_2)

$$y = X_1\tilde{\beta}_1 + u$$

$\tilde{\beta}_1 = (X_1'X)^{-1}X_1'X_2\beta_2$ — SESGO DE VARIABLE OMITIDA

- Regresión de las variables en X_2 parcializadas y luego puestas en conjunto en matriz de todas las variables X_1
- Efecto de las X_2 sobre y (tenemos una intuición sobre él)

Esto solo será cero si X_1 no está linealmente relacionado a las variables en X_2 ($X_1 \perp X_2$ ortogonales)

$\hat{\beta}_2 = 0$ (no significativo)

(A) RELACIÓN con correlacción con X_i y X_2 → la omitida. (B) RELACIÓN CON Y_i

→ Si alguna variable omitida cumple con [A] y/o [B] → $\hat{\beta}_1$ estará sesgada

$\delta_{1,2}$
- Pendientes de X_2 sobre todas las variables explicativas
- ~ $corr(X_1, X_2)$ → Si están correlacionadas ⊕ sesgo positivo si β_2⊕ ⊖ sesgo negativo si β_2⊖ + intuición sob signo β_2
- Corr entre cualquier predictor en X_1 y predictor omitido en X_2, segará vector de $\hat{\beta}_1$
- Permite entender mejor el sesgo de selección
  - Si grupos difieren en características no observables,
  - MCO entregará estimación sesgada del impacto causal → se irá al error

DIRECCIONES DE SESGO

- Si $\tilde{\beta}_1 > \beta_1$ (pensando $\beta_2 +$) ⇒ sesgo positivo, correlación positiva } Modelo inicial sobre-estimó el efecto de X_1
- Si $\tilde{\beta}_1 < \beta_1$ (pensando $\beta_2 +$) ⇒ sesgo negativo, correlacción negativa } Modelo inicial subestimó el efecto real de X_1
- Si $\tilde{\beta}_1 < \beta_1$ (pensando $\beta_2 -$) ⇒ sesgo negativo, correlacción negativa } Modelo inicial subestimó el efecto real de X_1
- Si $\tilde{\beta}_1 > \beta_1$ (pensando $\beta_2 -$) ⇒ sesgo positivo, correlación positiva } Modelo inicial sobre-estimó el efecto de X_1

Inclusión de variables irrelevantes

Sobre-especificado cuando se incluyen variables que no forman parte del modelo poblacional (→ último dígito del rut)

$$y = \beta_0 + \beta_1 X_1 + \ldots + \beta_K X_K + \beta_{K+1} X_{K+1} + u$$

, y esperamos que $\beta_{K+1} = 0$ *

↳ $E(\hat{\beta} - \beta) = 0$ dado que ** INSESGADO

↳ Si X_{K+1} relacionado con $X_{1\ldots K}$ MULTICOLINEALIDAD → varianzas de $\hat{\beta}$

Incluso en un experimento aleatorio es mejor incluir variables control pese a que * duda potencialmente generen ** mejoran la precisión al reducir varianza del error de β_1

$$se(\hat{\beta}_1) = \sqrt{Var(\hat{\beta}_1|X)} = \frac{\sigma^2}{\sqrt{SCT_1 \cdot (1 - R_1^2)}}$$

Si empeora a crecer al incorporar variables toda vez que estén esas

- también más mediciones

cada vez que estén esas variables relacionadas