FACULTAD DE CIENCIAS
ECONÓMICAS Y EMPRESARIALES

Ingeniería Informática + ADE

Universidad de Granada (UGR)

Autor: Ismael Sallami Moreno

Asignatura: Econometría

Índice

1 Ejercicios Propuestos

1.1. Tema 2

ECONOMETRÍA. GADE

Prácticas

Tema 2

Ejercicios propuestos

1. Detalle el orden de las matrices y vectores de un modelo econométrico con 3 variables explicativas más el termino constante y 50 observaciones.

2. Ponga un ejemplo de una matriz X con un término constante y tres variables explicativas de manera que el modelo no cumpla la condición del rango completo por columnas.

3. Razone que componentes del modelo econométrico tienen carácter estocástico (aleatorio) y cuáles tienen carácter determinista (fijo).

4. Ponga un ejemplo de una matriz de varianzas-covarianzas de las perturbaciones de un modelo con heterocedasticidad.

5. Ponga un ejemplo de una matriz de varianzas-covarianzas de las perturbaciones de un modelo con autocorrelación.

6. Se considera la posibilidad de introducir nuevas variables explicativas en el modelo de la curva de Phillips (EJERCICIO 1 de la relación de ejercicios resueltos) y para ello se recoge información sobre la renta nacional disponible neta a precios de mercado por habitante (X_2) y la renta nacional disponible neta (X_3).

Años	X_2	X_3
2006	18614	825737
2007	19403	877724
2008	19492	896295
2009	18719	867972
2010	18706	871015
2011	18229	851948
2012	17822	833445
2013	17691	824281
2014	18029	837556
2015	18828	873766

Se pide:

a) Estimar el modelo incluyendo la variable X_2. (Sol. $\hat{y} = 165,78 - 1,48X_1 - 0,0077X_2$)

b) Estimar el modelo incluyendo las variables X_2 y X_3. (Sol. $\hat{y} = 120,97 - 0,77X_1 - 0,017X_2 + 0,00025X_3$)

c) Comparar ambos modelos. (Sol. $\bar{R}_1^2 = 0,7097 < \bar{R}_2^2 = 0,96$; $AIC_1 = 2,83 > AIC_2 = 0,8920$)

Ejercicio seleccionado de [1].

7. En la siguiente tabla se recogen las ventas de seis empresas informáticas en función del número de comerciales:

v_t	109	111	132	140	169	180
c_t	12	15	17	18	19	20

Se pide:

a) Plantear el modelo econométrico y estimar los coeficientes por mínimos cuadrados ordinarios. Interpretación de los coeficientes estimados. (Sol. $\hat{v}_t = -13,56 + 9,132c_t$)

b) Calcular el coeficiente de determinación e interpretarlo. ($R^2 = 0,8294$)

c) Se estima un modelo alternativo añadiendo como variable explicativa el gasto en publicidad de cada empresa, obteniendo un coeficiente de determinación igual a 0,93363. Concluya de forma razonada si este modelo es mejor que el anterior. (Sol. $\bar{R}_1^2 = 0,78675 < \bar{R}_2^2 = 0,8893$)

Ejercicio seleccionado de [1].

8. Se tiene la siguiente información correspondiente al curso 2010/2011 sobre el número de becarios en la enseñanza universitaria (Y), el alumnado matriculado en estudios de 1er. y 2º ciclo y de grado (X_1) y el importe en miles de euros de las becas (X_2). Se pide:

CCAA	Y	X_1	X_2
Andalucía	97.105	234.851	266.222,60
Aragón	9.294	31.063	19.648,80
Asturias	6.882	23.746	16.048,20
Baleares	4.251	13.581	8.224,60
Canarias	19.125	45.146	45.084,30
Cantabria	4.087	10.516	8.013,00
Castilla y León	27.307	78.692	75.124,10
Castilla - La Mancha	14.988	30.431	39.696,80
Cataluña	51.965	179.639	107.077,10
Valencia	49.805	148.671	115.219,30
Extremadura	12.088	22.747	35.506,00
Galicia	24.500	64.262	66.019,40
Madrid	64.563	239.389	134.341,50
Murcia	15.549	42.573	37.177,90
Navarra	4.669	14.705	10.372,50
País Vasco	15.322	53.419	26.789,40
Rioja, La	1.761	8.119	3.345,10
A distancia	20.557	212.021	16.491,60

a) Estimar los parámetros. (Sol. $\hat{y} = 528,077 + 0,09X_1 + 0,2936X_2$)

b) Obtener el coeficiente de determinación. (Sol. $R^2 = 0,994$)

c) Obtener el coeficiente de determinación corregido. (Sol. $\bar{R}^2 = 0,993$)

d) Calcular el criterio de Akaike. (Sol. $AIC = 328{,}9392$)

Ejercicio seleccionado de [1].

9. En la siguiente tabla, podemos observar la cantidad demandada de un activo financiero F_t por un agente a lo largo de 10 años en función de su rendimiento I_t:

F_t	130	120	100	100	110	120	110	110	130	130
I_t	8	7	5	6	7	8	7	6	9	10

Se pide:

a) Plantear el modelo econométrico y estimar los coeficientes por mínimos cuadrados ordinarios. Sol.

$$\hat{\beta} = \begin{pmatrix} 64{,}4278 \\ 7{,}0646 \end{pmatrix}$$

b) Calcular el coeficiente de determinación. Sol. $R^2 = 0{,}8090$.

10. Para estimar el modelo $Y_t = \beta_1 + \beta_2 X_{2t} + \beta_3 X_{3t} + u_t$, se ha obtenido una muestra de la cual ha resultado:

$$X'X = \begin{pmatrix} 30 & 0 & 0 \\ 0 & 22 & 0 \\ 0 & 0 & 20 \end{pmatrix}, \quad X'Y = \begin{pmatrix} 17 \\ 22 \\ 21 \end{pmatrix}, \quad Y'Y = 96.$$

Estimar los coeficientes del modelo por MCO.
Sol.

$$\hat{\beta} = \begin{pmatrix} 0{,}5666 \\ 1 \\ 1{,}05 \end{pmatrix}$$

11. Considerando la siguiente información muestral que contiene el gasto en publicidad de 5 empresas (variable dependiente) y el número de empleados (variable independiente):

y_t	0	2	4	3	2
x_t	5	8	5	4	3

Estimar el modelo econométrico que explica y_t en función de x_t. Sol.

$$\hat{\beta} = \begin{pmatrix} 2{,}5571 \\ -0{,}0714 \end{pmatrix}$$

12. Se quieren estudiar las ventas de una empresa (y_t) en función de los gastos en publicidad (x_{2t}) y del número de empleados de dicha empresa (x_{3t}) con la información disponible para los últimos 10 años. La información muestral se resume a continuación:

$$\sum_{t=1}^{n} y_t = 436, \sum_{t=1}^{n} x_{2t} = 134, \sum_{t=1}^{n} x_{3t} = 101$$

$$\sum_{t=1}^{n} y_t^2 = 20386, \sum_{t=1}^{n} x_{2t}^2 = 2068, \sum_{t=1}^{n} x_{3t}^2 = 1177$$

$$\sum_{t=1}^{n} x_{2t}x_{3t} = 1534, \sum_{t=1}^{n} x_{2t}y_t = 6405, \sum_{t=1}^{n} x_{3t}y_t = 4786$$

a) Plantear las matrices $(X'X)$ y $X'Y$, necesarias para la estimación por MCO.

b) Sabiendo que:

$$(X'X)^{-1} = \begin{pmatrix} 0{,}798956 & -0{,}027501 & -0{,}032717 \\ -0{,}027501 & 0{,}015499 & -0{,}017840 \\ -0{,}032717 & -0{,}017840 & 0{,}026908 \end{pmatrix},$$

calcular los coeficientes del modelo por MCO.

$$\hat{\beta} = \begin{pmatrix} 15{,}61688 \\ 1{,}897671 \\ 0{,}252904 \end{pmatrix}.$$

13. En la siguiente tabla, podemos observar la cantidad demandada de un activo financiero F_t por un agente a lo largo de 6 meses en función de su rendimiento I_t:

F_t	16	18	16	12	15	16
I_t	4	6	3	3	5	7

Se pide:

a) Plantear el modelo econométrico y estimar los coeficientes por mínimos cuadrados ordinarios. (Sol. $\hat{F}_t = 12,35 + 0,675 I_t$)

b) Calcular el coeficiente de determinación e interpretarlo. (Sol. $R^2 = 0,3115$)

14. Para estimar el modelo $Y_t = \beta_1 + \beta_2 X_{2t} + \beta_3 X_{3t} + u_t$, se ha obtenido una muestra de la cual ha resultado:

$$X'X = \begin{pmatrix} 5 & 0 & 0 \\ 0 & 3 & 0 \\ 0 & 0 & 10 \end{pmatrix}, \quad X'Y = \begin{pmatrix} 11 \\ 12 \\ 14 \end{pmatrix}, \quad Y'Y = 100.$$

Se pide:

a) Estimar los coeficientes del modelo por MCO. (Sol. $\hat{Y}_t = 2,2 + 4x_{2t} + 1,4x_{3t}$)

b) Calcular el coeficiente de determinación. (Sol. $R^2 = 0{,}8918$)

c) Calcular los criterios de información. (Sol. $AIC = 22,66$; $BIC = 21,49$)

15. Considerando la siguiente información muestral:

y_t	0	2	1	1	0	4	3
x_t	5	12	10	9	6	15	14

a) Plantear y estimar el modelo econométrico que explica el número de errores en la contabilidad de una empresa (y_t) en función del número de apuntes contables (x_t) a lo largo de 7 meses. (Sol. $\hat{y}_t = -2,3322 + 0,384x_t$)

b) Obtener el coeficiente de determinación. (Sol. $R^2 = 0,9381$)

c) Obtener los residuos y comprobar que su suma es igual a 0. (Sol. $e_1 = 0,408; e_2 = -0,286; e_3 = -0,516; e_4 = -0,132; e_5 = 0,023; e_6 = 0,559; e_7 = -0,056$)

16. En un estudio de los determinantes de la inversión se usaron 20 datos anuales correspondientes a las siguientes variables: inversión anual en millones de euros (Y), tipo de interés en porcentaje (X_1), y variación anual del PIB en millones de euros (X_2). Se dispone de la siguiente información.

$$\sum x_{1t} = 100; \sum x_{1t}^2 = 680; \sum x_{1t}x_{2t} = 100$$

$$\sum x_{2t} = 24; \sum x_{2t}^2 = 48,8; \sum (y_t - \bar{y})^2 = 1200$$

$$\sum y_t = 5; \sum x_{1t}y_t = -255; \sum x_{2t}y_t = 146$$

Se pide:

a) Las estimacion de los parametros del modelo $y_t = \beta_1 + \beta_2 x_{1t} + \beta_3 x_{2t} + u_t$ mediante el método de mínimos cuadrados ordinarios. Sol. $\hat{\beta}' = \begin{pmatrix} -2,725 & -0,875 & 6,125 \end{pmatrix}$

b) Estudiar la bondad del ajuste realizado. Sol. ($R^2 = 0,91875$).

Ejercicio seleccionado de [5].

17. Se quiere estimar por MCO un modelo lineal entre las variables y_t y x_t utilizando 5 observaciones. En la siguiente tabla, se muestra la información de y_t y x_t:

t	1	2	3	4	5
y_t	27	40	54	18	22
x_t	23	24	17	5	10

a) Estimar el modelo por MCO. Sol.

$$\hat{\beta} = \begin{pmatrix} 16{,}7267 \\ 0{,}9793 \end{pmatrix}$$

b) Obtener la varianza estimada de los coeficientes estimados. Sol.

$$Var(\hat{\beta}) = \begin{pmatrix} 227{,}7692 & -11{,}8458 \\ -11{,}8458 & 0{,}7497 \end{pmatrix}$$

c) Obtener el coeficiente de determinación. Sol. $R^2 = 0{,}2989$.

18. Los trabajadores de cierta empresa consideran que existe discriminación salarial y por ello solicitan que se analice la retribución (Y) en función del genero (D_1 toma el valor 1 si el trabajador es hombre y cero si es mujer) y de la edad (X_2) de los 41 trabajadores obteniendo los siguientes datos:

$$\sum_{i=1}^{41} X_{2i} = 1735; \sum_{t=1}^{41} X_{2i}^2 = 77757;$$

$$\sum_{i=1}^{41} D_{1i} = 26; \sum_{i=1}^{41} y_i = 410,46276; \sum_{i=1}^{41} y_i^2 = 5849,20$$

$$\sum_{i=1}^{41} x_{2i}y_i = 18447,75; \sum_{i=1}^{41} D_{1i}y_i = 278,13; \sum_{i=1}^{41} X_{2i}D_{1i} = 1195$$

Se pide:

a) Estimar el modelo. (Sol. $\hat{y}_i = -0,7318 - 0,7686D_{1i} + 0,2653X_{2i}$)

b) Obtener el coeficiente de determinación. (Sol. $R^2 = 0,1565$)

Ejercicio seleccionado de [1].

19. La empresa del ejercicio anterior alega que el salario está relacionado con la tarea desempeñada por cada trabajador y presenta la siguiente estimación donde D_2 toma el valor 1 si el trabajador es operario de producción (0 en caso contrario), D_3 toma el valor 1 si el trabajador se ocupa del mantenimiento de la maquinaria de producción (0 en caso contrario), D_4 toma el valor 1 si el trabajador es encargado de línea de producción (0 en caso contario) y se toma como variable de referencia la variable D_5 que toma el valor 1 si los trabajadores realizan tareas de administración (0 en caso contrario). Se pide:

$$\hat{y} = 27,59 + 0,5353D_1 - 0,0014X_2 - 20,61D_2 - 19,85D_3 - 18,06D_4$$

$$R^2 = 0,846$$

a) Explique por qué no se introducen todas las variables en el modelo e interprete los estimadores de los coeficientes.

b) ¿Cuál de los dos modelos es mejor según la información disponible? (Sol. $\bar{R}_2^2 = 0,824 > \bar{R}_1^2 = 0,1121$)

Ejercicio seleccionado de [1].

Referencias

[1] García, C.B., Sánchez, J.M. y Salmerón, R. (2017) Econometría básica para la economía y la empresa. Ed. Fleming.

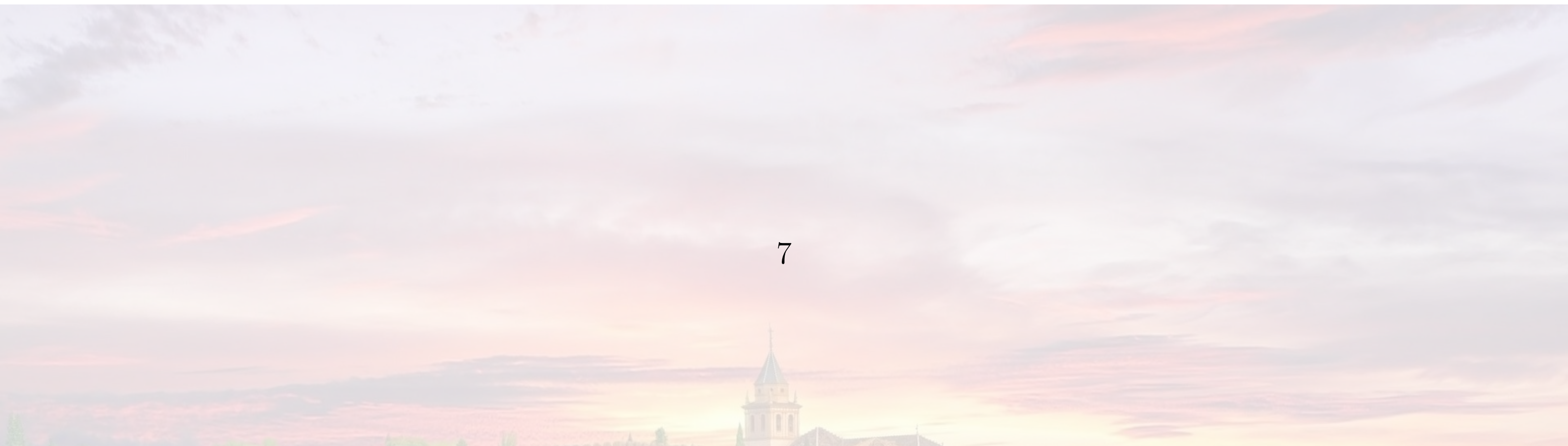
[2] García, J., Jiménez, J.F. y Cerrillo, J.R. Econometría práctica. Edo. Libreria Universitaria de Almería.

[3] Johnston, J. (1984) Métodos de econometría. Ed. Vicens Vives.

[4] Pena, B., Estavillo, J., Galindo, E., Leceta, M. y Zamora, M. (1999). Cien ejercicios de econometría. Ed. Pirámide.

[5] Sánchez, C., López, M.M. y García, T. (2015) Econometría. Ed. Fleming.

[6] Matilla García, m., Pérez Pascual, Pedro y Sanz Carnero, B. (2013) Econometría y predicción. Mc Graw Hill.



1.2. Tema 3

ECONOMETRÍA. GADE

Prácticas

Tema 3

Ejercicios propuestos

1. Dado un modelo econométrico que explique las ventas de licencias de un software contable en decenas de miles de euros (y_t) para una determinada empresa a lo largo de los últimos 5 años, en función del gasto en miles de euros en publicidad (x_1) y el precio del software en cientos de euros (x_2) se pide:

$$X'X = \begin{pmatrix} 5 & 15 & 25 \\ & 55 & 81 \\ & & 129 \end{pmatrix}; \bar{y} = 4; \sum_{t=1}^{5} y_t x_{1t} = 76; \sum_{t=1}^{5} y_t x_{2t} = 109; y'y = 108$$

   a) Contrastar la significación individual de las variables. (Sol. $|t_{exp}| = 2,8867; |t_{exp}| = 1,095$)

   b) Contrastar la significación global del modelo. Tabla ANOVA. (Sol. $F_{exp} = 17,66$)

   c) Contrastar al 99 % de confianza si los parámetros asociados a las variables x_{1t} y x_{2t} pueden tener el mismo efecto sobre la variable pero de signo contario. Datos complementarios: $R(X'X)^{-1}R' = 0,5$. (Sol. $F_{exp} = 2,666$)

   Ejercicio seleccionado de [1].

2. Se tienen datos para los años 2006-2015 sobre la evolución de las ventas (Y) del grupo Inditex, el número de tiendas (X_2), número de paises (X_3), número de marcas (X_4) y una variable ficticia que toma el valor 1 para los años en los que se oferta comercio electrónico (X_5). Se pide:

t	Y	X_2	X_3	X_4	X_5
2006	8196	3131	64	7	0
2007	9435	3691	66	7	0
2008	10407	4264	71	8	0
2009	11084	4607	73	8	0
2010	12527	5044	77	8	1
2011	13793	5527	82	8	1
2012	15946	6009	86	8	1
2013	16724	6340	87	8	1
2014	18117	6683	88	8	1
2015	20900	7013	88	8	1

   Se pide:

   a) Estimar el modelo. (Sol. $\hat{\beta}_1 = 18299,38; \hat{\beta}_2 = 5,14; \hat{\beta}_3 = -227,90; \hat{\beta}_4 = -1745,64; \hat{\beta}_5 = -101,65$)

b) Contrastar la signficación individual de las variables. Sol.

$$t_{exp}(\beta_1) = 2,28; t_{exp}(\beta_2) = 4,96; t_{exp}(\beta_3) = -1,28; t_{exp}(\beta_4) = -2,35; t_{exp}(\beta_5) = -0,10$$

c) Contrastar la significación global del modelo. (Sol. $F_{exp} = 104,69$)

d) Obtener el coeficiente de determinación. (Sol. $R^2 = 0,9882$)

e) Si se estima otro modelo alternativo en el que no se tiene en cuenta el número de marcas y se obtiene un coeficiente de determinación de 0,9751, ¿Se puede decir que este modelo es peor que el anterior? (Sol. $\bar{R}_1^2 = 0,9787; \bar{R}_2^2 = 0,9626$)

Ejercicio seleccionado de [1].

3. En la siguiente tabla se recogen las ventas de cinco empresas informáticas en función del número de comerciales:

v_t	109	111	132	140	169	180
c_t	12	15	17	18	19	20

Se pide:

a) ¿El número de comerciales influye significativamente en el volumen de ventas de la empresa? ¿Y la constante? (95 % de confianza). (Sol. $t_{exp} = -0,384; t_{exp} = 4,41$).

Ejercicio seleccionado de [1].

4. Se tienen datos sobre la retribución anual en miles de euros (Y) de 41 trabajadores, su genero (D_1 que toma el valor 1 si es un hombre y cero en caso contrario) y su ocupación en la empresa de manera que se introduce las siguientes variables D_2 que toma el valor 1 si el trabajador es operario de producción (0 en caso contrario), D_3 que toma el valor 1 si el trabajador se ocupa del mantenimiento de la maquinaria de producción (0 en caso contrario) y D_4 que toma el valor 1 si el trabajador es encargado de línea de producción (0 en caso contario). Se toma como variable de referencia la variable D_5 que toma el valor 1 si los trabajadores realizan tareas de administración (0 en caso contrario). A partir de dichos datos se ha realizado la siguiente estimación:

$$\begin{array}{lcccccc} \hat{y}_i = & 27,59+ & 0,5353D_{1i} & -0,0014X_{1i} & -20,61D_{2i} & -19,85D_{3i} & -18,06D_{4i} \\ & (2,94) & (1,07) & (0,052) & (1,6821) & (1,8973) & (1,7181) \end{array}$$

$$R^2 = 0,846$$

Se pide:

a) Contrastar la significación individual de cada una de las variables.

$$t_{exp}(\beta_1) = 9,384; t_{exp}(\beta_2) = 0,500; t_{exp}(\beta_3) = -0,027$$

$$t_{exp}(\beta_4) = -12,253; t_{exp}(\beta) = -10,462; t_{exp}(\beta) = -10,512$$

b) Sabiendo que la $SCE = 267,59$, contrastar las significación global del modelo. Realizar tabla ANOVA. (Sol. $F_{exp} = 38,45$)

c) Contrastar si es posible que los encargados de línea de producción ganen 25000 euros menos que los de administración. (Sol. Se rechaza $H_0 : \beta_6 = -25$)

d) Realizar un intervalo de confianza para el estimador de la variable D_1 y a la vista de los resultados concluir si es posible que un trabajador gane 1000 euros más que una trabajadora en esta empresa. (Sol. $-1,64; 2,71$)

e) ¿Se puede afirmar que los trabajadores de administración son los que más cobran? (Sol. Si)

Ejercicio seleccionado de [1].

5. Se pretende estimar el modelo que explique el número de vehículos vendidos en el año por la marca A (y) en función de la renta media per capita de la población expresada en miles de euros (x_1) y el precio medio de la marca A en miles de euros (x_2) a partir de la siguiente información:

$$X'X = \begin{pmatrix} 20 & 165,3 & 282,4 \\ & 1390,75 & 2314,72 \\ & & 4020,38 \end{pmatrix}; X'\bar{y} = \begin{pmatrix} 1038 \\ 8749,1 \\ 14501,3 \end{pmatrix};$$

$$(X'X)^{-1} = \begin{pmatrix} 26,8761 & -1,2541 & -1,1658 \\ & 0{,}0757 & 0{,}0445 \\ & & 0,0565 \end{pmatrix}; y'y = 55132$$

Se pide:

a) Estimar el modelo e interpretarlo comentando si los resultados esperados coinciden con los que cabría esperar. Sol. $\hat{\beta}' = \begin{pmatrix} 19,65 & 5,97 & -1,2127 \end{pmatrix}$

b) Estimar la matriz de varianzas covarianzas de los estimadores. Interpretar.
Sol. $\hat{var}(\hat{\beta}) = \begin{pmatrix} 88,47 & -4,13 & -3,84 \\ -4,13 & 0,25 & 0,15 \\ -3,84 & 0,15 & 0,19 \end{pmatrix}$

c) Comprobar la significación individual y conjunta de los parámetros. Sol.

$$t_{exp}(\beta_1) = 2,09; t_{exp}(\beta_2) = 11,96; t_{exp}(\beta_3) = -2,81; F_{exp} = 182,86$$

d) Contraste la hipótesis de que el parámetro que acompaña a la variable renta media (x_1) es igual a seis veces el parámetro que acompaña a la variable precio medio de la marca (x_2). Sol. $R(X'X)^{-1}R' = 1,57; F_{exp} = 33,82$

e) Se ha vuelto a estimar un modelo en el que se ha incorporado el precio medio de las demás marcas de vehículos (x_3) obteniéndose el siguiente modelo $\hat{y}_t = 21,8616 + 6,0112x_{1t} - 0,7681x_{2t} - 0{,}5604x_{3t}$ con un coeficiente de determinación $R^2 = 0{,}96745725$. Desde el punto de vista de la bondad del ajuste ¿qué modelo sería mejor? Sol. $\bar{R}_1^2 = 0,950355 < \bar{R}_2^2 = 0,961355$.

