1. (*Basado en Stock & Watson (2012), Ch. 5, ej.E5.1*) Con la base de datos **CPS08** ejecute una regresión de los ingresos salariales medios por hora (**AHE**) sobre la variable edad (**Age**) y realice los siguientes ejercicios.

   a) ¿Es estadísticamente significativo el coeficiente de la pendiente de la regresión estimado? ¿Cuál es el p-valor asociado al estadístico t del coeficiente?

   b) Construya un intervalo de confianza al 95% para el coeficiente de la pendiente.

   c) Repita el ejercicio a) utilizando sólo los datos de los graduados de escuela secundaria.

   d) Repita el ejercicio a) utilizando sólo los datos de los graduados universitarios.

   e) ¿Es distinto el efecto de la variable **Age** sobre los ingresos salariales para los graduados de secundaria que para los graduados univerrsitarios? Explíquelo.

2. (*Basado en Stock & Watson (2012), Ch. 7, ej. E7.1*) Utilice la base de datos **CPS08** para responder a las siguientes cuestiones.

   a) Realice una regresión de los ingresos medios por hora (**AHE**) sobre la variable de edad (**Age**). ¿Cuál es el intercepto estimado? ¿Cuál es la pendiente estimada?

   b) Realice una regresión de la variable **AHE** sobre la variable (**Age**), la variable género (**Female**), y la variable educación (**Bachelor**). ¿Cuál es el efecto estimado de la variable **Age** sobre los ingresos? Construya un intervalo de confianza al 95% para el coeficiente de la variable (**Age**) en la regresión.

   c) ¿Son los resultados de la regresión de b) sustancialmente diferentes de los resultados de a) con respecto a los efectos de la variable **Age** sobre la variable **AHE**? ¿Parece que exista en la regresión a) sesgo de variable omitida?

   d) ¿Son el género y la educación factores determinantes de los ingresos? Contraste la hipótesis nula de que la variable **Female** puede eliminarse de la regresión. Contraste la hipótesis nula de que la variable **Bachelor** puede eliminarse de la regresión. Contraste la hipótesis de que tanto la variable **Female** como la variable **Bachellor** se pueden eliminar de la regresión.

3. (*Basado en Stock & Watson (2012), Ch. 7, ej. E7.4*) Con la base de datos **Growth**, excluyendo los datos de Malta, lleve a cabo los siguientes ejercicios:

   a) Realice una regresión de la variable **Growth** sobre las variables **TradeShare**, **YearsSchool**, **RevCoups**, **Assessinations** y **RGDP60**. Construya un intervalo de confianza al 95% para el coeficiente de la variable **TradeShare**. ¿Es estadísticamente significativo el coeficiente al nivel del 5%?

   b) Compruebe si, como grupo, las variables **YearsSchool**, **RevCoups**, **Assessinations** y **RGDP60** pueden ser omitidas de la regresión. ¿Cuál es el p-valor del estadístico F?

4. (*Basado en Wooldridge (2009), Ch. 4, ej.C4.1*) El siguiente modelo puede ser usado para estudiar si los gastos de campaña política pueden afectar a los resultados de las elecciones:

$$voteA = \beta_0 + \beta_1 log(expendA) + \beta_2 log(expendB) + \beta_3 prtystrA + u$$

   donde **voteA** es el porcentaje de votos recibido por el Candidato A, **expendA** y **expendB** son gastos de campaña tanto del Candidato A como del Candidato B, y **prtystrA** es una medida sobre la fortaleza del partido del Candidato A (el porcentaje de los votos recibidos por el partido A en la última votación presidencial).

a) ¿Cuál es la interpretación de β_1?

b) En términos de los parámetros, establece la hipótesis nula de que un aumento de un 1% en gastos del Candidato A es compensado por un aumento de un 1% en los gastos del candidato B.

c) Estima el modelo anterior usando los datos de **VOTE1.RAW** y expresa los resultados de la forma habitual. ¿Puedes concluir que los gastos del Candidato A afectan al resultado? ¿Qué puedes decir sobre los gastos del Candidato B? ¿Puedes usar estos resultados para evaluar la hipótesis del apartado b)?

d) Estima un modelo que directamente nos proporciona el t-estadístico para evaluar la hipótesis del apartado b). ¿Qué puedes concluir? Usa una alternativa bilateral.

