

## Primer ejercicio

Por motivo de la grave situación económica en el país luego de la pandemia, el Gobierno lanzó un programa social destinado a mejorar los ingresos de hogares pequeños cultivadores de café. En particular, la ayuda consistió en una capacitación acerca de la elección y uso de fertilizantes, ofrecida al jefe de hogar, junto con la entrega de una cantidad fija de dinero. Sin embargo, los recursos para desarrollar esta política eran limitados, por lo que sólo se podía brindar el programa a algunos de los hogares registrados en las bases de datos del Gobierno (los hogares no se tenían que inscribir para ser candidatos a recibir el programa). Por lo tanto, buscando evitar favorecimientos indebidos, se eligió aleatoriamente un subconjunto de hogares que recibirían el programa. Todos las familias seleccionadas para recibir la ayuda participaron activamente.

En busca de evaluar la efectividad de la medida, el Gobierno recolectó información acerca de la cosecha posterior al programa para todos los  $n$  hogares que fueron elegibles para el programa (esto es, tanto para los que recibieron el programa como para aquellos que no). Entre la información recolectada, se cuenta con la cantidad de kilogramos de café producido por hectárea cultivada por cada hogar  $C_i$ , así como una dummy  $D_i$  que toma el valor de uno si el hogar participó en el programa y cero de lo contrario.

Confundiendo en las habilidades de los estudiantes de econometría avanzada, a ustedes los contrata el Departamento Nacional de Planeación para llevar a cabo la evaluación de impacto respectiva.

Su jefe les dice que una manera de modelar el problema es a través de un modelo de regresión lineal dado por

$$C_i = \alpha + \tau D_i + \epsilon_i \quad (1)$$

donde  $\epsilon_i$  es un componente aleatorio idiosincrásico de media cero ( $\mathbb{E}[\epsilon_i] = 0$ ) y con varianza  $\sigma^2$  ( $\mathbb{E}[\epsilon_i^2] = \sigma^2$ ).

- a) Sean  $C_i(1)$  y  $C_i(0)$  los resultados potenciales de haber participado o no en el programa respectivamente. Similarmente, sean  $\epsilon_i(1)$  y  $\epsilon_i(0)$  los resultados potenciales análogos del error idiosincrásico. Suponga además que  $\mathbb{E}[\epsilon_i(1)] = \mathbb{E}[\epsilon_i(0)] = 0$ .

- ¿Cuáles son las formas funcionales de  $C_i(1)$  y  $C_i(0)$  inducidas por el modelo (1)?
- ¿Cuál parámetro del modelo captura el ATT del programa? Justifique matemáticamente.
- Imagine un escenario donde el programa, en lugar de ser asignado aleatoriamente, se entregaba a las familias que viven en lugares con climas menos favorables para el cultivo. ¿En este escenario el parámetro del inciso ii) sigue capturando el ATT?

**Pista:** Recuerde que

$$C_i = D_i C_i(1) + (1 - D_i) C_i(0); \quad \epsilon_i = D_i \epsilon_i(1) + (1 - D_i) \epsilon_i(0)$$

- b) Bajo el cumplimiento de los supuestos del modelo clásico lineal, demuestre que el estimador de MCO  $\hat{\tau}$  es un estimador consistente del ATT y derive explícitamente su distribución asintótica (la varianza asintótica debe depender únicamente de  $n$ ,  $\sigma^2$  y  $p = \mathbb{P}(D_i = 1)$ ).

**Pista:** Recuerde el siguiente teorema visto en clase:

Una secuencia de vectores aleatorios  $\{x_N : N = 1, 2, \dots\}$  de  $K \times 1$  converge en distribución al vector aleatorio  $x$ , si para cualquier vector no aleatorio  $c$  de  $K \times 1$ ,

$$c^T x_N \xrightarrow{d} c^T x$$

Contentos por los reveladores hallazgos de los incisos a) y b), ustedes van por un tinto a la cafetería. En ella, encuentran dos colegas debatiendo acerca de cómo estimar correctamente el ATT. Uno de ellos argumenta que, debido a que hay aleatorización, el estimador debe ser una diferencia “ingenua” de medias de la variable dependiente entre los hogares participantes y aquellos que, aunque elegibles, no accedieron al programa. Su otro compañero lo contradice, pues afirma que la manera correcta de hacerlo es estimando  $\tau$  del modelo (1) por MCO dadas las conclusiones del inciso a) y b). ¿Quién tiene la razón?

