# Hoja de Ejercicios 3

## Econometría I

1. Utilizando los datos de GPA2.RAW sobre 4127 alumnos universitarios, estima la siguiente ecuación por MCO,

$$colgpa = \beta_0 + \beta_1 hsperc + \beta_2 sat + u$$

donde *colgpa* se mide sobre una escala de cuatro puntos, *hsperc* es el percentil de los alumnos de instituto que se gradúan ese año (definido de forma que, por ejemplo, $hsperc = 5$ se refiere al 5 por ciento de los *mejores* alumnos que se gradúan) y *sat* equivale a los resultados conjuntos en matemáticas y lengua en el test de aptitud escolar. Escribe la ecuación estimada de la forma habitual y proporciona el $R^2$.

   *a*) ¿Por qué tiene sentido que el coeficiente de *hsperc* sea negativo?

   *b*) ¿Qué nota media universitaria (*colgpa*) podemos predecir si $hsperc = 20$ y $sat = 1050$?

   *c*) Supongamos que dos alumnos, A y B, se gradúan en el instituto dentro del mismo percentil, pero que el resultado de A en el test SAT de aptitud escolar es 140 puntos más alto que el de B (aproximadamente una desviación estándar en la muestra). ¿Qué diferencia podemos predecir entre la nota media universitaria de ambos alumnos? ¿Es una diferencia importante?

   *d*) Si mantenemos *hsperc* fijo, ¿qué diferencia en los resultados *SAT* nos llevaría a predecir una diferencia de 0.50 puntos (medio punto) en *colgpa*? Razona tu respuesta

2. El siguiente modelo es una versión simplificada del modelo de regresión múltiple utilizado por Bidel y Hamermesh (1990) para estudiar la relación inversa entre el tiempo dedicado a dormir y el dedicado a trabajar, junto con otros factores que afectan al sueño:

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

donde el sueño (*sleep*) y el tiempo total de trabajo (*totwrk*) se miden en minutos por semana, y la formación académica (*educ*) y la edad (*age*) se miden en años.

   *a*) Si los adultos sustituyen tiempo de sueño por trabajo, ¿cuál será el signo de $\beta_1$?

   *b*) ¿Qué signo podemos pensar que tendrán $\beta_2$ y $\beta_3$?

   *c*) Estima la ecuación por MCO usando los datos de SLEEP75.
   Si alguien trabaja cinco horas más por semana, ¿en cuántos minutos se estima que disminuirá *sleep*? ¿Es una sustitución importante?

   *d*) Argumentar el signo y la magnitud del del coeficiente estimado de *educ*.

   *e*) ¿Explican *totwrk*, *educ* y *age* una parte importante de la variación en *sleep*? ¿Qué otros factores pueden afectar al tiempo dedicado a dormir? ¿Es probable que estén correlacionadas con *totwrk*?

3. En un estudio que relaciona la nota media universitaria con el tiempo empleado en diversas actividades, se distribuye una encuesta entre un grupo de estudiantes en la que se les pregunta cuántas horas emplean en cuatro actividades: estudiar, dormir, trabajar y ocio. Cualquier actividad debe incluirse en una de las cuatro categorías, de forma que las cuatro actividades deben sumar 168 horas para cada estudiante

*a*) En el modelo

$$GPA = \beta_0 + \beta_1 study + \beta_2 sleep + \beta_3 work + \beta_4 \ leisure + u,$$

¿tiene sentido mantener fijos *sleep*, *work* y *leisure*, y modificar *study*?

*b*) Explicar por qué este modelo viola el supuesto RLM.4

*c*) ¿Cómo se podría reformular el modelo para que los parámetros tengan una interpretación útil y satisfaga el supuesto RLM.4?

4. Considera el modelo de regresión múltiple bajo los supuestos RLM.1-4,

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + u.$$

Nos interesa estimar la suma de los parámetros en $x_1$ y $x_2$ : llamémoslo $\theta_1 = \beta_1 + \beta_2$. Demostrar que $\hat{\theta}_1 = \hat{\beta}_1 + \hat{\beta}_2$ es un estimador insesgado de $\theta_1$.

