RECOPILACION DE

EXAMENES DE

INTRODUCCION A LA

ECONOMETRIA

25 de septiembre de 2006

[0] CVS Id: $Id: e1por.tex,v 1.3 2003/09/18 15:55:11 etpdihei Exp

Autores:

Aurora Alonso	M. Paz Moral
Ignacio Díaz-Emparanza	Iñaki Murillo
M. Victoria Esteban	Ainhoa Oguiza
Ana I. Fernández	Susan Orbe
F. Javier Fernández	Marta Regúlez
Inmaculada Gallastegui	M. Paz Rojo
Beatriz Goitisolo	Jorge Virto
Petr Mariel	Ainhoa Zarraga
Juan I. Modroño	Marian Zubia

Índice

LE-2001.1 (Feb-2001)

Se cree que existe una relación lineal entre el Paro Registrado (P) y las variables Exportaciones (E) y Tasa de crecimiento salarial (T) en la economía vasca, disponiéndose para las variables de información anual desde 1981 hasta 1999 que se resume en los siguientes datos:

$$\begin{array}{lll} \sum_{t=1}^{19} P_t = 2526,99 & \sum_{t=1}^{19} E_t = 150,01 & \sum_{t=1}^{19} T_t = 132,26 \\ \sum_{t=1}^{19} P_t^2 = 348471,62 & \sum_{t=1}^{19} E_t^2 = 1529,31 & \sum_{t=1}^{19} T_t^2 = 1096,59 \\ \sum_{t=1}^{19} E_t T_t = 833,61 & \sum_{t=1}^{19} E_t P_t = 18712,38 & \sum_{t=1}^{19} T_t P_t = 17901,32 \end{array}$$

$$(X'X)^{-1} = \begin{bmatrix} 3,200 & -0,1767 & -0,25169 \\ -0,1767 & 0,01087 & 0,01305 \\ -0,25169 & 0,01305 & 0,02135 \end{bmatrix}$$

Apartado I: Se propone el siguiente modelo

$$P_t = \beta_0 + \beta_1 E_t + \beta_2 T_t + u_t \tag{1}$$

a) Estima los coeficientes del modelo utilizando la información proporcionada.

b) Estima la matriz de varianzas y covarianzas de los coeficientes estimados.

c) Obtén una medida de la bondad del ajuste.

d) Contrasta la significación individual y conjunta de las variables.

e) Valora e interpreta los resultados obtenidos.

f) Si la relación entre las variables no ha cambiado y te dijeran que las variables Exportaciones y Tasa salarial han tomado los siguientes valores : $E_{2000} = 18,6$ y $T_{2000} = 3,5$ ¿te parecería razonable esperar una cifra de Paro registrado de 121,17 en el año 2000?

Apartado II:

g) Se ha estimado también el modelo siguiente:

$$\underset{\text{(desv.típica)}}{\widehat{P}_t} = \underset{(15,12)}{120,78} + \underset{(1,99)}{1,754}T_t \tag{2}$$

Dados los resultados obtenidos en el Apartado I), ¿cómo explicas que se haya obtenido con los mismos datos la estimación anterior? Si tuvieras que elegir un modelo para explicar el comportamiento del paro registrado, ¿cuál de los dos elegirías? ¿Por qué?

h) Contrasta la hipótesis de que los coeficientes de las dos variables explicativas en el modelo (1) son iguales. Valora e interpreta el resultado obtenido. ¿Podrías obtener una estimación ”mejor” de los coeficientes? ¿Cómo? ¿En qué sentido sería ”mejor¿ Razona cuidadosamente tu respuesta.

[0] CVS Id: $Id: 01e1g.tex,v 1.1.1.1 2003/09/17 14:40:56 etpdihei Exp

Apartado III:

i) Se sospecha que la entrada de España en la Unión Europea ocurrida en 1986 puede haber tenido algún efecto sobre el paro registrado en el País Vasco. ¿Cómo incluirías este posible efecto en el modelo (1)? Escribe el modelo que propones, describiendo con claridad todos sus elementos e interpreta los coeficientes del mismo.

j) Si te dicen que para el modelo que has propuesto en el apartado anterior se ha obtenido un valor de $R^2 = 0,7330$, lleva a cabo un contraste de la hipótesis de que la entrada en la Unión Europea no ha tenido efecto sobre el paro registrado en el País Vasco.

LE-2001.2 (Jun-2001)

Con objeto de analizar el mercado automovilístico de un determinado país europeo, se especifica el siguiente modelo:

$$\mathrm{Y}_t = \beta_1 + \beta_2 X_{2t} + \beta_3 X_{3t} + u_t \qquad t = 1987, \ldots, 2000 \quad u \sim N(0, \sigma^2 I) \tag{1}$$

donde: Y_t es el número, en miles de unidades, de coches vendidos, X_{2t} es el precio medio, en millones, de un coche y X_{3t} es la renta media del país, en millones. Con la siguiente información muestral:

$$\begin{array}{lll}
\sum_t Y_t = 35,962 & \sum_t (Y_t - \bar{Y})(X_{2t} - \bar{X}_2) = 3,50938 & \sum_t (X_{2t} - \bar{X}_2)^2 = 6,8656 \\
\sum_t X_{2t} = 87,10002 & \sum_t (Y_t - \bar{Y})(X_{3t} - \bar{X}_3) = 82,93 & \sum_t (X_{3t} - \bar{X}_3)^2 = 821,31 \\
\sum_t X_{3t} = 463 & \sum_t (X_{3t} - \bar{X}_3)(X_{2t} - \bar{X}_2) = 48,0046 & \sum_t (Y_t - \bar{Y})^2 = 11,514
\end{array}$$

$$(X'X)^{-1} = \begin{pmatrix} 5,9331 & -1,0563 & 0,0215 \\ -1,0563 & 0,2463 & -0,0144 \\ 0,0215 & -0,0144 & 0,0020587 \end{pmatrix}$$

APARTADO I

1. Estima el modelo propuesto por MCO.
2. Interpreta las estimaciones obtenidas. ¿Crees que el modelo propuesto se corresponde con una función de oferta?, ¿y de demanda?
3. Calcula el R^2 e interprétalo.
4. ¿Son las variables explicativas individualmente significativas?, ¿y conjuntamente?
5. Si en el año 2002 el precio medio del automóvil es $5,5$ y la renta media del país es $22,9$, ¿entre qué valores se encuentra el número de coches que se venderán en ese año?

6. En el año 2002 desaparecerá la moneda del país, y tanto el precio de los coches como la renta se medirán en euros, ¿deberías volver a estimar el modelo?, ¿cambiaría la predicción que has realizado en el apartado anterior?

   APARTADO II

7. Contrasta la hipótesis de que el efecto de la renta y el del precio sobre el número medio de coches vendidos, es de la misma cuantía pero de signo opuesto.

8. Dado el resultado del contraste anterior, ¿hay algún modo de mejorar en la estimación?, ¿en qué sentido?

9. En el modelo propuesto, no se ha tenido en cuenta el posible efecto del precio del seguro del coche sobre el número de coches vendidos, ¿qué implicaciones tendría sobre las propiedades de los estimadores del modelo no tener en cuenta esta variable en la estimación?

10. ¿Qué hubiera sucedido si las variables X_2 y X_3 tuvieran un coeficiente de correlación simple del valor de $0,998$?

11. En el año 1996 el gobierno introdujo el Plan Renove para modernizar el parque de automóviles. Formula un modelo que recoja este hecho. Interpreta los coeficientes del modelo que has propuesto. Describe cada uno de los elementos necesarios para llevar a cabo un contraste sobre si el Plan Renove ha sido efectivo o no.

LADE-2001.1 (Feb-2001)

Un consultor que trabaja para una empresa de la llamada *Nueva Economía*, considera el siguiente modelo:

$$Y_t = \beta_1 + \beta_2 X_{2t} + \beta_3 X_{3t} + u_t \qquad t = 1996, \ldots, 2000 \qquad u_t \overset{iid}{\sim} N(0, \sigma_u^2) \tag{1}$$

Y_t = Número de accesos en el año t al portal de Internet MI-MUNDO.COM en decenas de millones.
X_{2t} = Número de ordenadores personales del país en cientos de miles de unidades.
X_{3t} = Precio de la llamada local en cientos de pesetas la hora.

Si se dispone de los siguientes datos:

Y	X_2	X_3
0,5	1	3
6,7	4	2
4,2	3	3
6,3	4	2
10,4	5	1

$$(X'X)^{-1} = \begin{pmatrix} 22,5 & -3 & -5,5 \\ & 0,4375 & 0,6875 \\ & & 1,4375 \end{pmatrix}$$

[0] CVS Id: $Id: 01e1e.tex,v 1.2 2003/09/18 07:47:11 etpdihei Exp

1. Estima el modelo (1) por el método MCO.
2. Interpreta los coeficientes del modelo.
3. Calcula e interpreta el coeficiente de determinación.
4. Estima la matriz de covarianzas de los estimadores del modelo.
5. ¿Existe evidencia en la muestra, para un nivel de significación del 5 %, de que el precio de la llamada local afecta negativamente al número de accesos del portal ($H_o : \beta_3 = 0 \;\; H_a : \beta_3 < 0$)?
6. Contrasta la hipótesis de significación conjunta de las variables del modelo.
7. Las previsiones para el año 2001 son que el número de ordenadores personales del país alcance las 700.000 unidades (**7 cientos de miles**), mientras que se espera que el precio de la llamada local se mantenga en 100 pesetas (**1 ciento**). ¿Cuál crees que será, basándote en esta información, el número de accesos en el año 2001? ¿Crees estadísticamente posible que el número de accesos pueda duplicarse en 2001 con respecto al año 2000?
8. Estima el modelo sujeto a la restricción de que $\beta_2 = 2$.
9. Sabiendo que la función de regresión poblacional es:

$$E(Y_t) = 1 + 2X_{2t} - X_{3t} \qquad t = 1, \ldots, T \tag{2}$$

   ¿Es el estimador calculado en el apartado anterior insesgado? Entre éste y el del modelo inicial (1), ¿cuál elegirías? Razona tu respuesta.

LADE-2001.2 (Feb-2001)

La siguiente tabla recoge el gasto anual por estudiante de tercer ciclo (Y) y el Producto Interior Bruto per capita (X), ambos en miles de dólares, para un conjunto de **24** países (datos de la publicación *OCDE en cifras 2000*).

Países	Y	X	X^2	XY	Y^2
Australia	11,2	24,4	595,36	273,28	125,44
Austria	10	24,6	605,16	246	100
Belgium	7,8	24,3	590,49	189,54	60,84
Canada	14,8	25,9	670,81	383,32	219,04
Czech Republic	5,4	13,1	171,61	70,74	29,16
Denmark	7,3	26,3	691,69	191,99	53,29
Finland	7,1	22,8	519,84	161,88	50,41
France	7,2	21,9	479,61	157,68	51,84
Germany	9,5	23,6	556,96	224,2	90,25
Greece	4	14,8	219,04	59,2	16
Hungary	5,4	10,9	118,81	58,86	29,16
Ireland	8,1	25,2	635,04	204,12	65,61

Japan	10,2	24,5	600,25	249,9	104,04
Korea	6,8	15,9	252,81	108,12	46,24
Mexico	4,5	8,1	65,61	36,45	20,25
Netherlands	10	25,1	630,01	251	100
Norway	10,1	27,6	761,76	278,76	102,01
Poland	4,4	8,1	65,61	35,64	19,36
Spain	5,2	18,1	327,61	94,12	27,04
Sweden	13	23	529	299	169
Switzerland	16,4	27,5	756,25	451	268,96
Turkey	2,4	6,3	39,69	15,12	5,76
United Kingdom	8,2	22,3	497,29	182,86	67,24
United States	17,5	33,9	1149,21	593,25	306,25
SUMAS	**206,5**	**498,2**	**11529,5**	**4816,03**	**2127,19**

Un investigador piensa que los países del area del euro (Austria, Belgium, Denmark, Finland, France, Germany, Ireland, Netherlands y Spain) tienen un comportamiento diferente de los del resto del mundo, y plantea el siguiente modelo:

$$Y_i = \alpha + \delta E_i + \beta X_i + \gamma E_i X_i + u_i \qquad u_i \overset{iid}{\sim} N(0, \sigma_u^2) \tag{3}$$

donde E_i es una variable ficticia que toma valor 1 si el país pertenece al area del euro y cero en caso contrario.

1. ¿Cuál es la interpretación de los coeficientes del modelo (3)?

2. Con los datos de la tabla se ha estimado el modelo (3):

$$\underset{(\hat{desv}(\hat{\beta}))}{\hat{Y}_i} = \underset{(1,135)}{-0,92} - \underset{(6,123)}{1,31}\, E_i + \underset{(0,0546)}{0,517}\, X_i - \underset{(0,26)}{0,082}\, X_i E_i \qquad R^2 = 0,8252$$

   ¿Es el término independiente significativamente distinto entre los países del area del euro y el resto? ¿Y la pendiente?

3. Utilizando el estadístico basado en las sumas de cuadrados residuales de los modelos restringido y no restringido **contrasta** si los países del area del euro tienen un comportamiento diferente del resto para un nivel de significación del 5 %. ¿Encuentras alguna contradicción con el resultado del apartado anterior? Si lo haces, da una posible explicación para esta contradicción.

4. Un investigador diferente piensa que el modelo adecuado no es el modelo (3), sino uno de la forma:

$$Y_i = \alpha_1 NE_i + \alpha_2 E_i + \beta_1 NE_i X_i + \beta_2 E_i X_i + u_i \tag{4}$$

   donde NE_i es una variable ficticia que toma valor 1 si el país **NO** pertenece al area del euro y cero en caso contrario. ¿Cuál de los dos modelos, el (3) o el (4), es preferible? Calcula las estimaciones de α_1, α_2, β_1 y β_2.

LADE-2001.3 (Jun-2001)

Un investigador desea analizar la relación entre la producción de un grupo de empresas de cierto sector industrial, Y, y los factores productivos que utilizan, capital, K, y trabajo, L. Para ello propone el siguiente modelo tipo Cobb-Douglas:

$$\ln Y_i = \beta_1 \ln K_i + \beta_2 \ln L_i + u_i \qquad u_i \sim NID(0, \sigma_u^2) \tag{1}$$

Para ello dispone de los datos, **en logaritmos**, de 6 empresas:

Empresa	$\ln Y_i$	$\ln K_i$	$\ln L_i$
1	3	4	2
2	-3	-4	-1
3	3	4	1
4	-1	-2	-1
5	2	2	1
6	-4	-4	-2

a) Estima los parámetros del modelo (1), β_1 y β_2, por MCO.

b) Contrasta la hipótesis nula de que existen rendimientos constantes a escala, es decir, $H_o : \beta_1 + \beta_2 = 1$ frente a la alternativa, $H_a : \beta_1 + \beta_2 \neq 1$.

c) **Estima** los parámetros del modelo (1), β_1 y β_2, imponiendo la restricción anterior, es decir, $\beta_1 + \beta_2 = 1$.

d) Si $\beta_1 + \beta_2 = 0,9$ (y no $\beta_1 + \beta_2 = 1$ como se ha impuesto en el apartado anterior) ¿cuáles son las propiedades del estimador restringido? Si queremos elegir entre los estimadores de los apartados a) y c), ¿qué criterio debemos utilizar? Razona tu respuesta.

LADE-2001.4 (Jun-2001)

Con el fin de explicar las importaciones (I_t) se analiza la relación entre éstas con el producto interior bruto (P_t) y el consumo (C_t) en un país desde 1964 hasta 1998. La estimación del modelo con esta muestra resulta ser:

$$\underset{(\text{Estadístico t})}{\hat{I}_t} = 5,7247 + \underset{(0,17238)}{0,03220}P_t + \underset{(0,8504)}{0,2427}C_t \qquad\qquad R^2 = 0,7396$$

Analiza la significatividad individual y conjunta de las variables P_t y C_t y comenta razonadamente cuál puede ser el problema si al incluir en la estimación los datos correspondientes al año 1999 se obtienen las siguientes estimaciones de los coeficientes:

$$\hat{I}_t = 25,421 + 0,092P_t + 0,1456C_t$$

LADE-2001.5 (Jun-2001)

En una muestra de 40 familias se ha estimado la renta disponible (R_i) en función del nivel de estudios máximo alcanzado por el cabeza de familia (primarios, medios y superiores) donde se supone que se satisfacen todas las hipótesis básicas:

$$\underset{(\widehat{desv}(\hat{\beta}_j))}{\hat{R}_i} = \underset{(102{,}68)}{1093{,}68} + \underset{(161{,}10)}{440{,}55}\,EM_i + \underset{(188{,}64)}{881{,}32}\,ES_i$$

donde EM_i (respectivamente ES_i) es una variable ficticia que toma el valor 1 si el cabeza de familia tiene estudios medios (respectivamente superiores) y 0 en caso contrario. Además se sabe que:

$$\sum_{i=1}^{40} R_i = 56525 \qquad \sum_{i=1}^{40}(R_i - \bar{R})^2 = 12067482{,}52 \qquad \sum_{i=1}^{40} \hat{u}_i^2 = 7412384{,}4$$

a) Interpreta los coeficientes estimados.

b) ¿Cuál es el valor medio estimado de la renta disponible si el cabeza de familia tiene estudios medios?

c) Calcula e interpreta el coeficiente de determinación.

d) Obtén un intervalo de confianza del 95 % para la renta media disponible de los hogares cuyo cabeza de familia tiene un nivel de estudios primarios.

e) Bajo la hipótesis nula de que el nivel de estudios alcanzado por el cabeza de familia no influye en la renta disponible:

- ¿Cuál sería el modelo restringido?
- Estímalo.
- Obtén su coeficiente de determinación e interpreta el resultado.

f) Utilizando las sumas de cuadrados residuales de los modelos restringido y no restringido, contrasta al nivel de significación del 5 % esa hipótesis nula.

LADE-2001.6 (Jun-2001)

Se realiza un estudio con el fin de analizar el consumo de café entre las familias de una determinada comunidad autónoma. Se recogen datos del consumo de café (Q), la renta (R), los precios del café (PC) y de la leche (PL), para un conjunto de 20 familias. En un primer análisis se estima la siguiente ecuación:

$$\underset{(\widehat{desv}(\hat{\beta}_j))}{\hat{Q}_i} = \underset{(6{,}178)}{1{,}1875} + \underset{(0{,}0000274)}{0{,}00011}\,R_i - \underset{(0{,}0495)}{0{,}1516}\,PC_i \qquad (2)$$

$$R^2 = 0{,}5575 \qquad \sum \hat{u}_i^2 = 53{,}46$$

a) Contrasta la significación individual de cada una de las variables explicativas.

b) Contrasta la significación conjunta de las variables. Un estudio posterior incluye como variable explicativa el precio de la leche, por lo que se vuelve a estimar el modelo con las mismas familias obteniéndose el siguiente ajuste con una bondad del $57,79\,\%$:

$$\hat{Q}_i = 8,821 + 0,00011R_i - 0,195PC_i + 0,084PL_i \tag{3}$$

c) Contrasta la significación individual del precio de la leche.

d) Discute cuales serían las ventajas e inconvenientes de elegir cada una de las dos especificaciones. Razona tu respuesta.

LE-2002.1 (Feb-2002)

El departamento de recursos humanos de una empresa de consultoría ha llevado a cabo un estudio sobre la relación entre la remuneración de sus empleados (Y), los años de estudios que poseen (X_1) y su experiencia laboral (X_2). Los resultados del estudio para una muestra de 30 empleados han sido los siguientes:

$$\hat{Y}_i = 73,034 + 0,178X_{1i} + 4,983X_{2i} \tag{1}$$

$$\hat{\sigma}^2_u = 47,7481 \qquad \sum_{i=1}^{30}(Y_i - \bar{Y})^2 = 5387,5301$$

$$(X'X)^{-1} = \begin{pmatrix} 0,604 & -0,025 & -0,0306 \\ & 0,00018 & -0,0018 \\ & & 0,0087 \end{pmatrix}$$

1. Interpreta el significado económico de **los coeficientes estimados**.

   $\boxed{\text{Solución:}}$ $\widehat{\beta_0}$ = Valor estimado para $\widehat{Y_i}$ cuando $X_{1i} = X_{2i} = 0$.
   $\widehat{\beta_1}$= Incremento en $\widehat{Y_i}$ cuando X_{1i} aumenta en una unidad y X_{2i} permanece constante. Como $\widehat{\beta_1} = 0,178$ la remuneración aumenta en $0,178$ unidades por cada año de estudio si la experiencia laboral no varía.
   $\widehat{\beta_2}$= Incremento en $\widehat{Y_i}$ cuando X_{2i} aumenta en una unidad y X_{1i} permanece constante. La remuneración aumenta en $4,983$ por cada año más de experiencia laboral si el nivel de estudio no varía.

   .

2. Calcula una medida de la bondad del ajuste. Interprétala.

   $\boxed{\text{Solución:}}$

$$R^2 = 1 - \frac{\Sigma\widehat{u_t}^2}{\Sigma(Y_t - \bar{Y})^2} = \quad 1 - \frac{\widehat{\sigma_u}^2(T-k-1)}{\Sigma(Y_t-\bar{Y})^2} = 1 - \frac{47,748(30-3)}{5387,5309} = 0,7607$$

[0]CVS Id: $Id: 02e1g.tex,v 1.3 2003/09/18 15:55:11 etpdihei Exp

El modelo explica el $76,07\,\%$ de las variaciones en la remuneración a través de los cambios en los años de estudio y la experiencia laboral.

3. ¿Puedes aceptar que el multiplicador de la experiencia laboral sobre el salario de los empleados de la empresa es igual a 6?

Solución: Dado que

$$\frac{\Delta Y}{\Delta X_2} |_{X_1=cte}= \beta_2$$

de donde la hipótesis a contrastar será

$$H_0 : \beta_2 = 6$$
$$H_a : \beta_2 \neq 6$$

Dado que

$$\frac{\widehat{\beta_2} - 6}{\sigma_{\widehat{\beta_2}}}$$

sigue una distribución $t(T - k - 1)$ bajo H_0, y puesto que

$$\widehat{\sigma_{\widehat{\beta_2}}} = \sqrt{\widehat{\sigma_u}^2 (X'X)^{-1}_{3,3}} = \sqrt{47,7481 \times 0,0087} = 0,6445$$

$\Rightarrow$

$$t = \frac{4,983 - 6}{0,6445} = -1,577$$

$$|t| \leq t^{0,025}_{T-k-1} = 2,052$$

Acepto la H_0, el multiplicador podría ser igual a 6.

4. ¿Son las variables explicativas conjuntamente significativas para explicar la remuneración de sus empleados?

Solución: La hipótesis a contrastar será:

$$H_0 : \beta_1 = \beta_2 = 0$$
$$H_a : \beta_1 \neq 0 \quad \text{y/o} \quad \beta_2 \neq 0$$

Bajo H_0:

$$\frac{\frac{R^2}{m}}{\frac{1-R^2}{T-k-1}}$$

sigue una distribución $F(m, T - k - 1)$. En este caso,

$$m = 2 \quad \frac{\frac{0,7607}{2}}{\frac{1-0,7607}{27}} = 42,91 \geq F^2_{27,0,05} \simeq 3,39$$

$\Rightarrow$ Se rechaza la H_0 al nivel de significación del $5\,\%$. Las dos variables son conjuntamente significativas.

Un estudio alternativo considera la experiencia laboral como la única variable relevante para explicar la remuneración de los empleados. Para la misma muestra se obtienen los siguientes resultados:

$$\hat{Y}_i = 74,79 + 5,108X_{2i} \tag{2}$$

$$R^2 = 0,7279$$

5. ¿Está justificada la exclusión de la variable años de estudio en el modelo? Realiza los contrastes oportunos basándote en la comparación del modelo restringido y no restringido.

   Solución:

   $$H_0 : \beta_1 = 0 \Rightarrow Y_i = \beta_0 + \beta_2 X_{2i} + u_i \quad MR(2)$$
   $$H_a : \beta_1 \neq 0 \Rightarrow Y_i = \beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + u_i \quad MNR(1)$$

   Basándonos en ambos modelos, podemos contrastar con el estadístico

   $$F = \frac{R^2_{NR} - R^2_R}{1 - R^2_{NR}} \frac{T - k - 1}{m}$$

   donde m es el número de restricciones a contrastar. El estadístico sigue, bajo la hipótesis nula, la distribución $F(m, T-k-1)$. Para este caso,

   $$F = \frac{\frac{(0{,}7607 - 0{,}7279)}{1}}{\frac{(1 - 0{,}7607)}{27}} = 3,70$$

   dado que R^2_{NR} = $0,7607$ y R^2_R=$0,7279$ y $m = 1 \Rightarrow 3,70 \leq 4,24 \simeq F^1_{27}(0,05)$
   $\Rightarrow$ Se acepta H_0 para $\alpha = 0{,}05$, la variable años de estudio no es individualmente significativa y su exclusión del modelo está justificada.

6. Dada la información de que dispones, ¿cuál de las dos especificaciones alternativas elegirías? Razona tu respuesta indicando claramente las ventajas del modelo que eliges frente al alternativo.

   Solución: El modelo no restringido incorpora una variable que no es individualmente significatva. No obstante, dado que $\frac{\widehat{\beta_1}}{\sigma_{\beta_1}}$=$\frac{0{,}178}{0{,}0927}$=$1,92$ es evidente que se capacidad explicativa no es nula. Tanto el R^2 como el $\bar{R}^2$ en el modelo (1) serán superiores a los del modelo restringido, ya que el cociente anterior es superior a la unidad. Podría existir colinealidad, puesto que $cov(\beta_1, \beta_2) \neq 0$. Si elegimos el modelo (1) que incluye una variable irrelevante perdemos precisión en la estimación (ya que el estimador será ineficiente) pero mantenemos la insesgadez de los coeficientes y la validez de los contrastes. Por otro lado, si elegimos el modelo (2) y la no significación de X_1 se debe a la existencia de colinealidad estaríamos sesgando la estimación y los contrastes. $\Rightarrow$ Elegiría el modelo (1).

7. Razona las siguientes afirmaciones indicando si son ciertas o falsas en la ecuación (1):

   a) $E(\hat{\beta}_2) = \beta_2$

   Solución: Como $\beta_2 \in \beta$ y como $\hat{\beta}$=$(X'X)^{-1}X'Y$, bajo las hipótesis básicas $E(u) = 0$ y X fijas $\Rightarrow E(\hat{\beta}) = \beta \Rightarrow E(\hat{\beta}_2) = \beta_2 \Rightarrow$ La afirmación es cierta.

b) $E(\hat{\sigma}_u^2) = \sigma_u^2$

$\boxed{\text{Solución:}}$ $\hat{\sigma}_u^2 = \frac{\hat{u}'\hat{u}}{T-k-1} = \frac{u'Mu}{T-k-1}$
$E(\hat{u}'\hat{u})=E(u'Mu)=\sigma^2(T-k-1) \Rightarrow E(\hat{\sigma^2}) = \sigma^2$ bajo las hipótesis básicas $\Rightarrow$ La afirmación es cierta.

c) $\sum_{i=1}^{30} \hat{u}_i = 0$

$\boxed{\text{Solución:}}$ Sabemos que $\hat{u}_i = Y_i - \hat{\beta}_0 - \hat{\beta}_1 X_{1i} - \hat{\beta}_2 X_{2i}$. Del método de estimación mínimo cuadrático obtenemos

$$Min \sum_{i=1}^{30} \hat{u}_i^2 \Rightarrow \frac{\partial}{\partial \beta_0} = -2\sum(Y_i - \hat{\beta}_0 - \hat{\beta}_1 X_{1i} - \hat{\beta}_2 X_{2i}) = 0 \Rightarrow \sum \hat{u}_i = 0$$

$\Rightarrow$ La afirmación es cierta por la primera ecuación normal, siempre que en el modelo exista una ordenada en el origen.

d) $\sum_{i=1}^{30}(X_{2i} - \bar{X}_2) = 0$

$\boxed{\text{Solución:}}$

$$\sum_{i=1}^{30}(X_{2i} - \bar{X}_2) = \sum_{i=1}^{30} X_{2i} - \sum_{i=1}^{30} \bar{X}_2 = \sum_{i=1}^{30} X_{2i} - \frac{30\sum_{i=1}^{30} X_{2i}}{30} = 0$$

$\Rightarrow$ La afirmación es cierta por definición de $\bar{X_{2i}}$.

LE-2002.2 (Feb-2002)

Una empresa ha otorgado becas a tres estudiantes para analizar el comportamiento de sus ventas en el periodo 1981-2000. El primero de ellos debe especificar un modelo adecuado para dicho análisis. Tras un largo proceso de estudio, se encuentra indeciso entre los siguientes modelos:

$$\widehat{V}_t = \underset{(87,94)}{5,76} - \underset{(-0,035)}{1,93}\, P_t + \underset{(4,02)}{3,8}\, R_t \qquad R^2 = 0,8245 \tag{3}$$

$$\widehat{V}_t = \underset{(7,53)}{5,32} - \underset{(-0,12)}{0,83}\, P_t \qquad R^2 = 0,1237 \tag{4}$$

$$\widehat{V}_t = \underset{(7,27)}{5,01} + \underset{(5,32)}{3,67}\, R_t \qquad R^2 = 0,7822 \tag{5}$$

donde V_t son las ventas en miles de unidades, P_t es el precio y R_t es la renta media de la población en el periodo t.

1) En base a la información proporcionada, decide cuál de los tres modelos es el más adecuado. Razona tu respuesta.

Solución: El modelo más adecuado será el mejor especificado. Por ello, deberíamos comenzar analizando la significatividad individual y conjunta de las variables en el modelo más general, ésto es, en (3).Se nos proporciona el valor del estadístico t habitual para el contraste $H_0 : \beta_i = 0$, de forma que podemos concluir que R es una variable individualmente significativa al nivel $\alpha = 0{,}05$ mientras que P no lo es, ya que $t_1^{0{,}025}7 = 2,11$. Ambas variables son conjuntamente significativas:

$$\begin{array}{l} H_0 : \beta_1 = \beta_2 = 0 \\ H_a : \beta_1 \neq 0 \quad \text{y/o} \quad \beta_2 \neq 0 \end{array}$$

ya que $F = \frac{\frac{0{,}8245}{2}}{\frac{(1-0{,}8245)}{17}} = 39,93 > F^2_{17,(0{,}025)} = 3,59$. Por tanto, parece que lo razonable es elegir el modelo (5) que incluye a la variable exógena R y no a P, ya que al ser P no significativa el modelo (3) incluye un regresor irrelevante. No obstante, la elección entre ambos modelos no es tan clara, porque P podría no resultar significativa en el modelo (3) debido a la existencia de colinealidad con R, en cuyo caso elegir (5) implicaría un error de especificación. El modelo (4) no deberíamos elegirlo en ningún caso, ya que excluye la variable R que, como se ha visto, es claramente relevante.

2) El segundo becario, por el contrario, propone el siguiente modelo

$$\widehat{V}_t = \underset{(7{,}88)}{5,78} - \underset{(-3.15)}{3,9}\, P_t + \underset{(5{,}61)}{4,3}\, R_t + \underset{(3{,}22)}{2,6}\, G_t \quad R^2 = 0,91121 \qquad (6)$$

donde G_t es el gasto realizado en publicidad.

Por último, el tercer becario dispone de toda la información anterior y dados los resultados obtenidos por los anteriores becarios, decide especificar las ventas como:

$$V_t = \alpha + \beta R_t + \gamma G_t + e_t$$

¿es correcta su decisión?, ¿por qué?

Solución: Analizando el modelo (6) se observa que todas las variables son individualmente significativas, ya que los valores del estadístico t entre paréntesis son mayores que el valor crítico $t_{16}^{(0{,}025)}$ para todas las variables. Además, son conjuntamente significativas ya que el estadístico para contrastar la hipótesis

$$\begin{array}{l} H_0 : \beta_1 = \beta_2 = \beta_3 = 0 \\ H_a : \beta_1 \neq 0 \quad \text{y/o} \quad \beta_2 \neq 0 \text{y/o} \beta_3 \neq 0 \end{array}$$

toma el valor $\frac{\frac{0{,}91121}{3}}{\frac{(1-0{,}91121}{16}} = 54,93 > F^3_{16(0{,}025)} = 3,24$. El tercer investigador no toma una decisión correcta, ya que excluye P en base a los contrastes realizados en el modelo (3) cuando si tomamos en cuenta los resultados del modelo (6), que incluye la variable gasto en publicidad, es significativa, por lo que la elección de un modelo que no incluya esta variable generará resultados de estimación y contraste sesgados y no válidos.

LE-2002.3 (Feb-2002)

Supongamos que disponemos de datos sobre la demanda de gasolina (Y) en Bilbao en sucesivos trimestres. Denotamos:

$$D_{jt} = \begin{cases} 1 & \text{si } t = j\text{-ésimo trimestre} \\ 0 & \text{en caso contrario} \end{cases}$$

1. ¿Qué interpretación tienen los parámetros en el siguiente modelo

$$Y_t = \alpha_1 D_{1t} + \alpha_2 D_{2t} + \alpha_3 D_{3t} + \alpha_4 D_{4t} + u_t$$

?

$\boxed{\text{Solución:}}$ $\alpha_i = E(Y_t|D_{it} = 0)$, es decir, cada uno de los parámetros representa y recoge el valor esperado de la demanda de gasolina para el trimestre correspondiente. Como sólo D_{it} toma el valor 1 en el trimestre i-ésimo, para $t = 1, ..., T$ si estamos en el primer trimestre $D_{1t} = 1, D_{2t} = D_{3t} = D_{4t} = 0 \Rightarrow Y_t = \alpha_1 + u_t$ por lo que α_1 representa la demanda esperada de gasolina en el primer trimestre. De igual manera, α_2,α_3, α_4 miden la demanda esperada de gasolina en los trimestres segundo, tercero y cuarto.

2. Escribe los valores de la matriz X para el modelo anterior en sus primeras 8 observaciones.

$\boxed{\text{Solución:}}$

$$X' = \begin{bmatrix} 1 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 & 0 & 1 \end{bmatrix}$$

Las observaciones son trimestrales, los trimestres se ordenan cronológicamente y hemos supuesto que las observaciones comienzan al principio de un año.

3. ¿Qué interpretación tienen los parámetros en el modelo

$$Y_t = \beta_0 + \beta_1 D_{1t} + \beta_2 D_{2t} + \beta_3 D_{3t} + u_t$$

?

$\boxed{\text{Solución:}}$ En este caso, se comprueba que

$$\begin{aligned} E(Y_t|D_{it} = 1) &= \beta_0 + \beta_i \quad \text{para i=1,2,3} \\ E(Y_t|D_{1t} = D_{2t} = D_{3t} = 0) &= \beta_0 \quad \text{i=4} \end{aligned}$$

Por lo tanto, β_0 representa la demanda esperada de gasolina en el cuarto trimestre, al tiempo que la demanda esperada en los trimestres primero, segundo y tercero viene dada por $\beta_0 + \beta_1$, $\beta_0 + \beta_2$,$\beta_0 + \beta_3$ respectivamente. Por lo tanto, β_1, β_2, β_3 representan el efecto diferencial en la demanda esperada de gasolina debida a estar en los trimestres primero, segundo y tercero respecto del cuarto trimestre.

4. Establece la relación entre las estimaciones de los parámetros del modelo del apartado a) y los del apartado c).

Solución: Dado lo expuesto en los apartados anteriores, se deduce que

$$\begin{aligned} E(Y_t|\text{trimestre}1) &= \alpha_1 = \beta_0 + \beta_1 \\ E(Y_t|\text{trimestre}2) &= \alpha_2 = \beta_0 + \beta_2 \\ E(Y_t|\text{trimestre}3) &= \alpha_3 = \beta_0 + \beta_3 \\ E(Y_t|\text{trimestre}4) &= \alpha_4 = \beta_0 \end{aligned}$$

Esta misma relación se mantiene, naturalmente, entre los coeficientes estimados, de manera que

$$\begin{aligned} \hat{\alpha_4} &= \bar{Y_4} = \hat{\beta_0} \\ \hat{\alpha_1} &= \bar{Y_1} = \hat{\beta_0} + \hat{\beta_1} \quad \Rightarrow \hat{\beta_1} = \bar{Y_1} - \bar{Y_4} \\ \hat{\alpha_2} &= \bar{Y_2} = \hat{\beta_0} + \hat{\beta_2} \quad \Rightarrow \hat{\beta_2} = \bar{Y_2} - \bar{Y_4} \\ \hat{\alpha_3} &= \bar{Y_3} = \hat{\beta_0} + \hat{\beta_3} \quad \Rightarrow \hat{\beta_1} = \bar{Y_1} - \bar{Y_4} \end{aligned}$$

5. Para cada uno de los modelos anteriores explica cómo contrastarías que no existen diferencias significativas en el consumo de gasolina en las distintas épocas del año. Para cada caso escribe la hipótesis nula y alternativa, el estadístico de contraste así como la regla de decisión. Explica claramente cómo obtendrías cada uno de los elementos del estadístico de contraste.

Solución: Para el modelo del apartado (a), contrastaría

$$\begin{aligned} &H_0 : \alpha_1 = \alpha_2 = \alpha_3 = \alpha_4 \\ &H_a : \alpha_1 \neq \alpha_2 \quad \text{y/o} \quad \alpha_1 \neq \alpha_3 \quad \text{y/o} \quad \alpha_1 \neq \alpha_4 \end{aligned}$$

y utilizaría el estadístico

$$\frac{\frac{\hat{u_R}'\hat{u_R} - \hat{u}'\hat{u}}{m}}{\frac{\hat{u}'\hat{u}}{T-K}}$$

que, bajo la hipótesis nula, sigue una distribución $F(m, T-k)$, donde m es el número de restricciones. Los residuos $\hat{u_R}$ son residuos restringidos procedentes del modelo restringido $Y_t = \alpha + u_t$ mientras que los residuos $\hat{u}$ son los residuos no restringidos procedentes del modelo en (a). Para este caso, $m = 3$, $k = 4$ y T es el tamaño muestral. Si el valor calculado del estadístico $F > F^q_{T-k,(\alpha)}$, rechazaremos H_0 y admitiremos que existen diferencias significativas en el consumo de gasolina según el trimestre. Para el modelo del apartado (c) la hipótesis a contrastar será

$$\begin{aligned} &H_0 : \beta_1 = \beta_2 = \beta_3 = 0 \\ &H_a : \beta_1 \neq 0 \quad \text{y/o} \quad \beta_2 \neq 0 \quad \text{y/o} \quad \beta_3 \neq 0 \end{aligned}$$

y utilizaría el mismo estadístico de contraste que antes, pero ahora los residuos $\hat{u_R}$ serán los procedentes del modelo restringido $Y_t = \beta_0 + u_t$ y los residuos $\hat{u}$ los que provengan del modelo del apartado (c). Para este caso también $m = 3$, $k = 4$ y T representa el tamaño muestral. La regla de decisión será la misma que antes y la misma también la interpretación si se rechaza la hipótesis nula.

LE-2002.4 (Jun-2002)

Se quiere analizar las ventas (en cientos de unidades) de los ciclomotores en Bilbao C_i en función del sexo del comprador y el precio del seguro en cientos de euros, P_i. Se dispone de datos correspondientes a cuatro años tanto para mujeres como para hombres:

C_i	19	21	24	30	45	51	55	48
$Sexo$	M	M	M	M	H	H	H	H
P_i	3,00	3,30	4,02	5,22	6,01	5,22	4,50	5,70

$\sum_{i=1}^{8} C_i = 293$ $\quad \sum_{i=1}^{8} C_i^2 = 12233$ $\quad \sum_{i=1}^{8} P_i = 36,97$ $\quad \sum_{i=1}^{8} S_iP_i = 15,54$

$\sum_{i=1}^{8} P_i^2 = 179,40$ $\quad \sum_{i=1}^{8} S_iC_i = 94$ $\quad \sum_{i=1}^{8} P_iC_i = 1437,15$

Para la siguiente especificación del comportamiento de las ventas,

$$C_i = \beta_1 + \beta_2 S_i + \beta_3 P_i + u_i \qquad i = 1, \dots, 8$$

siendo S_i la variable Sexo, que toma valor uno si el individuo es mujer y cero en caso contrario, y siendo

$$(X'X)^{-1} = \begin{pmatrix} 7,0465 & -2,1180 & -1,2686 \\ -2,1180 & 1,0134 & 0,3487 \\ -1,2686 & 0,3487 & 0,2368 \end{pmatrix}$$

1) Escribe la matriz X correspondiente a las 8 observaciones que aparecen en la tabla de arriba.

Solución:

$$X = \begin{bmatrix} 1 & 1 & 3 \\ 1 & 1 & 3,3 \\ 1 & 1 & 4,02 \\ 1 & 1 & 5,22 \\ 1 & 0 & 6,01 \\ 1 & 0 & 5,22 \\ 1 & 0 & 4,5 \\ 1 & 0 & 5,7 \end{bmatrix}$$

2) Estima el modelo propuesto por MCO y calcula una medida de bondad del ajuste.

Solución:

$$\begin{aligned} \hat{\beta}_{MCO} = (X'X)^{-1}X'Y &= \begin{pmatrix} 7,0465 & -2,1180 & -1,2686 \\ -2,1180 & 1,0134 & 0,3487 \\ -1,2686 & 0,3487 & 0,2368 \end{pmatrix} \begin{pmatrix} 293 \\ 94 \\ 1437,5 \end{pmatrix} = \\ &= \begin{pmatrix} 43,3646 \\ -24,1801 \\ 1,3951 \end{pmatrix} \end{aligned}$$

$$\hat{C}_i = 42,3640 - 24,180S_i + 1,3951P_i$$

$$R^2 = \frac{SCE}{SCT} = \frac{\hat{\beta}'X'Y - T\bar{Y}^2}{Y'Y - T\bar{Y}^2} = \frac{1413,5655}{1501,875} = 0,9412$$

$$\begin{aligned} SCE &= \hat{\beta}'X'Y - T\left(\frac{\sum C_i}{T}\right)^2 = \\ &= (42,3646 - 24,18011,3951)\begin{pmatrix} 293 \\ 94 \\ 1437,5 \end{pmatrix} - 10731,125 = 1413,56 \end{aligned}$$

$$SCT = Y'Y - T\left(\frac{\sum C_i}{T}\right)^2 = 12233 - 10731,125 = 1501,875$$

3) Interpreta los coeficientes estimados. ¿Te parece que tienen sentido desde el punto de vista económico?

