RECOPILACION DE EXAMENES DE INTRODUCCION A LA ECONOMETRIA

25 de octubre de 2005

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

Autores:

Aurora Alonso	Juan I. Modroño
Begoña Eguía	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	Jorge Virto
Beatriz Goitisolo	M. Paz Rojo
Petr Mariel	Marian Zubia

Índice

EXAMENES DE ECONOMETRIA
Plan antiguo
Temas de Introducción a la Econometría

[0] CVS Id: $Id: e1pv.tex,v 1.1.1.1 2003/09/17 14:40:56 etpdihei Exp

PV-G.1 (Febrero 1994)

1. En el modelo de regresión lineal general $Y = X\beta + U$, las propiedades del estimador MCO de β son independientes de la normalidad de las perturbaciones. Por lo tanto, los contrastes de hipótesis son independientes de este supuesto.

2. En el modelo econométrico siguiente:

$$Y_t = \alpha \, X_{1t}^{\beta_1} \, X_{2t}^{\beta_2} \, U_t$$

   no es posible estimar los parámetros α, β_1 y β_2 por MCO porque no se satisface la hipótesis de linealidad en la relación funcional.

3. Sea $Y = X\hat{\beta} + \hat{U}$. Deriva la distribución del vector de residuos mínimo cuadráticos $\hat{U}$. ¿Cuál es su vector de medias? ¿Cuál es su matriz de varianzas y covarianzas?

4. ¿Por qué nos interesa habitualmente disponer de una predicción por intervalo de la variable endógena en lugar de una predicción por punto?

PV-G.2 (Febrero 1995)

Se desea estudiar la función de consumo de energía eléctrica de las economías domésticas. Para ello se dispone de 6 datos sobre las siguientes variables:

Y_i: Consumo de energía eléctrica de la familia i.

X_i: Salario bruto de la familia i.

Sin embargo, se sospecha que las familias que están pagando un crédito hipotecario por la compra de su vivienda deben presentar una función de consumo de energía diferente al resto de las familias.

1. Comprueba si este hecho puede ser estadísticamente cierto con los siguientes datos:

   C={Conjunto de familias que están pagando un crédito hipotecario.}

   NC={Conjunto de familias que no están pagando un crédito hipotecario.}

   A={Conjunto de todas las familias.}

$$\sum_{i \in C} y_i^2 = 208{,}65 \qquad \sum_{i \in C} x_i^2 = 2314{,}65 \qquad \sum_{i \in C} x_i y_i = 529{,}33$$

$$\sum_{i \in NC} y_i^2 = 724{,}64 \qquad \sum_{i \in NC} x_i^2 = 7768{,}65 \qquad \sum_{i \in NC} x_i y_i = -2312{,}64$$

$$\sum_{i \in A} y_i^2 = 974{,}83 \qquad \sum_{i \in A} x_i^2 = 38649{,}34 \qquad \sum_{i \in A} x_i y_i = -748{,}32$$

[0]CVS Id: $Id: pargenc.tex,v 1.2 2004/07/21 09:58:18 etpdihei Exp

2. En lugar de la variable X_i se quiere considerar como variable explicativa la variable "sexo del cabeza de familia" (manteniendo la variable cred. hipot.). Plantea un modelo que recoja la relación entre el consumo de energía, la situación crediticia y el sexo del cabeza de familia. NOTA: Para ayudarte, considera que los datos son los de la siguiente tabla:

$Y_i - \overline{Y}_{total}$	Créd. Hipot.	Sexo
-4.83	sí	M
19.17	no	M
-16.83	no	F
-9.83	no	F
-1.83	sí	M
14.17	sí	F

3. Interpreta los parámetros del modelo anterior.
4. Plantea el contraste de la hipótesis de que el sexo del cabeza de familia no es una variable relevante para explicar el consumo de energía, especificando claramente cuál es la hipótesis nula, el estadístico de contraste, su distribución y la regla de decisión que se ha de utilizar.
5. Plantea el contraste de que ninguna de las dos variables cualitativas es relevante. Hazlo con el mismo detalle que en el apartado anterior.

PV-G.3 (Febrero 1996)

Considera el siguiente modelo de regresión,

$$Y_t = \beta_1 X_{1t} + \beta_2 X_{2t} + u_t$$

1. Escribe el criterio de estimación minimo-cuadrático de β_1 y β_2 sujeto a la restricción de que $\beta_1 + \beta_2 = 1$. Deriva las condiciones de primer orden. Resuelve el sistema de ecuaciones obteniendo las fórmulas explícitas para los estimadores de β_1 y β_2 por mínimos cuadrados restringidos.
2. Disponiendo de la siguiente información:
$$\sum (X_{1t} - X_{2t})^2 = 600 \qquad \sum (Y_t - X_{2t})(X_{1t} - X_{2t}) = 200$$
obtén las estimaciones de β_1 y β_2 tal que éstas satisfagan esa restricción.
3. Comenta la siguiente afirmación: "Aunque sea cierto que $\beta_1 + \beta_2 = 1$, las propiedades de los estimadores minimo-cuadráticos son las mismas impongamos o no la restricción." Razona tu respuesta.

[0] CVS Id: $Id: fingen1c.tex,v 1.2 2003/09/18 07:47:11 etpdihei Exp

PV-G.4 (Final febrero 1993)

Considera el siguiente modelo:

$$Y_t = \beta_0 + \beta_1 D_t + u_t \qquad\qquad u_t \sim \text{NID}(0, 100)$$

donde $D_t = 0$ para las primeras 20 observaciones y $D_t = 1$ para las 25 observaciones restantes.

1. Obtén los estimadores de β_0 y β_1 por el método de mínimos cuadrados ordinarios. Demuestra que el estimador mínimo-cuadrático de β_0 es igual a la media muestral de la variable dependiente para las primeras 20 observaciones.
2. Interpreta los coeficientes.
3. Calcula las varianzas de $\hat{\beta}_0$ y $(\hat{\beta}_0 + \hat{\beta}_1)$ y contrasta la hipótesis nula $H_0 : \beta_0 + \beta_1 = 1$.

PV-G.5 (Final junio 1993)

Un investigador observa que la demanda de sandías es mayor en periodos de menor precio. Se dispone de los siguientes datos sobre la cantidad de sandías vendidas (Y) y precio por unidad (X) durante 8 semanas en el verano de 1992 en una frutería de Bilbao:

Y (miles de unidades)	10	9	8	7	6	5	4	2
X (en miles)	3	4	5	5.5	6	6.5	7	7.5

donde suponemos que todas las sandías tienen igual peso.

1. Estima los parámetros de la ecuación de demanda.
2. Contrasta si un aumento de precio de las sandías afecta a la cantidad vendida.
3. ¿Cuál es el intervalo de confianza del 95 % para la cantidad vendida en la novena semana si se observó un precio $X_9 = 8$?
4. Dado que las 4 primeras observaciones pertenecen al mes de Julio y las 4 siguientes al mes de Agosto, especifica un modelo que recoja el hecho de que, en media, la cantidad de sandías vendida es menor en Agosto que en Julio. ¿Cómo contrastarías el supuesto anterior? Explica detalladamente el proceso de contraste.
5. ¿Cómo contrastarías un cambio en el conjunto de parámetros de la función de demanda de un mes a otro?

PV-G.6 (Final septiembre 1995)

Sea el modelo:

$$Y_t = \alpha + \beta_1 X_{1t} + \beta_2 X_{2t} + u_t$$

donde se cumplen todas las hipótesis básicas del MRLG, en particular $u_t \sim NID(0, \sigma^2)$.

Se dispone de la siguiente información muestral:

		$T = 10$
$\sum Y_t = 93$	$\sum Y_t^2 = 953$	$\sum X_{1t} X_{2t} = 38{,}3$
$\sum X_{1t} = 10{,}20$	$\sum X_{1t}^2 = 10{,}68$	$\sum Y_t X_{1t} = 92{,}5$
$\sum X_{2t} = 37$	$\sum X_{2t}^2 = 177$	$\sum Y_t X_{2t} = 382$

1. Estima por MCO los parámetros α, β_1 y β_2 y la matriz de varianzas y covarianzas de los estimadores MCO de β_1 y β_2.
2. Contrasta, al nivel de significación del 0.05, $H_0 : \beta_2 = 0$ frente a la alternativa $H_1 : \beta_2 > 0$.
3. Contrasta, al nivel de significación del 0.05, $H_0 : \beta_1 + 2\beta_2 = -7$ frente a la alternativa $H_1 : \beta_1 + 2\beta_2 \neq -7$.
4. Si aceptaras la hipótesis nula en c), ¿cómo utilizarías esa información para estimar los parámetros? ¿Por qué? Escribe la función objetivo a minimizar.
5. Explica claramente la diferencia entre la predicción por punto y por intervalo de Y_p dado X_{1p} y X_{2p}.

PV-G.7 (Final septiembre 1995)

Sea el modelo

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

De la regresión auxiliar de X_2 sobre X_3 se obtiene:

$$\widehat{X_{2t}} = 0{,}5 + 1{,}1875 X_{3t}$$

y además sabemos que $\sum x_{2t}^2 = 25$, $\sum x_{3t}^2 = 16$.

1. Expresa la varianza del estimador MCO de β_2 en función del coeficiente de correlación muestral entre las variables X_2 y X_3.
2. Calcula el coeficiente de correlación muestral entre X_2 y X_3.

3. ¿Puede surgir algún problema en la estimación por mínimos cuadrados ordinarios de los parámetros β_1, β_2 y β_3? Razona tu respuesta.

4. ¿Cómo cambia tu respuesta al apartado c) si de la regresión auxiliar se obtuviera, $\widehat{X_{2t}} = 0{,}5 + 0{,}125X_{3t}$?

PV-G.8 (Final febrero 1997)

Un cazador de tigres anota cada día el número de balas utilizadas (X) y el número de tigres cazados (Y). Sus últimas 7 anotaciones son:

Tigres	3	4	4	5	3	5	4
Balas	6	7	6	7	7	8	8

1. Con la muestra anterior estima el modelo siguiente:

$$Y_t = \alpha + \beta X_t + u_t \quad t = 1, \ldots, 7 \qquad u_t \sim \text{iid}(0, \sigma_u^2) \tag{1}$$

2. Contrasta la significatividad del coeficiente que acompaña a la variable exógena balas.

3. Comenta la bondad del ajuste.

4. Si te cuenta que en una día cazó con 10 balas 10 tigres ¿Te lo creerías?

El domingo se encuentra alrededor de la lumbre con unos amigos contando historias de cazadores, cuando, entre unos matorrales, surge un gran elefante enfurecido que a grandes pasos atraviesa el campamento, aplastando todo lo que pisa. Luego, sin dar tiempo a nadie a tomar sus armas, se interna de nuevo en la espesura.

Nuestro protagonista encuentra su rifle lleno de polvo y, aunque aparenta estar en perfecto estado, tiene la sospecha de que ha sido pisado por el elefante, en cuyo caso el cañón podría estar ligeramente desviado.

Con las anotaciones de los seis días siguientes estima de nuevo la ecuación (1) con los siguientes resultados:

$$\hat{Y}_t = \underset{(1{,}47)}{4,25} + \underset{(-0{,}32)}{-0,125} X_t \quad t = 8, \ldots, 13 \tag{2}$$

$$R^2 = 0,025 \qquad \sum_{t=8}^{13} \hat{u}_t^2 = 3,25$$

5. Suponiendo que la puntería de nuestro amigo y el resto de condiciones no han variado, contrasta la hipótesis de que el rifle no ha sido pisado por el elefante. Indica cual es la hipótesis nula, el modelo restringido y el no restringido. Además sabemos que la suma de cuadrados del modelo restringido es 8,73.

6. El cazador sabe que el número de tigres cazados puede depender en parte del viento que sople ese día. Explica a este cazador como recogerías en el modelo la influencia en el número de tigres cazados de que ese día haga o no viento.

PV-G.9 (Final febrero 1998)

La teoría de la curva de Phillips relaciona los salarios (Y) con la tasa de variación del desempleo (X), de forma que, a medida que se intenta reducir la tasa de variación del desempleo, se aumentan los salarios. Esta teoría suele formalizarse en los siguientes términos:

$$Y_t = \beta_0 + \beta_1 \frac{1}{X_t} + u_t, \qquad u_t \sim \text{NID}(0, \sigma_u^2) \tag{1}$$

Dada la siguiente información muestral referida a un país:

$$\sum_{t=1}^{17} Y_t = 81{,}4 \qquad \sum_{t=1}^{17} X_t^* = 12{,}1 \qquad \sum_{t=1}^{17} (Y_t - \bar{Y})^2 = 65{,}097$$

$$\sum_{t=1}^{17} (X_t^* - \bar{X}^*)^2 = 0{,}3286 \qquad \sum_{t=1}^{17} (Y_t - \bar{Y})(X_t^* - \bar{X}^*) = 2{,}9004$$

con $X_t^* = \dfrac{1}{X_t}$.

1. Estima los parámetros de la curva de Phillips. Dí qué método utilizas y porqué (sin demostrarlo).
2. Calcula el R^2 e interprétalo.
3. ¿Puedes obtener un estadístico para contrastar la $H_0 : \beta_1 = 0$ utilizando el R^2? Derívalo e interprétalo.
4. Contrasta $H_0 : \beta_1 = 0$ frenta a $H_a : \beta_1 > 0$. Interpreta el contraste en términos de la teoría que mantiene la curva de Phillips.
5. Supón que al final del año noveno tuvo lugar un cambio sociopolítico que afectó al mercado de trabajo a partir de $t = 10$. ¿Cómo recogerías y contrastarías que hubo un cambio en la estructura de la curva de Phillips tras este hecho?

PV-G.10 (Final septiembre 1998)

La cadena de perfumería LOPEZ, desea estudiar la relación entre las unidades vendidas del perfume **Clavel** y la publicidad del perfume realizada desde 1988 y 1997. Para ello, propone la siguiente relación:

$$Y_t = \alpha_1 + \alpha_2 X_t + u_t \qquad t = 1988, ..., 1997, \quad u_t \sim NID(0, \sigma^2), \quad X_t \text{ no estocástico,}$$

donde Y_t son las unidades vendidas (en cientos) del perfume en el año t y X_t es el gasto en publicidad (en 10^5 pesetas) del perfume en el año t.

Se dispone de la siguiente información muestral:

$$\sum Y_t = 440, \sum X_t = 48, \sum X_t Y_t = 2640, \sum X_t^2 = 294, \sum Y_t^2 = 23900.$$

1. Dispone de dos estimadores de α_2 para estudiar el efecto de la publicidad en las ventas: el estimador MCO y $\tilde{\alpha_2} = \frac{\sum Y_t}{\sum X_t}$. Obtén el sesgo de cada uno de ellos. ¿Qué criterio seguirías para elegir uno u otro?
2. Independientemente de la respuesta del apartado anterior y tomando el estimador MCO, contrasta si la publicidad es una variable significativa a un nivel de significación del 5 %.
3. ¿Cuál sería la predicción por punto y por intervalo, con una confianza del 95 %, de las ventas del perfume en el año 1998 si $X_{1998} = 3$?