5. Supongamos que la productividad media de los trabajadores en una fábrica (*avgprod*) depende de dos factores, la media de horas de capacitación laboral (*avgtrain*) y la habilidad media de los trabajadores (*avgabil*) :

$$avgprod = \beta_0 + \beta_1 avgtrain + \beta_2 avgabil + u.$$

Supongamos que esta ecuación satisface los supuestos de Gauss-Markov. Si se les dan subvenciones a aquellas fábricas cuyos trabajadores tienen una habilidad inferior a la media, de forma que *avgtrain* y *avgabil* están negativamente correlacionados, ¿cuál es el sesgo probable en $\tilde{\beta}_1$, obtenido a partir de una regresión simple de *avgprod* sobre *avgtrain*?

6. Suponemos que el modelo poblacional que determina $y$ es

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + u,$$

y que este modelo satisface los supuestos de Gauss-Markov. Sin embargo, estimamos el modelo que omite $x_3$. Supongamos que $\tilde{\beta}_0, \tilde{\beta}_1, \tilde{\beta}_2$ son los estimadores MCO de la regresión de $y$ sobre $x_1$ y $x_2$. Se pide demostrar que el valor esperado de $\tilde{\beta}_1$ (condicionado a los valores de las variables independientes en la muestra) es

$$E\left(\tilde{\beta}_1\right) = \beta_1 + \beta_3 \frac{\sum_{i=1}^n \hat{r}_{i1} x_{i3}}{\sum_{i=1}^n \hat{r}_{i1}^2},$$

donde $\hat{r}_{i1}$ son los residuos MCO de la regresión de $x_1$ sobre $x_2$. [Pista: usa la fórmula de $\tilde{\beta}_1$ en función de $\hat{r}_{i1}$ e introducir en la ecuación $y_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \beta_3 x_{i3} + u_i$].

7. Usar los datos de HPRICE1 para estimar el modelo

$$price = \beta_0 + \beta_1 sqrft + \beta_2 bdrms + u$$

donde *price* es el precio de la vivienda en miles de dólares, *sqrft* es la superficie de la vivienda en pies cuadrados y *bdrms* es el número de dormitorios.

*a*) Presentar los resultados en forma de ecuación

*b*) ¿Cuál será el aumento estimado en el precio de una vivienda con un dormitorio adicional, si mantenemos fija la superficie de la vivienda?

*c*) ¿Cuál será el aumento estimado en el precio de una vivienda con un dormitorio adicional de una superficie aproximada de 140 pies cuadrados? Comparar esta respuesta con la respuesta de (b).

*d*) ¿Qué porcentaje de la variación en el precio de la vivienda se explica por la superficie y por el número de dormitorios?

*e*) La primera vivienda de la muestra tiene $sqrft = 2438$ y $bdrms = 4$. Calcular el precio de venta estimado para esta vivienda a partir de la recta de regresión MCO.

*f*) El precio de venta verdadero de la primera vivienda de la muestra es de 300.000$ ( $price = 300$). Calcular el residuo para esta vivienda. ¿Sugiere esto que el comprador pagó un precio demasiado alto o demasiado bajo para la vivienda?

8. El archivo CEOSAL2 contiene datos sobre 177 directores generales que pueden usarse para examinar el efecto que tiene el rendimiento empresarial sobre el salario de los mismos.

*a*) Estimar un modelo que relaciona el salario con las ventas de la compañía y el valor de mercado. Especificar el modelo para que sea un modelo de elasticidad constante para ambas variables independientes. Presentar los resultados en forma de ecuación.

*b*) Añadir $profits$ al modelo de (a). ¿Por qué no se puede añadir esta variable en forma de logaritmo? ¿Explican estas variables del rendimiento empresarial la mayor parte de la variación de los salarios de los directores generales?

*c*) Añadir la variable *ceoten* al modelo de (b). ¿Cuál es el porcentaje de rendimiento estimado para cada año extra de permanencia en la empresa del director general, si mantenemos fijos el resto de los factores?

*d*) Calcular el coeficiente de correlación muestral entre las variables $\log(mktval)$ y $profits$. ¿Están estas variables estrechamente relacionadas? ¿Qué nos dice esto acerca de los estimadores MCO?