5. (*Basado en Wooldridge (2009), Ch. 4, ej.C4.2*) Usa los datos en **LAWSCH85.RAW** para este ejercicio.

a) Estima el siguiente modelo

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

y establece y evalúa la hipótesis nula de que posición alcanzada en el ranking de la facultad de derecho (siendo $rank = 1$ el mejor resultado) no tiene, ceteris paribus, efecto sobre el salario medio de inicio.

b) Siendo **LSTAT** la puntuación media obtenida en la carrera y **GPA** la nota media de la clase, ¿puedes concluir que existen características sobre las puntuaciones obtenidas por los estudiantes (basándonos en estas dos variables) que son individual o conjuntamente significativas para la variable a explicar **salary**? Asegúrate de tener en cuenta los datos no existentes para **LSAT** y **GPA**.

c) Evalúa si el tamaño de clase (**clsize**) o el tamaño de la facultad (**faculty**) debe ser añadido a esta ecuación llevando a cabo una sola prueba. Ten cuidado de tener en cuenta los datos inexistentes de **clsize** y **faculty**.

d) ¿Qué factores pueden influir sobre el ranking de la facultad de derecho que no están incluidas en la regresión de salario?

6. (*Basado en Wooldridge (2009), Ch. 4, ej.C4.3*) Usando el logaritmo de los precios de las casas como variable dependiente tenemos

$$log(price) = \beta_0 + \beta_1 sqrft + \beta_2 bdrms + u$$

a) Estima esta regresión y obtén un intervalo de confianza para el cambio porcentual en price cuando añadimos una habitación de 150 metros cuadrados a una casa. En forma decimal esto es $\theta_1 = 150\beta_1 + \beta_2$. Usa los datos de la base **HPRICE1.RAW** para estimar θ_1.

b) Expresa β_2 en términos de θ_1 y de β_1 e introdúcelo en la ecuación de **log(price)**.

c) Utiliza el apartado b) para obtener un error estándar para $\widehat{\theta}_1$ y utiliza este error estándar para construir un intervalo de confianza del 95%.

7. (*Basado en Wooldridge (2009), Ch. 4, ej.C4.4*) Usando los datos de **BWGHT.RAW** calcula el R-cuadrado de la regresión de **bwght** sobre **cigs**, **parity** y **faming** usando todas las observaciones. Compara este R-cuadrado con el que obtendrías si también tuviéramos en cuenta **motheduc** y **fatheduc**.

8. (*Basado en Wooldridge (2009), Ch. 4, ej.C4.5*) Usa los datos de **MLB1.RAW** para este ejercicio.

   a) Estima el modelo

   $$log(salary) = \beta_0 + \beta_1 years + \beta_2 gamesyr + \beta_3 hrunsyr + \beta_4 rbisyr + u$$

   ¿Qué ocurre con la significatividad estadística de **hrunsyr** si suprimimos la variable **rbisyr**? ¿Qué ocurre con el tamaño del coeficiente asociado a **hrunsyr**?

   b) Añade las variables **runsyr** (carreras por año), **fldperc** (porcentaje de carreras en el campo) y **sbasesyr** (balones robados al año) al modelo del apartado a). ¿Alguno de estos factores son individualmente significativos?

   c) En el modelo del apartado b), evalúa la significatividad conjunta de **bavg**, **fldperc** y **sbasesyr**.

