1. (*Basado en Stock y Watson (2012). Ch.4, ej.E.4.1*) Utilizando la base de datos **CPS04** de Stock y Watson (2012) que contiene datos sobre una muestra de trabajadores a tiempo completo en el año 2004, de entre 25 y 34 años, titulados en la escuela secundaria o licenciados/ingerieros (grado más alto de educación alcanzado).

   a) Determine cuál es el modelo de regresión a estimar cuando queremos explicar el impacto de la edad de un trabajador sobre sus ingresos salariales.

   b) Realice una regresión de los ingresos medios por hora **AHE** sobre la edad **Age**. ¿Cuál es el término independiente estimado? ¿Cuál es la pendiente estimada? ¿Cuánto aumentarán los ingresos al aumentar la edad de los trabajadores en un año?

   c) Bob es un trabajador de 26 años de edad. Prediga los ingresos de Bob a partir de la regresión estimada. Alexis es un trabajador de 30 años de edad. Prediga los ingresos de Alexis utilizando la regresión estimada.

   d) ¿Representa la edad una gran proporción de la varianza de los ingresos entre los individuos? Explíquelo.

2. (*Basado en Stock y Watson (2012). Ch.4. ej.E.4.2*) Utilizando la base de datos **TeachingEval** de Stock y Watson (2012) que contiene datos sobre las evaluaciones de una asignatura, las características de la asignatura y las características del profesor para 463 cursos de la Universidad de Texas en Austin. Una de las características es un índice de "belleza" del profesor de acuerdo con la clasificación de un jurado de seis jueces. En este ejercicio se investigará cómo las evaluaciones del curso están relacionadas con la "belleza" del profesor.

   a) Construya un diagrama de dispersión para las evaluaciones medias del curso **CourseEval** sobre la belleza del profesor **Beauty**. ¿Parece haber una relación entre las variables?

   b) Realice una regresión de las evaluaciones medias del curso **CourseEval** sobre la belleza del profesor **Beauty**. ¿Cuál es el término independiente estimado? ¿Cuál es la pendiente estimada? Explique por qué el término independiente estimado es igual a la media muestral de la variable **CourseEval**.

   c) ¿Explica **Beauty** una proporción grande de la varianza de las evaluaciones entre los cursos? Explíquelo.

3. (*Basado en Wooldridge (2005). Ch.2, ej.2.11*) El conjunto de datos en **CEOSAL2.RAW** contiene información sobre los directores ejecutivos (CEO) para empresas de U.S. La variable **salary** es el sueldo anual, en miles de dólares y **ceoten** es el número de años que lleva ocupando el cargo de CEO.

   a) Encuentra el salario medio y la antigüedad media de la muestra.

   b) ¿Para cuántos CEOs éste es su primer año como CEO (es decir, $ceoten = 0$)? ¿Cuál es la mayor antigüedad como CEO?

   c) Estime el siguiente modelo de regresión simple e indique los resultados de la forma habitual. ¿Cuál es el porcentaje predicho (aproximado) de aumento salarial dado un año más como CEO?

$$log(salary) = \beta_0 + \beta_1 ceoten + u.$$

4. (*Basado en Stock y Watson (2012). Ch.4, ej.E.4.3.*) Utilizando el archivo de datos **CollegeDistance** de Stock y Watson (2012) que contiene datos de una muestra aleatoria de alumnos del último año de secundaria entrevistados en 1980 y vueltos a entrevistar en 1986 tratamos de investigar la relación existente entre el número de años de educación completados por adultos jóvenes y la distancia entre la escuela secundaria de cada estudiante y la universidad más cercana. (Nota: La proximidad a la universidad reduce el coste de la educación, por lo que los estudiantes que viven más cerca de una unviersidad deberían, en promedio, completar más años de educación superior):

   a) Realice una regresión de los años completados de educación **ED** sobre la distancia a la universidad más cercana **Dist**, donde **Dist** está medida en decenas de millas. (Por ejemplo, $Dist = 2$ significa que la distancia es de 20 millas.) ¿Cuál es la estimación para el término independiente? ¿Cuál es la pendiente estimada? ¿Cuánto cambia el valor medio de los años de educación completados cuando las universidades se contruyen cerca de donde los estudiantes acuden a la escuela secundaria?

   b) La escuela secundaria de Bob estaba a 20 millas de la universidad más cercana. Prediga los años de educación completados por Bob utilizando la regresión estimada. ¿Cómo cambiarían los pronósticos sobre Bob si hubiera vivido a 10 millas de la universidad más cercana?

   c) ¿Explica la distancia a la universidad una proporción grande de la varianza de los logros educativos entre los individuos? Explíquelo.