6. Se quiere estimar por MCO un modelo lineal entre las variables y_t y x_t utilizando 5 observaciones. En la siguiente tabla, se muestra la información de y_t y x_t:

t	1	2	3	4	5
y_t	7	4	5	-4	3
x_t	5	4	4.5	0	3

   a) Estimar el modelo por MCO. (Sol. $\hat{y}_t = -3,892 + 2,088x_t$)
   b) Obtener la varianza estimada de los coeficientes estimados. (Sol. $v\hat{a}r(\hat{\beta}_1) = 0,3189$ y $v\hat{a}r(\hat{\beta}_2) = 0,0226$)
   c) Obtener el coeficiente de determinación. (Sol. $R^2 = 0{,}984$)

7. Se han estudiado las calificaciones de ECO 1 (Y) en relación con las notas obtenidas en TC I (X_1) y TC II (X_2) de 25 alumnos de la facultad. Se obtuvieron los siguientes resultados sobre el modelo $Y_i = \beta_1 + \beta_2 X_{1i} + \beta_3 X_{2i} + u_i$.

$$v\hat{a}r(\hat{\beta}) = \begin{pmatrix} 210,1392 & -9,9972 & -9,9972 \\ -9,9972 & 20,0060 & -20,0060 \\ -9,9972 & -20,0060 & 20,0060 \end{pmatrix}$$

$$X'Y = \begin{pmatrix} 312,5 \\ 1555,25 \\ 1650,75 \end{pmatrix}; e'e = 2551,5$$

   a) Obtener los estimadores mínimo cuadráticos del modelo. Sol.

$$\hat{\beta}' = \begin{pmatrix} 289,8626 & -43,4110 & -10,4637 \end{pmatrix}$$

   b) ¿Es el modelo significativo? Sol. $F_{exp} = 8,1386$
   c) Se intuye que $4\beta_3 = \beta_2$. ¿Podemos dar por cierta tal intuición? Sol. $F_{exp} = 0,0048$

8. En un modelo de regresión con término constante se cuenta con la siguiente información muestral:

$$y' = \begin{pmatrix} 3 & 1 & 8 & 3 & 5 \end{pmatrix}$$
$$x_2' = \begin{pmatrix} 3 & 1 & 5 & 2 & 4 \end{pmatrix}$$
$$x_3' = \begin{pmatrix} 5 & 4 & 6 & 4 & 6 \end{pmatrix}$$

   a) Estimar los parámetros del modelo y la matriz de varianzas covarianzas. Sol.

$$\hat{\beta}' = \begin{pmatrix} 4 & 2,5 & -1,5 \end{pmatrix}$$

$$v\hat{a}r(\hat{\beta}) = \begin{pmatrix} 20,025 & 3,375 & -6 \\ 3,375 & 0,75 & -1,125 \\ -6 & -1,125 & 1,875 \end{pmatrix}$$

b) Contraste la hipótesis nula $H_0 : \beta_2 = \beta_3 = 0$. Sol $F_{exp} = 17,66$

c) Contraste la hipótesis nula $H_0 : \beta_2 + \beta_3 = 0$. Sol $F_{exp} = 2,66$

d) Obtener un intervalo de confianza del 95 % para β_2. Sol $(-1,2265; 6,2265)$

e) Efectuar ANOVA. Sol. $F_{exp} = 17,66$.

Ejercicio seleccionado de [5].

9. Dado el modelo $Y_t = \beta_1 + \beta_2 X_{1t} + \beta_3 X_{2t} + u_t$, se obtuvieron los siguientes resultados correspondientes a 12 observaciones:

$$(X'X)^{-1} = \begin{pmatrix} 0,6477 & -0,041 & -0,0639 \\ -0,041 & 0,0071 & -0,0011 \\ -0,0639 & -0,0011 & 0,0152 \end{pmatrix}$$

$$X'Y = \begin{pmatrix} 91 \\ 699 \\ 448 \end{pmatrix}$$

$$SCT = 104,9167; Y'Y = 794,99$$

a) Obtener el estimador MCO de β.

$$Sol.\hat{\beta} = \begin{pmatrix} 1,6545 & 0,7391 & 0,2258 \end{pmatrix}$$

b) Calcular e interpretar la bondad del ajuste realizado. Sol. $R^2 = 0,7460347$

c) Contrastar la significación global del modelo. Sol. $F_{exp} = 13,2199$

d) ¿Podemos afirmar que $\beta_2 + \beta_3 = 1$? Sol. $F_{exp} = 0,0207$

10. Dado el siguiente modelo $Y_t = \beta_1 + \beta_2 x_{2t} + \beta_3 x_{3t} + u_t$, donde y_t es el vólumen de producción facturado en el periodo t en millones de unidades, x_{2t} es el precio del bien en el periodo t en unidades monetarias y x_{3t} es la renta media en el periodo t en miles de unidades monetarias, se dispone de los siguientes datos para el periodo 2000-2004.

$$\bar{y} = 1,8; \bar{x}_2 = 3; \bar{x}_3 = 2,8$$

$$\sum_{t=1}^{5} x_{2t}^2 = 51; \sum_{t=1}^{5} x_{3t}^2 = 54; \sum_{t=1}^{5} x_{2t} t x_{3t} = 49$$

$$\sum_{t=1}^{5} y_t x_{2t} = 27; \sum_{t=1}^{5} y_t x_{3t} = 27; y'y = 19$$

a) Estimar un intervalo de confianza al 95 % para los parámetros. Sol.

$$\beta_1 \in (-4,17; 8,15)$$

$$\beta_2 \in (-3,13; 2,504)$$

$$\beta_3 \in (-1,524; 2,067)$$

b) A la vista de los resultados justifique que variables son significativas. Sol. No son significativas porque todos los intervalos contienen al cero.

c) Contrastar la significación global del modelo. Tabla ANOVA. Sol. ($F_{exp} = 0,2173$)

d) Contrastar al 99 % de confianza si los parámetros asociados a las variables x_{2t} y x_{3t} pueden tener el mismo efecto sobre la variable Y pero de signo contrario. Sol. ($F_{exp} = 0,0103$)

11. El principal grupo distribuidor de diamantes en el mundo desea conocer la demanda de este producto para programar las extracciones que debe realizar en sus minas sudafricanas durante el próximo año. Para ello dispone del siguiente modelo:

$$Y_t = \beta_1 + \beta_2 x_{2t} + \beta_3 x_{3t} + u_t$$

donde y_t representa la cantidad demandada de diamantes en el mundo, x_2 representa el precio de los diamantes y x_3 es un indicador de la riqueza mundial. Suponga que usted es el consultor encargado de asesorar a este grupo distribuidor y a partir de la información disponible debe responder las cuestiones indicadas:

$$(X'X)^{-1} = \begin{pmatrix} 1/100 & 0 & 0 \\ 0 & 1/50 & 0 \\ 0 & 0 & 1/40 \end{pmatrix}$$

$$X'Y = \begin{pmatrix} 30 \\ 20 \\ 40 \end{pmatrix}; R^2 = 0,91875$$

a) Estimar por MCO los parámetros del modelo planteado. Sol.

$$\hat{\beta} = \begin{pmatrix} 0,3 & 0,4 & 1 \end{pmatrix}$$

b) ¿Puede considerarse significativo el indicador de riqueza mundial?. Sol. $t_{exp} = 30,23$

c) Realice un contraste de significación global utilizando un 95 % de confianza. Sol. $F_{exp} = 548,43$

d) Un modelo alternativo, que no considera el indicador de riqueza como variable explicativa, presenta los siguientes resultados

$$\hat{y}_i = 0,25 + 0,45 x_{2i}; R^2 = 0,89$$

¿Cuál de los dos modelos considera usted que es más adecuado para explicar la demanda mundial de diamantes? Sol. ($\bar{R}_1^2 = 0,917074 > 0,8887 = \bar{R}_2^2$)

Ejercicio seleccionado de [4].

12. Dado el modelo $Y_t = \beta_1 + \beta_2 x_{1t} + \beta_3 x_{2t} + u_t$, se obtuvieron los siguientes resultados correspondientes a 12 observaciones:

$$(X'X)^{-1} = \begin{pmatrix} 0,6477 & -0,041 & -0,0639 \\ -0,041 & 0,0071 & -0,0011 \\ -0,0639 & -0,0011 & 0,0152 \end{pmatrix}$$

$$X'Y = \begin{pmatrix} 91 \\ 699 \\ 448 \end{pmatrix}; SCT = 104,9167; Y'Y = 794,994$$

a) Obtener el intervalo de confianza para el valor esperado de Y cuando $X_1 = 2,5$ y $X_2 = 1$

13. Dado un modelo econométrico que explique las ventas de licencias de un software contable en decenas de miles de euros (y_t) para una determinada empresa a lo largo de los últimos 5 años, en función del gasto en miles de euros en publicidad (X_1) y el precio del software en cientos de euros (X_2) se pide:

$$X'X = \begin{pmatrix} 5 & 15 & 25 \\ & 55 & 81 \\ & & 129 \end{pmatrix}; \bar{y} = 4; \sum_{t=1}^{5} y_t x_{1t} = 76; \sum_{t=1}^{5} y_t x_{2t} = 109; y'y = 108$$

a) Obtener un intervalo de predicción para el valor medio de ventas suponiendo que el gasto en publicidad es 5.500 y el precio del software 700. (Sol. $(2,76; 11,73)$)

b) Suponiendo como cierta que los parámetros asociados a las variables x_{1t} y x_{2t} tienen el mismo efecto sobre la variable Y pero de signo contario, obtenga el estimador de mínimos cuadrados restringidos. Datos complementarios: $R(X'X)^{-1}R' = 0,5$. ($\hat{\beta}_{R1} = 11; \hat{\beta}_{R2} = 3,5; \hat{\beta}_{R3} = -3,5$).

Ejercicio seleccionado de [1].

14. Se pretende estimar el modelo que explique el número en miles de vehículos vendidos en el año por la marca A (y) en función de la renta media per capita de la población expresada en miles de euros (x_1) y el precio medio de la marca A en miles de euros (x_2) a partir de la siguiente información:

$$X'X = \begin{pmatrix} 20 & 165,3 & 282,4 \\ & 1390,75 & 2314,72 \\ & & 4020,38 \end{pmatrix}; X'\bar{y} = \begin{pmatrix} 1038 \\ 8749,1 \\ 14501,3 \end{pmatrix}; Y'Y = 55132$$

Se pide:

a) Estimar el modelo e interpretarlo comentando si los resultados esperados coinciden con los que cabría esperar. Sol. $\hat{\beta}' = \begin{pmatrix} 19,65 & 5,97 & -1,2127 \end{pmatrix}$

b) ¿Podríamos esperar que con una renta de 10000 euros y un precio medio de vehículos de la marca A de 10000 euros el número de vehiculos vendidos fuera 8500?. Sol. (64,29019;70,2284)

15. En la siguiente tabla se recogen las ventas de cinco empresas informáticas en función del número de comerciales:

v_t	109	111	132	140	169	180
c_t	12	15	17	18	19	20

Se pide:

a) ¿El número de comerciales influye significativamente en el volumen de ventas de la empresa? ¿Y la constante? (95 % de confianza). (Sol. $t_{exp} = -0,384; t_{exp} = 4,41$).

b) ¿Que volumen medio de ventas se espera que tenga una empresa que tiene 25 comerciales? (Sol. 214,74)

Ejercicio seleccionado de [1].

16. Sea el modelo lineal general $y_i = \beta_1 + \beta_2 x_{2i} + \beta_3 x_{3i} + u_i$, donde:

$$X'X = \begin{pmatrix} 10 & 6 & 8 \\ & 236 & 425 \\ & & 1027 \end{pmatrix} : X'y = \begin{pmatrix} 7 \\ 495 \\ 672 \end{pmatrix} ; y'y = 1250$$

Se pide:

a) Estimar los parámetros del modelo.

$$Sol.\hat{\beta} = \begin{pmatrix} -0,80848 \\ 3,64386 \\ -0,84729 \end{pmatrix}$$

b) Estimar σ^2 y obtener la matriz de varianzas covarianzas. Sol. $\hat{\sigma}^2 = 3,04679$

$$\hat{var}(\hat{\beta}) = \begin{pmatrix} 0,31034642 & -0,01388205 & 0,00332724 \\ & 0,05129633 & -0,02111964 \\ & & 0,01168063 \end{pmatrix}$$

c) Realizar una predicción individual del valor medio teórico de Y condicionado a que los valores de X_2 y X_3 son 120 y 110, respectivamente. Sol. $E[\hat{y}_0] = 343,25282$

d) Obtener las mismas predicciones del apartado anterior pero por intervalo. Sol. (300,93618; 385,56945)

Ejercicio seleccionado de [2].

17. Se cuenta con la siguiente información muestral relativa a la produccion de una factoría, que esta representada por el modelo $y_t = \beta_1 + \beta_2 x_{2t} + \beta_3 x_{3t} + u_t$. La variable y_t es el número de unidades producidas, x_{2t} es el trabajo medio en horas/trabajador y la variable x_{3t} representa el número de horas/maquina. Se pretende conocer si se pueden sustituir horas/trabajador por horas/maquinas sin menoscabo de la producción, es decir, si ambos factores son o no sustitutivos entre si. Se da la siguiente información:

$$X'X = \begin{pmatrix} 100 & 2000 & 2200 \\ & 10000 & 1500 \\ & & 9000 \end{pmatrix} ; X'y = \begin{pmatrix} 5000 \\ 15000 \\ 18000 \end{pmatrix} ; y'y = 100000$$

Se pide:

a) Estimar el modelo. Sol. $\hat{\beta} = \begin{pmatrix} 1,98238 \\ 0,89868 \\ 1,36564 \end{pmatrix}$

b) Estimar el modelo suponiendo que la hipótesis $\beta_2 = \beta_3$ es cierta. Sol. $\hat{\beta}_R = \begin{pmatrix} 1,85233 \\ 1,14637 \\ 1,14644 \end{pmatrix}$

c) Contrastar si la sustitución de un factor por otro sería adecuado.

Ejercicio seleccionado de [5].

18. Se desea analizar el beneficio (medido en millones de euros) de ocho empresas del sector financiero, $\overrightarrow{B}$, a partir del género del director general, $\overrightarrow{G}$ (toma el valor 1 cuando el puesto de director general esta ocupado por una mujer) y del salario del director/a general, $\overrightarrow{S}$ (medido en decenas de miles de euros). Para ello se especifica el siguiente modelo econométrico $B_i = \beta_1 + \beta_2 G_i + \beta_3 S_i + u_i$ y se cuenta con la siguiente información:

$$\left(X'X\right)^{-1} = \begin{pmatrix} 3,14 & 0,26 & -0,68 \\ & 0,59 & -0,12 \\ & & 0,16 \end{pmatrix}, \quad \overline{y} = 3,4375, \quad \sum_{i=1}^{8} B_i G_i = 19, \quad \sum_{i=1}^{8} B_i S_i = 127,5$$

$$\sum_{i=1}^{8} B_i^2 = 111,75$$

Se pide responder de forma razonada las siguientes cuestiones:

a) Estimar el modelo e interpretarlo.(Sol. $\hat{\hat{\beta}}' = \begin{pmatrix} 4,59 & 3,06 & -0,58 \end{pmatrix}$)

b) ¿Entre que valores puede variar el beneficio de una empresa en función del salario de su director/a general?. A la vista de su respuesta, razone si existen diferencias signficativas en el beneficio de una empresa en función del salario de su director/a general. (Sol. $(-0,04860; -1,11139)$, Si, al 95 % la variable salario del director/a general es significativa.)

c) Según este modelo y con un 99 % de confianza, contraste si el género del director/a general de una empresa puede aumentar en cinco millones el beneficio de la empresa.(Sol. $|t_{exp}| = 4,97$)

d) Comparar la bondad de ajuste del modelo anterior con un modelo alternativo en el que se incorpora como variable el número de empleados, obteniendo una suma de los cuadrados de los residuos igual a 0,25. (Sol. $\bar{R}_1^2 = 0{,}8908 < \bar{R}_2^2 = 0{,}9737$)

e) ¿Se puede afirmar, al 5 % de significación, que $\beta_2 = 2 \cdot \beta_3$?(Sol. $F_{exp} = 39{,}004$)

19. Una empresa de vehículos eléctricos con tiendas en distintas provincias desea analizar la demanda ($\overrightarrow{D}$) en cada ciudad (expresada en miles de unidades), en función del precio del producto ($\overrightarrow{P}$) en cada ciudad (expresado en miles de euros) y de la renta media ($\overrightarrow{R}$) de cada

ciudad (medida en miles de euros). Para ello se especifica el siguiente modelo econométrico $D_i = \beta_1 + \beta_2 P_i + \beta_3 R_i + u_i$ y se cuenta con la siguiente información:

$$\left(X'X\right) = \begin{pmatrix} 50 & 0 & 0 \\ & 25 & 0 \\ & & 80 \end{pmatrix}, \quad \sum_{i=1}^{n} D_i = 100, \quad \sum_{i=1}^{n} D_i P_i = -18 \quad \sum_{i=1}^{n} D_i R_i = 150 \quad \sum_{i=1}^{n} D_i^2 = 615$$

Se pide responder de forma razonada las siguientes cuestiones:

a) Estimar el modelo e interpretarlo. (Sol. $\hat{\beta}' = \begin{pmatrix} 2 & -0,72 & 1,875 \end{pmatrix}$)

b) Contraste si existen diferencias significativas en la demanda de vehículos eléctricos en función de su precio.(Sol. $t_{exp} = 2{,}24$)

c) Comparar la bondad de ajuste del modelo anterior con un modelo alternativo en el que se incorpora como variable la población de cada ciudad, obteniendo una suma de los cuadrados de los residuos igual a 92. (Sol. $\bar{R}_1^2 = 0{,}695 < \bar{R}_2^2 = 0{,}755$)

d) Razone con un nivel del confianza del 99 % cuál es el stock mínimo que debe tener la empresa en una ciudad en la que el precio del producto es de 10.000 euros y la renta media anual es de 24.000 euros. (Sol. 25.41 miles de unidades)

e) Contraste las siguientes hipótesis conjuntamente $\beta_2 = 0$ y $\beta_3 = 1$. (Sol. $F_{exp} = 14{,}50$)

Referencias

[1] García, C.B., Sánchez, J.M. y Salmerón, R. (2017) Econometría básica para la economía y la empresa. Ed. Fleming.

[2] García, J., Jiménez, J.F. y Cerrillo, J.R. Econometría práctica. Edo. Libreria Universitaria de Almería.

[3] Johnston, J. (1984) Métodos de econometría. Ed. Vicens Vives.

[4] Pena, B., Estavillo, J., Galindo, E., Leceta, M. y Zamora, M. (1999). Cien ejercicios de econometría. Ed. Pirámide.

[5] Sánchez, C., López, M.M. y García, T. (2015) Econometría. Ed. Fleming.

[6] Matilla García, m., Pérez Pascual, Pedro y Sanz Carnero, B. (2013) Econometría y predicción. Mc Graw Hill.

1.3. Tema 4

ECONOMETRÍA. GADE

Prácticas

Tema 4

1. Se ha estimado el siguiente modelo que explica la recaudación anual del cine (Y) en función del número de espectadores (X_2) y de la frecuencia (X_3):

$$\hat{Y}_t = -4,53789 \times 10^8 + 33805,8X_{2t} - 1,06687 \times 10^9 X_{3t}; R^2 = 0,964711$$

   a. ¿Consideras que los signos obtenidos son coherentes con los esperados? (Sol. El signo de la variable X_3 no parece lógico ya que a mayor frecuencia, mayor debería ser la recaudación.)

   b. Se ha obtenido la regresión auxiliar del número de espectadores en función de la frecuencia obteniéndose un $R^2 = 0,989596$. Calcule el Factor inflactor de la Varianza y concluya acerca de la posible existencia de colinealidad. (Sol. FIV=96,12)

2. Al estimar un modelo se obtiene los siguientes resultados:

$$\begin{array}{rllll} \hat{Y}_t = & 1,3006+ & 0,9452X_{2t}+ & 0,0162X_{3t}+ & 0,0318X_{4t} \\ & (0,2514) & (0,0856) & (0,0403) & (0,1511); \quad R^2 = 0,9994 \end{array}$$

   a) Analice el posible problema de multicolinealidad.

   b) A la vista de las siguientes regresiones auxiliares, ¿Qué podría decir sobre el modelo original? (Sol. FIV=75,188; FIV=63,2911: FIV=26,04)

$$\begin{array}{rlll} \hat{X}_{2t} = & 0,3731+ & 0,3840X_{3t}+ & 0,7535X_{4t} \\ & (1,1007) & (1,1032) & (0,6034); \quad R^2 = 0,9867 \end{array}$$

$$\begin{array}{rlll} \hat{X}_{3t} = & -0,1587+ & 1,7294X_{2t}+ & 0,6138X_{4t} \\ & (2,3542) & (0,4648) & (1,3970); \quad R^2 = 0,9842 \end{array}$$

$$\begin{array}{rlll} \hat{X}_{4t} = & 0,0070+ & 0,2417X_{2t}+ & 0,0437X_{3t} \\ & (0,6285) & (0,1935) & (0,0995); \quad R^2 = 0,9616 \end{array}$$

3. Se estima el siguiente modelo que explica el salario (Y) en función de la edad (X_2), la experiencia (X_3) y el género (X_4) que toma el valor 1 cuando se trata de un hombre.

$$\begin{array}{rllll} Y_t = & 978+ & 7,83X_{2t}+ & 7,08X_{3t}+ & 109X_{4t} \\ & (12,2) & (3,28) & (3,11) & (6,39) \quad R^2 = 0,959 \end{array}$$

   a) Concluir acerca de la existencia de colinealidad a partir de la matriz de correlaciones:

$$\begin{pmatrix} Y & X_2 & X_3 & X_4 & \\ 1 & 0,6210 & 0,6127 & 0,5147 & Y \\ & 1 & 0,9723 & -0,3170 & X_2 \\ & & 1 & -0,3267 & X_3 \\ & & & 1 & X_4 \end{pmatrix}$$

b) Analice la posible existencia de colinealidad sabiendo que el máximo autovalor de la matrix $X'X$ es igual a 3,350 y el mínimo es igual a 0,004. (Sol. CN=28,93)

4. Dadas las siguientes matrices analice la posible existencia de colinealidad, calculando el determinante de la matriz resultante $X'X$: (Sol. $|A'A| = 0; |B'B| = 6,71; |C'C| = 14,39$)

$$A = \begin{pmatrix} 1 & 2 & 4 \\ 1 & 3 & 5 \\ 1 & 5 & 7 \\ 1 & 4 & 6 \\ 1 & 8 & 10 \end{pmatrix} B = \begin{pmatrix} 1 & 2 & 3{,}9 \\ 1 & 3 & 5{,}1 \\ 1 & 5 & 6{,}8 \\ 1 & 4 & 6{,}1 \\ 1 & 8 & 9{,}8 \end{pmatrix} C = \begin{pmatrix} 1 & 2 & 3{,}8 \\ 1 & 3 & 5{,}2 \\ 1 & 5 & 6{,}9 \\ 1 & 4 & 6{,}2 \\ 1 & 8 & 9{,}9 \end{pmatrix}$$

5. Se ha estimado el siguiente modelo para analizar la relación entre la demanda de helado (D_t) desde marzo de 1951 a agosto de 1953 en función de los ingresos medios mensuales de las familias (I_t) y el precio del helado (P_t):

$$\hat{D}_t = 0{,}90 + 0{,}0002135 I_t - 2{,}03 P_t$$

a) Razone acerca de la posible existencia de colinealidad en este modelo sabiendo que el mayor autovalor de la matriz $X'X$ toma el valor 2,996 y el menor autovalor es igual a 0,0004. (Sol. $NC = 86,54$)

b) Razone acerca de la posible existencia de colinealidad a partir de la siguiente regresión. (Sol. $VIF = 1{,}01168$)

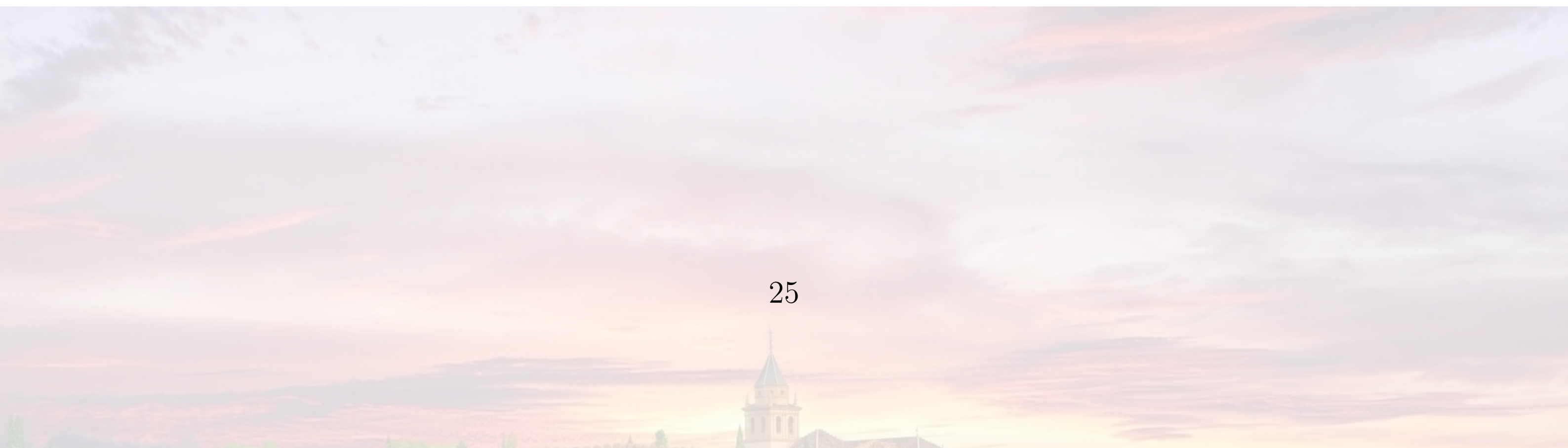
$$\hat{P}_t = 0{,}287446 - 0{,}000143564 I_t$$
$$R^2 = 0{,}011552$$

c) Razone acerca de las posibles incoherencias en la respuesta de los apartados a) y b). ¿Qué tipo de colinealidad existe? ¿Cómo lo resolvería?

