Hoja de Ejercicios 1
El Modelo de Regresión Múltiple

ECONOMETRÍA I. UC3M

1. Suponga que para establecer la relación lineal entre Y = variación porcentual del salario real y X = tasa de desempleo (en tanto por ciento) se considera la siguiente especificación:

$$Y = 8{,}33 - 0{,}84X + u.$$

   a) Interprete los coeficientes.

   b) Suponga ahora que se considera una función recíproca, de manera que consideramos $X' = 1/X$ (inversa de la tasa de desempleo):

$$Y = -0{,}12 + 0{,}983X' + u'.$$

   Interprete los coeficientes.

2. En la formulación alternativa del modelo de regresión clásico, ¿se podría reemplazar el supuesto $E(u|x) = 0$ por el supuesto $E(u) = 0$? ¿Son ambos supuestos equivalentes? ¿Es posible que $E(u) = 0$ y que $E(u|x) = 0$ para todo x? ¿Es posible que $E(u|x) = 0$ para todo x sin que $E(u) = 0$?

   (Pista: ¿es posible que la renta media en los Estados Unidos sea igual a 20000 dólares sin que la renta media de cada uno de los estados sea también de 20000 dólares? ¿Es posible que la renta media de cada estado sea 20000 dólares y que la renta media en los Estados Unidos no sea también de 20000 dólares?).

3. ¿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.

4. Suponga que $\log(wage)$ mide el logaritmo del salario mensual y *educ* el número de años de educación. Considere el modelo lineal

$$\log(wage) = b_0 + b_1 educ + v.$$

   a) ¿Es razonable suponer que $E(v|educ) = 0$? ¿Qué otros factores estarían incluidos en u?

   b) Considere el modelo

$$\log(wage) = \beta_0 + \beta_1 educ + \beta_2 abil + u,$$

donde *abil* el coeficiente de inteligencia (IQ) y u satisface el supuesto $E(u|educ, abil) = 0$. Interprete el coeficiente β_1. Dado que la muestra de la que inicialmente se dispone no contiene información sobre el coeficiente de inteligencia de los trabajadores, se estima el modelo de regresión simple

$$\begin{aligned} \log(\widehat{wage}) &= 5{,}97 + \underset{(0{,}006)}{0{,}06}\, educ \\ n &= 935, R^2 = 0{,}097, SCE = 149{,}52. \end{aligned}$$

Explique bajo qué condiciones la estimación MCO del parámetro de la variable *educ* es un estimador insesgado de β_1. Proporcione un intervalo de confianza al 95 % para la pendiente del modelo estimado.

c) Suponga ahora que se consigue información sobre el coeficiente de inteligencia de los trabajadores de la muestra, obteniéndose la siguiente estimación,

$$\begin{aligned} \log(\widehat{wage}) &= 5{,}66 + \underset{(0{,}007)}{0{,}04}\, educ + \underset{se(\hat{\beta}_2)}{0{,}0059 abil} \\ n &= 935, R^2 = 0{,}130 \end{aligned}$$

Contraste al nivel de significación del 5 % que la inteligencia no afecta el salario.

d) Obtenga la covarianza muestral entre el nivel educativo y el coeficiente de inteligencia. A la vista de los resultados, interprete el parámetro asociado a la educación en las dos regresiones.

e) Si $E(u^2|educ, abil) = \sigma^2(educ)$, donde $\sigma^2(educ)$ es una función positiva de *educ*, ¿cambiaría alguna de sus respuestas? ¿Cuáles serían las consecuencias del incumplimiento del supuesto de homocedasticidad?

5. La variable *rdintens* son los gastos en investigación y desarrollo (R&D) en porcentaje sobre las ventas. Las ventas, *sales*, se miden en millones de dólares. La variable *profmarg* son los beneficios en porcentaje sobre las ventas. Usando 32 empresas del sector químico, se ha obtenido la siguiente ecuación

$$\begin{aligned} \widehat{rdintens} &= 0{,}472 + \underset{(0{,}216)}{0{,}321} \log(sales) + \underset{(0{,}046)}{\hat{\beta}_2}\ profmarg \\ n &= 32, \quad R^2 = 0{,}098 \end{aligned}$$

a) Interprete el coeficiente de $\log(sales)$: si las ventas se incrementan en un 10 %, ¿cuál es el cambio promedio estimado, *ceteris paribus*, en *rdintens*? ¿Es un efecto económicamente importante? ¿Es estadísticamente significativo?

b) Se ha estimado el siguiente modelo alternativo utilizando la misma base de datos:

$$\begin{aligned} \widehat{rdintens} &= 1{,}104 + \underset{(0{,}216)}{,302} \log(sales) \\ n &= 32, \quad R^2 = 0{,}061 \end{aligned}$$

¿Es el coeficiente de *profmarg* en el modelo del apartado (*a*) significativo? Si sabemos que $\hat{\beta}_2$ es positivo, ¿cuál es su valor?

c) Considere este modelo que relaciona los beneficios con las ventas,

$$\begin{aligned} \widehat{profm}\arg &= 7{,}34 + \underset{se(\hat{\gamma}_1)}{\hat{\gamma}_1} \log(sales). \\ n &= 32, \quad R^2 = 0{,}0069 \end{aligned}$$

¿Cuál es el signo y el valor de $\hat{\gamma}_1$? ¿Es significativo? Obtenga $se(\hat{\gamma}_1)$.

