

Tarea 5: Correlación y Causalidad

Fecha de Entrega: Miercoles 18 de Septiembre, 6pm por Bloque Neón.

Introducción

En esta tarea vamos a pensar sobre el sesgo de variable omitida que puede inducir correlaciones espúreas. El sesgo de variable omitida (SVO) es el sesgo en un estimador de regresión que surge cuando hay una variable que no está incluida en la regresión y que está correlacionada con el regresor y es un determinante del resultado.

Ejercicio

Instrucciones: Para este ejercicio responda las preguntas junto a su equipo y al finalizar la clase suba un documento a Bloque Neón.

1. ¿Cuál piensa usted que es la verdadera relación causal entre el número de horas que un individuo estudia y sus calificaciones en los exámenes? Olvídense por un momento de los datos e imagine el siguiente experimento: usted esta en un laboratorio y puede mantener todo los factores constantes. Ahora asigna a ciertos estudiantes que estudien 3 horas, a otros 1 hora y a los restantes que no estudien (estudien 0 horas). Luego estos estudiantes toman un examen sobre el tema que estudiaron (o no) ¿Qué relación vería entre la calificación en el examen y las horas estudiadas? Describa en términos de signos esperados.
2. La cantidad y calidad de horas que duermen los individuos esta relacionada con su rendimiento diario, especialmente respecto al estudio. Formule una hipótesis sobre la relación entre las horas y la calidad de sueño y el estudio. Luego, sobre la relación entre entre las horas y la calidad de sueño y las calificaciones en los exámenes. (no mas de cinco lineas).
3. Descargue de Bloque Neón la base de datos llamada `notas_y_estudio.csv`. Esta base contiene información sobre 1,715 estudiantes, e incluye información sobre su edad (en años), género (1 si es mujer), los créditos obtenidos hasta el momento, las horas de estudio promedio (`horas_estudio`), las horas promedio de sueño (`horas_dormir`), y su calificación promedio (`calificacion`) expresada en porcentaje (0 a 100).
4. En su software favorito (**Stata**, **R**, etc) realice lo siguiente
 - (a) Grafique un diagrama de dispersión entre la calificación promedio (`calificacion`) y el número de horas de estudio promedio (`horas_estudio`). ¿Qué relación observa?

- (b) Estime una regresión simple entre la calificación promedio (**calificacion**) y el número de horas que un individuo estudia (**horas_estudio**). Describa el resultado obtenido en términos de magnitudes y significancia estadística. Comente en términos de su hipótesis en (1).
- (c) Grafique un diagrama de dispersión entre el número de horas que un individuo estudia (**horas_estudio**) y las horas de sueño (**horas_dormir**).
- (d) Estime una regresión simple entre el número de horas que un individuo estudia (**horas_estudio**) y las horas de sueño (**horas_dormir**). Describa el resultado obtenido en términos de magnitudes y significancia estadística. Comente en términos de su hipótesis en (2).
- (e) Estime una regresión múltiple entre sobre la calificación promedio (**calificacion**), el número de horas que un individuo estudia (**horas_estudio**) y las horas de sueño (**horas_dormir**). Describa el resultado obtenido en términos de magnitudes y significancia estadística. Comente en términos de su hipótesis en (1) y (2).

* Comandos Útiles en Stata

* Importar los datos a Stata

```
import delimited "notas_y_estudio.csv", clear
```

* Regresión Simple

```
reg y x
```

* Regresión Múltiple

```
reg y x z
```

* Diagrama de dispersión

```
twoway (scatter y x)
```

* Diagrama de dispersión con recta de valores ajustados

```
twoway (scatter y x) (lfit y x)
```

* Para obtener tablas prolijas puede usar outreg2

* Si necesita instalarlo, use el siguiente comando

```
ssc install outreg2
```

* outreg2 se llama luego de una regresión

* Corremos una regresión y guardamos los resultados en outreg2

que los escribe en el documento 'myreg.docx'.

La opción replace sirve para crear un archivo nuevo o sobrescribir uno existente.

La opción append adjunta los resultados en una columna contigua.

```
reg y x
```

```
outreg2 using myreg.doc, replace
```

```
reg y x z
```

```
outreg2 using myreg.doc, append
```

```
#####  
# Comandos Útiles en R  
#####  
  
#paquetes  
require("tidyverse")  
  
#leer datos  
dta<-read.csv("notas_y_estudio.csv")  
  
# Regresión Simple  
reg_simple<-lm(y~x, data=dta)  
  
# Regresión Múltiple  
reg_multiple<-lm(y~x+z, data=dta)  
  
#Diagrama de dispersión  
ggplot(dta,aes(x=horas,y=calificacion)) +  
  geom_point()  
  
# Diagrama de dispersión con recta de valores ajustados  
ggplot(dta,aes(x=horas,y=calificacion)) +  
  geom_point() +  
  geom_smooth(method = lm)  
  
# Para obtener tablas prolijas puede usar stargazer  
require("stargazer")  
stargazer(reg_simple,reg_multiple,type="text")
```