

1. (*Basado en Wooldridge (2005). Ch.3, ej.3.13*) Un problema de interés para los gobiernos es determinar los efectos del hábito de fumar durante el embarazo en la salud de los niños. Una medida de la salud de los niños es su peso en el momento de nacer, donde un peso demasiado bajo en el nacimiento puede poner a un niño en riesgo de contraer diversas enfermedades. Dado que otros factores que afectan al peso de los niños en el parto además de los cigarrillos fumados probablemente están correlacionados con el fumar, deberíamos tener en cuenta dichos factores. Por ejemplo, un ingreso mayor normalmente permite tener acceso a un mejor cuidado prenatal, así como a una mejor nutrición para la madre. Una ecuación que reconoce esto es

$$bwght = \beta_0 + \beta_1cigs + \beta_2faminc + u$$

- a) ¿Cuál es el signo más probable de β_2 ?
- b) ¿Crees que **cigs** y **faminc** probablemente estarán correlacionadas? Explica porqué la correlación puede ser positiva o negativa.
- c) Ahora, estima la ecuación con y sin **faminc** usando los datos de la base **BWGHT.RAW**. Escribe los resultados en forma de ecuación, incluyendo el tamaño muestral y el R-cuadrado. ¿Dirías que añadir **faminc** al modelo a estimar cambia sustancialmente el efecto estimado de **cigs** sobre **bwght**.
2. (*Basado en Stock y Watson (2012). Ch. 6, ej.E6.1*) Utilizando la base de datos **TeachingEval** conteste a las siguientes cuestiones:
- a) Realice una regresión del efecto que las calificaciones obtenidas en los exámenes (**CourseEval**) tienen sobre la variable que mide la belleza del profesor (**Beauty**). ¿Cuál es la pendiente estimada? Interprete su valor.
- b) Realice una regresión de la variable **CourseEval** sobre la variable **Beauty**, incluyendo algunas variables de control adicionales del tipo de curso y de las características del profesor (**Intro**, **OneCredit**, **Female**, **Minority**, **NNEnglish**). ¿Cuál es el efecto estimado de la variable **Beauty** sobre la variable **CourseEval**? ¿Presenta la regresión anterior un sesgo de variable omitida importante?
- c) Estime el coeficiente de la variable **Beauty** del modelo de regresión múltiple mediante el proceso de parcialización. Verifique que el coeficiente resultante es el mismo que el obtenido en el apartado anterior. Determine cada uno de los pasos que es necesario llevar a cabo.
- d) El profesor Smith es un hombre negro con un valor de la variable **Beauty** promedio y es angloparlante nativo. Es profesor de una asignatura de tres créditos del curso superior. Prediga la evaluación de la asignatura del profesor Smith.
3. (*Basado en Wooldridge (2005). Ch.3, ej.3.14*) Usando los datos de **HPRICE1.RAW** para estimar el modelo

$$price = \beta_0 + \beta_1sqrft + \beta_2bdrms + u$$

donde **price** representa el precio de los hogares medido en miles de dólares.

- a) Escribe los resultados en forma de ecuación. Interprete cada uno de los parámetros estimados.
- b) ¿Cuál es el aumento estimado en los precios para un hogar con una habitación más, manteniendo constante los metros cuadrados?

- c) ¿Cuál es el aumento estimado en el precio para un hogar con una habitación adicional y que tiene una dimensión de 140 metros cuadrados? Compara este resultado con tu respuesta en el apartado anterior.
 - d) ¿Qué porcentaje de la variación del precio es explicada por los metros cuadrados y cuál por el número de habitaciones?
 - e) La primera casa de la muestra tiene $sqrft = 2438$ y $bdrms = 4$. Encuentra el precio de venta predicho para esta casa partiendo de una línea de regresión MCO.
 - f) El precio actual de venta de la primera casa de la muestra era de 300.000\$ (de modo que $price = 300$). Encuentra el residuo para esta casa. ¿Esto nos sugiere que el vendedor fue pagado de más o de menos por esta casa?
4. Usa los datos de **MEAP93.RAW** para resolver esta cuestión.
- a) Estima el modelo

