Aleatorización

Recordemos que en este curso estudiamos herramientas que nos permitan hacer inferencia causal. Con lo que hemos visto hasta ahora...

¿Por qué no podemos hacer inferencia causal con MCO?

↳ correlación no implica causalidad

↳

En realidad sí podríamos hacer inferencia causal con una regresión lineal, sólo que necesitamos hacer un supuesto extra. Consideremos el siguiente ejemplo para ilustrar lo anterior. Nos interesa estudiar el efecto del ejercicio en el ritmo cardiaco de reposo (HR) de un individuo. Supongamos que el HR potencial de un individuo i que decide ejercitarse una cantidad e es lineal en la cantidad de ejercicio que, de hecho, se ejercita. Esto es:

$$HR_{ie} = \alpha + \beta e_i + \varepsilon_i \quad (1)$$

donde ε es el término de error que captura todos los demás elementos que podrían afectar al HR potencial.

¡Ojo! En realidad es imposible observar HR potencial y el HR potencial $\neq$ HR observado.

Recordemos que nos interesa estudiar el efecto causal del ejercicio en el HR potencial que tendría una persona si ésta se ejercitara +1 veces.

$$\delta_i = HR_{i(e+1)} - HR_{ie}$$

Si tomamos la esperanza de δ_i lo que tendríamos es el efecto causal promedio en el HR potencial. Ahora bien, como no podemos observar el HR potencial tendríamos que estimar un modelo:

$$HR_i = \alpha + \beta e_i + \varepsilon_i \quad (2)$$

¿Cuál es el problema con este modelo?

No podemos asegurar que $HR_i \perp\!\!\!\perp \varepsilon_i$, entonces, si estimamos (2) obtendríamos una β sesgada.

Una manera de estimar β insesgada es si el ejercicio se asignara de manera aleatoria. Es decir, podemos usar regresión lineal si hay aleatoriedad porque 1) elimina el sesgo de selección y 2) nos permite construir una suerte de contrafactual (que nunca observamos en la realidad).

En la práctica, es muy difícil que todo sea aleatorio, entonces recurrimos al supuesto de independencia condicional (CIA). Esto es que el HR potencial es independiente del ejercicio que deciden hacer, condicional a un set de regresores.

$HR_{ie} \perp\!\!\!\perp \varepsilon_i \mid X_i$

↳ $HR_i = \alpha + \beta e_i + X_i \delta + v_i \quad (3)$

Por construcción, en (3) v_i es ortogonal a X_i, es decir $E[v_i \mid X_i] = 0$.

Bueno, ¿y eso qué? Recordemos que queremos hacer inferencia causal entre el ejercicio y el HR potencial. Es decir nos interesa conocer la esperanza del HR potencial si i se ejercita e veces y lo estamos tratando de obtener a partir de lo que observamos que el individuo se ejercita y el HR que tiene. CIA nos permite:

lo quitamos porque $\perp\!\!\!\perp$

$$E[HR_{ie} \mid X_i, e_i] = E[HR_{ie} \mid X_i]$$

Notemos que al hacer lo anterior, estamos descomponiendo el término de error en X_i y V_i; pero, al ser éstos ortogonales para que:

$$E[HR_{ie} \mid X_i, e_i] = E[HR_{ie} \mid X_i] = \alpha + \beta e + X_i \delta$$

lo que implica que $cov(v_i, e_i \mid X_i) = 0$ y que β será un estimador insesgado del cual podremos inferir causalidad.

¿Y esto qué tiene que ver con aleatoridad?

En la práctica, muchas veces no podemos inferir causalidad por problemas de selección (el tratamiento no es aleatorio) ej. la decisión de entrar al mct laboral, tener una TDC, ir al doctor, etc.

Cuando el tratamiento es aleatorio (aka es indep. del outcome) tenemos:

$(y_{1i}, y_{0i}) \perp\!\!\!\perp D_i$ → aplicamos CIA

↳ notación

Entonces tenemos que

$$E[y_{1i} \mid D_i = 1] = E[y_{1i} \mid D_i = 0] = E[y_{1i}]$$

$$E[y_{0i} \mid D_i = 1] = E[y_{0i} \mid D_i = 0] = E[y_{0i}]$$

contrafactuales son idénticos → si no hubiera habido tratamiento el grupo de tratamiento y control se hubieran comportado igual

En este contexto, podemos estimar el efecto del tratamiento como una diferencia de medias:

$$\begin{aligned}
\Delta\mu &= E[y_i \mid D_i = 1] - E[y_i \mid D_i = 0] \\
&= E[y_{1i} \mid D_i = 1] - E[y_{0i} \mid D_i = 0] \\
&= E[y_{1i} \mid D_i = 1] - E[y_{0i} \mid D_i = 0] + \overbrace{E[y_{0i} \mid D_i = 1] - E[y_{0i} \mid D_i = 1]}^{=0} \\
&= E[y_{1i} \mid D_i = 1] - E[y_{0i} \mid D_i = 1] + E[y_{0i} \mid D_i = 1] - E[y_{0i} \mid D_i = 0] \\
&= E[y_{1i}] - E[y_{0i}]
\end{aligned}$$

ATE = Average treatment effect.

∴ La aleatorización elimina el sesgo de selección y nos permite estimar el ATE a partir de una dif. de medias obtenida por MCO.