Hoja de Ejercicios 3
El modelo de regresión lineal múltiple

Nota: En aquellos ejercicios en los que se incluyen estimaciones y referencia al archivo de datos utilizado, el estudiante debería comprobar los resultados obtenidos en Gretl.

1. Sea el modelo

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \varepsilon,$$

donde $E(\varepsilon|X_1, X_2) = 0$, y suponga que disponemos de una muestra de tamaño n.

(a) Derive las condiciones de primer orden de los estimadores MCO $\widehat{\beta}_0$, $\widehat{\beta}_1$, $\widehat{\beta}_2$ de los coeficientes β.

(b) Muestre que

$$\widehat{\beta}_0 = \overline{Y} - \widehat{\beta}_1 \overline{X}_1 - \widehat{\beta}_2 \overline{X}_2$$
$$\widehat{\beta}_1 = \frac{1}{D}(s_{22}s_{1y} - s_{12}s_{2y})$$
$$\widehat{\beta}_2 = \frac{1}{D}(s_{11}s_{2y} - s_{12}s_{1y})$$

donde

$$s_{11} = \tfrac{1}{n}\textstyle\sum_i x_{1i}^2, \quad s_{12} = \tfrac{1}{n}\sum_i x_{1i}x_{2i} = s_{21}, \quad s_{1y} = \tfrac{1}{n}\sum_i x_{1i}y_i$$
$$s_{22} = \tfrac{1}{n}\textstyle\sum_i x_{2i}^2, \quad s_{2y} = \tfrac{1}{n}\sum_i x_{2i}y_i$$
$$D = s_{11}s_{22} - s_{12}^2.$$

siendo $y_i = Y_i - \overline{Y}$, $x_{1i} = X_{1i} - \overline{X}_1$, $x_{2i} = X_{2i} - \overline{X}_2$.

(c) ¿Qué ocurre cuando $D = 0$? Interprete el resultado.

(d) Muestre que, cuando $s_{12} = 0$, el estimador $\widehat{\beta}_1$ coincide con el estimador $\widehat{\gamma}_1$ en la regresión simple

$$\widehat{Y} = \widehat{\gamma}_0 + \widehat{\gamma}_1 X_1,$$

e interprete el resultado.

2. ¿Cuál de las siguientes situaciones, si alguna, incumpliría los supuestos del modelo de regresión clásico?

(a) La variable X_2 es el recíproco de la variable X_1.

(b) La variable X_2 es el cuadrado de la variable X_1.

(c) La variable X_1 es una variable artificial que toma el valor 1 para mujeres y 0 para hombres, y la variable X_2 es una variable artificial que toma el valor 1 para hombres, y 0 para mujeres.

3. Aunque el vino es un bien de consumo, dadas las características de los grandes vinos de reserva puede tener sentido considerarlos como una inversión. En particular, disponemos de datos de los precios de subasta de miles de vinos tintos de reserva de Burdeos de añadas entre 1952 y 1980. Estos vinos se almacenan durante bastante tiempo antes de ser consumidos, lo que

incrementa su precio dado el coste de almacenaje, que supone un coste de oportunidad dada la posibilidad de inversiones alternativas. Nuestro fichero de datos `BORDEAUX.GDT` proporciona, para distintas añadas, información sobre LPR (logaritmo neperiano del precio del vino), *lluvinv* (Cantidad de lluvia caída en el invierno anterior a la cosecha), *tempmed* (Temperatura media durante el período de maduración de la uva), *lluvcos* (Cantidad de lluvia caída durante el período de maduración de la uva), *edad* (Número de años transcurridos desde la cosecha).

(a) Estime la proyección lineal de *lpr* sobre *edad.* Dados los resultados, ¿cuál sería la tasa de rentabilidad anual derivada de conservar el vino?

(b) Efectúe la regresión múltiple de *lpr* sobre *edad*, *lluvinv*, *lluvcos*, *tempmed.* ¿Cómo cambia la estimación de la tasa de rentabilidad de conservar el vino? ¿Cómo explica las diferencias? Explique qué (**Pista**: ¿Qué factores pueden afectar a la calidad del vino?).

4. Sean Y = logaritmo de la cantidad real de dinero, X_1 = logaritmo del PIB real, X_2 = logaritmo del tipo de interés de las Letras del Tesoro. Considere las siguientes regresiones:

$$\begin{array}{rccc} \widehat{Y} = & 2.3296 & +0.5573X_1 & -0.2032X_2 \\ & (0.2054) & (0.0264) & (0.0210) \end{array}$$

$$R^2 = 0.927 \quad s = 0.048 \quad \overline{Y} = 6.63$$

$$\begin{array}{rcc} \widetilde{Y} = & 2.9967 & +0.4356X_1 \\ & (0.3657) & (0.0438) \end{array}$$

$$\widetilde{R}^2 = 0.733 \quad \widetilde{s} = 0.091 \quad \overline{Y} = 6.63$$

$$\widehat{X}_2 = \quad -3.2839 \quad +0.5988X_1$$

Utilizando los resultados de estas regresiones, derive la pendiente de la "otra regresión corta", es decir, de la regresión de Y sobre X_2. Pista: desarrolle la regla de la variable omitida con los papeles de X_1 y X_2 cambiados. La pendiente y el R^2 de la regresión auxiliar pueden utilizarse para determinar la pendiente de la *otra regresión auxiliar.*