$$math10 = \beta_0 + \beta_1 \log(expend) + \beta_2 \lnchprg + u$$

- y escribe los resultados en la forma habitual, indicando el tamaño muestral y el R-cuadrado. ¿Son los signos de los coeficientes los que esperarías a priori? Explícalo.
- b) ¿Qué puedes decir sobre el intercepto estimado en el apartado anterior? ¿Este resultado tiene sentido para determinar que dos variables explicativas son cero? Recuerda que $\log(1) = 0$.
 - c) Ahora realiza la regresión simple de **math10** sobre **log(expend)** y compara el coeficiente de la pendiente estimado con el obtenido en el apartado anterior. ¿Es este efecto de gasto estimado ahora mayor o menor que en el primer apartado?
 - d) Encuentra la correlación entre **lexpend=log(expend)** y **lnchprg**. ¿Tiene sentido este resultado para ti?
 - e) Usa los resultados del apartado anterior para explicar los resultados obtenidos en el apartado c.
5. Use los datos de **DISCRIM.RAW** que contiene los precios de varios bienes de los restaurantes de comida rápida representados en código zip, así como características sobre la población también en código zip tanto en Nueva Jersey como en Pensilvania. La idea del estudio es ver si los restaurantes de comida rápida tienen mayores precios en áreas con una mayor concentración de personas negras.
- a) Encuentra los valores medios de **prpbck** e **income** en la muestra así como sus desviaciones estándar. Cuáles son las unidades de medida de **prpbck** e **income**?
 - b) Considera el siguiente modelo para explicar el precio de la soda, **psoda**, en términos de la proporción de la población que es blanca y del ingreso medio

$$psoda = \beta_0 + \beta_1 prpbck + \beta_2 income + u$$

Estima este modelo por MCO e indica los resultados en forma de ecuación, incluyendo el tamaño muestral y el R-cuadrado.

Interpreta el coeficiente asociado a **prpbck**. ¿Piensas que es económicamente relevante?

- c) Compara los coeficientes estimados obtenidos en el segundo apartado con los estimadores de la regresión simple de **psoda** sobre **prpbck**. ¿Es el efecto de discriminación mayor o menor que cuando controlamos por el ingreso?

- d) Un modelo con una elasticidad precio constante con respecto al ingreso puede ser más adecuado. Obtén los estimadores del siguiente modelo

$$\log(\text{psoda}) = \beta_0 + \beta_1 \text{prpblck} + \beta_2 \log(\text{income}) + u.$$