PV-E.1 (Febrero 1994)

Un investigador desea analizar la demanda de merluza en función del precio del pollo y de su propio precio:

$$Q_t = \beta_1 + \beta_2 P_t + \beta_3 M_t + u_t \qquad t = 1, ..., 5$$

donde
- Q_t : cantidad de merluza demandada
- P_t : precio del pollo en ptas
- M_t : precio de la merluza en ptas

Si se dispone de los siguientes datos:

Q_t	3	1	8	3	5
P_t	3	1	5	2	4
M_t	5	4	6	4	6

1. ¿Qué supuestos deberías hacer sobre los diferentes elementos de la regresión si deseas que los estimadores Mínimo Cuadráticos Ordinarios sean lineales, insesgados y de mínima varianza?
2. Estimar el modelo bajo las hipótesis consideradas en el apartado anterior.
3. Propón un estimador insesgado para la varianza de las perturbaciones, y calcúlalo en la muestra.
4. Estima la matriz de varianzas y covarianzas de los estimadores MCO.
5. En caso de conocer la verdadera matriz de varianzas y covarianzas de los estimadores ¿debería ésta coincidir con la estimada? ¿por qué?

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

6. Analiza la bondad de ajuste del modelo.
7. Contrasta la hipótesis de que un incremento en una unidad en el precio de la merluza disminuye la cantidad demandada de merluza en seis unidades.
8. Contrasta la hipótesis de que los precios influyen en la cantidad demandada de merluza en sentido opuesto pero en la misma cuantía.
9. Dada la respuesta del apartado anterior, ¿existe alguna posibilidad de estimar el modelo obteniendo menores varianzas de los coeficientes estimados?
10. Contrasta que el precio del pollo no influye en la cantidad demandada de merluza.
11. Obtener una predicción por punto y por intervalo de la cantidad de merluza demandada cuando ambos precios toman como valor 10 pts.
12. Un tendero dice que si el precio de la merluza fuera 5 pts y el del pollo 2 pts no se vendería nada de merluza. ¿Encuentras justificada esta opinión?

PV-E.2 (Febrero 1995)

Supón que un investigador está estudiando la relación entre consumo de energía eléctrica (Y_i) y renta (X_i). Una muestra obtenida con 20 familias proporciona la siguiente información:

$\sum_{i=1}^{20} Y_i = 1600$ $\quad\sum_{i=1}^{20} y_i^2 = 48600$ $\quad\sum_{i=1}^{20} x_i y_i = 1800$

$\sum_{i=1}^{20} X_i = 400$ $\quad\sum_{i=1}^{20} x_i^2 = 600$

1. Obtén los siguientes valores (escribiendo también las fórmulas correspondientes):
   a) $\overline{X}$
   b) $\overline{Y}$
   c) Pendiente de la regresión MCO.
   d) Ordenada de la regresión MCO.
   e) Varianza estimada de las perturbaciones.
   f) Coeficientes de determinación: R^2 y $\overline{R}^2$.
   g) Varianza estimada de la pendiente de la regresión.
2. Contrasta la hipótesis de que la pendiente poblacional es cero, al nivel de significación del 5 %. ¿Qué significa esta hipótesis?
3. Se sospecha que el hecho de que una vivienda tenga calefacción eléctrica aumenta significativamente su consumo energético. ¿Cómo recogerías este hecho en tu modelo si tuvieras datos sobre cuáles de estas familias tienen calefacción eléctrica?
4. ¿Cómo contrastarías la hipótesis de que, independientemente de la renta, las familias en cuya vivienda hay calefacción eléctrica tienen distinto consumo de energía? Especifica las hipótesis nula y alternativa, el estadístico de contraste y la regla de decisión.

5. Si la existencia de calefacción eléctrica es una variable significativa para explicar el consumo energético, ¿qué propiedades tienen los estimadores MCO de los apartados iii) y iv)?

PV-E.3 (Final febrero 1993)

Dado el siguiente modelo de regresión:

$$Y_t = \beta_o + \beta_1 X_{1t} + \beta_2 X_{2t} + \beta_3 X_{3t} + u_t$$

Su estimación por mínimos cuadrados ordinarios es:

$$(1) \qquad \hat{Y}_t = \underset{(18.44)}{-8{,}85} + \underset{(0.31)}{0{,}19}\, X_{1t} + \underset{(0.35)}{0{,}83}\, X_{2t} + \underset{(0.26)}{0{,}2}\, X_{3t} \qquad \hat{\sigma}_{\hat{\beta}}$$

$$R^2 = 0{,}798 \qquad T = 10 \qquad STC = 10{,}058$$

Si además se dispone de los siguientes resultados:

$$(2) \qquad \hat{Y}_t = \underset{(17.19)}{-9{,}93} + \underset{(0.23)}{0{,}8}\, X_{1t} \qquad R^2 = 0{,}602$$

$$(3) \qquad \hat{Y}_t = \underset{(13.43)}{2{,}86} + \underset{(0.20)}{1{,}02}\, X_{2t} \qquad R^2 = 0{,}759$$

$$(4) \qquad \hat{Y}_t = \underset{(26.09)}{47{,}54} + \underset{(0.47)}{0{,}34}\, X_{3t} \qquad R^2 = 0{,}590$$

realiza los siguientes contrastes sobre el modelo (1):

1. $H_o : \beta_1 = \beta_3 = 0$
2. $H_o : \beta_1 = 0 \qquad H_o : \beta_3 = 0$

PV-E.4 (Final febrero 1993)

Supongamos que el modelo correcto para explicar Y_t es:

$$(1) \qquad Y_t = \beta_1 + \beta_2 X_{2t} + \beta_3 X_{3t} + u_t$$

Un **primer investigador** propone los dos modelos siguientes:

$$(2) \qquad Y_t = \alpha_1 + \alpha_{2t} X_{2t} + v_t$$

$$(3) \qquad Y_t = \alpha_1^* + \alpha_2^* X_{3t} + v_t^*$$

[0] CVS Id: $Id: finemp1c.tex,v 1.1.1.1 2003/09/17 14:40:56 etpdihei Exp

y realiza las estimaciones con una muestra de 15 datos:

(2) $$\hat{Y}_t = \underset{(6.09)}{5{,}06} + \underset{(0.10)}{0{,}98}\,X_{2t} \qquad R^2 = 0{,}91$$

(3) $$\hat{Y}_t = \underset{(6.43)}{5{,}45} - \underset{(0.09)}{1{,}00}\,X_{3t} \qquad R^2 = 0{,}93$$

Un **segundo investigador** propone:

(4) $$Y_t = \gamma_1 + \gamma_2 X_{2t} + \gamma_3 X_{3t} + \gamma_4 X_{4t} + \gamma_5 X_{5t} + u_t$$

y estima con los mismos datos que el primer investigador:

$$\hat{Y}_t = \underset{(5.11)}{4{,}68} + \underset{(0.12)}{0{,}36}\,X_{2t} - \underset{(0.18)}{0{,}637}\,X_{3t} - \underset{(2.00)}{1{,}59}\,X_{4t} + \underset{(2.26)}{2{,}33}\,X_{5t}$$

1. ¿Qué podemos decir de la significatividad individual de los parámetros que acompañan a X_2 y X_3 en los modelos (2) y (3)?
2. ¿Se cumple que $E(\hat{\alpha}_2) = \beta_2$?
3. ¿Qué podemos decir de la significatividad individual de las variables en el modelo (4)?
4. Teniendo en cuenta la información que posees, ¿alguno de los procedimientos te parece más correcto? ¿Por qué?

PV-E.5 (Final junio 1993)

Un investigador social desea analizar la teoría de que el salario anual que recibe una mujer depende de dos hechos:

- Si su propia madre realizó alguna vez un trabajo remunerado.
- Si está casada.

Para ello define las variables:

$$m_i = \begin{cases} 1 & \text{para } i \text{ con madre trabajadora} \\ 0 & \text{en caso contrario} \end{cases} \qquad c_i = \begin{cases} 1 & \text{para } i \text{ casada} \\ 0 & \text{en caso contrario} \end{cases}$$

y, siendo W_i el salario anual, propone la relación:

$$W_i = \beta_1 + \beta_2 m_i + \beta_3 c_i + u_i$$

1. Interpreta los parámetros del modelo anterior.

2. Estima la ecuación anterior por M.C.O. con los siguientes datos:

W_i (miles de pesetas)	1622	153	777	1128	336	1079	395	1286
m_i	1	0	1	1	0	1	0	0
c_i	1	0	0	0	0	1	1	1

3. Contrasta para un nivel de significación del 5 %, si la regresión estimada en el apartado anterior es significativa.

4. Para el mismo nivel de significación, contrasta la hipótesis: $H_o : \beta_2 - 2\beta_3 = 0$.

5. Explica, sin volver a estimar, cómo se alterarían las estimaciones del vector β si en lugar de la variable c_i se utiliza:

$$c_i^* = \begin{cases} 50 & \text{para } i \text{ casada} \\ 0 & \text{en caso contrario} \end{cases}$$

PV-E.6 (Final junio 1993)

Una empresa produce las 24 horas del día utilizando tres relevos de trabajadores: mañana, tarde y noche. Los directivos argumentan que los salarios de esta empresa se determinan exclusivamente por la categoría profesional de los trabajadores: técnico, obrero y aprendiz.

Especifica un modelo que recoja las características de la empresa en cuanto a turnos y categorías ¿Cuál es la función del salario de un obrero que trabaja en turno de noche? Interpreta sus parámetros.

PV-E.7 (Final septiembre 1993)

Sea el modelo :

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

Explica brevemente qué parámetros serían estimables y porqué, en cada uno de los siquientes casos:

1. $\beta_1 = 2\beta_2$

2. $X_{1t} = 2X_{2t}$

3. $X_{1t} = 2X_{2t} + X_{3t}$ y $\beta_2 = 1 - 2\beta_1$

PV-E.8 (Final septiembre 1993)

Un estudio de tabaco en U.S.A. presenta el siguiente modelo:

$$\ln C_t = \beta_0 + \beta_1 \ln P_t + \beta_2 \ln X_{1t} + \beta_3 \ln X_{2t} + u_t$$

donde:

C_t : Ventas (en miles de millares de cigarrillos) de las principales empresas de tabaco

P_t : Precio (en dólares de 1958) del cigarrillo

X_{1t} : Gastos (en miles de dolares de 1958) en publicidad en cine, TV y radio

X_{2t} : Gastos (en miles de dolares de 1958) en publicidad en prensa escrita y vallas publicitarias.

Los resultados de la estimación con datos de 1930-1978 son:

$$\hat{\beta} = \begin{bmatrix} 2{,}489 \\ -0{,}303 \\ 0{,}044 \\ 1{,}445 \end{bmatrix} \qquad \widehat{Var}_{\hat{\beta}} = \begin{bmatrix} 1470 & -0{,}303 & 0{,}028 & -0{,}050 \\ & 0{,}068 & -0{,}004 & 0{,}003 \\ & & 0{,}002 & -0{,}003 \\ & & & 0{,}011 \end{bmatrix}$$

$$\hat{u}'\hat{u} = 0{,}924 \qquad\qquad R^2 = 0{,}933$$

1. Contrasta la hipótesis nula $H_o : \beta_2 = \beta_3$. Interpreta el resultado
2. Una posible medida para limitar el consumo de tabaco consiste en controlar el volumen de publicidad. Sin embargo, existe la opinión de que esta medida sería ineficaz, ya que la publicidad no afecta al consumo total de tabaco. ¿Cómo puedes contrastar si esto último es cierto?

PV-E.9 (Final febrero 1994)

Sea la función de producción agregada para la economía española

$$Y_t = \alpha_0 + \alpha_1 X_{1t} + \alpha_2 X_{2t} + u_t \qquad t = 1962, ..., 1990$$

donde	Y_t	=	$\ln PNB_t$	PNB:	producto nacional bruto
	X_{1t}	=	$\ln L_t$	L:	trabajo
	X_{2t}	=	$\ln K_t$	K:	capital

y con esa muestra se obtiene: $\hat{\alpha} = \begin{pmatrix} 3{,}5 \\ 0{,}7 \\ 0{,}4 \end{pmatrix}$

$$\sum_{t=1962}^{1990} (X_{1t} - \overline{X}_1)^2 = 118 \qquad \sum_{t=1962}^{1990} (Y_t - \overline{Y})^2 = 178{,}38$$

$$\sum_{t=1962}^{1990} (X_{2t} - \overline{X}_2)^2 = 56 \qquad \sum_{t=1962}^{1990} (X_{1t} - \overline{X}_1)(X_{2t} - \overline{X}_2) = 60$$

1. Especifica el modelo restringido por la hipótesis nula $\alpha_1 + \alpha_2 = 1$.
2. Contrasta la hipótesis nula anterior para un nivel de significación del $5\,\%$ sabiendo que al estimar el modelo restringido se obtiene $\sum_{t=1962}^{1990} \hat{u}_t^2 = 90$.

PV-E.10 (Final junio 1994)

Considera el siguiente modelo y su estimación:

$$Y_t = \beta_1 D_{1t} + \beta_2 D_{2t} + \beta_3 D_{3t} + \beta_4 D_{4t} + \beta_5 X_t + u_t \qquad t = 1, \ldots, T$$

$$\hat{Y}_t = -5{,}7D_{1t} - 1{,}2D_{2t} + 4{,}1D_{3t} + 3{,}2D_{4t} + 0{,}8X_t \qquad t = 1, \ldots, T$$

$$T = 55 \qquad \sum_{t=1}^{T} \hat{u}_t^2 = 450$$

$$D_{it} = \begin{cases} 1 & \text{si } t \in \text{i-ésimo trimestre} \\ 0 & \text{en caso contrario} \end{cases} \qquad i = 1, \ldots, 4$$

1. Plantea la hipótesis nula de que no existe efecto estacional.
2. Lleva a cabo el contraste sabiendo que:

$$\sum_{t=1}^{T} y_t^2 = 691{,}59 \qquad \sum_{t=1}^{T} x_t^2 = 291 \qquad \sum_{t=1}^{T} x_t y_t = 203{,}7$$

PV-E.11 (Final junio 1994)

Una compañía de seguros quiere examinar la relación entre las pólizas que contratan las unidades familiares y sus rentas. Para ello obtiene una muestra aleatoria de 20 familias que se puede resumir en los siguientes datos (en unidades monetarias):

$$\sum_{i=1}^{20} Y_i = 4739 \qquad \sum_{i=1}^{20} (Y_i - \bar{Y})^2 = 263860 \qquad \sum_{i=1}^{20} X_i Y_i = 344643$$

$$\sum_{i=1}^{20} X_i = 1186 \qquad \sum_{i=1}^{20} (X_i - \bar{X})^2 = 17259{,}2 \qquad \sum_{i=1}^{20} X_i^2 = 87589$$

El resultado de la estimación por MCO de una relación lineal entre la póliza contratada Y y la renta familiar X es:

$$\hat{Y}_i = 22{,}435 + 3{,}686X_i \qquad \hat{\sigma}_u^2 = 1630{,}274 \tag{1}$$

1. Predice por punto y por intervalo el valor de la póliza que contratará una familia con una renta de 10 unidades monetarias.