Referencias

[1] García, C.B., Sánchez, J.M. y Salmerón, R. (2017) Econometría básica para la economía y la empresa. Ed. Fleming.

1.4. Tema 5



ECONOMETRÍA. GADE

Prácticas

Tema 5

Ejercicios propuestos

1. Dados los siguientes datos contrastar la existencia de heterocedasticidad con el contraste de Golfeld-Quandt, Breusch-Pagan y Glesjer con $h = 1$ y $h = 2$:

Y	2	3	7	6	15	8	22
X	-3	-2	-1	0	1	2	3

   Especifique la matriz de transformación necesario para obtener estimadores eficientes. Sol. (GQ: $F_{exp} = 49$; BP: $\chi^2 = \frac{5{,}56}{2}$; Glesjer: $t_{exp}(h = 1) = 3,64$ y $t_{exp}(h = 2) = 0,2711$)

2. Suponiendo que existe heteroscedasticidad en un modelo lineal, $V_i = \beta_0 + \beta_1 P_i + u_i$, que estudia las ventas en función del precio y que la relación entre la varianza de las perturbaciones aleatorias y el precio es cuadrática, indique como transformaría el modelo para que sea homocedástico. (Sol.$Var(u_i) = E(u_i^2) = \sigma^2 \cdot P_i^2; P^{-1} = \frac{1}{P_i}$)

   Ejercicio seleccionado de [1].

3. En un modelo lineal para estimar el consumo (C) en función de la renta (R) se han ordenado los datos de menor a mayor y se han obtenido las siguientes regresiones:

$$\widehat{C}_i = 1,1 + 1,5 \cdot R_i, \quad \sum_{i=1}^{10} e_i^2 = 230,$$

$$\widehat{C}_i = 20,1 + 0,5 \cdot R_i, \quad \sum_{i=21}^{30} e_i^2 = 43230.$$

   Analice la presencia de heteroscedasticidad en el modelo. (Sol. $F_{exp} = 187,956 > 3,4381$)

   Ejercicio seleccionado de [1].

4. Se tiene el siguiente modelo sobre datos de una empresa aseguradora durante los últimos 38 años, donde y son las pólizas contratadas, x es el gasto en publicidad y z el número de comerciales.

$$\begin{array}{cccc} \hat{y} = & 0,989+ & 0,534x_t+ & 0,183z_t \\ & (1,387) & (0,147) & (0,136) \end{array}$$

$$R^2 = 0,537$$

   a) Si consideramos que $\sigma_u^2 = \sigma^2(x_t z_t)$, ¿Cuál sería la matriz de transformación necesaria para corregir el problema de la heterocedasticidad? (Sol. $P = \frac{1}{\sqrt{x_t z_t}}$)

b) Considerando que los datos para la primera observación son $y_1 = 4$, $x_1 = 11$ y $z_1 = 2$ realice la transformación de esa observación. (Sol. $y*_1 = 0,8525; x*_1 = 2,3452; z*_1 = 0,4264$).

Ejercicio seleccionado de [1].

5. Se dispone de datos anuales para el periodo 1985-2014 de consumo agregado, Y_t, y renta disponible agregada, X_t. Se considera el siguiente modelo de consumo para la economía española:

$$Y_t = \beta_1 + \beta_2 X_t + u_t$$

Teniendo en cuenta que X_t no ha dejado de crecer durante todo el periodo 1985-2014, se ha estimado el modelo para los periodos de 1985-1994 y 2005-2014, obteniéndose los siguientes resultados:

$$\begin{array}{lll} \hat{y}_t = 3{,}341,27 + 0,8X_t & R^2 = 0,985 & \sum_{t=1985}^{1994}(Y_t - \bar{Y})^2 = 148{,}456,17 \\ \hat{y}_t = 23{,}182,01 + 0,955X_t & R^2 = 0,305 & \sum_{t=2005}^{2014}(Y_t - \bar{Y})^2 = 21{,}418,47 \end{array}$$

Utiliza la información disponible para contrastar la hipótesis nula de que la varianza de las perturbaciones se ha mantenido constante en el tiempo desde 1985 hasta 2014. (Sol. $F_{exp} = 6,68$). *Ejercicio seleccionado de [1].*

6. Se tienen los siguientes datos para el empleo en el sector turístico (Y) en las distintas comunidades españoles para el año 2005 asi como del número de viajeros (X_1) y la estancia media (X_2). Se pide:

Provincias	Empleo	Viajeros	Estancia
Andalucía	28,4	11.902	3,1
Aragón	3,6	1.848	2,1
Asturias	2,4	1.088	2,3
Baleares	25,9	6.716	7,2
Canarias	27,2	4.875	7,8
Cantabria	2	933	2,4
Castilla y León	6,2	3647	1,7
Castilla Mancha	2,8	1805	1,7
Cataluña	23,5	10.771	3,4
Valencia	13,4	5.579	3,9
Extremadura	2,2	1000	1,7
Galicia	6,3	3.040	2,1
Madrid	10,7	5.748	2,1
Murcia	2	882	3
Navarra	1,1	557	2
Pais Vasco	3,2	1.540	1,9
La Rioja	0,7	446	1,8

Se pide:

a) Estimar el modelo e interpretrar los coeficientes. (Sol. $\hat{\beta}_0 = -5,7358; \hat{\beta}_1 = 0,0019873; \hat{\beta}_2 = 2,69218$)

b) Obtener el coeficiente de determinación. (Sol. $R^2 = 0,9836$)

c) Contrastar la existencia de heterocedasticidad mediante el contraste de Goldfeld-Quandt. (Sol. $F_{exp} = 1358,36278$)

d) Contrastar la existencia de heterocedasticidad mediante el test de Glesjer para $h = 1; 2; -1; 1/2$.

$$|\hat{e}_i| = 0,589 + 0,00010X_{1i}; t_{exp} = 2,07526; R^2 = 0,223$$

$$|\hat{e}_i| = 0,7654 + 8,87696 \cdot 10^{-09} X_{1i}^2; t_{exp} = 1,988; R^2 = 0,208572$$

$$|\hat{e}_i| = 1,22 - 348,03X_{1i}^{-1}; t_{exp} = -1,079; R^2 = 0,0720$$

$$|\hat{e}_i| = 0,2671 + 0,0132X_{1i}^{1/2}; t_{exp} = 1,9469; R^2 = 0,2017$$

e) A la vista de los resultados anteriores qué solución propondría para resolver la heterocedasticidad.

Ejercicio seleccionado de [1].

7. En una estación agrícola se desea ensayar el efecto de un determinado fertilizante sobre la producción de trigo. Para ello se han elegido veintidós parcelas de terreno de igual superficie y se ha planteado un modelo sobre la producción total (Y) en función de la cantidad de fertilizante (CF), que ha dado lugar a los siguientes resultados:

$$\begin{array}{rll} \hat{Y}_i = & 136,9029+ & 0,666648CF_i \\ & (567,3311) & (0,006312) \end{array}$$

$$R^2 = 0,99821$$

a) ¿Influye de forma significativa la cantidad de fertilizantes en la explicación de la variable dependiente?

b) Reordenando todas las variables de la regresión original de menor a mayor de acuerdo con la variable cantidad de fertilizantes, se ajustan dos regresiones por separado, una para los siete primeros datos y otra para los siete últimos, con el siguiente resultado:

$$\begin{array}{rll} \hat{Y}_i = & -49,20379 & +0,718676CF_i \\ & (115,8282) & (0,007438) \end{array}$$

$$\hat{\sigma}_u^2 = 251,2505$$

$$\begin{array}{rll} \hat{Y}_i = & 847,984 & +0,670859CF_i \\ & (1301,311) & (0,006854) \end{array}$$

$$SCR = 397588,8$$

Contrastar, utilizando toda la información disponible, la presencia de heteroscedasticidad. Sol. $F_{exp} = 316,48$.

c) Nuevos estudios han dado como resultado las siguientes regresiones adicionales

$$\begin{array}{rcc} e_i^2 = & 2405525+ & 80,51248CF_i \\ & (3846351) & (42,7965) \end{array}$$

$$R^2 = 0,662736$$

$$\begin{array}{rcc} e_i^2 = & -1407812+ & 57939,85CF_i^2 \\ & (4113921) & (21466,97) \end{array}$$

$$R^2 = 0,266898$$

$$\begin{array}{rcc} e_i^2 = & 4521{,}236+ & 0,000104\sqrt{CF_i} \\ & (3949,677) & (0,000149) \end{array}$$

$$R^2 = 0,023942$$

$$\begin{array}{rcc} e_i^2 = & 15252876 & -4,65\frac{1}{CF_i} \\ & (6159339) & (2,20) \end{array}$$

$$R^2 = 0,159635$$

Utilizando la nueva información: ¿existen problemas de heteroscedasticidad. En caso afirmativo, ¿cómo se estimarían los parámetros del modelo de forma eficiente? Especificar la matriz de transformación. Sol. $P^{-1} = \frac{1}{CF_i}$ *Ejercicio adaptado a partir de [2].*

8. Ante la próxima reunión de la Comisión de la Unión Europea sobre la concesión de subvenciones para la explotación agraria dedicadas a verduras y hortalizas, se ha establecido un modelo que explica la producción de verduras y hortalizas (Y) en función de la superficie cultivada medida en hectáreas (X) obteniendo el siguiente modelo estimado:

$$\hat{y}_i = 302,7623 + 0,171795x_i$$

Se ha obtenido la estimación maximo verosimil de la varianza de la perturbación $\hat{\sigma}^2_{MV} = 2924222,58824$ y se han obtenido los residuos estandarizados al cuadrado (g_i). De la regresión auxiliar en la que se explica g_i en función de la variable explicativa x_i se ha obtenido $SCR = 21,42911$ y $R^2 = 0,271035$. Use el contraste de Breusch-Pagan para contrastar la posible existencia de heterocedasticidad al 95 % de confianza. (Sol. $\chi^2 = 3,9837$)

Ejercicio adaptado a partir de [2].

9. Una empresa desea realizar un estudio sobre la relación entre el número de sucursales (S) y el numero de licencias comerciales (L) estimando la siguiente ecuación sobre una muestra de 50 observaciones:

$$\begin{array}{rcc} \hat{S} = & 5,88+ & 0,2L_t \\ & (2,3) & (0,02) \end{array}$$

$$R^2 = 0,58$$

a) Tras ordenar la muestra en función de los valores de la variable L se han estimado dos regresiones de la forma $S_t = \beta_0 + \beta_1 L_t + u_t$ con las primeras y últimas 12 observaciones obteniéndose SCR_1=1,58 y SCR_2=2,47 respectivamente. Contrasta la posible existencia de heterocedasticidad. Explica lo que haces y por qué lo haces. Sol. $F_{exp} = 1{,}5632$

b) Si suponemos que $E(uu') = \sigma^2 L_t$ escribe el modelo transformado en el que las perturbaciones sean homocedásticas.

10. (Septiembre, 2015) Se han analizado 11 producciones de Hollywood para relacionar el impacto que suponen los gastos de promoción (en millones de dolares) sobre las ganancias (en millones de dolares) el primer año de la película. Utilizando el método de mínimos cuadrados ordinarios se obtuvo el siguiente modelo:

$$GANAN_i = 33,35 + 9,91 GASTOSPROM_i$$

donde se utilizaron los siguientes datos y se obtuvieron los siguientes residuos:

$GANAN_i$	30	50	90	50	82	85	90	105	120	80	130
$GASTOSPROM_i$	1	2	8	3	4	5	5	6	8	3	10
e_i	-13.26	-3.17	-22.63	-13.08	9.01	2.1	7.1	12.19	7.37	16.92	-2.45

a) Aplicar el test de Goldfeld-Quandt obviando 3 observaciones en el análisis. Comentar los resultados obtenidos.

b) ¿Coinciden las conclusiones del apartado anterior con las que se obtienen aplicando el test de Breuch y Pagan? (Regresión auxiliar con $SCR = 12,5871$ y $\bar{R}^2 = -0,109768$).

11. Se quiere analizar la dependencia de las comisiones de los agentes de seguros en base a los años que lleva trabajando en la aseguradora (X_1), el número de pólizas contratadas de media al año (X_2), los siniestros medios por año (X_3) y los nuevos clientes captados el año anterior (X_4). Se tienen datos de 50 agentes. Utilizando el método de mínimos cuadrados ordinarios se obtuvo el siguiente modelo:

$$COMISION_i = 1182,18 - 1,94X_{1i} + 2,71X_{2i} - 0,27X_{3i} + 54,72X_{4i}$$

$$R^2 = 0,69$$

Con la siguiente información, y suponiendo que no existe autocorrelación en la pertubación aleatoria del modelo, ¿es eficiente la estimación anterior?

$$|\hat{e}_i| = \underset{(111,09)}{1590{,}59+} \; \underset{(66777{,}01)}{146209\tfrac{1}{\sqrt{X_{3i}}}} \; ; SCR = 24239$$

$$|\hat{e}_i| = \underset{(244{,}79)}{1483{,}10+} \; \underset{(8607{,}97)}{18633{,}43\tfrac{1}{X_{3i}}} \; ; SCR = 23935$$

$$|\hat{e}_i| = \underset{(195{,}76)}{330{,}47+} \; \underset{(0{,}07)}{0{,}059X_{3i}} \; ; SCR = 74323$$

$$|\hat{e}_i| = \underset{(542{,}55)}{1556{,}22+} \; \underset{(10{,}76)}{2{,}25\sqrt{X_{3i}}} \; ; SCR = 24515$$

Justifica la respuesta y menciona el método utilizado. En caso de no ser eficiente, ¿cómo realizarías la estimación de los coeficientes para que sí lo fuera?

12. (Parcial 2, 2018) Dado el modelo $Y_i = \beta_1 + \beta_2 X_i + u_i$ donde $\hat{var}(u_i) = \sigma^2 X_i^2$ y $E(u_i u_j) = 0$ se tiene la siguiente información acerca de una muestra de 800 observaciones:

$$\sum X_i = 330; \sum X_i^2 = 144; \sum \frac{1}{X_i} = 2058; \sum \frac{1}{X_i^2} = 5683$$

$$\sum Y_i = 2672; \sum Y_i^2 = 9576; \sum \frac{Y_i}{X_i} = 6835; \sum \frac{Y_i}{X_i^2} = 18755$$

Se pide transformar el modelo y estimar el modelo para obtener estimadores eficientes. Sol. (3.014; 0.7879)

13. Se tiene información sobre la relación del consumo anual (C_i medido en miles de euros) de 81 familias granadinas cuyo ingreso principal procede del sector turistico con el nivel de dichos ingresos (I_i medido en miles de euros) y el numero de personas (P_i) que constituyen la unidad familiar, obteniendose el siguiente modelo:

$$\hat{C}_i = \underset{(3,23)}{6,759+} \; \underset{(0,095)}{0,422 I_i} \; \underset{(0,947)}{+2,637 P_i}$$

 a) Sabiendo que el coeficiente de determinación es igual a 0,3532, contrastar la significación global del modelo. Sol. ($F_{exp} = 21,2968$)

 b) Usando un intervalo de confianza razone si es posible que el aumento de una persona a la unidad familiar aumente en 5.000 euros el consumo de la familia. Sol. ($IC(95\,\%)(0{,}75247; 4,5215)$)

 c) Razone a partir de las siguientes regresiones si el modelo presenta heterocedasticidad y cómo se podrían obtener estimadores eficientes.

$$\hat{|e|}_i = \underset{(0,19)}{2,69+} \; \underset{(0,33)}{1,35 I_i} \; ; SCR = 16,25$$

$$\hat{|e|}_i = \underset{(0,26)}{4,32+} \; \underset{(0,51)}{2,65 I_i^2} \; ; SCR = 12,38$$

Referencias

[1] García, C.B., Sánchez, J.M. y Salmerón, R. (2017) Econometría básica para la economía y la empresa. Ed. Fleming.

[2] Pena, B., Estavillo, J., Galindo, E., Leceta, M. y Zamora, M. (1999). Cien ejercicios de econometría. Ed. Pirámide.

1.5. Tema 6

ECONOMETRÍA. GADE

Prácticas

Tema 6

Ejercicios propuestos

1. Una empresa desea realizar un estudio sobre la relación entre el número de sucursales (S) y el numero de licencias comerciales (L) estimando la siguiente ecuación sobre una muestra de 50 observaciones:

$$\begin{array}{rll} \hat{S} = & 5,88+ & 0,2L_t \\ & (2,3) & (0,02) \end{array}$$

$$R^2 = 0,58$$

Si el investigador realiza la siguiente estimación:

$$\begin{array}{rlll} \hat{S}_t = & 10+ & 3,5L_t+ & 2,1S_{t-1} \\ & (8,3) & (2,7) & (0,005) \end{array}$$

$$dw = 0,63$$

Contrasta la existencia de autocorrelación en el modelo. Explica lo que haces y por qué lo haces.

2. Dado un modelo lineal del consumo en función del PIB a partir de estos datos:

t	1	2	3	4	5
C	-5	4	1	-2	2
PIB	2	0	-1	1	2

Contraste, explicando y justificando su respuesta, la existencia de auto correlación de primer orden sabiendo que la regresión del modelo original produce los siguientes residuos:

t	1	2	3	4	5
e	-3,41	2,94	-1,38	-1,73	?

(Sol. e_5=3,58; $D-W = 2,29$; $d_L = 0,61$; $d_U = 1,4$; No hay evidencias para rechazar la hipótesis nula y, por tanto, decimos que no existe autocorrelación.)

3. Se tiene el siguiente modelo sobre datos de una empresa aseguradora durante los últimos 38 años, donde y son las pólizas contratadas, x es el gasto en publicidad y z el número de comerciales.

$$\begin{array}{rlll} \hat{y} = & 0,989+ & +0,534x_t+ & +0,183z_t \\ & (1,387) & (0,147) & (0,136) \end{array}$$

$$R^2 = 0,537$$

Contraste la existencia de autocorrelacion en las perturbaciones en el modelo, sabiendo que $e_t = 0,327e_{t-1}$. (Sol. $DW = 1,346$). *Ejercicio seleccionado de [1].*

4. Se ha estimado un modelo que explica la recaudacion anual del cine en función del numero de espectadores y de la frecuencia, obtiendose los siguientes residuos:

2010	-7669280
2011	2097942
2012	22574579
2013	-5232218
2014	325068.0
2015	-12096091

Contraste si existe autocorrelación en el modelo usando el contraste de Durbin-Watson (Sol. $DW = 1{,}9798$) *Ejercicio seleccionado de [1].*

5. Se ha estimado el siguiente modelo que explica las ventas (Y_t) en función del numero de tiendas (X_1), el numero de ventas del año anterior (Y_{t-1}) y el numero de empleados (X_2) durante 20 años:

$$\begin{array}{lcccc} \hat{y}_t = & -56,51 & -3,70X_1 & 3,21Y_{t-1} & 0,164X_2 \\ & (162,90) & (0,8054) & (0,06017) & (0,02284) \end{array}$$

Se sabe que $D - W = 1,66$. Contraste la existencia de autocorrelación. (Sol. $H - durbin = 0,7839$)

Ejercicio seleccionado de [1].

6. El departamento de comercio de Canada está estudiando el comportamiento de sus importaciones. En particular, una de las comisiones de estudio ha decidido analizar las importaciones provenientes de las economías emergentes de Asia, para lo que cuenta con datos mensuales de estas dos variables desde enero de 1980 a octubre de 1994. El departamento se plantea distintos modelos en los que las importaciones asiaticas dependen del PNB:

$$ModeloI$$

$$\begin{array}{lcc} \hat{M}_t = & 46925,64+ & 4,171127PNB_t \\ & (2110,03) & (0,125471) \end{array}$$

$$R^2 = 0,862623; dw = 0,12013$$

$$ModeloII$$

$$\begin{array}{lccc} \hat{M}_t = & 14948,67+ & 2,133735PNB_t+ & 3,932413PIB_{t-1} \\ & (684,9666) & (1,552213) & (1,552912) \end{array}$$

$$R^2 = 0,977436; dw = 0,042266$$

$$ModeloIII$$

$$\begin{array}{lccc} \hat{M}_t = & 672,2529+ & 0,089347PNB_t+ & 0,98357M_{t-1} \\ & (179,1994) & (0,058276) & (0,009485) \end{array}$$

$$R^2 = 0,999339; dw = 1,699329$$

¿Qué modelo es más adecuado para la predicciones de las importaciones de productos asiaticos? Sol. Modelo I $DW = 0,121013$, modelo II $DW = 0,042266$, modelo III $H - D = 2,084569$. *Ejercicio seleccionado de [2].*

7. Al estudiar la relación entre las licencias vendidas por una empresa de software informático, el gasto en publicidad y el gasto en publicidad del año anterior se han obtenido los siguientes resultados correspondientes al periodo 1975-2014: $\sum e_t^2 = 6,2$ y $\sum e_t e_{t-1} = 2,29$. Introduciendo aquellas hipótesis que estime necesarias realice un contraste de autocorrelación utilizando solamente la información anterior. Justifique todos los pasos.

8. Se estima mediante Gretl un modelo que pretende explicar la producción agregada (q) en función del ratio de capital/trabajo agregado (k) y un índice tecnológico (A) para una empresa desde 1969 hasta 2009, obteniendo los siguientes resultados:

Variable dependiente: q

	Coeficiente	Desv. Típica	Estadístico t	valor p
const	−0.295445	0.0111377	−26.53	0.0000
k	0.0956250	0.00398550	23.99	0.0000
A	0.717229	0.00491408	146.0	0.0000

Media de la vble. dep.	0.905976	D.T. de la vble. dep.	0.201284
Suma de cuad. residuos	0.002500	D.T. de la regresión	0.008112
R^2	0.998457	R^2 corregido	0.998376
$F(2, 38)$	12295.91	Valor p (de F)	3.79e–54
Log-verosimilitud	140.7739	Criterio de Akaike	−275.5477
Criterio de Schwarz	−270.4070	Hannan–Quinn	−273.6758
$\hat{\rho}$	0.483540	Durbin–Watson	1.022892

Se pide:

a) Contrastar la significación individual de cada uno de los parámetros.

b) Contrastar la significación global del modelo.

c) Obtener un intervalo de confianza para el parámetro asociado a la variable índice tecnológico.

d) Dada la siguiente matriz obtener un intervalo de predicción suponiendo que el ratio de capital/trabajo agregado es 2,5 y el índice tecnológico 1,7:

$$(X'X)^{-1} = \begin{pmatrix} 1,885 & -0,5490 & -0,3145 \\ -0,5490 & 0,2414 & -0,0652 \\ -0,3145 & -0,0652 & 0,3670 \end{pmatrix}$$

e) Contrastar la existencia de autocorrelación en el modelo.

9. A partir de los datos utilizados por Stock and Watson (1993) se ha estimado el siguiente modelo que relaciona el logaritmo de la demanda de interes (*lmoney*) con el logaritmo de los ingresos (*lincome*) y la tasa de interés (*intrate*) de E.E.U.U. para los años 1900-1989. Se pide:

Variable dependiente: lmoney

	Coeficiente	Desv. Típica	Estadístico t	valor p
const	-0.771121	0.0429311	.	.
lincome	0.940341	0.0197292	.	.
intrate	-0.0829121	0.00541344	.	.

Media de la vble. dep.	0.977332	D.T. de la vble. dep.	0.692195
Suma de cuad. residuos	1.549942	D.T. de la regresión	0.133474
R^2	0.963653	R^2 corregido	0.962817
$F(2, 87)$		Valor p (de F)	
Log-verosimilitud	55.06717	Criterio de Akaike	-104.1343
Criterio de Schwarz	-96.63491	Hannan–Quinn	-101.1101
$\hat{\rho}$	0.717793	Durbin–Watson	

a) Contrastar la significación individual de cada uno de los parámetros.

b) Contrastar la significación global del modelo.

c) Obtener un intervalo de confianza para el parámetro asociado a la variable *intrate*.

d) Obtener un intervalo de predicción suponiendo que *lincome* toma el valor 4 y *intrate* toma el valor 8 dada la matriz:

$$(X'X)^{-1} = \begin{pmatrix} 0,10 & -3757,43 & -162,21 \\ -3757,43 & 2184,87 & -259,63 \\ -162,21 & -259,63 & 16449396,82 \end{pmatrix}$$

e) Contrastar la existencia de autocorrelación en el modelo.

10. Dado el modelo $Y_t = \beta_1 + \beta_2 X_{2t} + \beta_3 X_{3t} + u_t$ y suponiendo que se ha llegado a la conclusión de que $u_t = \rho u_{t-1} + \epsilon_t$ y $\hat{var}(u_t) = \sigma^2$, razone cuál sería la transformación necesaria para obtener estimadores eficientes y compruebe que el modelo transformado verifica las hipótesis básicas sobre la perturbación.

