

Hoja de Ejercicios 4 Regresión Múltiple

Estadística-II. INTRODUCCIÓN a la ECONOMETRÍA. UC3M

1. (Ejercicio 3.1, Wooldridge (2006)). Con los datos GPA2.RAW sobre 4.137 universitarios, se estimó la siguiente ecuación mediante MCO:

$$\begin{aligned}\widehat{colgpa} &= 1,392 - ,0135 \, hspc + ,00148 \, sat \\ n &= 4,137, \, R^2 = ,273\end{aligned}$$

donde *colgpa*, es el promedio de calificaciones de la universidad, se mide en una escala de cuatro puntos; *hspc* es el percentil de alumnos del instituto que se gradúan ese año (definido de tal modo que, por ejemplo, *hspc* = 5 se refiere al 5 por ciento de los mejores alumnos que se gradúan) y *sat* - Standard Aptitude Test (SAT) (prueba de aptitud académica en matemáticas y expresión verbal, ideada para alumnos de Estados Unidos)- es la combinación de las calificaciones de matemáticas y lengua en el test de aptitud escolar.

- (i) ¿Por qué tiene sentido que el coeficiente *hspc* sea negativo?
 - (ii) ¿Cuál es el nota media (*colgpa*) que podemos predecir si *hspc* = 20 y *sat* = 1050?
 - (iii) Supongamos que dos alumnos, A y B, se gradúan en el instituto dentro del mismo percentil, pero el resultado de A en el test *sat* de A es 140 puntos superior que el de B (alrededor de una desviación estándar en la muestra). ¿Qué diferencia podemos predecir entre la nota media universitaria de ambos estudiantes? ¿Es una diferencia importante?
 - (iv) Con *hspc* constante, ¿qué diferencia en los resultados *sat* lleva a predecir una diferencia en *univ* de 0,50 (medio punto) en *colgpa*? Razonar la respuesta.
2. (Ejercicio 3.3, Wooldridge (2006)). El modelo siguiente es una versión simplificada del modelo de regresión múltiple que emplearon Biddle y Hamermesh (1990) para estudiar el equilibrio entre el tiempo dedicado a dormir y a trabajar y analizar otros factores que influyen en el sueño:

$$sleep = \beta_0 + \beta_1 \, totwrk + \beta_2 \, educ + \beta_3 \, age + u$$

en el que *sleep* y *totwrk* (trabajo total) se miden en minutos por semana, en tanto que *educ* y *age* se miden en años (véase también el problema 2.12).

- (i) Si los adultos sustituyen tiempo de sueño por trabajo, ¿cuál es el signo de β_1 ?
- (ii) ¿Qué signos cree que tendrán β_2 y β_3 ?
- (iii) Con los datos de SUEÑO.RAW, la ecuación estimada es

$$\begin{aligned}\widehat{sleep} &= 3638,25 - ,148 \, totwrk - 11,13 \, educ + 2,20 \, age \\ n &= 706, \, R^2 = 0,113\end{aligned}$$

Si alguien trabaja cinco horas más por semana, ¿cuántos minutos se espera que disminuirá el sueño? ¿Es un cambio grande?

(iv) Analice el signo y la magnitud del coeficiente estimado de *educ*.

(v) ¿Explican *totwrk*, *educ* y *age* una parte importante de la variación en *sleep*? ¿Qué otros factores influirían en el tiempo dedicado a dormir? ¿Es probable que se correlacionen con *totwrk*?

3. (Ejercicio 3.4, Wooldridge (2006)). El salario inicial de los nuevos licenciados en derecho está determinado por

$$\log(\text{salary}) = \beta_0 + \beta_1 LSAT + \beta_2 GPA + \beta_3 \log(\text{libvol}) + \beta_4 \log(\text{cost}) + \beta_5 \text{rank} + u$$

en el que *LSAT* - en Estados Unidos, es la prueba de admisión en la facultad de derecho- es el resultado mediano del test *LSAT* para los alumnos de derecho que se gradúan ese año, *GPA* es la nota media de la universidad para ese mismo grupo, *libvol* es el número de volúmenes en la biblioteca de la facultad de derecho, *cost* es el coste anual de estudiar en la facultad de derecho y *rank* es el ranking de la facultad de derecho (*rank* = 1 es la mejor).

(i) Explique porqué esperamos que $\beta_5 \leq 0$.

(ii) ¿Qué signo esperaría para los otros parámetros? Justifique sus respuestas.

(iii) Con los datos de LEYES85.RAW, la ecuación estimada es

$$\begin{aligned} \log(\widehat{\text{salary}}) &= 8,34 + ,0047 LSAT + ,248 GPA + ,095 \log(\text{libvol}) + ,038 \log(\text{cost}) - ,0033 \text{rank} \\ n &= 136, R^2 = 0,842 \end{aligned}$$

¿Cuál es la diferencia esperada, *ceteris paribus*, en el salario entre facultades en las que la nota mediana difere en un punto? (Expresa su respuesta como porcentaje.)

(iv) Interprete el coeficiente de la variable $\log(\text{libvol})$.

(v) ¿Sería más recomendable ir a una facultad de derecho con un ranking superior? ¿En cuánto se ve afectado el salario inicial estimado por una diferencia de 20 puestos en el ranking?

4. (Ejercicio 3.5, Wooldridge (2006)). En un estudio que relaciona el nota media universitaria con el tiempo dedicado a diversas actividades, se distribuye una encuesta a varios estudiantes en la que se les pregunta cuántas horas destinan semanalmente a estas cuatro categorías: estudiar, dormir, trabajar y ocio. Cualquier actividad se sitúa en una de las categorías, de modo que la suma de horas de cada estudiante en las cuatro debe ser 168.

(i) En el modelo

$$GPA = \beta_0 + \beta_1 \text{study} + \beta_2 \text{sleep} + \beta_3 \text{work} + \beta_4 \text{leisure} + u$$

¿Tiene sentido mantener constantes *sleep*, *work* y *leisure* mientras cambia *study*?

(ii) Explique porque el modelo infringe el supuesto RLM.4.

(iii) ¿Cómo reformularía el modelo para que los parámetros tuvieran una interpretación útil y satisficieran el supuesto RLM.4?