2. Diez años atrás se observaba que una familia con una renta de 10 unidades monetarias elegía una póliza de 440 unidades monetarias. ¿Hay alguna evidencia que sugiera que la ecuación (1) no es adecuada para explicar lo que pasaba hace 10 años? Razona la respuesta.

3. Calcula un intervalo de confianza del 95 % para el parámetro de la renta. Interpreta su significado.

PV-E.12 (Final septiembre 1994)

La siguiente ecuación de regresión se ha estimado con 9 datos:

$$Y_i = \alpha + \beta X_i + \gamma Z_i + u_i \qquad i = 1, \ldots, 9 \tag{1}$$

$$X'X = \begin{bmatrix} 9 & 44 & 279 \\ 44 & 282 & 1995 \\ 279 & 1995 & 15061 \end{bmatrix} \qquad (X'X)^{-1} = \begin{bmatrix} 0,966 & -0,383 & 0,033 \\ -0,383 & 0,209 & -0,020 \\ 0,033 & -0,020 & 0,002 \end{bmatrix}$$

1. Calcula el coeficiente de correlación muestral entre X y Z.

2. El resultado de la estimación por MCO de la ecuación (1) ha sido:

$$\widehat{Y}_i = \underset{(0,14)}{0,98} + \underset{(5,98)}{18,64} X_i - \underset{(-5,66)}{1,80}\, Z_i \qquad R^2 = 0,86$$

   donde los números entre paréntesis son los estadísticos t.

   ¿Consideras que ha habido un problema de multicolinealidad imperfecta grave? Razona la respuesta.

PV-E.13 (Final febrero 1995)

Una empresa editorial publica una colección de 100 libros de bolsillo. Para tener información acerca del tipo de literatura más solicitada utiliza el siguiente modelo:

$$N_i = \beta_1 P_i + \beta_2 A_i + \beta_3 C_i + \beta_4 R_i + u_i \qquad i = 1, \ldots, 100$$

donde

- N_i es el número de ejemplares vendidos del volumen i-ésimo de la colección.
- $P_i = \begin{cases} 1 & \text{si la novela pertenece al género policíaco} \\ 0 & \text{en caso caso contrario.} \end{cases}$
- $A_i = \begin{cases} 1 & \text{si la novela pertenece al género de aventuras} \\ 0 & \text{en caso caso contrario.} \end{cases}$

- $C_i = \begin{cases} 1 & \text{si la novela es de ciencia-ficción} \\ 0 & \text{en caso caso contrario.} \end{cases}$

- $R_i = \begin{cases} 1 & \text{si la novela pertenece al género romántico} \\ 0 & \text{en caso caso contrario.} \end{cases}$

Las ventas para el año 1994 se resumen en:

$$\sum_{i \in \text{policíaca}} N_i = 380 \qquad \sum_{i \in \text{ aventuras}} N_i = 150 \qquad \sum_{i \in \text{ ciencia-ficción}} N_i = 70 \qquad \sum_{i \in \text{ romántica}} N_i = 500$$

$$\sum_{i=1}^{i=100} N_i^2 = 20000 \qquad \sum_{i=1}^{100} P_i = \sum_{i=1}^{100} A_i = \sum_{i=1}^{100} C_i = \sum_{i=1}^{100} R_i = 25$$

1. Interpreta los parámetros del modelo propuesto.
2. Estima los parámetros por MCO.
3. Contrasta si hay diferencias significativas en el volumen de ventas según el género literario.
4. Halla un intervalo de confianza del 95 % para el volumen esperado de ventas de novelas de ciencia-ficción.

PV-E.14 (Final febrero 1995)

Un economista estima el siguiente modelo:

$$Y_t = \beta_1 + \beta_2 X_{2t} + \beta_3 X_{3t} + u_t \qquad t = 1, \ldots, 36 \qquad u \sim N(0, \sigma_u^2 I)$$

Obteniendo los siguientes resultados:

$$\underset{(\text{estadístico } t \rightarrow)}{\hat{Y}_t} = 0{,}98 + \underset{(0.27)}{1{,}33}\, X_{2t} + \underset{(0.98)}{0{,}78}\, X_{3t} \qquad R^2 = 0{,}989$$

1. ¿Te parece que el modelo puede tener algún problema? ¿Cuál? ¿En qué basas tus respuestas?
2. ¿Cómo cambiarían tus conclusiones si en la estimación anterior el coeficiente de determinación hubiera sido $R^2 = 0{,}12$?

PV-E.15 (Final septiembre 1995)

Con una muestra de 30 observaciones se ha estimado por MCO la siguiente ecuación:

$$Y_t = \beta_1 + \beta_2 X_{2t} + \cdots + \beta_5 X_{5t} + u_t \tag{1}$$

obteniéndose un $R_1^2 = 0{,}76$

Posteriormente se ha introducido una nueva variable X_{6t} y se ha estimado el modelo:

$$Y_t = \beta_1 + \beta_2 X_{2t} + \cdots + \beta_5 X_{5t} + \beta_6 X_{6t} + v_t \tag{2}$$

obteniéndose $R_2^2 = 0{,}79$

1. ¿Cuál de los dos modelos presentará una SRC (Suma residual de cuadrados) menor?
2. Bajo el supuesto de que $v_t \sim NID(0, \sigma_v^2)$ contrasta, al nivel de significación del 5 % la hipótesis nula $H_0 : \beta_6 = 0$ contra la alternativa $H_a : \beta_6 \neq 0$
3. Tomando como base la SRC, ¿Cuál de los dos modelos se ajusta mejor a los datos?
4. Basándote en el resultado del apartado b), ¿Cuál de los dos modelos está mejor especificado?
5. ¿Podrías dar una explicación a la aparente contradicción entre las respuestas a los apartados c) y d)?

PV-E.16 (Final febrero 1996)

Supón que en el siguiente modelo se cumplen todas las hipótesis básicas:

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

Sobre estas variables se tiene la siguiente información:

t	Y_t	X_{2t}	X_{3t}
1	1	0	1
2	4	1	2
3	0	0	0
4	6	2	0
5	7	1	0
6	3	2	1

$$X'X = \begin{pmatrix} 6 & 6 & 4 \\ 6 & 10 & 4 \\ 4 & 4 & 6 \end{pmatrix} \qquad X'Y = \begin{pmatrix} 21 \\ 29 \\ 12 \end{pmatrix}$$

$$(X'X)^{-1} = \begin{pmatrix} 0,55 & -0,25 & -0,2 \\ -0,25 & 0,25 & 0 \\ -0,2 & 0 & 0,3 \end{pmatrix}$$

1. Estima los parámetros β_1, β_2 y β_3 por MCO, y obtén una estimación de σ_u^2
2. ¿Son los estimadores de los parámetros β_1, β_2 y β_3 del apartado a) insesgados y de mínima varianza?, ¿Por qué?
3. ¿Es el estimador de σ_u^2 del apartado a) insesgado?
4. Contrasta al nivel de significación del 5 % la hipótesis $\beta_2 + \beta_3 = 2$

5. Comprueba, con un nivel de confianza del 95 % si la observación ($Y = 6, X_2 = 1, X_3 = 3$) ha podido ser generada por el modelo (1).

6. Con los mismos datos de Y_t y X_{2t} del enunciado inicial estima por MCO los parámetros γ_1 y γ_2 de la ecuación:

$$Y_t = \gamma_1 + \gamma_2 X_{2t} + v_t \tag{2}$$

7. ¿Es casual la relación que hay entre $\widehat{\gamma_2}$ y $\widehat{\beta_2}$?

PV-E.17 (Final septiembre 1996)

Dadas las observaciones:

Y	X_2	X_3	X_4
1	Hombre	2	Casado
1	Mujer	-1	Soltera
-1	Hombre	0	Casado
0	Mujer	-1	Soltera
1	Hombre	2	Casado

1. Obtener las estimaciones MCO de los coeficientes del modelo lineal que relaciona a Y con una constante, X_2 y X_3. (Nota: **NO TE DESPISTES** no te estamos mencionando a X_4)

2. En el modelo anterior, estimar la varianza de la perturbación y la matriz de varianzas y covarianzas de los estimadores.

3. Calcular el coeficiente de determinación.

4. Comprobar que las variables explicativas son ortogonales a los residuos mínimo-cuadráticos.

5. Contrastar, al nivel de significación del 5 %, si el sexo es una variable relevante.

6. Si además se introduce como explicativa la variable X_4, ¿Se pueden estimar por MCO los parámetros de este nuevo modelo? ¿Por qué?

PV-E.18 (Final febrero 1997)

Sea el modelo de regresión siguiente:

$$Y_t = \beta_0 + \beta_1 X_{1t} + \beta_2 X_{2t} + u_t \qquad \text{con} \qquad u_t \sim NID(0, \sigma_u^2)$$

Para su estimación se dispone de los siguientes datos:

$$T = 6, \qquad\qquad Y'Y = 56$$

$$(X'X)^{-1} = \begin{pmatrix} 0,8 & & \\ -0,5 & 0,5 & \\ -0,7 & 0,5 & 0,8 \end{pmatrix}, \qquad X'Y = \begin{pmatrix} 12 \\ 10 \\ 4 \end{pmatrix}$$

1. Estimar los parámetros β_0, β_1 y β_2 por MCO.
2. Calcular una estimación de la varianza de la perturbación.
3. Calcular los coeficientes R^2 y $\overline{R}^2$.
4. Contrasta si las variables explicativas son conjuntamente significativas.
5. ¿Es β_1 significativamente menor que 1 ?
6. Contrasta la hipótesis nula $H_0 : \begin{bmatrix} \beta_1 = 2 \\ \beta_2 = 3 \end{bmatrix}$

PV-E.19 (Final febrero 1997)

Se desea estudiar cómo influye la existencia de ABS y de Airbag en la demanda de automóviles. Para ello se dispone de una base de datos formada por 100 modelos de los que se conoce el número de unidades vendidas y si contienen o no las opciones mencionadas.

1. Propón un modelo que permita analizar adecuadamente la influencia de cada opción sobre el número de unidades vendidas.
2. Interpreta los parámetros del modelo propuesto.
3. ¿Cómo contrastarías la hipótesis de que la tenencia o no de ABS no afecta a las ventas?

PV-E.20 (Final febrero 1997)

Consideremos el siguiente modelo:

$$Y_t = \beta_0 + \beta_1 X_{1t} + \beta_2 X_{2t} + u_t, \qquad \text{donde} \qquad u_t \sim NID(0, \sigma_u^2)$$

Con una muestra de 10 observaciones en la que $\sum x_{1t}^2 = 50$ hemos estimado el modelo, obteniendo $\widehat{\beta}_1 = 2$.

1. Demuestra que

$$Var(\widehat{\beta}_1) = \frac{\sigma_u^2}{\sum x_{1t}^2(1 - r_{12}^2)}$$

siendo r_{12} el coeficiente de correlación muestral simple entre las variables X_1 y X_2.

2. Si $\sigma_u^2 = 1$, contrasta la hipótesis de significatividad individual de la variable X_1 en los dos siguientes casos:

   a) $r_{12} = 0,5$

   b) $r_{12} = 0,999$

3. Explica detalladamente la diferencia entre las situaciones de los apartados i) y ii).

PV-E.21 (Final junio 1997)

Un economista especializado en temas de alimentación cree que la cantidad de buey consumida (Y) en toneladas a lo largo de un año, depende de las siguientes variables:

- El precio de la carne de buey, (X_1), en pesetas por kilo.
- El precio de la carne de cerdo, (X_2), en pesetas por kilo.
- El precio de la carne de pollo, (X_3), en pesetas por kilo.
- La Renta Familiar, (X_4), en miles de pesetas.

a través del modelo:

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

Estimando los coeficientes por mínimos cuadrados, con datos anuales desde 1968 hasta 1997, se obtuvo el siguiente modelo:

$$\underset{(Desv.\,tipicas\,\rightarrow)}{\widehat{log(Y_t)}} = -0{,}024 - \underset{(0{,}168)}{0{,}529} log(X_{1t}) + \underset{(0{,}103)}{0{,}217} log(X_{2t}) + \underset{(0{,}106)}{0{,}193} log(X_{3t}) + \underset{(0{,}163)}{0{,}416} log(X_{4t})$$

R^2 = 0,683

1. Interpreta los coeficientes β_1 y β_2.

2. Contrasta a un nivel de significación del 5 %, la hipótesis de que al aumentar la renta familiar se incrementa la cantidad de buey consumida.

3. Contrasta la hipótesis de que las cuatro variables ($log(X_{1t})$, $log(X_{2t})$, $log(X_{3t})$ y $log(X_{4t})$), no tienen en conjunto una influencia lineal sobre la cantidad de buey consumida.

   Un experto en alimentación opina que, debido a la epidemia de las vacas locas, el término constante ha de ser diferente en los años 1996 y 1997 con respecto a los años anteriores.

4. ¿Cómo especificarías un modelo en el que además de las variables explicativas ya mencionadas, se recogiera este efecto?

5. ¿Cómo contrastarías en este modelo que el experto en alimentación se equivoca?. Especifica la H_0 y la H_a, cuál sería el estadístico a utilizar y el criterio de decisión.

PV-E.22 (Final febrero 1998)

La asociación de comerciantes de la Margen Izquierda del Nervión desea analizar la evolución temporal de las ventas durante los últimos 8 años. Para ello, plantea el modelo siguiente:

$$Y_t = \alpha + \beta t + u_t \qquad t = 1, \ldots, 8 \tag{1}$$

donde:

- Y_t: ventas en la zona en decenas de miles de millones de pesetas.
- t: tiempo en años.

Disponiendo de las siguientes observaciones para las ventas:

Y_t	15	15	16	14	6	5	7	9

1. Estima el modelo (1) por MCO (utilizar al menos 4 decimales).
2. Obtén los residuos $\hat{u}_t : \; t = 1, 2, \ldots, 8$ y estima σ_u^2.
3. Contrasta la hipótesis de que las ventas siguen una trayectoria temporal decreciente (es decir, si t aumenta, Y_t disminuye).
4. Calcula e interpreta el coeficiente de determinación R^2.