11. Se ha estimado un modelo econométrico con datos trimestrales desde 1974 hasta 1996 para explicar el comportamiento de las exportaciones de cítricos (y) en función del Producto Interior Bruto (X_1), el nivel de precios (X_2), el volumen de credito oficial concedido a las exportaciones (X_3) y la superficie cultivada (X_4) obteniendo la siguiente recta de regresión muestral:

$$\hat{Y}_t = 692 + 0{,}81 X_{1t} - 0{,}38 X_{2t} + 2{,}1 X_{3t} + 3{,}9 X_{4t}$$

$$R^2 = 0,89; e'e = 18,48$$

a) Se sabe que $\sum(e_t - e_{t-1})^2 = 69,069$ y $\sum(e_t^2) = 21,817$. Contraste la posible existencia de autocorrelación AR(1). A la vista de los resultados, represente cómo esperaría que fuera el gráfico de los residuos frente a su primer retardo.

12. Se tiene información para los años 1996-2014 del producto nacional bruto (y_t), la masa monetaria (M_t), la inversion bruta del año anterior (I_{t-1}) y los gastos de gobierno (G_t) y se estima el siguiente modelo:

$$\hat{y}_t = -2{,}77 + 15{,}12M_t - 3{,}48I_{t-1} - 5{,}13G_t$$

Se ha obtenido un valor para el estadistico Durbin Watson igual a 0.82.

a) Contraste la posible existencia de autocorrelación. Sol. Autocorrelación positiva $DW = 0{,}82, D_L = 0{,}967; D_U = 1{,}685$

b) Suponiendo que exista correlación de tipo AR(1), indique el primer y segundo valor de la variable y transformada para obtener estimadores eficientes, sabiendo que dicha variable toma el valor 6,39 y 11,85 para los años 1996 y 1997, respectivamente. Sol. $y_1* = 5{,}1593; y_2* = 8{,}0799$

c) Se ha ordenado la muestra de menor a mayor en función de la masa monetaria y se han eliminado siete observaciones centrales. A continuación se ha realizado la regresión de los seis primeros datos obteniendo los siguientes residuos:

$$e_1 = -0{,}011759; e_2 = -0{,}67981; e_3 = 0{,}808758; e_4 = -0{,}03508; e_5 = 0{,}252570; e_6 =?$$

Se ha realizado igualmente la regresión de los seis ultimos datos obteniendo los siguientes residuos:

$$e_1 = -2{,}88582; e_2 = 1{,}413507; e_3 = 5{,}44764; e_4 = -0{,}420075; e_5 = -0{,}31586; e_6 =?$$

Se pide contrastar la posible existencia de heterocedasticidad explicando el procedimiento utilizado. A la vista de los resultados, represente cómo esperaría que fuera el gráfico de los residuos frente a la variable explicativa masa monetaria. Sol. $F_{exp} = 39{,}25 < F_{2,2}(0{,}95) = 19$

Referencias

[1] García, C.B., Sánchez, J.M. y Salmerón, R. (2017) Econometría básica para la economía y la empresa. Ed. Fleming.

[2] Pena, B., Estavillo, J., Galindo, E., Leceta, M. y Zamora, M. (1999). Cien ejercicios de econometría. Ed. Pirámide.

2 Soluciones Ejercicios Propuestos

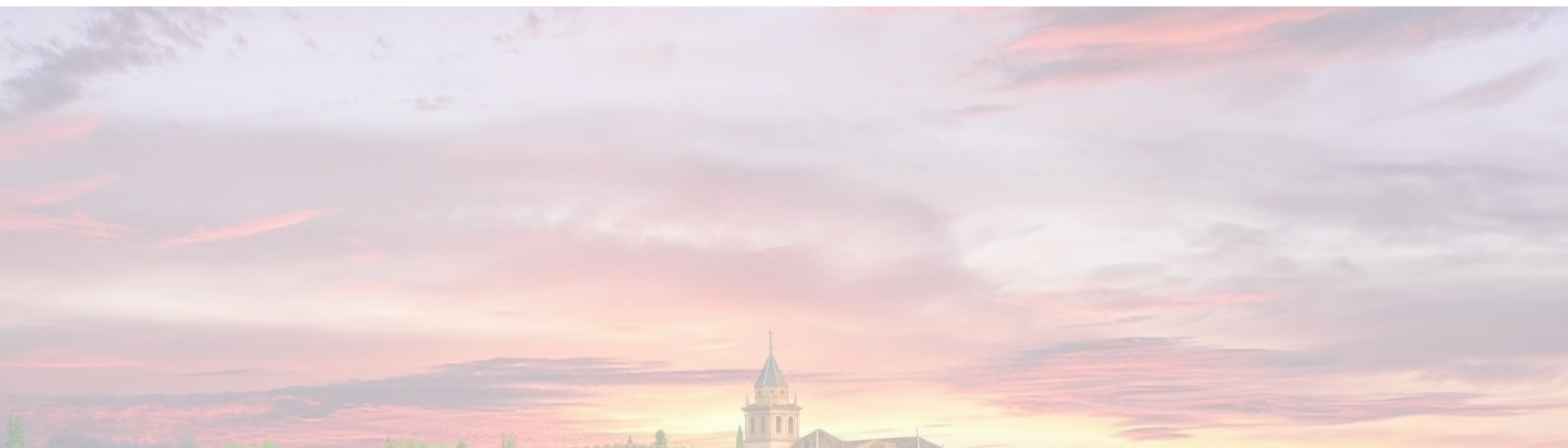
FACULTAD DE CIENCIAS
ECONÓMICAS Y EMPRESARIALES

Ingeniería Informática + ADE

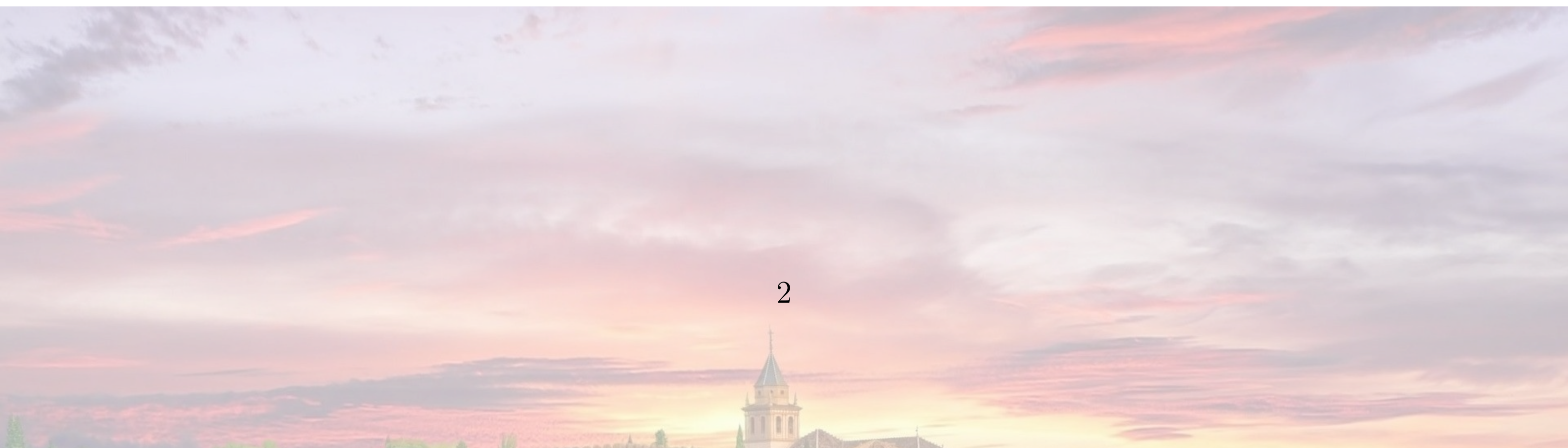
Universidad de Granada (UGR)

Autor: Ismael Sallami Moreno

Asignatura: Econometría: Ejercicios Propuestos Solucionados



Índice



1 Tema 2

1.1. Ejercicio 1

Enunciado

1. Detalle el orden de las matrices y vectores de un modelo econométrico con 3 variables explicativas más el termino constante y 50 observaciones.

Solución

Para ello vamos a suponer que la variable explicada es Y_t, si denotamos a las demás varibales explicativas con la simbología β_i, siendo i el índice de la variable explicativa, podemos nombra a la matriz de variables explicativas como X y a la matriz de coeficientes como β, y nos queda la siguiente ecuación:

$$Y_t = \beta_0 + \beta_1 X_{1t} + \beta_2 X_{2t} + \beta_3 X_{3t} + \beta_4 X_{4t} + u_t \tag{1}$$

De manera más sintética, nos quedaría:

$$\vec{y} = X\vec{\beta} + \vec{u} \tag{2}$$

Por lo que podemos decir:

- Orden de la matriz $\vec{y} = (50x1)$
- Orden de la matriz $X = (50x4)$
- Orden de la matriz $\vec{\beta} = (4x1)$
- Orden de la matriz $\vec{u} = (50x1)$

1.2. Ejercicio 2

Enunciado

2. Ponga un ejemplo de una matriz X con un término constante y tres variables explicativas de manera que el modelo no cumpla la condición del rango completo por columnas.

Solución

Una matriz X con un término constante y tres variables explicativas que no cumpla la condición del rango completo por columnas es aquella en la cual al menos una de las columnas es una combinación lineal de las demás. Por ejemplo:

$$X = \begin{pmatrix} 1 & 2 & 3 & 6 \\ 1 & 4 & 5 & 9 \\ 1 & 7 & 8 & 15 \end{pmatrix}$$

En este ejemplo, la cuarta columna es la suma de la segunda y la tercera columnas, lo que significa que el rango de la matriz no es completo.

1.3. Ejercicio 3

Enunciado

3. Razone qué componentes del modelo econométrico tienen carácter estocástico (aleatorio) y cuáles tienen carácter determinista (fijo).

Solución

En un modelo econométrico típico, hay tanto componentes deterministas como estocásticos.

- **Componentes deterministas (fijos):**

- **Término constante**: Este es el intercepto del modelo, representado generalmente como β_0, y es fijo.
- **Coeficientes de las variables explicativas**: Representados como $\beta_1, \beta_2, \ldots, \beta_k$, son los parámetros del modelo que permanecen fijos una vez estimados.

- **Componentes estocásticos (aleatorios):**

- **Término de error**: Representado como ϵ_i o u_i, este componente capta la variabilidad no explicada por el modelo y es aleatorio.

Por ejemplo, en el modelo lineal clásico:

$$Y_i = \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \ldots + \beta_k X_{ik} + \epsilon_i$$

Los componentes deterministas son $\beta_0, \beta_1, \ldots, \beta_k$ y los componentes estocásticos son los ϵ_i.

1.4. Ejercicio 4

Enunciado

Ponga un ejemplo de una matriz de varianzas-covarianzas de las perturbaciones de un modelo con heterocedasticidad.

Solución

En un modelo con heterocedasticidad, la varianza de los errores no es constante a lo largo de las observaciones. Un ejemplo de matriz de varianzas-covarianzas de las perturbaciones podría ser:

$$\Sigma = \begin{pmatrix} \sigma_1^2 & 0 & 0 \\ 0 & \sigma_2^2 & 0 \\ 0 & 0 & \sigma_3^2 \end{pmatrix}$$

donde σ_1^2, σ_2^2 y σ_3^2 son las varianzas de los errores en las diferentes observaciones y son diferentes entre sí, reflejando la presencia de heterocedasticidad.

Por ejemplo:

$$\Sigma = \begin{pmatrix} 4 & 0 & 0 \\ 0 & 9 & 0 \\ 0 & 0 & 16 \end{pmatrix}$$

En este caso, $\sigma_1^2 = 4$, $\sigma_2^2 = 9$, y $\sigma_3^2 = 16$, lo cual indica que las perturbaciones tienen varianzas distintas.

1.5. Ejercicio 5

Enunciado

Ponga un ejemplo de una matriz de varianzas-covarianzas de las perturbaciones de un modelo con autocorrelación.

Solución

En un modelo con autocorrelación, la matriz de varianzas-covarianzas de las perturbaciones podría tener la siguiente forma:

$$\Sigma = \begin{pmatrix} \sigma^2 & \rho\sigma^2 & \rho^2\sigma^2 \\ \rho\sigma^2 & \sigma^2 & \rho\sigma^2 \\ \rho^2\sigma^2 & \rho\sigma^2 & \sigma^2 \end{pmatrix}$$

donde σ^2 es la varianza de los errores y ρ es el coeficiente de autocorrelación.

Un ejemplo específico podría ser:

$$\Sigma = \begin{pmatrix} 1 & 0{,}5 & 0{,}25 \\ 0{,}5 & 1 & 0{,}5 \\ 0{,}25 & 0{,}5 & 1 \end{pmatrix}$$

En este caso, $\sigma^2 = 1$ y $\rho = 0{,}5$. Esto muestra que las perturbaciones están correlacionadas entre sí, reflejando la presencia de autocorrelación en el modelo.

1.6. Ejercicio 6

Enunciado

6. Se considera la posibilidad de introducir nuevas variables explicativas en el modelo de la curva de Phillips (EJERCICIO 1 de la relación de ejercicios resueltos) y para ello se recoge información sobre la renta nacional disponible neta a precios de mercado por habitante (X_2) y la renta nacional disponible neta (X_3).

Años	X_2	X_3
2006	18614	825737
2007	19403	877724
2008	19492	896295
2009	18719	867972
2010	18706	871015
2011	18229	851948
2012	17822	833445
2013	17691	824821
2014	18029	837556
2015	18828	873766

Se pide:

a) Estimar el modelo incluyendo la variable X_2. (**Sol.** $\hat{y} = 165,78 - 1,48X_1 - 0,0077X_2$)

b) Estimar el modelo incluyendo las variables X_2 y X_3. (**Sol.** $\hat{y} = 120,97 - 0,77X_1 - 0,017X_2 + 0,00025X_3$)

c) Comparar ambos modelos. (**Sol.** $R_1^2 = 0,7097 < R_2^2 = 0,96; AIC_1 = 2,83 > AIC_2 = 0,8920$)

Solución

- **Modelo incluyendo la variable X_2:**

$$\hat{y} = 165,78 - 1,48X_1 - 0,0077X_2$$

Para estimar este modelo, hemos utilizado una regresión lineal múltiple donde X_1 es la variable original del modelo de la curva de Phillips y X_2 es la nueva variable explicativa añadida (renta nacional disponible neta a precios de mercado por habitante). Los coeficientes $165,78$ (intercepto), $-1,48$ (coeficiente de X_1), y $-0,0077$ (coeficiente de X_2) han sido estimados a partir de los datos proporcionados usando el método de mínimos cuadrados ordinarios (MCO).

Para ello:

- Debemos de calcular los siguiente valores:

$$\sum_{i=1}^{10} X_{1_i}$$
$$\sum_{i=1}^{10} X_{1_i}^2$$
$$\sum_{i=1}^{10} X_{2_i}$$
$$\sum_{i=1}^{10} X_{2_i}^2$$
$$\sum_{i=1}^{10} X_{1_i} Y_t$$
$$\sum_{i=1}^{10} X_{2_i} Y_t$$
$$\sum_{i=1}^{10} Y_t$$

- Con estos valores calculamos la siguiente matriz:

$$X'X = \begin{pmatrix} n & \sum_{i=1}^{10} X_{1_i} & \sum_{i=1}^{10} X_{2_i} \\ \sum_{i=1}^{10} X_{1_i} & \sum_{i=1}^{10} X_{1_i}^2 & \sum_{i=1}^{10} X_{1_i} X_{2_i} \\ \sum_{i=1}^{10} X_{2_i} & \sum_{i=1}^{10} X_{1_i} X_{2_i} & \sum_{i=1}^{10} X_{2_i}^2 \end{pmatrix}$$

- Y calculamos de manera análoga la siguiente matriz:

$$X'Y = \begin{pmatrix} \sum_{i=1}^{10} Y_t \\ \sum_{i=1}^{10} X_{1_i} Y_t \\ \sum_{i=1}^{10} X_{2_i} Y_t \end{pmatrix}$$

- Por último, calculamos la matriz de del estimador, aplicando el estimador de MCO:

$$\hat{\beta} = (X'X)^{-1} X'Y$$

- **Modelo incluyendo las variables X_2 y X_3**:

$$\hat{y} = 120,97 - 0,77X_1 - 0,017X_2 + 0,00025X_3$$

Similarmente, para este modelo se ha llevado a cabo una regresión lineal múltiple incluyendo ahora tanto X_2 como X_3 (renta nacional disponible neta total). Los coeficientes $120,97$ (intercepto), $-0,77$ (coeficiente de X_1), $-0,017$ (coeficiente de X_2), y $0,00025$ (coeficiente de X_3) han sido obtenidos mediante el método de MCO.

- **Comparación de ambos modelos**:

- Coeficiente de determinación ajustado (R^2):

$$R_1^2 = 0,7097 \quad \text{(modelo con } X_2 \text{ solamente)} < R_2^2 = 0,96 \quad \text{(modelo con } X_2 \text{ y } X_3\text{)}$$

  El coeficiente de determinación ajustado (R^2) indica el porcentaje de variabilidad en la variable dependiente que es explicado por las variables independientes. En este caso, el modelo que incluye X_2 y X_3 explica el 96 % de la variabilidad en Y, mientras que el modelo con sólo X_2 explica el $70,97$ %.

- Criterio de Información de Akaike (AIC):

$$AIC_1 = 2,83 \quad \text{(modelo con } X_2 \text{ solamente)} > AIC_2 = 0,8920 \quad \text{(modelo con } X_2 \text{ y } X_3\text{)}$$

  El AIC es una medida de la calidad del modelo relativo a otros modelos. Un menor AIC indica un mejor ajuste del modelo. En este caso, el modelo con X_2 y X_3 tiene un AIC menor (0,8920) comparado con el modelo que incluye sólo X_2 (2,83), indicando que el modelo con ambas variables explicativas X_2 y X_3 es superior en términos de ajuste.

Para realizar el apartado c), hemos usado estan funciones:

$$R^2 = \frac{\frac{1}{n} \cdot \sum_{i=1}^{n} \left(\hat{Y}_i - \bar{Y}\right)^2}{\frac{1}{n} \cdot \sum_{i=1}^{n} \left(Y_i - \bar{Y}\right)^2} = \frac{\sum_{i=1}^{n} \left(\hat{Y}_i - \bar{Y}\right)^2}{\sum_{i=1}^{n} \left(Y_i - \bar{Y}\right)^2} = \frac{SCE}{SCT} \tag{3}$$

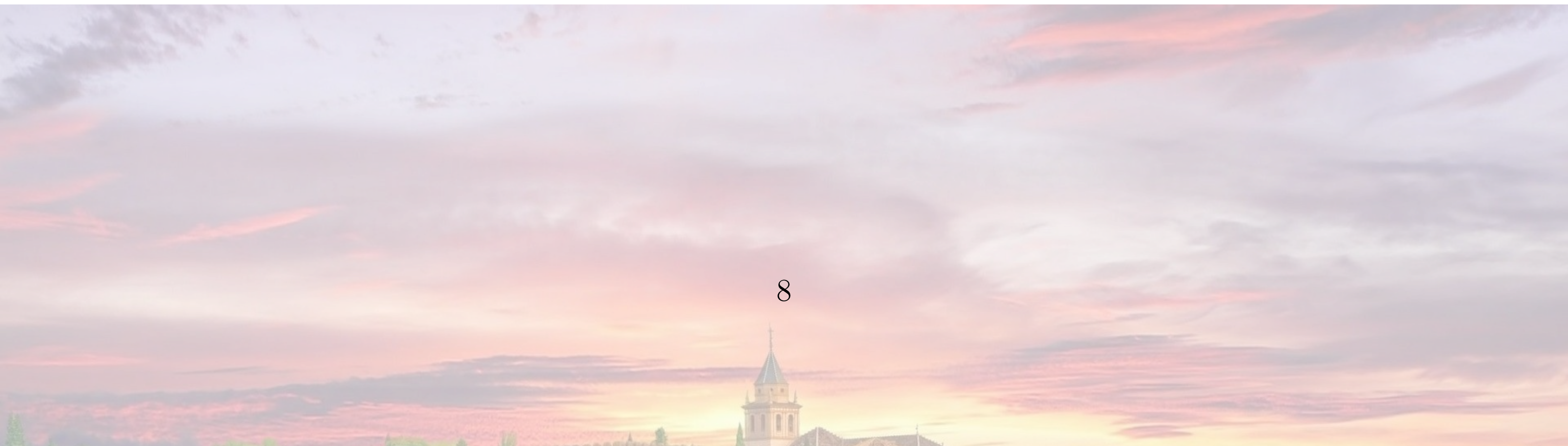
$$SCE = Adj(\hat{\beta}) \cdot X'Y - 10 \cdot \left(\frac{\sum_{i=1}^{10} Y}{n}\right)^2 \tag{4}$$

Y para el cálculo de AIC:

$$AIC = ln(\frac{SCR}{n}) \cdot \frac{2k}{n} \tag{5}$$

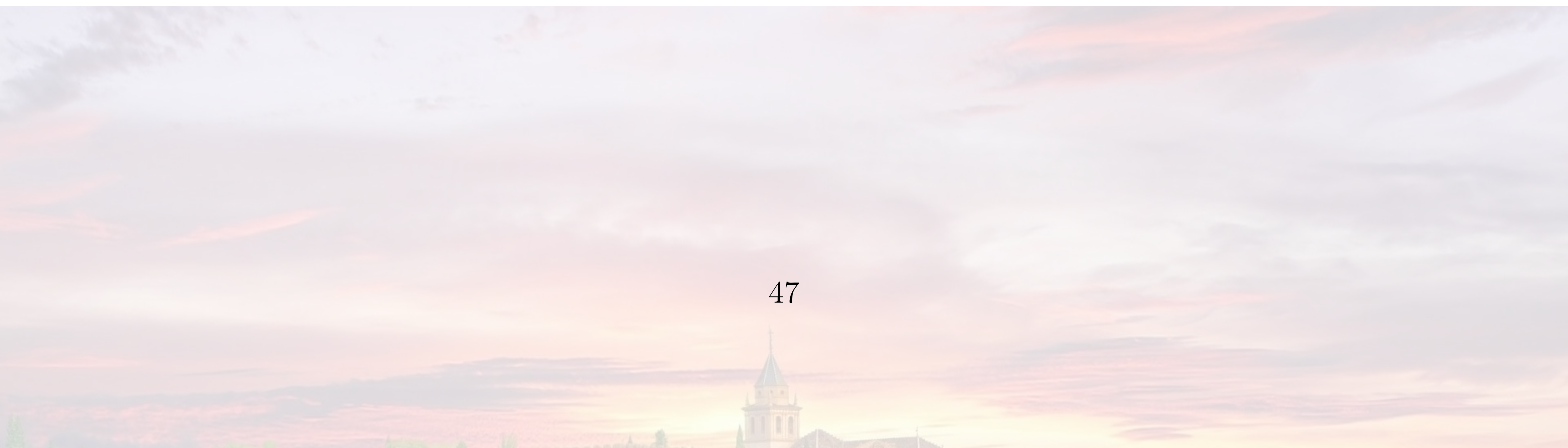
Siendo k el número de variables explicativas y n el número de observaciones. Para encontrar los calculos pincha aquí y visualiza el archivo en formato .xlsx

1.7. Ejercicio 7



3 Ejercicios Resueltos

3.1. Tema 2



ECONOMETRÍA. GADE

Prácticas

Tema 2

Ejercicios resueltos

1. A partir de los siguientes datos se pide representar matricialmente el modelo con todas las observaciones:

Año	Tasa de Desempleo	Inflación	Renta nacional disponible neta per capita (miles de euros)
2006	8,5	3,52	18,61
2007	8,2	2,78	19,403
2008	11,3	4,09	19,492
2009	17,9	-0,28	18,719
2010	19,9	1,8	18,706
2011	21,4	3,2	18,229
2012	24,8	2,44	17,822
2013	26,1	1,42	17,691
2014	24,4	-0,15	18,029
2015	22,1	-0,5	18,828

Resolución:
Inicialmente se puede plantear como:

$$Y_t = \beta_1 + \beta_2 X_{2t} + \beta_3 X_{3t} + u_t$$

donde Y es la tasa de desempleo, X_2 la inflación, X_3 la renta nacional disponible neta a precios de mercado por habitante (miles de euros) y u la perturbación aleatoria. Matricialmente se puede expresar como:

$$\begin{pmatrix} 8,5 \\ 8,2 \\ 11,3 \\ 17,9 \\ 19,9 \\ 21,4 \\ 24,8 \\ 26,1 \\ 24,4 \\ 22,1 \end{pmatrix} = \begin{pmatrix} 1 & 3,52 & 18,61 \\ 1 & 2,78 & 19,403 \\ 1 & 4,09 & 19,492 \\ 1 & -0,28 & 18,719 \\ 1 & 1,8 & 18,706 \\ 1 & 3,2 & 18,229 \\ 1 & 2,44 & 17,822 \\ 1 & 1,42 & 17,691 \\ 1 & -0,15 & 18,029 \\ 1 & -0,5 & 18,828 \end{pmatrix} \begin{pmatrix} \beta_1 \\ \beta_2 \\ \beta_3 \end{pmatrix} + \begin{pmatrix} u_1 \\ u_2 \\ u_3 \\ u_4 \\ u_5 \\ u_6 \\ u_7 \\ u_8 \\ u_9 \\ u_{10} \end{pmatrix}$$

Tambien se puede expresar matricialmente de forma abreviada de la siguiente forma:

$$\vec{y} = X\vec{\beta} + \vec{u}$$

Ejercicio adaptado de [1].

2. Detalla el orden de las matrices y vectores del ejercicio anterior:
   El orden de Y es 10×1, el orden de la matrix X es 10×3, el orden del vector $\vec{\beta}$ es 3×1 y, por último, el orden del vector $\vec{u}$ es 10×1.