Solución:

$\widehat{\beta}_1 = 42,36$ mide el valor estimado de las ventas de ciclomotores para los hombres cuando el precio del seguro es cero. Su interpretación es complicada ya que el supuesto $P_i = 0$ es dificil de aceptar.

$\widehat{\beta}_2 = -24,1801$ mide el diferencial estimado en las ventas de ciclomotores para las mujeres dado un mismo valor de P_i. Su signo negativo indica que en media las mujeres compran menos ciclomotores.

$\widehat{\beta}_3 = 1,3951$ indica en cuánto se incrementa la venta estimada de ciclomotores que ante un incremento en una unidad en el precio del seguro, lo cual no tiene mucho sentido económico.

4) Contrasta la significatividad conjunta de las variables.

Solución:

$H_o: \quad \beta_2 = \beta_3 = 0$
$H_a: \quad \beta_2 \neq 0$ y/o $\beta_3 \neq 0$

Estadístico: $F = \dfrac{R^2}{1-R^2}\dfrac{T-K}{q} \overset{H_o}{\sim} \mathcal{F}_{(q,T-K)}$

Criterio de decisión:

$$\frac{0,9412}{1-0,9412}\frac{8-3}{2} = 40,01 > 5,79 = \mathcal{F}_{(2,5)0,05}$$

Rechazamos la H_o para un nivel de significación de $5\,\%$ y por tanto ambas variables, sexo y precio del seguro, son conjuntamente significativas.

5) ¿Existen diferencias significativas en las ventas de ciclomotores entre los dos sexos?

Solución:

$$\begin{array}{ll} H_o: & \beta_2 = 0 \\ H_a: & \beta_2 \neq 0 \end{array} \quad \text{Estadístico: } \lambda = \frac{\widehat{\beta}_2}{\widehat{\sigma}_{\hat{\beta}_2}} \overset{H_o}{\sim} t_{(T-K)}$$

$$\hat{\sigma}^2 = \frac{\hat{u}'\hat{u}}{T-K} = \frac{1501,875 - 1413,5655}{8-3} = \frac{88,3095}{5} = 17,6619$$

$$\widehat{\sigma}_{\hat{\beta}_2} = \sqrt{\hat{\sigma}^2 \times a_{22}} = \sqrt{17,6619 \times 1,0134} = 4,2306$$

Criterio de decisión:

$$\mid \lambda \mid = \mid \frac{-24,1801}{4,2306} \mid = 5,71 > 2,571 = t_{(8-3)0,025}$$

Rechazamos la H_o para un nivel de significación de $5\,\%$ y por tanto sí existen diferencias significativas entre ambos sexos.

6) Considera la restricción $\beta_2 + 20\beta_3 = 0$.

a) Contrasta si la restricción es compatible con los datos.

Solución:

$$\begin{array}{ll} H_o: & \beta_2 + 20\beta_3 = 0 \\ H_a: & \beta_2 + 20\beta_3 \neq 0 \end{array} \quad \text{Estadístico: } \lambda = \frac{\widehat{\beta}_2 + 20\widehat{\beta}_3}{\widehat{\text{des}}(\widehat{\beta}_2 + 20\widehat{\beta}_3)} \overset{H_o}{\sim} t_{(T-K)}$$

$$\widehat{\beta}_2 + 20\widehat{\beta}_3 = -24,1801 + 20 \times 1,3951 = 3,7219$$

$$\widehat{\text{des}}(\widehat{\beta}_2 + 20\widehat{\beta}_3) = \sqrt{\widehat{\text{var}}\,(\widehat{\beta}_2) + 20^2 \times \widehat{\text{var}}\,(\widehat{\beta}_3) + 2 \times 20 \times \widehat{\text{covar}}\,(\widehat{\beta}_2, \widehat{\beta}_2)} =$$
$$\sqrt{17,6619 \times (1,0134 + 400 \times 0,2368 + 40 \times 0,3487)} = \sqrt{1937,1819} =$$

$$44,0134$$

Criterio de decisión:

$$\mid \lambda \mid = \frac{3,7219}{44,0134} = 0,08456 < 2,571 = t_{(8-3)0,025}$$

No rechazamos la H_o para un nivel de significación de $5\,\%$ y por tanto la restricción es compatible con los datos.

b) Si has aceptado la restricción ¿querrías reformular el modelo original? ¿Por qué?

Solución: Dado que se ha aceptado la restricción sí querría reestimar el modelo. Estimaría el modelo restringido:

$$C_i = \beta_1 + \beta_2 S_i + \beta_3 P_i + u_i \quad \text{donde} \quad \beta_2 = -20\beta_3$$

$$C_i = \beta_1 - 20\beta_3 S_i + \beta_3 P_i + u_i$$
$$C_i = \beta_1 - \beta_3(P_i - 20S_i) + u_i$$

La razón de optar por estimar el modelo restringido es que las varianzas de los estimadores $\hat{\beta}_1^R$, $\hat{\beta}_3^R$ y $\hat{\beta}_2^R = -20\hat{\beta}_3^R$ son menores que los de MCO.

c) A efectos del modelo ¿es equivalente la restricción anterior a la igualdad $S_i + 20P_i = 0$?

Solución: No, la restricción anterior era una restricción sobre los coeficientes mientras que la igualdad $S_i + 20P_i = 0$ recoge una combinación lineal de las variables explicativas S_i y P_i. Esta relación nos indica que existe un problema de multicolinealidad prefecta por lo que no se pueden estimar los coeficientes del modelo de forma individual. El modelo resultante de introducir esta combinación es: $C_i = \beta_1 + (\beta_3 - 20\beta_2)P_i + u_i$ por lo que solamente tenemos información para estimar β_1 y la combinación $\beta_3 - 20\beta_2$.

7) Dado que el precio del seguro es más bajo para las mujeres, se sospecha que ello tiene un efecto diferencial sobre la venta de ciclomotores. Especifica un modelo que te permita contrastar esta sospecha y explica qué procedimiento seguirías para realizar el contraste.

Solución: Si el precio del seguro es más bajo para las mujeres su efecto en C_i podría ser diferente. Sin embargo en el modelo propuesto el efecto del precio es β_3 siendo común para los hombres y las mujeres. para que este efecto sea distinto habría que discriminar por sexo en precios. Esto puede hacerse de la siguiente forma: $C_i = \beta_1 + \beta_2 S_i + \beta_3 P_i + \beta_4(S_i P_i) + u_i$ donde β_4 recoge el efecto diferencial en P_i por ser mujer. Para contrastar la existencia de este efecto, contrastamos:

$$\begin{array}{ll} H_o: & \beta_4 = 0 \\ H_a: & \beta_4 \neq 0 \end{array} \qquad t_r = \frac{\widehat{\beta}_4}{\widehat{\sigma}_{\widehat{\beta}_4}} \overset{H_o}{\sim} t_{(T-K)} \quad T = 8 \quad k = 4$$

Si $\mid t_r \mid >= t_{(4)0{,}025}$ rechazamos la H_o para un nivel de significación de $5\,\%$ y existiría un efecto diferencial.

8) Si fuera cierta la sospecha del apartado anterior ¿qué efectos tendría esto en los estimadores del apartado 2)? ¿Y en los contrastes que has realizado?

Solución: Si el modelo correcto fuera $C_i = \beta_1 + \beta_2 S_i + \beta_3 P_i + \beta_4(S_i P_i) + u_i$ y hemos estimado el modelo $C_i = \beta_1 + \beta_2 S_i + \beta_3 P_i + u_i$ habríamos omitido una variable relevante que no es ortogonal ni a S_i ni a P_i por lo que los estimadores MCO de los coeficientes serían sesgados. El estimador de la varianza de la perturbación también sería sesgado. La inferencia realizada en base a estos estimadores no es válida ya que los estadísticos t_c y F habituales no siguen las distribuciones t-Student y F-Snedecor.

LE-2002.5 (Jun-2002)

Se ha estimado un modelo para analizar el grado de ocupación hotelera (Y) como función de la calidad de los servicios prestados (X_1), el precio medio por habitación (X_2) y la puntuación obtenida por el hotel en la Guía de Hoteles del año correspondiente (X_3). Los resultados de la estimación por MCO para un conjunto de 11 hoteles han sido:

$$\hat{Y}_i = 66431,018 - \underset{(389,689)}{100,143}X_{1i} - \underset{(1660,464)}{3649,113}X_{2i} + \underset{(0,0776)}{0,8882}X_{3i} \qquad R^2 = 0,9876 \tag{1}$$

1) ¿Podría existir algún problema de colinealidad en el modelo anterior? Justifica adecuadamente tu respuesta.

Solución: Un resultado habitual cuando existe alta colinealidad entre regresores es obtener un R^2 alto junto con variables exógenas individualmente no significativas:

$$\begin{array}{ll} H_o: & \beta_i = 0 \\ H_a: & \beta_i \neq 0 \end{array} \qquad \text{Estadístico: } t_r^i = \frac{\widehat{\beta}_i}{\widehat{\sigma}_{\hat{\beta}_i}} \overset{H_o}{\sim} t_{(T-K)}$$

$\mid t_r^1 \mid = 0,2569 < 2,365 = t_{(11-4)0,025}$ no rechazo H_o para un nivel de $\alpha = 5\,\%$, X_{1i} no es significativa
$\mid t_r^2 \mid = 2,1976 < 2,365 = t_{(11-4)0,025}$ no rechazo H_o para un nivel de $\alpha = 5\,\%$, X_{2i} no es significativa
$\mid t_r^3 \mid = 11,4458 > 2,365 = t_{(11-4)0,025}$ rechazo H_o para un nivel de $\alpha = 5\,\%$, X_{3i} es significativa

$$\begin{array}{ll} H_o: & \beta_1 = \beta_2 = \beta_3 = 0 \\ H_a: & \beta_1 \neq 0 \ \beta_2 \neq 0 \text{ y/o } \beta_3 \neq 0 \end{array}$$

Estadístico: $F = \dfrac{R^2}{1-R^2}\dfrac{T-K}{q} \overset{H_o}{\sim} \mathcal{F}_{(q,T-K)}$

Criterio de decisión:

$$F = 185,83 > 4,35 = \mathcal{F}_{(3,7)0,05}$$

Rechazamos la H_o para un nivel de significación de $5\,\%$ y por tanto las tres variables explicativas son conjuntamente significativas.
A la vista de los resultados podemos **sospechar** la existencia de colinealidad entre las variables X_{1i} y X_{2i}. Sin embargo con la información disponible **no** podemos afirmarlo.

Además de la estimación anterior se conocen las siguientes estimaciones auxiliares:

$$\hat{Y}_i = 98116,978 + \underset{(1118,7)}{4038,84}X_{1i} + \underset{(6846,864)}{10798,901}X_{2i} \qquad R^2 = 0,7335 \tag{2}$$

$$\hat{Y}_i = 60071,404 + \underset{(455,46)}{192,823}X_{1i} + \underset{(0,0776)}{0,8882}X_{3i} \qquad R^2 = 0,9776 \tag{3}$$

$$\hat{Y}_i = 67310,275 - \underset{(1252,4)}{3503,14}X_{2i} + \underset{(0,04)}{0,8717}X_{3i} \qquad R^2 = 0,9874 \tag{4}$$

2) Dada toda la información disponible, elige el modelo más adecuado, explicando con claridad por qué es el más adecuado y cuáles son las propiedades de los estimadores en el modelo elegido así como en los no elegidos.

Solución:
En el modelo conjunto solamente X_3 es individualmente significativo, por lo tanto nunca elegiré el modelo (1). En (2), X_2 no es significativa pero sí lo es en (4). Dado que en X_1 no es significativa ni en (1) ni en (2), hay evidencia de que X_1 no es realmente significativa mientras que X_2 sí lo es. Podemos concluir que las variables X_2 y X_3 son significativas mientras que X_1 no lo es. Por tanto el modelo más adecuado es el (4) que no incluye una variable irrelevante ni excluye variables relevantes. Los estimadores MCO empleados en este modelo son lineales respecto a u, insesgados y de mínima varianza. Los estimadores MCO empleados en los modelos (2) y (3), que incluyen una variable irrelevante y omiten una relevante, son sesgados no siendo la inferencia válida. El estimador MCO empleado en el modelo (1), que incluye una variable irrelevante, es lineal e insesgado pero pierde eficiencia en la estimación debido a la pérdida de grados de libertad.

LADE-2002.1 (Feb-2002)

Sea la variable Y el gasto semanal en euros realizado por un individuo en cine, y X es el número de días libres por semana. Se dispone además de información relativa al sexo del individuo recogida en la variable S ($S_i = 1$ si el i-ésimo individuo es mujer y 0 en caso contrario). Con los datos anteriores se especifica el modelo:

$$Y_i = \alpha + \alpha_1 S_i + u_i \quad i = 1, 2, \ldots, N \quad u_i \sim N(0, \sigma_u^2). \tag{1}$$

a) Interpreta los coeficientes del modelo (1).

b) Obtén las expresiones teóricas de los estimadores de mínimos cuadrados ordinarios de los coeficientes α y α_1 en función de las medias muestrales del gasto en cine realizado por los hombres ($\bar{Y}_H$) y las mujeres ($\bar{Y}_M$).

c) Una vez encuestados los individuos se obtiene que $\bar{Y} = 9,91$, $\bar{Y}_H = 8,834$ e $\bar{Y}_M = 11,525$. Obtén las estimaciones MCO de los coeficientes α y α_1.

d) En la estimación del modelo (1) para una muestra de 100 observaciones se ha obtenido que $R^2 = 0,1962$. Obtén el modelo restringido a la hipótesis de que no hay diferencias entre hombres y mujeres en el gasto semanal esperado. Contrasta dicha hipótesis basándote en la comparación de los modelos restringido y no restringido.

Utilizando la misma muestra de 100 observaciones se ha estimado por MCO el siguiente modelo:

$$Y_i = \beta_0 + \beta_1 S_i + \beta_2 X_i + \varepsilon_i \quad i = 1, 2, \ldots, 100 \quad \varepsilon_i \sim N(0, \sigma_\varepsilon^2). \tag{2}$$

Los resultados de la estimación se resumen a continuación

$$(X'X)^{-1} = \begin{pmatrix} 0,0463 & -0,0169 & -0,0199 \\ -0,0169 & 0,0417 & 0,0002 \\ -0,0199 & 0,0002 & 0,0133 \end{pmatrix} \qquad \begin{array}{ll} \sum_{i=1}^{100} Y_i = 991 & \sum_{i=1}^{100} S_i = 40 \\ \sum_{i=1}^{100} X_i = 148,5 & \sum_{i=1}^{100} Y_i S_i = 461 \\ \sum_{i=1}^{100} Y_i X_i = 1679,5 & \sum_{i=1}^{100} Y_i^2 = 10707 \end{array}$$

e) Estimar los coeficientes del modelo (2) por mínimos cuadrados ordinarios.

f) Obtén su coeficiente de determinación e interprétalo.

g) Si te dicen que un hombre que tiene un día libre a la semana gasta 14 euros semanales en cine, ¿te lo creerías?

LADE-2002.2 (Feb-2002)

Los directivos de una empresa desean estudiar el absentismo laboral de sus 48 empleados. Para ello disponen, para cada empleado, de la siguiente información:

Y_i: Número de días que el i-ésimo empleado ha faltado al trabajo en el último año.

S_i: Variable ficticia que toma el valor 1 si el i-ésimo empleado es hombre.

E_i: Edad del i-ésimo empleado.

A_i: Antiguedad en la empresa del i-ésimo empleado.

W_i: Salario mensual en euros del i-ésimo empleado.

Un directivo considera que el sexo es la única variable que influye en el absentismo laboral y estima el siguiente modelo:

$$\underset{\text{(estad t)}}{\widehat{Y}_i} = \underset{(5,93)}{4,952} - \underset{(-0,73)}{0,804}\, S_i \qquad R^2 = 0,0113 \tag{3}$$

a) Contrasta, al nivel de significación del 5 %, si el sexo influye en el absentismo laboral.

Otro directivo estima los tres siguientes modelos:

$$\widehat{Y}_i = \underset{(8,92)}{13,441} + \underset{(3,02)}{2,165}\, S_i - \underset{(-0,88)}{0,042}\, E_i - \underset{(-1,25)}{0,149}\, A_i - \underset{(-6,17)}{0,045}\, W_i \qquad R^2 = 0,7477 \tag{4}$$

$$\widehat{Y}_i = \underset{(13,99)}{15,808} + \underset{(2,27)}{1,588}\, S_i - \underset{(-5,73)}{0,136}\, E_i - \underset{(-6,31)}{0,048}\, W_i \qquad R^2 = 0,718 \tag{5}$$

$$\widehat{Y}_i = \underset{(12,99)}{12,417} + \underset{(3,63)}{2,403}\, S_i - \underset{(-6,35)}{0,200}\, A_i - \underset{(-6,21)}{0,046}\, W_i \qquad R^2 = 0,743 \tag{6}$$

b) A la vista de estos resultados, ¿te parece que puede haber problemas de multicolinealidad en el modelo (4)?

c) Dados los resultados del apartado b), ¿consideras correcta la conclusión obtenida en el apartado a)? Razona tu respuesta.

LADE-2002.3 (Feb-2002)

En el siguiente modelo de regresión donde se cumplen todas las hipótesis básicas

$$Y_t = \alpha + \beta X_t + u_t \quad t = 1, 2, \ldots, T \tag{7}$$

se ha estimado la pendiente mediante el siguiente estimador:

$$\tilde{\beta} = \frac{\sum_{t=1}^{T} X_t Y_t}{\sum_{t=1}^{T} X_t^2}$$

a) Obtén la media y la varianza poblacional de $\tilde{\beta}$.

b) El Teorema de Gauss Markov asegura que el estimador MCO es **siempre** preferible a cualquier otro estimador. Comenta esta afirmación para un contexto general.

LADE-2002.4 (Jun-2002)

Se tienen datos de 141 planes de pensiones durante el año 2001. 98 son planes de pensiones individuales y 43 son planes de pensiones de empleo. Para estudiar su rentabilidad ese año se propone el siguiente modelo:

$$Y_i = \beta_1 + \beta_2 D_i + \beta_3 V_i + \beta_4 L_i + u_i \qquad i = 1, \ldots, 141. \tag{1}$$

Y_i es la rentabilidad del plan de pensiones i-ésimo, D_i vale 1 si i es un plan de pensiones individual y 0 en caso contrario, V_i es el porcentaje del patrimonio invertido en renta variable y L_i es el porcentaje del patrimonio mantenido como liquidez.

a) Interpreta los coeficientes del modelo(1).

Solución:

$$E(Y_i|D_i = 1) = E(Y_i|PPI) = (\beta_1 + \beta_2) + \beta_3 V_i + \beta_4 L_i$$
$$E(Y_i|D_i = 0) = E(Y_i|PPE) = \beta_1 + \beta_3 V_i + \beta_4 L_i$$

β_1:rentabilidad esperada de un PPE que tenga invertido un 0 % de su patrimonio en renta variable (RV) y un 0 % en liquidez.
β_2:diferencia en la rentabilidad esperada de un PPI respecto a un PPE con la misma composición de cartera.
β_3:incremento en la rentabilidad esperada de un PP cuando el porcentaje del patrimonio invertido en RV aumenta en una unidad, manteniendo constante el $\%$ en liquidez y el tipo de plan.
β_4:incremento en la rentabilidad esperada de un PP cuando el $\%$ de patrimonio mantenido en liquidez aumenta en una unidad, resto céteris paribus. Como resultado de la estimación de (1) por mínimos cuadrados ordinarios se ha obtenido:

$$\underset{(\hat{\sigma}_{\hat{\beta}_i})}{\hat{Y}_i} = \underset{(2,23)}{8,12} - \underset{(1,22)}{1,83}D_i - \underset{(0,37)}{2,31}V_i - \underset{(0,64)}{0,12}T_i \qquad SCE = 1615,63 \quad SCR = 5360,46$$

b) ¿Qué variables son individualmente significativas? Realiza los contrastes oportunos.

Solución:

$$\begin{array}{l} H_0 : \beta_2 = 0 \\ H_a : \beta_2 \neq 0 \end{array} \qquad t = \frac{\widehat{\beta_2}}{\widehat{\sigma(\hat{\beta}_2)}} \overset{H_0}{\sim} t_{T-K}$$

$t =\mid \frac{-1,83}{1,22} \mid = 1,5 \leq t^{0,025}(137) = 1,98$

Al nivel de significación del 5 $\%$ no se rechaza la H_0, de modo que la variable tipo de PP no es individualmente significativa.

$$\begin{array}{l} H_0 : \beta_3 = 0 \\ H_a : \beta_3 \neq 0 \end{array} \qquad t = \frac{\widehat{\beta_3}}{\widehat{\sigma(\hat{\beta}_3)}} \overset{H_0}{\sim} t_{T-K}$$

$t =\mid \frac{-2,31}{0,37} \mid = 6,24 \geq t^{0,025}(137) = 1,98$

Al nivel de significación del 5 $\%$ se rechaza la H_0, de modo que la variable $\%$ del patrimonio invertido en RV sí es individualmente significativa.

$$\begin{array}{l} H_0 : \beta_4 = 0 \\ H_a : \beta_4 \neq 0 \end{array} \qquad t = \frac{\widehat{\beta_4}}{\widehat{\sigma(\hat{\beta}_4)}} \overset{H_0}{\sim} t_{T-K}$$

$t =\mid \frac{-0,12}{0,64} \mid = 0,1875 \leq t^{0,025}(137) = 1,98$

Al nivel de significación del 5 $\%$ no se rechaza la H_0, de modo que la variable $\%$ del patrimonio mantenido como liquidez no es individualmente significativa.

c) Obtén el coeficiente de determinación. Interpreta su significado.

Solución:

$$R^2 = \frac{SCE}{SCT} = \frac{SCE}{SCE + SCR} = \frac{1615,63}{1615,63 + 5360,46} = 0,2316$$

El ajuste es malo: sólo el 23 $\%$ de las variaciones en la rentabilidad es explicado por la regresión.

d) Realiza el contraste de significatividad conjunta de la regresión.

Solución:

$$H_0 : \beta_2 = \beta_3 = \beta_4 = 0$$
$$H_a : \beta_2 \neq 0 \quad y/o \quad \beta_3 \neq 0 \quad y/o \quad \beta_4 \neq 0$$

$$F = \frac{R^2/(K-1)}{(1-R^2)/(T-K)} \overset{H_0}{\sim} F_{(K-1,T-K)}$$

$t = \frac{0{,}2316/3}{(1-0{,}2316)/(141-4)} = 13,765 \geq F^{0{,}05}_{(3{,}137)} = 2,68$

Al nivel de significación del 5 $\%$ se rechaza la H_0 en favor de la H_a, por tanto, las variables explicativas son conjuntamente significativas.

e) Al nivel de significación del 10 %, ¿se puede concluir que la rentabilidad en los planes de pensiones individuales es **menor** que en los planes de pensiones de empleo?

Solución:

$$H_0 : \beta_2 \geq 0 \qquad t = \frac{\widehat{\beta_2}}{\widehat{\sigma(\hat{\beta}_2)}} \overset{H_0}{\sim} t_{T-K}$$
$$H_a : \beta_2 < 0$$

$t = -1,5 \leq t^{0{,}10}_{137} = -1,28$

Al nivel de significación del 10 $\%$ se rechaza la H_0 en favor de la alternativa de que $\beta_2 < 0$. Al nivel de significación del 10 $\%$ se puede concluir que la rentabilidad es menor en los planes de pensiones individuales que en los planes de pensiones de empleo.

f) Supón que el 100 % del patrimonio de cada plan de pensiones debe estar distribuido entre renta variable, liquidez y renta fija. Describe qué habría sucedido en caso de incluir también esta última variable en el modelo (1).

Solución: El modelo

$$Y_i = \beta_1 + \beta_2 D_i + \beta_3 V_i + \beta_4 L_i + \beta_5 F_i + U_i \qquad i = 1, 2, \ldots, 141$$

tendría multicolinealidad exacta porque $F_i = 100 - V_i - L_i \quad i = 1, 2, \ldots, 141$. Es decir, existe una combinación lineal exacta entre las variables explicativas y por tanto $rg(X'X) = 4 < 5 = K \Rightarrow \nexists (X'X)^{-1} \quad \Rightarrow$ no se puede estimar de forma única los parámetros del nuevo modelo.

LADE-2002.5 (Jun-2002)

Con los datos relativos a la compra de 150 viviendas en una gran ciudad se especifica el siguiente modelo:

$$\mathrm{P}_i = \beta_1 + \beta_2 C_i + \beta_3 M_i + u_i \qquad i = 1, \ldots, 150,$$

donde P_i es el precio final de venta de la vivienda (en miles de euros), C_i es la tasación de la construcción (en miles de euros) y M_i es la tasación de las mejoras incluidas en la vivienda (en miles de euros).

Con dichos datos se obtiene la siguiente estimación del modelo:

$$\hat{P}_i = 502,36 + 2,172C_i + 1,562M_i \qquad SCR = 2530,76$$

a) Si además de los datos anteriores posees únicamente información sobre $(X'X)^{-1}$ ¿Cómo contrastarías que la tasación de la construcción y de las mejoras tienen el mismo efecto en el precio final de venta? Escribe la hipótesis nula, la alternativa, el estadístico de contraste, cómo obtendrías cada uno de los valores que necesitas y la regla de decisión.

Solución:

$$\begin{array}{l} H_0 : \beta_2 - \beta_3 = 0 \\ H_a : \beta_2 - \beta_3 \neq 0 \end{array} \quad F = \frac{(R\hat{\beta} - r)'[R(X'X)^{-1}R']^{-1}(R\hat{\beta} - r)/q}{SCR/(T-K)} \overset{H_0}{\sim} F_{(q,T-K)}$$

donde si $\quad H_0 : R\beta = r \qquad R = \begin{bmatrix} 0 & 1 & -1 \end{bmatrix} \qquad r = \begin{bmatrix} 0 \end{bmatrix}$

$$\widehat{\beta} = \begin{bmatrix} 502,36 \\ 2,172 \\ 1,562 \end{bmatrix} \quad ; \quad (T-K) = 150 - 3 = 147 \quad ; \quad F^{0,05}_{(1,147)} \simeq 3,9$$

Rechazaría la H_0 en favor de la H_a con un nivel de significación del 5 % si F>3,9.

Otra forma:

Sea $\quad w = \beta_2 - \beta_3$ de modo que $\Rightarrow$

$$\begin{array}{l} H_0 : w = 0 \\ H_a : w \neq 0 \end{array} \qquad t = \frac{\widehat{w}}{\widehat{\sigma(\hat{w})}} \overset{H_0}{\sim} t_{T-K} \quad \text{donde} \quad \hat{w} = \hat{\beta}_2 - \hat{\beta}_3 = 0,61$$

$$\hat{\sigma^2}_{\hat{w}} = \hat{\sigma^2}_{\hat{u}}(a_{22} + a_{33} - 2a_{23}) \quad \text{y} \quad t^{0,025}_{147} \simeq 1,97$$

donde a_{ij} es el elemento ij de la matriz $(X'X)^{-1}$. Se rechaza la H_0 con un nivel de significatividad del 5 % si $\mid t \mid > 1,97$

Con los mismos datos se ha obtenido también la siguiente estimación:

$$\hat{P}_i = 225{,}4 + 1{,}853T_i \qquad SCR = 4726{,}14$$

donde T_i es la tasación total ($T_i = C_i + M_i$) de la vivienda.

b) Realiza el contraste del apartado a).

Solución: El modelo restringido (MR) por $H_0 : \beta_2 = \beta_3$ es:

$P_i = \beta_1 + \beta_2 C_i + \beta_2 M_i + U_{Ri} \Rightarrow P_i = \beta_1 + \beta_2 T_i + U_{Ri}$

$$F = \frac{\hat{U}_R'\hat{U}_R - \hat{U}'\hat{U}/q}{\hat{U}'\hat{U}/(T-K/} \overset{H_0}{\sim} F_{(q,T-K)}$$

$$F = \frac{4726{,}14 - 2530{,}76}{\frac{2530{,}76}{150-3}} = 127,52 > 3,9 = F_{(1,147)}^{0,05}$$

Al nivel de significación del 5 % se rechaza la H_0 y se concluye que ambas variables NO tienen el mismo efecto sobre el precio final. $\overset{H_0}{\sim}$

c) ¿Cuál de las dos especificaciones te parece correcta? ¿Por qué? ¿Qué propiedades tienen los estimadores de la especificación no escogida?

Solución: Escogeríamos la primera especificación, es decir, el modelo no restringido porque el modelo del apartado (b) recoge una restricción que no es cierta. Por tanto, los estimadores del segundo modelo son lineales, sesgados y con varianzas menores que en el modelo inicial, pero éstos son lineales e insesgados bajo las hipótesis básicas.

LADE-2002.6 (Jun-2002)

Se quiere estimar un modelo de regresión lineal para la variable endógena Y = consumo de bienes duraderos. Para ello se cuenta con los datos de 100 personas sobre dicho consumo (Y) y sobre dos posibles variables explicativas: la renta (X) y la riqueza (Z).

a) En el caso de que no se supiera la especificación correcta del modelo, discute las **ventajas e inconvenientes** de:

a.1) Incluir únicamente una de las dos variables explicativas. ¿Cómo podrías decidir si incluir la renta o la riqueza?

Solución:

Ventajas:

- Simplicidad del modelo.
- Si $r_{xz} \simeq 1$ (algo que podría suceder por la fuerte relación entre renta y riqueza) se evitan los problemas de un alto grado de multicolinealidad ($\hat{\sigma}_{\hat{\beta}_i}$ grandes).

Desventajas:

- Se pierde poder explicativo respecto a la situación en que se incluyen ambas variables.
- Puede haber un problema de omisión de variable relevante $\Rightarrow \hat{\beta}_{MCO}$ sesgados y la inferencia no sería válida.

El criterio para elegir la variable: mejor ajuste del modelo (mayor R^2).

$$\left.\begin{array}{l} \text{(a)} Y_i = \alpha_1 + \beta_1 X_i + U_{1i} \Rightarrow R_x^2 \\ \text{(b)} Y_i = \alpha_2 + \beta_2 X_i + U_{2i} \Rightarrow R_z^2 \end{array}\right\} \Rightarrow \text{se elige (a) si } R_x^2 > R_z^2 \text{ y viceversa}$$

a.2) Incluir ambas variables explicativas.

$\boxed{\text{Solución:}}$

Ventajas:

- Mayor poder explicativo del modelo.
- Menor probabilidad de omitir variables relevantes.

Inconvenientes:

- Puede haber un alto grado de multicolinealidad si $r_{xz} \simeq 1 \Rightarrow \hat{\beta}_{MCO}$ imprecisas, y habría que tomar con cautela las conclusiones de los contrastes de significatividad individual.

b) Los resultados de la estimación por MCO del modelo anterior son:

$$\underset{}{\widehat{Y}_t} = \underset{(2,0)}{3,32} + \underset{(0,75)}{0,1632}\, X_t - \underset{(-3,2)}{0,2931}\, Z_t \quad R^2 = 0{,}92 \qquad \text{(a)}$$

$$\widehat{X}_t = \underset{(-1,5)}{8,31} + \underset{(5,3)}{3,01}\, Z_t \qquad \text{(b)}$$

¿Crees que puede existir algún problema de especificación ?

$\boxed{\text{Solución:}}$ Con los resultados de la regresión (b) se puede obtener r_{xz}:

$$\left(\frac{\hat{\beta}_2}{\hat{\sigma}_{\hat{\beta}_2}}\right)^2 = 6,5^2 = \frac{r_{xz}^2/1}{(1-r_{xz}^2)/(T-K)} \Rightarrow r_{xz} = 0,55$$

Hay un grado alto de relación lineal entre X y Z. Este hecho puede ocasionar que exista multicolinealidad de grado alto en la regresión (a) y hay síntomas de ello:

- Ni X ni Z son individualmente significativas en (a).

$$|\, t \,| < t_{(100-3)}^{\frac{0,05}{2}} = t_{97}^{0,025} \simeq 1,96$$

- X y Z son conjuntamente significativas en (a).

$$\begin{array}{l} H_0 : \beta_2 = \beta_3 = 0 \\ H_a : \beta_2 \neq 0 \quad \text{y/o} \quad \beta_3 \neq 0 \end{array} \qquad F = \frac{R^2/q}{(1-R^2)/T-K} \overset{H_0}{\sim} F_{(q,T-K)}$$

$$\frac{R^2/2}{(1-R^2)/97} = 557,75 > F_{(q,T-K)}^{0,05} \simeq 3,1$$

Al nivel de significación del 5 % las variables explicativas son conjuntamente significativas.

LE-2003.1 (Feb-2003)

Una empresa dedicada a la fabricación de juguetes quiere obtener una relación entre sus ventas (Y) (medida en millones de euros), un índice del precio medio de los juguetes (P) y la cantidad destinada a publicidad (R). Con este objeto se plantea la siguiente especificación:

$$Y_t = \beta_0 + \beta_1 P_t + \beta_2 R_t + u_t \qquad u_t \sim N(0, \sigma^2) \tag{1}$$

Basándote en la información muestral disponible, que corresponde a datos mensuales para seis años completos, y se resume en:

$$(X'X)^{-1} = \begin{bmatrix} 0,863 & -0,0566 & -0,0519 \\ -0,0566 & 0,00526 & 0,00135 \\ -0,0519 & 0,00135 & 0,00616 \end{bmatrix} \qquad \begin{array}{ll} \sum Y_t = 20929,6 & \sum Y_t P_t = 188617,3 \\ \sum Y_t R_t = 139648,8 & \sum Y_t^2 = 6302232,6 \end{array}$$

a) Estima el modelo propuesto por el método de mínimos cuadrados ordinarios.

Solución:

$$\widehat{\beta} = (X'X)^{-1}X'Y = \begin{bmatrix} 0,863 & -0,0566 & -0,0519 \\ -0,0566 & 0,00526 & 0,00135 \\ -0,0519 & 0,00135 & 0,00616 \end{bmatrix} \begin{bmatrix} 20929,6 \\ 188617,3 \\ 139648,8 \end{bmatrix}$$

$$\widehat{\beta} = \begin{bmatrix} 138,73 \\ -3,96 \\ 28,62 \end{bmatrix} \Rightarrow \hat{Y}_t = 138,73 - 3,96 P_t + 28,62 R_t$$

b) Calcula una medida de la bondad del ajuste e interprétala.

Solución:

$$R^2 = 1 - \frac{\sum_{t=1}^T \hat{u}_t{}^2}{\sum_{t=1}^T (Y_t - \bar{Y})^2}$$

$$\sum_{t=1}^T \hat{u}_t{}^2 = \sum_{t=1}^T Y_t{}^2 - \widehat{\beta}' X'Y = 6302232,6 - 6153387,5 = 148845,04$$

$$\bar{Y} = \frac{20929,6}{72} = 290,69 \qquad \sum_{t=1}^T (Y_t - \bar{Y})^2 = 6302232,6 - 72(290,69)^2$$

[0] CVS Id: $Id: 03e1g.tex,v 1.1 2003/09/18 07:47:11 etpdihei Exp

$$\Rightarrow R^2 = 1 - \frac{148845,04}{218183,92} = 0,317$$

Que el valor de R^2 sea igual a 0,317 significa que a través del modelo se explica el $31,7\,\%$ de las variaciones que experimentan las ventas de juguetes de esa empresa. Esta es la medida de la capacidad explicativa de las variables que se han incorporado al modelo: El índice de precios y el gasto en publicidad.

c) Estima la matriz de varianzas y covarianzas del estimador.

Solución:

$\widehat{V(\widehat{\beta})} = \widehat{\sigma^2}(X'X)^{-1}$ siendo $\widehat{\sigma^2} = \dfrac{\hat{u}'\hat{u}}{T-k-1} = \dfrac{148845,24}{69} = 2157,17 \Rightarrow \widehat{\sigma} = 46,44$

d) ¿Son las variables explicativas conjuntamente significativas?

Solución: La hipótesis que debemos contrastar es la siguiente:

$$H_0 : \beta_1 = \beta_2 = 0$$
$$H_a : \beta_1 \neq 0 \quad \text{y/o} \quad \beta_2 \neq 0$$

Para llevar a cabo el contraste, podemos utilizar el estadístico:

$F = \frac{R^2}{1-R^2}\frac{T-k-1}{m}$, donde m es el número de restricciones a contrastar. El estadístico sigue, bajo la hipótesis nula, la distribución $F(m, T-k-1)$.

Para este caso,

$F = \frac{0{,}317}{1-0{,}317}\frac{69}{2} = 16,01$; $F^{0,05}(2,69) \simeq 3,15 \Rightarrow 16,01 \geq 3,15$, la hipótesis nula se rechaza. Esto significa que, conjuntamente, las dos variables explicativas de la ecuación son significativas.

e) Contrasta al nivel de significación del $5\,\%$ que el precio de los juguetes afecta al nivel de ventas.

Solución: Se trata de contrastar

$$H_0 : \beta_1 = 0 \qquad H_a : \beta_1 \neq 0 \qquad \text{usando el estadístico} \quad t = \frac{\widehat{\beta_1}}{\hat{\sigma}(\widehat{\beta_1})}$$

que, bajo la hipotesis nula, se distribuye como una t de Student con 69 grados de libertad.

Dados los resultados obtenidos anteriormente, $t = \left| \frac{-3{,}96}{\sqrt{2157{,}17}\sqrt{0{,}000526}} \right| = 1,17$; $t^{0,025}(69) \simeq 2 \Rightarrow 1,17 \leq 2 \Rightarrow$ Se acepta Ho, el precio de los juguetes no resulta significativo para explicar las ventas.

f) Interpreta el resultado anterior, teniendo en cuenta que la correlación muestral entre los dos regresores, P y R, es igual a -0,23.

Solución: Aunque el modelo no tiene demasiada capacidad explicativa, aproximadamente el $32\,\%$ de las variaciones de Y, las dos variables conjuntamente son significativas. Individualmente, el precio (P) no lo es. Esto podría deberse, de estar R y P muy correlacionadas, a que al estar R en el modelo, la otra variable no añade suficiente información nueva, lo que representaría un problema de colinealidad. Sin embargo, dado que la correlación entre los dos regresores sólo es de $-0,23$, la colinealidad no es la causa del resultado obtenido. Simplemente, resulta que para estos datos el precio de los juguetes no parece una variable importante para explicar las ventas.

LE-2003.2 (Feb-2003)

El gerente de la empresa anterior sabe que en la época de Navidad el número de ventas se incrementa. Por esa razón, piensa que podría ser más adecuado el modelo:

$$Y_t = \alpha_0 + \alpha_1 N_t + \alpha_2 P_t + \alpha_3 R_t + u_t \tag{2}$$

donde N_t toma valor uno cuando el mes t es noviembre y diciembre y cero en otro caso.

a) Interpreta los coeficientes del modelo. ¿Qué diferencia existe entre las especificaciones (1) y (2)?

Solución: Dado que la única diferencia entre ambos modelos está en la inclusión de la variable N, que es dicotómica, podemos interpretar los coeficientes a través de

$$E(Y_t \mid N_t = 1) = \alpha_0 + \alpha_1 + \alpha_2 P_t + \alpha_3 R_t$$
$$E(Y_t \mid N_t = 0) = \alpha_0 + \alpha_2 P_t + \alpha_3 R_t$$

$\Rightarrow$ Por lo tanto, α_1 mide la diferencia en ventas esperadas en los meses de Navidad respecto del resto de los meses del año, para un mismo valor de P y R. $\Rightarrow$ El modelo (1) no tiene en cuenta este efecto de la Navidad en las ventas, que en el modelo (2) se incorpora a través de la variable N.
α_0 = Ventas medias para P=R=N=0, es decir, cuando las variables explicativas son cero y no estamos en Navidad.
α_2 (α_3) = incremento que se produce en las ventas ante un aumento de una unidad en P (R),*ceteris paribus* (si el resto de las variables permanecen constantes).

b) Contrasta si el efecto Navidad es significativo para un nivel de significatividad del $5\,\%$, dados los resultados siguientes:

$$\hat{Y}_t = \underset{(15,6)}{184,0} + \underset{(6,31)}{73,34} N_t - \underset{(1,17)}{3,37} P_t + \underset{(1,54)}{19,50} R_t \quad R^2 = 0,918$$

siendo en este caso la matriz inversa igual a

$$(X'X)^{-1} = \begin{bmatrix} 0,9337 & 0,1039 & -0,0560 & -0,0664 \\ 0,1039 & 0,1527 & 0,00088 & -0,0214 \\ -0,0560 & 0,00088 & 0,00527 & 0,00122 \\ -0,0664 & -0,0214 & 0,00122 & 0,00916 \end{bmatrix}$$

Solución: La hipótesis a contrastar en este caso es

$$\begin{array}{l} H_0 : \alpha_1 = 0 \\ H_a : \alpha_1 \neq 0 \end{array} \qquad \text{usando el estadístico} \quad t = \frac{\widehat{\alpha_1}}{\sigma(\widehat{\alpha_1})}$$

que, bajo la hipotesis nula, se distribuye como una t de Student con 68 grados de libertad.

Dados los resultados obtenidos anteriormente, $t = \mid \frac{73{,}34}{6{,}31} \mid = 11,62$; $t^{0{,}025}(68) \simeq 2 \Rightarrow 11,62 \geq 2 \Rightarrow$ Se rechaza la hipótesis nula $\Rightarrow$ El efecto de la Navidad sobre las ventas de juguetes es significativo.

c) De acuerdo con el discurso de fin de año del Presidente, durante la próxima campaña de Navidad, en el mes de diciembre, la empresa piensa gastar en publicidad 15 millones de euros, mantener el precio de los juguetes en 9 unidades y aspira a vender 550 millones de euros en juguetes. ¿Crees que el presidente está siendo realista con estos datos? Explica cuidadosamente cómo podrías responder a esta pregunta.