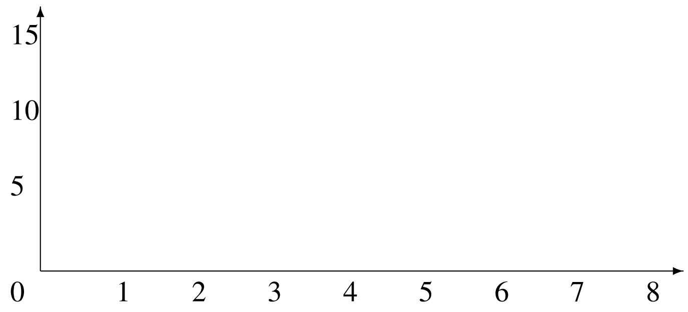
Un dirigente de la asociación señala el hecho de que hace 5 años (en $t = 4$) abrieron sus puertas varias grandes superficies en la zona que podrían haber afectado a los comerciantes de la asociación. Por ello, especifica el modelo

$$Y_t = \delta + \gamma D_t + v_t \qquad t = 1, 2, \ldots, 8 \tag{2}$$

$$D_t = \begin{cases} 1 & \text{si } t < 4 \\ 0 & \text{si } t \geq 4 \end{cases}$$

6. Interpreta los parámetros de este modelo. ¿Qué signo esperarías que tuviera el coeficiente γ estimado basándote en lo obtenido en el apartado c)?

7. Dibuja las funciones de regresión muestrales de los modelos (1) y (2), junto a las observaciones muestrales.



PV-E.23 (Final junio 1998)

Se quieren analizar las ventas de ropa vaquera en el País Vasco, Y, en función del sexo y de la edad del comprador. Los datos disponibles para una muestra de 10 personas son:

Y	59	60	20	15	30	48	36	35	22	43
Sexo	M	M	M	M	M	H	H	H	H	H
Edad	23	25	38	39	20	40	50	18	53	46

1. Especifica un modelo para el comportamiento de Y_i con la información que se da.
2. Escribe la matriz de datos correspondiente al modelo especificado en el apartado anterior.
3. En el supuesto de que se satisfagan las hipótesis básicas, incluyendo la normalidad, ¿qué método de estimación propondrías? Razona sus propiedades.
4. Si se considera que para la variable edad es suficiente con especificar si se es mayor de 26 años o no, especifica un modelo alternativo e interpreta sus coeficientes.
5. ¿Cómo contrastarías que, en términos medios, se vende más ropa vaquera a las mujeres que a los hombres?
6. ¿Qué sucedería si las ventas de este tipo de ropa dependieran solamente del salario de los compradores?
7. Propón un modelo en el que se cumpla la sospecha del apartado anterior. Interpreta sus coeficientes.

PV-E.24 (Final junio 1998)

Un país que tiene que importar todo el petróleo que consume, está interesado en realizar una política de construcción de viviendas tal que reduzca el consumo de petróleo destinado a calefacción.

Se tienen datos referidos a los últimos siete años de las siguientes variables: consumo anual de petróleo para calefacción, Y (en miles de bidones); temperatura promedio en invierno, X_1 (en grados centígrados), y cantidades de aislamiento exigidas en las paredes, X_2 (en pulgadas). Los datos se resumen en:

$$\begin{array}{lll}
\sum_{t=1}^{7} Y_t = 1479 & \sum_{t=1}^{7} X_{1t} = 73 & \sum_{t=1}^{7} X_{2t} = 42 \\
\sum_{t=1}^{7} Y_t^2 = 314465 & \sum_{t=1}^{7} X_{1t}^2 = 779 & \sum_{t=1}^{7} X_{2t}^2 = 274 \\
\sum_{t=1}^{7} Y_t X_{1t} = 15295 & \sum_{t=1}^{7} Y_t X_{2t} = 8978 & \sum_{t=1}^{7} X_{1t} X_{2t} = 431
\end{array}$$

1. Estimar la ecuación de regresión $Y_t = \beta_0 + \beta_1 X_{1t} + \beta_2 X_{2t} + u_t$, sabiendo además que

$$(X'X)^{-1} = \begin{bmatrix} T & \sum X_{1t} & \sum X_{2t} \\ & \sum X_{1t}^2 & \sum X_{1t}X_{2t} \\ & & \sum X_{2t}^2 \end{bmatrix}^{-1} = \begin{bmatrix} 11,6080 & -0,7966 & -0,5262 \\ & 0,0646 & 0,0205 \\ & & 0,0520 \end{bmatrix}.$$

2. Si las casas tuviesen 5,5 pulgadas de aislamiento y la temperatura promedio de invierno fuese de 9 grados, ¿cuánto estimas que sería el consumo de petróleo para calefacción?, ¿entre qué valores oscilaría el consumo de petróleo para calefacción, con una confianza del 95 %?

3. Supón que se realiza de nuevo la estimación añadiendo una variable referida al porcentaje de humedad en la atmósfera, y la suma de cuadrados de los residuos disminuye en dos unidades. Contrasta la hipótesis de que dicha variable no es relevante.

PV-E.25 (Final septiembre 1998)

Una consultoría fiscal desea estudiar los impuestos directos pagados por los contribuyentes, Y. Para hacerlo dispone de información sobre las variables:

X_1 = renta total anual
X_2 = ingresos anuales del trabajo
X_3 = deducciones totales
X_4 = estado civil (= 1 si casado y 0 en caso contrario, que llamaremos soltero)

El modelo inicial propuesto para realizar este estudio es:

$$Y_i = \beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + \beta_3 X_{3i} + \beta_4 X_{4i} + \beta_5 X_{1i}X_{4i} + \beta_6 X_{2i}X_{4i} + \beta_7 X_{3i}X_{4i} + u_i \quad N = 65$$

1. Plantea, paso a paso, el contraste de que no hay diferencia entre el valor medio de los impuestos directos pagados por los solteros y por los casados. Escribe la hipótesis nula, el estadístico de contraste y el modelo resultante si no se rechazara la hipótesis nula.

2. ¿Qué efectos adicionales se recogen en el modelo inicial propuesto respecto al modelo más sencillo

$$Y_i = \beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + \beta_3 X_{3i} + \beta_4 X_{4i} + u_i \quad ?$$

[0] CVS Id: $Id: e1pn.tex,v 1.1.1.1 2003/09/17 14:40:56 etpdihei Exp

EXAMENES DE INTRODUCCION A LA ECONOMETRIA, LE-LADE

LE-1997.1 (Feb-1997)

Con una muestra de 30 observaciones se ha estimado por Mínimos Cuadrados Ordinarios el siguiente modelo:

$$Y_t = \alpha + \beta X_t + \gamma Z_t + u_t \tag{1}$$

donde

- $Y_t =$ es el logaritmo del output de una empresa en el período t.
- $X_t =$ es el logaritmo del capital utilizado por la empresa en el período t.
- $Z_t =$ es el logaritmo del trabajo utilizado por la empresa en el período t.

1. Encuentra los valores que se han dejado en blanco al escribir los resultados de dicha estimación:

$$\hat{Y}_t = \underset{(......)}{0{,}4252} + \underset{(0.063)}{0{,}9491} X_t + \underset{(0.0659)}{......} Z_t$$

$$t - muestral : (3{,}0789) \qquad (......) \qquad (11{,}1588)$$

$$R^2 = 0{,}918 \qquad \bar{R}^2 = ... \qquad \sum_{t=1}^{T} \hat{u}_t^2 = 5{,}0232 \qquad \hat{\sigma} = ...$$

2. Lleva a cabo contrastes de significatividad individual de las variables X_t y Z_t.

3. En la estimación donde se omite la variable X_t se ha encontrado:

$$\hat{Y}_t = \underset{(0.314)}{-0{,}9378} + \underset{(0.1955)}{0{,}5648} Z_t$$

$$R^2 = 0{,}2296 \qquad \bar{R}^2 = ,2021 \qquad \sum_{t=1}^{T} \hat{u}_t^2 = 47{,}216 \qquad \hat{\sigma} = 1{,}2985$$

Si sabes que $\sum x_t^2 = 94{,}89$, $\sum z_t^2 = 77{,}38$ y $\sum x_t z_t = 31{,}45$, calcula el sesgo del estimador de γ en este modelo si el verdadero valor del parámetro β fuera 0.7; ¿Tu resultado implica que, en promedio, estas subestimando el verdadero valor del parámetro γ?

4. Si los datos de la variable Z_t los dividimos entre 10 ¿Cómo varían las estimaciones de los parámetros α, β, γ de la ecuación (1) ¿Y la $\sum_{t=1}^{T} \hat{u}_t^2$? Entonces, ¿qué ocurre con los contrastes anteriores?

[0] CVS Id: $Id: 97e1g.tex,v 1.1.1.1 2003/09/17 14:40:56 etpdihei Exp

LE-1997.2 (Feb-1997)

Contesta lo más detalladamente posible a las siguientes preguntas:

1. En el modelo:

$$Y_i = \beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + \beta_3 X_{3i} + u_i \qquad i = 1, ..., 100$$

   Escribe la matriz R, el vector r y q=número de restricciones, para cada una de las siguientes hipótesis nulas:

   a) $H_o : \beta_1 = \beta_2 = \beta_3$

   b) $H_o : \beta_1 = 5, \;\; \beta_2 = 0, \;\; \beta_3 = -1$

2. Escribe los modelos restringidos bajo las dos hipótesis del apartado anterior. Deriva los estimadores MCO de ambos modelos restringidos.

3. Si en el modelo anterior observamos que $5X_{1i} = X_{3i}, \forall i$, entonces tenemos multicolinealidad exacta y sólo se pueden estimar combinaciones lineales de los parámetros, pero esto no es cierto si observamos $X_{2i} = 0 \;\; \forall i$.

LE-1997.3 (Feb-1997)

Para analizar el salario de un profesor universitario se utiliza la ecuación de regresión:

$$Y_i = \alpha_1 + \alpha_2 D_{2i} + \alpha_3 D_{3i} + \alpha_4 (D_{2i} D_{3i}) + \beta X_i + u_i \qquad i = 1, \ldots N$$

donde Y es el salario anual del profesor, X los años de experiencia docente, D_2 es una variable ficticia que vale 1 si el profesor es hombre y 0 en los demás casos y D_3 es otra variable ficticia que vale 1 si el profesor es de raza blanca y 0 en los demás casos.

1. ¿Cuál es la función de salario de un profesor varón de raza blanca?

2. ¿Cómo contrastarías la hipótesis .en la determinación del salario hay discriminación por raza entre los hombres pero no entre las mujeres¿

3. El término $(D_{2i} D_{3i})$ recibe habitualmente el nombre de .efecto interacción". Justifica por qué aparece esa variable en la ecuación y por qué recibe ese nombre. ¿Qué significado tiene el coeficiente α_4?

LE-1997.4 (Jun-1997)

Comenta y justifica si las siguientes afirmaciones son verdaderas o falsas (analiza cada afirmación independientemente del resto)

1. .[El] modelo $Y = X\beta + U$ satisface todas las hipótesis clásicas del modelo de regresión, y además se sabe que $\beta_2 = \beta_3 = 0$. Si tomas 100 muestras diferentes y para cada una de ellas contrastas la hipótesis $\beta_2 = \beta_3 = 0$ utilizando el test $\mathcal{F}$ con $\alpha = 0{,}05$, entonces aproximadamente en un $5\,\%$ de los casos rechazarás la hipótesis nula."
2. .[Ex]iste un criterio que permite tomar decisiones sobre la elección de un estimador insesgado y otro sesgado pero de menor varianza."
3. "La correlación muestral entre los residuos mínimo cuadráticos y cada una de las variables explicativas es siempre positiva"
4. .[En] una regresión con cinco variables explicativas y tamaño muestral igual a 30, se obtuvo un $R^2 = 0{,}72$. Al incluir una nueva variable explicativa y reestimar el modelo con las seis variables, se obtuvo un $R^2 = 0{,}75$. La nueva variable explicativa no es significativa".

LE-1997.5 (Jun-1997)

Supón que eres el gerente de una empresa de vestidos de alta calidad, y tratas de analizar la incidencia que el nivel de renta (X) y el nivel de empleo (Z) pueden ejercer sobre el volumen de ventas (Y). Dispones de observaciones para el periodo comprendido entre el primer trimestre de 1992 y el cuarto trimestre de 1995, esto es, de una muestra de 16 datos trimestrales. A partir de esa muestra se han obtenido los siguientes datos:

$$x'x = \begin{pmatrix} 0.43 & 265.4 \\ & 175559 \end{pmatrix} \qquad (x'x)^{-1} = \begin{pmatrix} 34.7 & -0.05 \\ & 0.000085 \end{pmatrix}$$

$$x'y = \begin{pmatrix} 4.4 \\ 2690.1 \end{pmatrix} \qquad \sum y_t^2 = 110{,}8$$

Si la especificación del modelo es

$$Y_t = \beta_1 + \beta_2 X_t + \beta_3 Z_t + u_t$$

1. Halla las estimaciones MCO de β_2 y β_3, y sus varianzas y covarianzas estimadas respectivas.
2. Contrasta tanto individual como conjuntamente la significatividad de las variables explicativas X_t y Z_t. Interpreta los resultados.
3. Dado que los datos son trimestrales, se puede plantear la posibilidad de que la variable endógena presente estacionalidad. Ante este hecho se puede decidir especificar el modelo:

   $$Y_t = \alpha_0 + \alpha_1 D_{1t} + \alpha_2 D_{2t} + \alpha_3 D_{3t} + \alpha_4 D_{4t} + \alpha_5 X_t + \alpha_6 Z_t + u_t$$

   donde: $D_{it} = \begin{cases} 1 & \text{si t pertenece al trimestre i.} \\ 0 & \text{caso contrario.} \end{cases}$

   ¿Es posible obtener estimaciones MCO para todos los parámetros del modelo? Razona la respuesta.

4. ¿Podrías plantear algún otro modelo en el que se incorpore estacionalidad? Interpreta los coeficientes del modelo.
5. Supón que has estimado el modelo propuesto en el apartado anterior por mínimos cuadrados ordinarios y has obtenido que la suma de los residuos al cuadrado es 2.39. Contrasta la existencia de estacionalidad.
6. Dado el resultado de este contraste ¿Qué propiedades tienen los estimadores del primer apartado?

LE-1997.6 (Jun-1997)

Dos investigadores (A y B), trabajando independientemente, estiman por MCO la ecuación de regresión lineal:

$$Y_t = \beta_0 + \beta_1 X_{1t} + \beta_2 X_{2t} + U_t \qquad t = 1, \ldots, 20$$

X_1 y X_2 son variables no estocásticas y se supone que se verifican las hipótesis del modelo clásico de regresión lineal. De las muestras que cada investigador tenía y de los resultados obtenidos para cada uno se sabe:

	$\sum X_{1t}$	$\sum X_{2t}$	$\sum Y_t$	$\sum X_{1t}^2$	$\sum X_{2t}^2$	$\sum X_{1t}X_{2t}$	$\sum X_{1t}Y_t$	$\sum X_{2t}Y_t$	$\hat{\beta}_0$	$\hat{\beta}_1$	$\hat{\beta}_2$
A	100	120	500	600	740	640	2850	3160	-12.5	1.5	5
B	200	100	700	2400	620	1200	8400	4300	0	1	5

Posteriormente, los investigadores deciden intercambiar sus muestras (es decir, en lo que sigue ambos investigadores conocen todo el cuadro anterior)

1. A partir de toda la información disponible, el investigador A propone como estimador $\tilde{\beta}_1 = \frac{(\hat{\beta}_1^A + \hat{\beta}_1^B)}{2}$. Halla la estimación de $\tilde{\beta}_1$ e indica sus propiedades.
2. El investigador B propone una estimación para β_1 que se obtiene de la aplicación de MCO a toda la muestra. Deja indicada la matriz $X'X$ y el vector $X'Y$ de datos que dan lugar a esta estimación.
3. ¿Qué estimador crees tú que es preferible? ¿Por qué?