3. Razone si las siguientes matrices verifican la condición de rango completo por columnas:

$$A = \begin{pmatrix} 1 & 2 & 3 \\ 1 & 4 & 5 \\ 1 & 5 & 6 \\ 1 & 6 & 7 \\ 1 & 7 & 8 \end{pmatrix}$$

$$B = \begin{pmatrix} 1 & 2 & 3 \\ 1 & 4 & 5 \\ 1 & 5 & 6 \\ 1 & 6 & 7 \\ 1 & 7 & 9 \end{pmatrix}$$

$$C = \begin{pmatrix} 1 & 2 & 5 \\ 1 & 4 & 9 \\ 1 & 5 & 11 \\ 1 & 6 & 13 \\ 1 & 7 & 15 \end{pmatrix}$$

   El rango (orden del mayor determinante no nulo) de las matrices A y C es 2 por lo que es menor que $k = 3$ (número de columnas), por lo que en estos casos no se cumple la condición de rango completo por columnas. En el caso de la matriz B, el rango es igual a 3 que coincide con k por lo que en este caso si se cumple la condición.

4. Dado un modelo con 5 observaciones construye la matriz de varianzas-covarianzas de las perturbaciones aleatorias y detalla que valores tomaría en caso de que se verificasen las hipótesis básicas sobre la perturbación aleatoria.
   Resolución:

$$var - cov(u_i, u_j) = E(\vec{u}\vec{u}') = E\left(\begin{pmatrix} u_1 \\ u_2 \\ u_3 \\ u_4 \\ u_5 \end{pmatrix} \begin{pmatrix} u_1 & u_2 & u_3 & u_4 & u_5 \end{pmatrix} \right) =$$

$$\begin{pmatrix} E(u_1u_1) & E(u_1u_2) & E(u_1u_3) & E(u_1u_4) & E(u_1u_5) \\ E(u_2u_1) & E(u_2u_2) & E(u_2u_3) & E(u_2u_4) & E(u_2u_5) \\ E(u_3u_1) & E(u_3u_2) & E(u_3u_3) & E(u_3u_4) & E(u_3u_5) \\ E(u_4u_1) & E(u_4u_2) & E(u_4u_3) & E(u_4u_4) & E(u_4u_5) \\ E(u_5u_1) & E(u_5u_2) & E(u_5u_3) & E(u_5u_4) & E(u_5u_5) \end{pmatrix} = \begin{pmatrix} \sigma^2 & 0 & 0 & 0 & 0 \\ 0 & \sigma^2 & 0 & 0 & 0 \\ 0 & 0 & \sigma^2 & 0 & 0 \\ 0 & 0 & 0 & \sigma^2 & 0 \\ 0 & 0 & 0 & 0 & \sigma^2 \end{pmatrix} = \sigma^2 I_{(5\times5)}$$

   Si se cumplen las hipótesis básicas la diagonal tomaría un valor constante σ^2 y los valores de fuera de la diagonal tomarían el valor 0. beginenumerate

5. Dada las siguientes matrices se pide obtener $X'X$ y $X'y$:

$$X = \begin{pmatrix} 1 & 2 \\ 1 & 2 \\ 1 & 4 \\ 1 & 3 \\ 1 & 5 \end{pmatrix}$$

$$y = \begin{pmatrix} 3 \\ 4 \\ 4 \\ 3 \\ 6 \end{pmatrix}$$

RESOLUCIÓN:

Se puede obtener las matrices $X'X$ y $X'y$ de la siguiente forma:

$$X'X = \begin{pmatrix} 1 & 1 & 1 & 1 & 1 \\ 2 & 2 & 4 & 3 & 5 \end{pmatrix} \begin{pmatrix} 1 & 2 \\ 1 & 2 \\ 1 & 4 \\ 1 & 3 \\ 1 & 5 \end{pmatrix} = \begin{pmatrix} 5 & 16 \\ 16 & 58 \end{pmatrix}$$

$$X'y = \begin{pmatrix} 20 \\ 69 \end{pmatrix}.$$

6. Calcule a partir de las siguientes matrices, el determinante del producto $X'X$ y comente si el resultado podría provocar algún problema en la estimación:

   a)

$$X = \begin{pmatrix} 1 & 2 & 3 \\ 1 & 4 & 5 \\ 1 & 5 & 6 \\ 1 & 6 & 7 \\ 1 & 7 & 8 \end{pmatrix}$$

   RESOLUCIÓN:
   Al calcular $X'X$ se obtiene:

$$(X'X) = \begin{pmatrix} 5 & 24 & 29 \\ 24 & 130 & 154 \\ 29 & 154 & 183 \end{pmatrix}$$

   El determinante es cero, por lo que no se podrá calcular la inversa de la matrix $X'X$ y, por tanto, no se podrá realizar la estimación.

   b)

$$X = \begin{pmatrix} 1 & 2 & 3 \\ 1 & 4 & 5 \\ 1 & 5 & 6 \\ 1 & 6 & 7 \\ 1 & 7 & 9 \end{pmatrix}$$

Resolución:
Al calcular $X'X$ se obtiene:

$$(X'X) = \begin{pmatrix} 5 & 24 & 30 \\ 24 & 130 & 161 \\ 30 & 161 & 200 \end{pmatrix}$$

El determinante es igual a 35, por lo que se podrá calcular la inversa de la matriz $X'X$ y, por tanto, se podrá realizar la estimación.

c)

$$X = \begin{pmatrix} 1 & 2 & 5 \\ 1 & 4 & 9 \\ 1 & 5 & 11 \\ 1 & 6 & 13 \\ 1 & 7 & 15 \end{pmatrix}$$

Resolución:
Al calcular $X'X$ se obtiene:

$$(X'X) = \begin{pmatrix} 5 & 24 & 53 \\ 24 & 130 & 284 \\ 53 & 284 & 621 \end{pmatrix}$$

El determinante es cero, por lo que no se podrá calcular la inversa de la matrix $X'X$.

7. A partir de los siguientes datos se pide:

Año	Tasa de Desempleo (Y)	Inflación (X)
2006	8,5	3,52
2007	8,2	2,78
2008	11,3	4,09
2009	17,9	-0,28
2010	19,9	1,8
2011	21,4	3,2
2012	24,8	2,44
2013	26,1	1,42
2014	24,4	-0,15
2015	22,1	-0,5

a) Estimar la recta que explique la tasa de desempleo en función de la inflación. Resolución:
Realizando los siguientes sumatorios $\sum_{i=1}^{10} x_t = 18,32$, $\sum_{i=1}^{10} x_t^2 = 58,65$, $\sum_{i=1}^{10} y_t = 184,6$ y $\sum_{i=1}^{10} x_t y_t = 281,08$, se obtiene:

$$X'X = \begin{pmatrix} 10 & 18,32 \\ 18,32 & 58,65 \end{pmatrix}$$

$$X'y = \begin{pmatrix} 184,6 \\ 281,08 \end{pmatrix}$$

Y aplicando la expresión del estimador de mínimos cuadrados ordinarios (MCO) se obtiene:

$$\hat{\beta} = (X'X)^{-1}X'y = \begin{pmatrix} 22,629 \\ -2,276 \end{pmatrix}$$

b) Obtener los residuos y comprobar que su suma es cero.
RESOLUCIÓN:
Sustituyendo en la recta de regresión el valor real de la variable X, se obtiene el valor estimado de y, y realizando la diferencia entre el valor real y el valor estimado se obtienen los residuos:

$\hat{y}_t = 22,629 - 2,276 \cdot X_t$	$e = y - \hat{y}$
$\hat{y}_1 = 22,629 - 2,276 \cdot (3,52) = 14,617$	-6,12
$\hat{y}_2 = 22,629 - 2,276 \cdot (2,78) = 16,3$	-8,10
$\hat{y}_3 = 22,629 - 2,276 \cdot (4,09) = 13,32$	-2,02
$\hat{y}_4 = 22,629 - 2,276 \cdot (-0,28) = 23,26$	-5,37
$\hat{y}_5 = 22,629 - 2,276 \cdot (1,8) = 18,53$	1,37
$\hat{y}_6 = 22,629 - 2,276 \cdot (3,2) = 15,346$	6,05
$\hat{y}_8 = 22,629 - 2,276 \cdot (2,44) = 17,07$	7,72
$\hat{y}_8 = 22,629 - 2,276 \cdot (1,42) = 19,39$	6,70
$\hat{y}_9 = 22,629 - 2,276 \cdot (-0,15) = 22,97$	1,43
$\hat{y}_{10} = 22,629 - 2,276 \cdot (-0,5) = 23,76$	-1,66

c) Obtener el coeficiente de determinación.
RESOLUCIÓN:
Se obtiene el sumatorio $\sum_{i=1}^{10} y_i^2 = 3821,58$ que nos permite obtener la suma de los cuadrados totales (SCT):

$$SCT = 3821,58 - 10 \cdot 18,46^2 = 413,86$$

Igualmente, se obtiene la suma de los cuadrados explicados:

$$SCE = \begin{pmatrix} 22,629 & -2,276 \end{pmatrix} \begin{pmatrix} 184,6 \\ 281,08 \end{pmatrix} - 10 \cdot 18,46^2 = 129,98$$

Y por tanto, se obtiene:

$$R^2 = \frac{SCE}{SCT} = \frac{129,98}{413,86} = 0,3140$$

d) Obtener el coeficiente de determinación corregido.
RESOLUCIÓN:

$$\overline{R}^2 = 1 - (1 - 0,3140) \cdot \frac{10-1}{10-2} = 0,2283.$$

Ejercicio seleccionado de [1].

8. Se tiene información para el periodo 1996-2015 de las ventas del grupo Inditex (V_t), el número de tiendas (T_t), número de países (P_t), número de marcas (M_t), número de empleados (E_t) y una variable ficticia que toma el valor 1 para los años en los que se oferta comercio electrónico (D_1). A partir de la que se han obtenido las siguientes estimaciones en la que la variable explicada son las ventas y cuya $SCT =$ 732,784,211.
MODELO 1: $\hat{V}t = -616,54 + 2,76T_t$; $R^2 = 0,9889$
MODELO 2: $\hat{V}t = -66,37 + 3,05T_t - 26,28P_t$; $R^2 = 0,99$
MODELO 3: $\hat{V}t = 845,49 + 2,82T_t + 19,72P_t - 424,26M_t$; $R^2 = 0,9905$
MODELO 4: $\hat{V}t = 232,66 - 0,19T_t - 40,54P_t + 0,1647E_t$; $R^2 = 0,9967$
MODELO 5: $\hat{V}t = -324,81 + 0,12E_t + 1087,1D_{1t}$; $R^2 = 0,9963$

Se pide:

a) Obtener el coeficiente de determinación corregido para cada modelo y comparar.
RESOLUCIÓN:
A partir de la formula del coeficiente de determinación corregido se obtiene:
Modelo 1: $\bar{R}^2 = 1 - (1 - 0,9883) \cdot \frac{20-1}{20-2} = 0,9883$
Modelo 2: $\bar{R}^2 = 1 - (1 - 0,99) \cdot \frac{20-1}{20-3} = 0,9889$
Modelo 3: $\bar{R}^2 = 1 - (1 - 0,9905) \cdot \frac{20-1}{20-4} = 0,9887$
Modelo 4: $\bar{R}^2 = 1 - (1 - 0,9967) \cdot \frac{20-1}{20-4} = 0,9960$
Modelo 5: $\bar{R}^2 = 1 - (1 - 0,9963) \cdot \frac{20-1}{20-3} = 0,9959$
Por tanto, a la vista de los coeficientes de determinación corregidos el mejor modelo sería el cuarto seguido del quinto.

b) Obtener el criterio de información de Akaike para cada modelo y comparar.
RESOLUCIÓN:
Teniendo en cuenta que $R^2 = \frac{SCE}{SCT}$, y dado que se conoce el valor del R^2 de cada modelo y la SCT, se puede obtener la SCE. Posteriormente, dado que se cumple que $SCT = SCE + SCR$, se puede obtener la suma de los cuadrados de los residuos aplicando $SCR = SCT - SCE$:
Modelo 1: $SCR = 8062143,195$
Modelo 2: $SCR = 7265849,34$
Modelo 3: $SCR = 6960596,88$
Modelo 4: $SCR = 2411720,88$
Modelo 5: $SCR = 2645740,91$
Y, aplicando la formula criterio de información de Akaike se obtiene la siguiente tabla:

	Modelo 1	Modelo 2	Modelo 3	Modelo 4	Modelo 5
k	2	3	4	4	3
AIC	318.90	318.82	319.96	298.76	298.61

Ejercicio seleccionado de [1].

Referencias

[1] García, C.B., Sánchez, J.M. y Salmerón, R. (2017) Econometría básica para la economía y la empresa. Ed. Fleming.

[2] García, J., Jiménez, J.F. y Cerrillo, J.R. Econometría práctica. Edo. Libreria Universitaria de Almería.

[3] Johnston, J. (1984) Métodos de econometría. Ed. Vicens Vives.

[4] Pena, B., Estavillo, J., Galindo, E., Leceta, M. y Zamora, M. (1999). Cien ejercicios de econometría. Ed. Pirámide.

[5] Sánchez, C., López, M.M. y García, T. (2015) Econometría. Ed. Fleming.

[6] Matilla García, m., Pérez Pascual, Pedro y Sanz Carnero, B. (2013) Econometría y predicción. Mc Graw Hill.

3.2. Tema 3

ECONOMETRÍA. GADE

Prácticas

Tema 3

Ejercicios resueltos

1. Para los datos del ejercicio de la curva de Philips se pide:

   a) Estimar la varianza de las perturbaciones y la matriz de varianzas covarianzas de los estimadores:

$$X'X = \begin{pmatrix} 10 & 18,32 \\ 18,32 & 58,65 \end{pmatrix}; X'y = \begin{pmatrix} 184,6 \\ 281,08 \end{pmatrix}$$

$$y'y = 3821,58$$

$$\hat{\beta} = \begin{pmatrix} 22,629 \\ -2,276 \end{pmatrix}$$

RESOLUCIÓN:
Anteriormente se ha calculado el coeficiente de determinación, se sabe que $SCE = 129,98$ y $SCT = 413,86$ por lo que se puede despejar el valor de $SCR = 413,86 - 129,98 = 283,88$.
Puesto que $n = 10$ y $k = 2$, se obtiene:

$$\hat{\sigma}^2 = \frac{283,88}{8} = 35,485$$

Y entonces se calcula la matriz de varianzas-covarianzas como:

$$v\hat{a}r(\hat{\beta}) = \hat{\sigma}^2(X'X)^{-1} = 35,485 \begin{pmatrix} 10 & 18,32 \\ 18,32 & 58,65 \end{pmatrix}^{-1} = \begin{pmatrix} 8,29 & -2,59 \\ -2,59 & 1,414 \end{pmatrix}$$

   b) Contrastar la significación individual para la constante y la variable explicativa.
RESOLUCIÓN:
Para estimar la significación individual de la constante se plantea las siguientes hipótesis:

$$\begin{aligned} H_0 &: \beta_1 = 0 \\ H_1 &: \beta_1 \neq 0 \end{aligned}$$

Siendo el estadístico experimental:

$$t_{exp} = \frac{\hat{\beta}_1}{\sqrt{v\hat{a}r(\hat{\beta}_1)}} = \frac{22,63}{\sqrt{8,29}} = 7,85$$

Puesto que $t_{exp} > t_{n-k,1-\frac{\alpha}{2}} = t_{8,0{,}975} = 2{,}303$, se rechaza la H_0 y, por tanto se dice que la constante es significativa. Para estimar la significación individual de la variable inflación (X) se plantea las siguientes hipótesis:

$$\begin{aligned} H_0 &: \beta_2 = 0 \\ H_1 &: \beta_2 \neq 0 \end{aligned}$$

Siendo el estadístico experimental:

$$t_{exp} = \frac{\hat{\beta}_2}{\sqrt{v\hat{a}r(\hat{\beta}_2)}} = \frac{-2,27}{\sqrt{1,414}} = -1,913$$

Puesto que $|t_{exp}| < t_{n-k,1-\frac{\alpha}{2}} = t_8(0{,}975) = 2{,}303$, no se puede rechazar la H_0 y, por tanto se dice que la variable inflación no es significativa.

c) Obtener la tabla ANOVA y contrastar la significación global del modelo.
Resolución:
A partir de los datos del apartado anterior se puede elaborar la tabla ANOVA:

Fuente de variación	Suma de Cuadrados	Grados de Libertad	Medias
Explicada	$SCE = 129,98$	$2-1$	$129,98$
Residuos	$SCR = 283,88$	$10-2$	$35,485$
Total	$SCT = 413,86$	$10-1$	

A partir de la tabla se obtiene el estadístico experimental:

$$F_{exp} = \frac{129,98/(2-1)}{(283,88)/(10-2)} = 3,66 < F_{1,8,0{,}95} = 5,32$$

No se puede rechazar la H_0, por lo que el modelo globalmente no es válido.
Tengáse en cuenta que, en este caso, puesto que el modelo sólo cuenta con una variable explicativa, la hipótesis nula del contraste de significación global coincide con la hipótesis nula del contraste de significación individual de dicha variable, es decir:

$$\begin{aligned} H_0 &: \beta_2 = 0 \\ H_1 &: \beta_2 \neq 0. \end{aligned}$$

d) Obtener el intervalo de confianza para los estimadores y para la varianza de la perturbación en el ejemplo de la Curva de Philips.
Resolución:
En el caso del intervalo de confianza para los estimadores, se obtiene: Para $n = 10$, $k = 2$ y $\alpha = 0,1$ se obtiene:

$$\beta_1 \in (22,63 \pm 1,860 \cdot \sqrt{8,29}) = (17,27; 27,98)$$

$$\beta_2 \in (-2,276 \pm 1,860 \cdot \sqrt{1,414}) = (-4{,}4877; -0{,}06424)$$

Para $n = 10$, $k = 2$ y $\alpha = 0,05$ se obtiene:

$$\beta_1 \in (22,63 \pm 2,306 \cdot \sqrt{8,29}) = (15,99; 29,27)$$

$$\beta_2 \in (-2,276 \pm 2,306 \cdot \sqrt{1,414}) = (-5,018; 0,4663)$$

Para $n = 10$, $k = 2$ y $\alpha = 0,01$ se obtiene:

$$\beta_1 \in (22,63 \pm 3,355 \cdot \sqrt{8,29}) = (12,97; 32,289)$$

$$\beta_2 \in (-2,276 \pm 3,355 \cdot \sqrt{1,414}) = (-6,2654; 1,7134)$$

En el caso del intervalo de confianza para la varianza de la perturbación, se obtiene al 90, 95 y 99 % respectivamente:

$$\frac{(10-2)35,485}{\chi^2_8(15,507)}; \frac{(10-2)35,485}{\chi^2_8(2,733)} = (18,30; 103,87)$$

$$\frac{(10-2)35,485}{\chi^2_8(17,535)}; \frac{(10-2)35,485}{\chi^2_8(2,180)} = (16,189; 130,22)$$

$$\frac{(10-2)35,485}{\chi^2_8(21,995)}; \frac{(10-2)35,485}{\chi^2_8(1,344)} = (12,906; 211,22)$$

2. Se incluyen en la curva de Philips las variables Renta nacional disponible neta a precios de mercado por habitante (X_2) y renta nacional disponible neta (X_3), obteniendo la siguiente estimación:

$$\begin{array}{lcccc} \hat{y} = & 120,97 & -0,77X_1 & -0,0172X_2 & +0,0002567X_3 \\ & (15,54) & (0,300) & (0,0016) & (3,82 \cdot 10^{-5}) \end{array}$$

Se pide:

a) Contrastar la significación individual de cada variable.

RESOLUCIÓN:

Para estimar la significación individual de cada variable se plantea las siguientes hipótesis:

$$\begin{array}{rcl} H_0 & : & \beta_i = 0 \\ H_1 & : & \beta_i \neq 0 \end{array}$$

Siendo el estadístico experimental:

$$t_{exp} = \frac{\hat{\beta}_i}{\sqrt{\hat{var}(\hat{\beta}_i)}}$$

Por tanto, para β_1:

$$t_{exp} = \frac{120,97}{15,54} = 7,78$$

Para β_2:

$$t_{exp} = \frac{-0,77}{0,30} = -2,564$$

Para β_3:

$$t_{exp} = \frac{-0,0172}{0,0016} = -10,759$$

Para β_4:

$$t_{exp} = \frac{0,00025}{3,82 \cdot 10^{-5}} = 6,71$$

Puesto que en todos los casos $t_{exp} > t_{n-k,1-\frac{\alpha}{2}} = t_{6;0,975} = 2,4469$, se rechaza la H_0 y, por tanto se puede decir que todas las variables son significativas.

b) Sabiendo que R^2 es igual a 0,9735 contrastar la significación global del modelo.
RESOLUCIÓN:
A partir del coeficiente de determinación, se puede realizar el contraste de significación global a partir de la siguiente expresión:

$$F_{exp} = \frac{\frac{R^2}{k-1}}{\frac{1-R^2}{n-k}} = \frac{\frac{0,9735}{4-1}}{\frac{1-0,9735}{10-4}} = 73,49$$

Puesto que $F_{exp} > F_{4-1,10-4}(1-0,05) = 4,75$, se rechaza la hipótesis nula y por tanto se puede decir que el modelo es globalmente significativo.

Nota: Ejercicio resuelto en "Econometría básica para la economía y la empresa", Ed. Fleming.

3. La empresa Miranda quiere saber si la inversión en publicidad expresada en miles de euros (Y) está relacionada con el número de nuevos productos (X_1) y el número de comerciales (X_2), y para ello recoge la siguiente información de los ultimos seis años:

Y	X_1	X_2
51	20	8
36	20	10
32	15	6
56	25	10
45	20	12
48	30	10

Sabiendo que:

$$\hat{\beta} = \begin{pmatrix} 17,25 \\ 1,05 \\ 0,5 \end{pmatrix}$$

Se pide:

a) Estimar el modelo.

RESOLUCIÓN:

A partir de los datos anteriores, se crea la siguiente tabla:

Y	X_1	X_2	X_1^2	X_2^2	$X_1 \cdot X_2$	$X_1 \cdot Y$	$X_2 \cdot Y$	Y^2
51	20	8	400	64	160	1020	408	2601
36	20	10	400	100	200	720	360	1296
32	15	6	225	36	90	480	192	1024
56	25	10	625	100	250	1400	560	3136
45	20	12	400	144	240	900	540	2025
48	30	10	900	100	300	1440	480	2304
268	130	56	2950	544	1240	5960	2540	12386

A partir de dicha tabla se obtiene:

$$X'X = \begin{pmatrix} 6 & 130 & 56 \\ 130 & 2950 & 1240 \\ 56 & 1240 & 544 \end{pmatrix}$$

$$X'y = \begin{pmatrix} 268 \\ 5960 \\ 2540 \end{pmatrix}$$

Y aplicando la expresión de MCO, se obtiene:

$$\hat{\beta} = \begin{pmatrix} 17,25 \\ 1,05 \\ 0,5 \end{pmatrix}$$

b) Obtener un intervalo de predicción para el valor medio de inversión en publicidad suponiendo que el proximo año se saquen 35 nuevos productos y la empresa tenga 10 comerciales.

RESOLUCIÓN:

En primer lugar, se realiza la predicción puntual:

$$p_0 = x_0^t \cdot \widehat{\beta} = \begin{pmatrix} 1 & 35 & 10 \end{pmatrix} \cdot \begin{pmatrix} 17,25 \\ 1,05 \\ 0,5 \end{pmatrix} = 59$$

A continuación se calcula:

$$x_0^t (X'X)^{-1} = \begin{pmatrix} 1 & 35 & 10 \end{pmatrix} \cdot \begin{pmatrix} 5,25 & -0,1 & -0,3125 \\ -0,1 & 0,01 & -0,0125 \\ -0,3125 & -0,0125 & 0,0625 \end{pmatrix}$$

$$= \begin{pmatrix} -1,375 & 0,125 & -0,125 \end{pmatrix}$$

$$x_0^t (X'X)^{-1} x_0 = \begin{pmatrix} -1,375 & 0,125 & -0,125 \end{pmatrix} \begin{pmatrix} 1 \\ 35 \\ 10 \end{pmatrix} = 1,75$$

Se calcula la SCR como:

$$SCR = 12386 - \begin{pmatrix} 17,25 & 1,05 & 0,5 \end{pmatrix} \cdot \begin{pmatrix} 268 \\ 5960 \\ 2540 \end{pmatrix} = 12386 - 12151 = 235$$

Por tanto:

$$\hat{\sigma}^2 = \frac{SCR}{n-k} = \frac{235}{6-3} = 78,33$$

Finalmente, el intervalo sería:

$$59 \pm 3,182\sqrt{78,33}\sqrt{1,75} = (21,74; 96,25)$$

Nota: Ejercicio resuelto en [1].