Queriendo resolver esta encrucijada, a ustedes se les ocurre una idea: si logran probar que ambos estimadores son numéricamente equivalentes, entonces no debería importar cuál procedimiento utilicen.

- c) Prueben que la diferencia ingenua de medias es numéricamente equivalente al estimador de MCO del parámetro  $\tau$  del modelo (1).

**Ayuda:** Considere el modelo vectorial

$$C = X\beta + \epsilon$$

donde

$$X = \begin{pmatrix} 1 & D_1 \\ 1 & D_2 \\ \vdots & \vdots \\ 1 & D_n \end{pmatrix}; \quad C = \begin{pmatrix} C_1 \\ C_2 \\ \vdots \\ C_n \end{pmatrix}; \quad \epsilon = \begin{pmatrix} \epsilon_1 \\ \epsilon_2 \\ \vdots \\ \epsilon_n \end{pmatrix}; \quad \beta = \begin{pmatrix} \alpha \\ \tau \end{pmatrix}$$

Puede usar sin probar que

$$(X'X)^{-1} = \frac{1}{\left(\sum_{i=1}^n D_i\right) \left(\sum_{i=1}^n (1 - D_i)\right)} \begin{pmatrix} \sum_{i=1}^n D_i & -\sum_{i=1}^n D_i \\ -\sum_{i=1}^n D_i & n \end{pmatrix}$$

y que

$$X'C = \begin{pmatrix} \sum_{i=1}^n C_i \\ \sum_{i=1}^n D_i C_i \end{pmatrix}$$

Ahora una compañera le comenta que, según lo que ha leído, ella cree que el éxito del programa depende en gran medida de la calidad de la tierra donde se siembra. De manera que, pueden existir posibles *efectos heterogéneos* dependiendo de las dotaciones de este factor en cada hogar. Asimismo, le comenta que afortunadamente se cuenta con una variable  $Z_i$  en la base de datos que captura esta información. Por lo tanto, ella propone que un modelo más completo estaría dado por

$$C_i = \beta_0 + \beta_1 D_i + \beta_2 \tilde{Z}_i + \beta_3 D_i \tilde{Z}_i + \epsilon_i \quad (2)$$

donde  $\tilde{Z}_i = (Z_i - \bar{Z})$  y  $\bar{Z}$  es el promedio muestral. Finalmente, suponga que  $\mathbb{E}[\epsilon_i(1)|Z_i] = 0$  y  $\mathbb{E}[\epsilon_i(0)|Z_i] = 0$ , esto es, que la calidad de la tierra es exógena.

- d) Partiendo del modelo (2):

- i) Calcule el efecto esperado en los tratados como función de  $Z$ :

$$ATT(Z) = \mathbb{E}[C_i(1) - C_i(0)|D_i = 1, Z]$$

¿Cuál es la interpretación de  $\beta_1$  y  $\beta_3$ ?

- ii) Halle el parámetro que ahora captura el ATT del programa.  
iii) ¿Qué pasaría si  $Z_i$  no fuese exógena?

Finalmente, su jefe les tiene un último reto: probar si para al menos el 95 % de los hogares participantes el efecto de participar fue positivo.

e) Suponga que  $Z_i \sim \mathbb{N}(\mu, 1)$ . Usando la forma funcional del modelo (2):

- i) Encuentre la distribución de  $ATT(Z)$
- ii) Encuentre el percentil 5 ( $p_{0.05}$ ) de la distribución de  $ATT(Z)$  en términos de los parámetros del modelo.
- iii) Suponga que usted estimó por MCO que

$$\begin{pmatrix} \hat{\beta}_0 \\ \hat{\beta}_1 \\ \hat{\beta}_2 \\ \hat{\beta}_3 \end{pmatrix} = \begin{pmatrix} 0.6 \\ 2 \\ 0.3 \\ -0.03 \end{pmatrix}; \quad S = \begin{pmatrix} 0.6 & -0.26 & 0.33 & 0 \\ -0.26 & 0.7 & 0.5 & 0 \\ 0.33 & 0.5 & 0.35 & -0.25 \\ 0 & 0 & -0.25 & 0.08 \end{pmatrix}$$

donde  $S$  es la matriz de varianza-covarianza estimada de los parámetros.