Solución: Si llevamos a cabo una predicción para el mes de diciembre utilizando el modelo (2), tendremos:

$$\widehat{Y_p} = 184,0 + 73,34 - 3,37x9 + 19,50x15 = 519,5$$

que es la predicción por punto del modelo para la situación que se ha descrito en el enunciado. Sin embargo, para poder saber si el presidente tiene o no razón, tendríamos que calcular un intervalo de confianza para esa predicción. Para ello, debemos calcular

$$Pr[\widehat{Y_p} - t^{0{,}025}(68)\widehat{\sigma}\sqrt{1 + x_p'(X'X)^{-1}x_p} \leq Y_p \leq \widehat{Y_p} - t^{0{,}025}(68)\widehat{\sigma}\sqrt{1 + x_p'(X'X)^{-1}x_p}] = 0{,}95$$

donde

$$x_p' = \begin{bmatrix} 1 & 1 & 9 & 15 \end{bmatrix}$$

representa los valores de las variables explicativas para el período de predicción. La expresión $1+x_p'(X'X)^{-1}x_p$ toma el valor de $1,4852$ usando el valor de x_p y la matriz inversa proporcionada más arriba. Asimismo, es preciso calcular el valor de $\widehat{\sigma}$ para el modelo (2). Puede calcularse su valor a partir de la expresión

$$\widehat{\sigma^2} = \frac{\hat{u}'\hat{u}}{T-k-1} = \frac{(1-R^2)\sum_{t=1}^{T}(Y_t - \bar{Y})^2}{T-k-1} \Rightarrow \frac{(1-0,918)218183,92}{68} = 263,1 \rightarrow \widehat{\sigma} = 16,22$$

Por lo tanto, el intervalo de confianza requerido es el siguiente:

$$Pr[519,5 - 2\sqrt{263,1x1,4852} \leq Y_p \leq 519,5 + 2\sqrt{263,1x1,4852}] = 0{,}95$$

$$Pr[480,38 \leq Y_p \leq 558,6] = 0{,}95$$

De donde se deduce que el presidente tiene razón y puede ser que se vendan 550 millones de euros en juguetes en la próxima campaña de Navidad, dado que esta cifra pertenece al intervalo de confianza calculado.

LE-2003.3 (Feb-2003)

El director de marketing de la misma empresa considera que el modelo anterior no hace justicia a la eficacia de la publicidad, ya que, como se hace más publicidad en Navidad, debe tenerse en cuenta explícitamente este hecho. Por ello, propone y estima el siguiente modelo:

$$\hat{Y}_t = \underset{(15,33)}{204,48} - \underset{(46,6)}{99,53}N_t - \underset{(1,11)}{4,44}P_t + \underset{(5,30)}{38,62}R_tN_t + \underset{(1,49)}{17,74}R_tA_t \quad R^2 = 0,932 \tag{3}$$

donde A_t es una variable ficticia complementaria de N_t, esto es, A_t toma valor cero en noviembre y diciembre y valor uno en los restantes meses del año.

a) Escribe las doce primeras observaciones de la matriz X para el modelo (3), sabiendo que los valores de P y R han sido los siguientes:

P_t	9	8	11	9	10	10	9	7	9	9	8	10
R_t	44,7	45,3	46,2	46,6	48,2	48,6	50,5	51,3	58,1	61,9	72,2	81,4

Solución: Las observaciones serán las siguientes:

$$X = \begin{bmatrix} 1 & 0 & 9 & 0 & 44,7 \\ 1 & 0 & 8 & 0 & 45,3 \\ 1 & 0 & 11 & 0 & 46,2 \\ 1 & 0 & 9 & 0 & 46,6 \\ 1 & 0 & 10 & 0 & 48,2 \\ 1 & 0 & 10 & 0 & 48,6 \\ 1 & 0 & 9 & 0 & 50,5 \\ 1 & 0 & 7 & 0 & 51,3 \\ 1 & 0 & 9 & 0 & 58,1 \\ 1 & 0 & 9 & 0 & 61,9 \\ 1 & 1 & 8 & 72,2 & 0 \\ 1 & 1 & 10 & 81,4 & 0 \end{bmatrix}$$

b) Interpreta los coeficientes del modelo (3).

Solución: Los coeficientes que acompañan a las variables N y P, así como el término constante, tienen el mismo significado ene ste caso que en el modelo anterior. Por lo tanto, el modelo nos dice que se estima un decrecimiento de las ventas de -4,44 millones de euros por cada unidad de aumento en el precio, y un descenso en el volumen de ventas en los meses de noviembre y diciembre con respecto al resto de meses del año. Este es un resultado extraño, que puede explicarse si miramos a los valores esperados

$$E(Y_t \mid N_t = 1) = \gamma_0 + \gamma_1 + \gamma_2 P_t + \gamma_3 R_t$$
$$E(Y_t \mid N_t = 0) = \gamma_0 + \gamma_2 P_t + \gamma_4 R_t$$

Por lo tanto, γ_3 y γ_4, que son los coeficientes de las variables RN y RA, indican respectivamente el efecto que tiene el gasto en publicidad sobre las ventas, dependiendo de que se trate de meses navideños (aumento de $38,62$ millones en ventas por cada millón destinado a publicidad, si el precio permanece constante) o de meses no navideños (aumento de $17,74$ millones en ventas). Se observa así que, según el modelo, las ventas aumentan por efecto de la Navidad, pero lo hacen a través de su efecto en el gasto en publicidad.

c) Contrasta la hipótesis de que no existe interrelación entre el efecto de la Navidad y el efecto de la publicidad.

Solución: Si no existiera interrelación entre ambas variables, ello significaría que la publicidad y la Navidad tienen efectos completamente independientes entre sí, por lo que el efecto de la publicidad sobre las ventas sería el mismo con independencia de si estamos en época navideña o no. Formalmente, esto significa que

$$H_0 : \gamma_3 = \gamma_4$$
$$H_a : \gamma_3 \neq \gamma_4$$

Bajo la hipótesis nula, por lo tanto, el modelo correcto (Modelo restringido, MR) es el (2) mientras que bajo la alternativa el modelo correcto (Modelo no restringido, MNR) es el (3). Podemos llevar a cabo el contraste utilizando el estadístico $F = \frac{R^2_{NR} - R^2_R}{1 - R^2_{NR}} \frac{T-k-1}{m}$, donde m es el número de restricciones a contrastar. El estadístico sigue, bajo la hipótesis nula, la distribución $F(m, T-k-1)$.

Para este caso,

$F = \frac{0{,}932 - 0{,}918}{1 - 0{,}932} \frac{67}{1} = 13,79$; $F^{0{,}05}(1,67) \simeq 4 \Rightarrow 13,79 \geq 4 \Rightarrow$ la hipótesis nula se rechaza. Existe por lo tanto evidencia de que la publicidad y la Navidad se relacionan entre sí, con un efecto conjunto diferente de la suma de sus efectos individuales.

d) Contrasta en el modelo (3) que la Navidad no tiene ningún efecto sobre las ventas de esta empresa.

Solución: Si la Navidad no presenta ningún efecto, ello significa que la variable N no tiene que aparecer en el modelo de ninguna manera. En este caso, el modelo (3),(No Restringido), quedaría reducido al (1), (Restringido), por lo que podemos comparar entre sí estos dos modelos para contrastar la siguiente hipótesis nula:

$$H_0 : \gamma_1 = 0; \gamma_3 = \gamma_4$$
$$H_a : \gamma_1 \neq 0; \gamma_3 \neq \gamma_4$$

El estadístico sería el mismo que se ha escrito en el apartado anterior, aunque ahora el modelo restringido es el (1) y hay dos restricciones a contrastar, por lo que $m = 2$. Por lo tanto,

$F = \frac{0{,}932 - 0{,}317}{1 - 0{,}932} \frac{67}{2} = 302,97$; $F^{0{,}05}(2,67) \simeq 3,15 \Rightarrow 302,97 \geq 3,15 \Rightarrow$ Se rechaza la hipótesis nula; la Navidad tiene un claro efecto sobre las ventas de juguetes.

e) Dados los resultados obtenidos en el modelo (3) ¿qué propiedades tienen los estimadores y qué validez los contrastes que has realizado en el Ejercicio 1?

Solución: Dado que el modelo (1) incluye restricciones que no son ciertas (puesto que en él $\gamma_1 = 0$ y $\gamma_3 = \gamma_4$, restricciones que se rechazan, como acabamos de ver), esto significa que en ese modelo estamos cometiendo un error de especificación, por exclusión de variables relevantes. Tal como nos dice la teoría, en este caso tendremos $E(\widehat{\beta_I}) = \beta_I + (X_I'X_I)^{-1}X_I'X_{II}\beta_{II} \neq \beta_I$ salvo que las variables incluídas en el modelo (X_I) y las excluidas erróneamente (X_{II}) sean ortogonales entre sí. En este caso, no se produce ortogonalidad, por lo que los estimadores del modelo (1) serán sesgados. De la misma manera, la teoría nos dice que ante la omisión de variables relevantes en un modelo la suma de cuadrados de los residuos no sigue la distribución χ^2, con lo que los estadísticos que se utilizan para llevar a cabo los contrastes en los apartados 4. y 5. del ejercicio 1 no siguen las distribuciones que les atribuímos, invalidando los resultados que se han obtenido.

LE-2003.4 (Jun-2003)

Un investigador quiere analizar el mercado del automóvil en la Comunidad Autónoma Vasca. Con este objetivo propone la siguiente especificación:

$$P_t = \alpha_0 + \alpha_1 T_t + \alpha_2 C_t + u_t \qquad u_t \sim NID(0, \sigma^2) \tag{1}$$

donde

P_t es el precio final de venta, en logaritmos, del t-ésimo modelo.

T_t es la tara, en toneladas, del t-ésimo modelo.

C_t mide los caballos de potencia (CV), en cientos, del modelo t.

A continuación se resume la información muestral disponible correspondiente a 485 observaciones tomadas de distintos concesionarios.

$$(X'X)^{-1} = \begin{bmatrix} 0,9556 & -4,3345 & 4,8594 \\ -4,3345 & 20,2385 & -22,8140 \\ 4,8594 & -22,8140 & 25,7455 \end{bmatrix} \qquad \begin{array}{ll} \sum P_t = 1997,619 & \sum P_t C_t = 1973,397 \\ \sum P_t T_t = 2652,534 & \sum (P_t - \bar{P})^2 = 281,053 \end{array}$$

a) Estima el modelo propuesto por el método de mínimos cuadrados ordinarios.

Solución:

$$\begin{aligned} \widehat{\alpha} &= \begin{bmatrix} \hat{\alpha}_1 \\ \hat{\alpha}_2 \\ \hat{\alpha}_3 \end{bmatrix} = (X'X)^{-1}X'P = \\ & \begin{bmatrix} 0,9556 & -4,3345 & 4,8594 \\ -4,3345 & 20,2385 & -22,8140 \\ 4,8594 & -22,8140 & 25,7455 \end{bmatrix} \begin{bmatrix} 1997,619 \\ 2652,534 \\ 1973,397 \end{bmatrix} = \begin{bmatrix} 1,04 \\ 3,55 \\ -1,58 \end{bmatrix} \end{aligned}$$

$$\widehat{P_t} = 1,58 + 3,55T_t - 1,58C_t$$

b) Calcula una medida de la bondad del ajuste e interprétala.

Solución:

$$R^2 = \frac{\hat{\beta}' X' P - T\bar{P}^2}{\sum_{i=1}^{485}(P_t - \bar{P})^2} =$$

$$\frac{(1,04 \;\; 3,55 \;\; -1,58)\begin{bmatrix} 1997,619 \\ 2652,534 \\ 1973,397 \end{bmatrix} - 485(1997,619/485)^2}{281,053} =$$

$$\frac{148,25}{281,053} = 0,527$$

Que el valor de R^2 sea igual a 0,527 significa que a través de las variables explicativas T (tara) y C (potencia) se explica el $52,7\,\%$ de las variaciones que experimenta el precio final de venta del automóvil.

c) Estima la matriz de varianzas y covarianzas del estimador de los coeficientes.

Solución:

$$\widehat{V}(\widehat{\beta}) = \widehat{\sigma}^2 (X'X)^{-1} \qquad \text{siendo} \qquad \widehat{\sigma}^2 = \frac{\hat{u}'\hat{u}}{T-k-1} = \frac{132,803}{485-2-1} = 0,2755$$

$$\hat{u}'\hat{u} = SCT - SCE = \sum (P_t - \bar{P})^2 - \sum (\hat{P}_t - \bar{P})^2 = 281,053 - 148,25 = 132,803$$

$$\widehat{V}(\widehat{\beta}) = 0,2755 (X'X)^{-1} = \begin{pmatrix} 0,2632 & -1,1941 & 1,3387 \\ & 5,5757 & -6,2852 \\ & & 7,0928 \end{pmatrix}$$

d) ¿Son las variables explicativas conjuntamente significativas?

Solución:

La hipótesis que debemos contrastar es la siguiente:

$$H_0 : \alpha_1 = \alpha_2 = 0$$
$$H_a : \alpha_1 \neq 0 \quad \text{y/o} \quad \alpha_2 \neq 0$$

Para llevar a cabo el contraste, podemos utilizar el estadístico:

$F = \frac{R^2}{1-R^2}\frac{T-k-1}{m}$, donde m es el número de restricciones a contrastar. El estadístico sigue, bajo la hipótesis nula, la distribución $F(m, T-k-1)$.

Para este caso,

$F = \frac{0,527}{1-0,527}\frac{485-3}{2} = 268,5; \quad F^{0,05}(2,482) \simeq 3 \Rightarrow 268,5 > 3$, la hipótesis nula se rechaza. Esto significa que, las dos variables explicativas de la regresión son conjuntamente significativas a un nivel de significación del $5\,\%$.

e) Contrasta al nivel de significación del $5\,\%$ que la potencia del automóvil influye sobre su precio.

Solución:

Se trata de contrastar

$$\begin{array}{l} H_0 : \alpha_2 = 0 \\ H_a : \alpha_2 \neq 0 \end{array} \qquad \text{usando el estadístico} \quad t = \frac{\widehat{\alpha_2}}{\hat{\sigma}_{\hat{\alpha_2}}}$$

que, bajo la hipótesis nula, se distribuye como una t de Student con (485-3) grados de libertad.

Dados los resultados obtenidos anteriormente, $t = \mid \frac{-1{,}58}{\sqrt{0{,}2755 \times 25{,}7455}} \mid = 0,59$; $t^{0{,}05}(482) \simeq 1,96 \quad \Rightarrow 0,59 < 1,96 \quad \Rightarrow$ No se rechaza la Ho a un nivel de significación del $5\,\%$, la potencia del coche no influye sobre su precio final de venta.

f) Dado que los coches con mayor tara necesitan más potencia, ¿crees que podrían existir problemas de multicolinealidad? Razona tu respuesta.

Solución:

El ajuste no es muy bueno y la potencia no es individualmente significativa. Si realizamos el contraste de significatividad de la variable T:

$$\begin{array}{l} H_0 : \alpha_1 = 0 \\ H_a : \alpha_1 \neq 0 \end{array} \qquad \text{usando el estadístico} \quad t = \frac{\widehat{\alpha_1}}{\hat{\sigma}_{\hat{\alpha_1}}}$$

que, bajo la hipótesis nula, se distribuye como una t de Student con (485-3) grados de libertad.

Dados los resultados obtenidos anteriormente, $t = \mid \frac{3{,}55}{\sqrt{0{,}2755 \times 20{,}2385}} \mid = 1,5$; $t^{0{,}05/2}(482) \simeq 1,96 \quad \Rightarrow 1,5 < 1,96 \quad \Rightarrow$ No se rechaza la Ho a un nivel de significación del $5\,\%$, la tara del coche no es una variable individualmente significativa para explicar el precio final de venta de un coche. Sin embargo, conjuntamente ambas variables tienen capacidad explicativa. Parece claro que hay un problema de multicolinealidad imperfecta. Al estar ambas variables correlacionadas, ninguna de ellas aporta información relevante sobre la proporcionada por la otra variable, de ahí que no sean individualmente significativas.

LE-2003.5 (Jun-2003)

Más tarde el investigador cree que el tipo de combustible (gasoil o gasolina) que consume el automóvil puede ser una variable relevante para determinar el precio. Con este propósito se define D_t que toma el valor uno si el t-ésimo automóvil consume gasoil y cero en caso contrario. Los resultados de incluir esta variable son los siguientes:

$$\underset{(\hat{\sigma}_{\hat{\beta}_i})}{\hat{P}_t} = \underset{(0,2869)}{\overbrace{0,5649}^{\hat{\beta}_1}} + \underset{(1,319)}{\overbrace{3,3251}^{\hat{\beta}_2}} T_t + \underset{(1,4879)}{\overbrace{1,7106}^{\hat{\beta}_3}} C_t + \underset{(0,025)}{\overbrace{0,8312}^{\hat{\beta}_4}} D_t \quad \sum_{t=1}^{485} \hat{u}_t^2 = 41,6674 \tag{2}$$

a) Interpreta los coeficientes del modelo. ¿Qué diferencia existe entre las especificaciones (1) y (2)?

Solución:

⇒ El intercepto β_1 puede interpretarse a través de

$$E(P_t|T_t = C_t = D_t = 0) \quad = \quad \beta_1$$

es decir, el precio de venta esperado de un hipotético automóvil de gasolina con $T = C = 0$.
⇒ Por otro lado, dado que la única diferencia entre ambos modelos está en la inclusión de la variable D, que es dicotómica, podemos interpretar su coeficiente a través de

$$\begin{aligned} E(P_t|D_t = 1) &= \beta_1 + \beta_2 T_t + \beta_3 C_t + \beta_4 \\ E(P_t|D_t = 0) &= \beta_1 + \beta_2 T_t + \beta_3 C_t \\ \beta_4 &= E(P_t|D_t = 1) - E(P_t|D_t = 0) \end{aligned}$$

esto es, β_4 mide la diferencia en el precio final de venta esperado para un automóvil gasoil respecto de uno de gasolina, con un mismo valor de T y C.
⇒ Por lo demás, los coeficientes β_2 y β_3 indican el incremento que se produce en las ventas esperadas ante un aumento de una unidad en T y C respectivamente, siempre y cuando el resto de las variables permanezcan constantes.
⇒ El modelo (1) no tiene en cuenta el efecto del tipo de combustible en el precio final de venta de los automóviles, el cual se incorpora en el modelo (2) a través de la variable ficticia D.

b) Contrasta si la variable “tipo de combustible” es significativa y comenta las consecuencias de tu respuesta sobre los resultados obtenidos en el primer ejercicio.

Solución:

La hipótesis a contrastar en este caso es

$$\begin{array}{l} H_0 : \beta_4 = 0 \\ H_a : \beta_4 \neq 0 \end{array} \qquad \text{usando el estadístico} \quad t = \frac{\widehat{\beta}_4}{\widehat{\sigma}_{\beta_4}}$$

el cual, bajo la hipótesis nula, se distribuye como una t de Student con obs − params $= 481$ grados de libertad.

Dados los resultados obtenidos,
$t = \mid \frac{0{,}8312}{0{,}025} \mid = 33,25; \quad t_{0{,}025}(481) = 1,96 \quad \Rightarrow 33,25 \geq 1,96 \quad \Rightarrow$ Se rechaza la hipótesis nula ⇒ El efecto del tipo de combustible sobre el precio final de venta de automóviles es significativo.
⇒ Por tanto, la especificación (1) incurre en un error de especificación en la forma de omisión de una variable re
(tipo de combustible). En consecuencia, las estimaciones obtenidas en el primer ejercicio serán sesgadas y los contrastes realizados serán inválidos.

c) Una posible especificación alternativa a (2) para recoger el efecto del tipo de combustible es

$$P_t \quad = \quad \gamma_1 D_t + \gamma_2 G_t + \gamma_3 C_t + \gamma_4 T_t + u_t$$

donde G_t toma el valor uno si el automóvil consume gasolina y cero en caso contrario.

Estima la nueva especificación e interpreta los coeficientes γ_1 y γ_2.

Solución:

⇒ Dado que la única diferencia entre ambos modelos está en la substitución del intercepto por la nueva variable ficticia G, que es dicotómica y complementaria de D, tenemos que

$$\begin{aligned} E(P_t|D_t = 1, G_t = 0, C_t = T_t = 0) &= \beta_1 + \beta_4 = \gamma_1 \\ E(P_t|D_t = 0, G_t = 1, C_t = T_t = 0) &= \beta_1 = \gamma_2 \end{aligned}$$

de modo que

$$\begin{aligned} \hat{\gamma}_1 &= \hat{\beta}_1 + \hat{\beta}_4 = 0,5649 + 0,8312 = 1,3961 \\ \hat{\gamma}_2 &= \hat{\beta}_1 = 0,5649 \end{aligned}$$

$\Rightarrow$ Podemos interpretar los nuevos coeficientes a través de

$$\begin{aligned} E(P_t|D_t = 1, C_t = T_t = 0) &= \gamma_1 \\ E(P_t|G_t = 1, C_t = T_t = 0) &= \gamma_2 \end{aligned}$$

esto es, γ_1 y γ_2 son los interceptos correspondientes respectivamente a los automóviles bien de gasoil o bien de gasolina. En otras palabras, γ_1 es el precio de venta esperado de un hipotético automóvil de gasoil con $T = C = 0$ mientras que γ_2 es el precio de venta esperado de un hipotético automóvil de gasolina con $T = C = 0$.

LE-2003.6 (Jun-2003)

Finalmente el investigador especifica un modelo en el que se añade una nueva variable que clasifica el tipo de automóvil según su tamaño: utilitario, mediano o grande. Por ello, propone y estima el siguiente modelo:

$$\underset{(\hat{\sigma}_{\hat{\beta}_i})}{\hat{P}_t} = \underset{(1,4638)}{\overbrace{2,6895}^{\hat{\delta}_0}} \underset{(1,2253)}{\overbrace{-2,3924}^{\hat{\delta}_1}} T_t + \underset{(1,3216)}{\overbrace{3,3367}^{\hat{\delta}_2}} C_t + \underset{(0,0182)}{\overbrace{0,8318}^{\hat{\delta}_3}} D_t + \underset{(0,0446)}{\overbrace{0,7969}^{\hat{\delta}_4}} M_t + \underset{(0,0767)}{\overbrace{0,8416}^{\hat{\delta}_5}} L_t \quad \sum_{i=1}^{485} \hat{u}_t^2 = 22,0501 \quad (3)$$

donde M_t es una variable ficticia que toma valor uno si el automóvil t es de tamaño mediano y cero en caso contrario y L_t toma valor uno si el automóvil t es de tamaño grande y cero en caso contrario.

La matriz de varianzas y covarianzas estimada correspondiente al estimador MCO de los coeficientes de este modelo es

$$\hat{V}_{\hat{\beta}} = \hat{\sigma}^2 (X'X)^{-1} = \hat{\sigma}^2 \begin{bmatrix} 2,1437 & -8,1577 & 8,6386 & -3,9805 & 0,1889 & 0,3878 \\ & 32,6410 & -35,1173 & 0,0016 & -0,5761 & -1,2394 \\ & & 37,9734 & -0,0011 & 0,5520 & 1,2195 \\ & & & 0,0072 & 0,000006 & -0,00006 \\ & & & & 0,00433 & 0,0660 \\ & & & & & 0,1281 \end{bmatrix}$$

a) ¿Cuál es el precio medio estimado de un utilitario de 55CV y 970kg que consume gasolina? ¿Y si consume gasoil?

Solución:

Dado el modelo estimado:

$$\hat{P}_t = 2,6895 - 2,3924T_t + 3,3367C_t + 0,8318D_t + 0,7969M_t + 0,8416L_t$$

Si consume gasolina:

$$\hat{P}_t = 2,6895 - 2,3924 \times 0,970 + 3,3367 \times 0,55 + 0,8318 \times 0 + 0,7969 \times 0 + 0,8416 \times 0 = 2,2040$$

Si consume gasoil:

$$\hat{P}_t = 2,6895 - 2,3924 \times 0,970 + 3,3367 \times 0,55 + 0,8318 \times 1 + 0,7969 \times 0 + 0,8416 \times 0 = 3,0358$$

b) ¿Cuál es la diferencia en el precio medio estimado entre un coche grande y otro mediano?, ¿y si las demás características fueran iguales?

Solución:

Precio medio estimado coche grande, t=L:

$$\hat{P}_L = (2,6895 + 0,8416) - 2,3924T_L + 3,3367C_L + 0,8318D_L$$

Precio medio estimado coche mediano, t=M:

$$\hat{P}_M = (2,6895 + 0,7969) - 2,3924T_M + 3,3367C_M + 0,8318D_M$$

Diferencia:

$$\hat{P}_L - \hat{P}_M = (0,8416 - 0,7969) - 2,3924(T_L - T_M) + 3,3367(C_L - C_M) + 0,8318(D_L - D_M)$$

Si las demás características son iguales:
- misma tara, $T_L = T_M$
- misma potencia, $C_L = C_M$
- mismo combustible, $D_L = D_M$

Entonces la diferencia viene dada solamente por el hecho de tener diferente tamaño:

$$\hat{P}_L - \hat{P}_M = 0,8416 - 0,7969 = 0,0447$$

c) Contrasta la hipótesis de que el tamaño del coche no influye en su precio final.

Solución:

$$H_0 : \delta_4 = \delta_5 = 0$$
$$H_a : \delta_4 \neq \delta_5 = 0$$

Bajo la hipótesis nula, el modelo restringido (MR) es el (1) mientras que bajo la alternativa el modelo no restringido (MNR) es el (3). Podemos llevar a cabo el contraste utilizando el estadístico $F = \frac{\hat{u}'\hat{u}_R - \hat{u}'\hat{u}_{NR}}{\hat{u}'\hat{u}_{NR}} \frac{T-k-1}{m}$, donde m es el número de restricciones a contrastar. El estadístico sigue, bajo la hipótesis nula, la distribución $F(m, T-k-1)$.

Para este caso,

$F = \frac{41{,}6674-22{,}0501}{22{,}0501} \frac{485-5-1}{2} = 213,0756 > 3 \simeq F^{0{,}05}(2,479) \Rightarrow$ la hipótesis nula se rechaza al nivel de significación del $5\,\%$. Existe por lo tanto evidencia de que el tamaño del automóvil sí afecta a su precio de venta final.

d) Contrasta la hipótesis de que en cuanto al tamaño se refiere, solamente influye sobre el precio final de venta el que sea utilitario o no.

Solución:

$$\begin{array}{lll} H_0 : \delta_4 = \delta_5 & H_0 : \lambda = 0 & \text{con} \lambda = \delta_4 - \delta_5 \\ H_a : \delta_4 \neq \delta_5 & H_a : \lambda \neq 0 & \end{array}$$

usando el estadístico $t = \frac{\widehat{\lambda}}{\hat{\sigma}_{\hat{\lambda}}}$ que, bajo la hipótesis nula, se distribuye como una t de Student con (485-6) grados de libertad.

Dados los resultados Obtenidos anteriormente,

$$\hat{\lambda} = \hat{\delta}_4 - \hat{\delta}_5 = 0,7969 - 0,8416 = -0{,}0447$$

$$\hat{\sigma}_{\hat{\lambda}} = \hat{\sigma}^2_{\hat{\delta}_4} + \hat{\sigma}^2_{\hat{\delta}_5} - 2 \times \hat{\sigma}_{\hat{\delta}_4 \hat{\delta}_5} = 0,00433 + 0,1281 - 2 \times 0,0660 = 0,00043$$

$$t = \mid \frac{-0,0447}{\sqrt{0,0207}} \mid = 2,15 > 1,96 \simeq t^{0{,}025}(485-6)$$

Se rechaza la hipótesis nula con un nivel de significación del $5\,\%$, por tanto, el efecto que el tamaño de los coches medianos y grandes tienen sobre el precio final de venta (ceteris paribus) no es el mismo.

e) Dados todos los resultados obtenidos anteriormente, ¿cómo especificarías el modelo? Razona tu respuesta.

Solución: Dado que en el apartado anterior hemos rechazado la hipótesis nula, optaría por la especificación (3), en el que se discrimina los tres tamaños de coche. El estimador MCO de esta especificación es lineal en u, insesgado y de mínima varianza. Las especificaciones (1) y (2) omiten las variables 'tamaño' y 'tipo de combustible' por lo que los estimadores MCO en estas especificaciones son lineales en u pero no son insesgados por lo que tanto la inferencia como la predicción no son válidas. Si por el contrario optamos por el modelo restringido a $\delta_4 = \delta_5$

$$P_t = \delta_0 + \delta_1 T_t + \delta_2 C_t + \delta_3 D_t + \delta_4 MG_t + \delta_5 + u_t \tag{4}$$

donde MG_t es una variable ficticia que toma valor uno si el coche es mediano o grande (no un utilitario) y cero en otro caso. El estimador MCO de este modelo restringido es lineal, sesgado porque la restricción no se satisface.

[0] CVS Id: $Id: 03e1e.tex,v 1.2 2003/09/18 14:26:07 etpdihei Exp

LADE-2003.1 (Feb-2003)

El gobierno de un determinado país quiere realizar un estudio sobre la venta de viviendas. Para ello, propone el siguiente modelo:

$$Y_t = \beta_1 + \beta_2 P_t + \beta_3 I_t + u_t \qquad t = 1, ..., 30 \qquad u_t \sim NID(0, \sigma^2)$$

donde:

Y_t es el número de viviendas vendidas en el año t (medido en miles de unidades).

P_t es el precio medio de la vivienda en el año t (en 10^4 unidades monetarias del país).

I_t es el tipo de interés medio de los préstamos hipotecarios en el año t (medido en $\%$).

Se dispone de la siguiente información muestral para los años 1971 a 2000:

$\bar{Y} = 38,40 \qquad \sum Y_t^2 = 48600,92 \qquad \sum Y_t P_t = 6400,22$

$\bar{P} = 6,51 \qquad \sum P_t^2 = 1466,04 \qquad \sum Y_t I_t = 4462,26$

$\bar{I} = 4,83 \qquad \sum I_t^2 = 893,89 \qquad \sum P_t I_t = 1134,75$

$$(X'X)^{-1} = \begin{pmatrix} 0,725 & -0,326 & 0,296 \\ -0,326 & 0,186 & -0,183 \\ 0,296 & -0,183 & 0,185 \end{pmatrix}$$

a) Estima el modelo propuesto por Mínimos Cuadrados Ordinarios.

Solución:

$$\widehat{\beta}_{MCO} = (X'X)'X'Y = (X'X)^{-1} \begin{pmatrix} \sum_{t=1}^T Y_t \\ \sum_{t=1}^T Y_t P_t \\ \sum_{t=1}^T Y_t I_t \end{pmatrix} =$$

$$(X'X)^{-1} \begin{pmatrix} 1152 \\ 6400,22 \\ 4462,26 \end{pmatrix} = \begin{pmatrix} 69,557 \\ -1,704 \\ -4,73 \end{pmatrix}$$

Modelo estimado:

$\hat{Y}_t = 69,557 - 1,704 P_t - 4,73 I_t \quad t = 1, 2, \dots, 30$

b) Obtén una medida de la bondad del ajuste e interprétala.

Solución:

$$R^2 = 1 - \frac{SCR}{SCT} = 1 - \frac{483,72}{4364,12} = 0,8891$$

$$SCT = \sum Y_t^2 - T\bar{Y}^2 = 48600,92 - 30(38,40)^2 = 4364,12$$

$$SCE = \sum Y_t^2 - \hat{\beta}' X'Y = (69,557 - 1,704 - 4,73) \begin{pmatrix} 1152 \\ 6400,22 \\ 4462,26 \end{pmatrix} = 483,7$$

Se explica de forma lineal casi el $90\,\%$ de la variación total del número de viviendas vendidas mediante la variación del precio medio de la vivienda y del tipo de interés medio de los préstamos hipotecarios.

c) Estima la matriz de varianzas y covarianzas de los estimadores MCO.

Solución:

$$\widehat{\sigma}^2 = \frac{SCR}{T-K} = 4837,230 - 3 = 17,915$$

$$\widehat{var}(\hat{\beta}) = \widehat{\sigma}^2 (X'X)^{-1} =$$

$$17,915(X'X)^{-1} = \begin{pmatrix} 12,988 & -5,84 & 5,302 \\ & 3,332 & -3,278 \\ & & 3,314 \end{pmatrix}$$

d) ¿Es el tipo de interés medio de los préstamos hipotecarios una variable significativa en el modelo propuesto? ¿Son las variables explicativas del modelo conjuntamente significativas?

Solución:

$$\begin{cases} H_o: & \beta_3 = 0 \\ H_a: & \beta_3 \neq 0 \end{cases} \qquad t_c = \frac{\hat{\beta}_3}{\hat{\sigma}_{\hat{\beta}_3}} \overset{H_o}{\sim} t_{(T-K)}$$

$$\mid t_c \mid = \mid \frac{-4,73}{\sqrt{3,314}} \mid = 2,598 > 2,052 = t_{(27)0,025}$$

Rechazo H_o para un nivel de significatividad del $5\,\%$. El interés medio de los préstamos hipotecarios es una variable significativa para explicar el número de viviendas vendidas.

$$\begin{cases} H_o: & \beta_2 = \beta_3 = 0 \\ H_a: & \beta_2 \neq 0 \text{y/o} \beta_3 \neq 0 \end{cases} \qquad F = \frac{R^2}{1-R^2} \frac{T-K}{q} \overset{H_o}{\sim} F_{(q,T-K)}$$

$$F = \frac{0,8891/2}{(1-0,8891)/27} = 108,231 > 3,35 = F_{(2,27)0,05}$$

Rechazo H_o para un nivel de significatividad del $5\,\%$. Las variables explicativas P_t e I_t son conjuntamente significativas para explicar el número de viviendas vendidas.

e) El gobierno cree que el efecto del tipo de interés sobre la venta media de viviendas es el doble que el del precio y en el mismo sentido. Realiza el contraste adecuado para determinar si el gobierno está en lo cierto.

Solución:

$$\left\{\begin{array}{ll} H_o: & \beta_3 = 2\beta_2 \\ H_a: & \beta_3 \neq 2\beta_2 \end{array}\right. \rightarrow \left\{\begin{array}{ll} H_o: & \beta_3 - 2\beta_2 = 0 \\ H_a: & \beta_3 - 2\beta_2 \neq 0 \end{array}\right. \rightarrow \left\{\begin{array}{ll} H_o: & w = 0 \\ H_a: & w \neq 0 \end{array}\right.$$

$$t_c = \frac{\hat{w}}{\hat{\sigma}_{\hat{w}}} \overset{H_o}{\sim} t_{(T-K)}$$

$$\hat{w} = 2\hat{\beta}_3 - 2\hat{\beta}_2 = -4,73 - 2 \times (-1,704) = -1,322$$

$$v\hat{a}r(\hat{w}) = v\hat{a}r(\hat{\beta}_3) + 4v\hat{a}r(\hat{\beta}_2) - 4c\hat{o}v(\hat{\beta}_3, \hat{\beta}_2) = 3,314 + 4 \times 3,332 - 4 \times (-3,278) = 29,755$$

$$\mid t_c \mid = \mid \frac{1,322}{\sqrt{29,755}} \mid = 0,2423 < 2,052 = t_{(27)0,025}$$

No rechazo H_o para un nivel de significatividad del $5\,\%$. Por tanto el gobierno está en lo cierto.

f) Para el año 2003 el gobierno prevé que el tipo de interés baje 2 puntos con respecto al del año 2000 (que fue 4,5) y que el precio medio de la vivienda se mantenga estable en el mismo nivel que en el año 2000 (que fue 10). Con estas previsiones, el gobierno afirma que la venta de viviendas en el año 2003 se triplicará con respecto al nivel alcanzado en el año 2000, que fue 8. ¿Es razonable esta afirmación?

Solución:

$$I_{03} = 2,5 \quad P_{03} = 10 \quad Y_{03} = 3 \times 8 = 24?$$

$$IC(Y_{03})_{1-\alpha} = \left[\hat{Y}_{03} \pm t_{(T-K)\alpha}\hat{\sigma}\sqrt{1 + X_p'(X'X)^{-1}X_p}\right]$$

$$\hat{Y}_{03} = 69,557 - 1,704 \times 10 - 4,73 \times 2,5 = 40,692$$

$$t_{(27)0,05} = 2,052 \quad X_p' = (1102,5)' \quad X_p'(X'X)^{-1}X_p = 6,29125$$

$$IC(Y_{03})_{0,95} = \left[40,692 \pm 2,052\sqrt{17,915}\sqrt{1 + 6,29125}\right] = [17,262, 64,122]$$

LADE-2003.2 (Feb-2003)

Una empresa editorial ha recogido la siguiente información:

i	1	2	3	4	5	6	7	8	9	10	11	12	13
Y_i	8	10	12	12	14	20	22	5	6	10	12	20	24
H_i	0	0	0	0	0	0	0	1	1	1	1	1	1
ES_i	0	0	0	1	1	0	0	0	0	1	1	0	0
EU_i	0	0	0	0	0	1	1	0	0	0	0	1	1

donde Y_i es el número de horas semanales dedicadas a la lectura por el individuo i-ésimo, H_i es una variable que toma el valor 1 si el individuo i-ésimo es Hombre, ES_i es una variable que toma el valor 1 si el nivel máximo de estudios alcanzado por el individuo i-ésimo es la Enseñanza Secundaria y EU_i es una variable que toma el valor 1 si el nivel máximo de estudios alcanzado por el individuo i-ésimo es la Enseñanza Universitaria.

De dicha información se deduce que:

$$\bar{Y} = \frac{\sum_1^{13} Y_i}{13} = 13,4615 \quad \sum_1^{13}(Y_i - 13,4615)^2 = 457,24 \quad \sum_H Y_i = 77 \quad \sum_{ES} Y_i = 48 \quad \sum_{EU} Y_i = 86$$

$$\sum_{M,EP} Y_i = 30 \quad \sum_{M,ES} Y_i = 26 \quad \sum_{M,EU} Y_i = 42 \quad \sum_{H,EP} Y_i = 11 \quad \sum_{H,ES} Y_i = 22 \quad \sum_{H,EU} Y_i = 44$$

donde M=Mujer, EP= el nivel máximo de estudios alcanzado es Enseñanza Primaria.

a) El 4º individuo de la muestra ¿qué número de horas semanales dedica a la lectura? ¿Cuáles son su sexo y nivel de estudios?

Solución: $Y_4 = 12 \quad H_4 = 0 \quad ES_4 = 1$ El cuarto individuo es mujer con nivel de estudios máximos de Enseñanza Secundaria y dedica 12 horas semanales a la lectura.

b) Obtén la media muestral del número de horas semanales dedicadas a la lectura de los grupos (M,ES) y (H,EU).

Solución:

$$\bar{Y}_{(M,ES)} = \frac{\sum_{i\in(M,ES)} Y_i}{T_{(M,ES)}} = \frac{26}{2} = 13 \quad \bar{Y}_{(H,EU)} = \frac{\sum_{i\in(H,EU)} Y_i}{T_{(H,EU)}} = \frac{44}{2} = 22$$

Un economista de la editorial desea saber si la política publicitaria de la empresa debe centrarse en las mujeres, por lo que ha estimado el siguiente modelo:

$$\hat{Y}_i = \underset{(1,04)}{9,00} - \underset{(1,163)}{2,00}\, H_i + \underset{(1,401)}{4,00}\, ES_i + \underset{(1,401)}{13,50}\, EU_i \qquad (1)$$

$$R^2 = 0,9147 \qquad \sum \hat{u}_i^2 = 39,00$$

c) Contrasta si realmente la política publicitaria de la empresa debe ser diferente para las mujeres.

Solución:

$$Y_i = \beta_0 + \beta_1 H_i + \beta_2 ES_i + \beta_3 EU_i + u_i$$

$$\begin{cases} H_o: & \beta_1 = 0 \\ H_a: & \beta_1 \neq 0 \end{cases} \qquad t_c = \frac{\hat{\beta}_1}{\hat{\sigma}_{\hat{\beta}_1}} \overset{H_o}{\sim} t_{(T-K)}$$

$$|\, t_c \,| = \left| \frac{-2}{\sqrt{1,163}} \right| = 1,72 < 2,262 = t_{(9)0,025}$$

No rechazo H_o para un nivel de significatividad del $5\,\%$. No hay evidencia de que haya diferencia en el número de horas semanales dedicadas a la lectura por razón de sexo. Por tanto la política publicitaria no debería centrarse en las mujeres.

Otro economista de la editorial opina que, dadas las medias muestrales, una especificación más adecuada es:

$$\hat{Y}_i = \underset{(1,03)}{10,00} - \underset{(1,636)}{4,50}\, H_i + \underset{(1,636)}{3,00}\, ES_i + \underset{(1,636)}{11,00}\, EU_i + \underset{(2,427)}{2,50}\, (H_i ES_i) + \underset{(2,427)}{5,50}\, (H_i EU_i)$$

$$R^2 = 0,95079 \qquad \sum \hat{u}_i^2 = 22,50 \qquad (2)$$

d) ¿Cuál es la diferencia fundamental entre las especificaciones de los modelos (1) y (2)? Escribe los datos de las variables $H_i ES_i$ y $H_i EU_i$.

Solución: En (2) se introduce un efecto interacción entre las variables sexo y nivel máximo de estudios. Con ello se recoge la posibilidad de que el efecto del nivel de estudios pueda ser diferente para hombres y mujeres.

$$H_i ES_i = (0000000001100)$$
$$H_i EU_i = (0000000000011)$$

e) ¿Cuáles son los efectos diferenciales estimados entre los niveles educativos Enseñanza Universitaria y Enseñanza Secundaria en el modelo (2)? ¿Son iguales para los dos sexos?

Solución:

	EP	ES	EU
H	$10-4,5$	$10-4,5+3+2,5$	$10-4,5+11+5,5$
M	10	$10+3$	$10+11$

Para hombres tenemos (10-4,5+11+5,5)-(10-4,5+3+2,5)=11 y para mujeres tenemos (10+11)-(10+3)=8 por lo que no son iguales para ambos sexos.

f) Contrasta al nivel de significación del $5\,\%$ si el efecto diferencial del sexo es común para todos los niveles educativos.

Solución:

$$\begin{cases} H_o: & \beta_4 = \beta_5 = 0 \\ H_a: & \beta_4 \neq 0 \text{y/o} \beta_5 \neq 0 \end{cases} \qquad F = \frac{\hat{u}'_R \hat{u}_R - \hat{u}'\hat{u}}{\hat{u}'\hat{u}} \frac{T-K}{q} \overset{H_o}{\sim} F_{(q,T-K)}$$

$$MR: (1) \quad \hat{u}'_R \hat{u}_R = 39$$
$$MNR: (2) \quad \hat{u}'\hat{u} = 22,5$$

$$F = \frac{39 - 22,5}{22,5} \frac{13-6}{2} = 2,56 < 4,74 = F_{(2,7)0,05}$$

No rechazo H_o para un nivel de significatividad del $5\,\%$. El efecto diferencial del sexo es común para todos los niveles educativos.

g) En el modelo (2), contrasta al nivel de significación del $5\,\%$ que ni el sexo ni el nivel educativo son relevantes para explicar el número de horas dedicadas semanalmente a la lectura.