4. Se tiene información sobre ocho empresas aseguradoras sobre el volumen de polizas (Y), el gasto en publicidad en televisión (X_1) y el gasto en publicidad en internet (X_2) todo expresado en decenas de miles de euros, tal que:

$$X'X = \begin{pmatrix} 8 & 139 & 183 \\ 139 & 2533 & 3285 \\ 183 & 3285 & 4303 \end{pmatrix}$$

$$X'y = \begin{pmatrix} 319 \\ 5894 \\ 7636 \end{pmatrix}$$

$$y`y = 13907$$

Se pide:

a) Contrastar si es posible que el efecto del gasto en publicidad en televisión sea tres veces el efecto del gasto en publicidad en internet.
Resolución:
Si se plantea el modelo como $Y_i = \beta_1 + \beta_2 X_{1i} + \beta_3 X_{3i} + u_i$, la hipótesis a contrastar se define como:

$$\begin{aligned} H_0 &: \beta_2 - 3 \cdot \beta_3 = 0 \\ H_1 &: \beta_2 - 3 \cdot \beta_3 \neq 0 \end{aligned}$$

Entonces se define la siguiente matriz de restricciones:

$$R = \begin{pmatrix} 0 & 1 & -3 \end{pmatrix}$$

y la siguiente matriz de resultados $r = 0$.
A continuación se obtiene:

$$F_{exp} = \left(R\widehat{\beta} - r\right)^t \cdot \frac{\left[R\left(X^tX\right)^{-1}R^t\right]^{-1}}{q \cdot \widehat{\sigma}^2} \cdot \left(R\widehat{\beta} - r\right)$$

Se observa que es necesario estimar previamente el modelo aplicando la expresión de MCO:

$$\hat{\beta} = \begin{pmatrix} 8 & 139 & 183 \\ 139 & 2533 & 3285 \\ 183 & 3285 & 4303 \end{pmatrix}^{-1} \cdot \begin{pmatrix} 319 \\ 5894 \\ 7636 \end{pmatrix} = \begin{pmatrix} -19,93 \\ 2 \\ 1,09 \end{pmatrix}$$

A continuación, se realizan los siguientes calculos:

$$R\widehat{\beta} = \begin{pmatrix} 0 & 1 & -3 \end{pmatrix} \cdot \begin{pmatrix} -19,93 \\ 2 \\ 1,09 \end{pmatrix} = -1,27$$

$$R(X'X)^{-1} = \begin{pmatrix} 1,113 & 0,162 & -0,171 \end{pmatrix}$$

$$R(X'X)^{-1}R' = \begin{pmatrix} 1,113 & 0,162 & -0,171 \end{pmatrix} \begin{pmatrix} 0 \\ 1 \\ -3 \end{pmatrix} = 0,6772$$

Por último, se calcula $\hat{\sigma}^2 = \frac{SCR}{n-k}$:

$$SCR = 13907 - \begin{pmatrix} -19,93 & 2 & 1,09 \end{pmatrix} \cdot \begin{pmatrix} 319 \\ 5894 \\ 7636 \end{pmatrix} = 112,41$$

Y por tanto:

$$\hat{\sigma}^2 = \frac{112,41}{8-3} = 22,48$$

A partir de todos estos calculos, se obtiene el siguiente estadístico experimental:

$$F_{exp} = (-1,27)\frac{(0,6772)^{-1}}{1 \cdot 22,48}(-1,27) = 0,10617$$

Dado que $F_{exp} < F_{1,8-3}(1-0,05) = 6,60$ no se puede rechazar la hipótesis nula y por tanto, se puede decir que el efecto del gasto en publicidad en televisión puede ser el triple que el efecto de gasto en publicidad en internet.

b) Suponiendo que la hipótesis anterior es cierta, obtener los estimadores minimos cuadrados restringidos.
Resolución:
Se aplica la formula de MCO restringidos tal que:

$$\widehat{\beta}_R = \begin{pmatrix} -19,93 \\ 2 \\ 1,09 \end{pmatrix} - \begin{pmatrix} 1,113 \\ 0,162 \\ -0,171 \end{pmatrix} \cdot (0,677)^{-1} \cdot (-1,2714)$$

$$= \begin{pmatrix} -17,85 \\ 2,3088 \\ 0,7696 \end{pmatrix}$$

Nota: Ejercicio resuelto en [1].

Referencias

[1] García, C.B., Sánchez, J.M. y Salmerón, R. (2017) Econometría básica para la economía y la empresa. Ed. Fleming.

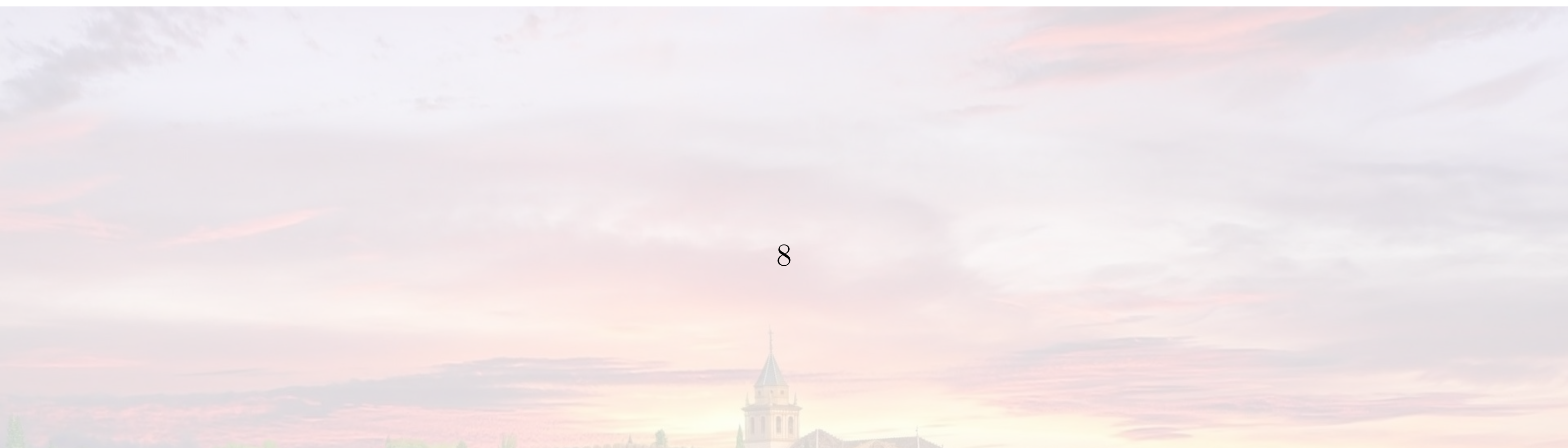
[2] García, J., Jiménez, J.F. y Cerrillo, J.R. Econometría práctica. Edo. Libreria Universitaria de Almería.

[3] Johnston, J. (1984) Métodos de econometría. Ed. Vicens Vives.

[4] Pena, B., Estavillo, J., Galindo, E., Leceta, M. y Zamora, M. (1999). Cien ejercicios de econometría. Ed. Pirámide.

[5] Sánchez, C., López, M.M. y García, T. (2015) Econometría. Ed. Fleming.

[6] Matilla García, m., Pérez Pascual, Pedro y Sanz Carnero, B. (2013) Econometría y predicción. Mc Graw Hill.



3.3. Tema 4

ECONOMETRÍA. GADE

Prácticas

Tema 4

Ejercicios resueltos

1. Proponga un modelo para explicar la recaudación anual del cine (Y) en el periodo 1985-2015, en función del número de entradas vendidas los días de diario (X_1), el número de entradas vendidas los días festivos (X_2) y el número de entradas totales vendidas (X_3). ¿Considera que puede existir colinealidad en este modelo?. ¿De que tipo?
   RESOLUCIÓN:
   El modelo se puede formular como $Y_t = \beta_1 + \beta_2 X_{1t} + \beta_3 X_{2t} + \beta_4 X_{3t} + u_t$. Este modelo tendrá colinealidad perfecta ya que el número de entradas totales vendidas (X_3) se calcula como la suma del número de entradas vendidas los días de diario (X_1) y el número de entradas vendidas los días festivos (X_2).

2. Se tienen datos sobre la evolución de las ventas (Y) del grupo Inditex, el número de tiendas (X_1), número de países (X_2), número de marcas (X_3) y número de empleados (X_4). Se pide:

   a) Analice la posible existencia de colinealidad a partir de la siguiente matriz de correlaciones:

Y	X_1	X_2	X_3	X_4	
1	0,9945	0,9376	0,8253	0,9971	Y
	1	0,9529	0,8500	0,9977	X_1
		1	0,9606	0,9551	X_2
			1	0,8530	X_3
				1	X_4

   RESOLUCIÓN:
   A la vista de los coeficientes de correlación se detecta que:

   - La variable X_1 esta fuertemente relacionada con las variables X_2 y X_4 ya que los coeficientes de correlación son respectivamente 0,9529 y 0,9977. Tambien esta relacionada con la variable X_3 ya que el coeficiente de correlación es igual 0,85.
   - La variable X_2 esta muy relacionada con las variables X_3 y X_4 ya que los coeficientes de correlación son 0,9606 y 0,9551, respectivamente.
   - La variable X_3 esta relacionada con la variable X_4 ya que el coeficiente de correlación es igual a 0,8530.

   Todas las variables independientes estan relacionadas entre sí, por lo que se puede decir que existe problemas de colinealidad.

b) Analice la posible existencia de colinealidad a partir de las siguientes regresiones auxiliares:
RESOLUCIÓN:

$$\begin{array}{lllll} \hat{X}_1 = & 108+ & 4,83X_2 & -43,9X_3+ & 0,0471X_4 \\ & (299) & (15,0) & (129) & (0,00445) \quad R^2 = 0,995 \end{array}$$

$$\begin{array}{lllll} \hat{X}_2 = & -14,8+ & 0,00133X_1 & +8,09X_3+ & 0,000213X_4 \\ & (3,32) & (0,00413) & (0,708) & (0,000202) \quad R^2 = 0,99 \end{array}$$

$$\begin{array}{lllll} \hat{X}_3 = & 2,11 & -0,000164X_1 & +0,110X_2 & -1{,}91 \times 10^{-5} X_4 \\ & (0,240) & (0,000482) & (0,00963) & (2,39 \times 10^{-5}) \quad R^2 = 0,97 \end{array}$$

$$\begin{array}{lllll} \hat{X}_4 = & 2,52 \times 10^3+ & 18,6X_1 & +305X_2 & -2{,}02 \times 10^3 X_3 \\ & (5,93 \times 10^{-5}) & (1,76) & (290) & (2,52 \times 10^3) \quad R^2 = 0,996 \end{array}$$

A la vista de los datos aportados, se puede calcular el FIV como $FIV = \frac{1}{1-R^2}$ para cada una de las variables explicativas obteniéndose los siguientes resultados:

$$FIV(X_1) = \frac{1}{1-0,995} = 200$$

$$FIV(X_2) = \frac{1}{1-0,99} = 100$$

$$FIV(X_3) = \frac{1}{1-0,97} = 33,33$$

$$FIV(X_4) = \frac{1}{1-0,996} = 250$$

Todos los valores son mayores que 10, por lo que se concluye que existe colinealidad en el modelo.

c) Analice la posible existencia de colinealidad sabiendo que el maximo autovalor de la matrix $X'X$ es igual a 4.735 y el minimo es igual a 0.001. RESOLUCIÓN:
Teniendo en cuenta que el número de condición se calcula como:

$$k(X) = \sqrt{\frac{\lambda_{max}}{\lambda_{min}}} = \sqrt{\frac{4{,}735}{0{,}001}} = 68{,}81.$$

se observa que toma un valor superior a 30 y por tanto se concluye que existe colinealidad en el modelo.

3. Calcule a partir de las siguientes matrices, la inversa y el determinante y concluya acerca de la relación entre las variables:

$$X'X = \begin{pmatrix} 1 & 0,95 \\ 0,95 & 1 \end{pmatrix}$$

$$X'X = \begin{pmatrix} 1 & 0,995 \\ 0,995 & 1 \end{pmatrix}$$

Resolución:
A partir de los datos anteriores, se obtiene:

$$(X'X)^{-1} = \begin{pmatrix} 10,26 & -9,74 \\ -9,74 & 10,26 \end{pmatrix}; |X'X| = 0{,}0975$$

$$(X'X)^{-1} = \begin{pmatrix} 100,25 & -99,75 \\ -99,75 & 100,25 \end{pmatrix}; |X'X| = 0,009975$$

4. A partir del fichero de muestra data7-3 de Ramanathan que contiene datos para 14 viviendas sobre el precio de venta de las viviendas (***price***), se pide estimar el precio en funcion de la superficie en pies cuadrados (***sqft***) , el numero de dormitorios (***bedrms***) y el numero de baños (***baths***). Concluye sobre la posible existencia de colinealidad a partir de:

   a) Los coeficientes de correlación muestral. Resolución:

Coeficientes de correlación, usando las observaciones 1 - 14
valor crítico al 5 % (a dos colas) = 0.5324 para n = 14

price	sqft	bedrms	baths	
1,0000	0,9058	0,3156	0,6696	price
	1,0000	0,4647	0,7873	sqft
		1,0000	0,5323	bedrms
			1,0000	baths

   b) El coeficiente de determinación de las regresiones de cada una de las variables explicativas sobre el resto de las variables independientes del modelo. Resolución:

$$\hat{sqft} = -658 + 74{,}0 * bedrms + 975 * baths; R^2 = 0{,}623$$

$$\hat{bedrms} = 2{,}30 + 0{,}000103 * sqft + 0{,}488 * baths; R^2 = 0{,}289$$

$$\hat{baths} = 0{,}647 + 0{,}000532 * sqft + 0{,}191 * bedrms; R^2 = 0{,}655$$

   c) A partir del FIV. Resolución:

$$FIV(sqft) = 2{,}651$$

$$FIV(bedrms) = 1{,}406$$

$$FIV(baths) = 2{,}900$$

d) A partir del NC Se obtiene que el minimo autovalor es igual a 0.008 y el maximo igual a 3.938, por lo tanto el numero de condición es igual:

$$K(X) = \sqrt{\frac{3{,}938}{0{,}008}} = 22{,}422$$

Notese que a partir de los FIV se concluiria que no existe multicolinealidad, mientras que el NC presenta un valor superior a 20 que indicaría multicolinealidad. Si se obtienen los estadisticos principales de las variables se observa que el rango de la variable ***baths*** es muy pequeño (varia entre 3 y 4) presentando un coeficiente de variación muy pequeño.

Estadísticos principales, usando las observaciones 1 - 14

Variable	Media	Mediana	Mínimo	Máximo
price	317.49	291.50	199.90	505.00
sqft	1910.9	1835.0	1065.0	3000.0
bedrms	3.6429	4.0000	3.0000	4.0000
baths	2.3571	2.2500	1.7500	3.0000

Variable	Desv. Típ.	C.V.	Asimetría	Exc. de curtosis
price	88.498	0.27874	0.65346	−0.52983
sqft	577.76	0.30234	0.48526	−0.67212
bedrms	0.49725	0.13650	−0.59628	−1.6444
baths	0.44629	0.18934	0.33161	−1.3902

Variable	porc. 5 %	porc. 95 %	Rango IQ	Observaciones ausentes
price	indefinido	indefinido	154.50	0
sqft	indefinido	indefinido	832.75	0
bedrms	indefinido	indefinido	1.0000	0
baths	indefinido	indefinido	0.81250	0

5. A partir del fichero de muestra data4-7 de Ramanathan se dispone de una base de datos anuales sobre las tasas de mortalidad por enfermedades coronarias y sus determinantes para el periodo de 1947 a 1980 en E.E.U.U. Se pide:

   a) Especifica el modelo y razona sobre el signo esperado en cada coeficiente.

   b) Estima el modelo por MCO. Resolución:

Modelo 1: MCO, usando las observaciones 1947–1980 ($T = 34$)
Variable dependiente: chd

	Coeficiente	Desv. Típica	Estadístico t	valor p
const	226.002	146.830	1.539	0.1363
cal	−69.9825	78.5568	−0.8909	0.3815
unemp	−0.613405	1.58641	−0.3867	0.7023
cig	10.1164	5.07126	1.995	0.0571
edfat	2.80992	1.66827	1.684	0.1046
meat	0.111592	0.243007	0.4592	0.6500
spirits	21.7156	8.45724	2.568	0.0166
beer	−3.46661	1.29774	−2.671	0.0131
wine	−4.56184	16.2472	−0.2808	0.7812

Media de la vble. dep.	354.8147	D.T. de la vble. dep.	14.94605
Suma de cuad. residuos	1980.298	D.T. de la regresión	8.900109
R^2	0.731364	R^2 corregido	0.645401
$F(8, 25)$	8.507846	Valor p (de F)	0.000016
Log-verosimilitud	−117.3428	Criterio de Akaike	252.6857
Criterio de Schwarz	266.4229	Hannan–Quinn	257.3705
$\hat{\rho}$	0.252080	Durbin–Watson	1.333807

c) Interpreta los coeficientes del modelo anterior.

d) Comenta los resultados obtenidos de la estimación en terminos de bondad de ajuste, significatividad y signos de los coeficientes estimados. Razona si te parecen adecuados.

e) Detecta la posible existencia de multicolinealidad a partir del FIV y del NC.

$$FIV(cal) = 5{,}786$$

$$FIV(unemp) = 1{,}967$$

$$FIV(cig) = 11{,}756$$

$$FIV(edfat) = 25{,}570$$

$$FIV(meat) = 6{,}250$$

$$FIV(spirits) = 16{,}703$$

$$FIV(beer) = 36{,}218$$

$$FIV(wine) = 49{,}066$$

Se obtiene que el minimo autovalor es igual a 0,00008 y el maximo igual a 8.644, por lo tanto el numero de condición es igual:

$$K(X) = \sqrt{\frac{8{,}644}{0{,}0008}} = 328{,}84$$

Si centramos las variables originales, el valor de los FIV se mantienen pero el del NC seria:

$$K(X) = \sqrt{\frac{6{,}502}{0{,}012}} = 23{,}63$$

Se habria corregido la multicolinealidad no esencial, pero seguiria existiendo colinealidad esencial.

6. A partir de los datos de [10] actualizados desde Junio de 2008 a abril de 2019 estimar el rendimiento a 52 semanas (c_{52}) en función del rendimiento a 13 y 26 semanas ((c_{13} y c_{26}, respectivamente). Analizar la posible existencia de multicolinealidad y estimar mediante el metodo ortogonal. Este ejemplo es desarrollado completamente en [9]. Se especifica el siguiente modelo:

$$\mathbf{c}_{52} = \beta_1 + \beta_2\ \mathbf{c}_{13} + \beta_3\ \mathbf{c}_{26} + \mathbf{u}, \tag{1}$$

Obteniendose la siguiente matriz de correlación:

$$\begin{array}{c} \\ \mathbf{c}_{52} \\ \mathbf{c}_{13} \\ \mathbf{c}_{26} \end{array} \begin{array}{c} \begin{array}{ccc} \mathbf{c}_{52} & \mathbf{c}_{13} & \mathbf{c}_{26} \end{array} \\ \begin{pmatrix} 1{,}000 & & \\ 0{,}981 & 1{,}000 & \\ 0{,}995 & 0{,}993 & 1{,}000 \end{pmatrix} \end{array}.$$

Al estimarlo por MCO se obtiene:

MCO, usando las observaciones ($T = 131$)
Variable dependiente: r52

	Coeficiente	Desv. Típica	Estadístico t	valor p
const	0.0504730	0.00853647	5.913	0.0000
r13	−0.556909	0.0654995	−8.502	0.0000
r26	1.56218	0.0625815	24.96	0.0000

Media de la vble. dep.	0.696565	D.T. de la vble. dep.	0.781464
Suma de cuad. residuos	0.515394	D.T. de la regresión	0.063455
R^2	0.993508	R^2 corregido	0.993407
$F(2, 128)$	9794.297	Valor p (de F)	9.8e–141
Log-verosimilitud	176.8594	Criterio de Akaike	−347.7189
Criterio de Schwarz	−339.0933	Hannan–Quinn	−344.2139
$\hat{\rho}$	0.352018	Durbin–Watson	1.279550

Notese que el signo del parámetro β_2 no es coherente con la matriz de correlaciones. Si se calcula el FIV se obtiene un valor de 71.516 y para el NC se obtiene un valor igual a 23.233, que indicaría la existencia de multicolinealidad. En relación a la multicolinealidad no esencial se concluye que no existe ya que los coeficientes de variación de las variables no son bajos: ($\text{CV}(\mathbf{c}_{13}) = 1{,}486997$ y $\text{CV}(\mathbf{c}_{26}) = 1{,}280229$). Si se aplica el método de estimación ortogonal, en primer lugar, hay que estimar la siguiente regresión auxiliar:

$$\mathbf{c}_{26} = \alpha_1 + \alpha_2\ \mathbf{c}_{13} + \mathbf{v}, \tag{2}$$

y, posteriormente, guardar dichos residuos y usarlos como variable explicativa sustityendo a la variable c_{26}, de manera que se estima el siguiente modelo:

MCO, usando las observaciones ($T = 131$)
Variable dependiente: r52

	Coeficiente	Desv. Típica	Estadístico t	valor p
const	0.183112	0.00668113	27.41	0.0000
r13	1.06664	0.00774526	137.7	0.0000
e	1.56218	0.0625815	24.96	0.0000

Media de la vble. dep.	0.696565	D.T. de la vble. dep.	0.781464
Suma de cuad. residuos	0.515394	D.T. de la regresión	0.063455
R^2	0.993508	R^2 corregido	0.993407
$F(2, 128)$	9794.297	Valor p (de F)	9.8e–141
Log-verosimilitud	176.8594	Criterio de Akaike	−347.7189
Criterio de Schwarz	−339.0933	Hannan–Quinn	−344.2139
$\hat{\rho}$	0.352018	Durbin–Watson	1.279550

Notese que en este modelo se ha corregido el simbolo negativo del estimador del parámetro de la variable c_{13}. Si se calcula el FIV se obtendrá un valor igual a 1 (el minimo posible ya que ahora las variables explicativas son ortogonales) y el NC tendrá un valor de 1.878.

Referencias

[1] García, C.B., Sánchez, J.M. y Salmerón, R. (2017) Econometría básica para la economía y la empresa. Ed. Fleming.

[2] García, J., Jiménez, J.F. y Cerrillo, J.R. Econometría práctica. Edo. Libreria Universitaria de Almería.

[3] Johnston, J. (1984) Métodos de econometría. Ed. Vicens Vives.

[4] Pena, B., Estavillo, J., Galindo, E., Leceta, M. y Zamora, M. (1999). Cien ejercicios de econometría. Ed. Pirámide.

[5] Sánchez, C., López, M.M. y García, T. (2015) Econometría. Ed. Fleming.

[6] Matilla García, m., Pérez Pascual, Pedro y Sanz Carnero, B. (2013) Econometría y predicción. Mc Graw Hill.

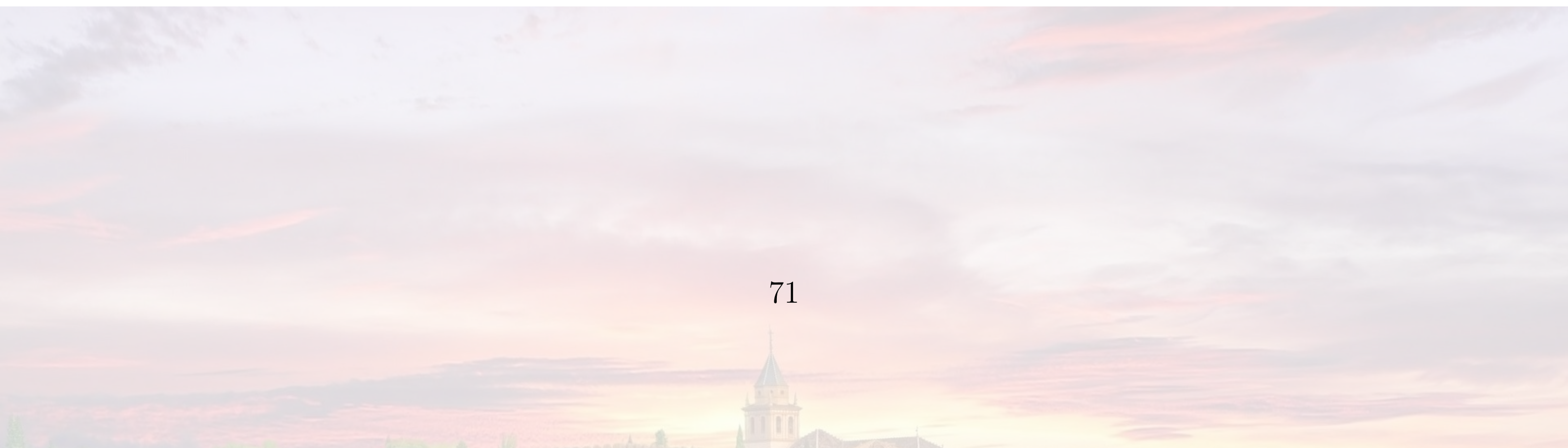
[7] Olva Maldonado, H. (2009). *Análisis de la función de producción Cobb-Douglas y su aplicación en el sector productivo Mexicano.* Tesis profesional. Universidad Autónoma de Chapingo, México. Dirección web: http://bit.ly/1pFtQnnhttp://bit.ly/1pFtQnn.

[8] Wikipedia: Función de producción de Cobb-Douglas. Dirección web: http://bit.ly/1gfKeFPhttp://bit.ly/1gfKeFP.

[9] García, C. B., Salmerón, R., García, C., García, J. (2019). *Residualization: justification, properties and application.* Journal of Applied Statistics, 1-21.

[10] Wooldridge, J. (2008). Introducción a la econometría. Un enfoque moderno, 2nd ed., Thomson Paraninfo, Madrid.