9. (*Basado en Wooldridge (2009), Ch. 4, ej.C4.6*) Usa los datos de **WAGE2.RAW** para realizar este ejercicio.

   a) Considera la siguiente ecuación de salarios estándar

   $$log(wage) = \beta_0 + \beta_1 educ + \beta_2 exper + \beta_3 tenure + u$$

   Establece la hipótesis nula de que un año adicional de experiencia laboral general tiene el mismo efecto sobre **log(wage)** que un año adicional de experiencia en el trabajo actual.

   b) Evalúa la hipótesis nula del apartado a) contra una alternativa bilateral, a un nivel de significatividad del 5%, construyendo un intervalo de confianza del 95%. ¿Qué puedes concluir de los resultados?

10. (*Basado en Wooldridge (2009), Ch. 4, ej.C4.7*) Usando los datos de **TWOYEAR.RAW** responde a las siguientes cuestiones.

    a) La variable **phsrank** es el percentil obtenido en el instituto por los alumnos. (NOTA: un número mayor es mejor. Por ejemplo, 90 implica que estás mejor clasificado que el 90% del resto de alumnos de la clase). Encuentra los valores menor, mayor y medio de **phsrank** en la muestra.

    b) Estima el siguiente modelo

    $$log(wage) = \beta_0 + \beta_1 jc + \beta_2 totcoll + \beta_3 exper + u$$

    y expresa los estimadores MCO del modo habitual.

    b) Estima el siguiente modelo

    $$log(wage) = \beta_0 + \beta_1 jc + \beta_2 totcoll + \beta_3 exper + \beta_4 phsrank + u$$

    y expresa los estimadores MCO del modo habitual. ¿Puedes concluir que **phsrank** es estadísticamente significativa? ¿Cuál es el valor de 10 puntos porcentuales en el ranking del bachillerato en términos de salario?

    c) Si comparamos los resultados, especialmente los relacionados con **phsrank**, de los modelos estimados en los apartados a) y b), ¿cambian sustancialmente las conclusiones sobre los rendimientos de los alumnos con 2 o 4 años de escolarización? Explícalo.

d) El conjunto de datos contiene una variable llamada **id**. Explica por qué si añadimos **id** a los dos modelos estimados en los apartados a) y b) esperamos que ésta sea estadísticamente insignificante. ¿Cuál es el p-valor de dos colas?

11. (*Basado en Wooldridge (2009), Ch. 4, ej.C4.8*) El conjunto de datos **401KSUBS.RAW** contiene información sobre la riqueza financiera neta (**nettfa**), la edad de la persona que responde la encuesta (**age**), ingreso familiar anual (**inc**), tamaño de la familia (**fsize**) y la participación en ciertos planes de pensiones en los Estados Unidos. Las variables riqueza e ingreso están medidas en miles de dólares. Para esta cuestión, sólo usamos los datos sobre hogares constituidos por una única persona (de modo que $fsize = 1$).

a) ¿Cuántos hogares formados por una única familia están en la base de datos?

b) Usa MCO para estimar el siguiente modelo

$$nettfa = \beta_0 + \beta_1 inc + \beta_2 age + u$$

e indica los resultados usando el formato habitual. Asegúrate de usar sólo los hogares formados por una única persona en la muestra. Interpreta los coeficientes de las pendientes. ¿Encuentras sorpresas en los estimadores de las pendientes?

c) ¿Tiene el intercepto de la regresión del apartado b) un significado interesante? Explícalo.

d) Encuentra el p-valor para el test $H_0 : \beta_2 = 1$ contra $H_1 : \beta_2 < 1$. ¿Rechazarías H_0 a un nivel de significatividad del 1%?

e) Si realizas una regresión simple de **nettfa** sobre **inc**, ¿el coeficiente estimado de **inc** es muy diferente del estimador obtenido en el apartado b)? ¿Por qué si o por qué no?