Solución:

$$\begin{cases} H_o: & \beta_1 = \ldots = \beta_5 = 0 \\ H_a: & \text{alguna igualdad no se cumple} \end{cases} \qquad F = \frac{R^2}{1-R^2} \frac{T-K}{q} \overset{H_o}{\sim} F_{(q,T-K)}$$

$$F = \frac{0,95079/5}{(1-0,95079)/9} = 27,0495 > 3,97 = F_{(5,7)0,05}$$

Rechazo H_o para un nivel de significatividad del $5\,\%$. Las variables explicativas sexo y nivel máximo de estudios son conjuntamente significativas para explicar el número de horas semanales dedicados a la lectura.

LADE-2003.3 (Feb-2003)

Un estudiante desea analizar el crecimiento de la producción en la Unión Europea (Y_t). Para ello propone el siguiente modelo

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

donde X_{2t} es el tipo de interés nominal, X_{3t} es el tipo de interés real y X_{4t} es la tasa de inflación.

a) Si el estudiante calcula el tipo de interés real como la diferencia entre el tipo de interés nominal y la tasa de inflación, ¿qué problema encontrará a la hora de estimar los coeficientes del modelo por MCO? ¿Y si a la hora de obtener el tipo de interés real ha cometido un error de cálculo de forma que la relación anterior entre tipo de interés nominal, real e inflación no se cumple de forma exacta, sino que el coeficiente de correlación entre los tipos de interés nominal y real es 0,95? Describe ambos casos con detalle.

Solución:

$$\mathbf{1})\ X_{3t} = X_{2t} - X_{4t} \quad \rightarrow \quad \begin{aligned} Y_t &= \beta_1 + \beta_2 X_{2t} + \beta_3 (X_{2t} - X_{4t}) + \beta_4 X_{4t} + u_t \\ Y_t &= \beta_1 + (\beta_2 + \beta_3) X_{2t} + (\beta_4 - \beta_3) X_{4t} + u_t \end{aligned}$$

Hay un problema de multicolinealidad perfecta. No podemos invertir $(X'X)$ por lo que no podemos estimar los coeficientes del modelo de manera individual sino combinaciones lineales de ellos. En concreto obtenemos: $\hat\beta_1$, $(\beta_2 \hat{+} \beta_3)$ y $(\beta_4 \hat{-} \beta_3)$. Sin embargo el coeficiente de determinación (R^2) y la predicción no se ven alterados.

2) Si $r_{23} = 0,95$ hay un problema de multicolinealidad alto. Se puede estimar cada coeficiente del modelo pero obtendremos estimaciones imprecisas ya que la varianza de los estimadores es grande. Por ello, será más facil no rechazar hipótesis nulas de no siginificatividad individual de las variables.

b) En los dos casos descritos en a), si se sabe con certeza que $\beta_3 = \beta_2 - \beta_4$, ¿cómo estimarías los coeficientes del modelo? Descríbelo en detalle y comenta las propiedades del estimador propuesto.

Solución:

1) Estimaría por MCR, lo que equivale a estimar por MCO el modelo restringido:

$$\begin{aligned} Y_t &= \beta_1 + (\beta_2 + \beta_2 - \beta_4) X_{2t} + (\beta_4 - \beta_2 + \beta_4) X_{4t} + u_t \\ Y_t &= \beta_1 + (2\beta_2 - \beta_4) X_{2t} + (2\beta_4 - \beta_2) X_{4t} + u_t \end{aligned}$$

Como resultado obtendríamos $\hat\beta_{1R}$, $(2\widehat{\beta_2 - \beta}_4)_R$ y $2\widehat{\beta_4 - \beta}_{2R}$. Dado que tenemos tres ecuaciones y tres incógnitas los coeficientes pueden ser estimados individualmente: $\hat\beta_1^R$, $\hat\beta_2^R$, $\hat\beta_4^R$ y $\hat\beta_3^R = \hat\beta_2^R - \hat\beta_4^R$. Los estimadores serán lineales en u, insesgados (ya que la restricción $\beta_3 = \beta_2 - \beta_4$ es cierta) y tendrán menor varianza que si no incorporáramos la restricción.

2) Estimaría por MCR, lo que equivale a estimar por MCO el modelo restringido:

$$Y_t = \beta_1 + \beta_2 X_{2t}(\beta_2 - \beta_4)X_{3t} + \beta_4 X_{4t} + u_t$$
$$Y_t = \beta_1 + \beta_2(X_{2t} + X_{3t}) + \beta_4(X_{4t} - X_{3t} + u_t$$

Obtendríamos $\hat{\beta}_1^R$, $\hat{\beta}_2^R$, $\hat{\beta}_4^R$ y $\hat{\beta}_3^R = \hat{\beta}_2^R - \hat{\beta}_4^R$. Los estimadores serán lineales en u, insesgados (ya que la restricción $\beta_3 = \beta_2 - \beta_4$ es cierta) y tendrán menor varianza que si no incorporáramos la restricción.

LADE-2003.4 (Jun-2003)

Una agencia de viajes quiere realizar un estudio sobre el presupuesto que asignan las familias a la realización de viajes al extranjero. Ha observado que las familias con más renta destinan más dinero a los viajes, pero a medida que aumenta el número de hijos en la familia, esa asignación presupuestaria disminuye. Para ello, propone un modelo en el que analiza el presupuesto familiar destinado a este tipo de viajes, Y_i, en función de la renta familiar, R_i, y el número de hijos, H_i.

$$Y_i = \beta_1 + \beta_2 R_i + \beta_3 H_i + u_i \quad i = 1, 2, ..., 8 \quad u \sim N(0, \sigma^2 I) \tag{1}$$

Se dispone de la siguiente información muestral, donde el presupuesto y la renta están medidos en miles de euros:

i	Y_i	R_i	H_i
1	2	18,5	3
2	4	24	2
3	3	22,5	2
4	4	27,5	3
5	6	31,5	1
6	6	36	3
7	3	28,5	2
8	4	33,5	2

$$(X'X)^{-1} = \begin{bmatrix} 5,63655 & -0,1313 & -0,830 \\ -0,1313 & 0,00424 & 0,00606 \\ -0,830 & 0,00606 & 0,2944 \end{bmatrix} \qquad (X'Y) = \begin{bmatrix} 32 \\ 935 \\ 70 \end{bmatrix}$$

a) Estima el modelo propuesto por mínimos cuadrados ordinarios. ¿Tienen los coeficientes los signos esperados?

Solución:

$$\widehat{\beta} = \begin{bmatrix} \hat{\beta}_1 \\ \hat{\beta}_2 \\ \hat{\beta}_3 \end{bmatrix} = (X'X)^{-1}X'Y =$$
$$\begin{bmatrix} 5,63655 & -0,1313 & -0,830 \\ -0,1313 & 0,00424 & 0,00606 \\ -0,830 & 0,00606 & 0,2944 \end{bmatrix} \begin{bmatrix} 32 \\ 935 \\ 70 \end{bmatrix} = \begin{bmatrix} -0{,}4959 \\ 0{,}187 \\ -0{,}2859 \end{bmatrix}$$

$$\widehat{Y}_i = -0{,}4959 + 0{,}187R_i - 0{,}2859H_i$$

Por lo que se refiere a los signos, β_2: Incremento en el valor esperado de Y_i cuando la variable R_i se incrementa en una unidad, manteniéndose constante el resto de las variables. La agencia esperaba un signo positivo, y es el encontrado.
β_3:Incremento en el valor esperado de Y_i cuando la variable H_i se incrementa en una unidad, manteniéndose constante el resto de las variables. La agencia esperaba un signo negativo, y es el encontrado.

b)Obtén una medida de la bondad del ajuste e interpreta el resultado.

Solución:

$$R^2 = \frac{SCE}{SCT} = \frac{\hat{\beta}'X'Y - T\bar{Y}^2}{Y'Y - T\bar{Y}^2} = \frac{138{,}9632 - 8 \times 16}{142 - 8 \times 16} = \frac{10{,}963}{14} = 0{,}7830$$

el 78.30 % de la variación de la variable presupuesto destinado a viajes es explicado por la regresión.

c) ¿Crees que las variables explicativas son conjuntamente significativas? Realiza el contraste que creas oportuno.

Solución:

$$H_0 : \beta_2 = \beta_3 = 0$$
$$H_a : \beta_2 \neq 0 \quad y/o \quad \beta_3 \neq 0$$

$$F = \frac{R^2/(K-1)}{(1-R^2)/(T-K)} \overset{H_0}{\sim} F_{(K-1,T-K)}$$

$F = \frac{0{,}7830/2}{(1-0{,}7830)/(8-3)} = 9{,}0207 \geq F^{0{,}05}_{(2{,}5)} = 5{,}79$

Al nivel de significación del 5 % se rechaza la H_0 en favor de la H_a, por tanto, las variables explicativas son conjuntamente significativas.

d) ¿Crees que es aceptable la idea de que una familia con una renta de 36.000 euros y 4 hijos, destine a viajes al extranjero 7.000 euros?

Solución: Si realizamos una predicción para una familia de 4 hijos y con una renta de 36000 euros utilizando el modelo anterior, obtenemos:

$$\widehat{Y_p} = -0{,}4959 + 0{,}187 \times 36 - 0{,}2859 \times 4 = 5{,}0925$$

Para ver si es posible que esta familia destine a viajes 7000 euros, debemos hacer un intervalo de confianza para esa predicción:

$$Pr[\widehat{Y_p} - t_{0{,}025}\sqrt{1 + X_p'(X'X)^{-1}X_p} \leq Y_p \leq \widehat{Y_p} - t_{0{,}025}\widehat{\sigma}\sqrt{1 + X_p'(X'X)^{-1}X_p}] = 0{,}95$$

donde

$$X_p{}' = \left[\begin{array}{ccc} 1 & 36 & 4 \end{array} \right]$$

representa los valores de las variables explicativas para la familia para la que realizamos la predicción. La expresión $1 + X_p{}'(X'X)^{-1}X_p$ toma el valor de $2{,}49367$ usando el valor de X_p y la matriz inversa proporcionados. Asimismo, es preciso calcular el valor de $\widehat{\sigma}$ a partir de la siguiente expresión

$$\widehat{\sigma^2} = \frac{\hat{u}'\hat{u}}{T-k} = \frac{14 - 10{,}9632}{8-3} = 0{,}60736,$$

$\widehat{\sigma^2} \times (1 + X_p{}'(X'X)^{-1}X_p) = 0{,}60736(1 + 1{,}49367) = 1{,}5145$. Por lo tanto, el intervalo de confianza requerido es el siguiente:

$$Pr[5{,}0925 - 2{,}571\sqrt{1{,}5145} \leq Y_p \leq 5{,}0925 + 2{,}571\sqrt{1,5145}] = 0{,}95$$

$$Pr[1{,}928 \leq Y_p \leq 8{,}256] = 0{,}95$$

De donde se deduce que una familia con una renta de 36000 euros y 4 hijos si es posible que dediquen a viajes 7000 euros, dado que esta cifra pertenece al intervalo de confianza calculado.

e)Contrasta conjuntamente si el incremento esperado en el presupuesto cuando la renta aumenta en 1.000 euros es 0,25 **y** el incremento esperado en el presupuesto cuando el matrimonio tiene un hijo más es -0,5.

Solución:

$$\begin{array}{l} H_0 : \beta_2 = 0{,}25 \quad \beta_3 = -0{,}5 \\ H_a : \beta_2 \neq 0{,}25 \quad y/o \quad \beta_3 \neq 0{,}5 \end{array}$$

el estadístico de contrate en este caso es

$$F = \frac{(R\hat{\beta} - r)(R(X'X)^{-1}R')^{-1}(R\hat{\beta} - r)/q}{\hat{\sigma}^2} \overset{H_0}{\sim} F_{(q,T-K)}$$

donde

$$R = \left[\begin{array}{ccc} 0 & 1 & 0 \\ 0 & 0 & 1 \end{array} \right] \quad r = \left[\begin{array}{c} 0{,}25 \\ -0{,}5 \end{array} \right] \quad R(X'X)^{-1}R' = \left[\begin{array}{cc} 0{,}00424 & 0{,}00606 \\ 0{,}00606 & 0{,}2944 \end{array} \right] \quad R\hat{\beta} - r = \left[\begin{array}{c} -0{,}063 \\ 0{,}2141 \end{array} \right]$$

$F = \frac{1{,}2598/2}{0{,}60736} = 1{,}037 \leq F^{0{,}05}_{(2{,}5)} = 5{,}79$

Por lo tanto no se rechaza la hipótesis propuesta al nivel de significación del 0.05.

f) Suponiendo que fuese cierta la hipótesis del contraste del apartado e), **propón y estima** un modelo que le permita recoger dicha hipótesis en su estudio a la agencia de viajes. ¿Aconsejarías trabajar con este modelo a partir de ahora? ¿Por qué?

$\boxed{\text{Solución:}}$ Propondríamos el siguiente modelo restringido:

$$Y_i = \beta_1 + 0{,}25R_i - 0{,}5H_i + u_i^R$$

que podemos escribir como:

$$\underbrace{Y_i - 0{,}25R_i + 0{,}5H_i}_{Y_i^*} = \beta_1 + u_i^R$$

de donde $\hat{\beta}_1^R = \frac{\sum Y_i^*}{T} = \bar{Y}^* = \bar{Y} - 0{,}25\bar{R} + 0{,}5\bar{H} = -1{,}8125$ y por lo tanto $\hat{\beta}_2^R = 0{,}25$ y $\hat{\beta}_3^R = -0{,}5$. Los estimadores MCR son lineales, insesgados y eficientes (la diferencia de matrices de varianzas y covarianzas es semidefinida positiva), por lo que como la hipótesis de contraste suponemos que es cierta es más eficiente estimar el modelo restringido, y por lo tanto, aconsejaría a la agencia trabajar con el modelo restringido.

LADE-2003.5 (Jun-2003)

Un investigador quiere analizar la inversión (I_i) realizada en planes de pensiones en la Comunidad Autónoma Vasca en función del salario percibido (W_i) y el sector (privado o público) en el que se trabaja. Con una muestra de 500 individuos, de los cuales la mitad trabaja en el sector público, se ha estimado por MCO el siguiente modelo:

$$\underset{(t-est)}{\hat{I}_i} = 2,7 + \underset{(0,22)}{0,31}PU_i + \underset{(3,7)}{0,47}W_i \qquad R^2 = 0,71 \tag{2}$$

donde PU_i toma valor 1 si el individuo i-ésimo trabaja en el sector público y cero en caso contrario.
a) **Deriva** los estimadores utilizados en la estimación del modelo (2).

$\boxed{\text{Solución:}}$ El modelo que consideramos es

$$I_i = \beta_1 + \beta_2 PU_i + \beta_3 W_i + u_i$$

el criterio de MCO consiste en

$$\min_{\beta_i} \sum_i \hat{u}_i^2 = \min_{\beta_i} \sum_i (I_i - \beta_1 - \beta_2 PU_i - \beta_3 W_i)^2 = L(\beta_1, \beta_2, \beta_3)$$

las condiciones de primer orden son:

$$\frac{\partial L}{\partial \beta_1} = 0 \Longrightarrow -2\sum_i (I_i - \hat{\beta}_1 - \hat{\beta}_2 PU_i - \hat{\beta}_3 W_i) = 0$$

$$\frac{\partial L}{\partial \beta_2} = 0 \Longrightarrow -2\sum_i (I_i - \hat{\beta}_1 - \hat{\beta}_2 PU_i - \hat{\beta}_3 W_i)PU_i = 0$$

$$\frac{\partial L}{\partial \beta_3} = 0 \Longrightarrow -2\sum_i (I_i - \hat{\beta}_1 - \hat{\beta}_2 PU_i - \hat{\beta}_3 W_i)W_i = 0$$

de la resolución simultánea de las tres ecuaciones obtendremos los estimadores de MCO para β_1, β_2 y β_3.

b) Interpreta los coeficientes del modelo.

Solución:

$$E(I_i|PU = 0, W_i = 0) \quad = \quad \beta_1$$

es decir, β_1 es la inversión media de un trabajador del sector privado cuando el salario es cero.

$$E(I_i|PU = 1, W_i = 0) \quad = \quad \beta_1 + \beta_2$$

es decir, $\beta_1 + \beta_2$ es la inversión media de un trabajador del sector público cuando el salario es cero, por lo tanto β_2 es la diferencia en la inversión media de un trabajador del sector público respecto del trabajador del sector privado. Por otra parte,

$$\frac{\partial E(I_i)}{\partial W_i} \quad = \quad \beta_3$$

asi, β_3 es el incremento que se produce en la inversión media de un trabajador ante un aumento unitario del salario.

c) Contrasta la significatividad individual de las variables explicativas.

Solución:

$$\begin{array}{l} H_0 : \beta_2 = 0 \\ H_a : \beta_2 \neq 0 \end{array} \qquad t = \frac{\widehat{\beta_2}}{\widehat{\sigma(\hat{\beta}_2)}} \overset{H_0}{\sim} t_{T-K}$$

$t = 0{,}22 \leq t^{0{,}025}(497) = 1,96$

Al nivel de significación del 5 % no se rechaza la H_0, de modo que la variable PU no es individualmente significativa. Por lo que se refiere al salario

$$\begin{array}{l} H_0 : \beta_3 = 0 \\ H_a : \beta_3 \neq 0 \end{array}$$

$t = 3{,}7 \geq t^{0{,}025}(497) = 1,96$

Al nivel de significación del 5 % se rechaza la H_0, de modo que la variable W sí es individualmente significativa.

d)Dados los resultados de los contrastes, ¿qué propiedades tienen los estimadores de los coeficientes del modelo (2)?, ¿por qué?

Solución: Podríamos pensar que hay un error de especificación. Si existe una inclusión de variables irrelevantes , los estimadores MCO serían insesgados y los contrastes válidos. Si hay omisisón de variables relevantes, MCO serían sesgados y los contrastes no serían válidos

Más tarde, el investigador sospecha que la variable sexo puede afectar al incremento de la inversión media ante aumentos unitarios en el salario. Con este objetivo estima los siguientes modelos:

$$\underset{(t-est)}{\hat{I}_i} = 3,87 + \underset{(13,8)}{2,28} PU_i + \underset{(7,26)}{0,63} W_i + \underset{(17,4)}{0,21} W_i S_i \qquad R^2 = 0,82 \tag{3}$$

$$\underset{(t-est)}{\hat{I}_i} = 4,7 + \underset{(5,82)}{0,32} W_i + \underset{(1,87)}{0,15} W_i S_i \qquad R^2 = 0,75 \tag{4}$$

donde S_i toma valor 1 si el individuo i-ésimo es hombre y cero en caso contrario.

e) Dadas las regresiones (2), (3) y (4), selecciona la especificación más adecuada para analizar la inversión en los planes de pensiones. Razona detalladamente tu proceso de selección.

Solución: En el modelo (2) la variable W es individualmente significativa y PU es no significativa, sin embargo por lo que se refiere a la significatividad conjunta:

$$H_0 : \beta_2 = \beta_3 = 0$$
$$H_a : \beta_2 \neq 0 \quad y/o \quad \beta_3 \neq 0$$

$F = \frac{R^2/k-1}{(1-R^2)/T-K} = 405{,}6 \geq F^{0{,}05}_{(2{,}497)} = 3.$

Al nivel de significación del 5 $\%$ se rechaza la H_0 en favor de la H_a, por tanto, las variables explicativas son conjuntamente significativas.

En relación la modelo (3) la situación es que las tres variables explicativas son individualmente significativas:

$$H_0 : \beta_i = 0 \qquad \text{usando el estadístico} \quad t = \frac{\widehat{\beta}_i}{\hat{\sigma}_{\hat{\beta}_i}} \overset{H_0}{\sim} t_{T-K}$$
$$H_a : \beta_i \neq 0$$

que, bajo la hipótesis nula, se distribuye como una t de Student con $(500-3)$ grados de libertad y toma el valor $1,96$.

Dados los resultados del enunciado, $t_2 = 13{,}8 > 1{,}96$, $t_3 = 7{,}26 > 1{,}96$,$t_4 = 17{,}4 > 1{,}96$ se rechaza la Ho a un nivel de significación del $5\,\%$ en los tres casos.

Por lo que se refiere al modelo (4) y utilizando el mismo estadístico, se rechaza la significatividad de la variable W, no se rechaza la significatividad de la variable WS, y conjuntamente son significativas. Por lo tanto, de los resultados de los contrastes anteriores elegiremos el modelo (3) como el más adecuado.

f) ¿Qué sucedería en la estimación de los modelos (3) y (4) si la muestra estuviera compuesta sólo por hombres?

$\boxed{\text{Solución:}}$ Si la muestra estuviera compuesta sólo por hombres $S_i = 1 \forall i$ tendríamos un problema de multicolinealidad perfecta, no podriamos estimar todos los parámetros de forma única, sino combinaciones lineales de ellos. Si nos proponemos estimar el modelo (3):

$$Y_i = \beta_1 + \beta_2 PU_i + (\beta_3 + \beta_4) W_i + u_i$$

y estimaríamos $\hat{\beta}_1$, $\hat{\beta}_2$, $\widehat{\beta_3 + \beta_4}$. En el caso del modelo (4)

$$Y_i = \gamma_1 + (\gamma_2 + \gamma_3) W_i + u_i$$

y estimariamos $\hat{\gamma}_1$, $\widehat{\gamma_2 + \gamma_3}$.

g) Suponiendo que eliges el modelo (4) ¿qué propiedades tienen los estimadores?

$\boxed{\text{Solución:}}$ En este modelo se produce omisión de la variable relevante PU_i. El estimador de MCO de los coeficientes del modelo es sesgado y el estimador de la varianza de las perturbaciones es sesgado. Por lo tanto, la inferencia queda invalidada.

LE-2004.1 (Feb-2004)

Se desea estudiar el comportamiento de las exportaciones industriales vascas (Y) considerando que las variables que pueden influir en el mismo son el nivel de actividad interior (medido a través del Valor Añadido vasco, X_1), el nivel de actividad externo (medido a través del Valor Añadido europeo, X_2) y el precio relativo de los productos exportados (medido a través de un cociente de precios entre el precio de los productos vascos y el precio medio europeo, X_3). Se ha propuesto por ello el siguiente modelo

$$Y_t = \beta_0 + \beta_1 X_{1t} + \beta_2 X_{2t} + \beta_3 X_{3t} + u_t \qquad u_t \sim N(0, \sigma^2) \tag{1}$$

Basándote en la información muestral disponible que corresponde a 34 observaciones anuales desde 1970 a 2003 y se resume en:

$$(x'x)^{-1} = \begin{pmatrix} 0,07357 & -0,000287 & 0,04267 \\ & 0,000760 & -0,000448 \\ & & 0,03179 \end{pmatrix}$$

$\sum_{t=1}^T x_{1t} y_t = 318,08$ $\quad \sum_{t=1}^T x_{2t} y_t = 409,13$ $\quad \sum_{t=1}^T x_{3t} y_t = -456,01$ $\quad \sum_{t=1}^T Y_t^2 = 36849,03$

$\sum_{t=1}^T Y_t = 1089,13$ $\quad \sum_{t=1}^T X_{1t} = 138,36$ $\quad \sum_{t=1}^T X_{2t} = 677,30$ $\quad \sum_{t=1}^T X_{3t} = 96,82$

a) Estima el modelo propuesto por el método de mínimos cuadrados ordinarios.

[0] CVS Id: $Id: lefeb04.tex,v 1.1 2004/07/21 09:54:56 etpdihei Exp

Solución:

$$\widehat{\beta}=(X'X)^{-1}X'Y=\begin{pmatrix}0,07357 & -0,000287 & 0,04267\\ & 0,000760 & -0,000448\\ & & 0,03179\end{pmatrix}\begin{pmatrix}318,08\\409,13\\-456,01\end{pmatrix}=\begin{pmatrix}3,825\\0,423\\-1,107\end{pmatrix}$$

$$\begin{aligned}\widehat{\beta}_0 &= \bar{Y}-\widehat{\beta}_1\bar{X}_1-\widehat{\beta}_2\bar{X}_2-\widehat{\beta}_3\bar{X}_3=\\ &= \frac{1}{34}(1089,13-3,825\times 138,6-0,423\times 677,3+1,107\times 96,82)=11,19\end{aligned}$$

$$\hat{Y}_t=11,19+3,825X_{1t}+0,423X_{2t}-1,107X_{3t}$$

b) Calcula una medida de la bondad del ajuste e interprétala.

Solución:

$$\begin{aligned}\widehat{u}\prime\widehat{u} &= \sum(Y_t-\bar{Y})^2-\widehat{\beta}'X'Y=\\ &= (36849,03-34x32,03^2)-\begin{pmatrix}3,825 & 0,423 & -1,107\end{pmatrix}\begin{pmatrix}318,08\\409,13\\-456,01\end{pmatrix}\end{aligned}$$

= $1967,67-1894,52=73,14\Rightarrow R^2$ = $1-\frac{\widehat{u}'\widehat{u}}{\sum(Y_t-\bar{Y})^2}=1-\frac{73,14}{1967,67}=0,9268$ Este resultado significa que el 96,28 % de la variabilididad de las exportaciones se explica a partir de la variabilidad de las tres variables incorporadas al modelo. Se trata de un buen ajuste.

c) Estima la matriz de varianzas y covarianzas de los coeficientes $\hat{\beta}_1$, $\hat{\beta}_2$, $\hat{\beta}_3$.

Solución:

$$\widehat{V}\begin{pmatrix}\widehat{\beta}_1\\\widehat{\beta}_2\\\widehat{\beta}_3\end{pmatrix}=\widehat{\sigma}^2(x'x)^{-1}=\frac{\widehat{u}'\widehat{u}}{T-K-1}(x'x)^{-1}=\frac{76,14}{30}\begin{pmatrix}0,07357 & -0,000287 & 0,04267\\ & 0,000760 & -0,000448\\ & & 0,03179\end{pmatrix}$$

d) Contrasta al nivel de significación del 5 % que las variables son individualmente significativas.

Solución:

$$\begin{array}{ll}H_0:\beta_1=0 & \\ H_a:\beta_1\neq 0 & \end{array}\qquad \frac{\widehat{\beta}_1}{s_{\widehat{\beta}_1}}\overset{H_o}{\sim}t_{30}$$

$$\mid\frac{\widehat{\beta}_1}{s_{\widehat{\beta}_1}}\mid=\frac{3,825}{\sqrt{2,438\times 0,07357}}=9,04\Rightarrow 9,04>2,042=t_{30}^{0,025}$$

$\Rightarrow$ Rechazar H_0 a un nivel de significación del $5\,\%$

$$\begin{array}{ll}H_0:\beta_2=0 & \\ H_a:\beta_2\neq 0 & \end{array}\qquad \frac{\widehat{\beta}_2}{s_{\widehat{\beta}_2}}\overset{H_o}{\sim}t_{30}$$

$$| \frac{\widehat{\beta}_2}{s_{\widehat{\beta}_2}} |= \frac{0,423}{\sqrt{2,438 \times 0,00076}} = 9,83 \Rightarrow 9,83 > 2,042 = t_{30}^{0,025}$$

$\Rightarrow$ Rechazar H_0 a un nivel de significación del $5\,\%$.

$$\begin{array}{ll} H_0 : \beta_3 = 0 & \\ H_a : \beta_3 \neq 0 & \end{array} \qquad \frac{\widehat{\beta}_3}{s_{\widehat{\beta}_3}} \overset{H_o}{\sim} t_{30}$$

$$| \frac{\widehat{\beta}_3}{s_{\widehat{\beta}_3}} |=| \frac{-1{,}107}{\sqrt{2,438 \times 0,03179}} |= 3,982 \Rightarrow 3{,}982 > 2,042 = t_{30}^{0,025}$$

$\Rightarrow$ Rechazar H_0 a un nivel de significación del $5\,\%$

Las tres variables explicativas, X_1, X_2 y X_3 son individualmente significativas

e) ¿Cómo contrastarías, al nivel de significación del $10\,\%$, que el nivel de actividad, tanto si es interior como si es exterior, no tiene efecto sobre las exportaciones? Indica claramente la hipótesis nula, la alternativa y el estadístico de contraste a utilizar así como la regla de decisión.

Solución: La hipótesis a contrastar es

$$\begin{array}{l} H_0 : \beta_1 = \beta_2 = 0 \\ H_a : \beta_1 \neq 0 \text{ y/o } \quad \beta_2 \neq 0 \end{array} \Rightarrow \begin{array}{l} H_0 : R\beta = s \\ H_a : R\beta \neq s \end{array}$$

usando las matrices

$$R = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \end{pmatrix}, s = \begin{pmatrix} 0 \\ 0 \end{pmatrix}$$

El estadístico a utilizar sería

$$E = \frac{(R\widehat{\beta} - s)'[R(x'x)^{-1}R']^{-1}(R\widehat{\beta} - s)/2}{\widehat{\sigma}^2} \overset{H_0}{\sim} F_{30}^2$$

El criterio de decisión consiste en rechazar la H_0 si $E > F^2_{30,0,10}$ y no rechazar la H_0 en caso contrario, ambos a un nivel de significación de α.

LE-2004.2 (Feb-2004)

a) Otra hipótesis que se ha considerado es la de que el efecto del nivel de actividad en el País Vasco y el nivel de actividad europeo sobre las exportaciones vascas es el mismo. Por esa razón, se ha propuesto estimar el modelo

$$Y_t = \beta_0 + \beta_1(X_{1t} + X_{2t}) + \beta_3 X_{3t} + u_t. \tag{2}$$

Dada la información de que dispones, contrasta si la hipótesis anterior es razonable y establece, en consecuencia, las propiedades que se obtendrían para los estimadores del modelo 2.

Solución: La hipótesis a contrastar es

$$\begin{array}{l} H_0 : \beta_1 = \beta_2 \\ H_a : \beta_1 \neq \beta_2 \end{array} \Rightarrow \begin{array}{l} H_0 : \beta_1 - \beta_2 = 0 \\ H_a : \beta_1 - \beta_2 \neq 0 \end{array}$$

que puede contrastarse utilizando el estadístico $t = \frac{\widehat{\beta}_1 - \widehat{\beta}_2}{\sqrt{\widehat{V}(\widehat{\beta}_1 - \widehat{\beta}_2)}}$ que, bajo H_0, seguirá una distribución t_{30}. Calculando $\widehat{\beta}_1 - \widehat{\beta}_2 = 3,825 - 0,423 = 3,402$, V($\widehat{\beta}_1 - \widehat{\beta}_2$)= $\sigma^2(0,07357 + 0,00076 - 2(-0,00827)) \Rightarrow \widehat{V}(\widehat{\beta}_1 - \widehat{\beta}_2) = \widehat{\sigma}^2 0,0759 = 0,1852$. De donde $t = \frac{3,402}{0,430} = 7,91 > 2,042 = t_{30}^{0,025} \Rightarrow$ Rechazar H_0.
Puesto que se rechaza la restricción, el modelo restringido (2) incorpora una restricción que no es cierta y ésto hace que la estimación de los coeficientes en este modelo sea sesgada, ya que el estimador restringido sólo es insesgado cuando la restricción es cierta. Sí tendrán menor varianza que los del modelo (1) pero, al ser sesgados, no deberíamos compararlos en términos de varianza con el estimador insesgado del modelo (1). Por ello, elegiremos la estimación obtenida en el modelo (1).

b) Con los datos del problema anterior (LE-2004.1) se ha estimado también la siguiente regresión auxiliar:

$$X_{1t} = \gamma_0 + \gamma_1 X_{3t} + v_t$$

con el siguiente resultado:

$$\hat{X}_{1t} = \underset{(0,191)}{5,726} - \underset{(0,054)}{0,582} X_{3t} \qquad R^2 = 0,779.$$

A la vista de la importante relación existente entre ambas variables, se ha propuesto y estimado el siguiente modelo, en el que se ha eliminado la variable X_1:

$$\hat{Y}_t = \underset{(1,744)}{32,81} + \underset{(0,080)}{0,438} X_{2t} - \underset{(0,246)}{3,344} X_{3t} \quad R^2 = 0,865. \tag{3}$$

¿Qué opinión te merece este procedimiento? ¿Por qué? Razona cuidadosamente tu respuesta, utilizando para ello todos los datos a tu disposición.

Solución: El resultado de la primera regresión nos informa de que las variables X_1 y X_3 están muy correlacionadas entre sí, ya que su coeficiente de correlación simple puede calcularse en este caso como $r_{1,3} = \sqrt{R^2} = \sqrt{0{,}779} = 0{,}882$. Como ambas variables son regresores en el modelo, podemos pensar en la existencia de un posible problema de multicolinealidad. Para juzgar la importancia del problema hacemos referencia a los resultados obtenidos en el problema 1, d) y comprobamos que todos los regresores han resultado individualmente significativos en el modelo (1). Por lo tanto, que X_1 y X_3 estén correlacionadas entre sí no ha hecho que la estimación de los coeficientes en (1) sea imprecisa.
En estas condiciones, ambos regresores son, de acuerdo con los datos, relevantes para explicar Y. Por eso el modelo (3) incurre en un error de especificación, ya que omite del modelo

un regresor relevante. Esto hace que la estimación de los coeficientes sea sesgada (a no ser que X_1 fuera incorrelacionada con X_2 y X_3, cosa que obviamente no ocurre en este caso). Además, al haber sesgado también la estimación de la varianza de las perturbaciones (porque los residuos contienen ahora la información de X_1 no incluida en el modelo) los resultados que se proporcionan en (3) no permiten llevar a cabo inferencia correcta. Por eso, el modelo (3) es incorrecto y deberemos elegir el modelo (1) y los resultados que de él se desprenden.

LE-2004.3 (Feb-2004)

En 1986, España entró a formar parte de la Unión Europea, lo que cambió de manera sustancial las condiciones de los intercambios comerciales con todos los países europeos. Se cree que ese cambio ha podido tener un efecto considerable en las exportaciones por lo que debe tenerse en cuenta en el modelo.

a) Define una variable que permita tener en cuenta la entrada en la Unión Europea a partir de 1986.¿Qué valor toma la variable que has definido en la primera observación? ¿Y en 1985? ¿Y en 1986? ¿Y en 2003?

Solución: Podemos definir una variable, D_t, de la siguiente manera:

$$D_t = \begin{cases} 1 & \text{para} \quad t < 1986 \\ 0 & \text{para} \quad t \geq 1986 \end{cases}$$

Con esta definición la variable distingue entre los períodos anteriores a la entrada en la Unión Europea y los posteriores a este hecho. Así

$$\begin{array}{rcl} t = 1 & \Rightarrow & D_{1970} = 0 \\ t = 1985 & \Rightarrow & D_{1985} = 0 \\ t = 1986 & \Rightarrow & D_{1986} = 1 \\ t = 2003 & \Rightarrow & D_{2003} = 1 \end{array}$$

b) Propón un modelo que incorpore esta variable e interpreta el valor que toma el coeficiente que le acompaña.

Solución: Un modelo posible es el modelo (1) añadiéndole además la variable D_t, es decir,

$$Y_t = \beta_0 + \beta_1 X_{1t} + \beta_2 X_{2t} + \beta_3 X_{3t} + \gamma D_t + u_t.$$

En este caso, el modelo establece que

$$\begin{array}{ll} E(Y_t \mid t < 1986) & = \beta_0 + \beta_1 X_{1t} + \beta_2 X_{2t} + \beta_3 X_{3t} \\ E(Y_t \mid t \geq 1986) & = (\beta_0 + \gamma) + \beta_1 X_{1t} + \beta_2 X_{2t} + \beta_3 X_{3t} \end{array}$$

Por lo tanto, si las variables X_1,X_2 y X_3 tomaran el mismo valor en ambos casos, la diferencia entre el valor esperado de las exportaciones antes y después de entrar en la Unión Europea viene dada por la diferencia entre β_0 y $(\beta_0 + \gamma)$. Luego, en este modelo, el coeficiente γ recoge la diferencia esperada debido a la entrada en la Unión Europea.

c) En la estimación del modelo que has propuesto en a) se ha obtenido un valor de $R^2 = 0,998$. Contrasta la hipótesis de que la entrada en la Unión Europea es un factor relevante para explicar las exportaciones vascas, señalando con claridad cuál es la hipótesis nula, la alternativa y el estadístico que utilizas para llevar a cabo el contraste.

Solución: Para el modelo del apartado anterior contrastar la hipótesis de que la entrada en la Unión Europea no es un factor relevante para explicar las exportaciones vascas supone contrastar

$\begin{array}{c} H_0 : \gamma = 0 \\ H_a : \gamma \neq 0 \end{array}$ Bajo H_0 el modelo restringido es el (1) mientras bajo H_a el modelo no restringido es el (4). Por lo tanto, podemos usar para llevar a cabo el contraste el estadístico $E = \frac{(\widehat{u}_r'\widehat{u}_r - \widehat{u}'\widehat{u})/m}{\widehat{\sigma}^2}$ que, bajo la hipótesis nula, sigue una distribución F^m_{T-K-1}. En este caso, $m = 1$, $T - K - 1 = 29$ en el modelo no restringido y el estadístico puede escribirse en la forma

$$E = \frac{(R^2_{nr} - R^2_r)/1}{(1 - R^2_{nr})/29} = \frac{0{,}998 - 0{,}9628}{0{,}002/29} = 510{,}4 > 4{,}18 = F^1_{29,0,05}$$

$\Rightarrow$ Rechazar H_0. La entrada en la Unión Europea ha sido un factor relevante para explicar la evolución de las exportaciones vascas.

LE-2004.4 (Jun-2004)

El dueño de una cadena de cines en una ciudad costera pretende saber cómo influyen en el número de espectadores de sus salas , N (*miles*), dos factores: el empleo existente en la ciudad , E (*cientos de miles*), y el número de otros espectáculos que se programen, O (*miles*). Propone el siguiente modelo

$$N_t = \alpha_0 + \alpha_1 E_t + \alpha_2 O_t + u_t \qquad u_t \sim NID(0, \sigma^2) \tag{1}$$

Dispone de información trimestral relativa a las tres variables que va desde el primer trimestre de 1985 hasta el primer trimestre de 2002, ambos inclusive. El resumen de dicha información aparece a continuación:

$$(X'X)^{-1} = \begin{bmatrix} 1{,}5334 & -0{,}2122 & 0{,}00734 \\ & 0{,}0320 & -0{,}00379 \\ & & 0{,}00329 \end{bmatrix} \qquad \begin{array}{ll} \sum N_t = 2279,4 & \sum N_t E_t = 16838{,}6 \\ \sum N_t O_t = 14193{,}8 & \sum (N_t - \bar{N})^2 = 2435 \end{array}$$

a) Estima el modelo propuesto por el método de mínimos cuadrados ordinarios.

Solución:

$$\widehat{\beta} = (X'X)^{-1}X'Y = \begin{bmatrix} 1{,}5334 & -0{,}2122 & 0{,}00734 \\ & 0{,}0320 & -0{,}00379 \\ & & 0{,}00329 \end{bmatrix} \begin{pmatrix} 2279,4 \\ 16838,6 \\ 14193,8 \end{pmatrix} = \begin{pmatrix} 26,26 \\ 1,352 \\ -0,389 \end{pmatrix}$$

[0] CVS Id: $Id: lejun04.tex,v 1.1 2004/07/21 09:54:56 etpdihei Exp

$$\hat{N}_t = 26,26 + 1,352E_t - 0,389O_t$$

b) Calcula una medida de la bondad del ajuste e interprétala.

Solución:

$$\begin{aligned} SCE &= \widehat{\beta}' X' N - T\bar{N}^2 = \\ &= (26,26 \;\; 1,352 \;\; -0,389) \begin{pmatrix} 2279,4 \\ 16838,6 \\ 14193,8 \end{pmatrix} - 69 \left(\frac{2279,4}{69} \right)^2 = \\ &= 77101,443 - 75299,483 = 1801,96 \end{aligned}$$

$$R^2 = \frac{SCE}{SCT} = \frac{\widehat{\beta}' X' N - T\bar{N}^2}{\sum (N_t - \bar{N})^2} = \frac{1801,96}{2435} = 0,74$$

Este resultado significa que el 74 % de la variabilididad del número de espectadores se explica a partir de la variabilidad de las dos variables incorporadas al modelo. Se trata de un buen ajuste.

c) Estima la matriz de varianzas y covarianzas del estimador de los coeficientes.

Solución:

$$\widehat{V} \begin{pmatrix} \widehat{\alpha}_0 \\ \widehat{\alpha}_1 \\ \widehat{\alpha}_2 \end{pmatrix} = \widehat{\sigma}^2 (X'X)^{-1} = 9,59 \begin{bmatrix} 1{,}5334 & -0{,}2122 & 0{,}00734 \\ & 0{,}0320 & -0{,}00379 \\ & & 0{,}00329 \end{bmatrix}$$

$$\widehat{\sigma}^2 = \frac{\widehat{u}' \widehat{u}}{T - K - 1} = \frac{y'y - \widehat{y}' \widehat{y}}{T - K - 1} = \frac{2435 - 1801,96}{69 - 2 - 1} = \frac{633,04}{66} = 9,59$$

d) ¿Son las variables explicativas conjuntamente significativas?

Solución:

$H_0 : \alpha_1 = \alpha_2 = 0$

$H_a : \alpha_1 \neq 0$ y/o $\alpha_2 \neq 0$

$$F = \frac{R^2}{1-R^2} \frac{T-K-1}{m} \overset{H_o}{\sim} \mathcal{F}_{(m, T-K-1)}$$

$$F = \frac{0,74}{1 - 0,74} \frac{69 - 2 - 1}{2} = 93,92 > 3,15 = \mathcal{F}^{0{,}05}_{(2{,}66)}$$

Se rechaza H_0 a un nivel de significación del $5\,\%$ y por tanto las variables explicativas, E y O son conjuntamente significativas

e) Contrasta al nivel de significación del $5\,\%$ que la programación de otros espectáculos tiene influencia sobre el número de espectadores en las salas de cine.

Solución:

$H_0 : \alpha_2 = 0$

$H_a : \alpha_2 \neq 0$

$$\frac{\widehat{\alpha}_2}{\sigma_{\widehat{\alpha}_2}} \overset{H_o}{\sim} t_{T-K-1}$$

$$| \frac{\widehat{\alpha}_2}{\sigma_{\widehat{\alpha}_2}} |=| \frac{-0,389}{\sqrt{9,59 \times 0,00329}} |= 2,18 > 2 = t_{66}^{0,025}$$

$\Rightarrow$ Se rechaza H_0 a un nivel de significación del $5\,\%$ y por tanto la variable O es individualmente significativa

f) Para poder planificar la próxima campaña, el dueño hace el supuesto de que va a tener 34.500 espectadores, dado que el número de empleos será de 754.000 y que están previstos otros 8.800 espectáculos. Compruébalo.

