Exogeneidad y Causalidad en Regresion

Walter Sosa-Escudero

Universidad de San Andres y CONICET

Exogeneiedad

$$E(u_i|X) = 0$$

- $E(u_i) = 0, \qquad i = 1, \ldots, n$. Cuidado: no en la otra direccion!

  Prueba: por LIE $E(u) = E[E(u|X)] = E(0) = 0.]$

- $E(x_{jk}u_i) = 0, \qquad j, i = 1, \ldots, n; \;\; k = 1, \ldots, K$

  Prueba: idem anterior.

- $E(x_{ik}u_i) = 0$, ortognalidad. Si hay intercepto, con $E(u_i) = 0$ implica $Cov(x_i, u_i) = 0$. $E(u_i|x_{ik}) = 0$ es mas fuerte: u_i no correlacionada con ninguna funcion (medible) de x_i.
- $E(u|X) = 0$ implica $E(u_i X_i) = 0$. Pedimos un poquito menos para consistencia.

Ejemplo: $Y_t = \beta_1 + \beta_2 Y_{t-1} + u_t, \quad t = 1, 2, \ldots$ $E(u_t Y_{t-1}) = 0$ (consistencia), pero $E(u|Y_{-1}) \neq 0$ (por que?). MCO consistente y sesgado!

Causalidad?

$$Y_i = \alpha + \beta D_i + u_i$$

- Droga sobre temperatura corporal, dieta sobre peso corporal, AUH sobre asistencia al secundario.
- En que sentido β mide el efecto que D tiene sobre Y?
- En que sentido $\hat{\beta}$ en base a $(D_i, Y_i), i = 1, \ldots, n$ estima el efecto que D tiene sobre Y?

Causa y efecto en base a observables

- $D = 0, 1$, 'causa', 'tratamiento'.
- Notacion $D_1 \equiv (D = 1), \;\; D_0 \equiv (D = 0)$.
- Y es un resultado.
- $Y|D_1 =$ resultado observable bajo tratamiento, $Y|D_0$ si no.

Ideas:

- $Y|D_1 - Y|D_0$ en general es una medida *sesgada* del efecto causal (tratados vs no, antes y despues).
- Existen condiciones bajo las cuales $Y|D_1 - Y|D_0$ mide efectos causales

Causa y efecto en base a contrafacticos

- Resultados potenciales:

$$\begin{array}{ll} Y_0 & \text{si } D = 0 \\ Y_1 & \text{si } D = 1 \end{array}$$

  independientemente de si hubo o no tratamiento.
- Ej: Y_1 temperatura si *tomase* un analgesico. Y_0 ingreso si no *recibiese* la AUH

Efecto causal: $\beta \equiv Y_1 - Y_0$

Ej: caida en la fiebre si tomases una aspirina con respecto a que no la tomes. En terminos de diferencias entre resultados potenciales.

Inobservabilidad de contrafacticos

- Problema: se observa Y_1 o Y_0 *pero nunca ambos*.
- D implica haber eliminado una ruta observable. Ambas rutas potenciales 'existen'.
- 'El tiempo se bifurca perpetuamente hacia innumerables futuros. En uno de ellos soy su enemigo'. (J.L. Borges, en 'El jardin de senderos que se bifurcan)

Sesgo de seleccion

En la practica se observa Y

$$Y = \begin{cases} Y_1 & \text{si } D = 1 \\ Y_0 & \text{si } D = 0 \end{cases}$$

O, alternativamente:

$$Y = Y_0 + (Y_1 - Y_0)\, D$$

Sesgo de seleccion

$$
\begin{aligned}
Y|D_1 - Y|D_0 &= \left[Y|D_1 - Y_0|D_1\right] + \left[Y_0|D_1 - Y|D_0\right] \\
&= \left[Y_1|D_1 - Y_0|D_1\right] + \left[Y_0|D_1 - Y_0|D_0\right]
\end{aligned}
$$

$$
\begin{array}{ccccc}
Y|D_1 - Y|D_0 & = & \beta & + & S \\
\text{Dif Observables} & = & \text{Efecto causal} & + & \text{Sesgo}
\end{array}
$$

Sesgo por seleccion: $S \equiv Y_0|D_1 - Y_0|D_0$

- Diferencia en peso potencial sin tratamiento, entre tratados y no tratados. Quien hace dieta / toma analgesicos?
- Con datos observacionales $S \neq 0$.
- Sesgo: la comparacion entre tratados y no tratados estima el efecto causal MAS el sesgo.

Aleatorizacion al rescate

Tratamiento aleatorio: D es independiente de Y_1 y Y_0

$$\begin{aligned} Y|D_1 - Y|D_0 &= \beta + \left[Y_0|D_1 - Y_0|D_0\right] \\ E\left[Y|D_1 - Y|D_0\right] &= \beta + E\left[Y_0|D_1\right] - E\left[Y_0|D_0\right] \\ &= \beta + E\left[Y_0|D_1\right] - E\left[Y_0|D_1\right] \\ &= \beta \end{aligned}$$

Paso clave: bajo tratamiento aleatorio $E\left[Y_0|D_1\right] = E\left[Y_0|D_0\right]$

Resultado: *el tratamiento aleatorio elimina el sesgo*.

Tratamiento aleatorio?

