

1. (Basado en Wooldridge (2009), Ch. 7, ej.C7.2) Usa los datos en **WAGE2.RAW** para estimar el siguiente modelo

$$\begin{aligned} \log(\text{wage}) &= \beta_0 + \beta_1 \text{educ} + \beta_2 \text{exper} + \beta_3 \text{tenure} + \beta_4 \text{married} + \beta_5 \text{black} + \beta_6 \text{south} \\ &+ \beta_7 \text{urban} + u \end{aligned}$$

- a) Indica los resultados del modo habitual. Manteniendo todos los factores fijos, ¿cuál es la diferencia aproximada en salario mensual entre hombres blancos y negros? ¿Es esta diferencia estadísticamente significativa?
- b) Añade las variables exper^2 y tenure^2 a la ecuación y muestra que son conjuntamente insignificantes incluso al nivel del 10%.
- c) Extiende el modelo original para permitir que el rendimiento educativo dependa de la raza y evalúa si el rendimiento educativo no depende de la raza.
- d) De nuevo, comienza con el modelo original, pero ahora permite que los salarios difieran entre 4 grupos de personas: casadas y negras, casadas y blancas, solteras y negras y solteras y blancas. ¿Cuál es el diferencial salarial estimado entre mujeres casadas negras y mujeres casadas blancas?
2. (Basado en Wooldridge (2009), Ch. 7, ej.C7.4) Usa los datos en **GPA2.RAW** para este ejercicio.

- a) Considera la siguiente ecuación

$$\text{colgpa} = \beta_0 + \beta_1 \text{hsize} + \beta_2 \text{hsize}^2 + \beta_3 \text{hsperc} + \beta_4 \text{sat} + \beta_5 \text{female} + \beta_6 \text{athlete} + u$$

donde **colgpa** es la puntuación media en la carrera universitaria, **hsize** es el tamaño de la clase de los graduados en el instituto (en miles), **hsperc** es el percentil académico en la clase que se gradúa, **sat** es la puntuación SAT (selectividad), **female** es una variable de género binario y **athlete** es una variable binaria que es uno para los estudiantes que son atletas. ¿Cuáles son tus expectativas para los coeficientes de esta ecuación? ¿Sobre cuál de ellos no estás seguro?

- b) Estima la ecuación del apartado a) y reporta los resultados del modo habitual. ¿Cuál es el diferencial de GPA estimado entre atletas y no atletas? ¿Es estadísticamente significativo?
- c) Elimina **sat** del modelo y reestima la ecuación. Ahora, ¿cuál es el efecto estimado de ser atleta? Discute porqué el estimador es distinto del obtenido en el apartado b).
- d) En el modelo del apartado a), permite que el efecto de ser atleta difiera por género y evalúa la hipótesis nula de que no existe, ceteris paribus, diferencia entre atletas mujeres y no atletas mujeres.
- e) ¿El efecto de **sat** sobre **colgpa** difiere por género? Justifica tu respuesta.
3. (Basado en Wooldridge (2009), Ch. 7, ej.C7.6) Usa los datos en **SLEEP75.RAW** para este ejercicio. La ecuación de interés es

$$\text{sleep} = \beta_0 + \beta_1 \text{totwrk} + \beta_2 \text{educ} + \beta_3 \text{age} + \beta_4 \text{age}^2 + \beta_5 \text{yngkid} + u.$$

- a) Estima esta ecuación de manera individual para hombres y mujeres y reporta los resultados en la forma habitual. ¿Hay diferencias notables en las dos ecuaciones estimadas?

- b) Calcula el test de Chow para la igualdad de los parámetros en la ecuación de sueño para hombres y mujeres. Usa la forma del test que añade **male** y los términos de interacción $\text{textbf{male}} * \text{totwkr}, \dots, \text{male} * \text{yngkid}$ y usa el conjunto total de observaciones. ¿Cuáles son los grados de libertad relevantes para este test? ¿Deberías rechazar la hipótesis nula al nivel del 5%?
- c) Ahora, permite un término de intercepto distinto para hombres y mujeres y determina si los términos de interacción relacionados con **male** son conjuntamente significativos.
- d) Dados los resultados de los apartados b) y c), ¿cuál sería tu modelo final?
4. (Basado en Wooldridge (2009), Ch. 7, ej. C7.10) Un modelo que analiza que el salario de los jugadores de beisbol de la liga de primera división difiera en función de su posición en el campo es el siguiente:

$$\begin{aligned} \log(\text{salary}) &= \beta_0 + \beta_1 \text{years} + \beta_2 \text{gamesyr} + \beta_3 \text{bavg} + \beta_4 \text{hrunsyr} + \beta_5 \text{rbisyr} \\ &+ \beta_6 \text{runsyr} + \beta_7 \text{fldperc} + \beta_8 \text{allstar} + \beta_9 \text{frstbase} \\ &+ \beta_{10} \text{scndbase} + \beta_{11} \text{thrdbase} + \beta_{12} \text{shrtstop} + \beta_{13} \text{catcher} + u \end{aligned}$$

donde “jardinero” es el grupo base.

- a) Establece la hipótesis nula de que, controlando por otros factores, **cátchers** y **lanzadores** ganan, en media, la misma cantidad. Evalúa esta hipótesis usando los datos en **MLB1.RAW** y comenta el tamaño del diferencial salarial estimado.
- b) Establece y evalúa la hipótesis nula de que no existe diferencia en el salario medio entre las posiciones, una vez que el resto de factores han sido controlados.
- c) ¿Son los resultados de los apartados a) y b) consistentes? Si no es así explica qué está pasando.
5. (Basado en Wooldridge (2009), Ch. 7, ej. 7.1) Con ayuda de los datos del archivo **SLEEP75.RAW** estime el siguiente modelo

$$\text{sleep} = \beta_0 + \beta_1 \text{totwrk} + \beta_2 \text{educ} + \beta_3 \text{age} + \beta_4 \text{age}^2 + \beta_5 \text{male} + u,$$

donde la variable **sleep** es la cantidad total de minutos por semana dormidos durante la noche, **totwrk** es la cantidad total de minutos que se trabajó por semana, **educ** y **age** (edad) están dadas en años y **male** (hombre) es una binaria para género.

- a) Permaneciendo todo lo demás constante, ¿hay alguna evidencia de que los hombres duerman más que las mujeres? ¿Qué tan fuerte es esa evidencia?
- b) ¿Hay un coste de oportunidad estadísticamente significativo entre trabajar y dormir? ¿Cuál es el coste de oportunidad estimado?
- a) ¿Qué otra regresión necesita correr para probar la hipótesis nula de que, manteniendo todos los demás factores constantes, la edad no afecta al dormir?
6. (Basado en Wooldridge (2009), Ch. 7, ej. 7.2) Las ecuaciones siguientes se estimaron empleando los datos del archivo **BWGHT.RAW**:

$$\begin{aligned} \widehat{\log(\text{bwght})} &= 4.65 - 0.044 \log(\text{faminc}) + 0.0093 \log(\text{faminc}) + 0.016 \text{parity} + 0.027 \text{male} \\ &\quad (0.22) \quad (0.0009) \quad (0.0059) \quad (0.006) \quad (0.01) \\ &+ 0.055 \text{white}, \quad n = 1388, \quad R^2 = 0.0472 \\ &\quad (0.013) \end{aligned}$$

$$\begin{aligned} \widehat{\log(\text{bwght})} &= 4.65 - 0.0052 \text{cigs} + 0.011 \log(\text{faminc}) + 0.017 \text{parity} + 0.034 \text{male} \\ &\quad (0.38) \quad (0.001) \quad (0.0085) \quad (0.006) \quad (0.011) \\ &+ 0.045 \text{white} - 0.003 \text{motheduc} + 0.0032 \text{fatheduc}, \quad n = 1191, \quad R^2 = 0.0493 \\ &\quad (0.015) \quad (0.003) \quad (0.0026) \end{aligned}$$

donde **male** es una variable binaria que indica si el niño es varón y **white** indica si el niño se clasifica como de raza blanca.