Solución:

$$\begin{aligned}
&N_p = 34,5 \quad E_p = 7,54 \quad O_p = 8,8\\
&IC(N_p)_{1-\alpha} = \left[\hat{N}_p \pm t_{(T-K-1)\alpha/2}\hat{\sigma}\sqrt{1 + X_p'(X'X)^{-1}X_p}\right]\\
&\hat{N}_p = 26,26 - 1,352 \times 7,54 + 0,389 \times 8,8 = 33,03\\
&t_{(66)0,025} = 2 \quad X_p' = (17,548,8)' \quad X_p'(X'X)^{-1}X_p = 0,03368\\
&IC(N_p)_{0,95} = \left[33,03 \pm 2\sqrt{9,59}\sqrt{1 + 0,03368}\right] = [26,74; 39,32]
\end{aligned}$$

Como $34,5 \in [26{,}74; 39{,}32]$ la planificación del dueño es correcta.

LE-2004.5 (Jun-2004)

Analizando los resultados con un amigo suyo, observan una pauta de comportamiento diferente en los distintos trimestres. Por ello, piensan que no ha tenido en cuenta que la época del año marca también la asistencia al cine, debido a que se va menos cuando hace mejor tiempo. Nuestro hombre decide incluir este factor.

a) Define las variables que creas necesarias para tener en cuenta este factor y propón un modelo que incluya el posible efecto de la época del año.

Solución: $D1_t$ es una variable ficticia que toma el valor 1 si la observación corresponde al primer trimestre (invierno) y cero en los demás casos, $D2_t$ es una variable ficticia que toma el valor 1 si la observación corresponde al segundo trimestre (primavera) y cero en los demás casos, $D3_t$ es una variable ficticia que toma el valor 1 si la observación corresponde al tercer trimestre (verano) y cero en los demás casos, $D4_t$ es una variable ficticia que toma el valor 1 si la observación corresponde al cuarto trimestre (otoño) y cero en los demás casos, es decir:

$$D3_t = \begin{cases} 1, & t \in \text{ tercer trimestre (verano);}\\ 0, & t \in \text{ otro caso.}\end{cases}$$

$$D1_t = \begin{cases} 1, & t \in \text{ primer trimestre (invierno);}\\ 0, & t \in \text{ otro caso.}\end{cases}$$

$$D2_t = \begin{cases} 1, & t \in \text{ segundo trimestre (primavera);} \\ 0, & t \in \text{ otro caso.} \end{cases}$$

$$D4_t = \begin{cases} 1, & t \in \text{ cuarto trimestre (otoño);} \\ 0, & t \in \text{ otro caso.} \end{cases}$$

$$N_t = \alpha_0 + \alpha_1 E_t + \alpha_2 O_t + \beta_1 D1_t + \beta_2 D2_t + \beta_3 D3_t + u_t$$

b) Interpreta los coeficientes de tu modelo. ¿Cuántas de las variables que has definido incluyes en el mismo? ¿Por qué?

Solución:

α_0 : $E[N_t|D4_t = 1; E_t = O_t = 0] = \alpha_0 \equiv$ número esperado de espectadores N (*miles*) en otoño si el empleo E y el número de otros espectáculos que se programen O son cero.

α_1 : $\dfrac{\partial E[N_t]}{\partial E_t} = \alpha_1 \equiv$ variación (aumento o disminución) en el número esperado de espectadores N (*miles*) debido a un aumento de $100\,000$ en el empleo E (viene medida en *cientos de miles*), manteniendo constantes el resto de variables.

α_2 : $\dfrac{\partial E[N_t]}{\partial O_t} = \alpha_2 \equiv$ variación (aumento o disminución) en el número esperado de espectadores N (*miles*) debido a un aumento de $1\,000$ en el número de otros espectáculos que se programen O (viene medida en *miles*), manteniendo constantes el resto de variables.

β_1 : $E[N_t|D1_t = 1] - E[N_t|D4_t = 1] = \beta_1 \equiv$ diferencial del número esperado de espectadores N (*miles*) en invierno respecto de otoño, para cualquier nivel de empleo E y número de otros espectáculos que se programen O.

β_2 : $E[N_t|D2_t = 1] - E[N_t|D4_t = 1] = \beta_2 \equiv$ diferencial del número esperado de espectadores N (*miles*) en primavera respecto de otoño, para cualquier nivel de empleo E y número de otros espectáculos que se programen O.

β_3 : $E[N_t|D3_t = 1] - E[N_t|D4_t = 1] = \beta_3 \equiv$diferencial del número esperado de espectadores N (*miles*) en verano respecto de otoño, para cualquier nivel de empleo E y número de otros espectáculos que se programen O.

A la hora de estimar el modelo elijo la alternativa que incluye SOLO tres variables ficticias además de E y O, es decir:

$$N_t = \alpha_0 + \alpha_1 E_t + \alpha_2 O_t + \beta_1 D1_t + \beta_2 D2_t + \beta_3 D3_t + u_t$$

para no caer en un problema de multicolinealidad perfecta entre las variables ficticias y la constante.

LE-2004.6 (Jun-2004)

A los pocos días, su amigo le plantea la siguiente duda: a su ciudad acuden muchos turistas en verano, lo que aumenta el empleo temporal que suele cubrirse con gente joven, en general más aficionada al cine. Si ésto es cierto, la composición del empleo en verano puede tener influencia sobre la asistencia al cine.

a) ¿Cómo definirías una variable para tener en cuenta esta interrelación? ¿Cómo la incluirías en el modelo?

Solución:

$$EM3_t = D3_t \times E_t$$

donde $D3_t$ es una variable ficticia que toma el valor 1 en el tercer trimestre (verano) y cero en los demás casos, y E_t es el empleo existente en la ciudad:

$$N_t = \alpha_0 + \alpha_1 E_t + \alpha_2 O_t + \delta EM3_t + u_t$$

En línea con lo planteado, el empresario escoge y estima tres modelos diferentes de entre todos los posibles, con los siguientes resultados:

$$\underset{(\hat{\sigma}_{\hat{\beta}_i})}{\hat{N}_t} = \underset{(2{,}92)}{\overbrace{23{,}78}^{\hat{\beta}_1}} + \underset{(0{,}421)}{\overbrace{1{,}371}^{\hat{\beta}_2}} E_t - \underset{(0{,}135)}{\overbrace{0{,}317}^{\hat{\beta}_3}} O_t + \underset{(0,658)}{\overbrace{4{,}571}^{\hat{\beta}_4}} D3t, \quad \sum_{t=1}^{69} \hat{u}_t^2 = 359{,}89 \tag{2}$$

$$\underset{(\hat{\sigma}_{\hat{\beta}_i})}{\hat{N}_t} = \underset{(2{,}895)}{\overbrace{25{,}0}^{\hat{\delta}_0}} + \underset{(0{,}418)}{\overbrace{1{,}204}^{\hat{\delta}_1}} E_t - \underset{(0{,}134)}{\overbrace{0{,}316}^{\hat{\delta}_2}} O_t + \underset{(0,088)}{\overbrace{0{,}624}^{\hat{\delta}_4}} EM3_t, \quad \sum_{i=1}^{69} \hat{u}_t^2 = 355{,}23 \tag{3}$$

$$\underset{(\hat{\sigma}_{\hat{\beta}_i})}{\hat{N}_t} = \underset{(3{,}404)}{\overbrace{25{,}43}^{\hat{\mu}_0}} + \underset{(0{,}483)}{\overbrace{1{,}146}^{\hat{\mu}_1}} E_t - \underset{(0{,}135)}{\overbrace{0{,}316}^{\hat{\mu}_2}} O_t - \underset{(6{,}552)}{\overbrace{1{,}613}^{\hat{\mu}_3}} D3_t + \underset{(0,887)}{\overbrace{0,841}^{\hat{\mu}_4}} EM3_t, \quad \sum_{i=1}^{69} \hat{u}_t^2 = 354{,}90 \tag{4}$$

donde

- $D3_t$ es una variable ficticia que toma el valor 1 en el tercer trimestre (verano) y cero en los demás casos.
- $EM3_t$ es una variable que incorpora el efecto estacional del empleo que se ha discutido en el apartado anterior (y tendrá la interpretación que le hayas atribuido anteriormente).

b) Dados los resultados anteriores, ¿hay evidencia de que existe efecto estacional? Si el efecto estacional existe ¿qué forma toma?, ¿cuál de los modelos propuestos te parece óptimo para recogerlo?, ¿por qué? Lleva a cabo los contrastes precisos.

Solución: El efecto estacional está recogido en el modelo (2) por la variable $D3_t$, en el modelo (3) por la variable $EM3_t$ y en el modelo (4) por ambas. Por lo tanto, si alguna de estas variables resulta significativa habrá evidencia de estacionalidad. Todos los contrastes los haremos a un nivel de significación del $5\,\%$.

- En el modelo (2) contrastamos $H_0 : \beta_4 = 0 \quad H_a : \beta_4 \neq 0$ utilizando el estadístico $t = \frac{4{,}571}{0{,}658} = 6{,}946 > 2 \cong t_{65}^{0{,}025} \Rightarrow$ Rechazamos H_0; existe efecto estacional en el tercer trimestre.
- En el modelo (3) contrastamos $H_0 : \delta_4 = 0 \quad H_a : \delta_4 \neq 0$ utilizando el estadístico $t = \frac{0{,}624}{0{,}088} = 7{,}090 > 2 \cong t_{65}^{0{,}025} \Rightarrow$ Rechazamos H_0; existe efecto estacional a través del empleo en el tercer trimestre.
- En el modelo (4) contrastamos $H_0 : \mu_3 = \mu_4 = 0$ $H_a : \mu_3 \neq 0$ *y/o* $\mu_4 \neq 0$ utilizando el estadístico

$$\frac{(\widehat{u}_r'\widehat{u}_r - \widehat{u}'\widehat{u})/2}{\widehat{u}'\widehat{u}/64} = \frac{(633{,}04 - 354{,}9)/2}{354{,}9/64} = 25{,}07 > 3{,}15 = \mathcal{F}_{(2,64)}^{0{,}05} \quad \Rightarrow$$

  Rechazamos H_0; ambas variables son conjuntamente significativas; existe claramente efecto estacional. Para saber qué forma toma este efecto estacional, realizamos los siguientes contrastes en el modelo (4):
  - $H_0 : \mu_3 = 0 \quad H_a : \mu_3 \neq 0$ utilizando el estadístico $\mid t \mid = \mid \frac{-1{,}613}{6{,}552} \mid = 0{,}253 < 2 \cong t_{64}^{0{,}025} \Rightarrow$ Aceptamos H_0; cuando la variable $EM3_t$ está en el modelo no hay evidencia de que la variable $D3_t$ sea significativa.
  - $H_0 : \mu_4 = 0 \quad H_a : \mu_4 \neq 0$ utilizando el estadístico $t = \frac{0{,}841}{0{,}887} = 0{,}948 < 2 \cong t_{64}^{0{,}025}$ $\Rightarrow$ Aceptamos H_0; cuando la variable $D3_t$ está en el modelo no hay evidencia de que la variable $EM3_t$ sea significativa.

  Por lo tanto, existe efecto estacional pero no puede determinarse con exactitud la forma que toma.

c) A la vista de los resultados obtenidos hasta el momento ¿es correcto utilizar los valores de los coeficientes que has obtenido en (1)?, ¿por qué?, ¿qué propiedades tienen?

Solución: En el apartado anterior hemos visto que, aunque no estemos seguros de cuál es la forma que el efecto estacional toma, existe evidencia clara en los tres modelos propuestos de que la o las variables añadidas tienen capacidad explicativa. Por lo tanto, el modelo (1) que no contiene a ninguna de las dos incurrirá en un error de especificación por omisión de variables relevantes.
Cuando se comete este error, sabemos que lo habitual es que los estimadores MC del modelo mal especificado sean sesgados; la única excepción se da cuando las variables omitidas son ortogonales a las variables del modelo (incluyendo el término constante); en este caso eso no se produce, ya que los coeficientes estimados en los modelos (2) y (3) son diferentes de los estimados en el modelo (1), mientras que si las variables $D3_t$ y/o $EM3_t$ fueran ortogonales a E_t y O_t los coeficientes estimados serían idénticos (resultado que se deriva precisamente de la ortogonalidad). Por lo tanto, los estimadores obtenidos en el modelo (1) serán sesgados.
Además, en modelos con variables omitidas el estimador de la varianza de las perturbaciones σ^2 es siempre sesgado, lo que hace que los estadísticos que calculamos para los contrastes habituales no sigan las distribuciones que les suponemos; en consecuencia, la inferencia que hemos realizado utilizando los resultados del modelo (1) no es valida.

d) ¿Cómo explicas los resultados del modelo (4) a la vista del (2) y el (3)? ¿Te parece lógico obtener este resultado?, ¿cuál puede ser su causa?

Solución: Al incluir las variables $D3_t$ y $EM3_t$ en el modelo (4) se gana en capacidad explicativa de la variable N_t, lo que indica que el efecto estacional que a través de ellas

se recoge tiene importancia en la explicación del numero de espectadores al cine. Por otro lado, en el contraste individual de cada una de ellas, ninguna de las dos resulta significativa. Esto suele ser un síntoma de multicolinealidad, aunque también podría suceder que una de las dos no tuviera capacidad explicativa tomada de manera individual. Los modelos (2) y (3), por su parte, nos indican que cualquiera de las dos tiene capacidad explicativa cuando se considera ella sola. Por lo tanto, hay evidencia clara de multicolinealidad, lo que no es sorprendente ya que ambas variables tratan de recoger el mismo efecto (aunque de diferente manera) y existe entre ellas una relación ya que $EM3_t = D3_t E_t$.
Si fuera necesario elegir alguno de los tres modelos, el (4) sería el mejor especificado pero tiene la dificultad de que ninguno de los dos coeficientes relativos a la estacionalidad se estima con precisión. Claramente, una de las dos variables es redundante en presencia de la otra. Comparando a su vez los modelos (2) y (3), cosa que podemos hacer a través de su R^2 puesto que tienen el mismo número de regresores, encontramos que el (3) es algo mejor, debido a que proporciona una suma de cuadrados de residuos menor (y muy próxima a la del modelo (4)).

LADE-2004.1 (Feb-2004)

Considera el modelo de regresión:

$$Y_t = \beta_1 + \beta_2 X_{2t} + \beta_3 X_{3t} + u_t \qquad\qquad u_t \sim N(0, \sigma_u^2) \tag{1}$$

A partir de una muestra de 10 observaciones se han obtenido los siguientes resultados:

$$\sum_{t=1}^{10} Y_t = 61,5 \qquad \sum_{t=1}^{10} X_{2t} = 530 \qquad \sum_{t=1}^{10} X_{3t} = 58,8$$

$$\sum_{t=1}^{10} Y_t^2 = 385,21 \qquad \sum_{t=1}^{10} X_{2t}^2 = 30196 \qquad \sum_{t=1}^{10} X_{3t}^2 = 353,62$$

$$\sum_{t=1}^{10} X_{2t}Y_t = 3376,2 \qquad \sum_{t=1}^{10} X_{3t}Y_t = 368,97 \qquad \sum_{t=1}^{10} x_{2t}^2 = 2106$$

$$\sum_{t=1}^{10} x_{3t}^2 = 7{,}876 \qquad \sum_{t=1}^{10} x_{2t}y_t = 116{,}7 \qquad \sum_{t=1}^{10} x_{3t}y_t = 7{,}35$$

$$\sum_{t=1}^{10} x_{2t}x_{3t} = 125{,}5$$

1) Comprueba que la Función de Regresión Muestral es:

$$\hat{Y}_t = 0,5023 - 0,0039 X_{2t} + 0,9959 X_{3t} \tag{2}$$

Solución:

$$\begin{bmatrix} \hat\beta_2 \\ \hat\beta_3 \end{bmatrix} = \begin{bmatrix} \sum x_{2t}^2 & \sum x_{2t}x_{3t} \\ \sum x_{3t}x_{2t} & \sum x_{3t}^2 \end{bmatrix}^{-1} \begin{bmatrix} \sum x_{2t}y_t \\ \sum x_{3t}y_t \end{bmatrix} = \begin{bmatrix} 2106 & 125{,}5 \\ 125{,}5 & 7{,}876 \end{bmatrix}^{-1} \begin{bmatrix} 116{,}7 \\ 7{,}35 \end{bmatrix}$$

[0]CVS Id: $Id: ladefeb04.tex,v 1.1 2004/07/21 09:54:56 etpdihei Exp

$$
\begin{aligned}
&= \begin{bmatrix} 0{,}0094 & 0{,}15 \\ 0{,}15 & 2{,}5173 \end{bmatrix} \begin{bmatrix} 116{,}7 \\ 7{,}35 \end{bmatrix} = \begin{bmatrix} -0{,}0039 \\ 0{,}9959 \end{bmatrix} \\
\hat{\beta}_1 &= \bar{Y} - \hat{\beta}_2\bar{X}_2 - \hat{\beta}_3\bar{X}_3 = 6{,}15 + 0{,}0039 \cdot 53 - 0{,}9959 \cdot 5{,}88 = 0{,}5008
\end{aligned}
$$

2) Calcula una medida de la bondad del ajuste realizado, empleando los coeficientes estimados indicados en (2). Interpreta el resultado.

Solución: $SCT = \sum y_t^2 = \sum Y_t^2 - T\bar{Y}^2 = 385{,}21 - 10 \cdot (6{,}15)^2 = 6{,}985$

$$SCE = \begin{bmatrix} \hat{\beta}_2 & \hat{\beta}_3 \end{bmatrix} \begin{bmatrix} \sum x_{2t}y_t \\ \sum x_{3t}y_t \end{bmatrix} = (-0{,}0039) \cdot 116{,}7 + 0{,}9959 \cdot 7{,}35 = 6{,}8647$$

$R^2 = \frac{SCE}{SCT} = \frac{6{,}8647}{6{,}985} = 0{,}9866$ esto significa que el $98{,}66\,\%$ de la variación total de Y se puede explicar mediante las variables X_2 y X_3 conjuntamente.

3) Contrasta la significatividad individual y conjunta de las variables explicativas X_2 y X_3.

Solución: $SCR = SCT - SCE = 6{,}985 - 6{,}8647 = 0{,}1203$;
$\hat{\sigma}_u^2 = \frac{SCR}{T-k} = \frac{0{,}1203}{10-7} = 0{,}01718$

$$\begin{array}{ll} H_0: & \beta_2 = 0 \\ H_a: & \beta_2 \neq 0 \end{array} \qquad t = \frac{\hat{\beta}_2}{\sqrt{\widehat{Var}(\hat{\beta}_2)}} \overset{H_0}{\sim} t(T-k)$$

Regla de decisión: si $|t| > t(T-k)_{\alpha/2} \rightarrow$ Rechazar H_0 al nivel de significación α. Tomando $\alpha = 0{,}05$, $|t| = |\frac{-0{,}0039}{\sqrt{0{,}01718 \cdot 0{,}0094}}| = 0{,}3068 < t(7)_{0{,}025} = 2{,}365$, por lo que no rechazamos la hipótesis nula al nivel de significación del 5 %. Utilizamos ahora el mismo estadístico para contrastar $H_0: \ \beta_3 = 0$ frente a $H_a: \ \beta_3 \neq 0$. En este caso $|t| = \frac{0{,}9959}{\sqrt{0{,}01718 \cdot 2{,}5173}} = 4{,}7903 > t(7)_{0{,}025} = 2{,}365$ por lo que rechazamos la H_0 al nivel de significación del 5 %.

4) Dados los resultados de los contrastes anteriores, ¿podemos pensar en la existencia de un problema de multicolinealidad? Razona tu respuesta.

Solución: La correlación entre X_2 y X_3 es bastante alta, $r_{23} = \frac{\sum x_{2t}x_{3t}}{\sqrt{\sum x_{2t}^2 \cdot \sum x_{3t}^2}} = \frac{125{,}5}{\sqrt{2106 \cdot 7{,}876}} = 0{,}9745$, sin embargo parece que el modelo se ajusta muy bien, dando unas perturbaciones con una varianza muy pequeña $\hat{\sigma}_u^2 = 0{,}01718$, esto compensa la correlación entre las variables, de manera que a pesar de ella el estadístico t correspondiente a la variable X_3 es alto y así la variable es significativa. Por tanto no hay un problema *grave* de multicolinealidad entre estas dos variables, aunque podría haberlo entre una de ellas (X_2) y el término constante.

5) ¿Crees posible que un aumento unitario en X_3 haga aumentar la media poblacional de Y en una unidad? Realiza el contraste oportuno.

Solución:

$$\begin{array}{ll} H_0: & \beta_3 = 1 \\ H_a: & \beta_3 \neq 1 \end{array} \qquad t = \frac{\hat{\beta}_3 - 1}{\sqrt{\widehat{Var}(\hat{\beta}_3)}} \overset{H_0}{\sim} t(T-k)$$

Regla de decisión: si $|t| > t(T-k)_{\alpha/2} \rightarrow$ Rechazar H_0 al nivel de significación α. Tomando $\alpha = 0{,}05$, $|t| = |\frac{0{,}9959-1}{\sqrt{0{,}01718 \cdot 2{,}5173}}| = 0{,}01972 < t(7)_{0{,}025} = 2{,}365$, por lo que no rechazamos la hipótesis nula al nivel de significación del 5 %. A este nivel de significación, podemos afirmar que es posible que un aumento unitario en X_3 haga aumentar la media poblacional en una unidad.

6) Si impones que $\beta_3 = 1$:

6.1) ¿Cuál es el modelo restringido?

Solución: Si $\beta_3 = 1$ el modelo quedaría $Y_t = \beta_1 + \beta_2 X_{2t} + X_{3t} + u_t \Longrightarrow$ $Y_t^* = Y_t - X_{3t} = \beta_1 + \beta_2 X_{2t} + u_t$ es decir, $Y_t^* = \beta_1 + \beta_2 X_{2t} + u_t$ siendo $Y_t^* = Y_t - X_{3t}$.

6.2) Dada la información muestral estímalo. Calcula la Suma de los Residuos al Cuadrado (SRC), la Suma Total de Cuadrados (STC) y la Suma Explicada de Cuadrados (SEC).

Solución: $\sum x_{2t} y_t^* = \sum [x_{2t}(y_t - x_{3t})] = \sum x_{2t} y_t - \sum x_{2t} x_{3t} = 116{,}7 - 125{,}5 = -8{,}8$
$\hat\beta_2 = \frac{\sum x_{2t} y_t^*}{\sum x_{2t}^2} = \frac{-8{,}8}{2106} = -0{,}004178; \bar{Y}^* = \frac{\sum Y_t - \sum X_{3t}}{T} = \frac{61{,}5 - 58{,}8}{10} = 0{,}27$
$\hat\beta_1 = \bar{Y}^* - \hat\beta_2 \bar{X}_2 = 0{,}27 + 0{,}004178 \cdot 53 = 0{,}4914$
$STC = \sum y_t^{*2} = \sum (y_t - x_{3t})^2 = \sum y_t^2 + \sum x_{3t}^2 - 2\sum x_{3t} y_t = 6{,}985 + 7{,}876 - 2 \cdot 7{,}35 = 0{,}161; \quad SEC = \hat\beta_2 \sum x_{2t} y_t^* = -0{,}004178 \cdot (-8{,}8) = 0{,}03677$
$SRC = STC - SEC = 0{,}161 - 0{,}03677 = 0{,}12423$

6.3) ¿**Cómo estimarías** la matriz de varianzas y covarianzas de los estimadores restringidos de β_1 y β_2?

Solución: $\widehat{Var}(\hat\beta) = \hat\sigma^2 \begin{pmatrix} T & \sum X_{2t} \\ \sum X_{2t} & \sum X_{2t}^2 \end{pmatrix}^{-1}$ $\hat\sigma^2 = \frac{SRC}{T-k} = \frac{0{,}12423}{10-2} = 0{,}01553$ $\widehat{Var}(\hat\beta) =$
$0{,}01553 \begin{pmatrix} 10 & 530 \\ 530 & 30196 \end{pmatrix}^{-1} = \begin{pmatrix} 0{,}1553 & 8{,}2430 \\ 8{,}2430 & 469{,}63 \end{pmatrix}$

LADE-2004.2 (Feb-2004)

Se dispone de una muestra de 100 personas de las que se conocen sus gastos de consumo total anual (en miles de euros), su *género* y su *actitud ante el tabaco* (fumar o no fumar).

1) Plantea el modelo que te permita recoger la posible influencia del *género* y la *actitud ante el tabaco*, sobre los gastos totales bajo la hipótesis de independencia entre estas dos variables. Describe con detalle todos los elementos.

Solución:

Definimos una variable ficticia para el género y otra para la variable *fumar/no fumar*:

$$H_i = \begin{cases} 1 & \text{Si } i \text{ es hombre} \\ 0 & \text{si } i \text{ es mujer} \end{cases} \quad ; \quad F_i = \begin{cases} 1 & \text{Si } i \text{ fuma} \\ 0 & \text{si } i \text{ no fuma} \end{cases}$$

Siendo G_i los gastos de consumo total anual, planteamos el modelo:

$$G_i = \beta_0 + \beta_1 H_i + \beta_2 F_i + u_i \tag{3}$$

Donde u_i es un término de perturbación aleatoria serialmente incorrelacionado y con varianza constante y media cero. β_0, β_1 y β_2 son los parámetros de este modelo.

$E(G_i/i \text{ es hombre fumador}) = E(\beta_0 + \beta_1 + \beta_2 + u_i) = \beta_0 + \beta_1 + \beta_2$
$E(G_i/i \text{ es hombre no fumador}) = E(\beta_0 + \beta_1 + u_i) = \beta_0 + \beta_1$

$E(G_i/i \text{ es mujer fumadora}) = E(\beta_0 + \beta_2 + u_i) = \beta_0 + \beta_2$
$E(G_i/i \text{ es mujer no fumadora}) = E(\beta_0 + u_i) = \beta_0$

Así que β_0 representa al gasto medio de una mujer no fumadora, β_1 es el diferencial de gasto medio de un hombre respecto a una mujer (fume o no fume) y β_2 es el diferencial de gasto medio de un fumador respecto a un no fumador (sea hombre o mujer).

2) Explica **cómo contrastarías** la hipótesis de que ser fumador hace aumentar los gastos totales.

Solución:

$$\begin{array}{ll} H_0: & \beta_2 = 0 \\ H_a: & \beta_2 > 0 \end{array} ; \qquad t = \frac{\hat{\beta}_2}{\sqrt{\widehat{Var}(\hat{\beta}_2)}} \overset{H_0}{\sim} t_{(T-k)}$$

El contraste debe realizarse por la cola derecha de la distribución. Regla de decisión: si $t > t_{(T-k)\alpha}$ rechazaremos H_0 y concluimos que fumar hace aumentar los gastos totales.

3) ¿Cómo recogerías la posible dependencia entre el *genero* y la variable *fumar/no fumar*? Explica **cómo contrastarías** la hipótesis de que las dos variables son independientes.

Solución: Añadiendo al modelo un término de interacción:

$$G_i = \gamma_0 + \gamma_1 H_i + \gamma_2 F_i + \gamma_3 H_i F_i + u_i$$

El parámetro γ_3 recoge la diferencia entre el gasto medio de hombres y mujeres del efecto diferencial que supone fumar o no fumar. Para contrastar la independencia entre el *género* y la *actitud ante el tabaco* realizaríamos el contraste de $H_0: \gamma_3 = 0$ contra $H_a: \gamma_3 \neq 0$ este contraste se puede realizar mediante el estadístico $t = \frac{\hat{\gamma}_3}{\sqrt{\widehat{Var}(\hat{\gamma}_3)}}$, que bajo H_0 sigue una distribución t de Student con $T-k$ grados de libertad. Si $|t| > t_{(T-k)\alpha/2}$ rechazaríamos H_0 al nivel de significación α, concluyendo así que las dos variables no son independientes.

4) Si el resultado del contraste anterior te indica que se rechaza la hipótesis nula, ¿qué propiedades tendrían los estimadores de los coeficientes del modelo propuesto en el apartado 1)?

Solución: Si no se rechaza la hipótesis nula anterior, el término de interacción $H_i F_i$ sería significativo y por tanto, en el modelo (3) se estaría incurriendo en un error de especificación por omisión de una variable relevante. Esto hace que los estimadores MCO de β y $\hat{\sigma}^2$ sean sesgados, por lo que los procedimientos estándar de inferencia no son válidos y el contraste del apartado 2) no es correcto.

5) En base al modelo que has propuesto en el apartado 3), ¿**cómo calcularías** el valor esperado del gasto de una mujer fumadora? ¿y un intervalo del 95 % de confianza para el gasto de consumo total de una mujer fumadora?

Solución: $E(G_i/i \text{ mujer fumadora}) = E(\gamma_0 + \gamma_2 + u_i) = \gamma_0 + \gamma_2$
Intervalo de confianza:
$I_{1-\alpha}(G_p) = \left[\hat{G}_p \pm t_{(T-k)\alpha/2}\hat{\sigma}\sqrt{C'(X'X)^{-1}C + 1}\right]$ Siendo G_p el gasto de consumo de una mujer fumadora y $C' = (1, 0, 1)$

LADE-2004.3 (Feb-2004)

Considera el siguiente modelo de regresión

$$Y_t = \beta X_t + u_t \qquad u_t \sim N(0, \sigma^2) \tag{4}$$

donde se cumplen todas las hipótesis básicas del MRLG. Para estimar el parámetro desconocido β, se proponen dos estimadores,

$$\hat{\beta}^* = \frac{\sum X_t Y_t}{\sum X_t^2 + \bar{X}^2} \qquad\qquad \hat{\beta}_{MCO} = \frac{\sum X_t Y_t}{\sum X_t^2}$$

1) **Deriva** el estimador $\hat{\beta}_{MCO}$ propuesto.

Solución: $Min_{\hat{\beta}} \sum (Y_t - \hat{\beta} X_t)^2 = Min_{\hat{\beta}} (\sum Y_t^2 - \hat{\beta}^2 \sum X_t^2 - 2\hat{\beta} \sum X_t Y_t)$

$$\frac{\partial (\sum Y_t^2 - \hat{\beta}^2 \sum X_t^2 - 2\hat{\beta} \sum X_t Y_t)}{\partial \hat{\beta}} = 2\hat{\beta} \sum X_t^2 - 2 \sum X_t Y_t$$

Para el valor de $\hat{\beta}$ que minimiza la suma de cuadrados residual, esta derivada es cero, así que se obtiene

$$\hat{\beta} \sum X_t^2 - \sum X_t Y_t = 0 \quad \Longrightarrow \quad \hat{\beta} = \frac{\sum X_t Y_t}{\sum X_t^2}$$

2) ¿Son ambos estimadores insesgados?

Solución:

$$\hat{\beta} = \frac{\sum X_t Y_t}{\sum X_t^2} = \frac{\sum (\beta X_t + u_t) X_t}{\sum X_t^2} = \beta + \frac{\sum u_t X_t}{\sum X_t^2} \Rightarrow E(\hat{\beta}) = \beta + \frac{\sum \overbrace{E(u_t)}^{=0} X_t}{\sum X_t^2} = \beta$$

así que el estimador MCO es insesgado.

$$\hat{\beta}^* = \frac{\sum X_t Y_t}{\sum X_t^2 + \bar{X}^2} = \frac{\sum (\beta X_t + u_t) X_t}{\sum X_t^2 + \bar{X}^2} = \beta \frac{\sum X_t^2}{\sum X_t^2 + \bar{X}^2} + \frac{\sum u_t X_t}{\sum X_t^2 + \bar{X}^2} \Longrightarrow$$

$$E(\hat{\beta}^*) = \beta \frac{\sum X_t^2}{\sum X_t^2 + \bar{X}^2} + \frac{\sum \overbrace{E(u_t)}^{=0} X_t}{\sum X_t^2 + \bar{X}^2} = \beta \frac{\sum X_t^2}{\sum X_t^2 + \bar{X}^2} = \beta - \frac{\beta \bar{X}^2}{\sum X_t^2 + \bar{X}^2}$$

así que generalmente (siempre que $\bar{X}$ sea distinta de cero), el estimador $\hat{\beta}^*$ será sesgado.

3) Obtén la varianza de $\hat{\beta}^*$.

Solución:

$$\begin{aligned} Var(\hat{\beta}^*) &= E\left(\frac{\sum X_t Y_t}{\sum X_t^2 + \bar{X}^2} - \beta \frac{\sum X_t^2}{\sum X_t^2 + \bar{X}^2} \right)^2 = \frac{E(\sum X_t Y_t - \beta \sum X_t^2)^2}{(\sum X_t^2 + \bar{X}^2)^2} = \\ &= \frac{E\left(\sum X_t(\beta X_t + u_t) - \beta \sum X_t^2\right)^2}{(\sum X_t^2 + \bar{X}^2)^2} = \frac{\sum X_t^2 \overbrace{E(u_t^2)}^{=\sigma^2}}{(\sum X_t^2 + \bar{X}^2)^2} = \frac{\sigma^2 \sum X_t^2}{(\sum X_t^2 + \bar{X}^2)^2} \end{aligned}$$

4) En base al teorema de Gauss Markov el estimador de MCO es el de varianza mínima, ¿podría ocurrir que $V(\hat{\beta}^*)$ fuera menor que $V(\hat{\beta}_{MCO})$? Razona qué estimador elegirías.

Solución:

El teorema de Gauss-Markov garantiza que los estimadores MCO son los de mínima varianza dentro de la clase de los estimadores lineales e insesgados, sin embargo hemos visto que $\hat{\beta}^*$ es sesgado, por tanto no está dentro de esta clase y puede tener una varianza menor que la de MCO. A partir de la expresión obtenida en el apartado anterior, la varianza de $\hat{\beta}^*$ se puede expresar también como

$$Var(\hat{\beta}^*) = \frac{\sigma^2}{\sum X_t^2 - T\bar{X}^2 + (T+2)\bar{X}^2 + \frac{\bar{X}^4}{\sum X_t^2}}$$

Como los dos últimos términos del denominador son siempre positivos (suponiendo $\bar{X} \neq 0$), esta varianza será siempre menor que la varianza del estimador MCO.

Dado que $\hat{\beta}$ es un estimador insesgado y $\hat{\beta}^*$ es sesgado, para elegir entre ambos, utilizaríamos el error cuadrático medio, que es una medida que tiene en cuenta tanto la varianza como el sesgo. Si $\hat{\theta}$ es un estimador cualquiera, $ECM(\hat{\theta}) = Var(\hat{\theta}) + \text{sesgo}(\hat{\theta})^2$. Elegiríamos aquél estimador que presente menor *Error Cuadrático Medio*.

LADE-2004.4 (Jun-2004)

Una asociación de productores de txakoli quiere estudiar la influencia que tiene la cantidad de empleados (X_1) y de abono (X_2) utilizados por hectárea sobre la producción de litros de txakoli por hectárea (Y), por lo que proponen la siguiente relación funcional:

$$lnY_i = \beta_0 + \beta_1 lnX_{1i} + \beta_2 lnX_{2i} + u_i \quad i = 1, 2, \ldots, N \quad u_i \overset{iid}{\sim} N(0, \sigma_u^2) \tag{1}$$

A partir de una muestra de 10 productores, se ha obtenido la siguiente información muestral:

$$\begin{array}{lll}
\sum_{i=1}^{10} lnY_i = 11{,}8 & \sum_{i=1}^{10} lnY_i lnX_{1i} = 7{,}1 & \sum_{i=1}^{10} (lnY_i)^2 = 19{,}34 \\
\sum_{i=1}^{10} lnX_{1i} = 2 & \sum_{i=1}^{10} lnY_i lnX_{2i} = 4{,}1 & \sum_{i=1}^{10} (lnX_{1i})^2 = 7 \\
\sum_{i=1}^{10} lnX_{2i} = 2 & \sum_{i=1}^{10} lnX_{1i} lnX_{2i} = 1 & \sum_{i=1}^{10} (lnX_{2i})^2 = 7
\end{array}$$

$$\begin{array}{ll}
\sum_{i=1}^{10} (lnY_i - lnX_{1i})(lnX_{2i} - lnX_{1i}) = 3 & \sum_{i=1}^{10} (lnY_i - lnX_{2i})(lnX_{1i} - lnX_{2i}) = 9 \\
\sum_{i=1}^{10} (lnX_{2i} - lnX_{1i})^2 = 12 & \sum_{i=1}^{10} (lnX_{1i} - lnX_{2i})^2 = 12 \\
\sum_{i=1}^{10} (lnY_i - lnX_{1i})^2 = 12{,}14 & \sum_{i=1}^{10} (lnY_i - lnX_{2i})^2 = 18{,}14
\end{array}$$

[0] CVS Id: $Id: ladejun04.tex,v 1.1 2004/07/21 09:54:56 etpdihei Exp

a) Estima los coeficientes del modelo propuesto por Mínimos Cuadrados Ordinarios e interpreta los mismos si:

$$(X'X)^{-1} = \frac{1}{432} \begin{bmatrix} 48 & -12 & -12 \\ -12 & 66 & -6 \\ -12 & -6 & 66 \end{bmatrix}$$

b) Obtén una medida de la bondad del ajuste e interprétala.

c) ¿Es la variable X_2 significativa en el modelo propuesto? ¿Son las variables explicativas del modelo conjuntamente significativas?

d) ¿Crees posible la idea de que un productor con dos empleados y tres unidades de abono produzca 15 litros de txakoli por hectárea?

e) Realiza el contraste adecuado para verificar la existencia de rendimientos constantes de escala: $\beta_1 + \beta_2 = 1$.

f) Suponiendo que la hipótesis del apartado anterior es cierta:

   i) Propón y estima un modelo que recoja dicha hipótesis en el estudio (usa álgebra matricial en la estimación).

   ii) ¿Cuáles son las propiedades de estos nuevos estimadores?

   iii) Calcula la suma de los residuos al cuadrado y la suma de cuadrados total.

LADE-2004.5 (Jun-2004)

Comenta y **razona** la veracidad o no de las siguientes cuestiones:

1) Si en el modelo $Y_t = \alpha + \beta X_t + \gamma Z_t + u_t$ queremos contrastar $\beta = \gamma = 7$ podemos utilizar el estadístico:

$$\frac{\widehat{\beta} - \widehat{\gamma} - 7}{\sqrt{\widehat{var}(\widehat{\beta} - \widehat{\gamma} - 7)}} \sim t_{(T-K)} \qquad (2)$$

2) Si se quisiera estimar el salario del empleado i-ésimo (S_i) en función del turno de trabajo (Mañanas o Tardes) y se define:

$$T_i = \begin{cases} 1 & \text{si el empleado i-ésimo trabaja por las Tardes} \\ 0 & \text{en caso contrario} \end{cases}$$

$$M_i = \begin{cases} 1 & \text{si el empleado i-ésimo trabaja por las Mañanas} \\ 0 & \text{en caso contrario} \end{cases}$$

i) Razona, en términos del salario medio, si se puede emplear cualquiera de las dos siguientes especificaciones:

$$S_i = \alpha_1 + \alpha_2 M_i + u_i \quad i = 1, 2 \ldots, N \tag{3}$$

$$S_i = \beta_1 M_i + \beta_2 T_i + u_i \quad i = 1, 2 \ldots, N \tag{4}$$

ii) La equivalencia entre ambas especificaciones es: $\alpha_1 = \beta_1$ y $\alpha_2 = \beta_2$.

iii) Si $N_j, j = M, T$ indican el número de trabajadores en la muestra que tienen turno de Mañana y turno de Tarde respectivamente, ¿son los estimadores MCO de los coeficientes de los modelos (3) y (4) las siguientes?

$$\widehat{\beta}_1 = \frac{\sum^{N_M} S_i}{N_M} = \widehat{\alpha}_1 + \widehat{\alpha}_2$$

$$\widehat{\beta}_2 = \frac{\sum^{N_T} S_i}{N_T} = \widehat{\alpha}_1$$

LADE-2004.6 (Jun-2004)

El gerente de una empresa pide a tres estudiantes en prácticas analizar el comportamiento de las ventas de la empresa en el periodo $1981 - 2000$. El primer estudiante decide presentar los siguientes resultados:

$$\underset{(7,94)}{\widehat{V}_t = 5,76} + \underset{(4,02)}{3,8} \; RM_t - \underset{(-0,035)}{1,93} \; P_t \quad R^2 = 0,85367 \tag{5}$$

donde V_t son las ventas en miles de unidades, RM_t es la renta media de la población y P_t es el precio en el año t. El segundo estudiante presenta las siguientes estimaciones:

$$\underset{(7,88)}{\widehat{V}_t = 5,78} + \underset{(5,61)}{4,3} \; RM_t + \underset{(3,22)}{2,6} \; G_t - \underset{(-3,15)}{3,9} \; P_t \quad R^2 = 0,91121 \tag{6}$$

donde G_t es el gasto realizado por la empresa en publicidad. Por último, el tercer estudiante, que ha visto los resultados de sus compañeros, decide especificar las ventas a través de la relación:

$$V_t = \alpha + \beta \quad RM_t + \gamma \quad G_t + u_t \quad t = 1981, \ldots, 2000 \tag{7}$$

i) ¿Por qué no aconsejarías a este estudiante estimar los parámetros del modelo (7) por MCO? Razona tu respuesta realizando los contrastes necesarios.

ii) Demuestra las propiedades que tendrían los estimadores MCO de los parámetros del modelo (7).

iii) ¿Son las propiedades de estos estimadores las mismas que las del modelo (5)? ¿Qué concluyes sobre los estadísticos t del modelo (5)?

LE-2005.1 (Feb-2005)

Se desea estudiar la función de gasto en espectáculos de las economías familiares. Para ello se dispone de observaciones para 100 familias en el año 2004 sobre las siguientes variables:

Y_i: Gasto en espectáculos (cine, teatro y ópera) de la familia i, medida en cientos de euros.
X_{2i}: Renta disponible, medida en miles de euros, de la familia i.
X_{3i}: Gasto en compra y alquiler de vídeos y DVD en cientos de euros de la familia i.
X_{4i}: Años del cabeza de familia.