5. Dado el modelo

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_1^2 + \varepsilon,$$

donde $E(\varepsilon | X_1) = 0$, se ha obtenido la siguiente ecuación por MCO:

$$\begin{array}{c} \widehat{Y} = 2.613 + 0.30X_1 - 0.090X_1^2 \\ (0.429) \quad (0.14) \quad (0.037) \\ n = 32, \quad R^2 = 0.1484 \end{array}$$

(a) Dados los resultados de la estimación, ¿a partir de qué valor de X_1 el efecto causal de X_1 sobre Y comienza a ser negativo? Justifique la respuesta.

(b) Queremos contrastar si conviene mantener el término cuadrático en el modelo. Establezca la hipótesis nula y la hipótesis alternativa en términos de los parámetros del modelo, y escriba el correspondiente estadístico de contraste y su distribución aproximada. En caso de disponer de información, realice dicho contraste e indique cuál es su conclusión.

6. Consideramos una ecuación para explicar los salarios de los directores generales de las empresas en función de las ventas anuales de las empresas, el rendimiento de pagarés (*roe*, en porcentaje) y el rendimiento de las acciones de la empresa (*ros*, en porcentaje):

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

(a) En términos de los parámetros del modelo, especificar la hipótesis nula de que, una vez tomada en cuenta la influencia de *sales* y *roe*, *ros* no influye en el salario de los directores generales. Especifique como alternativa que la mejora en el rendimiento del mercado de valores incrementa el salario del director general.

(b) Considere los siguientes resultados de la estimación de la ecuación anterior por MCO:

$$\begin{array}{l} \widehat{log(salary)} = \underset{(0.32)}{4.32} + \underset{(0.035)}{0.280} log(sales) + \underset{(0.0041)}{0.0174} roe + \underset{(0.00054)}{0.00024} ros \\ n = 209 \qquad R^2 = 0.283 \end{array}$$

¿En qué porcentaje se predice que aumentaría la variable *salary*, si *ros* aumentase en 50 puntos?

(c) Contraste, al 10%, la hipótesis nula de que *ros* no tiene efecto sobre *salary*, contra la alternativa de que *ros* tiene un efecto positivo.

(d) ¿Incluiría *ros* en el modelo final que explica la remuneración del director general en función del rendimiento empresarial? Justifique la respuesta.

7. Considere la siguiente estimación por MCO:

$$\begin{array}{l} \widehat{sleep} = \underset{(235.22)}{3840.83} - \underset{(0.018)}{0.163} totwrk - \underset{(5.86)}{11.71} educ - \underset{(11.21)}{8.70} age + \underset{(0.134)}{0.128} age2 + \underset{(34.33)}{87.75} male \\ n = 706 \qquad R^2 = 0.117 \end{array}$$

La variable *sleep* son los minutos dedicados a dormir a la semana, *totwrk* son los minutos dedicados a trabajar, *educ* y *age* están medidos en años, y *male* es una variable binaria que toma el valor 1 si el individuo es hombre y 0 en caso contrario.

(a) Manteniendo el resto de los factores constantes, ¿hay evidencia de que los hombres duermen más que las mujeres? ¿Se trata de una evidencia fuerte?

(b) ¿Es estadísticamente significativa la disyuntiva entre trabajar y dormir? ¿Cuál es la estimación de esta disyuntiva?

(c) ¿Qué otra regresión se necesita para contrastar la hipótesis nula de que la edad no afecta al tiempo dedicado a dormir (manteniendo el resto constante)?

8. El archivo RELOJES.GDT incluye datos de una subasta anual de relojes antiguos organizada por la compañía alemana Triberg Clock. Se ha considerado el siguiente modelo:

$$P = \beta_0 + \beta_1 A + \beta_2 C + \beta_3 A^2 + \beta_4 C^2 + \beta_5 (A \times C) + \varepsilon, \tag{R1}$$

donde P es el precio, en cientos de euros, de la subasta ganadora, A es la antigüedad del reloj (en años) y C es el número de postores o pujadores. Además, el término de error verifica, para cualquier edad y para cualquier número de postores, $E(\varepsilon|\, A, C) = 0$ y $V(\varepsilon|\, A, C) = \sigma^2$. Supondremos que las correspondientes características poblacionales de dichas variables coinciden con sus análogos muestrales, y que podemos aplicar resultados asintóticos para hacer inferencia aproximada. Tenga en cuenta que para algunas preguntas puede ser necesario conocer dichas características poblacionales, particularmente el rango de valores que toma cada una de las variables de interés.