Construya un estimador consistente de  $p_{0.05}$  y diseñe un test que le permita probar:

$$\begin{cases} H_0 : p_{0.05} \leq 0 \\ H_a : p_{0.05} > 0 \end{cases}$$

Use los datos disponibles para ejecutar su test. Concluya para un nivel de significancia  $\alpha = 0.05$ .

- iv) Verdadero o falso: Si el  $p$ -valor del test es mayor a 0.05, entonces existe evidencia estadística de que el efecto de que el hogar en el percentil 5 de la distribución de efectos fue negativo o nulo. Justifique.

## Segundo ejercicio

Una curva de aprendizaje se puede pensar como el cambio en la productividad o eficiencia con la que se hacen las cosas en el transcurso del tiempo. Este concepto es utilizado en Economía, por ejemplo, para modelar el cambio de los costos reales de las firmas según el tiempo que llevan operando. En particular, se supone que las firmas van aprendiendo a producir de manera más eficiente, por lo que el costo real de producción debería, *ceteris paribus*, decrecer.

Supongan que la curva generalizada de aprendizaje de una firma hipotética está dada por

$$C_i = C_0 N_i^{\frac{\alpha}{\gamma}} Y_i^{\frac{1-\alpha}{\gamma}} \exp(u_i), \quad (1)$$

donde  $C_i$  corresponde a los costos reales unitarios que enfrenta una firma  $i$ ;  $Y_i$  es su nivel de producción;  $N_i$  es la producción acumulada a lo largo del tiempo;  $C_0$  corresponde a una medida de costos iniciales; y  $u_i$  es un término estocástico de media cero desconocido para el econometrista. El parámetro  $\alpha$  determina la dirección de la elasticidad del costo unitario con respecto a la producción acumulada. Este es el principal parámetro de interés. Finalmente, el parámetro  $\gamma \in \mathbb{R}^+$  caracteriza los retornos a escala de la función de aprendizaje: si  $\gamma = 1$  la curva tiene retornos constantes a escala, si  $\gamma < 1$  la curva tiene retornos decrecientes a escala, y si  $\gamma > 1$  la curva tiene retornos crecientes a escala.<sup>1</sup>

Usted, como investigador, está interesado en estimar la curva de aprendizaje que enfrentan las firmas de la industria manufacturera colombiana. Para ello, cuenta con información de  $C_i$ ,  $N_i$  y  $Y_i$  para  $N$  firmas de la industria.

- a) Propongan un modelo de regresión lineal que capture la forma funcional dada en (1). Muestre cómo los parámetros de su nuevo modelo dependen de los parámetros de la curva de aprendizaje. Mencionen y discutan los supuestos necesarios para que los estimadores por MCO de los parámetros del modelo sean insesgados y consistentes. ¿Son estos supuestos plausibles en el contexto del problema?

<sup>1</sup>Si la tecnología de producción muestra rendimientos constantes a escala, los costos unitarios reales no deben variar con el nivel de producción. Por el contrario, si los rendimientos son crecientes a escala, los costos unitarios deberían disminuir a medida que aumenta el nivel de producción, y si son decrecientes, se esperaría lo contrario.

- b) Supongan que se cumplen los supuestos que ustedes discutieron en el inciso anterior. Propongan un estimador consistente de  $\alpha$  y  $\gamma$ .
- c) Propongan estimadores de las desviaciones estándar de  $\hat{\alpha}$  y  $\hat{\gamma}$ . Para esto, supongan que la matriz de varianzas y covarianzas asintótica de  $\sqrt{N}(\hat{\beta} - \beta)$  y su respectivo estimador están dados por

$$V = Avar(\sqrt{N}(\hat{\beta} - \beta)) = \begin{pmatrix} \sigma_1^2 & \sigma_{12} & \sigma_{13} \\ \sigma_{12} & \sigma_2^2 & \sigma_{23} \\ \sigma_{13} & \sigma_{23} & \sigma_3^2 \end{pmatrix} \text{ y } \hat{V} = \widehat{Avar}(\sqrt{N}(\hat{\beta} - \beta)) = \begin{pmatrix} \hat{\sigma}_1^2 & \hat{\sigma}_{12} & \hat{\sigma}_{13} \\ \hat{\sigma}_{12} & \hat{\sigma}_2^2 & \hat{\sigma}_{23} \\ \hat{\sigma}_{13} & \hat{\sigma}_{23} & \hat{\sigma}_3^2 \end{pmatrix} \quad (3)$$

respectivamente. Asegúrense de que sus expresiones estén en términos de  $\hat{\beta}_1, \hat{\beta}_2$  y los elementos que componen  $\hat{V}$ .