- Experimento o cuasi experimento.
- D se mueve en forma exogena ('causa').
- Auge de la aproximacion experimental en medicina. Economia?
- Experimento: control de la variabilidad exogena.

tanto la administración empírica deberá ser conducida con sentido común. El empirismo en la medicina tiende a simplificar la interpretación de los hechos en conceptos semejantes al siguiente, atribuido a Galeno: "Todos los que toman este remedio se curan, excepto aquellos a quienes el remedio no ayuda y fallecen. Es obvio que es eficaz, salvo para los casos incurables"[11].

Félix Roberto Shardonofsky

Centro Respiratorio, Hospital de Niños, Buenos Aires

1. Milner AD, Henry RL: Acute airway obstruction in children under 5. *Thorax* 37: 641, 1982.

Regresion?

Que informacion contiene $Y_i = \alpha + \beta D_i + u_i$ en esta historia?

$$
\begin{aligned}
Y &= Y_0 + (Y_1 - Y_0)D \\
&= E(Y_0) + \beta D + \left[Y_0 - E(Y_0)\right] \\
Y &= \alpha + \beta D + u
\end{aligned}
$$

con $\alpha \equiv E(Y_0)$ y $u \equiv Y_0 - E(Y_0)$

- Supongamos que tenemos una muestra $(Y_i, D_i), i = 1, \ldots, n$
- Para que $\hat{\beta}$ sea insesgado necesitamos $E(u_i|D_i) = 0$.

$$
\begin{aligned}
E(u_i|D_i) &= E\big[Y_0 - E(Y_0) \ \mid D_i\big] \\
&= E(Y_0|D_i) - E(Y_0) \\
&= E(Y_0) - E(Y_0) \\
&= 0,
\end{aligned}
$$

ya que bajo aleatorizacon $E(Y_0) = E(Y_0|D_i)$, de modo que $\hat{\beta}$ en base a datos observables es insesgado para el efecto causal.

Conclusion: *Bajo aleatorizacion de tratamiento,* $Y = \alpha + \beta D + u$ *tiene una interpretacion causal.* $\hat{\beta}$ *es insesgado para los datos observacionales (no hace falta ver los potenciales).*

Resumiendo

- Causalidad: relacion entre contrafacuales. Uno no es observable.
- Bajo aleatorizacion de tratamiento, $Y = \alpha + \beta D + u$ tiene una interpretacion causal. $\hat{\beta}$ es insesgado.
- Rol de $E(u|D) = 0$: D varia en forma exogena.
- Relevancia del razonamiento experimental.
- Cuestion muy importante en las ciencias sociales en los ultimos tiempos.
- Big data?

Referencias

- Angrist, J. y Pischke, J., 2014, *Mastering Metrics: the Path from Cause to Effect*, Cap. 2, Princeton University Press, Princeton.
- Sosa Escudero, W., 2014, *Que es (y que no es) la Estadistica*, Siglo XXI Editores, Buenos Aires. Capitulo 3: El huevo y la gallina: causalidades y casualidades.
- Borges, J.L., 1944, El jardin de senderos que se bifurcan, en *Ficciones*, Sudamericana, Buenos Aires.

"A diferencia de Newton y de Schopenhauer, su antepasado no creia en un tiempo uniforme, absoluto. Creia en infinitas series de tiempos, en una red creciente y vertiginosa de tiempos divergentes, convergentes y paralelos. Esa trama de tiempos que se aproximan, se bifurcan, se cortan o que secularmente se ignoran, abarca todas la posibilidades. No existimos en la mayoria de esos tiempos; en algunos existe usted y no yo; en otros, yo, no usted; en otros, los dos. En este, que un favorable azar me depara, usted ha llegado a mi casa; en otro, usted, al atravezar el jardn, me ha encontrado muerto; en otro, yo digo estas mismas palabras, pero soy un error, un fantasma."

J.L. Borges, 1944, El jardin de senderos que se bifurcan

Apendice: $\hat{\beta}$ como diferencia de medias

$$Y_i = \alpha + \beta D_i + u_i, \quad i = i, \ldots, N$$

Notacion

- $T =$ tratados, $N - T =$ no tratados.
- $\bar{Y}_T$, $\bar{Y}_{N-T}$, promedios tratados y no tratados.
- $\sum_T Y_i \equiv \sum D_i Y_i, \quad \sum_{N-T} \equiv \sum (1-D) Y_i$

Resultado: $\hat{\beta} = \bar{Y}_T - \bar{Y}_{N-T}$

Prueba (aburrida): Sea M la matriz que genera errores de regresar cualquier variable en un vector de n unos (el intercepto). Por TFWL:

$$\hat{\beta} = (D'M'MD)^{-1}D'M'MY = (D'M'MD)^{-1}D'M'Y,$$

ya que M es idempotente. Es facil verificar que el $i-$esimo elemento de MD es $d_i \equiv D_i - \bar{D}$
Entonces, reemplazando:

$$\hat{\beta} = \frac{\sum d_i Y_i}{\sum d_i^2}, \quad d_i \equiv D_i - \bar{D}$$

Denominador:

$$\sum d_i^2 = \sum (D_i - \bar{D})^2 = \sum D_i^2 - N\bar{D}^2 = \sum D_i - N\ T^2/N^2 = T - T^2/N = T\ (1 - T/N)$$

Numerador:

$$\sum d_i Y_i = \bar{Y}_T\ T(1 - T/N)\ -\ \bar{Y}_{N-T}\ T(1 - T/N) = T(1 - T/N)\ \left(\bar{Y}_T - \bar{Y}_{N-T}\right)$$

Reemplazando y simplificando se obtiene el resultado.

Ejercicio: derivar $\hat{\alpha}$ para este caso.