3.4. Tema 5



ECONOMETRÍA. GADE

Prácticas

Tema 5

Ejercicios resueltos

1. Utilizando una muestra de 25 observaciones anuales se estima el siguiente modelo de demanda:

$$D_t = \beta_1 + \beta_2 Y_t + \beta_3 PR_t + u_t$$

Utilizando sólo las 10 primeras observaciones se obtiene la siguiente ecuación estimada [1]:

$$\begin{array}{ccccccc} \widehat{D}_t & = & 80{,}50 & + & 0{,}93 Y_t & - & 0{,}87 PR_t \\ & & (86{,}17) & & (1{,}06) & & (1{,}9) \end{array}$$

donde se ha obtenido $SCR_1 = 125{,}7$. Del mismo modo, y utilizando las 10 últimas observaciones, se obtiene la siguiente ecuación estimada

$$\begin{array}{ccccccc} \widehat{D}_t & = & 20{,}61 & + & 0{,}53 Y_t & - & 0{,}105 PR_t \\ & & (221{,}44) & & (0{,}29) & & (2{,}41) \end{array}$$

con $SCR_2 = 498{,}94$. Detectar mediante el test de Goldfeld y Quandt la existencia o no de heteroscedasticidad.

Procediendo de la forma indicada se tiene que la hipótesis nula del contraste es

$$H_0 : \sigma_1^2 = \sigma_2^2 = \sigma^2$$

En esta situación el estadístico de contraste es:

$$F_{exp} = \frac{SCR_2}{SCR_1} = \frac{498{,}94}{125{,}7} = 3{,}9693$$

y se verifica que $F_{exp} \sim F_{\frac{n-c}{2}-k, \frac{n-c}{2}-k} = F_{\frac{25-5}{2}-3, \frac{25-5}{2}-3} = F_{7,7}$, donde se ha tenido en cuenta que el número de observaciones que han sido omitidas son $c = 5$ y el número de parámetros a estimar $k = 3$.

Como $F_{exp} > 3{,}79$ entonces se rechaza la hipótesis nula a un nivel de significación del 5 %. Por tanto, concluimos que existe heteroscedasticidad.

2. A partir de los siguientes datos se pide que se realice el contraste de Breusch Pagan. Para ello, se realizan los siguientes pasos:

[1] Los valores que aparecen entre paréntesis corresponden a las desviaciones típicas estimadas de cada uno de los parámetros que describen el modelo.

y_i	x_i	$\widehat{y}_i$	e_i	g_i	$\widehat{g}_i$
55	80	60.31	-5.31	0.3582	0.0625
65	100	73.07	-8.07	0.8275	0.2638
70	85	63.5	6.5	0.5368	0.1128
80	110	79.45	0.55	0.0038	0.3644
79	120	85.82	-6.82	0.5910	0.4650
84	115	82.64	1.36	0.0235	0.4147
98	130	92.2	5.8	0.4274	0.5656
95	140	98.58	-3.58	0.1628	0.6663
90	125	89.01	0.99	0.0125	0.5153
75	90	66.69	8.31	0.8774	0.1631
74	105	76.26	-2.26	0.0649	0.3141
110	160	111.34	-1.34	0.0228	0.8675
113	150	104.96	8.04	0.8213	0.7669
125	165	114.52	10.48	1.3955	0.9178
108	145	101.77	6.23	0.4931	0.7166
115	180	124.09	-9.09	1.0498	1.0688
140	225	152.79	-12.79	2.0784	1.5216
120	200	136.85	-16.85	3.6074	1.2700
145	240	162.36	-17.36	3.8291	1.6725
130	185	127.28	2.72	0.0940	1.1191
152	220	149.6	2.4	0.0732	1.4713
144	210	143.23	0.77	0.0075	1.3706
175	245	165.55	9.45	1.1346	1.7228
180	260	175.11	4.89	0.3038	1.8738
135	190	130.47	4.53	0.2607	1.1694
140	205	140.04	-0.04	0.0000	1.3203
178	265	178.3	-0.3	0.0011	1.9241
191	270	181.49	9.51	1.1491	1.9744
137	230	155.98	-18.98	4.5771	1.5719
189	250	168.74	20.26	5.2153	1.7731

- Se ha estimado el modelo de regresión $\overrightarrow{y} = X\overrightarrow{\beta} + \overrightarrow{u}$ obteniendo, una vez aplicado el estimador mínimo cuadrático, $\overrightarrow{\widehat{y}} = 9{,}29031 + 0{,}637785X$. Y se han obtenido los residuos realizando la operación $e_i = y_i - \overrightarrow{\widehat{y}}_i$.
- A partir de dicha información, se tiene que la Suma de los Cuadrados de los residuos generados es 2361.153. Luego, de forma inmediata se tendría que $\tilde{\sigma}^2 = \frac{SCR}{n} = \frac{2361{,}153}{30} = 78{,}7051$.
- Dividiendo los residuos al cuadrado entre $\tilde{\sigma}^2 = 78{,}7051$ se obtienen los residuos cuadráticos medios, g_i, a partir de los cuales se obtiene la estimación del modelo $g_i = \overrightarrow{z}_i^t\overrightarrow{\alpha} + u_i$, donde se considera que la variable que es la causante de la heteroscedasticidad es X_i. La estimación obtenida es $\widehat{g}_i = -0{,}742509 + 0{,}0100626X_i$.
- El modelo estimado nos permite calcular la SCE como $\sum_{i=1}^{30}(\widehat{g}_i - \overline{g})^2 = 10{,}4267$.
- Así pues, el estadístico de contraste tomaría el valor

$$\chi^2_{exp} = \frac{SCE}{2} = \frac{10{,}4267}{2} = 5{,}21335$$

  Considerando un nivel de significación del 5 %, se tiene que como $\chi^2_{exp} > \chi^2_{k-1,1-\alpha} = \chi^2_{1,0{,}95} = 3{,}8414$ se rechaza la hipótesis nula. Así pues, no se puede rechazar el supuesto de heteroscedasticidad.

3. A partir de los datos del ejemplo anterior, realizar el contraste de Glesjer sabiendo que:

y_i	55	65	70	80	79	84	...	178	191	137	189
X_i	80	100	85	110	120	115	...	265	270	230	250
$\|e_i\|$	5.31	8.07	6.5	0.55	6.82	1.36	...	0.3	9.51	18.98	20.26

Se realizan los siguientes pasos:

- Para $h = \pm 1, \pm 2, \pm \frac{1}{2}$ se estima el modelo de regresión en el cual la variable endógena es el valor absoluto de los errores de estimación, $|e_t|$, y la variable exógena es la variable que pensamos que puede ser la causante del problema de heteroscedasticidad, X_i.

$$
\begin{array}{llll}
h = 1 & |\widehat{e}_t| = 1.1126 + 0.0331 \cdot z_t & & R^2 = 0.1188 \\
 & \qquad (3.6348) \quad (0.01988) & & \\
h = -1 & |\widehat{e}_t| = 11.0619\text{-}639.379 \cdot z_t^{-1} & & R^2 = 0.0814 \\
 & \qquad (1.9876) \quad (282.044) & & \\
h = 2 & |\widehat{e}_t| = 3.6566{+}9.56{\cdot}10^{-5} \cdot z_t^2 & & R^2 = \mathbf{0.1233} \\
 & \qquad (1.3159) \quad (3.34{\cdot}10^{-5}) & & \\
h = -2 & |\widehat{e}_t| = 8.5926\text{-}35036.3 \cdot z_t^{-2} & & R^2 = 0.0575 \\
 & \qquad (1.1709) \quad (18621.3) & & \\
h = 1/2 & |\widehat{e}_t| = \text{-}3.8895{+}0.8289 \cdot z_t^{1/2} & & R^2 = 0.1126 \\
 & \qquad (4.0212) \quad (0.3056) & & \\
h = -1/2 & |\widehat{e}_t| = 16.0441\text{-}115.249 \cdot z_t^{-1/2} & & R^2 = 0.0932 \\
 & \qquad (3.8293) \quad (47.1948) & &
\end{array}
$$

- Utilizando el estadístico de contraste $|t_{exp}| = |\frac{\widehat{\delta}_1}{\sqrt{\text{var}[\widehat{\delta}_1]}}|$ (recordemos que los valores que aparecen entre paréntesis son las desviaciones típicas estimadas de cada uno de los parámetros del modelo) se estudia la significatividad del parámetro δ_1 de la regresión $|\overrightarrow{e}_t| = \delta_0 + \delta_1 z_t^h + v_t$. Si se rechaza la hipótesis nula entonces estamos ante un problema de heteroscedasticidad.

$$
\begin{array}{lll}
h = 1 & |t_{exp}| = 1{,}6649 \not> t_{28,0{,}975} = 2{,}048 & \text{Se mantiene } H_0 \\
h = -1 & |t_{exp}| = 2{,}2669 > t_{28,0{,}975} = 2{,}048 & \text{Se rechaza } H_0 \\
h = 2 & |t_{exp}| = 2{,}8713 > t_{28,0{,}975} = 2{,}048 & \text{Se rechaza } H_0 \\
h = -2 & |t_{exp}| = 1{,}8815 \not> t_{28,0{,}975} = 2{,}048 & \text{Se mantiene } H_0 \\
h = 1/2 & |t_{exp}| = 2{,}7124 > t_{28,0{,}975} = 2{,}048 & \text{Se rechaza } H_0 \\
h = -1/2 & |t_{exp}| = 2{,}4419 > t_{28,0{,}975} = 2{,}048 & \text{Se rechaza } H_0
\end{array}
$$

  Luego, con las regresiones auxiliares para $h = -1; 2; 1/2; -1/2$ concluimos que existe heteroscedasticidad, siendo el modelo $|\widehat{e}_t| = 3{,}6566 + 9{,}56 \cdot 10^{-5} \cdot z_t^2$ el mejor ya que es el que presenta el mayor coeficiente de determinación. En esta situación, podemos pensar que $E[u_t^2] = \sigma^2 z_t^2$.

  En los casos primero y cuarto se concluye que no existe heteroscedasticidad ya que en ambos se mantiene la hipótesis nula.

4. Supongamos que deseamos analizar los dividendos de una empresa en función de sus beneficicios disponemos de las 20 observaciones de la siguiente tabla: Tras realizar la estimación por MCO se obtiene que:

$$
\widehat{D}_t = 10'2229 + 0'0638456 \cdot B_t
$$
$$
(1\text{'}35823) \qquad (0\text{'}011223)
$$

con $R^2 = 0$'642 y donde entre paréntesis se especifica la desviación típica estimada de cada coeficiente estimado. Tras realizar la estimación por MCO del modelo original (ejemplo anterior)

Empresas	Dividendos	Beneficios
1	13'2	61
2	15	78
3	22'2	158
4	15'2	110
5	16'1	85
6	18'5	150
7	15'5	140
8	15	70
9	20	122
10	15	70
11	21	140
12	16'2	91
13	18'5	105
14	17	115
15	17'5	115
16	22	160
17	18	165
18	23	170
19	17	130
20	17	90

hemos obtenido los residuos a partir de los cuales hemos ajustado las siguientes regresiones auxiliares:

$$|\widehat{e}_t| = \underset{(0'3191)}{2'15499} - \underset{(2703'4)}{8633'51} \cdot B_t^{-2} \qquad R^2 = 0'3617$$

$$|\widehat{e}_t| = \underset{(0'5479)}{3'1728} - \underset{(55'1884)}{197'755} \cdot B_t^{-1} \qquad R^2 = 0'4163$$

$$|\widehat{e}_t| = \underset{(1'0576)}{5'212} - \underset{(10'8788)}{40'6989} \cdot B_t^{-1/2} \qquad R^2 = 0'4374$$

$$|\widehat{e}_t| = \underset{(1'09025)}{-2'93457} + \underset{(0'1011)}{0'39747} \cdot B_t^{1/2} \qquad R^2 = 0'4619$$

$$|\widehat{e}_t| = \underset{(0'5783)}{-0'891826} + \underset{(0'004778)}{0'01888} \cdot B_t \qquad R^2 = 0'4646$$

$$|\widehat{e}_t| = \underset{(0'3424)}{0'138792} + \underset{(0'0000205)}{0'000079} \cdot B_t^2 \qquad R^2 = 0'4536$$

En todos los casos rechazamos H_0, por lo que hay heteroscedasticidad. Además, como el mejor modelo es el quinto (mayor coeficiente de determinación), pensamos que $E[u_t^2] = \sigma^2 B_t$.
Gracias al test de Glesjer hemos supuesto que $E[u_t^2] = \sigma^2 \cdot B_t$, es decir, la perturbación aleatoria depende directamente de los beneficios. Por tanto, lo que en teoría se denotaba como w_t corresponde a la variable beneficios, B_t. Luego para transformar los datos habrá que dividir por la raíz cuadrada de dicha variable:

$$D_t^* = \frac{D_t}{\sqrt{B_t}}, \quad cte_t^* = \frac{1}{\sqrt{B_t}}, \quad B_t^* = \frac{B_t}{\sqrt{B_t}} = \sqrt{B_t},$$

con $t = 1, \ldots, 20$.

A partir de los datos transformados se obtiene la siguiente estimación por MCO del modelo transformado: $\widehat{D}_t^* = \underset{(1'1196)}{10'2147} \cdot cte_t^* + \underset{(0'011)}{0'0639} \cdot B_t^*$ con $R^2 = 0'9931$ (adviértase que los valores numéricos de las nuevas estimaciones no difieren mucho de las originales).

Nota: Se facilita estimación hecha en excel en Prado.

Empresa	B_t^*	D_t^*	cte_t^*
1	7'81025	1'690087	0'1280369
2	8'83176	1'698416	0'1132277
3	12'56981	1'766137	0'07956
4	10'48809	1'449263	0'09535
5	9'21954	1'746290	0'1084652
6	12'24745	1'510519	0'08165
7	11'83216	1'309989	0'08452
8	8'36660	1'792843	0'1195229
9	11'04536	1'810715	0'09054
10	8'36660	1'792843	0'1195229
11	11'83216	1'774824	0'08452
12	9'53939	1'698221	0'1048285
13	10'24695	1'805415	0'09759
14	10'72381	1'585258	0'09325
15	10'72381	1'631883	0'09325
16	12'64911	1'739253	0'07906
17	12'84523	1'401298	0'07785
18	13'03840	1'764019	0'07670
19	11'40175	1'490999	0'08771
20	9'48683	1'791957	0'1054093

Referencias

[1] García, C.B., Sánchez, J.M. y Salmerón, R. (2017) Econometría básica para la economía y la empresa. Ed. Fleming.

3.5. Tema 6

ECONOMETRÍA. GADE

Prácticas

Tema 6

Ejercicios resueltos

1. Se tiene el siguiente modelo sobre datos de una empresa aseguradora durante los últimos 38 años, donde y son las pólizas contratadas, x es el gasto en publicidad y z el número de comerciales.

$$\begin{array}{cccc} \hat{y} = & 0,989+ & 0,534x_t+ & 0,183z_t \\ & (1,387) & (0,147) & (0,136) \end{array}$$

$$R^2 = 0,537$$

Contraste la existencia de autocorrelacion en las perturbaciones en el modelo, sabiendo que $e_t = 0,327e_{t-1}$. (Sol. $DW = 1,346$). *Ejercicio seleccionado de [1].*
A partir de la relación $e_t = 0,327e_{t-1}$ se sabe que el coeficiente de coeficiente de correlación es 0,327, por lo que se puede estimar el valor de D-W como:

$$DW \simeq 2(1-\rho) = 2(1-0{,}327) = 1{,}346$$

Dado que $n = 38$ y $k* = 2$, se obtiene de las tablas estadísticas de Durbin-Watson que $d_L = 1{,}373$ y $d_U = 1{,}594$, por lo que el valor del estadístico DW se situa en la zona en la que se rechaza la hipótesis nula de no existencia de autocorrelación concluyendo con la existencia de autocorrelación positiva.

2. A partir de los siguientes datos contrastar la posible existencia de autocorrelación en el siguiente modelo estimado por MCO:

$$\begin{array}{cccccccc} \widehat{Y_t} & = & 1'3 & + & 0'97 \cdot Y_{t-1} & + & 2'31 \cdot X_t & n=21 \qquad d=1'21 \\ & & (0'8) & & (0'07) & & (1'2) & \end{array}$$

A partir de la h de Durbin:

$$h = 0'394 \cdot \sqrt{\frac{21}{1-21\cdot 0'0049}} = 0'395 \cdot 4'83826 = 1'911113,$$

donde $\rho \simeq 1 - \frac{d}{2} = 1 - \frac{1'21}{2} = 0'395$ y $var = 0'07^2 = 0'0049$, se tiene que no se rechaza la hipótesis nula de incorrelación ya que $|h| \not> 1'96$. Luego no existe autocorrelación en el modelo.

3. El número de pequeños accidentes, Y, ocurridos en las calles de una ciudad y el número de coches que han sido matriculados, X, en la misma durante 10 años están recogidos en la Tabla 1. Dado el modelo $Y_t = \beta_1 + \beta_2 X_t + u_t$, se pide:

   a) Contraste la hipótesis de no autocorrelación por medio de Durbin-Watson.

b) En caso de detectar problemas de autocorrelación, obtenga una estimación aplicando MCG.

Cuadro 1: Datos iniciales

Y	X
25	510
27	520
28	528
32	540
33	590
36	650
38	700
40	760
41	800
45	870

Primero estimamos el modelo $Y_t = \beta_1 + \beta_2 X_t + u_t$ empleando el método de mínimos cuadrados ordinarios.

$$\widehat{Y}_t = 2{,}56755 + 0{,}0493699 X_t$$

con $R^2 = 0{,}942095$.

Cuadro 2: Estimación y residuos

Y	$\widehat{Y}$	e_t	e_{t-1}	$(e_t - e_{t-1})^2$	e_t^2
25	27.7462	-2.7462			7.5416
27	28.2399	-1.2399	-2.7462	2.2689	1.5374
28	28.6349	-0.6349	-1.2399	0.3661	0.4030
32	29.2273	2.7727	-0.6349	11.6115	7.6879
33	31.6958	1.3042	2.7727	2.1565	1.7010
36	34.6580	1.3420	1.3042	0.0014	1.8010
38	37.1265	0.8735	1.3420	0.2195	0.7630
40	40.0887	-0.0887	0.8735	0.9258	0.0079
41	42.0635	-1.0635	-0.0887	0.9502	1.1310
45	45.5194	-0.5194	-1.0635	0.2961	0.2697
				18.7960	22.8435

Utilizando la información de la Tabla 3, el estadístico de Durbin-Watson es

$$DW = \frac{\sum_{t=2}^{n}(e_t - e_{t-1})^2}{\sum_{t=1}^{n} e_t^2} = \frac{18{,}7960}{22{,}8435} = 0{,}8228 \tag{1}$$

Empleando la tabla de Durbin-Watson se tiene que como $k^* = 2 - 1 = 1$ y $n = 10$, los límites inferior y superior, al 5 % de significación, son $d_L = 0{,}879$ y $d_U = 1{,}320$. Como $DW < d_L$ entonces se concluye a un nivel de significación del 5 % que existe autocorrelación positiva.

Una vez obtenida la estimación del modelo por MCO, se estima la regresión $e_t = \rho e_{t-1} + \varepsilon_t$ se puede estimar:

$$\hat{\rho} = \frac{\sum e_t \cdot e_{t.-1}}{\sum e_{t-1}^2}$$

Empleando los residuos de la Tabla 3 se tiene

$$\widehat{e}_t = 0{,}4226 e_{t-1}$$

luego $\widehat{\rho} = 0{,}4226$. Como el tamaño de la muestra es pequeño utilizamos la Modificación Prais-Winsten realizando la transformación comentada anteriormente sobre las variables en estudio

$$Y^* = \begin{cases} \sqrt{1-(0{,}4226)^2}\,Y_t & t = 1 \\ Y_t - 0{,}4226 Y_{t-1} & t > 1 \end{cases}$$
$$Cte_t^* = \begin{cases} \sqrt{1-(0{,}4226)^2}\,Cte_t & t = 1 \\ Cte_t - 0{,}4226 Cte_{t-1} & t > 1 \end{cases}$$
$$X_t^* = \begin{cases} \sqrt{1-(0{,}4226)^2}\,X_t & t = 1 \\ X_t - 0{,}4226 X_{t-1} & t > 1 \end{cases} \tag{2}$$

Cuadro 3: Estimación y residuos

Y^*	cte^*	X^*
22.6587	0.9063	462.2198
16.4348	0.5773	304.4708
16.5896	0.5773	308.2447
20.1670	0.5773	316.8638
19.4765	0.5773	361.7926
22.0539	0.5773	400.6622
22.7861	0.5773	425.3059
23.9409	0.5773	464.1756
24.0957	0.5773	478.8192
27.6731	0.5773	531.9148

A partir de los nuevos datos se obtiene que la estimación del modelo $Y_t^* = \beta_3 + \beta_4 X_t^* + u_t^*$ es $\widehat{Y}^* = 2{,}1071 + 0{,}04975 X^*$.

NOTA: Como el valor de ρ ha sido estimado deberíamos realizar el proceso de forma iterativa hasta que converjan los valores de los coeficientes de autocorrelación estimados.

Otra alternativa es calcular de nuevo, con el modelo estimado utilizando las variables transformadas, el estadístico de Durbin-Watson y comprobar si se ha eliminado el problema de autocorrelación. Si optamos por esta opción se tiene que $DW = 1{,}5226$, indicando pues que la autocorrelación ha desaparecido.

4. A partir de los siguientes datos para las ventas en millones de euros (Y) y el numero de anuncios en television (X) para el periodo 2000-2008, se pide:

Y	X
2	16
3	16
3	18
3,2	20
4	19
4	25
5	26
5,5	29
6	30

   a) Estimar el modelo. (Sol. $\hat{\beta}_1 = -1,04; \hat{\beta}_2 = 0,2266$)
   b) Contrastar autocorrelación (Sol. $DW = 3,02$)
   c) Estimar el modelo para corregir la autocorrelación. (Sol. $\hat{\beta*}_1 = -1,0601; \hat{\beta*}_2 = 0,2280$)
   d) Comprobar que se ha resuelto la autocorrelación. (Sol. $DW = 2,4671$)

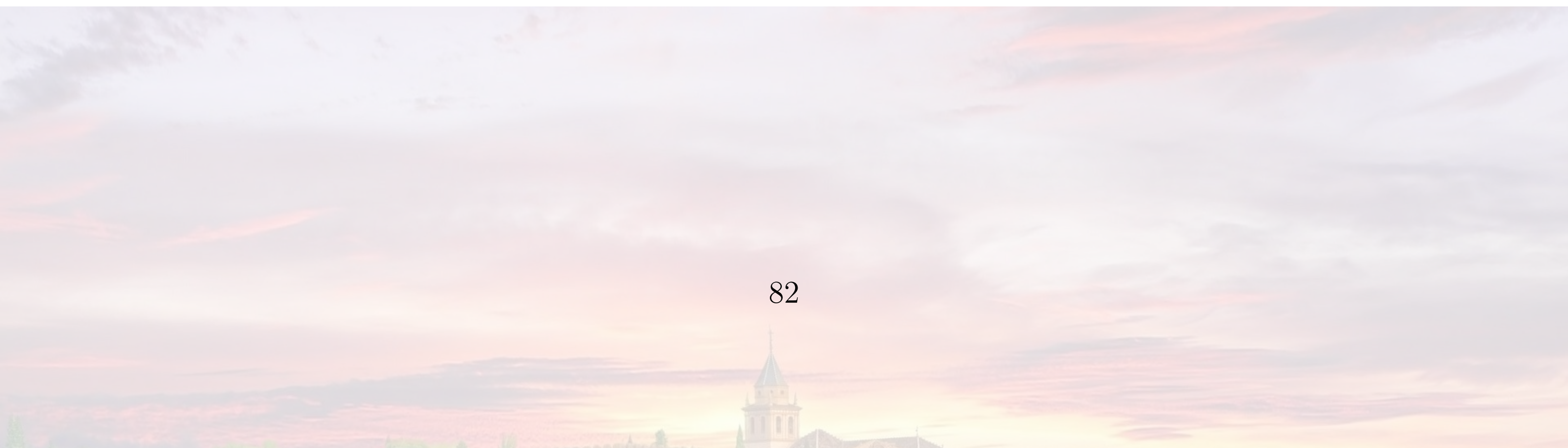
   Ejercicio seleccionado de [1]. Resuelto en Excel.

Referencias

[1] García, C.B., Sánchez, J.M. y Salmerón, R. (2017) Econometría básica para la economía y la empresa. Ed. Fleming.

[2] Pena, B., Estavillo, J., Galindo, E., Leceta, M. y Zamora, M. (1999). Cien ejercicios de econometría. Ed. Pirámide.