5. (*Basado en Wooldridge (2005). Ch.2, ej.2.14*) Para la población de empresas en la industria farmacéutica, siendo **rd** los gastos anuales en investigación y desarrollo, y siendo **sales** las ventas anuales (ambas en millones de dólares).

   a) Escribe un modelo (no una ecuación estimada) donde aparezca una elasticidad constante entre **rd** y **sales**. ¿Qué parámetro es la elasticidad?

   b) Estima el modelo usando los datos de **RDCHEM.RAW**. Escribe la ecuación estimada en la forma habitual. ¿Cuál es la elasticidad estimada de **rd** con respecto a **sales**? Explica en palabras lo que la elasticidad implica.

6. (*Basado en Wooldridge (2005). Ch.2, ej.C.6*) Usando los datos de **MEAP93.RAW** queremos explorar la relación existente entre la tasa de aprobados en matemáticas **math10** y el gasto por estudiante **expend**.

   a) ¿Piensas que cada dólar adicional de gasto tiene el mismo efecto sobre la tasa de aprobados o un efecto de disminución parece ser más apropiado? Explícalo.

   b) En el modelo poblacional

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

   argumenta si $\beta_1/10$ es el cambio asociado a una modificación de un punto porcentual en **math10** dado un aumento de un 10% en **expend**.

   c) Usa los datos de **MEAP93.RAW** para estimar el modelo del apartado b). Indica la ecuación estimada del modo habitual, incluyendo el tamaño muestral y el R^2.

   d) ¿Cómo de grande es el efecto del gasto estimado? Si el gasto aumenta en un 10%, ¿cuál es el aumento porcentual estimado en **math10**?

   e) Uno podría preocuparse sobre si el análisis de regresión podría producir valores ajustados para **math10** mayores que 100. ¿Por qué en este conjunto de datos la preocupación no es muy grande?

7. (*Basado en Stock y Watson (2012). Ch.4, ej.4.1*) Suponga que un investigador utiliza datos sobre el tamaño de las clases **TC** y los promedios de las calificaciones obtenidas en los exámenes para 100 clases de tercer curso, para estimar la regresión MCO

$$\widehat{cal.examen} = 520,4 - 5,82TC, \qquad R^2 = 0,08.$$

   a) Un aula tiene 22 estudiantes. ¿Cuál es la predicción de la regresión para la calificación media en el examen de esa clase?

   b) El año pasado el aula tenía 19 estudiantes, y este año cuenta con 23 alumnos. ¿Cuál es la predicción de la regresión para la variación en la media de las calificaciones en el examen para la clase?

   c) La media muestral del tamaño de la clase para 100 aulas es de $21,4$. ¿Cuál es la media muestral de las calificaciones en el examen entre las 100 aulas?

   d) ¿Cuál es la desvación típica muestral de las calificaciones en los exámenes entre las 100 aulas?

8. (*Basado en Wooldridge (2005). Ch.2. ej.2.12*) Usa los datos de **SLEEP75.RAW** de Biddle y Hamermesh (1990) para estudiar si existe una relación entre el tiempo destinado a dormir a la semana y el tiempo empleado en un trabajo remunerado. Podríamos usar cualquiera de ellas como variable dependiente. Para concretar, estimamos el modelo

$$sleep = \beta_0 + \beta_1 totwrk + \epsilon$$

   donde **sleep** es el tiempo (medido en minutos) destinado a dormir por semana y **totwrk** representa los minutos totales trabajados durante la semana.

   a) Indica los resultados de la ecuación así como el número de observaciones y el R^2. ¿Qué implica el intercepto en esta ecuación?

   b) Si **totwrk** aumenta en 2 horas, ¿en cuánto se estima que **sleep** vaya a caer? ¿Consideras que este es un efecto grande?