- e) Si **prpblck** aumenta en un 0.20 (20 puntos porcentuales), ¿cuál es el cambio porcentual estimado en **psoda**?
- f) Ahora añada la variable **prppov** a la regresión del apartado d). ¿Qué ocurre con $\hat{\beta}_{\text{prpblck}}$?
- g) Encuentra la correlación entre **log(income)** y **prppov**. ¿Es mayor que la que esperabas?
- h) Comenta la siguiente afirmación: “*Dado que **log(income)** y **prppov** están altamente correlacionadas, no tiene ninguna relevancia que ambas aparezcan en la misma regresión*”.
6. (Basado en Stock y Watson (2012). Ch.6, ej6.2) Con la base de datos **CollegeDistance** (western) realice los siguientes ejercicios:
- a) Realice una regresión de la variable años de educación completados **yrsed** sobre la variable de la distancia a la universidad más cercana **dist**. ¿Cuál es la pendiente estimada?
- b) Realice una regresión de la variable **yrsed** sobre la variable **dist**, pero incluyendo algunos regresores adicionales de control sobre las características del estudiante, la familia del estudiante, y el mercado laboral local. En concreto, incluya como regresores adicionales las variables **Bytest**, **Female**, **Black**, **Hispanic**, **Incomehi**, **Ownhome**, **Cue80** y **Stwmfg80**. ¿Cuál es el efecto estimado de la variable **Dist** sobre la variable **yrsed**?
- c) ¿Es sustancialmente distinto el efecto estimado de la variable **Dist** sobre la variable **ED** en la regresión de (b) de la regresión en a)? En base a esto, ¿parece que la regresión (a) presenta un sesgo de variable omitida importante?
- d) El valor del coeficiente de la variable **DadColl** es positivo. ¿Qué mide este coeficiente?
- e) Explique por qué las variables **Cue80** y **Swmfg80** aparecen en la regresión. ¿Cuáles cree que son los signos de sus coeficientes estimados? Interprete la magnitud de estos coeficientes.
- f) Bob es un hombre negro. Su escuela secundaria estaba a 20 millas de la universidad más cercana. Su calificación en la prueba **Bytest** fue de 58. Su renta familiar en 1980 fue de 26.000\$, y su familia poseía una casa. Su madre acudió a la universidad, pero su padre no. La tasa de desempleo en su condado era de 7.5%, y el promedio del salario por hora manufacturero en su estado era de 9,75\$. Estime el número de años completados de estudio por Bob utilizando la regresión de (b).
- g) Jim tiene las mismas características que Bob, salvo que su escuela secundaria estaba a 40 millas de la universidad más cercana. Estime los años completados de estudio por Jim utilizando la regresión de b).

7. (Basado en Wooldridge (2005). Ch.3, ej.3.1) Utilizando la base de datos **GPA2.RAW** formada por 4137 alumnos universitarios, se estimó la siguiente ecuación por MCO:

$$\widehat{\text{colgpa}} = 1392 - 0.0135\text{hsperc} + 0.00148\text{sat}, \quad n = 4137, \quad R^2 = 0.273.$$

donde **colgpa** es el promedio de calificaciones (medida sobre una escala de cuatro puntos), **hsperc** es el percentil en la clase de bachillerato que se gradúa (por ejemplo, $\text{hsperc} = 5$ significa el 5% superior de la clase), y **sat** es la puntuación combinada en matemáticas y habilidades verbales en la prueba de habilidad de los alumnos.

- ¿Por qué es lógico que el coeficiente asociado a **hsperc** sea negativo?
 - ¿Cuál es el promedio de calificaciones universitarias cuando $\text{hsperc} = 20$ y $\text{sat} = 1050$?
 - Suponga que dos graduados de bachillerato, A y B, se gradúan en el mismo percentil de bachillerato, pero el puntaje SAT del estudiante A es 140 puntos más alto (aproximadamente una desviación estándar en la muestra). ¿Cuál es la diferencia predicha en el promedio de calificaciones universitarias para estos dos alumnos? ¿Es grande la diferencia?
 - Manteniendo **hsperc** constante, ¿qué diferencia en las puntuaciones SAT conduce a una diferencia estimada de 0.50? Comente su respuesta.
8. (Basado en Wooldridge (2005). Ch.3, ej.3.12) Los datos en el archivo **WAGE.RAW** sobre trabajadores hombres se utilizan para estimar la ecuación siguiente:

$$\widehat{\text{educ}} = 10.36 - 0.094\text{sibs} + 0.131\text{meduc} + 0.210\text{feduc}, \quad n = 722, \quad R^2 = 0.214$$

donde **educ** es años de escolaridad, **sibs** es el número de hermanos, **meduc** es años de escolaridad de la madre y **feduc** años de escolaridad del padre.

- ¿Tiene **sibs** el efecto esperado? Explíquelo. Manteniendo constantes **meduc** y **feduc**, ¿cuánto tiene que aumentar **sibs** para tener una reducción de un año en los años de educación predichos?
 - Explique la interpretación del coeficiente de **meduc**.
 - Suponga que el hombre A no tiene hermanos, y que su madre y su padre tienen cada uno 12 años de escolaridad. El hombre B no tiene hermanos y su madre y su padre tienen cada uno 16 de escolaridad. ¿Cuál es la diferencia entre B y A en años predichos de escolaridad?
9. (Basado en Wooldridge (2005). Ch.3, ej.3.3) El siguiente modelo es una versión simplificada del modelo de regresión múltiple utilizado por Biddle y Hamermesh (1990) para estudiar el intercambio entre tiempo dedicado al sueño y dedicado al trabajo, así como ver otros factores que afectan el sueño:

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

donde **sleep** y **totwrk** (trabajo total) se miden en minutos por semana y **educ** y **age** se miden en años.