3.6. Tablas DW



Estadístico de Durbin-Watson - Puntos críticos de d_L y d_u al nivel de significación del 5 %
k^* corresponde al número de regresores del modelo excluido el término independiente (es decir, $k^* = k - 1$)

	$k^* = 1$		$k^* = 2$		$k^* = 3$		$k^* = 4$		$k^* = 5$		$k^* = 6$	
n	d_L	d_u	d_L	d_u	d_L	d_u	d_L	d_u	d_L	d_u	d_L	d_u
6	0.610	1.400										
7	0.700	1.356	0.467	1.896								
8	0.763	1.332	0.559	1.777	0.368	2.287						
9	0.824	1.320	0.629	1.699	0.455	2.128	0.296	2.588				
10	0.879	1.320	0.697	1.641	0.525	2.016	0.376	2.414	0.243	2.822		
11	0.927	1.324	0.658	1.604	0.595	1.928	0.444	2.283	0.316	2.645	0.203	3.005
12	0.971	1.331	0.812	1.579	0.658	1.864	0.512	2.177	0.379	2.506	0.268	2.832
13	1.010	1.340	0.861	1.562	0.715	1.816	0.574	2.094	0.445	2.390	0.328	2.692
14	1.045	1.350	0.905	1.551	0.767	1.779	0.632	2.030	0.505	2.296	0.389	2.572
15	1.077	1.361	0.946	1.543	0.814	1.750	0.685	1.977	0.562	2.220	0.447	2.472
16	1.106	1.371	0.982	1.539	0.857	1.728	0.734	1.935	0.615	2.157	0.502	2.388
17	1.133	1.381	1.015	1.536	0.897	1.710	0.779	1.900	0.664	2.104	0.554	2.318
18	1.158	1.391	1.046	1.535	0.933	1.696	0.820	1.872	0.710	2.060	0.603	2.257
19	1.180	1.401	1.074	1.536	0.967	1.685	0.859	1.848	0.752	2.023	0.649	2.206
20	1.201	1.411	1.100	1.537	0.998	1.676	0.894	1.828	0.792	1.991	0.692	2.162
21	1.221	1.420	1.125	1.538	1.026	1.669	0.927	1.812	0.829	1.964	0.732	2.124
22	1.239	1.429	1.147	1.541	1.053	1.664	0.958	1.797	0.863	1.940	0.769	2.090
23	1.257	1.437	1.168	1.543	1.078	1.660	0.986	1.785	0.895	1.920	0.804	2.061
24	1.273	1.446	1.188	1.546	1.101	1.656	1.013	1.775	0.925	1.902	0.837	2.035
25	1.288	1.454	1.206	1.550	1.123	1.654	1.038	1.767	0.953	1.886	0.868	2.012
26	1.302	1.461	1.224	1.553	1.143	1.652	1.062	1.759	0.979	1.873	0.897	1.992
27	1.316	1.469	1.240	1.556	1.162	1.651	1.084	1.753	1.004	1.861	0.925	1.974
28	1.328	1.476	1.255	1.560	1.181	1.650	1.104	1.747	1.028	1.850	0.951	1.958
29	1.341	1.483	1.270	1.563	1.198	1.650	1.124	1.743	1.050	1.841	0.975	1.944
30	1.352	1.489	1.284	1.567	1.214	1.650	1.143	1.739	1.071	1.833	0.998	1.931
31	1.363	1.496	1.297	1.570	1.229	1.650	1.160	1.735	1.090	1.825	1.020	1.920
32	1.373	1.502	1.309	1.574	1.244	1.650	1.177	1.732	1.109	1.819	1.041	1.909
33	1.383	1.508	1.321	1.577	1.258	1.651	1.193	1.730	1.127	1.813	1.061	1.900
34	1.393	1.514	1.333	1.580	1.271	1.652	1.208	1.728	1.144	1.808	1.080	1.891
35	1.402	1.519	1.343	1.584	1.283	1.653	1.222	1.726	1.160	1.803	1.097	1.884
36	1.411	1.525	1.354	1.587	1.295	1.654	1.236	1.724	1.175	1.799	1.114	1.877
37	1.419	1.530	1.364	1.590	1.307	1.655	1.249	1.723	1.190	1.795	1.131	1.870
38	1.427	1.535	1.373	1.594	1.318	1.656	1.261	1.722	1.204	1.792	1.146	1.864
39	1.435	1.540	1.382	1.597	1.328	1.658	1.273	1.722	1.218	1.789	1.161	1.859
40	1.442	1.544	1.391	1.600	1.338	1.659	1.285	1.721	1.230	1.786	1.175	1.854
45	1.475	1.566	1.430	1.615	1.383	1.666	1.336	1.720	1.287	1.776	1.238	1.835
50	1.503	1.585	1.462	1.628	1.421	1.674	1.378	1.721	1.335	1.771	1.291	1.822
55	1.528	1.601	1.490	1.641	1.452	1.681	1.414	1.724	1.374	1.768	1.334	1.814
60	1.549	1.616	1.514	1.652	1.480	1.689	1.444	1.727	1.408	1.767	1.372	1.808
65	1.567	1.629	1.536	1.662	1.503	1.696	1.471	1.731	1.438	1.767	1.404	1.805
70	1.583	1.641	1.554	1.672	1.525	1.703	1.494	1.735	1.464	1.768	1.433	1.802
75	1.598	1.652	1.571	1.680	1.543	1.709	1.515	1.739	1.487	1.770	1.458	1.801
80	1.611	1.662	1.586	1.688	1.560	1.715	1.534	1.743	1.507	1.772	1.480	1.801
85	1.624	1.671	1.600	1.696	1.575	1.721	1.550	1.747	1.525	1.774	1.500	1.801
90	1.635	1.679	1.612	1.703	1.589	1.726	1.566	1.751	1.542	1.776	1.518	1.801
95	1.645	1.687	1.623	1.709	1.602	1.732	1.579	1.755	1.557	1.778	1.535	1.802
100	1.654	1.694	1.634	1.715	1.613	1.736	1.592	1.758	1.571	1.780	1.550	1.803
150	1.720	1.746	1.706	1.760	1.693	1.774	1.679	1.788	1.665	1.802	1.651	1.817
200	1.758	1.778	1.748	1.789	1.738	1.799	1.728	1.810	1.718	1.820	1.707	1.831

4 Ejemplo Datos Inditex

5 Modelo Econométrico

5.1. Análisis_de_Multicolinealidad

Análisis de Multicolinealidad

José Ángel Carretero Montes
Ismael Sallami Moreno
Fernando José Gracia Choin

Noviembre 2024

1 Análisis de la Multicolinealidad

1.1 Obtención de las principales medidas de diagnóstico de la multicolinealidad

El análisis de multicolinealidad se llevó a cabo utilizando dos indicadores principales:

- **Número de Condición (Cond. No.)**: Calculado como $\sqrt{\lambda_{\max}/\lambda_{\min}}$, donde $\lambda_{\max}$ y $\lambda_{\min}$ representan los valores propios máximo y mínimo respectivamente de la matriz de covarianzas de las variables independientes. En este caso, el número de condición obtenido fue 28.49, lo cual es inferior al umbral crítico de 30. Por lo tanto, se considera que los niveles generales de multicolinealidad son aceptables.

- **Factor de Inflación de la Varianza (FIV)**: Calculado para cada variable como $FIV = \frac{1}{1-R_i^2}$, donde R_i^2 es el coeficiente de determinación de la regresión de la variable i sobre todas las demás variables. Los valores obtenidos fueron los siguientes:

Variable	**FIV**
Gender	760.64
Age	1.89
Height	1.93
family_history_with_overweight	2.18
FAVC	1.33
FCVC	1.16
NCP	1.15
CAEC	1.11
SMOKE	1.19
CH2O	1.04
SCC	1.12
FAF	1.10
TUE	1.22
CALC	1.13
MTRANS	1.67
NObeyesdad	1.33

Table 1: Valores de FIV para cada variable

1.2 Clasificación de la multicolinealidad existente

A partir del análisis del FIV, se pueden clasificar las variables de la siguiente manera:

- **Baja multicolinealidad**: Las variables con $FIV < 5$, como la mayoría de las incluidas en el análisis, presentan baja multicolinealidad y no representan un problema para el modelo.

- **Alta multicolinealidad**: La variable `Gender` muestra un valor de FIV extremadamente alto (760.64), indicando una fuerte dependencia lineal con otras variables. Este nivel de multicolinealidad es crítico y debe ser abordado.

1.2.1 Decisión de no eliminar la variable `Gender` a pesar de su alto índice de FIV

La variable `Gender` fue identificada como un factor crítico de multicolinealidad en el modelo, con un Factor de Inflación de la Varianza (FIV) extremadamente alto de 760.64. Este valor excede significativamente el umbral comúnmente aceptado de 10, lo que indica una fuerte dependencia lineal entre `Gender` y otras variables independientes del modelo. La inclusión de esta variable, por tanto, genera múltiples problemas de multicolinealidad que detallaremos más adelante.

Sin embargo, al evaluar el impacto de eliminar `Gender` en el desempeño general del modelo, observamos una reducción significativa en el coeficiente de determinación (R^2). Dado que R^2 mide la proporción de la variabilidad explicada por el modelo, esta disminución sugiere que `Gender` aporta información valiosa para las predicciones. Por lo tanto, decidimos mantener esta variable en el modelo, priorizando su capacidad explicativa a pesar de los problemas de multicolinealidad.

- Aunque la eliminación de `Gender` reduce el **Número de Condición** (Cond. No.) a niveles aceptables y mejora los valores de **FIV** de las demás variables, la pérdida en R^2 comprometería la calidad general del modelo.
- Se consideran alternativas como el uso de técnicas de regularización (ridge o lasso) para mitigar los efectos de la multicolinealidad sin eliminar la variable.

En conclusión, mantener la variable `Gender` permite preservar la capacidad explicativa del modelo, aunque conlleva un compromiso en términos de multicolinealidad. Este balance entre interpretabilidad y desempeño predictivo es fundamental para abordar los objetivos del análisis.

1.2.2 Matriz de correlación

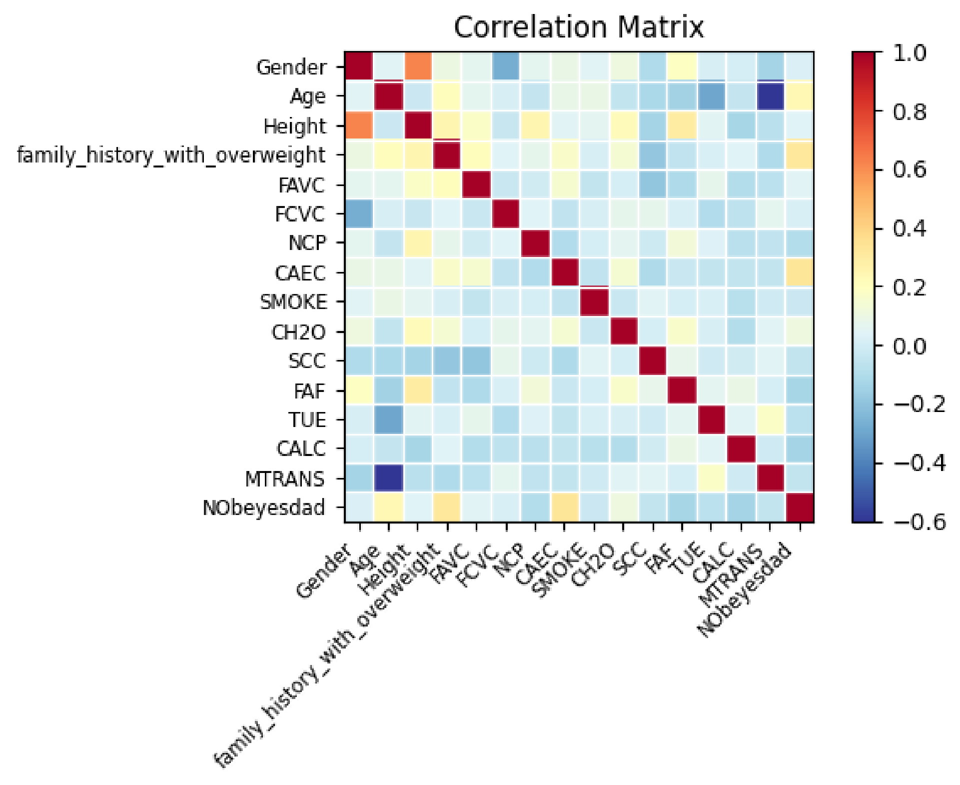


Figure 1: Matriz de correlación

Interpretación de la Matriz de Correlación

En la matriz de correlación observamos que la variable `Height` presenta un valor de correlación relativamente alto con respecto otras variables del modelo. Aunque este nivel de correlación podría sugerir la posibilidad de eliminar esta variable para mejorar la multicolinealidad, al llevar a cabo este procedimiento se observó una disminución significativa en el coeficiente de determinación ajustado (R^2). Esta reducción implica una pérdida de capacidad explicativa del modelo, lo que contrarrestaría los posibles beneficios de eliminar `Height`. Por lo tanto, se decidió mantener esta variable dentro del modelo para preservar su capacidad predictiva y explicativa.

1.3 Análisis de las posibles consecuencias derivadas de la existencia de multicolinealidad

La presencia de alta multicolinealidad puede tener las siguientes consecuencias en el modelo:

- **Inestabilidad de los coeficientes de regresión**: Los coeficientes estimados para las variables multicolineales pueden presentar grandes variaciones ante pequeños cambios en los datos, reduciendo la fiabilidad del modelo.
- **Incremento de la varianza**: Los altos valores de FIV implican un aumento en la varianza de los coeficientes estimados, lo que puede llevar a errores estándar más grandes y a una disminución en la significancia estadística de las variables.
- **Redundancia**: Las variables altamente correlacionadas aportan información redundante, lo que puede complicar la interpretación del modelo.

1.4 Mitigación de la multicolinealidad

Para mitigar los efectos de la multicolinealidad, se pueden aplicar las siguientes medidas:

- **Eliminación de variables redundantes**: Dado el alto FIV de la variable `Gender`, se podría evaluar la posibilidad de excluirla del modelo si su aportación al análisis no es crítica.
- **Transformaciones**: Aplicar transformaciones lineales como Análisis de Componentes Principales (PCA) para reducir la dimensionalidad y eliminar la correlación entre las variables.
- **Recolección de nuevos datos**: Ampliar el tamaño de la muestra puede reducir los efectos de la multicolinealidad, especialmente si las variables redundantes se distribuyen de forma más heterogénea en los nuevos datos.
- **Regularización**: Métodos como la regresión ridge o lasso pueden ser útiles para manejar la multicolinealidad al penalizar los coeficientes altamente correlacionados.

5.2. Modelo_Preliminar

Estimación del Modelo Preliminar

José Ángel Carretero Montes
Ismael Sallami Moreno
Fernando José Gracia Choin

Octubre 2024

1 Estimación del Modelo Preliminar

El modelo de regresión lineal múltiple estimado se representa de la siguiente forma:

$$Y_i = \beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + \cdots + \beta_k X_{ki} + u_i$$

donde:

- Y_i es la variable dependiente (Peso en este caso),
- $X_{1i}, X_{2i}, \ldots, X_{ki}$ son las variables independientes (Altura, Edad, Género, etc.),
- β_0 es el término constante,
- u_i es el término de error aleatorio.

Los resultados preliminares de la regresión OLS para este modelo se presentan en la siguiente tabla:

```
                            OLS Regression Results
==============================================================================
Dep. Variable:                 Weight   R-squared:                       0.576
Model:                            OLS   Adj. R-squared:                  0.572
Method:                 Least Squares   F-statistic:                     177.5
Date:                Tue, 22 Oct 2024   Prob (F-statistic):               0.00
Time:                        07:09:09   Log-Likelihood:                -8983.6
No. Observations:                2111   AIC:                         1.800e+04
Df Residuals:                    2094   BIC:                         1.810e+04
Df Model:                          16
Covariance Type:            nonrobust
====================================================================================================
                                      coef    std err          t      P>|t|      [0.025      0.975]
----------------------------------------------------------------------------------------------------
const                            -207.9919     10.281    -20.230      0.000    -228.155    -187.829
Gender                             -4.7575      1.026     -4.639      0.000      -6.769      -2.746
Age                                 0.7403      0.082      9.069      0.000       0.580       0.900
Height                            125.2342      5.894     21.247      0.000     113.675     136.793
family_history_with_overweight     17.0101      1.115     15.259      0.000      14.824      19.196
FAVC                                7.9764      1.252      6.373      0.000       5.522      10.431
FCVC                                9.0480      0.750     12.070      0.000       7.578      10.518
NCP                                 1.0882      0.506      2.151      0.032       0.096       2.080
CAEC                                8.2303      0.869      9.470      0.000       6.526       9.935
SMOKE                              -0.2063      2.664     -0.077      0.938      -5.430       5.017
CH2O                                1.3156      0.645      2.041      0.041       0.052       2.580
SCC                                -6.9609      1.878     -3.706      0.000     -10.645      -3.277
FAF                                -2.7221      0.484     -5.619      0.000      -3.672      -1.772
TUE                                -1.6760      0.652     -2.571      0.010      -2.954      -0.398
CALC                               -4.1794      0.764     -5.472      0.000      -5.677      -2.682
MTRANS                              3.6899      0.382      9.658      0.000       2.941       4.439
NObeyesdad                          2.3519      0.220     10.680      0.000       1.920       2.784
==============================================================================
Omnibus:                       37.671   Durbin-Watson:                   0.741
Prob(Omnibus):                  0.000   Jarque-Bera (JB):               21.704
Skew:                          -0.047   Prob(JB):                     1.94e-05
Kurtosis:                       2.512   Cond. No.                         812.
==============================================================================

Notes:
[1] Standard Errors assume that the covariance matrix of the errors is correctly specified.
```

2 Interpretación de los Estimadores

A continuación, se ofrece una interpretación de los coeficientes más relevantes del modelo:

- **Constante**: El coeficiente de la constante es -207.99, lo que indica el valor estimado del peso cuando todas las variables independientes son 0. Aunque en la práctica, esta interpretación rara vez es útil, ya que variables como la altura no pueden ser cero.
- **Género**: El coeficiente de Género es -4.757. Esto indica que, manteniendo todas las demás variables constantes, ser mujer (asumiendo que el género está codificado como 1 para mujeres y 0 para hombres) está asociado con una reducción de 4.76 unidades en el peso.
- **Edad**: El coeficiente de 0.7403 indica que por cada año adicional de edad, el peso aumenta en promedio 0.74 unidades, manteniendo las demás variables constantes.
- **Altura**: El coeficiente de altura es 125.23, lo que sugiere que, por cada unidad adicional de altura, el peso aumenta en 125.23 unidades, manteniendo todas las demás variables constantes.
- **Historial Familiar de Sobrepeso**: El coeficiente es 17.01, indicando que tener antecedentes familiares de sobrepeso está asociado con un aumento promedio de 17.01 unidades en el peso.
- **FAVC (Comida grasa frecuente)**: El coeficiente positivo de 7.976 indica que consumir frecuentemente comida grasa aumenta el peso en 7.976 unidades, manteniendo todo lo demás constante.
- **FCVC (Consumo de verduras)**: El coeficiente de 9.048 indica que un mayor consumo de verduras está asociado con un aumento de 9.05 unidades en el peso, lo cual es inesperado y podría requerir mayor análisis.
- **NCP (Número de comidas diarias)**: El coeficiente de 1.088 sugiere que, por cada comida adicional consumida al día, el peso aumenta en promedio 1.088 unidades.
- **CAEC (Comer entre comidas)**: El coeficiente de 8.230 sugiere que comer entre comidas aumenta el peso en 8.23 unidades, manteniendo el resto de factores constantes.
- **Fumar (SMOKE)**: El coeficiente es -0.206, indicando que fumar no tiene un impacto estadísticamente significativo en el peso, dado su alto valor p ($P > 0.05$).
- **Consumo de agua (CH2O)**: El coeficiente de 1.315 indica que por cada unidad adicional de agua consumida, el peso aumenta en 1.315 unidades, lo cual puede ser contraintuitivo y merece un análisis más profundo.
- **SCC (Conteo de calorías)**: El coeficiente de -6.961 sugiere que el control de calorías está asociado con una reducción del peso en 6.96 unidades, lo cual es consistente con lo esperado.
- **FAF (Actividad física)**: El coeficiente de -2.722 indica que un aumento en la actividad física reduce el peso en promedio 2.72 unidades, manteniendo constantes las demás variables.
- **TUE (Uso de tecnología)**: El coeficiente de -1.676 sugiere que más tiempo usando tecnología está asociado con una reducción del peso, lo cual puede requerir más análisis.
- **CALC (Consumo de alcohol)**: El coeficiente de -4.179 indica que consumir alcohol está asociado con una reducción del peso, lo cual es un resultado inesperado.
- **MTRANS (Medios de transporte)**: El coeficiente de 3.689 indica que el uso de ciertos medios de transporte se asocia con un aumento del peso en 3.689 unidades.
- **NObeyesdad (Nivel de obesidad)**: El coeficiente de 2.351 indica que un mayor nivel de obesidad previo está asociado con un aumento de 2.351 unidades en el peso.

3 Medidas de Bondad de Ajuste

Para evaluar la calidad del modelo estimado, utilizamos diversas medidas de bondad de ajuste que nos permiten juzgar cómo de bien el modelo ajusta a los datos observados. Las medidas que se consideran son el coeficiente de determinación R^2, el coeficiente de determinación ajustado R^2_{ajustado}, y los criterios de información de Akaike (AIC) y Schwarz (BIC).

3.1 Coeficiente de Determinación R^2

El coeficiente de determinación R^2 mide la proporción de la varianza total de la variable dependiente que es explicada por el modelo. En otras palabras, indica qué tan bien las variables independientes explican las variaciones en la variable dependiente.

$$R^2 = 1 - \frac{\text{Suma de Cuadrados de los Residuos (SCR)}}{\text{Suma Total de Cuadrados (SCT)}}$$

Donde:

- La **Suma Total de Cuadrados (SCT)** mide la variabilidad total de la variable dependiente en torno a su media:
$$\text{SCT} = \sum_{i=1}^{n} (Y_i - \bar{Y})^2$$
- La **Suma de Cuadrados de los Residuos (SCR)** mide la variabilidad que no ha sido explicada por el modelo:
$$\text{SCR} = \sum_{i=1}^{n} (Y_i - \hat{Y}_i)^2$$

Para este modelo, se ha obtenido un R^2 de **0.576**, lo que significa que el **57.6%** de la variabilidad del peso puede explicarse por las variables independientes incluidas en el modelo. Aunque este valor indica que el modelo captura más de la mitad de la variación en el peso, hay un **42.4%** de la variabilidad que no está siendo explicada, lo que sugiere que existen otros factores que no se han tenido en cuenta o que pueden requerir un tratamiento más sofisticado.

3.2 Coeficiente de Determinación Ajustado $R^2_{\textbf{ajustado}}$

El coeficiente de determinación ajustado R^2_{ajustado} ajusta el R^2 para el número de predictores en el modelo. Esto es importante porque el R^2 tiende a aumentar al agregar más variables, incluso si esas variables no tienen una relación significativa con la variable dependiente. El R^2_{ajustado} penaliza el uso de variables adicionales, lo que permite obtener una mejor medida del poder explicativo real del modelo.

$$R^2_{\text{ajustado}} = 1 - \frac{\frac{\text{SCR}}{n-k}}{\frac{\text{SCT}}{n-1}}$$

Donde:

- n es el número de observaciones,
- k es el número de variables explicativas incluidas en el modelo.

El valor obtenido para el R^2_{ajustado} es de **0.572**, lo cual es muy cercano al valor de R^2. Esto indica que el modelo no está sobreajustado (es decir, no ha añadido variables innecesarias), ya que el R^2_{ajustado} no disminuye considerablemente respecto al R^2 original.

3.3 Criterio de Información de Akaike (AIC)

El criterio de información de Akaike (AIC) es una medida que penaliza la complejidad del modelo. Es útil cuando se comparan varios modelos, ya que tiene en cuenta tanto la bondad del ajuste (medida por la suma de cuadrados de los residuos) como el número de parámetros estimados. Cuanto menor sea el valor del AIC, mejor es el modelo.

$$AIC = \ln\left(\frac{\text{SCR}}{n}\right) + \frac{2k}{n}$$

Donde:

- n es el número de observaciones,
- k es el número de parámetros estimados (incluyendo la constante).

En este modelo, el valor de **AIC** es $\mathbf{1.800 \times 10^4}$. Para comparar diferentes modelos, se elegiría aquel que minimice este valor. Un menor AIC indica que el modelo no solo ajusta mejor los datos, sino que también evita la inclusión innecesaria de parámetros.

3.4 Criterio de Información de Schwarz (BIC)

El criterio de información bayesiano de Schwarz (BIC) es otra medida que, como el AIC, penaliza la complejidad del modelo, pero de manera más estricta. El BIC introduce una penalización más fuerte por la inclusión de más variables en el modelo, lo que lo hace útil para evitar el sobreajuste.

$$BIC = \ln\left(\frac{\text{SCR}}{n}\right) + \frac{k}{n} \cdot \ln(n)$$

En este modelo, el valor del **BIC** es $\mathbf{1.810} \times \mathbf{10}^4$. Al igual que con el AIC, un menor valor de BIC es preferible, y su objetivo es identificar el modelo que ofrece el mejor ajuste sin sobrecargarlo con demasiadas variables.

3.5 Comparación y Selección del Mejor Modelo

Las medidas de bondad de ajuste nos indican diferentes aspectos de la calidad del modelo. En resumen:

- Un R^2 de 0.576 indica que el modelo es razonablemente bueno para explicar la variabilidad del peso.
- El R^2_{ajustado} de 0.572 sugiere que no hay un sobreajuste significativo.
- Los criterios de información AIC y BIC nos permiten comparar este modelo con otros posibles modelos. Si generamos otros modelos, escogeremos aquel con los valores más bajos de AIC y BIC.

En función de estas medidas, este modelo tiene un ajuste aceptable, pero se puede mejorar probando otros modelos que optimicen las medidas de AIC y BIC.

4 Conclusiones y Selección del Modelo

El modelo preliminar parece explicar bien la variabilidad del peso, con un R^2 razonable y estimadores significativos. Para mejorar el modelo, podemos considerar variables adicionales o la eliminación de aquellas que no resulten significativas en otros ensayos.

Además, el uso de AIC y BIC será clave para la selección del mejor modelo en función de los diferentes criterios de ajuste que hemos estudiado.

6 Referencias

- Diapositivas de clase.