LE-1997.7 (Jun-1997)

Comenta por qué pueden ocurrir los siguientes resultados:

1. Al eliminar una variable de una regresión cambian drásticamente las estimaciones de los parámetros, así como las varianzas estimadas, pasando a ser las variables muy significativas cuando anteriormente no lo eran.

2. Al eliminar unas pocas observaciones de una regresión ocurre el mismo fenómeno del apartado anterior.
3. Al eliminar una variable de una regresión obtengo prácticamente las mismas estimaciones de los parámetros que cuando está incluida y tambíen prácticamente las mismas varianzas estimadas.

LADE-1997.1 (Ene-1997)

Bilbao Iniciativas Turísticas (BIT), tras 10 años de fomentar el turismo en Bilbao, quiere estudiar el éxito de su gestión. Para ello propone el siguiente modelo:

$$Y_t = \alpha + \beta X_t + u_t \qquad t = 1987, ..., 1996 \tag{1}$$

donde Y_t es el número de visitantes llegados a Bilbao, y X_t son los gastos en promoción.

A partir de la siguiente información muestral:

$$\begin{array}{ccc} \sum X_t = 37,2 & \sum X_t^2 = 147,18 & \sum Y_t^2 = 597,03 \\ \sum X_t Y_t = 295,95 & \sum Y_t = 75,5 & \end{array}$$

1. Estima el modelo propuesto por Mínimos Cuadrados Ordinarios.
2. Interpreta los coeficientes del modelo (1).
3. Calcula el coeficiente de determinación e interprétalo.
4. ¿Crees que los gastos en publicidad aumentan el número de visitantes? Haz el contraste oportuno.
5. Calcula un intervalo de confianza para el número de personas que visitarán Bilbao en 1997, si los gastos de promoción son 2,5.
6. El analista que realiza el estudio cree que hay otra variable que influye en el número de visitantes, el clima. Así, cree que si el año que se considera es lluvioso el número de visitantes será distinto que si es soleado. (NOTA: Supón que el año sólo puede ser lluvioso o soleado.)
   a) Formula un modelo que recoga también este hecho.
   b) ¿Cómo contrastarías si el clima influye en el número de visitantes? Escribe la hipótesis nula y el estadístico de contraste.

LADE-1997.2 (Ene-1997)

Sea el modelo:

$$Y_t = \gamma_0 + \gamma_1 Z_t + \gamma_2 X_t + \gamma_3 W_t + u_t \quad t = 1, \ldots, T$$

donde las variables explicativas no son estocásticas y $u_t \sim N(0, \sigma_u^2)$.

[0] CVS Id: $Id: 97e1e.tex,v 1.1.1.1 2003/09/17 14:40:56 etpdihei Exp

1. Si **no** se rechazaran las hipótesis nulas de que $H_o: \gamma_1 = 0$, $H_o: \gamma_2 = 0$, $H_o: \gamma_3 = 0$ ¿podemos afirmar que el modelo correcto es $Y_t = \gamma_0 + u_t$?

2. Indica, para cada uno de los siguientes casos, qué problema presenta, cuáles son los parámetros estimables, cuál sería el método de estimación que emplearías y por qué lo propones.

   a) $\gamma_1 = \gamma_2 = \gamma_3$

   b) $W_t = 3X_t + 5Z_t - 7$

LADE-1997.3 (Ene-1997)

Compara las siguientes dos ecuaciones, estimadas con datos anuales de 1960 a 1994 ($T = 35$):

$$\widehat{QD}_t = -60{,}5 - \underset{(0.07)}{0{,}45}\ P_t + \underset{(0.05)}{0{,}12}\ PC_t + \underset{(1.2)}{12{,}2}\, YD_t \qquad \overline{R}^2 = 0,984 \tag{2}$$

$$\widehat{QD}_t = -80{,}7 - \underset{(0.06)}{0{,}34}\ P_t + \underset{(0.42)}{15}\ YD_t \qquad \overline{R}^2 = 0,981 \tag{3}$$

donde QD_t es la cantidad demandada de pollo en el año t, P_t su precio, PC_t un índice del precio de otras carnes, e YD_t la renta disponible.

1. ¿Te parece que PC_t es una variable relevante omitida en la ecuación (3) ? Explica con detalle en qué basas tu respuesta.

2. ¿Crees que el estadístico t del coeficiente asociado a la variable YD_t del modelo (3) $t = 15/0{,}42$ es adecuado para contrastar la significación individual de esta variable? Razónalo brevemente.

LADE-1997.4 (Jun-1997)

La empresa PLASTIPLAS S.A. se dedica a la fabricación de tarrinas para helados y vasos de plástico. Cuenta con una cadena de producción en la que intervienen 30 máquinas. Los datos de producción de los últimos años son:

Año	1990	1991	1992	1993	1994	1995
P_t	67	83	92	74	37	91
H_t	4.5	3.7	4.2	4.3	2.9	4.4

donde P_t es la cantidad producida, en millones de unidades, y H_t es el tiempo anual de funcionamiento de la cadena productiva, en miles de horas. Para explicar la producción propone el siguiente modelo:

$$P_t = \alpha + \beta H_t + u_t \qquad t = 1990, ..., 1995 \tag{1}$$

1. Estima el modelo propuesto por Mínimos Cuadrados Ordinarios.

2. Interpreta los coeficientes del modelo (1).

3. Calcula el coeficiente de determinación e interprétalo.

4. ¿Crees que el tiempo de funcionamiento de las máquinas influye en la cantidad producida? Haz el contraste oportuno.

5. El director de producción solicita a la empresa de maquinaria mecánica MAMECA un estudio para la modernización de su cadena de producción. Esta empresa le propone el alquiler de una nueva máquina, cuya incorporación a la cadena productiva se espera aumente la producción hasta los 135 millones de unidades si la cadena funciona 12 horas diarias durante 360 días al año.

   a) ¿Crees que la incorporación de la máquina supondrá un aumento significativo de la producción?

   b) Si el precio por unidad vendida es de 1.5 pts (durante todos los años) y el precio de alquiler de la máquina es de 85 millones de pts ¿Recomendarías el alquiler de esta máquina? ¿Por qué? (Supón que la depreciación física es nula)

6. En 1993 se sustituyo parte de la plantilla, jubilados anticipadamente, por trabajadores jóvenes sin experiencia. Propón un modelo que recoja que la producción ha podido cambiar debido a este hecho.

LADE-1997.5 (Jun-1997)

Sea el modelo:

$$Y_t = \gamma_0 + \gamma_1 Z_t + \gamma_2 X_t + u_t \qquad t = 1, \ldots, T$$

donde las variables explicativas no son estocásticas y $u_t \sim NID(0, \sigma_u^2)$.

1. ¿Como estimarías el modelo si $Z_t = X_t$?

2. ¿Cambiaría tu respuesta si $\gamma_1 = \gamma_2$?

LADE-1997.6 (Jun-1997)

Para la estimación del parámetro asociado a Z_t en el modelo $Y_t = \beta Z_t + u_t$ donde $u_t \sim NID(0, \sigma_u^2)$, y Z_t es una variable no estocástica, se han propuesto dos estimadores:

- $\hat{\beta} = \frac{\sum Y_t}{\sum Z_t}$

- $\tilde{\beta} = \frac{Y_1}{2} + \frac{Y_T}{2}$

1. Obtén la media y la varianza de cada uno de los estimadores.
2. Deriva el estimador de MCO de β y obtén su varianza.
3. ¿Es preferible este estimador, el de MCO, a los dos anteriores? ¿Por qué?

LE/LADE-1998.1 (Feb-1998)

El gerente del Museo de Bellas Artes de Bilbao quiere analizar el número de visitantes del museo, (Y). Opina que las variables de interés son el precio de la entrada al museo (X_1), y los gastos realizados en publicidad (X_2). Dispone de observaciones mensuales desde marzo de 1996 hasta diciembre de 1997. A partir de la siguiente información muestral:

$$\begin{array}{lll} \sum X_{1t} = 1{,}02 & \sum X_{1t}^2 = 0{,}1068 & \sum Y_t X_{1t} = 9{,}26 \\ \sum Y_t = 93 & \sum Y_t^2 = 953 & \sum Y_t X_{2t} = 38{,}2 \\ \sum X_{2t} = 3{,}7 & \sum X_{2t}^2 = 1{,}77 & \sum X_{1t} X_{2t} = 0{,}383 \end{array}$$

$$(X'X)^{-1} = \begin{bmatrix} 1{,}30 & -9{,}4 & -0{,}64 \\ -9{,}4 & 100{,}0 & -2{,}40 \\ -0{,}64 & -2{,}4 & 2{,}40 \end{bmatrix}$$

1. Propón un modelo que permita recoger la opinión del gerente. Estímalo por MCO.
2. Interpreta los coeficientes estimados del modelo.
3. Calcula el coeficiente de determinación e interprétalo.
4. ¿Crees que los gastos en publicidad es una variable relevante para explicar el número de visitantes? Haz el contraste oportuno.
5. ¿Cuál es el intervalo del número esperado de visitantes que recibirá el museo si el precio de la entrada es 0.25 y los gastos en publicidad 0.5?
6. En una reunión del Consejo de Administración del museo, un consejero sugiere que no se ha tenido en cuenta que en septiembre de 1997 se inauguró el Guggenheim, y que ésto ha podido influir en el número de visitantes del museo. Para recogerlo se ha estimado el siguiente modelo:
$$\hat{Y}_t = 10{,}2 - 41{,}5X_{1t} + 11{,}01X_{2t} + 0{,}34D_t$$
donde D_t es una variable ficticia que toma el valor 1 si la observación es posterior a septiembre de 1997 y 0 en caso contrario. El coeficiente R^2 es 0,8975.
    a) Analiza si este hecho ha producido un cambio importante en el número de visitantes.
    b) ¿En qué medida afecta este resultado a las conclusiones obtenidas en los apartados a) y d)?

[0] CVS Id: $Id: 98e1.tex,v 1.1.1.1 2003/09/17 14:40:56 etpdihei Exp

LE/LADE-1998.2 (Feb-1998)

Considera el siguiente modelo :

$$Y_t = \alpha + \beta X_t + \gamma Z_t + u_t \qquad t = 1, .., T$$

1. Explica qué problemas se plantean y cómo los tratarías en los siguientes casos:
   a) $\beta + \gamma = 1$
   b) $X_t + Z_t = 1$
   c) Al regresar X_t sobre Z_t obtienes un $R^2 = 0{,}9897$.
2. ¿Qué estadístico utilizarías para contrastar las siguientes hipótesis?
   a) $H_o : \frac{\beta}{\gamma} = 1$
   b) $H_o : \beta = 8 + 5\gamma$ **y** $3\beta + \gamma = 0$
3. Si Z_t es una variable significativa y la eliminas de la regresión, ¿qué propiedades tienen los estimadores de MCO?

LE/LADE-1998.3 (Jun-1998)

Una compañía aérea tiene el siguiente modelo para analizar la evolución del número de pasajeros transportados en el periodo 1991-1997:

$$Y_t = \alpha + \beta X_t + u_t \qquad t = 1991, ..., 1997$$

donde Y_t es el número de pasajeros transportados en el año t, X_t el precio medio del billete en pesetas constantes de 1991 y u_t satisface las hipótesis habituales.

Se pide, con los datos siguientes (en decenas de miles):

Y_t	5	4	5	6	5	6	7
X_t	7	6	7	4	5	4	2

1. Estimar α y β por MCO.
2. Calcular la estimación por intervalo, al 95 % de confianza, de la pendiente β e interpretar el resultado.
3. Contrastar la hipótesis nula $\mathbf{H_o} : \alpha = \mathbf{10}\beta$ al nivel de significación del 5 %.
4. Calcular e interpretar el coeficiente de determinación R^2.
5. En el pasado, ha resultado estadísticamente significativa la inclusión de la variable Z_t, que representa el precio de los carburantes, en modelos de líneas aéreas como el indicado. ¿Cómo crees que puede afectar esto a las estimaciones obtenidas en (a)? ¿y a lo obtenido en (b)? Demuéstralo.

6. Si el precio del billete X_t estuviese determinado principalmente (por ejemplo, en un 80 %) por el precio del carburante Z_t ¿qué otras consecuencias crees que podría tener la inclusión de Z_t? Razónalo.

LE/LADE-1998.4 (Jun-1998)

Una firma automovilística ha recogido información sobre 62 compradores de vehículos suyos. En concreto, sabe el precio pagado en cientos de dólares (P), la renta anual del comprador en **cientos de dólares** (R), su sexo ($D = 1$ si es hombre, $D = 0$ si es mujer) y si tiene o no educación superior ($E = 1$ si tiene educación superior y 0 en caso contrario). Con esta información, se propone el modelo

$$P_i = \beta_1 + \beta_2 R_i + \beta_3 D_i + \beta_4 E_i + u_i. \tag{1}$$

Parte de la información muestral es:

$$P = \begin{bmatrix} 136 \\ 105 \\ 140 \\ 180 \\ \vdots \\ 98 \end{bmatrix} \qquad X = \begin{bmatrix} 1 & 400 & 1 & 0 \\ 1 & 165 & 0 & 1 \\ 1 & 180 & 1 & 1 \\ 1 & 420 & 1 & 1 \\ \vdots & \vdots & \vdots & \vdots \\ 1 & 440 & 1 & 0 \end{bmatrix}$$

$$(X'X)^{-1} = \begin{bmatrix} 0{,}109564 & -0{,}000115 & -0{,}0353 & -0{,}0268 \\ -0{,}000115 & 0{,}000001 & -0{,}000115 & -0{,}000091 \\ -0{,}0353 & -0{,}000115 & 0{,}102446 & 0{,}023971 \\ -0{,}0268 & -0{,}000091 & 0{,}023971 & 0{,}083184 \end{bmatrix} \qquad X'P = \begin{bmatrix} 9558 \\ 4880937 \\ 7396 \\ 6552 \end{bmatrix}$$

1. Interpreta los coeficientes del modelo.
2. El tercer individuo de la muestra ¿qué precio ha pagado por el coche, cuál ha sido su renta y qué sexo y nivel educativo tiene?
3. Si $\widehat{\sigma}_u^2 = 30106$ ¿cuál es la desviación típica estimada de $\widehat{\beta}_3$?
4. ¿Cuál es, **según el modelo 1**, el precio medio pagado por una **compradora** con una renta anual de 80000 dólares que no tiene educación superior?
5. **Dada la información muestral**, ¿a qué es igual la media del precio pagado por los compradores, (tanto hombres como mujeres)? Si en la muestra hay 40 hombres, ¿cuál es la media del precio pagado por los hombres?
6. ¿Qué consecuencias tendría para la estimación MCO que observásemos que **ninguna** mujer tiene educación superior y **todos** los hombres sí la tienen?

