

Problem Set 3: Fuentes de sesgo e imprecisión

Fecha de entrega: 6 de septiembre de 2024

Paola Llamas y Tomás Pacheco

Instrucciones

- Agreguen los nombres y apellidos de los integrantes en el encabezado del archivo que entregan.
- El título de los archivos entregados debe tener la siguiente estructura:
PS3_Garcia_Gomez_Gonzalez_Perez.pdf
- **UN ÚNICO** miembro del grupo debe subir el PDF y *do-file* al Campus Virtual en la sección "Problem Set 3".
- El código lo tendrán que entregar de dos formas distintas. Primero deberán exportar el *do-file* como PDF y unirlo al informe ([ayuda](#)). Luego, deberán entregar el código en formato .do en el lugar indicado en el Campus Virtual.
- El directorio de trabajo debe tener tres carpetas: programs, input y output. El *do-file* debe ser entregado de tal manera que los tutores podamos correr el código **solamente** modificando el global con el *main directory* (**esto es muy importante**).
- Recuerden que siempre deben presentar la información en tablas y gráficos estéticos. Nunca presenten una tabla o figura sin dedicarle unas líneas a comentar lo que se observa.

Ejercicio 1

Repitan la simulación hecha en clase incluyendo modificaciones menores para mostrar los siguientes puntos. Si quieren, pueden crear un modelo propio.

1. ¿Qué sucede con los errores estándar de los regresores si aumenta el tamaño muestral?
2. ¿Qué sucede con los errores estándar de los regresores si aumenta la varianza de μ (término de error)?
3. ¿Qué sucede con los errores estándar de un regresor si aumenta la varianza de X ?
4. ¿Cuánto vale la suma de los residuos? Para responder a esta pregunta, pueden elegir cualquier modelo.
5. ¿Son los residuos ortogonales a los regresores?
6. ¿Cómo afecta la alta multicolinealidad a la estimación de Y ?
7. ¿Qué sucede si corren una regresión con un error no aleatorio en X ? ¿Y si ese error fuera aleatorio?
8. ¿Qué sucede si corren una regresión con un error no aleatorio en Y ? ¿Y si ese error fuera aleatorio?

Ejercicio 2

Supongan que están interesados en estimar el efecto causal de X_1 en Y . Cuentan con datos de dos variables de control, X_2 y X_3 . Pueden suponer que Y es la nota en un examen de matemática, X_1 la asistencia a clases, X_2 el promedio del alumno y X_3 la cantidad de horas que estudia el alumno por semana.

Llamaremos $\tilde{\beta}_1$ al estimador de la regresión de Y en X_1 :

$$score_i = \beta_0 + \beta_1 attend_i + \mu_i$$

y $\hat{\beta}_1$ al coeficiente de la asistencia de la regresión de Y en X_1 , X_2 y X_3 :

$$score_i = \beta_0 + \beta_1 attend_i + \beta_2 cgpa_i + \beta_3 study_i + \mu_i$$

Respondan las siguientes preguntas:

1. Si X_1 está altamente correlacionada con X_2 y con X_3 , y X_2 y X_3 tiene mucha relación con Y , ¿esperan que $\hat{\beta}_1$ y $\tilde{\beta}_1$ sean similares o distintos? Expliquen.
2. Si X_1 no está correlacionada con X_2 y X_3 pero X_2 y X_3 están altamente correlacionadas entre ellas, ¿esperan que $\hat{\beta}_1$ y $\tilde{\beta}_1$ sean similares o distintos? Expliquen.
3. Si $\hat{\beta}_1$ es el coeficiente de asistencia a clase en la regresión de Y en X_1 , X_2 , X_3 y X_4 , en donde X_4 es el consumo de chocolate del alumno, ¿esperan que $\hat{\beta}_1$ y $\tilde{\beta}_1$ sean similares o distintos? Expliquen.
4. Si X_1 está altamente correlacionada con X_2 y X_3 pero X_2 y X_3 tienen poca relación con Y , ¿esperan que $\hat{\beta}_1$ y $\tilde{\beta}_1$ sean similares o distintos? Expliquen.
5. ¿Cómo esperan que sea la relación entre los errores estándar de los coeficientes $\tilde{\beta}_1$ y $\hat{\beta}_1$ si: X_1 esta incorrelacionada con X_2 y X_3 , X_2 y X_3 tienen grandes efectos marginales en Y y X_2 y X_3 están altamente correlacionados? Expliquen.
6. ¿Cómo esperan que sea la relación entre los errores estándar de $\hat{\beta}_1$ y $\tilde{\beta}_1$? (expliquen los supuestos que hacen sobre el consumo de chocolate y la correlación con otras variables).