**Pista:** El método delta puede resultar útil.

Usted cuenta con la base de datos “manufacturaCol.dta” para implementar la estimación de los parámetros propuesta en los incisos anteriores. La base de datos cuenta con información de las siguientes variables para cada una de las 11,500 firmas de la muestra:

- *costos*: Costos unitarios reales de producción que enfrenta la firma en el momento que se levantaron los datos medido en miles de millones pesos.
- *producto*: Producción de la firma en miles de millones pesos en el momento que se levantaron los datos.
- *producto\_acum*: Producción acumulada histórica de la firma en el momento en que se levantaron los datos medida en miles de millones de pesos.

Resuelvan los siguientes incisos a partir de los datos disponibles en la base de datos.

- d) A manera de estadísticas descriptivas, representen mediante un gráfico de dispersión y su respectiva línea de ajuste las siguientes relaciones: 1. La relación de  $\log(C_i)$  con  $\log(N_i)$  una vez se ha “removido” el efecto de  $\log(Y_i)$ . 2. La relación de  $\log(C_i)$  con  $\log(Y_i)$  una vez se ha “removido” el efecto de  $\log(N_i)$ .

**Ayuda:** para lograr esto, apliquen la intuición de partialling out o residualización de los modelos de regresión múltiples. Interprete los gráficos de dispersión.

- e) Finalmente, estimen por MCO la ecuación propuesta en el inciso a). A partir de estos resultados, estimen  $\alpha, \gamma$ , y sus respectivos errores estándar siguiendo el planteamiento de los incisos b) y c). Presenten en una tabla los parámetros estimados por MCO. ¿Son estos parámetros consistentes con las gráficas que presentaron en el inciso anterior? Discutan brevemente. Adicionalmente, en una segunda tabla presenten los parámetros  $\alpha$  y  $\gamma$  que ustedes estimaron y sus respectivos errores estándar. Interpreten estos últimos. ¿Hay rendimientos crecientes, decrecientes o constantes a escala? ¿Qué tipo de curva de aprendizaje hay en la industria?

## Tercer ejercicio

Aunque las propiedades asintóticas de los estimadores son de gran utilidad para el ejercicio práctico de la estadística, lo cierto es que muchas veces encontrar resultados teóricos en este ámbito es extremadamente complicado. En estas ocasiones, es usual recurrir a otras herramientas, como lo son los procedimientos de Monte Carlo. En breve, los métodos de Monte Carlo aplicados a la estadística buscan explorar las propiedades de los estimadores (insesgamiento, consistencia, eficiencia, suficiencia, etc.) al observar el comportamiento de los mismos en varias muestras aleatorias simuladas. Este ejercicio los guiará a través de una serie de actividades que les permitirán entender algunos métodos clásicos de simulación de muestras aleatorias para, posteriormente, usarlos para evaluar el comportamiento de uno de los estimadores más importantes que se encontrarán en sus carreras: el estimador de MCO.

Suponga que el Ministerio de Educación ha decidido lanzar un programa de preparación y acompañamiento para la presentación de la prueba ICFES a estudiantes que se encuentren en su último año de bachillerato. En particular, suponga que usted conoce que el puntaje ICFES estandarizado potencial  $Y_i(\cdot)$  obedece el siguiente

modelo:

$$Y_i(0) \sim N(0, 1)$$

$$Y_i(1) = Y_i(0) + 3$$

donde  $Y_i(1)$  y  $Y_i(0)$  son los resultados potenciales del puntaje ICFES en caso de participar y de no participar respectivamente, y donde  $Y_i(0)$  tiene una distribución normal estándar.