- Si los adultos intercambian sueño por trabajo, ¿cuál es el signo de β_1 ?
- Utilizando los datos del archivo **SLEEP75.RAW**, la ecuación estimada es

$$\widehat{\text{sleep}} = 3628.25 - 0.148\text{totwrk} + 11.13\text{educ} + 2.20\text{age}, \quad n = 706, \quad R^2 = 0.113.$$

Si una persona trabaja cinco horas más a la semana, ¿cuántos minutos se predice que disminuya **sleep**? ¿Es este un intercambio grande?

- c) Analice el signo y la magnitud del coeficiente estimado para **educ**.
- d) ¿Diría que **totwrk**, **educ** y **age** explican gran parte de la variación en **sleep**? ¿Qué otros factores podrían afectar al tiempo dedicado al sueño? ¿Es probable que estén correlacionados con **totwrk**?
10. (Basado en Wooldridge (2005). Ch.3, ej.3.12) Responda a las siguientes cuestiones:

- a) Considere el modelo de regresión simple $y = \beta_0 + \beta_1 x + u$ bajo los primeros cuatro supuestos de Gauss-Markov. Para una función $g(x)$, por ejemplo $g(x) = x^2$ o $g(x) = \log(1 + x^2)$, define $z_i = g(x_i)$. Siendo el estimador de la pendiente

$$\tilde{\beta}_1 = \frac{\sum_{i=1}^n (Z_i - \bar{Z}) Y_i}{\sum_{i=1}^n (Z_i - \bar{Z}) X_i},$$

muestre que $\tilde{\beta}_2$ es un estimador lineal e insesgado.

- b) Agregue el supuesto de homocedasticidad **RLM.5** y demuestre que

$$\text{Var}(\tilde{\beta}_1) = \frac{\sigma^2 \sum_{i=1}^n (Z_i - \bar{Z})^2}{(\sum_{i=1}^n (Z_i - \bar{Z}) X_i)^2}.$$

- c) Muestre de manera directa que, bajo los supuestos de Gauss-Markov, $\text{Var}(\hat{\beta}_1) \leq \text{Var}(\tilde{\beta}_1)$, donde $\hat{\beta}_1$ es el estimador de MCO.

11. (Basado en Wooldridge (2005). Ch.6, ej.6.9) Usando los datos de WAGE1.RAW conteste a las siguientes cuestiones:

- a) Utilice el método de MCO para estimar la siguiente ecuación y muestre los resultados del modo habitual.

$$\log(\text{wage}) = \beta_0 + \beta_1 \text{educ} + \beta_2 \text{exper} + \beta_3 \text{exper}^2 + u.$$

- b) Interprete los coeficientes.
- c) ¿En qué punto el efecto marginal de **educ** se convierte en negativo?
- d) ¿Cuál es el efecto aproximado de 20 años de experiencia sobre los salarios?

12. (Basado en Wooldridge (2005). Ch.6, ej. 6.12) Utilizando los datos de HPRICE1.RAW, responda a las siguientes cuestiones.

- a) Estime el siguiente modelo y muestre los resultados del modo habitual.

$$\log(\text{price}) = \beta_0 + \beta_1 \log(\text{lotsize}) + \beta_2 \log(\text{sqrft}) + \beta_3 \text{bdrms} + u$$

- b) Encuentre el valor predicho de **log(price)**, cuando $\text{lotsize} = 20000$, $\text{sqrft} = 2500$ y $\text{bdrms} = 4$.