LE/LADE-1998.5 (Jun-1998)

Demuestra que si se estima por MCO

$$Y_i = \alpha + \beta X_i + v_i \qquad i = 1, ..., N$$

cuando el **verdadero** modelo es $Y_i = \alpha + u_i$, entonces $\widehat{\alpha}$ y $\widehat{\beta}$ son estimadores insesgados de los verdaderos parámetros poblacionales.

LE/LADE-1999.1 (Feb-1999)

La Comunidad Económica Europea ha decidido analizar el sector lácteo con objetivo de revisar los límites máximos de producción permitidos a cada país. Los técnicos encargados del estudio proponen el siguiente modelo

$$Q_t = \beta_1 + \beta_2 L_t + \beta_3 P_t + u_t \quad t = 1994, \ldots, 1998 \tag{1}$$

donde

Q_t: es la cantidad de leche producida en el momento t por las distintas explotaciones (en millones de litros).

L_t: es el precio de venta del litro de leche en el momento t (en pesetas constantes).

P_t: es el precio por kilogramo de pienso (en pesetas constantes).

Si se dispone de la siguiente información muestral

Q_t	L_t	P_t
30	30	50
10	10	40
80	50	60
30	20	40
50	40	60

$\sum L_t = 150$ $\quad \sum L_t^2 = 5500$

$\sum P_t L_t = 8100$ $\quad \sum P_t = 250$

$\sum P_t^2 = 12900$ $\quad \sum Q_t L_t = 7600$

$\sum Q_t = 200$ $\quad \sum Q_t^2 = 10800$

$\sum Q_t P_t = 10900$

$$(X'X)^{-1} = \begin{bmatrix} 26,7 & 0,45 & -0,8 \\ 0,45 & 0,01 & -0,015 \\ -0,8 & -0,015 & 0,025 \end{bmatrix}$$

a) Explica en qué consiste el criterio mínimo cuadrático ordinario y desarróllalo obteniendo las ecuaciones normales.

b) Estima el modelo empleando el método de mínimos cuadrados ordinarios. ¿Tienen los coeficientes los signos esperados? Explica por qué.

c) Calcula el coeficiente de determinación. Interpreta el resultado.

d) ¿Cuál es el valor de la covarianza estimada entre $\hat{\beta}_2$ y $\hat{\beta}_3$?, ¿y de la varianza estimada de $\hat{\beta}_3$?

e) ¿Crees que el precio de la leche es una variable relevante?

[0] CVS Id: $Id: 99e1.tex,v 1.2 2003/09/18 07:47:11 etpdihei Exp

f) Contrasta la hipótesis de que los precios influyen en la cantidad producida de leche en la misma cuantía, pero en sentido opuesto.

g) Si en el año 1999 el precio de la leche se mantiene igual al del año anterior y el precio del pienso aumenta en 10 unidades, ¿crees razonable una producción de leche de 100 millones de litros?

LE/LADE-1999.2 (Feb-1999)

Un profesor de COU quiere analizar la nota que los estudiantes obtienen en el exámen de selectividad, Y. Cree que las variables explicativas relevantes son, la nota media que obtienen en el curso de COU (en décimas por encima de cinco), X, si han cursado sus estudios en un centro privado o en uno público y el idioma (castellano o euskera) en el que se realiza el examen.

a) Especifica el modelo que recoja la opinión de este profesor y explica detalladamente el significado de los coeficientes.

b) ¿Cómo contrastarías que sólo la nota de COU influye en la nota de selectividad?

c) Si en el contraste anterior no rechazaras la hipótesis nula, ¿qué consecuencias tiene este hecho sobre las propiedades de los estimadores MCO de los coeficientes del modelo del apartado a)?

d) Si la nota media de todos los estudiantes que se presentan al examen es de 6, ¿crees que la estimación del modelo presentaría algún problema?

LE/LADE-1999.3 (Feb-1999)

Un investigador está interesado en conocer la respuesta del consumo a cambios en la renta, para los consumidores vascos. Para obtener la estimación de la misma se ha propuesto el siguiente modelo:

$$C_t = \beta_0 + \beta_1 R_t + \beta_2 RI_t + \beta_3 I_t + u_t \quad t = 1, \ldots, T$$

donde C representa el consumo privado, R la renta de los consumidores vascos, RI la riqueza que poseen e I el tipo de interés a corto plazo. El modelo presentado ha sido estimado con datos de la CAV para el periodo 1980-1997, los resultados obtenidos son los siguientes:

$$\hat{C}_t = \underset{(15.1)}{120{,}1} + \underset{(0.20)}{0{,}27}\, R_t + \underset{(0.02)}{0{,}012}\, RI_t - \underset{(0.97)}{1{,}12}\, I_t \quad R^2 = 0{,}963 \tag{2}$$

a) Analiza el modelo propuesto así como los resultados obtenidos. ¿Crees que $0{,}27$ es una estimación precisa del parámetro asociado a la variable renta?, ¿en qué te basas para responder?

b) Se han estimado también los siguientes modelos:

$$\hat{C}_t = \underset{(12.7)}{114{,}3} + \underset{(0.16)}{0{,}29}\, R_t + \underset{(0.008)}{0{,}015}\, RI_t \quad R^2 = 0{,}951 \tag{3}$$

$$\hat{C}_t = \underset{(12.7)}{109{,}2} + \underset{(0.01)}{0{,}023} RI_t - \underset{(0.98)}{1{,}07}\, I_t \quad R^2 = 0{,}902 \tag{4}$$

$$\hat{C}_t = \underset{(11.9)}{97{,}4} + \underset{(0.11)}{0{,}25}\, R_t - \underset{(0.99)}{0{,}97}\, I_t \quad R^2 = 0{,}947 \tag{5}$$

Compara entre sí los tres modelos. ¿Cuál te parece mejor? ¿Por qué?

c) Compara el modelo (2) con el (3) y el (5). Te proporcionan tres estimaciones distintas para el coeficiente de la renta, 0,27, 0,29 y 0,25. Explica cuál de las tres elegirías razonando cuidadosamenta tu respuesta.

LE-1999.1 (Jun-99)

Una importante asociación de joyeros quiere analizar sus ventas (V) en función de la renta (R) y el precio del oro (AU):

$$V_t = \beta_1 + \beta_2 R_t + \beta_3 AU_t + u_t \quad t = 1, \ldots, 10 \tag{1}$$

Si se dispone de la siguiente información muestral

$$(X'X)^{-1} = \begin{bmatrix} 0,5 & -0,5 & -0,5 \\ -0,5 & 0,7 & 0,5 \\ -0,5 & 0,5 & 0,834 \end{bmatrix} \quad X'V = \begin{bmatrix} 18 \\ 16 \\ 1 \end{bmatrix} \quad \sum (V_t - \bar{V})^2 = 27,6$$

1. Estima el modelo por MCO.
2. Interpreta los coeficientes. ¿Tienen los signos esperados?
3. Estima las varianzas de los coeficientes estimados.
4. Comenta la bondad del modelo. ¿Es la variable renta significativa?
5. Dada la información muestral disponible, ¿podrías verificar si $\beta_3 = -0,5$ es un valor posible?
6. Si sabemos que el modelo que **realmente** determina las ventas es

$$V_t = \beta_1 + \beta_2 R_t + \beta_3 AU_t + \beta_4 AG_t + u_t \tag{2}$$

donde AG_t es el precio de la plata en el momento t.

   a) ¿Podrías esperar que el R^2 del primer modelo es menor que el R^2 del segundo?
   b) ¿Sucede lo mismo con los coeficientes de determinación corregidos?
   c) Si solamente estuvieras interesado en estimar el coeficiente β_1, ¿podrías basarte en el modelo (1) y evitar las operaciones adicionales que supone estimar el modelo (2)?

[0] CVS Id: $Id: 99e1g.tex,v 1.2 2003/09/18 07:47:11 etpdihei Exp

LE-1999.2 (Jun-99)

Una familia quiere analizar su consumo en función de la renta

$$C_t = \alpha + \beta R_t + u_t \quad u_t \sim N(0, \sigma^2) \tag{3}$$

Con una muestra que comprende los años de convivencia (desde que se casaron hasta hoy en dia) estiman el modelo por MCO obteniendo el siguiente resultado:

$$\hat{C}_t = 202,23 + 0,895 R_t \quad t = 1981, \ldots 1999 \quad R^2 = 0{,}9974 \sum_{1981}^{1999} c_t^2 = 6403269,6$$

Sin embargo, dado que en 1988 tuvieron a su primer hijo, sospechan que ésto pudo influir en su función de consumo.

1. ¿Qué modelo propondrías de forma que se tenga en cuenta el nacimiento del niño? Interpreta los coeficientes del modelo que has propuesto.
2. Posteriormente consideran que lo más apropiado es separar los periodos 1981-1987 y 1988-1999 y estimar su consumo de forma separada. Los resultados obtenidos son

$$\hat{C}_t = 337,24 + 0,8846 R_t \quad t = 1981, \ldots 1987 \quad R^2 = 0{,}9954 \sum_{1981}^{1987} \hat{u}_t^2 = 3679,020 \tag{4}$$

$$\hat{C}_t = 1416,21 + 0,3663 R_t \quad t = 1988, \ldots 1999 \quad R^2 = 0{,}9121 \sum_{1988}^{1999} \hat{u}_t^2 = 2379,315 \tag{5}$$

   - Si el modelo que has propuesto en el apartado (1) no es equivalente a estas dos ecuaciones, propón uno que lo sea. Interpretalo.
   - Comprueba si el nacimiento del hijo ha tenido efecto en su función de consumo.

3. Si la renta de la familia hubiese permanecido estable durante el periodo muestral, ¿hubieras podido obtener los resultados anteriores?, ¿por qué?

LADE-1999.1 (Jun-1999)

Las cifras que aparecen en la tabla corresponden a las pesadas de un cerdo en Kg. efectuadas a intervalos semanales.

Semana t=	1	2	3	4	5	6	7	8	9
Peso Y_t=	21,77	24,50	27,22	30,39	34,48	39,01	42,64	47,18	50,81
Semana t=	10	11	12	13	14	15	16	17	
Peso Y_t=	56,26	60,79	65,33	71,68	77,13	82,12	87,11	92,55	

[0] CVS Id: $Id: 99e1e.tex,v 1.1.1.1 2003/09/17 14:40:56 etpdihei Exp

Al ajustar una curva de crecimiento polinomial de segundo grado mediante mínimos cuadrados ordinarios se ha obtenido:
$$\widehat{Y}_t = 18,14 + 2,88 \cdot t + 0,09 \cdot t^2$$

1. Teniendo en cuenta que
$$\begin{pmatrix} 17 & \sum_{t=1}^{17} t & \sum_{t=1}^{17} t^2 \\ & \sum_{t=1}^{17} t^2 & \sum_{t=1}^{17} t^3 \\ & & \sum_{t=1}^{17} t^4 \end{pmatrix}^{-1} = \begin{pmatrix} 17 & 153 & 1{,}785 \\ 153 & 1{,}785 & 23{,}409 \\ 1{,}785 & 23{,}409 & 327{,}369 \end{pmatrix}^{-1} =$$
$$\begin{pmatrix} 0,67647 & -0,15441 & 0,00735 \\ -0,15441 & 0,04424 & -0,00232 \\ 0,00735 & -0,00232 & 0,00013 \end{pmatrix} \quad \text{y} \quad \begin{pmatrix} \sum_{t=1}^{17} Y_t \\ \sum_{t=1}^{17} t \cdot Y_t \\ \sum_{t=1}^{17} t^2 \cdot Y_t \end{pmatrix} = \begin{pmatrix} 910,97 \\ 10{,}039,48 \\ 129{,}486,46 \end{pmatrix}$$
¿Qué porcentaje de la variación en el peso se puede explicar mediante el polinomio temporal?

2. A un 95 % de confianza, ¿se puede considerar que las 2 primeras potencias de la variable tiempo son conjuntamente relevantes para explicar el peso?

3. Contrasta la igualdad de los coeficientes de t y t^2.

4. Con un nivel de confianza del 95 % ¿entre qué valores estará el peso del cerdo en la semana 18?

5. Como se desea presentar un informe en EE.UU. es necesario expresar el peso en libras (1kg=2,204lb). Al cambiar las unidades de peso a libras, ¿qué consecuencias tendrá esto sobre las estimaciones MCO de los coeficientes de regresión?

LADE-1999.2 (Jun-1999)

Con el fin de estudiar la discriminación salarial en EE.UU., se plantea el siguiente modelo de regresión:
$$W_i = \beta_1 + \beta_2 S_i + \beta_3 A_i + \beta_4 E_i + \beta_5 H_i + \beta_6 O_i + u_i \qquad i = 1, \ldots, 206 \qquad u \sim \mathcal{N}(0, \sigma_u^2 I)$$

$$\begin{aligned} W_i &= \text{salario cobrado por el individuo i, en dólares por hora} \\ S_i &= \text{1 si el indiv. i es mujer y 0 si es hombre} \\ A_i &= \text{años de educación del indiv. i} \\ E_i &= \text{edad del indiv. i} \\ H_i &= \text{1 si el indiv. i es hispano y 0 en caso contrario} \\ O_i &= \text{1 si el indiv. i no es ni blanco ni hispano y 0 en caso contrario} \end{aligned} \qquad (1)$$

El modelo estimado por MCO es:
$$\hat{W}_i = \underset{(-3,38)}{-6,41} - \underset{(-4,61)}{2,76}\, S_i + \underset{(8,54)}{0,99}\, A_i + \underset{(4,63)}{0,12}\, E_i + +\underset{(0,22)}{0,24}\, H_i - \underset{(-1,07)}{1,06}\, O_i \qquad R^2 = 0,367$$

1. Explica qué significado tienen los coeficientes β_1, β_5 y β_6 de este modelo.

2. El salario por hora medio estimado para los hombres es de 6,87 dólares. ¿Cuál es el de las mujeres? (Explica cómo lo has calculado).
3. En vista de los resultados, ¿hay discriminación salarial **en contra** de las mujeres?
4. Un observador externo asegura que el salario por hora no depende del conjunto de las variables explicativas propuestas. ¿Estás de acuerdo con él? Explica en qué te basas.

LADE-1999.3 (Jun-1999)

Suponer que se ha estimado una ecuación de demanda de alimentos con una muestra de las observaciones correspondientes a 100 familias como la siguiente:

$$\widehat{\ln q} = 2 - 0,33 \ln p + 0,76 \ln y \qquad R_1^2 = 0,91$$

siendo q la cantidad, p el precio e y la renta de cada familia.