Se propone el siguiente modelo:

$$Y_i = \beta_1 + \beta_2 X_{2i} + \beta_3 X_{3i} + u_i \qquad u_i \sim N(0, \sigma^2) \tag{1}$$

El resumen de la información disponible es el siguiente:

$$(X'X) = \begin{bmatrix} 100 & 200,48 & 17,97 \\ 200,48 & 468,15 & 35,54 \\ 17,97 & 35,54 & 3,6 \end{bmatrix} \qquad (X'X)^{-1} = \begin{bmatrix} 0,1736 & -0,0342 & -0,5294 \\ -0,0342 & 0,0152 & 0,0200 \\ -0,5294 & 0,0200 & 2,7232 \end{bmatrix}$$

$$\sum Y_i = 3125 \quad \sum Y_i X_{2i} = 7092 \quad \sum Y_i^2 = 112262,89 \quad \sum Y_i X_{3i} = 551$$

1. Interpreta los parámetros del modelo propuesto. ¿Qué signos esperarías que tuvieran?

   Solución:
   $\beta_1 = E(Y_i / X_{2i} = X_{3i} = 0)$: valor esperado del gasto en espectáculos cuando la renta es cero y el gasto en alquiler y compra de videos y DVD es cero.
   $\beta_2 = \frac{\partial E(Y_i)}{\partial X_{2i}}$: incremento en el valor esperado del gasto en espectáculos cuando la renta disponible se incrementa en una unidad (1000 euros) y el resto de variables permanecen constantes. Esperamos que sea positivo.
   $\beta_3 = \frac{\partial E(Y_i)}{\partial X_{3i}}$: incremento en el valor esperado del gasto en espectáculos cuando el gasto en alquiler y compra de videos y DVD se incrementa en una unidad (100 euros) y el resto de variables permanecen constantes. Esperamos que sea negativo.

2. Escribe la función a minimizar que se corresponde con el criterio mínimo cuadrático ordinario.

   Solución: El criterio MCO es:

$$\min_{\hat\beta_1,\, \hat\beta_2,\, \hat\beta_3} \sum_{i=1}^{100} \hat u_i^2 = \min_{\hat\beta_1,\, \hat\beta_2,\, \hat\beta_3} \sum_{i=1}^{100} (Y_i - \hat\beta_1 - \hat\beta_2 X_{2i} - \hat\beta_3 X_{3i})^2$$

3. Estima el modelo propuesto por el método de mínimos cuadrados ordinarios.

   Solución:

$$\widehat\beta_{MCO} = \begin{bmatrix} \hat\beta_1 \\ \hat\beta_2 \\ \hat\beta_3 \end{bmatrix} = (X'X)^{-1} X'Y =$$

$$\begin{bmatrix} 0,1736 & -0,0342 & -0,5294 \\ -0,0342 & 0,0152 & 0,0200 \\ -0,5294 & 0,0200 & 2,7232 \end{bmatrix} \begin{bmatrix} 3125 \\ 7092 \\ 551 \end{bmatrix} = \begin{bmatrix} 8{,}2542 \\ 11{,}9434 \\ -12,0518 \end{bmatrix}$$

$$\widehat{Y}_i = 8{,}2542 + 11{,}9434X_{2i} - 12,0518X_{3i}$$

4. Calcula una medida de la bondad del ajuste e interprétala.

Solución:

$$R^2 = 1 - \frac{SCR}{SCT} = 1 - \frac{8406,464}{14606,62} = 0{,}4244$$
$$SCT = \sum_i Y_i^2 - T\bar{Y}^2 = 112262{,}89 - 100\left(\frac{3125}{100}\right)^2 = 14606{,}64$$
$$SCR = \sum_i Y_i^2 - \hat{\beta}'X'Y = (8{,}2542 \;\; 11{,}9434 \;\; -12,0518)\begin{bmatrix} 3125 \\ 7092 \\ 551 \end{bmatrix} =$$
$$= 112262{,}89 - 103856{,}426 = 8406{,}464$$

Interpretación: El 42,44 % de la variación en el gasto en espectáculos (cine, teatro, ópera) es explicado por la variación de la renta disponible y la variación del gasto en compra y alquiler de videos y DVD.

5. Estima la matriz de varianzas y covarianzas del estimador de los coeficientes del modelo.

Solución:

$$\hat{V}(\hat{\beta}) = \hat{\sigma}^2(X'X)^{-1}$$
$$\hat{\sigma}^2 = \frac{\hat{u}'\hat{u}}{100-3} = \frac{8406,464}{100-3} = 86,6645$$
$$\hat{V}(\hat{\beta}) = 86,6645\;(X'X)^{-1} = \begin{bmatrix} 15{,}0449 & -2{,}9639 & -45{,}880 \\ & 1{,}3173 & 1{,}7332 \\ & & 236{,}004 \end{bmatrix}$$

6. Contrasta la significatividad individual y conjunta de las variables explicativas.

Solución: Significatividad individual:

$$\begin{array}{ll} H_0 : \beta_2 = 0 & \\ H_a : \beta_2 \neq 0 & \end{array} \qquad \frac{\widehat{\beta_2}}{\hat{\sigma}_{\hat{\beta}_2}} \sim t(100-3)$$

Dados los resultados obtenidos anteriormente, $\mid \frac{11{,}9434}{\sqrt{1{,}3173}} \mid = 10{,}4060 > 1{,}96 \simeq t(100-3)_{0{,}05/2}$ por lo que se rechaza la Ho a un nivel de significación del $5\,\%$, la renta disponible sí es una variable significativa para explicar el gasto en espectáculos.

$$\begin{array}{ll} H_0 : \beta_3 = 0 & \\ H_a : \beta_3 \neq 0 & \end{array} \qquad \frac{\widehat{\beta_3}}{\hat{\sigma}_{\hat{\beta}_3}} \sim t(100-3)$$

Dados los resultados obtenidos anteriormente, $\mid \frac{-12{,}0518}{\sqrt{236{,}004}} \mid = 0{,}7844 < 1{,}96 \simeq t(100-3)_{0{,}05/2}$ por lo que no se rechaza la Ho a un nivel de significación del $5\,\%$, el gasto en compra y alquiler de videos y DVD no es una variable significativa para explicar el gasto

en espectáculos.
Significatividad conjunta de las variables:

$$H_0 : \beta_2 = \beta_3 = 0$$
$$H_a : \beta_2 \neq 0 \quad \text{y/o} \quad \beta_3 \neq 0$$

Para llevar a cabo el contraste, podemos utilizar el estadístico:

$F = \frac{R^2}{1-R^2}\frac{100-3}{2} \sim \mathcal{F}(2, 100-3)$, donde $q = 2$ es el número de restricciones a contrastar. Como el valor del estadístico $F = \frac{0{,}4244}{1-0{,}4244}\frac{100-3}{2} = 35{,}784$ es mayor que el valor de las tablas $\mathcal{F}(2, 100-3)_{0{,}05} \simeq 3,07$ rechazamos la hipótesis nula a un nivel de significación del $5\,\%$. Esto significa que, las dos variables explicativas de la regresión son conjuntamente significativas a un nivel de significación del $5\,\%$.

7. Contrasta $H_0 : \beta_2 = -\beta_3$. A la vista de los resultados del contraste, ¿qué concluyes sobre la especificación del modelo?

Solución: La hipótesis nula $\begin{array}{l} H_0 : \beta_2 = -\beta_3 \\ H_a : \beta_2 \neq -\beta_3 \end{array}$ puede escribirse $\begin{array}{l} H_0 : \beta_2 + \beta_3 = 0 \\ H_a : \beta_2 + \beta_3 \neq 0 \end{array}$ Sea $w = \beta_2 + \beta_3$, el contraste requiere comprobar si $\frac{\widehat{w}}{\hat{\sigma}_{\widehat{w}}}$ pertenece o no a la región crítica. Como $\widehat{w} = 11{,}9434 - 12{,}0518 = -0{,}1084$, y $\widehat{\sigma^2}_{\widehat{w}} = \widehat{var}(\hat{\beta}_2) + \widehat{var}(\hat{\beta}_3) + 2\widehat{cov}(\hat{\beta}_2, \hat{\beta}_3) = 1{,}3173 + 236{,}004 + 2 \times 1{,}7352$ entonces $\mid \frac{-0{,}1084}{15{,}4613} \mid = 0{,}007011 < 1{,}96 \simeq t_{97}^{0{,}025}$ No se rechaza la hipótesis nula al $5\,\%$ de significación; por lo tanto, la restricción es aceptada por los datos. Cuando una restricción es cierta, incluirla en el modelo proporciona un estimador de los coeficientes insesgado y de menor varianza que el que se obtiene sin tener en cuenta la restricción. Por lo tanto, debemos en este caso especificar y estimar el modelo restringido.

8. Estima el modelo restringido correspondiente a la hipótesis nula anterior.

Solución: El modelo restringido será el que incluye la restricción

$$Y_i = \beta_1 + (-\beta_3)X_{2i} + \beta_3 X_{3i} + u_i$$
$$Y_i = \beta_1^R + \beta_3^R(X_{3i} - X_{2i}) + u_i$$

Definiendo $Z_i = X_{3i} - X_{2i}$, se trata ahora de un modelo de regresión simple y la estimación de sus coeficientes se deduce de las siguientes expresiones:

$$\widehat{\beta}_3^R = \frac{\sum(Z_i - \bar{Z})(Y_i - \bar{Y})}{\sum(Z_i - \bar{Z})^2} = \frac{\sum Z_i Y_i - T\bar{Z}\bar{Y}}{\sum Z_i^2 - T\bar{Z}^2} = \frac{-837{,}5625}{67{,}5709} = -12{,}3953,$$

ya que $\bar{Y} = 3125/100 = 31{,}25$, $\quad \bar{Z} = \bar{X}_3 - \bar{X}_2 = 17{,}97/100 - 200{,}48/100 = -1{,}8251$, $\sum Z_i^2 = \sum(X_{3i} - X_{2i})^2 = \sum X_{3i}^2 + \sum X_{2i}^2 - 2\sum X_{2i}X_{3i} = 3{,}6 + 468{,}15 - 2 \times 35{,}54 = 400{,}67$, $\sum Z_i Y_i = \sum(X_{3i} - X_{2i})Y_i = \sum X_{3i}Y_i - \sum X_{2i}Y_i = 551 - 7092 = -6541$. Por otro lado,

$$\widehat{\beta}_1^R = \bar{Y} - \widehat{\beta}_3^R\bar{Z} = 31{,}25 - (-12{,}3953)(-1{,}8251) = 8{,}6273.$$

Por último, el coeficiente que acompaña a la variable X_2 toma el valor de $\widehat{\beta}_2^R = \widehat{\beta}_3^R = -12{,}3953$.

A continuación se reestima el modelo (1) incluyendo la variable X_4 para la misma muestra de familias.

$$Y_i = \gamma_1 + \gamma_2 X_{2i} + \gamma_3 X_{3i} + \gamma_4 X_{4i} + u_i \tag{2}$$

Los resultados de reestimar el modelo por MCO:

$$\underset{\text{(estadístico-t)}}{\hat{Y}_i} = \underset{(0{,}1508)}{0,7911} + \underset{(10{,}8581)}{12,4656} X_{2i} - \underset{(\text{-}8{,}3215)}{12,7423} X_{3i} + 0,1829 X_{4i} \qquad R^2 = 0,7158$$

9. ¿Por qué aconsejarías especificar el modelo (2) en vez del (1)? Razona tu respuesta.

Solución: La diferencia entre el modelo (1) y el (2) radica en la inclusión de la variable X_4. Los resultados obtenidos en (2) indican que las variables X_2 y X_3 resultan individualmente significativas. Por lo tanto, en el contraste conjunto de significación en este modelo se rechazará la hipótesis de que ninguna de las variables explicativas sea significativa. Además, nos interesa saber si la nueva variable añadida, X_4, es significativa o no. Para saberlo, llevamos a cabo el contraste:

$$H_0 : \gamma_4 = 0 \quad \text{Modelo restringido = Modelo (1)}$$
$$H_a : \gamma_4 \neq 0 \quad \text{Modelo no restringido = Modelo (2)}$$

El contraste puede entonces llevarse a cabo usando el estadístico $\frac{(R^2_{NR} - R^2_R)/q}{(1 - R^2_{NR})/(100-4)}$ y comparándolo con el valor de una variable con distribución F^1_{100-4}. El resultado, para este caso, es $\frac{(0{,}7158 - 0{,}4244)/1}{(1 - 0{,}7158)/96} = 98{,}43 > 3{,}07 = F^1_{96(0{,}05)}$. Por lo tanto, se rechaza la hipótesis nula y resulta significativa la variable X_4 en el modelo (2). Esto significa que en el modelo (1) hemos incurrido en un error por omisión de variable relevante, al no incluir esta variable, y por lo tanto la estimación de los coeficientes y de la varianza obtenidas en los primeros apartados resulta sesgada y la inferencia que hemos realizado no es válida. Consecuentemente, elegiremos el modelo (2) que, además, presenta un valor del R^2 bastante mayor como consecuencia de la inclusión de X_4 junto al resto de variables.

10. Dadas las estimaciones de γ_2 y γ_3 en el modelo (2) se ha reespecificado el modelo obteniendo los siguientes resultados:

$$\underset{\text{(estadístico-t)}}{\hat{Y}_i} = \underset{(0{,}3153)}{0,9043} + \underset{(16{,}3898)}{12,4526} (X_{2i} - X_{3i}) + \underset{(3{,}1775)}{0,1798} X_{4i} \qquad R^2 = 0,7060 \tag{3}$$

¿cuál de los tres modelos propuestos está mejor especificado? ¿Por qué?

Solución: Según acaba de verse, el modelo (1) no es un modelo correctamente especificado, ya que omite una variable que resulta ser relevante, por lo que sus resultados son sesgados y no debemos utilizarlos. La elección está, por lo tanto, entre los modelos (2) y (3). La diferencia entre ambos está en la restricción que se ha introducido en (3). Esa misma restricción la hemos contrastado en (1) y ha resultado aceptada. No obstante, ese resultado no es válido debido a la mala especificación del modelo. Además, el modelo (2) es diferente del inicial, por lo que se debe contrastar si la restricción es cierta en este modelo. Para ello, contrastamos

$$H_0 : \gamma_2 = -\gamma_3 \quad \text{Modelo restringido = Modelo (3)}$$
$$H_a : \gamma_2 \neq -\gamma_3 \quad \text{Modelo no restringido = Modelo (2)}$$

contraste que puede llevarse a cabo con el mismo estadístico que se ha utilizado en el apartado anterior. Dados los modelos restringido y no restringido en este caso, el resultado es: $\frac{(R^2_{NR} - R^2_R)/q}{(1 - R^2_{NR})/(100-4)} = \frac{(0{,}7158 - 0{,}7060)/1}{(1 - 0{,}7158)/96} = 3{,}31 > 3{,}07 = F^1_{96(0{,}05)}$. Por lo tanto, se rechaza la hipótesis nula a un nivel de significación del $5\,\%$, lo que significa que no hay evidencia de que la restricción sea cierta. En este caso, sabemos que si se incorpora al modelo una restricción que no es cierta se sesgan los resultados de la estimación. Por lo tanto, no debemos incorporar la restricción y elegiremos el modelo (2). En éste, la capacidad explicativa de las tres variables incorporadas es bastante satisfactoria y cada una de las variables resulta individualmente significativa, no apreciándose ningún problema en los resultados obtenidos.

11. Suponiendo que todas las familias acuden a la cadena de cines Cinesa S.A. y conocido el precio de una entrada de cine en esta cadena, queremos incorporar ésta variable a la ecuación (1). Escribe las cuatro primeras observaciones de la matriz de regresores del modelo en el que hayas incorporado la variable precio de la entrada de cine. ¿Te permite el modelo propuesto estudiar si el precio de la entrada de cine es una variable significativa para el gasto en espectáculos de las familias? Razona tú respuesta y si tienes algún problema propón una solución al mismo.

    Solución: En el modelo (1) se explica el gasto en espectáculos de las 100 familias de la muestra, en función de su renta disponible y su gasto en compra y alquiler de vídeos. Realmente, dados los resultados anteriores, deberíamos incluir también en el modelo la variable edad del cabeza de familia, y utilizar el modelo (2). Pero la cuestión que aquí se plantea es común a cualquiera de los dos modelos.
    Si todas las familias van a un mismo cine (o cadena de cines), todos pagan el mismo precio por la entrada. Si queremos incorporar esta variable al modelo (1) (o al modelo (2)), nos encontramos con una variable que toma el mismo valor en todas las informaciones. Así, si llamamos P_i a la nueva variable, resulta que $P_i = P \quad \forall i$. Por lo tanto, las primeras cuatro observaciones de las variables explicativas en el modelo $Y_i = \delta_1 + \delta_2 X_{2i} + \delta_3 X_{3i} + \delta_4 P_i + u_i$ serán

$$\begin{bmatrix} 1 & X_{21} & X_{31} & P \\ 1 & X_{22} & X_{32} & P \\ 1 & X_{23} & X_{33} & P \\ 1 & X_{24} & X_{34} & P \\ \vdots & \vdots & \vdots & \vdots \end{bmatrix}$$

    lo que significa que la cuarta columna es combinación lineal de la primera, ya que Columna $4 = P \times$ Columna 1. Tendremos multicolinealidad exacta en el modelo, porque lo mismo sucederá en todas las demás observaciones. Eso significa que el modelo que he escrito más arriba no se puede estimar, debido a que la matriz inversa $(X'X)^{-1}$ no existe. El problema se deriva de que no hay variabilidad en los datos de precio. Como el coeficiente δ_4 en el modelo representa la variación en Y ante variaciones en P, es imposible captar el efecto de esta variable cuando sus valores no cambian entre las diferentes observaciones.

LE-2005.2 (Feb-2005)

4 Para realizar un estudio sobre el ahorro de un colectivo de familias se propone el siguiente modelo:

$$Y_i = \alpha_1 D_{1i} + \alpha_2 D_{2i} + \gamma_1 F_{1i} + \gamma_2 F_{2i} + \beta X_i + u_i \qquad i = 1, \ldots, N \tag{4}$$

donde

$$D_{1i} = \begin{cases} 1 & \text{si en la familia i la mujer trabaja fuera del hogar} \\ 0 & \text{en caso contrario} \end{cases} \qquad F_{1i} = \begin{cases} 1 & \text{si i es una familia numerosa} \\ 0 & \text{en caso contrario} \end{cases}$$

$$D_{2i} = \begin{cases} 1 & \text{si en la familia i la mujer no trabaja fuera del hogar} \\ 0 & \text{en caso contrario} \end{cases} \qquad F_{2i} = \begin{cases} 1 & \text{si i no es una familia numerosa} \\ 0 & \text{en caso contrario} \end{cases}$$

X_i: Renta disponible $\qquad$ Y_i: Ahorro familiar

1. ¿Es el modelo (4) estimable? ¿Por qué? Propón un modelo estimable e interpreta sus parámetros.

   Solución: No. El modelo (4) presenta multicolinealidad exacta ya que se han incluido todas las categorías posibles de los dos conjuntos de variables ficticias. Para evitarla bastaría con quitar una variable ficticia de cada conjunto e incluir un término constante. Así, un posible modelo estimable sería:

   $$Y_i = \mu + \alpha_2^\star D_{2i} + \gamma_1^\star F_{1i} + \beta^\star X_i + u_i \qquad i = 1, \ldots, N$$

   donde:

   - μ es la base de comparación, $\mu = E[Y_i/D_{2i} = 0, F_{1i} = 0, X_i = 0]$ por tanto μ es el valor esperado del ahorro familiar de una familia no numerosa, cuya mujer trabaja fuera del hogar siendo la renta disponible cero.
   - $\alpha_2^\star$ es el diferencial en el ahorro familiar esperado cuando la mujer no trabaja fuera del hogar.
   - $\gamma_1^\star$ es el diferencial en el ahorro familiar esperado cuando la familia es numerosa.
   - $\beta^\star = \dfrac{\partial E(Y_i)}{\partial X_i}$ es el incremento en el valor esperado del ahorro familiar cuando la renta disponible aumenta en una unidad, permaneciendo el resto constante.

2. Si se dispone de la siguiente información: $\hat{Y}_i = 40,5 + 10{,}5 D_{2i} - 7,6 F_{1i} + 0{,}4 X_i$, calcula las estimaciones de los coeficientes del modelo: $Y_i = \alpha_1 D_{1i} + \gamma_1 F_{1i} + \gamma_2 F_{2i} + \beta X_i + u_i$.

   Solución: Necesitamos comparar ambos modelos, para ello denotamos:

   $$\hat{Y}_i = 40,5 + 10{,}5 D_{2i} - 7,6 F_{1i} + 0{,}4 X_i, \Longleftrightarrow Y_i = \mu + \alpha_2^\star D_{2i} + \gamma_1^\star F_{1i} + \beta^\star X_i + u_i \tag{5}$$

   $$Y_i = \alpha_1 D_{1i} + \gamma_1 F_{1i} + \gamma_2 F_{2i} + \beta X_i + u_i \tag{6}$$

	(5)	(6)
$E(Y_i/D_{2i}=0,F_{1i}=0,X_i=0)$	μ	$\alpha_1+\gamma_2$
$E(Y_i/D_{2i}=0,F_{1i}=1,X_i=0)$	$\mu+\gamma_1^\star$	$\alpha_1+\gamma_1$
$E(Y_i/D_{2i}=1,F_{1i}=1,X_i=0)$	$\mu+\alpha_2^\star+\gamma_1^\star$	γ_1
$E(Y_i/D_{2i}=1,F_{1i}=0,X_i=0)$	$\mu+\alpha_2^\star$	γ_2
$\frac{\partial E(Y_i)}{\partial X_i}$	$\beta^\star$	β

De (5) sabemos:

$$\hat{\mu}=40,5;\hat{\alpha}_2^\star=10,5;\hat{\gamma}_1^\star=-7,6;\hat{\beta}^\star=0,4 \qquad \text{de donde:}$$

$$\left.\begin{array}{l}\hat{\mu}=40,5\\ \hat{\mu}+\hat{\gamma}_1^\star=40,5-7,6=32,9\\ \hat{\mu}+\hat{\alpha}_2^\star+\hat{\gamma}_1^\star=40,5+10,5-7,6=43,4\\ \hat{\mu}+\hat{\alpha}_2^\star=40,5+10,5=51\\ \hat{\beta}^\star=0,4\end{array}\right\}\text{por lo que en (6)}\Longrightarrow\left\{\begin{array}{l}\hat{\gamma}_2=51\\ \hat{\gamma}_1=43,4\\ \hat{\alpha}_1=-10,5\\ \hat{\beta}=0,4\end{array}\right.$$

3. ¿Cómo contrastarías que el hecho de que la mujer trabaje fuera del hogar no influye en el ahorro familiar? Describe claramente todos los elementos del contraste.

   Solución: En el modelo (5) $\quad Y_i=\mu+\alpha_2^\star D_{2i}+\gamma_1^\star F_{1i}+\beta^\star X_i+u_i$ contrastamos:

   $$\left.\begin{array}{l}H_0:\alpha_2^\star=0\\ H_a:\alpha_2^\star\neq 0\end{array}\right\} \quad \text{con el estadístico} \quad \frac{\hat{\alpha}_2^\star}{\widehat{desv}(\hat{\alpha}_2^\star)}\overset{H_0}{\sim}t_{(N-4)}$$

   donde:

   $\hat{\alpha}_2^\star=10,5$ se obtiene como el segundo elemento del vector de parámetros estimados, $\hat{\beta}=(\hat{\mu}\ \ \hat{\alpha}_2^\star\ \ \hat{\gamma}_1^\star\ \ \hat{\beta}^\star)^T$

   $\widehat{desv}(\hat{\alpha}_2^\star)=\hat{\sigma}\sqrt{a_{22}}$ donde el elemento a_{22} es el segundo elemento en la diagonal de $(X'X)^{-1}$ y $\hat{\sigma}=\sqrt{\hat{\sigma}^2}$ siendo $\hat{\sigma}^2=\frac{\hat{u}'\hat{u}}{N-4}$

   Si $t_c>t_{(\alpha/2)}(N-4)$ para un nivel de significatividad α dado rechazo la H_0, $\alpha_2^\star\neq 0$ y el hecho de que la mujer trabaje fuera del hogar si influye en el ahorro familiar.

   En el modelo (6) $\quad Y_i=\alpha_1 D_{1i}+\gamma_1 F_{1i}+\gamma_2 F_{2i}+\beta X_i+u_i \quad$ la hipótesis nula equivalente es:

   $$\left.\begin{array}{l}H_0:\alpha_1=0\\ H_a:\alpha_1\neq 0\end{array}\right\} \quad \text{con el estadístico} \quad \frac{\hat{\alpha}_1}{\widehat{desv}(\hat{\alpha}_1)}\overset{H_0}{\sim}t_{(N-4)}$$

   donde:

   $\hat{\alpha}_1=-10,5$ se obtiene como el primer elemento del vector de parámetros estimados, $\hat{\beta}=(\hat{\alpha}_1\ \ \hat{\gamma}_1\ \ \hat{\gamma}_2\ \ \hat{\beta})^T$

   $\widehat{desv}(\hat{\alpha}_1)=\hat{\sigma}\sqrt{a_{11}}$ donde el elemento a_{11} es el primer elemento en la diagonal de $(X'X)^{-1}$ y $\hat{\sigma}=\sqrt{\hat{\sigma}^2}$ siendo $\hat{\sigma}^2=\frac{\hat{u}'\hat{u}}{N-4}$

   El criterio de decisión sería el mismo que el anterior.

   • En el modelo (4) no se puede realizar el contraste ya que no es estimable.

LE-2005.3 (Jun-2005)

La evolución de la relación entre producción, trabajo y capital para la empresa METALICAS VIZCAINAS de 1978 a 2004 queda recogida por la siguiente función:

$$Q_t = AL_t^{\beta_1} K_t^{\beta_2} e^{u_t} \tag{1}$$

donde Q se mide en términos del valor añadido, L es el valor del trabajo y K es el valor bruto de la planta y maquinaria, ambas en miles de euros,

1. ¿Incumple la ecuación (1) alguna hipótesis básica del Modelo de Regresión Lineal General? ¿Cuál? ¿Qué consecuencias tiene? ¿Cómo solucionarías el problema?

   Solución: Sí. El modelo no es lineal en los coeficientes. En consecuencia, dadas las técnicas de estimación que conocemos, el modelo no es estimable. En este caso el problema puede solucionarse porque el modelo es linealizable si tomamos logaritmos neperianos.

2. Propón un modelo que permita estimar los parámetros A, β_1 y β_2 por Mínimos Cuadrados Ordinarios, Escribe el modelo y la función a minimizar, Interpreta los parámetros del modelo propuesto.

   Solución: Tomando logaritmos neperianos tenemos:

$$ln(Q_t) = ln(AL_t^{\beta_1} K_t^{\beta_2} e^{u_t})$$

$$ln(Q_t) = \beta_0 + \beta_1\, ln(L_t) + \beta_2\, ln(K_t) + u_t$$

   donde $\beta_0 = ln(A)$. Una vez estimados los coeficientes por MCO, se puede recuperar A como $\hat{A} = exp(\hat{\beta}_0)$. La función a minimizar es la siguiente:

$$\min_{\hat{\beta}_0,\, \hat{\beta}_1,\, \hat{\beta}_2} \sum_{t=1}^{T} (ln(Q_t) - \hat{\beta}_0 - \hat{\beta}_1\, ln(L_t) - \hat{\beta}_2\, ln(K_t))^2.$$

   Interpretación de los coeficientes:
   $\beta_0 = ln(A) = E(ln(Q_t)/ln(L_t) = ln(Q_t) = 0) = E(ln(Q_t)/L_t = Q_t = 1)$: neperiano del valor añadido medio cuando el neperiano del valor del trabajo y el neperiano del valor de planta y maquinaria son cero (cuando el valor del trabajo y el el valor de planta y maquinaria son uno).
   $\beta_1 = \frac{\partial E(ln(Q_t))}{\partial ln(L_t)} = E\frac{\partial Q_t}{\partial L_t}\frac{L_t}{Q_t} \rightarrow$ el incremento esperado en $ln(Q_t)$ ante un incremento unitario en $ln(L_t)$ manteniendo el resto constante. *O BIEN:* La elasticidad del trabajo esperada sobre el valor añadido de la empresa manteniéndose constante el resto. *O BIEN:* Variación porcentual esperada en el valor añadido de la empresa ante una variación del $1\,\%$, manteniéndose el resto constante.
   $\beta_2 = \frac{\partial E(ln(Q_t))}{\partial ln(K_t)} = E\frac{\partial Q_t}{\partial K_t}\frac{K_t}{Q_t} \rightarrow$ el incremento esperado en $ln(Q_t)$ ante un incremento unitario en $ln(K_t)$ manteniendo el resto constante. *O BIEN:...*

   La apropiada transformación para la estimación del modelo (1) se corresponde con la siguiente ecuación:

$$Q_t^* = \alpha + \beta_1 L_t^* + \beta_2 K_t^* + u_t \qquad u_t \sim NID(0,\sigma^2) \tag{2}$$

para la cual disponemos de la siguiente información:

$$(X'X) = \begin{bmatrix} 27 & 155,6186 & 201,0399 \\ 155,6186 & 908,1284 & 1173,5062 \\ 201,0399 & 1173,5062 & 1521,3144 \end{bmatrix} \qquad (X'X)^{-1} = \begin{bmatrix} 3,0094 & -0,5590 & 0,0335 \\ -0,5590 & 0,4471 & -0,2710 \\ 0,0335 & -0,2710 & 0,2053 \end{bmatrix}$$

t	Q	L	K
1	657,29	162,31	279,99
2	935,93	214,43	542,50
⋮	⋮	⋮	⋮
27	745,67	137,00	768,59

$\sum Q_t^* = 200,9780$ $\quad \sum Q_t^* L_t^* = 1170,6730$

$\sum Q_t^* K_t^* = 1514,5429$ $\quad \sum Q_t^{*2} = 1593,6446$

3. Estima la ecuación (2) por el método de mínimos cuadrados ordinarios.

Solución:

$$\widehat{\beta} = (X'X)^{-1}X'Y = (X'X)^{-1}\begin{pmatrix} 200{,}9780 \\ 1170{,}6730 \\ 1514{,}5429 \end{pmatrix} = \begin{pmatrix} 1{,}154 \\ 0{,}62 \\ 0{,}41 \end{pmatrix}$$

$$\widehat{Q}_t^* = 1{,}154 + 0{,}62 L_t^* + 0{,}41 K_t^*$$

4. En la ecuación (2)

a) Calcula el valor del residuo correspondiente al año 1979.

Solución:

$$\hat{u}_{1979} = Q_{1979}^* - \hat{\beta}_0 - \hat{\beta}_1 L_{1979}^* - \hat{\beta}_2 K_{1979}^* =$$
$$ln(Q_{1979}) - \hat{\beta}_0 - \hat{\beta}_1 ln(L_{1979}) - \hat{\beta}_2 ln(K_{1979}) =$$
$$ln(935{,}93) - 1{,}154 - (0{,}62\ ln(214{,}43)) - (0{,}62\ ln(542{,}50)) = -0{,}2599$$

b) ¿Cuánto vale $\sum_{t=1}^T \hat{u}_t$? ¿Por qué?

Solución: $\sum_{t=1}^T \hat{u}_t = 0$. Sabemos, bien por las ecuaciones normales que se derivan de la estimación MCO o bien por las propiedades de la recta de regresión que: $X'\hat{u} = 0$. Si a esto le añadimos que en nuestro modelo tenemos un término constante, entonces el primer elemento del vector $X'\hat{u} = 0$ es $\sum_{t=1}^T \hat{u}_t = 0$.

5. Calcula una medida de la bondad del ajuste e interprétala,

Solución:

$$R^2 = 1 - \frac{SCR}{SCT} = 1 - \frac{15{,}7401}{97{,}6396} = 0{,}8335$$

$$SCR = Q^{*'}Q^* - \hat{\beta}' X' Q^* =$$

$$
\begin{aligned}
&= 1593{,}6446 - (1{,}154\ 0{,}62\ 0{,}41) \begin{pmatrix} 200{,}9780 \\ 1170{,}6730 \\ 1514{,}5429 \end{pmatrix} \\
&= 1593{,}6446 - 1577{,}9045 = 15{,}7401
\end{aligned}
$$

$$
SCT = Q^{*\prime} Q^* - T\bar{Q^*}^2 = 1593{,}6446 - 27 \left(\frac{200{,}978}{27} \right)^2 = 97{,}6396
$$

Explicamos, de forma lineal, el $83{,}35\,\%$ de la variación del neperiano del valor añadido de la empresa mediante las variaciones del neperiano del valor del trabajo y las variaciones del neperiano del valor de la planta y la maquinaria.

6. Contrasta la significatividad individual y conjunta de las variables explicativas.

Solución:
Varianza de la perturbación: $\widehat{\sigma^2} = \frac{\widehat{u}'\widehat{u}}{T-K-1} = \frac{15{,}7401}{27-2-1} = 0{,}6558$
Contraste significatividad individual:
$H_0 : \beta_1 = 0$ $\quad \frac{\widehat{\beta}_1 - 0}{\widehat{des}(\beta_1)} \overset{H_o}{\sim} t_{(T-K-1)}$
$H_a : \beta_1 \neq 0$

$$
\mid \frac{\widehat{\beta}_1}{\widehat{des}(\beta_1)} \mid = \frac{0{,}62}{\sqrt{0{,}6558 \times 0{,}4471}} = 1{,}1449 < 2,064 = t_{27-2-1}^{0{,}025}
$$

No se rechaza H_0 a un nivel de significatividad del $5\,\%$. La variable L^* no es significativa para explicar Q^*.

$H_0 : \beta_2 = 0$ $\quad \frac{\widehat{\beta}_2}{\widehat{des}(\beta_2)} \overset{H_o}{\sim} t_{(T-K-1)}$
$H_a : \beta_2 \neq 0$

$$
\mid \frac{\widehat{\beta}_2}{\widehat{des}(\beta_2)} \mid = \frac{0{,}41}{\sqrt{0{,}6558 \times 0{,}2053}} = 1{,}1173 < 2,064 = t_{27-2-1}^{0{,}025}
$$

No se rechaza H_0 a un nivel de significatividad del $5\,\%$. La variable K^* no es significativa para explicar Q^*.

Contraste de significatividad conjunto:
$H_0 : \beta_1 = \beta_2 = 0$ $\quad \frac{R^2}{1-R^2} \frac{T-K-1}{q} \overset{H_o}{\sim} F_{(q,T-K-1)}$
$H_a : \beta_1 \neq 0 \wedge \vee \beta_2 = 0$

$$
F = \frac{0{,}8335}{1 - 08335} \frac{27 - 2 - 1}{2} = 60{,}072 < 3{,}40 = F_{(2,27-3)}^{0{,}05}
$$

Se rechaza H_0 a un nivel de significatividad del $5\,\%$. Las variables L^* y K^* son conjuntamente significativas para explicar Q^*.

7. ¿Hay evidencia de algún problema muestral? Interpreta los resultados de los contrastes anteriores

Solución: Sí, parece haber un problema de multicolinealidad imperfecta entre las variables L^* y K^*, ya que individualmente no son significativas pero conjuntamente sí lo son. Debido a este problema muestral, y a pesar de que el estimador MCO tiene buenas propiedades (lineal en u, insesgado y de mínima varianza), las varianzas estimadas son elevadas. Esta es la razón por la cual se tiende a no rechazar H_o, por lo que en el apartado anterior hemos obtenido que las variables no son individualmente significativas.

8. ¿Tiene la empresa rendimientos constantes a escala?($H_0 : \beta_2 + \beta_3 = 1$)

Solución:

$H_0 : \beta_1 + \beta_2 = 1$ $\quad \frac{\widehat{\beta}_1+\widehat{\beta}_2-1}{\widehat{des}(\hat{\beta}_1+\hat{\beta}_2)} \overset{H_o}{\sim} t_{(T-K-1)}$

$H_a : \beta_1 + \beta_2 \neq 1$

$$\begin{aligned}\widehat{des}(\hat{\beta}_1+\hat{\beta}_2) &= \sqrt{\widehat{var}(\hat{\beta}_1)+\widehat{var}(\hat{\beta}_2)+2\widehat{cov}(\hat{\beta}_1,\hat{\beta}_2)} \\ &= \sqrt{0{,}6558(0{,}4471+0{,}2053-2\times 0{,}2710)} = 0{,}2690\end{aligned}$$

$$\mid \frac{\widehat{\beta}_1+\widehat{\beta}_2-1}{\widehat{des}(\widehat{\beta}_1+\widehat{\beta}_2)} \mid = \frac{0{,}62+0{,}41-1}{0{,}26901} = 1{,}111 < 2,064 = t^{0{,}025}_{27-2-1}$$

No se rechaza H_0 a un nivel de significatividad del $5\,\%$. La empresa tiene rendimientos constantes a escala

9. El gabinete de gerencia fija el valor añadido de la empresa para el 2005 en 1789,45 ¿Es un valor factible para un valor del factor trabajo de 267,890 miles de euros y un valor bruto de la planta y maquinaria de 1123,45 miles de euros?

Solución:

$$\begin{aligned}&Q^*_p = ln(Q_p) = ln(1789{,}45) = 7{,}4896\\ &L^*_p = ln(L_p) = ln(267{,}890) = 5{,}59\\ &K^*_p = ln(K_p) = ln(1123{,}45) = 7{,}024\\ &\hat{Q}^*{}_p = 1{,}154 + (0{,}62\times 5{,}59) + (0{,}41\times 7{,}024) = 7{,}4956\\ &t_{(27-2-1)0{,}025} = 2{,}064; \qquad X'_p = (1\ 5{,}59\ 7{,}024)' \quad X'_p(X'X)^{-1}X_p = 0{,}046\\ &IC(Q^*_p)_{1-\alpha} = \left[\hat{Q}^*{}_p \pm t_{(T-K-1)\alpha/2}\hat{\sigma}\sqrt{1+X'_p(X'X)^{-1}X_p}\right]\\ &IC(Q^*_p)_{0{,}95} = \left[74956 \pm 2{,}064\sqrt{0{,}6558}\sqrt{1+0{,}046}\right] = [5{,}7862; 9{,}205]\end{aligned}$$

Como $Q^*_p = ln(Q_p) = ln(1789{,}45) = 7{,}4896 \in [5{,}7862; 9{,}205]$, se trata de un valor posible.

LE-2005.4 (Jun-2005)

Continuando en la misma empresa del problema 1, se cree que la tecnología (T_t) puede también ser un factor importante para explicar la producción de la empresa, Por ello se ha estimado la siguiente especificación:

$$Q^*_t = \gamma_0 + \gamma_1 L^*_t + \gamma_2 K^*_t + \gamma_3 T^*_t + u_t \tag{3}$$

$$\underset{(\hat{\sigma}_{\hat{\beta}_i})}{\hat{Q}_t^*} = \underset{(1,07)}{\overbrace{2,05}^{\hat{\gamma}_0}} + \underset{(0,12)}{\overbrace{0,51}^{\hat{\gamma}_1}} L_t^* + \underset{(0,09)}{\overbrace{0,32}^{\hat{\gamma}_2}} K_t^* + \underset{(0,076)}{\overbrace{0,22}^{\hat{\gamma}_3}} T_t^*$$

donde T_t^* es la variable T_t transformada de forma similar al resto de variables,

10. ¿Es la nueva variable significativa? Dados los resultados obtenidos en el apartado anterior, ¿varían las conclusiones obtenidas hasta el momento? Razona tu respuesta,

Solución: Contrastamos

$$\begin{array}{l} H_0 : \gamma_3 = 0 \\ H_a : \gamma_3 \neq 0 \end{array} \qquad \frac{\widehat{\gamma_3}}{\widehat{des\gamma_3}} \overset{H_0}{\sim} t_{(T-K-1)} \Rightarrow \frac{0{,}22}{0{,}076} = 2{,}894 > 2{,}069 = t_{23}^{0{,}025} \Rightarrow \text{Rechazar } H_0$$

Por lo tanto, la nueva variable resulta significativa. Si realizamos el contraste individual para las otras variables explicativas, comprobaremos que en el modelo (3) también son significativas $L*$ y K^*, por lo que este modelo parece estar bien especificado. Esto significa que el modelo (2) no está bien especificado, al omitirse un regresor relevante. Por lo tanto, y dado que la variable T^* no es ortogonal a los otros dos regresores (cambia el valor de los coeficientes estimados), la estimación realizada en (2) es sesgada tanto para los coeficientes como para la varianza de la perturbación, lo que hace que los contrastes realizados sean inválidos. No podemos mantener las conclusiones obtenidas en los apartados anteriores.

11. ¿Cómo estimarías el modelo (3) si resultara que $2K_t^* = T_t^*$? ¿qué propiedades tiene tu estimador?

Solución: En este caso, existirá colinealidad perfecta entre dos regresores del modelo, K_t^* y T_t^*, lo que implica que la matriz $X'X$ es singular y no puede obtenerse un único valor para cada coeficiente del modelo a partir de la información muestral. Para comprobarlo, introducimos la relación en el modelo:

$$\begin{aligned} Q_t^* &= \gamma_0 + \gamma_1 L_t^* + \gamma_2 K_t^* + \gamma_3 (2K_t^*) + u_t \\ Q_t^* &= \gamma_0 + \gamma_1 L_t^* + (\gamma_2 + 2\gamma_3) K_t^* + u_t \end{aligned}$$

lo que significa que sólo pueden estimarse tres coeficientes, γ_0, γ_1 y $(\gamma_2 + 2\gamma_3)$ y no puede obtenerse un valor único para γ_2 ni para γ_3. Por otro lado, el resto de las hipótesis básicas del modelo no han cambiado, por lo que las propiedades del estimador MCO se mantienen, siendo insesgados y de mínima varianza.

12. ¿Cambiarías la respuesta del apartado anterior si supieras que, además de cumplirse que $2K_t^* = T_t^*$, hay rendimientos constantes a escala ($\gamma_1 + \gamma_2 + \gamma_3 = 1$)?, ¿cómo estimarías el modelo?, ¿qué propiedades tiene tu estimador?

Solución: Si a la relación anterior entre variables se le une esta restricción de los coeficientes, podemos introducirla en el modelo y comprobar si resuelve el problema que ha planteado la multicolinealidad. Dado que $\gamma_1 + \gamma_2 + \gamma_3 = 1 \Rightarrow \gamma_2 = 1 - \gamma_1 - \gamma_3$, por lo que

$$\begin{aligned} Q_t^* &= \gamma_0 + \gamma_1 L_t^* + (\gamma_2 + 2\gamma_3) K_t^* + u_t \\ Q_t^* &= \gamma_0 + \gamma_1 L_t^* + (1 - \gamma_1 - \gamma_3 + 2\gamma_3) K_t^* + u_t \\ \Rightarrow Q_t^* - K_t^* &= \gamma_0 + \gamma_1 (L_t^* - K_t^*) + \gamma_3 K_t^* + u_t \end{aligned}$$

Por lo tanto, ahora pueden estimarse todos los coeficientes: $\hat{\gamma}_0$, $\hat{\gamma}_1$ y $\hat{\gamma}_3$ a través del modelo transformado y $\hat{\gamma}_2$ = 1- $\hat{\gamma}_1$-$\hat{\gamma}_3$. El modelo obtenido en este apartado es un modelo restringido; por lo tanto los estimadores serán insesgados sólo si la restricción es cierta; si lo es, serán además más eficientes que los no restringidos.