6. Para estudiar el efecto de la asistencia a clase sobre las notas finales de una asignatura, se han utilizando datos de 680 alumnos de Introducción a Microeconomía en una universidad americana para ajustar el siguiente modelo de regresión:

$$\begin{aligned} stndfnl &= \beta_0 + \beta_1 atndrte + \beta_2 priGPA \\ &\quad + \beta_3 priGPA^2 + \beta_4 ACT^2 + \beta_5 priGPA \cdot atndrte + U, \end{aligned} \tag{1}$$

los resultados de la estimación son los siguientes

$$\widehat{stndfnl} = \underset{(1{,}36)}{2,05} - \underset{(0{,}0102)}{0,0067} atndrte - \underset{(0{,}48)}{1,63} priGPA$$

$$+\underset{(0{,}101)}{0,296} priGPA^2 + \underset{(0{,}0022)}{0,0045} ACT^2 + \underset{(0{,}0022)}{0,0045} priGPA \cdot atndrte, \ R^2 = 0,222,$$

donde *stndfnl* es el resultado de un examen final estandarizado, *atndrte* es el porcentaje de asistencia a clase, *priGPA* es la nota media obtenida en los cursos anteriores y *ACT* es la nota de acceso a la Universidad.

a) ¿Cuál es el efecto parcial estimado de la asistencia a clase sobre el examen final? ¿Se puede concluir que el efecto es significativo para un alumno con nota media igual a 3 en cursos anteriores si $\widehat{Cov}\left(\hat{\beta}_1, \hat{\beta}_5\right) = 0{,}0001$?

b) Si añadimos el término $\beta_6\ ACT * atndrte^2$ a la ecuación (1), ¿Cuál será el efecto parcial de asistir a clase en el modelo en términos de los parámetros desconocidos?

c) Explique cómo contrastaría que el modelo es lineal en todas las variables. Esto es, que $\beta_3 = \beta_4 = \beta_5 = 0$. Exprese con claridad todos los pasos que ha de seguir para realizar el contraste: 1) Estadístico a utilizar, explicando sus componentes con claridad y 2) Regla de decisión.

7. El análisis de regresión puede utilizarse para contrastar si el mercado hace un uso eficiente de la información a la hora de valorar las acciones. En concreto, sea *return* el rendimiento total de las acciones de una empresa a lo largo de un periodo de cuatro años, desde finales de 1990 hasta finales de 1994. La hipótesis de eficiencia del mercado dice que este rendimiento no debería estar relacionado de manera sistemática con la información conocida en 1990. Si las características de la empresa conocidas al principio del periodo fuesen de ayuda para predecir el rendimiento del mercado, entonces podríamos usar esta información para seleccionar unas acciones u otras.

Para 1990, sea *dkr* el cociente del endeudamiento de la empresa en relación a su capital, sean *eps* las ganancias por acción, *netinc* la renta neta y *salary* la remuneración total del director general.

a) Usando los datos de RETURN, estime la siguiente ecuación:

$$return = \beta_0 + \beta_1 dkr + \beta_2 eps + \beta_3 netinc + \beta_4 salary + u.$$

Contraste si las variables explicativas son conjuntamente significativas al 5 %. Contraste si *netinc* y *salary* son conjuntamente significativas. ¿Hay alguna variable explicativa que sea individualmente significativa?

b) Reestime ahora el modelo que aplica logaritmos a *netinc* y *salary*,

$$return = \beta_0 + \beta_1 dkr + \beta_2 eps + \beta_3 \log(netinc) + \beta_4 \log(salary) + u.$$

¿Cómo cambian las conclusiones del apartado *(a)*?

c) ¿Por qué no se aplican logaritmos a *dkr* y *eps* en el apartado *(b)*?

d) En términos generales, ¿la evidencia a favor de la predictibilidad del rendimiento de las acciones es fuerte o débil?

8. Dadas las variables $Alim$ = Gasto familiar anual (en euros) en alimentación, GT = Gasto familiar total (en euros), considere los siguientes modelos alternativos:

$$\begin{array}{l} \text{(i) } \ln(Alim) = 3{,}67 + 0{,}48\ln(GT) \\ \text{(ii) } Alim = -164567 + 2163\ln(GT) \\ \text{(iii) } (Alim/GT) \times 100 = 156{,}89 - 13{,}32\ln(GT) \end{array}$$

(El último modelo se conoce como especificación "Working-Lesser", que considera los determinantes del gasto en alimentación como porcentaje del gasto total)
Interprete los coeficientes que se obtienen en cada uno de los modelos.

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 $\beta_4 = \beta_5$.