12. (*Basado en Wooldridge (2009), Ch. 4, ej.C4.9*) Usa los datos de **DISCRIM.RAW** para contestar a esta cuestión.

a) Usa MCO para estimar el modelo

$$log(psoda) = \beta_0 + \beta_1 prpblck + \beta_2 log(income) + \beta_3 prppov + u$$

e indica los resultados en la forma habitual. ¿Es $\widehat{\beta}_1$ estadísticamente diferente de cero al nivel del 5% contra una alternativa bilateral? ¿Qué podemos decir si trabajamos a un nivel del 1%?

b) ¿Cuál es la correlación existente entre **log(income)** y **prppov**? ¿Es cada variable estadísticamente significativa? Indica los p-valor bilaterales.

c) A la regresión del apartado a), añade la variable **log(hseval)**. Interpreta su coeficiente y reporta el p-valor bilateral para $H_0 : \beta_{log(hseval)} = 0$.

d) En la regresión del apartado c), ¿qué ocurre con la significatividad estadística individual de **log(income)** y **prppov**? ¿Son estas variables conjuntamente significativas? (Calcula un p-valor.) ¿Qué puedes decir sobre estas respuestas?

e) Dados los resultados obtenidos en las regresiones anteriores, ¿cuál reportarías como el resultado más realista a la hora de determinar si la raza de los individuos de un determinado área influye sobre los precios locales de la comida rápida?

13. (*Basado en Wooldridge (2009), Ch. 4, ej.C4.10*) Usa los datos de **ELEM94_95** para contestar a esta pregunta. La variable dependiente **lavgsal** es el logaritmo del salario medio de los profesores y **bs** es el ratio de las ganancias medias respecto del salario medio (por escuela).

   a) Realiza la regresión simple de **lavgsal** sobre **bs**. ¿Es la pendiente estimada estadísticamente distinta de cero? ¿Es estadísticamente distinta de -1?

   b) Añade las variables **lenrol** y **lstaff** a la regresión del apartado a). ¿Qué ocurre con el coeficiente asociado a **bs**?

   c) ¿Qué ocurre con el error estándar del coeficiente de **bs**?, ¿cómo puede ser menor en el apartado b) que en el apartado a)? (NOTA: ¿Qué ocurre con la varianza del error versus la multicolinealidad cuando **lenrol** y **lstaff** son añadidas?)

   d) ¿Qué indica el hecho de que el coeficiente de **lstaff** sea negativo? ¿Es una magnitud grande?

   e) Ahora añade la variable **lunch** a la regresión. Manteniendo el resto de factores fijo, ¿podemos concluir que los profesores están siendo compensados por enseñar a los estudiantes de entornos desfavorecidos?

14. (*Basado en Wooldridge (2009), Ch. 6, ej.C6.3*) Considera un modelo en el que los rendimientos educativos dependen de la cantidad de experiencia laboral (y viceversa):

$$log(wage) = \beta_0 + \beta_1 educ + \beta_2 exper + \beta_3 educ * exper + u$$

   a) Muestra que el rendimiento generado por un año más de educación (en forma decimal), manteniendo exper fijo, es $\beta_1 + \beta_3 exper$.

   b) Establece la hipótesis nula de que el rendimiento educativo no depende del nivel de expereriencia. ¿Piensas que es una alternativa adecuada?

   c) Usa los datos en **WAGE2.RAW** para evaluar la hipótesis nula en b) contra la alternativa que establezcas.

   d) Denotando θ_1 el rendimiento educativo (en forma decimal), cuando $exper = 10$; $\theta_1 = \beta_1 + 10\beta_3$. Obtén $\widehat{\theta}_1$ y un intervalo de confianza del 95% para θ_1.

15. (*Basado en Wooldridge (2009), Ch. 4, ej.4.1*) Considere una ecuación para explicar los sueldos de los directores generales o CEO en términos de las ventas anuales de la empresa, el rendimiento sobre capital (**roe**, en forma de porcentaje), y el rendimiento de las acciones de la empresa (**ros**, en forma de porcentaje):

$$log(salary) = \beta_0 + \beta_1 log(sales) + \beta_2 roe + \beta_3 roos + u.$$

   a) En términos de los parámetros del modelo, establezca la hipótesis nula de que, controlando **sales** y **roe**, **ros** no tiene efecto en el sueldo de los CEO. Establezca la alternativa de que un mejor desempeño de las acciones de la empresa incrementa el sueldo de los CEO.