- a) Interprete el coeficiente de la variable **cigs** en la primera ecuación. En particular, ¿qué efecto tiene sobre el peso al nacer que la madre haya fumado 10 cigarrillos más por día?
 - b) En la primera ecuación, permaneciendo todos los demás factores constantes, ¿cuánto más se predice que pese un niño blanco en comparación con uno no blanco? ¿Es esta diferencia significativa?
 - c) Analice el efecto estimado y la significancia estadística de **mothereduc**.
 - d) Con base en la información dada, ¿por qué no puede usted calcular el estadístico F para la significancia conjunta de **mothereduc** y **fathereduc**? ¿Qué tendría que hacer para calcular este estadístico F ?
7. (Basado en Wooldridge (2009), Ch. 7, ej.7.3) Con ayuda de los datos del archivo **GPA2.RAW** estime el siguiente modelo:

$$sat = \beta_0 + \beta_1 hsize + \beta_2 hsize^2 + \beta_3 female + \beta_4 black + \beta_5 female * black + u$$

donde la variable **sat** es la puntuación obtenida en el examen de admisión (SAT), **hsize** es la cantidad de alumnos (en cientos) que en la escuela del estudiante terminaron con él el bachillerato, **female** es una variable binaria para el género femenino y **black** es una variable binaria para raza, que es igual a uno para negros y cero para los no negros.

- a) ¿Existe alguna evidencia fuerte de que **hsize** deba ser incluida en el modelo? Con base en esta ecuación, ¿cuál es el tamaño óptimo de una escuela?
 - b) Manteniendo **hsize** constante, ¿cuál es la diferencia estimada en la puntuación del SAT entre mujeres no negras y hombres no negros? ¿Qué tan estadísticamente significativa es esta diferencia?
 - c) ¿Cuál es la diferencia estimada en la puntuación del SAT entre hombres no negros y hombres negros? Pruebe la hipótesis nula de que no hay diferencia entre estas puntuaciones contra la alternativa de que sí hay diferencia.
 - d) ¿Cuál es la diferencia estimada en la puntuación del SAT entre mujeres negras y mujeres no negras? ¿Qué se necesita hacer para probar si esta diferencia es estadísticamente significativa?
8. (Basado en Wooldridge (2009), Ch. 7, ej.7.4) Una ecuación que explica el sueldo de los presidentes de consejos de administración es

$$\log(salary) = \beta_0 + \beta_1 \log(sales) + \beta_2 roe + \beta_3 finance + \beta_4 consprod + \beta_5 utility + u$$

donde **finance**, **consprod** y **utility** son variables binarias que corresponden a las industrias financiera, de productos de consumo y servicios. La industria que se ha omitido es la del transporte.

- a) Utilice los datos de **CEOSAL1.RAW** para estimar el modelo anterior y exprese los resultados del modo habitual.
- b) Calcule la diferencia porcentual aproximada entre los sueldos estimados de las industrias de servicios y de transporte, manteniendo **sales** y **roe** constantes. ¿Es esta diferencia estadísticamente significativa al nivel de significación de 1%?

- c) Emplee la ecuación $100 \cdot [\exp(\hat{\beta}_1) - 1]$ para obtener una diferencia porcentual exacta entre los sueldos estimados de las industrias de servicios y de transporte, y compare el resultado con el obtenido en el apartado a).
- d) ¿Cuál es la diferencia porcentual aproximada entre los sueldos estimados de la industria de productos de consumo y en el sector financiero? Dé una ecuación que permita probar si esta diferencia es estadísticamente significativa.
9. (Basado en Wooldridge (2009), Ch. 7, ej.7.8) Suponga que mediante una encuesta recolecta usted datos sobre salarios, educación, experiencia y género. Solicita también información sobre uso de la marihuana. La pregunta original es “¿cuántas veces fumó marihuana el mes pasado”?
- a) Dé una ecuación que permita estimar el efecto de fumar marihuana sobre el salario, controlando los demás factores. La ecuación deberá permitir hacer afirmaciones como “se estima que fumar marihuana cinco veces al mes hace que el salario varía $x\%$ ”.
- b) Formule una ecuación que permita probar si el uso de drogas tiene efectos diferentes sobre los salarios de hombres y mujeres. ¿Cómo puede probarse que los efectos del uso de drogas no son diferentes entre hombres y mujeres?
- c) Suponga que considera que es mejor medir el uso de marihuana clasificando a las personas en cuatro categorías: no usuario, usuario suave (1 a 5 veces por mes), usuario moderado (6 a 10 veces por mes) y usuario fuerte (más de 10 veces por mes). Ahora, diseñe un modelo que permita estimar el efecto del uso de la marihuana sobre el salario.
- d) Usando el modelo del apartado c), explique detalladamente cómo probar la hipótesis nula de que fumar marihuana no tiene ningún efecto sobre el salario. Sea muy específico y dé una lista detallada de los grados de libertad.
- e) ¿Cuáles son algunos de los problemas potenciales de hacer inferencias causales a partir de los datos recolectados por usted?