

# Auxiliar 9

## Repaso OLS

**Profesor: Raimundo Undurraga**

Auxiliares: Brandon Galarza, Camila Jáuregui, Leonardo Meneses, Francisca Monetta,  
Matías Reyes, Bastián Urzúa, Antonia Villegas.

## Parte 1: Comentes

1. Si los supuestos del teorema de Gauss-Markov se cumplen, entonces no es posible hallar un estimador con error cuadrático medio menor que MCO.
2. Suponga que quiere construir un modelo OLS para determinar los factores que influyen en el desempeño académico de los estudiantes que cursan segundo año medio. Para lo cual, piensa utilizar, como variables independientes, el nivel de ingresos mensuales recibidos en el hogar del estudiante y su calificación en el Registro Social de Hogares. Con lo cual, se tiene el siguiente modelo:

$$DesempeñoAcadémico_i = \beta_0 + \beta_1 IngresosMensuales_i + \beta_2 CalificaciónRSH_i + \epsilon_i$$

Antes de ejecutar sus modelo en R, anticipése comentando qué problemas podría presentar el estimador que construyó. (Hint: investigue cómo se determina el RSH de un hogar)

3. Para su práctica, lo contratan desde un centro de salud, para que construya un modelo causal que explique el efecto del consumo frecuente de alcohol en los jóvenes, sobre a el estado de salud que tienen cuando superan los 40 años. Así que construye el siguiente modelo:

$$Salud40_i = \beta_0 + \beta_1 ConsumoCig_i + \epsilon_i$$

Donde Salud40 es un indicador de la salud de la persona  $i$  cuando esta cumple 40 años. El cual toma valores en el rango  $[1,10]$ , donde 1 representa un estado de salud precario, mientras que un 10 significa una salud perfecta.

Al correr su modelo en R, tiene que  $\beta_1 = -9.5$ . Su tutora, una enfermera experimentada, dice estar de acuerdo con tener un efecto negativo asociado al consumo, pero le parece

exagerada su magnitud. Entonces, usted piensa que su coeficiente puede estar sesgado. En ese sentido, comente qué está pasando exactamente con su modelo y cómo mejorar el coeficiente asociado.

## Parte 2: Diagnóstico de OLS

La ecuación de Mincer, que busca explicar el retorno a la educación (es decir, cuánto más rentable es cada año adicional de estudio) se usa y define tradicionalmente como:

$$\ln(\text{ingresoporhora}_i) = \beta_0 + \beta_1 \text{escolaridad}_i + \beta_2 \text{experiencia}_i + \beta_3 \text{experiencia}_i^2 + \mu_i$$

Donde  $\text{experiencia}_i$  representa la experiencia potencial del individuo  $i$ . A partir de la Ecuación de Mincer, realice lo siguiente:

- Estudie si agregar un término cúbico de experiencia potencial mejora el modelo propuesto. Para las siguientes partes, trabaje con el modelo elegido.
- Estudie si el modelo elegido presenta heterocedasticidad.
- Estudie si el modelo presenta multicolinealidad.
- Estudie si el modelo presenta endogeneidad.
- De acuerdo a sus resultados, responda: ¿Qué tan adecuado es el modelo de regresión para explicar el salario obtenido por una persona?

## Parte 3: Inferencia Causal

El gobierno está evaluando el impacto que podría causar un nuevo programa social que llaman "Escoge ser saludable", el cual buscaría inculcar los hábitos de vida sana en los niños, para así reducir los niveles de obesidad en los menores. Dicha campaña consta en impartir talleres deportivos a estos niños, y otorgarles una canasta mensual de alimentos saludables.

Para lo cual, disponen del siguiente modelo de regresión lineal:

$$IMC_i = \beta_0 + \beta_1 \text{Inscripción}_i + \epsilon_i$$

Donde el  $IMC_i$  es el Índice de Masa Corporal del niño  $i$ . E  $\text{Inscripción}_i$  es una variable binaria que indica si el niño  $i$  se inscribe en el programa.

Responda:

1. Si se decide que la inscripción será voluntaria, o sea, que cada apoderado del niño  $i$  se autoselecciona para participar en el programa, ¿a qué problema nos enfrentaremos? Justifique.

2. Explique por qué una asignación aleatoria de los niños al programa solucionaría el problema anterior. O sea, el gobierno decide, al azar, qué niños formarán un grupo de tratamiento y otro de control.