   b) Con los datos de CEOSAL1.RAW, estime el modelo anterior y escribe los resultados del modo habitual. ¿Cuál es el porcentaje de aumento de salary que se pronostica si **ros** aumenta 50 puntos? ¿Tienen **ros** un efecto práctico grande sobre **salary**?

   c) Pruebe la hipótesis nula que dice que **ros** no tiene efecto sobre **salary** contra la alternativa que dice que **ros** tiene efecto positivo. Realice la prueba al nivel de significancia de 10%.

d) ¿Incluiría usted **ros** en el modelo final que explica las compensaciones de los CEO en términos del desempeño de la empresa? Explique.

16. (*Basado en Wooldridge (2009), Ch. 4, ej.4.3*) La variable **rdintens** representa el gasto en investigación y desarrollo (I&D) dado como porcentaje de las ventas. Las ventas (**sales**) se miden en millones de dólares. La variable **profmarg** representa la ganancia como porcentaje de las ventas. Empleando los datos del archivo **RDCHEM.RAW** de 32 empresas de la industria química, se estime la ecuación siguiente:

$$rdintens = \beta_0 + \beta_1 log(sales) + \beta_2 profmarg + u$$

a) Interprete el coeficiente de **log(sales)**. En particular, si sales aumenta 10%, ¿cuál es la variación estimada en puntos porcentuales en **rdintens** predicho? ¿Es este efecto económicamente grande?

b) Pruebe la hipótesis de que la intensidad de la $I\&D$ no varía con **sales** contra la alternativa de que aumenta con las ventas. Realice la prueba a los niveles de significancia de 5 y 10%.

c) Interprete el coeficiente de **profmarg**. ¿Es este coeficiente económicamente grande?

d) ¿Tiene **profmarg** un efecto estadístico significativo sobre **rdintens**?

17. (*Basado en Wooldridge (2009), Ch. 4, ej.4.5*) Utilice la base de datos **GPA1.RAW** para explicar el promedio general de calificaciones (**GPA**) obtenido en la universidad en función de la nota media obtenida en el bachillerato (**hsGPA**), la puntuación obtenida (**ACT**) y la cantidad media de faltas de asistencia a clase por semana (**skipped**).

a) Empleando la aproximación normal estándar, encuentre el intervalo de confianza de 95% para β_{hsGPA}.

b) ¿Se puede rechazar la hipótesis $H_0 : \beta_{hsGPA} = 0.4$ contra la alternativa de dos colas al nivel de 5%?

c) ¿Se puede rechazar la hipótesis $H_0 : \beta_{hsGPA} = 1$ contra la alternativa de dos colas al nivel de 5%?

18. (*Basado en Wooldridge (2009), Ch. 4, ej.4.8*) Considere el siguiente modelo de regresión múltiple con tres variables independientes, bajo los supuestos RLM.1 a RLM.6 del modelo lineal clásico:

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

Usted desea probar la hipótesis nula $H_0 : \beta_1 - 3\beta_2 = 1$.

a) Sean $\widehat{\beta}_1$ y $\widehat{\beta}_2$ los estimadores de MCO de β_1 y β_2. Encuentre $Var(\widehat{\beta}_1 - 3\widehat{\beta}_2)$ en términos de las varianzas de $\widehat{\beta}_1$ y $\widehat{\beta}_2$ de la covarianza entre ellas. ¿Cuál es el error estándar de $\widehat{\beta}_1 - 3\widehat{\beta}_2$?

b) Dé el estadístico t para probar $H_0 : \beta_1 - 3\beta_2 = 1$.

c) Defina $\theta_1 = \beta_1 - 3\beta_2$ y $\widehat{\theta}_1 = \widehat{\beta}_1 - 3\widehat{\beta}_2$. Escriba una ecuación de regresión en la que se incluyan β_0, θ_1, β_2 y β_2 que le permita obtener directamente $\widehat{\theta}_1$ y su error estándar.