- a) Escriba el puntaje ICFES observado,  $Y_i$ , en términos de los resultados potenciales y la exposición al tratamiento,  $D_i$ . Suponga que el Ministerio escoge aleatoriamente a los participantes del programa y que todo individuo seleccionado obligatoriamente participa. Proponga un modelo de regresión lineal que le permita estimar el efecto del tratamiento sobre el puntaje, llámese  $\delta$ .
- b). Inicialmente, buscaremos validar a través de simulaciones algunas de las propiedades asintóticas del estimador por MCO. En particular, a usted le interesa saber si a medida que aumenta el tamaño de muestra sus estimativos del efecto del programa se tornan más precisos.
  - I. Simule 100 muestras  $\{(Y_i(0), Y_i(1), D_i)\}_{i=1}^n$  de tamaño  $n = 10, 20, \dots, 1000$  donde  $D_i \sim \text{Bernoulli}(0.3)$ . Para cada muestra, estime el efecto de la política del Ministerio sobre el puntaje ICFES y almacene su estimado.
  - II. Obtenga la media y la varianza muestral de los estimadores almacenados para los distintos tamaños de muestra.
  - III. Haga las siguientes dos gráficas: 1) Grafique el promedio muestral de los estimadores contra el tamaño de muestra. 2) Grafique las varianzas muestrales contra el tamaño de muestra correspondientes. ¿Qué puede decir acerca del comportamiento de las estimaciones a medida que aumenta el tamaño de muestra? ¿Qué propiedades del estimador de MCO se ven reflejadas en el ejercicio?

Uno de los propósitos principales de los modelos de regresión lineal es poder estudiar las relaciones que existen entre la variable dependiente y las variables independientes. Más precisamente, nos interesa dilucidar el tamaño y el signo de dicha relación. No obstante, en la práctica, esto es difícil de establecer puesto que desconocemos el proceso generador de los datos y en general, contamos con una sola muestra para nuestra estimación. Así las cosas, si nuestro objetivo es hacer inferencia, debemos preguntarnos qué tan precisas son nuestras estimaciones, esto es, qué tan dependientes son de la muestra particular que tenemos. Para lograrlo, es útil entender a los estimadores como variables aleatorias, que dependen de una muestra, pero que tienen una distribución definida. En particular, es de nuestro interés dicha distribución cuando el tamaño de muestra es grande, pues conocer la distribución exacta para muestras reducidas puede ser complicado.

- c) En este inciso vamos a aproximar la distribución asintótica del estimador de MCO. Para ello, realicen las siguientes instrucciones:

- I. Siguiendo los pasos expuestos en el punto b), simule 1000 muestras  $\{(Y_i(0), Y_i(1), D_i)\}_{i=1}^n$  de tamaño  $n = 10, 20, 100, 1000$  donde  $D_i \sim \text{Bernoulli}(0.3)$ . Para cada muestra, estime el efecto del programa del Ministerio y calcule y almacene

$$a_{k,n} = \sqrt{n}(\hat{\delta}_{k,n} - \delta)$$

donde  $k$  indexa las muestras de un determinado tamaño.

- II. Grafique las densidades estimadas de los  $a_{k,n}$  para cada  $n$ . ¿Qué puede apreciar a medida que aumenta  $n$ ? ¿A qué se debe este resultado?

Suponga ahora que usted sabe que existen factores ajenos a las políticas del Ministerio que inciden sobre los resultados potenciales de los alumnos. Por ejemplo, sabe que existen alumnos que, independientemente de si son o no seleccionados, igual se inscribirían en cursos de preparación para presentar el ICFES. Similarmente, hay alumnos que por condiciones adversas (enfermedades, salones menos adecuados para la presentación del examen) exhiben un desempeño muy inferior a lo esperado. Estos factores producen una mayor volatilidad en los resultados potenciales de los alumnos.

- d) En este sentido, usted sabe que es más plausible considerar que los resultados potenciales en realidad obedecen la ley  $Y_i(0) \sim \text{Cauchy}$  (**Distribución Cauchy estándar**) en vez de una normal estándar. Esto es, en ocasiones usted observa datos atípicos que no parecen corresponder con el patrón general de la muestra. Repita los incisos I y II del c) bajo estas condiciones. ¿Por qué en este caso no se satisface el Teorema del Límite Central? ¿Qué consecuencias tendría esto para la inferencia de los parámetros de MCO si usáramos los procedimientos estadísticos usuales en este caso?

Finalmente, suponga que usted sabe que el programa implementado por el ministerio tiene el objetivo de nivelar los conocimientos de los estudiantes que en el participan. Por dicha razón, los resultados de las pruebas de los estudiantes que no participan en el programa son mas volátiles.

Una manera de modelar esta observación es a través del modelo:

$$Y_i = 3 * D_i + v_i$$

donde  $v_i \sim N(0, 2 - D_i)$ .