Con una muestra (independiente de la anterior) con datos de otras 100 familias, se estima:

$$\widehat{\ln q} = 1,52 - 0,47 \ln p + 0,35 \ln y \qquad R_2^2 = 0,98$$

La varianza muestral de $\ln q$ en ambos modelos es 624,2 y 736,8 respectivamente.

Para la muestra completa n=200 se estima:

$$\widehat{\ln q} = 2,1 - 0,28 \ln p + 0,6 \ln y \qquad R^2 = 0,93$$

siendo la varianza muestral de $\ln q = 800$.

¿Es estable la ecuación de demanda de alimentos en ambas submuestras?

LADE-1999.4 (Jun-1999)

1. Escribe el criterio que seguirías si tuvieras que elegir entre un estimador sesgado y otro insesgado y explica por qué se suele utilizar ese criterio.
2. Razonar y responder si son ciertas o falsas las siguientes afirmaciones:
   a) Los parámetros del modelo $Y_t = A \cdot X_t^B \cdot u_t$ no pueden estimarse por MCO, por tratarse de un modelo no lineal.
   b) En un MRLG es cierto que $E(u_t) = \frac{1}{T}\sum_{t=1}^{T} u_t = 0$.

[0] CVS Id: $Id: 00e1g.tex,v 1.1.1.1 2003/09/17 14:40:56 etpdihei Exp

LE-2000.1 (Feb-2000)

Comenta y justifica si las siguientes afirmaciones son ciertas o falsas (analiza cada afirmación independientemente del resto):

1. En la estimación por Mínimos Cuadrados Ordinarios de

$$Y_t = \alpha + \beta X_t + u_t \qquad t = 1, ..., T$$

   a) Para obtener la estimación del coeficiente β solamente hace falta conocer el coeficiente de correlación simple entre X e Y.

   b) Cuanto mayor sea $\sum_{t=1}^{T}(X_t - \bar{X})^2$ mayor será la significatividad de X en la regresión.

   c) Si llamamos $\hat{u}_t$ al residuo mínimo cuadrático en t, puede suceder en este caso que $\sum_{t=1}^{T}\hat{u}_t$ sea distinta de 0.

2. En la estimación por Mínimos Cuadrados Ordinarios de

$$Y_t = \alpha + \gamma X_t + \delta Z_t + u_t \qquad t = 1, ..., T$$

   a) Estimaré con mayor precisión el coeficiente δ si la correlación entre las variables explicativas X y Z es menor.

   b) Puedo contrastar la hipótesis de que $\delta = 0$ utilizando los coeficientes de determinación de las regresiones estimadas con y sin la variable Z en el modelo.

   c) Dado que el imponer la restricción $\delta = \gamma$ empeora el ajuste, **no se gana nada** al estimar los parámetros imponiendo esta restricción.

LE-2000.2 (Feb-2000)

Se quiere estudiar la relación de los años de educación de los jóvenes (E) con la renta familiar (R) y la procedencia geográfica, Rural (RUR) o Urbana (URB).

Se consideran las siguientes especificaciones para el modelo:

$$E_t = \beta_1 + \beta_2 URB_t + \beta_3 R_t + u_t \tag{1}$$

$$E_t = \gamma_1 URB_t + \gamma_2 RUR_t + \gamma_3 R_t + u_t \tag{2}$$

para t=1,...,T. (Si te resulta más cómodo, puedes cambiar los subíndices a i=1,..., N).

1. Con objeto de acabar de definir el modelo
   a) Define las variables URB y RUR para que puedas utilizarlas en el modelo. ¿Qué valores toman?
   b) Para el primero de los modelos propuestos, escribe la matrices X, $X'X$ y $X'Y$.
2. Para poder estimar este modelo se cuenta con la siguiente información muestral correspondiente al modelo (1):

$$(X'X) = \begin{bmatrix} 8 & 4 & 68 \\ 4 & 4 & 42 \\ 68 & 42 & 652 \end{bmatrix} \qquad X'Y = \begin{bmatrix} 113 \\ 66 \\ 1024 \end{bmatrix}$$

   Basándote en la respuesta al apartado (b) anterior, responde a las siguientes preguntas:
   a) ¿Cuántos individuos tienes en la muestra? ¿Por qué?
   b) ¿Cuantos individuos pertenecen a la zona urbana? ¿Por qué?
   c) ¿Cuál es el valor de la media muestral de la Renta familiar? ¿Por qué?
3. Con los datos que se te han facilitado arriba más los que se añaden ahora

$$(X'X)^{-1} = \begin{bmatrix} 1,256 & 0,369 & -0,15476 \\ 0,369 & 0,881 & -0,0952 \\ -0,15476 & -0,0952 & 0,0238 \end{bmatrix} \qquad \begin{array}{l} \sum_{t=1}^{T}(Y_t - \bar{Y})^2 = 64,875 \\ \sum_{t=1}^{T} Y_t^2 = 1661 \end{array}$$

   a) Estima los coeficientes β_1, β_2 y β_3 en el modelo (1).
   b) Contrasta en el modelo (1) la hipótesis de que la procedencia geográfica no es un factor relevante en la determinación de los años de educación (escribe la hipótesis nula y la alternativa, el estadístico de contraste y la región crítica, además del resultado del contraste).
   c) Contrasta en el modelo (1) la hipótesis de que ni la renta ni la procedencia geográfica tienen efecto sobre los años de educación.
   d) Predice con el modelo (1) los años de educación para un joven cuya familia tiene una renta de $10,5$ y vive en zona urbana. Calcula esta predicción por punto y por intervalo.

4. Refiriéndote ahora al modelo (2)

   a) Utilizando las estimaciones anteriores de β_1, β_2 y β_3, obtén las estimaciones de γ_1, γ_2 y γ_3 en el modelo (2), razonando la respuesta. Interpreta los coeficientes de este modelo.

   b) Escribe la hipótesis nula, la alternativa y el estadístico de contraste para contrastar en el modelo (2) que la procedencia geográfica no es un factor relevante para explicar los años de educación.

LE-2000.3 (Feb-2000)

En la estimación del modelo

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

utilizando una muestra de 15 observaciones, se han obtenido los siguientes resultados:

$$\underset{\text{(desv.típica)}}{\widehat{Y}_t} = \underset{(13{,}82)}{-27,5} + \underset{(0{,}0594)}{0,033}\, X_{1t} + \underset{(5{,}65)}{10,35}\, X_{2t} \underset{(4{,}6)}{-0,127}\, X_{3t} \tag{4}$$

$$R^2 = 0,814 \qquad \bar{R}^2 = 0,764 \qquad \sum_{t=1}^{15} \hat{u}_t^2 = 950,53$$

Dado que la variable X_3 resulta claramente no significativa se ha eliminado del modelo, reestimándolo con el siguiente resultado:

$$\underset{\text{(desv.típica)}}{\widehat{Y}_t} = \underset{(13{,}15)}{-27,7} + \underset{(0{,}0074)}{0,034}\, X_{1t} + \underset{(3{,}88)}{10,00}\, X_{2t} \tag{5}$$

$$R^2 = 0,812 \qquad \bar{R}^2 = 0,771 \qquad \sum_{t=1}^{15} \hat{u}_t^2 = 952,89$$

1. Valora los resultados de ambas estimaciones. ¿Cómo decidirías si eliges el modelo (4) o el (5)? Explica con claridad tus razones.

2. Si has elegido el modelo (5) qué sucedería con tu decision si te dijeran que el modelo correcto es el (4)? ¿Por qué?

3. Si te aseguran que el modelo (4) es el correcto ¿qué razón podría existir para haber obtenido los resultados anteriores?

LE-2000.4 (Jun-2000)

Se ha estimado el modelo siguiente que relaciona el logaritmo de los salarios (Y_t) con el logaritmo de la productividad del trabajo (X_{1t}) y el logaritmo de la inflación (X_{2t}):

$$\widehat{Y}_t = 4 + 0,4X_{1t} + 0,9X_{2t} \qquad (1)$$

$$R^2 = 0,134 \qquad \sum_{t=1}^{29} \hat{u}_t^2 = 530 \qquad T = 29$$

$$\sum_{t=1}^{29} X_{1t} = \sum_{t=1}^{29} X_{2t} = 0 \qquad \sum_{t=1}^{29} X_{1t}^2 = 50 \qquad \sum_{t=1}^{29} X_{1t}X_{2t} = 10 \qquad \sum_{t=1}^{29} X_{2t}^2 = 50$$

1. Utilizando los datos que se proporcionan, contrasta la hipótesis de que las dos pendientes suman 1.
2. Explica **detalladament**e cuál es la forma más eficiente de estimar el modelo, dados los resultados obtenidos en el apartado a).
3. Interpreta el resultado en términos de tu modelo.

LE-2000.5 (Jun-2000)

En una regresión con cuatro variables explicativas (además de un término constante) y tamaño muestral igual a 30, se obtuvo un $R^2 = 0,750$. Al incluir una nueva variable explicativa y reestimar el modelo con las cinco variables, se obtuvo un $R^2 = 0,757$.

1. Contrasta la hipótesis nula de que la nueva variable explicativa no es relevante. Utiliza el nivel de significación del $5\,\%$.
2. Utilizando el coeficiente de determinación, ¿qué modelo se ajusta mejor a los datos? ¿Por qué?
3. Calcula, para los dos casos, el coeficiente de determinación corregido. Utilizando este coeficiente ¿qué modelo se ajusta mejor a los datos? ¿Por qué?
4. Utilizando el resultado del contraste, ¿qué modelo crees está mejor especificado? ¿Por qué?
5. ¿Existe una contradicción entre las respuestas que has dado a los apartados anteriores? Razona la respuesta basándote en el estadístico utilizado en el contraste.

LE-2000.6 (Jun-2000)

Se dispone de los siguientes datos relativos a diez personas, aseguradas en distintas compañías de seguros, sobre la prima pagada por el seguro a todo riesgo de accidentes de tráfico y su sexo (M=Mujer, H=Hombre)

Prima en miles de ptas	59	60	20	15	48	36	30	35	22	43
Sexo	H	H	M	M	H	H	M	H	M	H

1. Especifica un modelo de regresión para la prima de seguros en función del sexo.
2. Interpreta los coeficientes del modelo de regresión propuesto y calcula sus estimaciones por el método de Mínimos Cuadrados Ordinarios.
3. Dibuja la nube de puntos de los datos sobre la prima del seguro e indica gráficamente qué te recoge la regresión estimada en b).
4. Contrasta si en media la prima del seguro voluntario de accidentes pagada por una mujer es igual a la de un hombre, indicando la hipótesis nula y la alternativa. Especifica y muestra todos los cálculos que sean necesarios además de todos los supuestos para que el contraste sea válido.
5. Obtén una predicción por intervalo del 95 % de confianza para la prima de seguro pagada por un hombre.

LADE-2000.1 (Feb-2000)

Para analizar el número de accidentes de tráfico en un país europeo, se propone el siguiente modelo:

$$\mathrm{Y}_t = \beta_1 + \beta_2 X_{2t} + \beta_3 X_{3t} + u_t \qquad t = 1967, \ldots, 1999,$$

donde Y_t es el número de accidentes de tráfico (en decenas de miles), X_{2t} es la velocidad media (en miles de kms./hora) y X_{3t} es la tasa de alcoholemia permitida.

Con la siguiente información muestral:

$$X'X = \begin{bmatrix} 33 & 1 & 2 \\ 1 & 4 & 5 \\ 2 & 5 & 12 \end{bmatrix} \quad X'Y = \begin{bmatrix} 132 \\ 42 \\ 80 \end{bmatrix} \quad \sum (Y_t - \bar{Y})^2 = 480.$$

1. Obtén la función de regresión muestral.
2. Estima la desviación típica de $\widehat{\beta}_2$ y contrasta la hipótesis de que la velocidad media no es una variable significativa.
3. Haz el contraste de nuevo, paso a paso, utilizando esta vez el estadístico de las sumas de cuadrados residuales.
4. Si en el año 2000 la velocidad media fuera de 0,1 (miles de kms./hora) y la tasa de alcoholemia 0,2 ¿cuál crees que sería el número medio de accidentes? Calcula un intervalo de confianza del 95 %.

[0] CVS Id: $Id: 00e1e.tex,v 1.1.1.1 2003/09/17 14:40:56 etpdihei Exp

LADE-2000.2 (Feb-2000)

El Gobierno Vasco decide emprender una política de concesión de subvenciones a las empresas en dificultades para conseguir incrementar el número de empleados en estas empresas. Un asesor del Gobierno Vasco trata de analizar si realmente la política es eficaz y para ello lleva a cabo un estudio de los determinantes del empleo en estas empresas, concluyendo que depende de los siguientes factores:

- La cuantía de la subvención pública, X_i
- La localización geográfica de la empresa (urbana o rural)
- El grado de cualificación del empresario (sin estudios, estudios medios o estudios superiores).

1. Especifica un modelo que recoja la relación entre el número de empleados de cada empresa y los factores que en él inciden. Interpreta los coeficientes.
2. ¿Cómo contrastarías la hipótesis de que el grado de cualificación del empresario afecta al empleo en esas empresas?
3. ¿Tendría alguna incidencia sobre la estimación del modelo el hecho de que todos los empresarios considerados tuvieran estudios superiores?

LADE-2000.3 (Feb-2000)

La relación $Y_i = \beta_1 + \beta_2 X_{2i} + \beta_3 X_{3i} + u_i \quad (1)$ se ha estimado por MCO con 33 observaciones. Algunos datos muestrales son:

$$(X'X)^{-1} = \begin{bmatrix} 1,2 & 0,3 & 0,1 \\ 0,3 & 0,6 & -0,2 \\ 0,1 & -0,2 & 0,8 \end{bmatrix} \qquad \begin{array}{c} \sum Y_i = 30 \\ \sum Y_i X_{2i} = -30 \\ \sum Y_i X_{3i} = 30 \end{array}$$

1. ¿Hay indicios de multicolinealidad exacta? Razona la respuesta.
2. Imagina que sobre el modelo del enunciado realizas el contraste de significación conjunta y obtienes que el valor del estadístico de contraste es 45. Calcula los coeficientes R^2 y $\bar{R}^2$ utilizando esta información, el número de observaciones y el número de coeficientes estimados (y nada más: NO puedes usar el valor de $(X'X)^{-1}$ ni de $X'Y$).
3. Si en realidad el modelo inicial está mal especificado debido a que el modelo correcto es $Y_i = \beta_1 + \beta_3 X_{3i} + v_i$, ¿es fiable el contraste de significatividad conjunta que se hizo en (b)? Razona la respuesta.