   (a) Calcule los estadísticos descriptivos de las variables P, A y C.
   (b) Estime el modelo (R1) por MCO. Proponga y calcule estimaciones consistentes de $V(P)$ y $V(P|\, A, C)$.
   (c) Dados los resultados de (R1), calcule el efecto *ceteris paribus* de la antigüedad sobre el precio del reloj. ¿Es dicho efecto constante? Justifique la respuesta.
   (d) Considere la siguiente afirmación: "El precio no depende de la antigüedad ni del número de postores". Escriba la hipótesis nula (H_0) y la hipótesis alternativa (H_1) en términos de los parámetros del modelo (R1). Escriba el estadístico de contraste e indique su distribución aproximada bajo H_0. Si es posible, realice el contraste e indique cuál es su conclusión. En caso de que falte información para realizar el contraste, indique qué necesitaría.
   (e) Considere la siguiente especificación:

$$P = \delta_0 + \delta_1 (4A + C) + \delta_2 A^2 + \delta_3 C^2 + \delta_4 (A \times C) + \varepsilon. \tag{R2}$$

   Estime (R2) por MCO.
   (f) Considere la hipótesis nula $H_0 : \beta_1 = 4\beta_2$ frente a la alternativa $H_1 : \beta_1 \neq 4\beta_2$. Escriba el modelo restringido (bajo H_0), y relaciónelo con el modelo (R2).

9. Considere el modelo de regresión lineal

$$Y_i = \beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + \beta_3 X_{3i} + \beta_4 X_{4i} + \beta_5 X_{5i} + \varepsilon_i.$$

Explique exactamente cómo contrastaría las siguientes hipótesis:

   (a) $\beta_1 = 0$.
   (b) $\beta_1 = 0$ y $\beta_4 = \beta_5$
   (c) $\beta_1 = 0$, $\beta_3 = 2$, y $\beta_4 = \beta_5$.

10. Para contrastar una única restricción asociada a una relación lineal entre parámetros del modelo, siempre es posible reparametrizar el modelo no restringido, mediante una representación equivalente, pero que permita contrastar la hipótesis nula mediante un contraste de que un determinado coeficiente es igual a cero. Este ejercicio ilustra una posible reparametrización

para una restricción lineal que afecta a varios parámetros del modelo no restringido.
Considere el modelo de regresión múltiple

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + \varepsilon, \tag{1}$$

que verifica los supuestos **1** a **4**. Se quiere contrastar la hipótesis nula $H_0 : \beta_1 - 3\beta_2 = 1$.

(a) Sean $\widehat{\beta}_1$ y $\widehat{\beta}_2$ los estimadores MCO respectivos de β_1 y β_2. Exprese $V\left(\widehat{\beta}_1 - 3\widehat{\beta}_2\right)$ en términos de $V\left(\widehat{\beta}_1\right)$, $V\left(\widehat{\beta}_2\right)$ y $C\left(\widehat{\beta}_1, \widehat{\beta}_2\right)$.

(b) Escriba el estadístico t para contrastar $H_0 : \beta_1 - 3\beta_2 = 1$.

(c) Definiendo $\theta = \beta_1 - 3\beta_2$ y su estimador correspondiente (basado en los estimadores MCO de β_1 y β_2), $\widehat{\theta} = \widehat{\beta}_1 - 3\widehat{\beta}_2$, escriba una especificación equivalente a (1) en la que aparezcan β_0, θ, β_2 y β_3 que permita obtener directamente a partir de una muestra de datos, $\widehat{\theta}$ y su error estándar.

(d) Explique una estrategia, alternativa al apartado (b), de contrastar $H_0 : \beta_1 - 3\beta_2 = 1$.

11. Considere la siguiente especificación para una función de producción:

$$y_i = \beta_0 + \beta_1 l_i + \beta_2 k_i + \varepsilon_i \quad (i = 1, \dots, n), \tag{2}$$

donde y = logaritmo del output, l = logaritmo del factor trabajo, k = logaritmo del capital.
Suponga además que $E\left(\varepsilon_i | \, l_i, k_i\right) = 0$ para todo l_i, k_i.
Se quiere contrastar la hipótesis de rendimientos constantes a escala, es decir, $\beta_1 + \beta_2 = 1$.
Explique cómo se haría el contraste:

(a) Si se tienen las estimaciones MCO de la regresión (2) y de su correspondiente matriz de varianzas y covarianzas de los parámetros estimados.

(b) Si se tienen las estimaciones MCO de la regresión de $(y_i - k_i)$ sobre una constante, $(l_i - k_i)$ y k_i.

(c) Si se tienen la suma de los cuadrados de los residuos de la estimación MCO de (2) y de la estimación MCO de la regresión de $(y_i - k_i)$ sobre una constante y $(l_i - k_i)$.