13. Interpreta en términos económicos el modelo que has obtenido tras tener en cuenta la información de los apartados 11 y 12, ¿En qué se parece y en qué se diferencia del modelo inicial (1)?

Solución: Dada la definición de las variables, y teniendo en cuenta que la transformación que se ha introducido en el apartado 2 es la logarítmica, el modelo anterior puede escribirse

$$ln\left(\frac{Q_t}{K_t}\right) = \gamma_0 + \gamma_1 ln\left(\frac{L_t}{K_t}\right) + \gamma_3 lnK_t + u_t \quad (4)$$

$$\Rightarrow \frac{Q_t}{K_t} = e^{\gamma_0}\left(\frac{L_t}{K_t}\right)^{\gamma_1} K_t^{\gamma_3} e^{u_t} \quad (5)$$

Comparando el modelo (5) con el modelo (1) resulta evidente que no son iguales; ambos tienen la misma forma funcional y contienen las mismas variables, pero la forma en que éstas aparecen es diferente. Si interpretamos el modelo restringido, vemos que la variable dependiente es la productividad del capital (valor añadido por unidad de capital) que, según el modelo, depende del stock de capital existente y de la inversa de la relación capital/producto. De esta manera, la forma en que los inputs productivos influyen en el output es más elaborada en el modelo (5), que resulta menos genérico que el (1).

LE-2005.5 (Jun-2005)

En una empresa de consultoría se quiere analizar la dependencia del salario con respecto de los años de experiencia, Se dispone de información sobre las siguientes variables: Salario (W_i) y Nivel de experiencia clasificado en tres categorías (sin experiencia, poca experiencia, mucha experiencia), Para una muestra de 18 empleados (de los cuales cuatro no tienen experiencia, seis tienen poca y los restantes ocho tienen mucha experiencia) se ha obtenido la siguiente información:

$$\sum W_i S_i = 827 \quad \sum W_i P_i = 1654 \quad \sum W_i M_i = 2481 \quad \sum W_i^2 = 1398506$$

donde S_i toma valor uno cuando el empleado i no tiene experiencia y cero en caso contrario, P_i toma valor uno cuando el empleado i tiene poca experiencia y cero en caso contrario, finalmente, M_i toma valor uno cuando el empleado i tiene mucha experiencia y cero en caso contrario,

1. Propón un modelo explicativo del salario en función del nivel de experiencia, Interpreta sus coeficientes,

Solución: Pueden proponerse varios modelos estimables, según qué restricción impongamos para evitar la colinealidad exacta. En este caso, elegimos

$$W_i = \beta_1 S_i + \beta_2 P_i + \beta_3 M_i + u_i.$$

La interpretación de los coeficientes es la siguiente:

- $E(W_i|S_i = 1)$ = β_1 = Salario esperado de los trabajadores sin experiencia
- $E(W_i|P_i = 1)$ = β_2 = Salario esperado de los trabajadores con poca experiencia
- $E(W_i|M_i = 1)$ = β_3 = Salario esperado de los trabajadores con mucha experiencia

2. ¿Es la experiencia un factor determinante para el salario?

Solución: En el caso de que la experiencia no sea relevante, eso significaría que ningún grado de experiencia tendría influencia en el salario. Dado que ésta es la única variable explicativa incorporada en el modelo, en ese caso el modelo quedaría reducido a $W_i = \beta + u_i$. La hipótesis a contrastar es

$$H_0 : \beta_1 = \beta_2 = \beta_3; \Rightarrow MR : W_i = \beta + u_i.$$
$$H_a : \beta_1 \neq \beta_2 \text{ y/o } \beta_1 \neq \beta_3; \Rightarrow MNR : W_i = \beta_1 S_i + \beta_2 P_i + \beta_3 M_i + u_i$$

Para poder llevar a cabo el contraste, tenemos que comenzar por estimar el modelo.

$$\hat{\beta} = (X'X)^{-1}X'Y = \begin{bmatrix} \frac{1}{4} & 0 & 0 \\ 0 & \frac{1}{6} & 0 \\ 0 & 0 & \frac{1}{6} \end{bmatrix} \begin{bmatrix} 827 \\ 1654 \\ 2481 \end{bmatrix} = \begin{bmatrix} 206{,}75 \\ 276{,}66 \\ 310{,}12 \end{bmatrix}$$

$$\hat{u}'\hat{u}_R = 1398506 - \left(\frac{827 + 1654 + 2481}{18}\right)^2 18 = 30648$$

$$\hat{u}'\hat{u}_{NR} = 1398506 - \begin{bmatrix} 206{,}75 & 276{,}66 & 310{,}12 \end{bmatrix} \begin{bmatrix} 827 \\ 1654 \\ 2481 \end{bmatrix} = 2174{,}39$$

$$\frac{\hat{u}'\hat{u}_R - \hat{u}'\hat{u}_{NR}}{\hat{u}'\hat{u}_{NR}} \frac{15}{2} = \frac{30648 - 2174{,}39}{2174{,}39} \frac{15}{2} = 98{,}2 > 3{,}6 = F_{2,15}^{0,05}$$

Por lo tanto, se rechaza H_0 con un nivel de significación del 5 %, la experiencia es un factor determinante para fijar el salario.

3. A la vista de las estimaciones obtenidas, podríamos pensar que en cuanto al nivel de experiencia, únicamente es relevante tenerla o no. Contrástalo.

Solución: Si el nivel de experiencia no es importante, sólo si se tiene alguna experiencia, el modelo debería ser $W_i = \beta_1 S_i + \beta^* C_i + u_i$, siendo C_i una variable que toma el valor 1 si el individuo tiene experiencia y 0 si no. Por lo tanto, lo que tenemos que contrastar es:

$$H_0 : \beta_2 = \beta_3 \Rightarrow \quad MR : W_i = \beta_1 S_i + \beta^* C_i + u_i$$
$$H_a : \beta_2 \neq \beta_3 \Rightarrow \quad MNR : W_i = \beta_1 S_i + \beta_2 P_i + \beta_3 M_i + u_i$$

Para este caso, podemos proponer como estadístico de contraste el basado en $\frac{\hat{\beta}_2 - \hat{\beta}_3}{\hat{\sigma}_{(\hat{\beta}_2 - \hat{\beta}_3)}} \overset{H_o}{\sim} t_{(15)}$. Dado que $\hat{\sigma}^2_{(\hat{\beta}_2 - \hat{\beta}_3)} = \hat{\sigma}^2((X'X)^{-1}_{22} + (X'X)^{-1}_{33} - 2(X'X)^{-1}_{23}) = \frac{2174{,}369}{15}(\frac{1}{6} + \frac{1}{8}) = 42{,}28$, resulta $\mid \frac{275{,}66 - 310{,}12}{\sqrt{42{,}88}} \mid = 5{,}26 > 2{,}131 = t^{0{,}025}_{(15)}$. Por lo tanto, se rechaza la hipótesis nula: el nivel de experiencia es relevante.

4. Interpreta los resultados obtenidos en los apartados anteriores, ¿Qué puedes decir respecto de la relación entre experiencia y salarios en esa empresa?

Solución: Hemos visto que la experiencia es un factor determinante para el salario, en el punto 2. Además, no se acepta que $\beta_2 = \beta_3$ en el apartado 3. Por lo tanto, no sólo influye en el salario el hecho de tener o no experiencia sino que también influye el nivel de experiencia que se tenga, comprobándose además a partir de la estimaciones que mayor nivel de experiencia implica mayor salario.

LADE-2005.1 (Feb-2005)

En los EE.UU. se analizó en 1980 el gasto que los diferentes estados realizaban en las escuelas públicas. Sea Y_i el gasto por alumno **(en cientos de dólares)** que el estado i-ésimo realizó en sus escuelas públicas en 1980 y X_i la renta per capita **(en miles de dólares)** de dicho estado en el mismo periodo. Los resultados de la regresión MCO han sido:

$$\widehat{Y}_i = \underset{(16{,}576)}{0{,}0309} - \underset{(-3{,}553)}{0{,}025}\ X_i + \underset{(30{,}050)}{0{,}154}\ X_i^2$$

$$N = 50 \qquad R^2 = 0{,}9937 \qquad \widehat{\sigma}^2 = 0{,}0000277 \qquad \hat{u}'\hat{u} = 0{,}001303$$

$$(X'X) = \begin{pmatrix} 50 & 23{,}08 & 16{,}545 \\ & 16{,}545 & 15{,}298 \\ & & 16{,}481 \end{pmatrix} \qquad (X'X)^{-1} = \begin{pmatrix} 0{,}125 & -0{,}413 & 0{,}257 \\ & 1{,}789 & -1{,}245 \\ & & 0{,}957 \end{pmatrix}$$

$$X'Y = \begin{pmatrix} 3{,}532 \\ 2{,}669 \\ 2{,}681 \end{pmatrix}$$

1) Interpretar la medida de bondad del ajuste.

   Solución: La variabilidad de la renta *per capita* y su cuadrado explican el $99{,}37\,\%$ de la variabilidad del gasto en escuelas públicas.

2) ¿Es posible que un estado con una renta per capita igual a $8267\$$ realice un gasto por alumno de $440\$$?

   Solución: Será posible a partir de este modelo si ese gasto está incluido en el intervalo de confianza correspondiente a su predicción, $IC_{1-\alpha}(Y_P) = (\hat{Y}_P \pm t_{T-K}^{\alpha/2}\hat{\sigma}\sqrt{1 + X_P'(X'X)^{-1}X_P})$.

$$\hat{Y}_P = 0{,}0309 - 0{,}025 \cdot 8{,}267 + 0{,}154 \cdot (8{,}267)^2 = 10{,}349$$

$$t_{47}^{0{,}025} \simeq 2{,}01 \quad \hat{\sigma} = \sqrt{0{,}0000277} \quad X_P' = (1\ 8{,}267\ 8{,}267^2) \Rightarrow X_P'(X'X)^{-1}X_P = 3213{,}8168$$

$$\Rightarrow IC_{95\,\%}(Y_P) = (10{,}349 \pm 0{,}5998) = (9{,}749; 10{,}948)$$

   Como $4{,}40 \notin IC_{95\,\%}(Y_P)$, con un nivel de confianza del $95\,\%$ no es posible que un estado con una renta *per capita* de 8267$ gaste 440$ por alumno.

3) Se desea contrastar si los coeficientes teóricos asociados a la Renta son iguales pero de sentido contrario. Contrastar al nivel de significación del $5\,\%$ dicha hipótesis.

   Solución:

$$\left\{ \begin{array}{ll} H_0: & \beta_2 = -\beta_3 \\ H_a: & \beta_2 \neq -\beta_3 \end{array} \right. \Leftrightarrow \left\{ \begin{array}{ll} H_0: & w = \beta_2 + \beta_3 = 0 \\ H_a: & w = \beta_2 + \beta_3 \neq 0 \end{array} \right. \qquad t = \frac{\hat{w}}{\sqrt{\widehat{Var}(\hat{w})}} \overset{H_0}{\sim} t_{T-K}$$

$$\hat{w} = \hat{\beta}_2 + \hat{\beta}_3 = 0{,}129$$

$$v\hat{a}r(\hat{w}) = v\hat{a}r(\hat{\beta}_2) + v\hat{a}r(\hat{\beta}_3) + 2 \cdot c\hat{o}v(\hat{\beta}_2, \hat{\beta}_3) = \hat{\sigma}^2(1{,}789 + 0{,}957 + 2 \cdot (-1{,}245)) = 7{,}0912 \cdot 10^{-6}$$

   $\Rightarrow |t| = 48{,}44 > t_{50-3}^{0{,}025} \simeq 2{,}01 \Rightarrow$ Rechazo H_0 de que los coeficientes teóricos asociados a la renta son iguales pero de sentido contrario, para un nivel de significación del $5\,\%$.

4) Escribir el modelo restringido bajo la hipótesis nula del apartado anterior y llevar a cabo la estimación MCO de dicho modelo. ¿Cuáles son las propiedades del estimador MCO así obtenido?

$\boxed{\text{Solución:}}$ Si $Y_i = \beta_1 + \beta_2 X_i + \beta_3 X_i^2 + u_i$, el modelo restringido sujeto a $\beta_2 = -\beta_3$ es $Y_i = \beta_1 + \beta_3 X_i^* + u_i$, donde $X_i^* = X_i^2 - X_i$. El estimador restringido es:

$$\hat{\beta^R} = \begin{pmatrix} \beta_1^R \\ \beta_3^R \end{pmatrix} = (X^{*\prime} X^*)^{-1} X^{*\prime} Y = \begin{pmatrix} N & \sum X_i^* \\ & \sum X_i^{*2} \end{pmatrix}^{-1} \begin{pmatrix} \sum Y_i \\ \sum Y_i X_i^* \end{pmatrix}$$

donde

$$\begin{aligned} \sum X_i^* &= \sum (X_i^2 - X_i) = \sum X_i^2 - \sum X_i = 16{,}545 - 23{,}08 = -6{,}535 \\ \sum X_i^{*2} &= \sum (X_i^2 - X_i)^2 = \sum X_i^4 - \sum X_i^2 - 2\sum X_i^3 = 16{,}481 + 16{,}545 - 2 \cdot 15{,}298 = 2{,}43 \\ \sum Y_i X_i^* &= \sum Y_i (X_i^2 - X_i) = \sum Y_i X_i^2 - \sum Y_i X_i = 2{,}681 - 2{,}669 = 0{,}012 \end{aligned}$$

de forma que

$$\hat{\beta^R} = \begin{pmatrix} 50 & -6{,}535 \\ & 2{,}43 \end{pmatrix}^{-1} \begin{pmatrix} 3{,}532 \\ 0{,}012 \end{pmatrix} = \begin{pmatrix} 0{,}11 \\ 0{,}29 \end{pmatrix}$$

PROPIEDADES: $\hat{\beta}^R$ es lineal en u pero sesgado puesto que la restricción es falsa con un $95\,\%$ de confianza. sin embargo, $\hat{\beta}^R$ tiene menor varianza que los estimadores MCO del enunciado.

LADE-2005.2 (Feb-2005)

Un profesor desea conocer si las notas que se obtienen en un examen, (N), dependen del sexo del alumno. Para analizar esta cuestión el profesor dispone de información sobre 100 alumnos y con ella ha obtenido las siguientes estimaciones MCO:

$$\hat{N}_i = 5,36 + 0,858 M_i \qquad R^2 = 0,0472 \tag{1}$$

donde M_i toma valor 1 si el alumno i-ésimo es mujer y cero en caso contrario.

1. Interpretar los coeficientes de este modelo.

   $\boxed{\text{Solución:}}$ $\hat{\beta}_0 = \hat{E}(N_i | i \in H) = 5{,}36$ es la nota media estimada de un alumno (hombre). $\hat{\beta}_0 + \hat{\beta}_1 = \hat{E}(N_i | i \in M) = 5{,}36 + 0{,}858 = 6{,}218$ es la nota media estimada de una alumna (mujer), de forma que $\hat{\beta}_1 = 0{,}858$ es la diferencia en la nota media estimada de una alumna (mujer) con respecto a un alumno (hombre).

2. ¿Es el sexo es una variable relevante a la hora de explicar la nota obtenida?

   $\boxed{\text{Solución:}}$

$$\left.\begin{array}{ll} H_0: & \beta_1 = 0 \\ H_a: & \beta_1 \neq 0 \end{array}\right\} \Rightarrow F = \frac{R^2/(K-1)}{(1-R^2)/(T-K)} \overset{H_0}{\sim} \mathcal{F}_{K-1,T-K}$$

Dado que $R^2 = 0{,}0472$, $K - 1 = 1$ y $T - K = 100 - 2 = 98$, entonces $F = 4{,}8547 > 3{,}92 \simeq \mathcal{F}_{1,98}^{0,05} \Rightarrow$ Rechazo H_0 de que la variable sexo no es relevante para un nivel de significación del $5\,\%$.

Otro profesor piensa que otra variable relevante para explicar la nota es el haber realizado los controles voluntarios que se han hecho a lo largo del curso. Por esta razón ha estimado el siguiente modelo:

$$\underset{(t-est)}{\hat{N}_i} = 5,07 + \underset{(1,64)}{0,57}M_i + \underset{(5,42)}{2,30}C_i \qquad R^2 = 0,2688 \tag{2}$$

donde C_i toma valor 1 si el alumno i-ésimo ha realizado alguno de los controles voluntarios y cero en caso contrario.

3. Interpretar los coeficientes de este modelo.

   Solución: $\hat{\beta}_0 = 5{,}07 = \hat{E}(N_i|i \in H, i \notin C)$ es la nota media estimada de un alumno que no ha hecho el control. $\hat{\beta}_1 = 0{,}57$ es la diferencia en la nota media estimada de una alumna respecto de un alumno (si ambos han hecho o ambos han dejado de hacer el control). $\hat{\beta}_2 = 2{,}3$ es la diferencia en la nota media estimada de un alumno (hombre o mujer) que ha hecho control respecto a otro que no lo ha hecho.

4. ¿Es relevante la nueva variable añadida?

   Solución:

$$\left.\begin{array}{ll} H_0: & \beta_2 = 0 \\ H_a: & \beta_2 \neq 0 \end{array}\right\} \Rightarrow t = \frac{\hat{\beta}_2}{\sqrt{\widehat{Var}(\hat{\beta}_2)}} \overset{H_0}{\sim} t_{T-K}$$

   $|t| = 5{,}42 > t_{100-3}^{0,025} \simeq 1{,}98 \Rightarrow$ Rechazo H_0 para $\alpha = 5\,\% \Rightarrow$ Es relevante.

5. A la vista de los resultados de los apartados anteriores comenta la relevancia del sexo a la hora de explicar la nota obtenida. ¿Existe alguna contradicción entre las conclusiones de las dos especificaciones? Comentar detalladamente.

   Solución:

$$\left.\begin{array}{ll} H_0: & \beta_1 = 0 \\ H_a: & \beta_1 \neq 0 \end{array}\right\} \Rightarrow |t| = 1{,}64 < t_{97}^{0,025} \simeq 1{,}98 \Rightarrow \text{No se rechaza } H_0 \text{ al } \alpha = 5\,\%.$$

   La variable sexo no es significativa. No existe ninguna contradicción, ya que en el modelo (1) se omite una variable relevante (C_i); por tanto, el estimador MCO de los coeficientes en (1) es sesgado, al igual que $\hat{\sigma}^2 \Rightarrow$ la inferencia queda invalidada.

6. Dado el análisis anterior se propone el modelo $N_i = \beta_0 + \beta_1 C_i + u_i$, ¿cuál es la razón de especificar finalmente este modelo? Estímalo usando la siguiente información:

   De los 100 alumnos 40 son hombres y 60 mujeres. Además, 20 de ellos han realizado alguno de los controles, y por tanto, 80 no han realizado ningún control. De los 20 que han realizado los controles 5 son hombres y 15 son mujeres. Además $\sum_1^{100}(N_i - 5{,}876)^2 = 374{,}2724$. Por último, la suma de las notas obtenidas para las distintas clases de alumnos se recoge en la siguiente tabla:

	Todos los alumnos	Hombres	Mujeres	No hacen control	Hacen control
$\sum N_i$	587,6	214,45	373,15	431,5	156,1

Solución: Especificamos el modelo $N_i = \beta_0 + \beta_1 C_i + u_i$ puesto que la variable sexo no es relevante, mientras que C sí lo es.

$$(X'X) = \begin{pmatrix} N & \sum C_i \\ & \sum C_i^2 \end{pmatrix} = \begin{pmatrix} 100 & 20 \\ & 20 \end{pmatrix} \qquad X'Y = \begin{pmatrix} \sum N_i \\ \sum C_i N_i \end{pmatrix} = \begin{pmatrix} 587{,}6 \\ 156{,}1 \end{pmatrix}$$

$$\hat{\beta} = \begin{pmatrix} \hat{\beta}_0 \\ \hat{\beta}_1 \end{pmatrix} = (X'X)^{-1}X'Y = \begin{pmatrix} 5{,}3937 \\ 2{,}4112 \end{pmatrix} \qquad \hat{N}_i = 5{,}3937 + 2{,}4112 C_i$$

7. Estimar la varianza de las perturbaciones.

$$\hat{\sigma}^2 \frac{\hat{u}'\hat{u}}{T-K} = \frac{\sum n_i^2 - (\hat{\beta}'X'Y - T\overline{Y}^2)}{T-K} = \frac{374{,}2724 - [3545{,}742 - 100(\frac{587{,}6}{100})^2]}{100-2} = 2{,}87$$

8. A los profesores les interesa contrastar si el efecto de la tarea es mayor que los puntos que añade al examen y para ello, quieren contrastar que el efecto diferencial en la nota de los alumnos que hacen la tarea frente a los que no la hacen es mayor que 1. Realizar el contraste.

Solución:

$$\left.\begin{array}{ll} H_0: & \beta_1 = 1 \\ H_a: & \beta_1 > 1 \end{array}\right\} \Rightarrow t = \frac{\hat{\beta}_1 - 1}{\sqrt{\widehat{Var}(\hat{\beta}_1)}} \overset{H_0}{\sim} t_{T-K}$$

$\widehat{Var}(\hat{\beta}_1) = \hat{\sigma}^2 \cdot a_{22} = 2{,}87\frac{100}{1600} = 0{,}1793 \Rightarrow |t| = 5{,}42 > t_{100-3}^{0{,}025} \simeq 1{,}98 \Rightarrow$ Rechazo H_0 para $\alpha = 5\,\% \Rightarrow$ El efecto diferencial es mayor que 1.

(Nota: Este ejercicio se ha preparado con datos reales de la asignatura Elementos de Probabilidad y Estadística).

LADE-2005.3 (Feb-2005)

El gerente de una empresa de un determinado país encarga a varios de sus empleados analizar el comportamiento de sus ventas anuales en el período 1995-2004. Se presentan las siguientes propuestas:

$$\hat{Y}_t = \underset{(20{,}93)}{38{,}3} - \underset{(-6{,}33)}{1{,}55}\, P_t - \underset{(-6{,}03)}{7{,}28}\, D_t \qquad R^2 = 0{,}9985 \tag{3}$$

$$\hat{Y}_t = \underset{(0{,}30)}{5{,}01} - \underset{(-1{,}23)}{0{,}92}\, P_t + \underset{(1{,}83)}{0{,}29}\, R_t \qquad R^2 = 0{,}9921 \tag{4}$$

donde Y_t son las ventas de la empresa en el año t, P_t el precio de venta del producto, R_t la renta anual media de la población y D_t una variable que toma el valor 1 para periodos con depresión económica y 0 para los de expansión económica del país.

i) Utilizar sólo el modelo (3) para decidir si el ciclo económico es una variable relevante para explicar las ventas.

Solución: Si β_2 es el coeficiente asociado a D_t, decidimos a partir del contraste

$$\left.\begin{array}{ll} H_0: & \beta_2 = 0 \\ H_a: & \beta_2 \neq 0 \end{array}\right\} \Rightarrow t = \frac{\hat{\beta}_2}{\sqrt{\widehat{Var}(\hat{\beta}_2)}} \overset{H_0}{\sim} t_{T-K}$$

$|t| = |-6{,}03| > t_{10-3}^{0{,}025} = 2{,}36 \Rightarrow$ Rechazo H_0 para $\alpha = 5\,\%$ $\Rightarrow$ El ciclo económico es una variable relevante.

ii) Analizar el modelo (4), sin tener en cuenta el modelo anterior (3) . ¿Existe algún problema muestral en dicho modelo? Razonar la respuesta. ¿Qué propiedades presentan estos estimadores?

Solución: Los contrastes individuales de la t de Student para los coeficientes de las variables P_t y R_t del modelo (4) producen $|t_1| = 1{,}23$ y $|t_2| = 1{,}83$, ambos menores que $t_{10-3}^{0{,}025} = 3{,}365 \Rightarrow$ ninguna de las dos variables es significativa individualmente. El contraste de significación conjunta queda:

$$F = \frac{R^2/(K-1)}{(1-R^2)/(T-K)} = \frac{0{,}9921/2}{(1-0{,}9921)/7} = 439{,}53 > \mathcal{F}_{2,10}^{0{,}05} = 4{,}10$$

Que indica que ambas variables son significativas conjuntamente. Por tanto, podríamos tener un problema de multicolinealidad aproximada. En esta caso, la varianza de los estimadores es alta $\Rightarrow$ los estadísticos t son pequeños $\Rightarrow$ mayor probabilidad de no rechazar hipótesis nulas falsas en contrastes individuales. Sin embargo, los estimadores son lineales, insesgados y de mínima varianza.

Además, se presenta una tercera propuesta:

$$\hat{Y}_t = \underset{(5{,}63)}{24{,}20} - \underset{(-6{,}8)}{1{,}16}\,P_t + \underset{(3{,}36)}{0{,}13}\,R_t - \underset{(-8{,}71)}{6{,}20}\,D_t \qquad R^2 = 0{,}9997 \tag{5}$$

Teniendo en cuenta ahora la información de las tres propuestas presentadas:

iii) Explicar razonadamente si cambiaría las respuestas de los apartados i) y ii).

Solución: En el modelo (5) las tres variables son individualmente significativas (comparar $8{,}71$, $6{,}8$ y $3{,}36$ con $t_6^{0{,}025} = 2{,}447$), con lo que es prácticamente seguro que son todas relevantes. Por tanto, en los modelos (3) y (4) hay omisión de una variable relevante, respectivamente R_t y D_t. Por tanto, los estimadores de los coeficientes y $\hat{\sigma}^2$ son sesgados en (3) y (4). Como la inferencia entonces no es válida, las conclusiones de los apartados anteriores, basadas en esos modelos, no son válidas.

iv) ¿Cuál de los tres modelos es el apropiado?, ¿cuáles son las propiedades de los estimadores del modelo elegido?

Solución: El modelo (5) es el apropiado, puesto que incluye todas las variables relevantes. Sus estimadores son lineales, insesgados y de mínima varianza.

LADE-2005.4 (Jun-2005)

Una empresa quiere analizar sus ventas de agua mineral, Y, en una determinada región en función de la temperatura media, X_2, (medida en grados centígrados) y el precio, X_3, (medido en euros por litro). Para ello especifica la siguiente función de regresión lineal:

$$Y_t = \beta_1 + \beta_2 X_{2t} + \beta_3 X_{3t} + u_t \qquad u_t \overset{iid}{\sim} N(0, \sigma^2) \tag{1}$$

A continuación se resume la información obtenida a partir de una muestra de las últimas 100 observaciones trimestrales (primavera, verano, otoño, invierno):

$$(X'X) = \begin{bmatrix} 100 & 198,20 & 497,52 \\ 198,20 & 432,05 & 986,47 \\ 497,52 & 986,47 & 2497,50 \end{bmatrix} \qquad (X'X)^{-1} = \begin{bmatrix} 1,2147 & -0,0483 & -0,2229 \\ -0,0483 & 0,0255 & -0,0004 \\ -0,2229 & -0,0004 & 0,0450 \end{bmatrix}$$

$$\sum Y_t = 713 \quad \sum Y_t X_{2t} = 1913 \quad \sum Y_t X_{3t} = 3400 \quad \sum Y_t^2 = 15041,56$$

a) Deriva las ecuaciones normales del criterio de estimación mínimo cuadrático.

Solución: Modelo: $Y = X\beta + u$. Residuos: $\hat{u} = Y - X\hat{\beta} \Rightarrow$ criterio MCO: $\text{mín}_\beta\, \hat{u}'\hat{u} = (Y - X\hat{\beta})'(Y - X\hat{\beta})$. $\Rightarrow \frac{\partial \hat{u}'\hat{u}}{\partial \hat{\beta}} = X'X\hat{\beta} - X'Y = 0 \Rightarrow X'X\hat{\beta} = X'Y$

b) Interpreta los coeficientes β_2 y β_3 del modelo.

Solución:

β_2 = Incremento esperado en las ventas de agua ante un aumento en 1°C de la temperatura.

β_3 = Incremento esperado en las ventas de agua ante un aumento de 1 € en el precio del litro de agua.

c) Estima el modelo por el método de mínimos cuadrados ordinarios. ¿Te parecen los signos razonables?, ¿por qué?

Solución:

$$\hat{\beta} = (X'X)^{-1}X'Y = \begin{bmatrix} 1,2147 & -0,0483 & -0,2229 \\ -0,0483 & 0,0255 & -0,0004 \\ -0,2229 & -0,0004 & 0,0450 \end{bmatrix} \begin{bmatrix} 713 \\ 1913 \\ 3400 \end{bmatrix} = \begin{bmatrix} 15,8232 \\ 12,9836 \\ -6,6929 \end{bmatrix}$$

Los signos parecen adecuados dado que parece razonable pensar que, a mayor temperatura, mayores serán las ventas $\beta_2 > 0$ y que las ventas serán menores cuanto mayores sean los precios $\beta_3 < 0$.

d) Estima la matriz de varianzas y covarianzas del estimador de los coeficientes del modelo.

Solución: $\hat{u}'\hat{u} = Y'Y - \hat{\beta}'X'Y = 15041,56 - (15,8232 \cdot 713 + 12,9836 \cdot 1913 - 6,6929 \cdot 3400) = 1677,8516 \Rightarrow \hat{\sigma}^2 = \frac{\hat{u}'\hat{u}}{T-K} = \frac{1677,8516}{97} = 17,2974 \Rightarrow$

$$\widehat{Var}(\hat{\beta}) = 17,2974 \begin{bmatrix} 1,2147 & -0,0483 & -0,2229 \\ -0,0483 & 0,0255 & -0,0004 \\ -0,2229 & -0,0004 & 0,0450 \end{bmatrix} = \begin{bmatrix} 21,0112 & & \\ -0,8355 & 0,4411 & \\ -3,8555 & -0,0069 & 0,7784 \end{bmatrix}$$

e) Dados los resultados obtenidos en las estimaciones de los coeficientes, se piensa que el efecto de la temperatura media sobre las ventas pudiera ser el doble que el efecto del precio, pero de signo contrario. Contrástalo.

Solución:

$$\begin{cases} H_0: & \beta_2 = -2\beta_3 \\ H_a: & \beta_2 \neq -2\beta_3 \end{cases} \Rightarrow w = \beta_2 + 2\beta_3$$

$\Rightarrow \hat{w} = \hat{\beta}_2 + 2\hat{\beta}_3 = -0{,}4022.$ $\widehat{Var}(\hat{w}) = \widehat{Var}(\hat{\beta}_2) + 4\widehat{Var}(\hat{\beta}_3) + 4\widehat{Cov}(\hat{\beta}_2, \hat{\beta}_3) = 3{,}5876$ Entonces, bajo H_0, $t = \frac{\hat{w}}{\hat{\sigma}_{\hat{w}}} \sim t_{T-K}$. Dado que $|t| = \left|\frac{-0{,}4022}{\sqrt{3{,}5876}}\right| = 0{,}2123 < 1{,}98 \approx t_{97}^{0{,}05/2} \Rightarrow$ no se rechaza la H_0 a un nivel de significación del 5 %.

f) Estima el modelo restringido correspondiente a la hipótesis nula anterior.

Solución: MR: $Y_t = \beta_1 + (-2\beta_3)X_{2t} + \beta_3 X_{3t} + u_t \Rightarrow Y_t = \beta_1 + \beta_3(X_{3t} - 2X_{2t}) + u_t \Rightarrow$ $Y_t = \beta_1 + \beta_3 X_{3t}^* + u_t$. Entonces

$$\hat{\beta}^* = (X^{*\prime}X^*)^{-1}X^{*\prime}Y = \begin{bmatrix} T & \sum X_{3t}^* \\ \sum X_{3t}^* & \sum (X_{3t}^*)^2 \end{bmatrix} \begin{bmatrix} \sum Y_t \\ \sum X_{3t}^* Y_t \end{bmatrix}$$

donde $\sum X_{3t}^* = \sum X_{3t} - 2\sum X_{2t} = 497{,}52 - 2 \cdot 198{,}2 = 101{,}12$
$\sum (X_{3t}^*)^2 = \sum X_{3t}^2 + 4\sum X_{2t}^2 - 4\sum X_{2t}X_{3t} = 2497{,}5 + 4(432{,}05) - 4(986{,}47) = 279{,}82$
$\sum X_{3t}^* Y_t = \sum X_{3t}Y_t - 2\sum X_{2t}Y_t = 3400 - 2(1930) = -426$ y entonces

$$\begin{aligned} \hat{\beta}^* &= \begin{bmatrix} 100 & 101{,}12 \\ 101{,}12 & 279{,}82 \end{bmatrix}^{-1} \begin{bmatrix} 713 \\ -426 \end{bmatrix} = \frac{1}{17756{,}74} \begin{bmatrix} 279{,}82 & -101{,}12 \\ -101{,}12 & 100 \end{bmatrix}^{-1} \begin{bmatrix} 713 \\ -426 \end{bmatrix} = \\ &= \begin{bmatrix} 13{,}6617 \\ -6{,}4593 \end{bmatrix} \end{aligned}$$

g) Teniendo en cuenta los resultados del contraste del apartado e), ¿qué propiedades tiene el estimador empleado en el apartado c)?

Solución: En (e) no he rechazado la hipótesis nula $H_0 : \beta_2 = -2\beta_3$ con lo que hay indicios de que la restricción impuesta en (f) puede ser verdadera. Si esto es así, los estimadores del apartado (c) tendrán mayor varianza que los de (f). En ambos casos, los estimadores son lineales e insesgados.

A continuación se piensa que el precio medio de los refrescos, X_4, puede afectar a las ventas, por lo que se reestima el modelo (1) incluyendo esta variable:

$$\underset{}{\hat{Y}_t} = \underset{(1{,}1901)}{4{,}0518} + \underset{(28{,}6606)}{13{,}8581}X_{2t} - \underset{(-11{,}8459)}{7{,}5349}X_{3t} + \underset{(0{,}0364)}{0{,}2922}X_{4t} \tag{2}$$

h) ¿Aconsejarías emplear el modelo (2) para calcular las ventas correspondientes al siguiente trimestre si se espera una temperatura media de 32 grados, que el precio medio por litro de agua mineral sea 0,4 € y el de los refrescos 1,2 €? Razona tu respuesta. **Explica cómo** obtendrías otros posibles valores para las ventas del siguiente trimestre con varianza más pequeña.

Solución: El modelo (2) incluye una variable irrelevante (X_4) dado que ésta no es individualmente significativa $(|t| = 0{,}0364)$. Este error de especificación implica que tenemos estimadores con varianza mayor que los del modelo correctamente especificado, con lo que, siendo el resto de propiedades las mismas, hacen que sean preferibles los del modelo inicial. La menor precisión consiguiente de las estimaciones hace que la predicción (por punto y por intervalo) sea también menos precisa y que sea, por tanto, preferible realizarla sobre el modelo (1) o, mejor aún, sobre el estimado en el apartado (f). P. ej., la predicción por punto sería $\hat{Y}_P = 13{,}6617 - 6{,}4593(X^*_{3P}) = 13{,}6617 - 6{,}4593(0{,}4 - 2 \cdot 32)$.

LADE-2005.5 (Jun-2005)

Con objeto de analizar el salario de un grupo de economistas, un estudiante ha recogido información de 17 individuos sobre su salario (Y), experiencia (X) y género. Ha estimado varias especificaciones alternativas que se muestran a continuación:

$$\hat{Y}_i = \underset{(14{,}06)}{1{,}7932} + \underset{(9{,}60)}{0{,}0709}\, X_i \tag{3}$$

$$R^2 = 0{,}8602 \qquad \textstyle\sum_i \hat{u}_i^2 = 1{,}2086$$

$$\hat{Y}_i = \underset{(10{,}84)}{2{,}6733}\, H_i + \underset{(11{,}45)}{2{,}9950}\, M_i \tag{4}$$

$$R^2 = 0{,}0506 \qquad \textstyle\sum_i \hat{u}_i^2 = 8{,}2100$$

$$\hat{Y}_i = \underset{(12{,}17)}{1{,}7618}\, H_i + \underset{(11{,}52)}{1{,}8556}\, M_i + \underset{(9{,}16)}{0{,}0701}\, X_i \tag{5}$$

$$R^2 = 0{,}8644 \qquad \textstyle\sum_i \hat{u}_i^2 = 1{,}1725$$

donde H_i es una variable ficticia que toma el valor 1 si el individuo es hombre y 0 en caso contrario, M_i es una variable ficticia que toma el valor 1 si el individuo es mujer y 0 en caso contrario. **Entre paréntesis aparecen los estadísticos t muestrales**.

a) Decide qué modelo es el adecuado para explicar los salarios. Contesta la pregunta realizando los contrastes de significación individual de la variable género y de la variable experiencia.

   Solución: La variable experiencia (X) es significativa puesto que lo es $(|t| = 9{,}16)$ en el modelo más parametrizado (5). Con respecto a la variable género y, nuevamente, en el modelo más parametrizado,

   $$\begin{cases} H_0: & \beta_H = \beta_M \\ H_a: & \beta_H \neq \beta_M \end{cases}$$

   equivale a contrastar el modelo (5) frente al modelo (3), lo que puede hacerse mediante el estadístico basado en sumas de residuos al cuadrado de MR y MNR, donde $\hat{u}_r'\hat{u}_r = 1{,}2086$ y $\hat{u}'\hat{u} = 1{,}1725$. Así, $F = 0{,}4336 < \mathcal{F}^{0{,}05}_{1,14}$, con lo que no se rechaza la hipótesis nula al nivel del 5 % y el modelo adecuado es el (3).

b) ¿Qué problema se habría producido en los contrastes anteriores si la muestra hubiera estado formada por 17 hombres?

Solución: No dispondríamos de información muestral sobre la variable género de forma que sólo tendría sentido plantear el modelo (3).

c) Otro estudiante le sugiere que no ha tenido en cuenta la siguiente estimación:

$$\hat{Y}_i = \underset{(11,52)}{1{,}8556} \underset{(-0,65)}{-0{,}0937} H_i + \underset{(9,16)}{0{,}0701} X_i \tag{6}$$

$$R^2 = 0{,}8644 \qquad \sum_i \hat{u}_i^2 = 1{,}1725$$

i) Explica detalladamente a este segundo estudiante que el primero **sí ha tenido en cuenta** dicha estimación.

Solución: Los modelos (5) y (6) son equivalentes, representan dos especificaciones alternativas de una misma relación en la que aparece un factor cualitativo. En (5), si denotamos β_1 y β_2 a los parámetros que acompañan a H_i y M_i respectivamente, y en (6) denotamos α_2 al coeficiente asociado a H_i y α_1 al término independiente, entonces $\beta_1 = \alpha_1 + \alpha_2$ y $\beta_2 = \alpha_1$.

ii) ¿Cómo se realizaría el contraste de significación de la variable género en el modelo (6)? Haz el contraste y comprueba que obtienes el mismo resultado que el contraste realizado en el modelo (5).

Solución: Para contrastar $H_0 : \alpha_3 = 0$ frente a $H_a : \alpha_3 \neq 0$, tenemos $|t| = |-0{,}65| = 0{,}65 < t_{14}^{0{,}025} \Rightarrow$ no rechazamos la H_0 a un nivel de significación del $5\,\%$, igual que en el apartado (a). Este contraste puede realizarse alternativamente mediante el estadístico F a partir de sumas de residuos al cuadrado de MR y MNR.

LADE-2005.6 (Jun-2005)

El gerente de una estación invernal de esquí del Pirineo, con datos anuales desde 1991 hasta 2003 de las variables:

$$\begin{aligned}
E_t &= \text{Número de esquiadores por año, en miles, en el año } t \\
P_t &= \text{Precio del forfait en € constantes de 1991} \\
F_t &= \text{Precio medio del forfait en estaciones del Pirineo francés en € constantes de 1991} \\
PIB_t &= \text{Tasa de crecimiento del PIB en el año } t \text{ respecto al año anterior} \\
S_t &= \text{\% de días soleados en la temporada de esquí del año } t
\end{aligned}$$

ha obtenido las siguientes estimaciones MCO:

$$\widehat{E}_t = \underset{(106,47)}{1029{,}57} + \underset{(2,22)}{0{,}91}\, P_t + \underset{(1,50)}{2{,}41}\, PIB_t - \underset{(-3,55)}{0{,}64}\, S_t \tag{7}$$

$$R^2 = 0{,}680$$

$$\underset{(99{,}05)}{\widehat{E}_t = 1026{,}22} + \underset{(2{,}36)}{0{,}97\, F_t} + \underset{(1{,}47)}{2{,}30\, PIB_t} - \underset{(-3{,}55)}{0{,}64\; S_t} \tag{8}$$

$$R^2 = 0{,}695$$

$$\underset{(68{,}44)}{\widehat{E}_t = 1018{,}38} - \underset{(-0{,}75)}{2{,}94\; P_t} + \underset{(0{,}99)}{3{,}99\, F_t} + \underset{(1{,}18)}{1{,}96\, PIB_t} - \underset{(-3{,}44)}{0{,}62\; S_t} \tag{9}$$

$$R^2 = 0{,}715$$

a) Dados los tres modelos estimados, se deduce que (9) tiene un problema muestral, ¿cuál es dicho problema?, ¿cuáles son las propiedades de los estimadores de dicho modelo? ¿Se concluye que son válidos los contrastes que se realicen en el mismo?

Solución: Este modelo tiene un R^2 alto mientras que todas las variables son no significativas, excepto S_t; al mismo tiempo, los coeficientes asociados a las variables P_t y F_t oscilan mucho de una estimación a otra, siendo el resto relativamente estables. Esto sugiere un posible problema de multicolinealidad aproximada, probablemente por una alta correlación entre esas dos variables. Los estimadores seguirán siendo lineales, insesgados y de mínima varianza, pero tendremos un valor de $|X'X|$ relativamente pequeño, de forma que las varianzas de los estimadores serán altas y los estadísticos t para los contrastes individuales poco significativos, de manera que no son muy fiables.

b) Si ahora además se tiene en cuenta el siguiente modelo estimado:

$$\underset{(86{,}32)}{\widehat{E}_t = 984{,}10} + \underset{(0{,}43)}{1{,}01\, P_t} - \underset{(-0{,}22)}{0{,}55\; F_t} + \underset{(2{,}39)}{2{,}20\, PIB_t} + \underset{(3{,}26)}{1{,}96\, S_t} - \underset{(-4{,}36)}{0{,}03\; S_t^2} \tag{10}$$

$$R^2 = 0{,}9233$$

¿Qué propiedades tienen los estimadores aquí empleados? ¿Qué modelo de entre los presentados te parece el adecuado para explicar el número de esquiadores? Razona detalladamente.