- e) Para este nuevo modelo, realice el siguiente procedimiento:

- I. Simule 1000 muestras  $\{(Y_i(0), Y_i(1), D_i)\}_{i=1}^{1000}$ . Para cada muestra, recupere los intervalos de confianza clásicos al 90 %, 95 % y 99 % obtenidos para  $\delta$  y codifique en una matriz si dicho intervalo contiene o no a  $\delta$ .
- II. ¿Qué porcentaje de los intervalos de cada nivel contiene al parámetro verdadero? Presente sus resultados en una tabla ¿A qué se deben estos resultados? ¿Qué le sugiere esto sobre su forma de computar los intervalos?
- III. Repita los numerales I y II, esta vez empleando errores estándar de White (robustos) en la construcción de sus intervalos. Concluya.

## Punto doctorado

Usted se encuentra cursando el curso de seminario de investigación doctoral. En las primeras semanas le piden que presente algunas relaciones de causalidad que tenga pensado abordar en su disertación. Su profesora de seminario lo invita a que avance en el planteamiento de sus ideas a través del uso de gráficos acíclicos dirigidos (Directed Acyclic Graph - DAG).<sup>2</sup> En breve, los DAGs se componen de dos elementos: flechas y variables. Las flechas entre variables indican causalidad en el sentido en el que esta apuntando la flecha. Un **Camino** es cualquier conexión entre dos variables realizada por flechas, sin importar su dirección o si existen otras variables intermedias. Un **Camino por la puerta trasera** de la variable  $X$  a la variable  $Y$  es un camino que empieza con una flecha dirigida hacia  $X$ . Un **Colisionador** es una variable a la que apuntan dos flechas en un camino.

- a) Usando puntos negros para graficar las variables observadas y circunferencias para graficar las variables no observadas ( $X$ ,  $Z$  y  $Y$ ), realice los DAGs que representan:
  - i)  $X$  tiene una relación de causalidad hacia  $Y$ . A su vez,  $Z$  tiene una relación de causalidad hacia  $X$  y otra hacia  $Y$ .
  - ii)  $X$  tiene una relación de causalidad hacia  $Z$ . A su vez,  $Z$  tiene una relación de causalidad hacia  $Y$ .
  - iii) La variable no observada:  $U$  tiene una relación de causalidad hacia  $Z$  y otra hacia  $Y$ . A su vez,  $X$  tiene una relación de causalidad hacia  $Y$  y otra hacia  $Z$ .
  - iv) La variable  $X$  tiene una relación de causalidad hacia  $Z$  y otra hacia  $Y$ . A su vez,  $Y$  tiene una relación de causalidad hacia  $Z$ .
  - v) La variable no observada:  $U$  tiene una relación de causalidad hacia  $Z$  y otra hacia  $Y$ . A su vez,  $X$  tiene una relación de causalidad hacia  $Y$ , mientras que  $Z$  tiene una relación de causalidad hacia  $X$ .
- b) Ahora usted quiere implementar un modelo de regresión lineal simple para evaluar los efectos causales de  $X$  sobre  $Y$ . Con este objetivo, usted está determinando la conveniencia de incluir la variable  $Z$  como variable de control en su modelo. Argumente, para cada caso representado por los DAGs, si la inclusión de  $Z$ , como control en el modelo, es apropiada. Para sus respuestas, tenga en cuenta que un variable es

<sup>2</sup>Si no se encuentra familiarizado con los DAGs, una excelente referencia y explicación se encuentra en el tercer capítulo del libro de Scott Cunningham: Causal Inference: The mixtape, del año 2021. <https://mixtape.scunning.com/>

un mal control si: bloquea caminos causales entre  $X$  y  $Y$  o abre otros caminos que no son causales entre  $X$  y  $Y$ .

*(Pista: Bloquear un camino es equivalente a controlar por variables que no son Colisionadoras o no controlar por las Colisionadoras. Abrir un camino es equivalente a controlar por variables que son Colisionadoras o no controlar por no Colisionadoras)*

- c) Finalmente, usted sabe que la variable  $X$  es independiente de los resultados potenciales de  $Y$  y de  $Z$ . Por otro lado, usted sospecha que  $Z$  podría ser un mal control y que este hecho puede generarse porque la variable  $Z$  es, a su vez, una variable de resultado en su modelo. Plantee una forma de aproximarse empíricamente para evaluar si este hecho puede estar ocurriendo