9. (*Basado en Wooldridge (2005). Ch.2. ej.2.13*) Usa los datos en **WAGE2.RAW** para estimar una regresión simple que explica el salario mensual **wage** en términos de la puntuación **IQ**.

   a) Encuentra el salario medio y el **IQ** medio de la muestra. ¿Cuál es la desviación estándar de **IQ** de la muestra? (Las puntuaciones de **IQ** están estandarizadas de modo que la media poblacional es 100 y la desviación estándar igual a 15).

   b) Estima un modelo de regresión simple donde un aumento de un punto en **IQ** cambia **wage** en 1 dólar. Usa el modelo para encontrar el aumento predicho en **wage** asociado con un aumento en **IQ** de 15 puntos. ¿Explica entonces **IQ** gran parte de la variación en **wage**?

   c) Estima un modelo donde cada aumento en un punto en **IQ** tiene el mismo efecto porcentual en **wage**. Si **IQ** aumenta en 15 puntos, ¿cuál es el aumento porcentual aproximado en el **wage** predicho?

10. (*Basado en Wooldridge (2005). Ch.2. ej.2.7*) Considere la función de ahorro

$$sav = \beta_0 + \beta_1 inc + u, \qquad u = \sqrt{inc} \cdot e$$

   donde es una variable aleatoria con $E(e) = 0$ y $Var(e) = \sigma_\epsilon^2$. Suponga que e es independiente de **inc**.

a) Muestre que $E(u|inc) = 0$, de manera que el supuesto clave de media condicional cero (supuesto RLS.4) se satisface.

b) Muestre que $Var(u|inc) = \sigma_\epsilon^2 inc$, de manera que se viola el supuesto RLS.5 de homocedasticidad. En particular, la varianza de **sav** aumenta con **inc**.

11. (*Basado en Wooldridge (2005). Ch.2. ej.2.4*) La base de datos **BWGHT.RAW** contiene cifras sobre los hijos nacidos de mujeres en Estados Unidos. Las dos variables de interés son la variable independiente, peso en onzas del niño al nacer **bwght** y la variable explicativa, cantidad diaria de cigarros consumidos por la madre durante el embarazo **cigs**. La siguiente ecuación de regresión simple se estimó con datos de $n = 1388$ nacimientos:

$$\widehat{bwght} = 119.77 - 0.514cigs$$

a) ¿Cuál es el peso al nacer que se predice si $cigs = 0$? ¿Y cuando $cigs = 20$ (un paquete diario)? Analice la diferencia.

b) ¿Capta esta ecuación de regresión simple una relación causal entre el peso del niño al nacer y el hábito de fumar de la madre? Explique.

c) Para que el peso al nacer predicho sea de 125 onzas, ¿cuál tiene que ser el valor de **cigs**? Explique.

d) La proporción de mujeres en la muestra que no fumaron durante el embarazo es aproximadamente 0.85. ¿Ayuda esto a entender sus hallazgos del apartado c)?

12. (*Basado en Wooldridge (2005), Ch.2. ej.2.6*) Usando los datos de Kiel y McClain (1995) sobre las casas vendidas en 1988 en Andover, Massachusetts, en la ecuación siguiente se relaciona el precio de las casas **price** con la distancia a una incineradora de basura construida recientemente **dist**:

$$\widehat{log(price)} = 9.40 + 0.310log(dist), \quad n = 135, \quad R^2 = 0.162.$$

a) Interprete el coeficiente de **log(dist)**. ¿Es el signo de esta estimación el que se esperaba?

b) ¿Considera que la regresión simple proporciona un estimador insesgado de la elasticidad *ceteris paribus* de **price** respecto a **dist**?

c) ¿Qué otros factores relacionados con una casa afectan a su precio? ¿Pueden estos factores estar correlacionados con la distancia a la incineradora?