Solución: La variable S_t^2 es claramente significativa $(|t| = 4{,}36)$, incluso a pesar de la posible existencia de multicolinealidad aproximada. Esto significa que los modelos anteriores incurren en la omisión de una variable relevante y sus estimadores MCO son sesgados. Por lo tanto, este modelo es, si bien no sabemos con seguridad si el modelo verdadero, al menos más completo que los anteriores y a falta de otra información, su estimador MCO es lineal, insesgado y de mínima varianza. Por todo ello, este modelo (10) resulta el más adecuado de todos los presentados.

LADE-2006.1 (Feb-2006)

Un estudiante de empresariales desea analizar el consumo de gasolina de su automóvil. Para ello, propone el siguiente modelo:

$$G_t = \alpha + \beta_1\, A_t + \beta_2\, C_t + u_t \qquad t = 1, \ldots, T \qquad u_t \sim N(0, \sigma^2) \tag{1}$$

donde:

G_t es el número de litros consumidos en la semana t.

A_t son los kilómetros realizados por **autopista** en la semana t, en cientos de kilómetros.

C_t son los kilómetros realizados por **ciudad** en la semana t, en cientos de kilómetros.

1. Escribe en qué consiste el criterio mínimo cuadrático ordinario (MCO) y desarróllalo para este modelo obteniendo las ecuaciones normales.

   Solución:

   El criterio de MCO consiste en

$$\min_{\hat{\beta}} \sum_t \hat{u}_t^2 = \min_{\{\hat{\alpha},\hat{\beta}_1,\hat{\beta}_2\}} \sum_t (G_t - \hat{\alpha} - \hat{\beta}_1 A_t - \hat{\beta}_2 C_t)^2 = L(\hat{\alpha}, \hat{\beta}_1, \hat{\beta}_2).$$

   Las condiciones de primer orden son:

$$\frac{\partial L}{\partial \hat{\alpha}} = 0 = -2\sum_t (G_t - \hat{\alpha} - \hat{\beta}_1 A_t - \hat{\beta}_2 C_t) = 0 \Rightarrow \quad \sum G_t = T\hat{\alpha} + \hat{\beta}_1 \sum A_t + \hat{\beta}_2 \sum C_t$$

$$\frac{\partial L}{\partial \hat{\beta}_1} = 0 = -2\sum_t (G_t - \hat{\alpha} - \hat{\beta}_1 A_t - \hat{\beta}_2 C_t) A_t = 0 \Rightarrow \quad \sum G_t A_t = \hat{\alpha} \sum A_t + \hat{\beta}_1 \sum A_t^2 + \hat{\beta}_2 \sum C_t A_t$$

$$\frac{\partial L}{\partial \hat{\beta}_2} = 0 = -2\sum_t (G_t - \hat{\alpha} - \hat{\beta}_1 A_t - \hat{\beta}_2 C_t) C_t = 0 \Rightarrow \quad \sum G_t C_t = \hat{\alpha} \sum C_t + \hat{\beta}_1 \sum A_t C_t + \hat{\beta}_2 \sum C_t^2.$$

2. Una de las propiedades de la Función de Regresión Muestral es que la media aritmética de los residuos de la regresión es cero. ¿Se cumple en este caso? Razona el resultado utilizando las ecuaciones normales del apartado anterior.

   Solución:

   Sí se cumple en este caso porque existe término independiente en el modelo. A partir de la primera ecuación normal,

$$\sum_t (G_t - \hat{\alpha} - \hat{\beta}_1 A_t - \hat{\beta}_2 C_t) = 0$$

$$\sum (G_t - \hat{G}_t) = 0 \quad \Rightarrow \quad \sum \hat{u}_t = 0 \quad \Rightarrow \quad \bar{\hat{u}} = 0$$

Se dispone de las primeras siete observaciones semanales a partir de la compra del coche. La información recogida se resume a continuación:

t	G_t	A_t	C_t
1	49	3	3
2	40	5,4	0
3	43	4,4	1
4	38	5,4	0
5	42	4,4	1
6	46	3,6	2
7	43	6	0

$\sum A_t = 32{,}2$ $\quad \sum C_t = 7$ $\quad \sum G_t = 301$

$\sum A_t^2 = 155$ $\quad \sum C_t^2 = 15$ $\quad \sum G_t^2 = 13023$

$\sum A_t C_t = 25$ $\quad \sum G_t A_t = 1365{,}8$ $\quad \sum G_t C_t = 324$

$$(X'X)^{-1} = \begin{pmatrix} 75{,}89286 & -13{,}75 & -12{,}50 \\ -13{,}75000 & 2{,}50 & 2{,}25 \\ -12{,}50000 & 2{,}25 & 2{,}15 \end{pmatrix}$$

3. Obtén los siguientes valores:

   - ¿Cuántos litros de gasolina ha consumido el coche en la tercera semana de su compra?

     Solución:

     $t = 3 \quad \rightarrow \quad G_3 = 43$ litros.

   - ¿Cuántos kilómetros ha conducido por la ciudad desde la compra del coche?

     Solución:

     $C_1 + C_2 + \ldots + C_7 = 7 \quad \rightarrow$ 700 Km.

   - ¿Cuál es la **media muestral** de kilómetros conducidos por autopista?

     Solución:

$$\bar{A} = \frac{\sum A_t}{T} = \frac{32{,}2}{7} = 4{,}6 \quad \rightarrow \quad \text{460 Km.}$$

Los resultados de la estimación mínimo cuadrática del modelo (1) son:

$$\hat{G}_t = 14 + 4,75A_t + 7,15C_t \qquad t = 1, \ldots, 7 \qquad R^2 = 0{,}93937$$

4. ¿Cuál es el residuo MCO cometido al estimar los litros de gasolina consumidos en la última semana?

   Solución:

   $t = 7 \quad \rightarrow \qquad \hat{u}_7 = G_7 - \hat{G}_7 = 43 - 14 - 4{,}75A_7 - 7{,}15C_7 = 43 - 42{,}5 = 0{,}5$

5. ¿Cuál es el consumo medio estimado de gasolina por cada cien kilómetros recorridos cuando **sólo** se circula por ciudad?, ¿y si **sólo** se circula por autopista?

   Solución:

   Sólo por autopista: $(\hat{G}_t | A_t = 1\,; C_t = 0) = 14 + 4{,}75 \cdot 1 = 18{,}75$ litros.

   Sólo por ciudad: $(\hat{G}_t | A_t = 10\,; C_t = 1) = 14 + 7{,}15 \cdot 1 = 21{,}15$ litros.

6. Estima de forma insesgada la matriz de varianzas y covarianzas de los estimadores de MCO.

   Solución:

$$X'Y = \begin{pmatrix} \sum G_t \\ \sum G_t A_t \\ \sum G_t C_t \end{pmatrix} = \begin{pmatrix} 301 \\ 1365{,}8 \\ 324 \end{pmatrix}$$

$$\hat{\sigma}^2 = \frac{SCR}{T - K} = \frac{Y'Y - \hat{\beta}X'Y}{7 - 3} = \frac{13023 - 13018}{4} = \frac{4{,}85}{4} = 1,2125$$

$$\widehat{Var}(\hat{\beta}) = \hat{\sigma}^2 (X'X)^{-1} = 1{,}2125 \begin{pmatrix} 75{,}89286 & -13{,}75 & -12{,}50 \\ -13{,}75000 & 2{,}50 & 2{,}25 \\ -12{,}50000 & 2{,}25 & 2{,}15 \end{pmatrix}$$

7. Supón que al recorrer 100 kilómetros adicionales con el coche, el consumo medio de gasolina aumenta el doble si se conduce por ciudad en lugar de conducir por autopista. **Estima** el modelo (1) cuando se tiene en cuenta esta información y obtén la suma de cuadrados residual.

Solución:

Sustituyendo $\beta_2 = 2\beta_1$ en la ecuación (1) obtenemos el modelo restringido:

$$\begin{aligned} G_t &= \alpha + \beta_1 \, A_t + 2\beta_1 \, C_t + u_t^R \\ G_t &= \alpha + \beta_1 \, (A_t + 2C_t) + u_t^R \\ G_t &= \alpha + \beta_1 X_t + u_t^R \quad \text{donde } X_t = A_t + 2C_t \end{aligned}$$

$$\begin{aligned} \begin{pmatrix} \hat{\alpha} \\ \hat{\beta}_1 \end{pmatrix}_R &= (X_*' X_*)^{-1} X_*' Y = \begin{pmatrix} T & \sum X_t \\ & \sum X_t^2 \end{pmatrix}^{-1} \begin{pmatrix} \sum G_t \\ \sum G_t X_t \end{pmatrix} = \begin{pmatrix} 7 & 46{,}2 \\ & 315 \end{pmatrix}^{-1} \begin{pmatrix} 301 \\ 2013{,}8 \end{pmatrix} \\ &= \begin{pmatrix} 25{,}19 \\ 2{,}698 \end{pmatrix} \quad \text{donde se ha usado que} \end{aligned}$$

$$\begin{aligned} \sum X_t &= \sum A_t + 2 \sum C_t = 46{,}2 \\ \sum X_t^2 &= \sum A_t^2 + 4 \sum C_t^2 + 4 \sum A_t C_t = 315 \\ \sum G_t X_t &= \sum G_t A_t + 2 \sum G_t C_t = 2013{,}8 \end{aligned}$$

$$\hat{u}_R' \hat{u}_R = \sum G_t^2 - \begin{pmatrix} \hat{\alpha} & \hat{\beta}_1 \end{pmatrix}_R \begin{pmatrix} \sum G_t \\ \sum G_t X_t \end{pmatrix} = 13023 - 13015{,}42 = 7{,}5776$$

8. Realiza el contraste oportuno para determinar si la información anterior es cierta.

Solución:

$$\begin{cases} H_o : & \beta_2 = 2\beta_1 \\ H_a : & \beta_2 \neq 2\beta_1 \end{cases} \qquad F = \frac{(\hat{u}_R' \hat{u}_R - \hat{u}'\hat{u})/q}{\hat{u}'\hat{u}/_{(T-K)}} \overset{H_o}{\sim} \mathcal{F}_{(q,T-K)}$$

$$F = \frac{7{,}5776 - 4{,}85}{1{,}2125} = 0{,}90528 < \mathcal{F}_{(1,4)}^{0{,}05} = 7{,}71 \Rightarrow$$

No se rechaza H_o para un nivel de significación del $5\,\%$.

9. Teniendo en cuenta el resultado del contraste anterior, ¿qué propiedades tienen los estimadores empleados en la estimación del modelo del apartado 7?

Solución:

Como no hemos rechazado la hipótesis nula en el apartado anterior, el estimador MCR de β es lineal en las perturbaciones, insesgado, y con menor varianza que el estimador de MCO ($Var(\hat{\beta}_i^R) < Var(\hat{\beta}_i^{MCO})$).

LADE-2006.2 (Feb-2006)

Con objeto de analizar el salario de un grupo de trabajadores de una empresa, se ha recogido información de 60 individuos sobre salario en miles de euros (Y), años de experiencia (X) y género (S) y se han estimado las siguientes especificaciones:

$$\underset{\text{(t-estad.)}}{\hat{Y}_i} = \underset{(12{,}17)}{1{,}7618}\, S1_i + \underset{(11{,}52)}{1{,}8556}\, S2_i + \underset{(9{,}16)}{0{,}0701}\, X_i \tag{2}$$

$$R^2 = 0{,}8644 \qquad \sum_i \hat{u}_i^2 = 1{,}1725$$

$$\underset{\text{(t-estad.)}}{\hat{Y}_i} = \underset{(11{,}52)}{1{,}8556} \underset{(-0{,}65)}{-0{,}0937}\, S1_i + \underset{(9{,}16)}{0{,}0701}\, X_i \tag{3}$$

$$R^2 = 0{,}8644 \qquad \sum_i \hat{u}_i^2 = 1{,}1725$$

donde $S1_i$ es una variable ficticia que toma el valor 1 si el individuo es hombre y cero en caso contrario, en cambio, la variable $S2_i$ toma el valor 1 si el individuo es mujer y cero en otro caso.

1. ¿Por qué el coeficiente de determinación y la suma de cuadrados residual coinciden en ambas especificaciones? Razona tu respuesta.

   Solución:

   Porque al fin y al cabo se trata de modelos equivalentes, para recoger el salario de los trabajadores en función del sexo y de la experiencia laboral. Denotando:

$$\left.\begin{array}{ll} (2) & Y_i = \beta_1 S1_i + \beta_2 S2_i + \beta_3 X_i + u_i \\ (3) & Y_i = \alpha_1 + \alpha_2 S1_i + \alpha_3 X_i + u_i^* \end{array}\right\} \quad \begin{array}{l} \beta_3 = \alpha_3, \quad \beta_2 = \alpha_1 \\ \beta_1 = \alpha_1 + \alpha_2 \quad \rightarrow \quad \alpha_2 = \beta_1 - \beta_2 \end{array}$$

2. Contrasta la significación **individual** de la variable género **y** de la variable experiencia.

   Solución:

   En el modelo (3) contrastamos:

$$\begin{array}{l} H_0 : \alpha_2 = 0 \\ H_a : \alpha_2 \neq 0 \end{array} \quad t = \frac{\hat{\alpha}_2}{\hat{\sigma}(\hat{\alpha_2})} \overset{H_0}{\sim} t_{T-K} \quad \rightarrow \quad |t| = 0{,}65 < t_{0{,}025}(57) = 2$$

   Al nivel de significación del 5 % no se rechaza la H_0, de modo que el género no es una variable relevante.

$$\begin{array}{l} H_0 : \alpha_3 = 0 \\ H_a : \alpha_3 \neq 0 \end{array} \quad t = \frac{\hat{\alpha}_3}{\hat{\sigma}(\hat{\alpha_3})} \overset{H_0}{\sim} t_{T-K} \quad \rightarrow \quad |t| = 9{,}16 > t_{0{,}025}(57) = 2$$

   Al nivel de significación del 5 % se rechaza la H_0, de modo que la variable experiencia sí es una variable relevante.

Posteriormente, y a la vista de los resultados obtenidos, se consideran las siguientes estimaciones:

$$\underset{\text{(t-estad.)}}{\hat{Y}_i} = \underset{(14{,}06)}{1{,}7932} + \underset{(9{,}60)}{0{,}0709}\, X_i \tag{4}$$

$$R^2 = 0{,}8602 \qquad \sum_i \hat{u}_i^2 = 1{,}2086$$

$$\underset{\text{(t-estad.)}}{\hat{Y}_i} = \underset{(10{,}39)}{2{,}2499} \underset{(-2{,}43)}{-0{,}4882}\, ES_i + \underset{(5{,}01)}{0{,}05141}\, X_i \tag{5}$$

$$R^2 = 0{,}9018 \qquad \sum_i \hat{u}_i^2 = 0{,}8487$$

$$\underset{\text{(t-estad.)}}{\hat{Y}_i} = \underset{(9,82)}{2{,}2808} + \underset{(4,85)}{0{,}05129}\, X_i \underset{(-0,49)}{-0{,}06227}\, S1_i \underset{(-2,30)}{-0{,}4770}\, ES_i \qquad (6)$$

$$R^2 = 0{,}9037 \qquad \sum_i \hat{u}_i^2 = 0{,}8329$$

donde ES_i es una variable ficticia que toma el valor 1 si el individuo tiene estudios primarios o secundarios y 0 si tiene estudios universitarios.

3. Basándote en los resultados del modelo (6), ¿cuál es el salario medio estimado de un hombre con estudios secundarios y sin experiencia?

   Solución:
   $\widehat{E(Y_i|\text{hombre}}, ES_i = 1, X_i = 0) = 2{,}2808 - 0{,}06227 - 0{,}4770 = 1{,}74153$ miles de euros.

   El salario medio estimado de un hombre con estudios secundarios y sin experiencia es 1741.53 euros.

4. Contrasta la significación conjunta de las variables estudios y género.

   Solución:
   Denotamos al MNR $(6) \quad Y_i = \gamma_1 + \gamma_2 X_i + \gamma_3 S1_i + \gamma_4 ES_i + u_i \quad \rightarrow \quad \hat{u}'\hat{u} = 0{,}8329$

$$\left.\begin{array}{ll} H_o : \gamma_3 & = \gamma_4 = 0 \\ H_a : \gamma_3 & \neq 0 \text{ y / o } \gamma_4 \neq 0 \end{array}\right\} \qquad F = \frac{(\hat{u}'_R \hat{u}_R - \hat{u}'\hat{u})/q}{\hat{u}'\hat{u}/_{(T-K)}} \overset{H_o}{\sim} \mathcal{F}_{(q,T-K)}$$

   El MR es (4) $\quad Y_i = \lambda_1 + \lambda_2 X_i + u_i^R \quad \rightarrow \quad \hat{u}'_R \hat{u}_R = 1{,}2086$

$$F = \frac{(1{,}2086 - 0{,}8329)/2}{0{,}8329/(60-4)} = 12{,}63 > \mathcal{F}^{0,05}_{(2,56)} = 3{,}15 \Rightarrow$$

   Se rechaza H_o para un nivel de significación del $5\,\%$ $\Rightarrow$ las variables estudios y género son conjuntamente significativas.

5. Basándote en la información disponible hasta el momento, ¿qué especificación te parece más correcta?, ¿por qué?

   Solución:
   La del modelo (5). En el modelo (6), la variable género no es individualmente significativa:

$$\left.\begin{array}{ll} H_o : \gamma_3 & = 0 \\ H_a : \gamma_3 & \neq 0 \end{array}\right\} \qquad t = \frac{\hat{\gamma}_3}{\hat{\sigma}(\hat{\gamma}_3)} \overset{H_0}{\sim} t_{T-K} \quad \rightarrow \quad |t| = 0{,}49 < t_{0,025}(56) = 2$$

   Al eliminar esta variable del modelo (6) se llega al modelo (5). En este modelo, todas las variables explicativas son individualmente significativas, ya que los estadísticos t de sus coeficientes son mayores que 2 en valor absoluto. Y como consecuencia, hay omisión de variables relevantes en los modelos anteriores, (4), (3) y (2).

6. ¿Crees que podría existir algún problema en la estimación de los efectos de las variables experiencia, género y estudios sobre el salario si la muestra disponible estuviera formada por 60 individuos con estudios universitarios?

Solución:

En este caso habría un problema de multicolinealidad exacta. En efecto, la matriz de diseño del modelo (6) sería:

$$X = \begin{pmatrix} 1 & X_1 & S1_1 & 0 \\ 1 & X_2 & S1_2 & 0 \\ 1 & X_3 & S1_3 & 0 \\ \vdots & \vdots & \vdots & \vdots \\ 1 & X_{60} & S1_{60} & 0 \end{pmatrix}$$

de modo que $rg(X) = 3 < K \Rightarrow |X'X| = 0$ y no se podrían estimar de forma única los coeficientes del modelo.

El modelo estimable sería $Y_i = \alpha_1 + \alpha_2 S1_i + \alpha_3 X_i + u_i \quad \rightarrow$ no podemos estimar el efecto del nivel de estudios (lo cual es lógico, dado que esta variable toma el mismo valor para todos los individuos de la muestra).

LADE-2006.3 (Jun-2006)

Se desea analizar el gasto en teléfono en función de la renta familiar y del tamaño de familia. Para ello, se propone el siguiente modelo:

$$Y_i = \beta_1 + \beta_2\, R_i + \beta_3\, T_i + u_i \qquad i = 1, \ldots, 15 \qquad u_i \sim NID(0, \sigma^2) \tag{1}$$

donde:

- Y_i: gasto anual realizado por la unidad familiar en telefonía fija, medido en euros.
- R_i: renta familiar anual, medida en miles de euros.
- T_i: número de miembros que componen la unidad familiar.

Con los datos recogidos para 15 familias se dispone de la siguiente información:

$$(X'X)^{-1} = \begin{pmatrix} 0{,}2782 & -0{,}00409 & -0{,}03859 \\ -0{,}00409 & 0{,}0002918 & -0{,}001955 \\ -0{,}03859 & -0{,}001955 & 0{,}04139 \end{pmatrix}$$

$$\begin{array}{lll} \sum Y_i = 5475 & \sum Y_i R_i = 195375 & \sum Y_i^2 = 2340325 \\ \sum R_i = 445 & \sum Y_i T_i = 15670 & \sum R_i^2 = 18213 \\ \sum T_i = 35 & \sum R_i T_i = 1275 & \sum T_i^2 = 117 \end{array}$$

1. Estima el modelo propuesto por el método de mínimos cuadrados ordinarios. Interpreta los coeficientes estimados. ¿Tienen los signos esperados?

Solución:

$$\widehat{\beta}_{MCO} = \begin{pmatrix} \hat{\beta}_1 \\ \hat{\beta}_2 \\ \hat{\beta}_3 \end{pmatrix} = (X'X)^{-1}X'Y =$$

$$\begin{pmatrix} 0{,}2782 & -0{,}00409 & -0{,}03859 \\ -0{,}00409 & 0{,}0002918 & -0{,}001955 \\ -0{,}03859 & -0{,}001955 & 0{,}04139 \end{pmatrix} \begin{pmatrix} 5475 \\ 195375 \\ 15670 \end{pmatrix} = \begin{pmatrix} 119{,}356 \\ 3{,}983 \\ 55{,}343 \end{pmatrix}$$

$\widehat{\beta}_1 = 119{,}356$ euros es el valor esperado estimado del gasto anual en telefonía fija cuando la renta es cero y el número de miembros de la familia es cero.
$\widehat{\beta}_2 = 3{,}983$ euros es el incremento estimado en el valor esperado del gasto en telefonía fija cuando la renta familiar se incrementa en una unidad (1000 euros), permaneciendo constante el número de miembros de la familia. Es positivo, como era de esperar, indicando que a más renta más gasto.
$\widehat{\beta}_3 = 55{,}343$ euros es el incremento estimado en el valor esperado del gasto en telefonía fija cuando el número de miembros de la familia se incrementa en una unidad, permaneciendo constante la renta familiar. Es positivo, como era de esperar, indicando que a más miembros en la familia más gasto.

2. Calcula e interpreta una medida de la bondad del ajuste realizado.

Solución:

$$R^2 = 1 - \frac{SCR}{SCT} = 1 - \frac{41483{,}1}{341950} = 0{,}8787$$

$$SCT = \sum_i Y_i^2 - T\bar{Y}^2 = 2340325 - 15\left(\frac{5475}{15}\right)^2 = 341950$$

$$SCR = \sum_i Y_i^2 - \hat{\beta}'X'Y = 2340325 - (119{,}356 \;\; 3{,}983 \;\; 55{,}343) \begin{bmatrix} 5475 \\ 195375 \\ 15670 \end{bmatrix} =$$

$$= 2340325 - 2298841{,}9 = 41483{,}1$$

Interpretación: El 87,87 % de la variación en el gasto en telefonía fija es explicado por la variación de la renta familiar y la variación en el número de miembros de la familia.

3. Demuestra teóricamente a qué es igual la matriz de varianzas y covarianzas del estimador de MCO del modelo (1). Escribe las hipótesis básicas que necesites para la demostración.

Solución:

Hipótesis básicas: $u \sim (0, \sigma^2 I_T)$, el modelo está bien especificado y la matriz X es no aleatoria. Además, usamos:

$$\begin{aligned} \hat{\beta} &= (X'X)^{-1}X'Y = (X'X)^{-1}X'(X\beta + u) = \beta + (X'X)^{-1}X'u \\ E(\hat{\beta}) &= E(\beta + (X'X)^{-1}X'u) = E(\beta) + (X'X)^{-1}X'E(u) = \beta \end{aligned}$$

La definición de matriz de varianzas y covarianzas del estimador de MCO es:
$Var(\hat{\beta}) = E[(\hat{\beta} - E(\hat{\beta}))(\hat{\beta} - E(\hat{\beta}))']$, por tanto:

$$Var(\hat{\beta}) = E[(\hat{\beta} - E(\hat{\beta}))(\hat{\beta} - E(\hat{\beta}))'] = E[(\hat{\beta} - \beta)(\hat{\beta} - \beta)'] = E\left[\left[(X'X)^{-1}X'u\right]\left[(X'X)^{-1}X'u\right]'\right] =$$

$$= \quad E[(X'X)^{-1}X'uu'X(X'X)^{-1}] = (X'X)^{-1}X'E[uu']X(X'X)^{-1} =$$
$$= \quad (X'X)^{-1}X'\sigma^2 I_T X(X'X)^{-1} = \sigma^2(X'X)^{-1}X'X(X'X)^{-1} = \sigma^2(X'X)^{-1}$$

4. Estima la varianza de las perturbaciones.

   Solución:

   $\hat{\sigma}^2 = \frac{\hat{u}'\hat{u}}{T-K} = \frac{41483{,}1}{15-3} = 3456{,}92$

5. ¿Son las variables explicativas individualmente significativas? ¿y conjuntamente? Realiza los contrastes oportunos.

   Solución:

   Para contrastar si la variable renta familiar es individualmente significativa contrastamos si el parámetro que la acompaña, β_2, es igual a cero (no significativa) o no (significativa):

$$\left.\begin{array}{ll} H_0: & \beta_2 = 0 \\ H_a: & \beta_2 \neq 0 \end{array}\right\} \Rightarrow t = \frac{\hat{\beta}_2}{\sqrt{\widehat{Var}(\hat{\beta}_2)}} \overset{H_0}{\sim} t_{T-K}$$

   $|t| = |\frac{3{,}983}{\sqrt{3456{,}92 \cdot 0{,}0002918}}| = 3{,}94 > t_{15-3}^{0{,}025} = 2{,}18 \Rightarrow$ Rechazo H_0 para $\alpha = 5\,\%$ $\Rightarrow$ La renta familiar es una variable relevante.

   Realizamos el mismo ejercicio para la variable número de miembros de la unidad familiar:

$$\left.\begin{array}{ll} H_0: & \beta_3 = 0 \\ H_a: & \beta_3 \neq 0 \end{array}\right\} \Rightarrow t = \frac{\hat{\beta}_3}{\sqrt{\widehat{Var}(\hat{\beta}_3)}} \overset{H_0}{\sim} t_{T-K}$$

   $|t| = |\frac{55{,}343}{\sqrt{3456{,}92 \cdot 0{,}04139}}| = 4{,}62 > t_{15-3}^{0{,}025} = 2{,}18 \Rightarrow$ Rechazo H_0 para $\alpha = 5\,\%$ $\Rightarrow$ La variable número de miembros de la unidad familiar es una variable relevante.

   Para analizar la significatividad conjunta de ambas variables, contrastamos:

$$\left.\begin{array}{ll} H_0: & \beta_2 = \beta_3 = 0 \\ H_a: & \beta_2 \neq 0 \;\; \text{y/o} \;\; \beta_3 \neq 0 \end{array}\right\} \Rightarrow F = \frac{R^2/(K-1)}{(1-R^2)/(T-K)} \overset{H_0}{\sim} \mathcal{F}_{2,12}$$

   $F = \frac{0{,}8787/2}{(1-0{,}8787)/12} = 43{,}46 > \mathcal{F}_{2,12}^{0{,}05} = 3{,}89 \Rightarrow$ Rechazo H_0 para $\alpha = 5\,\%$ $\Rightarrow$ Ambas variables son conjuntamente significativas.

6. Contrasta que el efecto unitario del tamaño de la familia sobre el gasto medio anual en telefonía fija es 10 veces superior al de la renta familiar anual.

   Solución:

$$\left\{\begin{array}{ll} H_0: & \beta_3 = 10\beta_2 \\ H_a: & \beta_3 \neq 10\beta_2 \end{array}\right. \Leftrightarrow \left\{\begin{array}{ll} H_0: & w = \beta_3 - 10\beta_2 = 0 \\ H_a: & w = \beta_3 - 10\beta_2 \neq 0 \end{array}\right. \qquad t = \frac{\hat{w}}{\sqrt{\widehat{Var}(\hat{w})}} \overset{H_0}{\sim} t_{T-K}$$

   $\hat{w} = \hat{\beta}_3 - 10\hat{\beta}_2 = 55{,}343 - 10 \cdot 3{,}983 = 15{,}513$

   $v\hat{a}r(\hat{w}) = v\hat{a}r(\hat{\beta}_3) + 10^2 v\hat{a}r(\hat{\beta}_2) - 2 \cdot 10 \cdot c\hat{o}v(\hat{\beta}_2, \hat{\beta}_3) = \hat{\sigma}^2(0{,}04139 + 100 \cdot 0{,}0002918 - 20 \cdot (-0{,}001955)) = 379{,}12$

   $\Rightarrow |t| = \frac{15{,}513}{\sqrt{379{,}12}} = 0{,}79 < t_{15-3}^{0{,}025} = 2{,}18 \Rightarrow$ No rechazo la H_0 de que el efecto unitario del tamaño de la familia sobre el gasto medio anual en telefonía fija es 10 veces superior al de la renta familiar anual, para un nivel de significación del $5\,\%$.

7. ¿Crees posible que una familia formada por 2 miembros y con una renta anual de 30.000 euros realice un gasto anual en teléfono de 700 euros? Razona tu respuesta.

   Solución: Será posible a partir de este modelo si ese gasto está incluido en el intervalo de confianza correspondiente a su predicción, $IC_{1-\alpha}(Y_P) = (\hat{Y}_P \pm t_{T-K}^{\alpha/2} \hat{\sigma} \sqrt{1 + X_P'(X'X)^{-1}X_P})$.

$$\hat{Y}_P = 119{,}356 + 3{,}983 \cdot 30 + 55{,}343 \cdot 2 = 349{,}53$$

$$t_{12}^{0,025} = 2{,}18 \quad \hat{\sigma} = \sqrt{3456{,}92} \quad X_P' = (1\ 30\ 2) \Rightarrow X_P'(X'X)^{-1}X_P = 0{,}07202$$

$$\Rightarrow IC_{95\,\%}(Y_P) = (349{,}53 \pm 132{,}70) = (216{,}83; 482{,}23)$$

   Como $700 \notin IC_{95\,\%}(Y_P)$, con un nivel de confianza del $95\,\%$ rechazamos que una familia formada por 2 miembros y con una renta anual de 30.000 euros realice un gasto anual en teléfono de 700 euros.

   Se piensa que la disposición de teléfono móvil puede afectar al gasto realizado en telefonía fija. Por ello, también se realiza la siguiente regresión:

$$\underset{\text{(t-estad.)}}{\widehat{Y}_i} = \underset{(2,95)}{110,372} + \underset{(3,52)}{4,095}R_i + \underset{(2,41)}{46,642}T_i + \underset{(0,64)}{17,36}M_i \qquad R^2 = 0{,}883 \qquad (2)$$

   donde M_i es el número de teléfonos móviles disponibles en la unidad familiar.

8. ¿Es significativa la nueva variable incorporada en el modelo, M_i? ¿Cuáles son las propiedades de los estimadores del modelo (2)? Teniendo en cuenta tus respuestas, ¿qué modelo elegirías para explicar el gasto en telefonía fija?

   Solución:
   Para contrastar si M_i es individualmente significativa contrastamos:

$$\left.\begin{array}{ll} H_0: & \beta_4 = 0 \\ H_a: & \beta_4 \neq 0 \end{array}\right\} \Rightarrow t = \frac{\hat{\beta}_4}{\sqrt{\widehat{Var}(\hat{\beta}_4)}} \overset{H_0}{\sim} t_{T-K}$$

   $|t| = 0{,}64 < t_{15-4}^{0,025} = 2{,}20 \Rightarrow$ No rechazo H_0 para $\alpha = 5\,\%$ $\Rightarrow$ La variable M_i no es relevante.
   El estimador MCO empleado en el modelo (2) es lineal e insesgado pero pierde eficiencia en la estimación debido a la inclusión de la variable irrelevante M_i. Aún así, los contrastes en este modelo serían válidos, pero menos eficientes que en el modelo (1). Por tanto, elegiré el modelo (1) para explicar el gasto en telefonía fija.

LADE-2006.4 (Jun-2006)

Se tienen datos de 32 modelos de coche sobre las variables consumo de combustible, C_i, (en litros cada 100 km.), peso, P_i, (en Tn.) y sobre si son de 4, 6 u 8 cilindros. Para recoger esta última característica, se definen las variables ficticias:

$$D4_i = \begin{cases} 1 & \text{si } i \text{ es un coche de 4 cilindros} \\ 0 & \text{si } i \text{ es un coche de 6 u 8 cilindros} \end{cases}$$

$$M4_i = \begin{cases} 1 & \text{si } i \text{ es un coche de 6 u 8 cilindros} \\ 0 & \text{si } i \text{ es un coche de 4 cilindros} \end{cases}$$

$$D6_i = \begin{cases} 1 & \text{si } i \text{ es un coche de 6 cilindros} \\ 0 & \text{si } i \text{ es un coche de 4 u 8 cilindros} \end{cases}$$

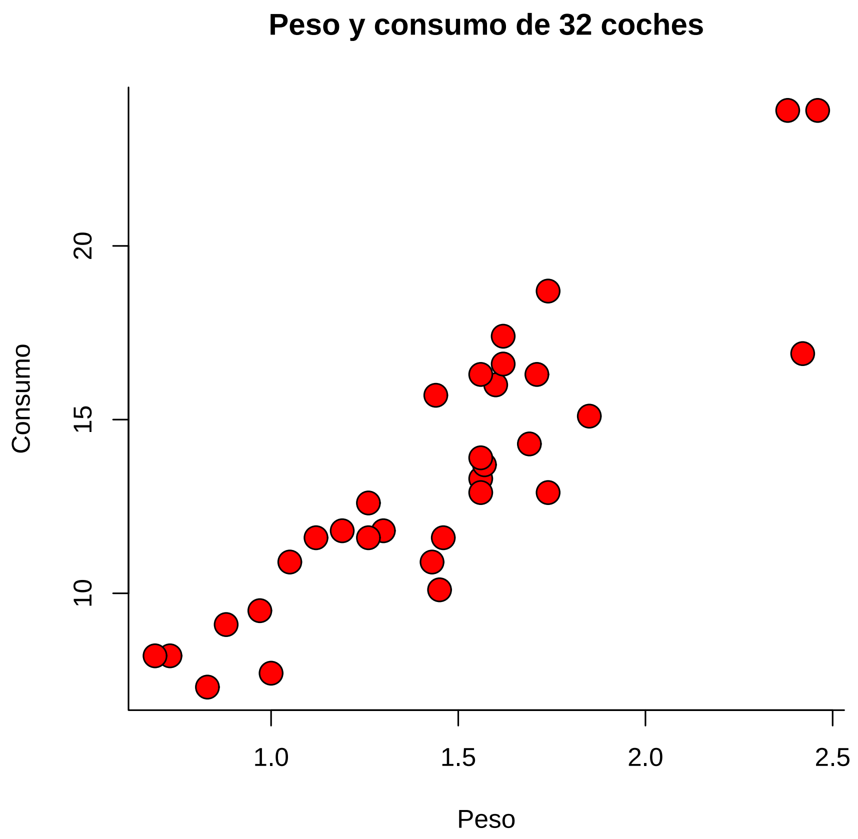
$$D8_i = \begin{cases} 1 & \text{si } i \text{ es un coche de 8 cilindros} \\ 0 & \text{si } i \text{ es un coche de 4 o 6 cilindros} \end{cases}$$

Inicialmente se estiman dos modelos:

$$\underset{\text{(t-estad.)}}{\widehat{C}_i} = \underset{(1{,}29)}{1,51} + \underset{(10{,}63)}{8,19}\, P_i \qquad \sum \hat{u}_i^2 = 118,3948 \tag{3}$$

$$\underset{\text{(t-estad.)}}{\widehat{C}_i} = \underset{(1{,}71)}{2,12}\, D4_i + \underset{(1{,}89)}{3,46}\, M4_i + \underset{(6{,}74)}{7,17}\, P_i \qquad \sum \hat{u}_i^2 = 101,7830 \tag{4}$$

1. Dibuja las rectas de regresión muestrales (3) y (4) en la siguiente nube de puntos:



2. Plantea en el modelo (4) el contraste de que la cilindrada de un coche no afecta al consumo. Realízalo con la información disponible.

Solución: Plantear en el modelo $C_i = \beta_1 D4_i + \beta_2 M4_i + \beta_3 P_i + u_i$ el contraste de que la cilindrada de un coche no afecta al consumo consiste en contrastar la hipótesis $\beta_1 = \beta_2$:

$$\begin{array}{ll} H_0: & \beta_1 = \beta_2 \Rightarrow MR: C_i = \beta \overbrace{(D4_i + M4_i)}^{=1} + \beta_3 P_i + u_i^R \Rightarrow C_i = \beta + \beta_3 P_i + u_i^R \\ H_a: & \beta_1 \neq \beta_2 \Rightarrow MNR: C_i = \beta_1 D4_i + \beta_2 M4_i + \beta_3 P_i + u_i^{NR} \end{array}$$

Podemos llevar a cabo el contraste utilizando el estadístico $F = \frac{\hat{u}'\hat{u}_R - \hat{u}'\hat{u}_{NR}}{\hat{u}'\hat{u}_{NR}} \frac{T-k}{m}$, donde m es el número de restricciones a contrastar. El estadístico sigue, bajo la hipótesis nula, la distribución $F(m, T-k)$.

En este caso, el modelo restringido (MR) coincide con el modelo estimado (3), y el modelo no restringido (MNR) con el (4). Por tanto:

$F = \frac{118{,}3948 - 101{,}7830}{101{,}7830} \frac{32-3}{1} = 4{,}73 > 4{,}18 \simeq F^{0{,}05}(1, 29) \Rightarrow$ la hipótesis nula se rechaza al nivel de significación del $5\,\%$. Existe, por lo tanto, evidencia de que la cilindrada de un coche afecta al consumo.

Considera ahora el modelo:

$$C_i = \alpha_1 D4_i + \alpha_2 D6_i + \alpha_3 D8_i + \beta P_i + u_i \quad i = 1, \ldots, 32 \tag{5}$$

3. Interpreta los coeficientes α_2 y α_3.

Solución:
$\alpha_2 = E(C_i/D4_i = 0, D6_i = 1, D8_i = 0, P_i = 0)$ es el consumo esperado de un coche de 6 cilindros con un peso de 0 toneladas.
$\alpha_3 = E(C_i/D4_i = 0, D6_i = 0, D8_i = 1, P_i = 0)$ es el consumo esperado de un coche de 8 cilindros con un peso de 0 toneladas.

El resultado de estimar el modelo (5) por MCO es:

$$\underset{\text{(t-estad.)}}{\widehat{C}_i} = \underset{(2{,}53)}{3,33}\,D4_i + \underset{(2{,}33)}{4,13}\,D6_i + \underset{(2{,}82)}{6,06}\,D8_i + \underset{(5{,}19)}{6,00}\,P_i \qquad \sum \hat{u}_i^2 = 88,44 \tag{6}$$

4. Dados los resultados, ¿cuál de las tres especificaciones te parece mejor? Razona la respuesta.

Solución:
En el apartado 2. hemos decidido que la cilindrada de un coche afecta al consumo. Por tanto, en base al contraste realizado en ese apartado, descartamos el modelo (3) y tendremos que elegir entre los modelos (4) y (5). El modelo (4) coincide con el modelo (5) salvo que en este se impone la restricción $\alpha_2 = \alpha_3$. Así, para elegir entre ambos modelos tenemos que ver si aceptamos o rechazamos esta restricción:

$$\begin{array}{ll} H_0: & \alpha_2 = \alpha_3 \Rightarrow MR\,[el(4)] : C_i = \alpha_1 D4_i + \alpha \overbrace{(D6_i + D8_i)}^{=M4_i} + \beta P_i + u_i^R \\ H_a: & \alpha_2 \neq \alpha_3 \Rightarrow MNR\,[el(5)] : C_i = \alpha_1 D4_i + \alpha_2 D6_i + \alpha_3 D8_i + \beta P_i + u_i^{NR} \end{array}$$

Podemos llevar a cabo este contraste utilizando el estadístico $F = \frac{\hat{u}'\hat{u}_R - \hat{u}'\hat{u}_{NR}}{\hat{u}'\hat{u}_{NR}} \frac{T-k}{m}$, donde m es el número de restricciones a contrastar. El estadístico sigue, bajo la hipótesis nula, la distribución $F(m, T-k)$.

Como ya hemos comentado, el modelo restringido (MR) coincide con el modelo estimado (4), y el modelo no restringido (MNR) con el modelo (5), cuya estimación la tenemos en el modelo (6). Por tanto:

$F = \frac{101{,}7830 - 88{,}44}{88{,}44} \frac{32-4}{1} = 4{,}22 > 3{,}34 \simeq F^{0{,}05}(1, 28) \Rightarrow$ la hipótesis nula se rechaza al nivel de significación del $5\,\%$. Por lo tanto, es preferible el modelo no restringido, el modelo (5).

5. Explica la diferencia entre los modelos (4) y (5).

Solución:
La diferencia está en que el modelo (4) recoge el efecto de tener un coche de 6 u 8 cilindros con una única variable ficticia, obligando al modelo a que el efecto en el consumo de los dos tipos de coche sea el mismo. El modelo (5), por contra, distingue el consumo entre coches de 6 y 8 cilindros (al proponer una variable ficticia para cada tipo) permitiendo una mayor flexibilidad en el modelo.

6. Con el modelo que has elegido en el apartado 4, ¿en cuánto estimas el consumo medio de combustible por cada 100 km. de un coche de 4 cilindros? ¿Y de un coche de 8 cilindros? ¿Y si ambos pesan 1,4 Tn.?

Solución:
El consumo medio estimado de combustible por cada 100 km. de un coche de 4 cilindros es de $\widehat{C}_i = 3,33 + 6,00P_i$ litros. El de un coche de 8 cilindros es de $\widehat{C}_i = 6,06 + 6,00P_i$

litros.

Si los coches pesan 1,4 Tn., entonces los consumos medios estimados son:

$\widehat{C}_i = 3,33 + 6,00 \cdot 1{,}4 = 11{,}73$ litros de consumo medio estimado para el coche de 4 cilindros
$\widehat{C}_i = 6,06 + 6,00 \cdot 1{,}4 = 14{,}46$ litros de consumo medio estimado para el coche de 8 cilindros