LADE-2000.4 (Jun-2000)

Se quiere analizar las ventas del producto de una determinada empresa (Y) en función de las horas trabajadas (X_2) y del precio de dicho producto (X_3). Se dispone de datos trimestrales de 1989 a 1996, y se especifica el siguiente modelo:

$$\text{Y}_t = \beta_1 + \beta_2 X_{2t} + \beta_3 X_{3t} + u_t \qquad t = 1, \ldots, 32.$$

Con la siguiente información muestral:

$$\begin{array}{lll}
\sum_t Y_t = 344 & \sum_t Y_t X_{2t} = 2939 & \sum_t X_{2t}^2 = 456 \\
\sum_t X_{2t} = 56 & \sum_t Y_t X_{3t} = 1713 & \sum_t X_{3t}^2 = 224 \\
\sum_t X_{3t} = 40 & \sum_t X_{3t} X_{2t} = 292 & \sum_t Y_t^2 = 19824
\end{array}$$

1. Estima el modelo por MCO e interpreta los coeficientes estimados.
2. ¿Cuál es la varianza estimada de $\widehat{\beta}_2$? ¿Y la covarianza estimada entre $\widehat{\beta}_2$ y $\widehat{\beta}_3$? ¿crees que estas covarianzas estimadas coincidirán con las teóricas?
3. Contrasta la hipótesis nula de que un aumento unitario en el precio del producto disminuye las ventas en ocho unidades.
4. El 1 de enero de 1993 se introdujo en la empresa un máquina de tecnología punta y se cree que las ventas fueron mayores a partir de ese año. Por ello se especifica el siguiente modelo,

   $$\text{Y}_t = \alpha_1 D_{1t} + \alpha_2 D_{2t} + \beta_2 X_{2t} + \beta_3 X_{3t} + u_t \qquad t = 1, \ldots, 32,$$

   donde D_{1t} toma valor 1 si estamos en $1989 - 1992$ y D_{2t} toma valor 1 en $1993 - 1996$. Sabemos además que:

   $$\sum Y_t D_{1t} = 102 \qquad \sum Y_t D_{2t} = 242$$

   a) ¿Cómo contrastarías si la introducción de la máquina en la empresa es realmente significativa?

   b) **Dada la información muestral**, ¿cuáles son las ventas medias totales? ¿Y las ventas medias de los cuatro primeros años?

LADE-2000.5 (Jun-2000)

Considera la siguiente función de producción Cobb-Douglas:

$$\text{Q}_t = A L_t^{\beta_2} K_t^{\beta_3} e^{u_t} \qquad t = 1987, \ldots 1998$$

donde Q_t es la producción, L_t es el factor trabajo y K_t el factor capital. Los resultados de la estimación son:

variable	coef.	$\widehat{\text{desv}}$
constante	3,4202522	4,4296310
$\ln(L_t)$	0,4256658	0,6352997
$\ln(K_t)$	0,2497068	0,1246899

$R^2 = 0{,}965608$

$\sum_t \hat{u}_t^2 = 0{,}001970$

1. Comprueba la posible existencia de multicolinealidad imperfecta.
2. Con el fin de contrastar la posible existencia de rendimientos constantes a escala ($\beta_2 + \beta_3 = 1$) se estima por MCO un modelo restringido cuya suma de cuadrados de los residuos resulta ser 0,002058
   a) Escribe el modelo restringido.
   b) Contrasta la existencia de rendimientos constantes a escala.
   c) A la vista de los resultados, ¿qué consecuencias tendrá sobre las propiedades de los estimadores imponer la restricción?, ¿y sobre el problema de multicolinealidad?

<u>LADE-2000.6</u> (Jun-2000)

Considera la siguiente relación:

$$Y_t = \alpha + \beta W_t + \gamma Z_t + u_t \quad t = 1, \ldots, T$$

donde $u_t \sim N(0, \sigma_u^2)$ y $\sum w_t z_t = 0$. *(NOTA: $w_t = W_t - \bar{W}$, $y_t = Y_t - \bar{Y}$ y $z_t = Z_t - \bar{Z}$)*

1. Demuestra que los estimadores MCO de β y γ son:

$$\hat{\beta} = \frac{\sum w_t y_t}{\sum w_t^2} \qquad \hat{\gamma} = \frac{\sum z_t y_t}{\sum z_t^2}$$

2. Si W_t y Z_t son variables relevantes, y en la estimación del modelo se omite Z_t, ¿es el estimador MCO de β insesgado? Demuéstralo.

<u>LE-2001.1</u> (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:

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

$\sum_{t=1}^{19} P_t = 2526,99$ $\quad \sum_{t=1}^{19} E_t = 150,01$ $\quad \sum_{t=1}^{19} T_t = 132,26$
$\sum_{t=1}^{19} P_t^2 = 348471,62$ $\quad \sum_{t=1}^{19} E_t^2 = 1529,31$ $\quad \sum_{t=1}^{19} T_t^2 = 1096,59$
$\sum_{t=1}^{19} E_tT_t = 833,61$ $\quad \sum_{t=1}^{19} E_tP_t = 18712,38$ $\quad \sum_{t=1}^{19} T_tP_t = 17901,32$

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

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.

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):

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

   ¿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:

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

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:

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

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

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:

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

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

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

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

   .

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

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

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

   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:

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

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

$$R^2 = 0,7279$$

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.

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.

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

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

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

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

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

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:

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

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

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

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.

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

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

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?

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

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

   ?

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

3. ¿Qué interpretación tienen los parámetros en el modelo

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

   ?

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

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.

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,

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

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.

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

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

4) Contrasta la significatividad conjunta de las variables.

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

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

   a) Contrasta si la restricción es compatible con los datos.
   b) Si has aceptado la restricción ¿querrías reformular el modelo original? ¿Por qué?
   c) A efectos del modelo ¿es equivalente la restricción anterior a la igualdad $S_i + 20P_i = 0$?

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.

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?

<u>LE-2002.5</u> (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.

   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.

<u>LADE-2002.1</u> (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:

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

Los resultados de la estimación se resumen a continuación

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

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:

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

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

Otro directivo estima los tres siguientes modelos:

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

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

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

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).

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.

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

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

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?

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).

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:

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

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.

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).

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

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?

   a.2) Incluir ambas variables explicativas.

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

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

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

   ¿Crees que puede existir algún problema de especificación ?

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:

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

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

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.

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

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

d) ¿Son las variables explicativas conjuntamente significativas?

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

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)?

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

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

siendo en este caso la matriz inversa igual a

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

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.

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

b) Interpreta los coeficientes del modelo (3).

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

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

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?

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.

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

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

d) ¿Son las variables explicativas conjuntamente significativas?

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

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

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)?

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.

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

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

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.

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?

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

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

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

[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:

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

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.

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

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

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?

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.

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?

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?

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

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:

$$\hat{Y}_i = \underset{(1,04)}{9,00} - \underset{(1,163)}{2,00}\, H_i + \underset{(1,401)}{4,00}\, ES_i + \underset{(1,401)}{13,50}\, EU_i \tag{1}$$
$$R^2 = 0,9147 \qquad \sum \hat{u}_i^2 = 39,00$$

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

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

$$\hat{Y}_i = \underset{(1,03)}{10,00} - \underset{(1,636)}{4,50}\, H_i + \underset{(1,636)}{3,00}\, ES_i + \underset{(1,636)}{11,00}\, EU_i + \underset{(2,427)}{2,50}\, (H_i ES_i) + \underset{(2,427)}{5,50}\, (H_i EU_i)$$
$$R^2 = 0,95079 \qquad \sum \hat{u}_i^2 = 22,50 \tag{2}$$

d) ¿Cuál es la diferencia fundamental entre las especificaciones de los modelos (1) y (2)? Escribe los datos de las variables H_iES_i y H_iEU_i.

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?

f) Contrasta al nivel de significación del 5 % si 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.

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.

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.

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?

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

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

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?

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.

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?

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).

b) Interpreta los coeficientes del modelo.

c) Contrasta la significatividad individual de las variables explicativas.

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

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.

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

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

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

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

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

d) Contrasta al nivel de significación del $5\,\%$ que las variables 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.

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.

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.

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?

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

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.

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.

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

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

d) ¿Son las variables explicativas conjuntamente significativas?

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

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.

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.

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.

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

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?

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

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

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

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

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.

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?

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?

LADE-2004.1 (Feb-2004)

Considera el modelo de regresión:

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

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

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

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

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

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

$\sum_{t=1}^{10} x_{2t}x_{3t} = 125{,}5$

1) Comprueba que la Función de Regresión Muestral es:

$$\hat{Y}_t = 0,5023 - 0,0039X_{2t} + 0,9959X_{3t} \qquad (2)$$

[0] CVS Id: $Id: ladefeb04.tex,v 1.1 2004/07/21 09:54:56 etpdihei Exp

2) Calcula una medida de la bondad del ajuste realizado, empleando los coeficientes estimados indicados en (2). Interpreta el resultado.

3) Contrasta la significatividad individual y conjunta de las variables explicativas X_2 y X_3.

4) Dados los resultados de los contrastes anteriores, ¿podemos pensar en la existencia de un problema de multicolinealidad? Razona tu respuesta.

5) ¿Crees posible que un aumento unitario en X_3 haga aumentar la media poblacional de Y en una unidad? Realiza el contraste oportuno.

6) Si impones que $\beta_3 = 1$:

   6.1) ¿Cuál es el modelo restringido?

   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).

   6.3) ¿**Cómo estimarías** la matriz de varianzas y covarianzas de los estimadores restringidos de β_1 y β_2?

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.

2) Explica **cómo contrastarías** la hipótesis de que ser fumador 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.

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)?

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?

LADE-2004.3 (Feb-2004)

Considera el siguiente modelo de regresión

$$Y_t = \beta X_t + u_t \qquad u_t \sim N(0, \sigma^2) \tag{3}$$

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.

2) ¿Son ambos estimadores insesgados?

3) Obtén la varianza de $\hat{\beta}^*$.

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.

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:

$$\begin{array}{lll} \sum_{i=1}^{10} lnY_i = 11{,}8 & \sum_{i=1}^{10} lnY_i lnX_{1i} = 7{,}1 & \sum_{i=1}^{10} (lnY_i)^2 = 19{,}34 \\ \sum_{i=1}^{10} lnX_{1i} = 2 & \sum_{i=1}^{10} lnY_i lnX_{2i} = 4{,}1 & \sum_{i=1}^{10} (lnX_{1i})^2 = 7 \\ \sum_{i=1}^{10} lnX_{2i} = 2 & \sum_{i=1}^{10} lnX_{1i} lnX_{2i} = 1 & \sum_{i=1}^{10} (lnX_{2i})^2 = 7 \end{array}$$

$$\begin{array}{ll} \sum_{i=1}^{10} (lnY_i - lnX_{1i})(lnX_{2i} - lnX_{1i}) = 3 & \sum_{i=1}^{10} (lnY_i - lnX_{2i})(lnX_{1i} - lnX_{2i}) = 9 \\ \sum_{i=1}^{10} (lnX_{2i} - lnX_{1i})^2 = 12 & \sum_{i=1}^{10} (lnX_{1i} - lnX_{2i})^2 = 12 \\ \sum_{i=1}^{10} (lnY_i - lnX_{1i})^2 = 12{,}14 & \sum_{i=1}^{10} (lnY_i - lnX_{2i})^2 = 18{,}14 \end{array}$$

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.

[0] CVS Id: $Id: ladejun04.tex,v 1.1 2004/07/21 09:54:56 etpdihei Exp

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:

$$\frac{\widehat{\beta} - \widehat{\gamma} - 7}{\sqrt{\widehat{var}(\widehat{\beta} - \widehat{\gamma} - 7)}} \sim t_{(T-K)} \tag{2}$$

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?

$$\widehat{\beta}_1 = \frac{\sum^{N_M} S_i}{N_M} = \widehat{\alpha}_1 + \widehat{\alpha}_2$$

$$\widehat{\beta}_2 = \frac{\sum^{N_T} S_i}{N_T} = \widehat{\alpha}_1$$

<u>LADE-2004.6</u> (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:

$$\widehat{V}_t = \underset{(7{,}94)}{5,76} + \underset{(4{,}02)}{3,8}\ RM_t - \underset{(-0{,}035)}{1,93}\ P_t \quad R^2 = 0,85367 \tag{5}$$

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:

$$\widehat{V}_t = \underset{(7{,}88)}{5,78} + \underset{(5{,}61)}{4,3}\ RM_t + \underset{(3{,}22)}{2,6}\ G_t - \underset{(-3{,}15)}{3,9}\ P_t \quad R^2 = 0,91121 \tag{6}$$

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)?

<u>LE-2005.1</u> (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:

$$(X'X) = \begin{bmatrix} 100 & 200,48 & 17,97 \\ 200,48 & 468,15 & 35,54 \\ 17,97 & 35,54 & 3,6 \end{bmatrix} \qquad (X'X)^{-1} = \begin{bmatrix} 0,1736 & -0,0342 & -0,5294 \\ -0,0342 & 0,0152 & 0,0200 \\ -0,5294 & 0,0200 & 2,7232 \end{bmatrix}$$

$$\sum Y_i = 3125 \qquad \sum Y_i X_{2i} = 7092 \qquad \sum Y_i^2 = 112262,89 \qquad \sum Y_i X_{3i} = 551$$

1. Interpreta los parámetros del modelo propuesto. ¿Qué signos esperarías que tuvieran?

2. Escribe la función a minimizar que se corresponde con el criterio mínimo cuadrático ordinario.

3. Estima el modelo propuesto por el método de mínimos cuadrados ordinarios.

4. Calcula una medida de la bondad del ajuste e interprétala.

5. Estima la matriz de varianzas y covarianzas del estimador de los coeficientes del modelo.

6. Contrasta la significatividad individual y conjunta de las variables explicativas.

7. Contrasta $H_0 : \beta_2 = -\beta_3$. A la vista de los resultados del contraste, ¿qué concluyes sobre la especificación del modelo?

8. Estima el modelo restringido correspondiente a la hipótesis nula anterior.

   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:

   $$\underset{\text{(estadístico-t)}}{\hat{Y}_i} = \underset{(0,1508)}{0,7911} + \underset{(10,8581)}{12,4656} X_{2i} - \underset{(-8,3215)}{12,7423} X_{3i} + 0,1829 X_{4i} \qquad R^2 = 0,7158$$

9. ¿Por qué aconsejarías especificar el modelo (2) en vez del (1)? Razona tu respuesta.

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?

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.

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 Y_i: Ahorro familiar

1. ¿Es el modelo (4) estimable? ¿Por qué? Propón un modelo estimable e interpreta sus parámetros.
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: $Y_i = \alpha_1 D_{1i} + \gamma_1 F_{1i} + \gamma_2 F_{2i} + \beta X_i + u_i$.
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.

LE-2005.1 (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:

$$Q_t = A L_t^{\beta_1} K_t^{\beta_2} e^{u_t} \tag{1}$$

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?
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.

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

$\sum Q_t^* = 200,9780$ $\qquad \sum Q_t^* L_t^* = 1170,6730$

$\sum Q_t^* K_t^* = 1514,5429$ $\qquad \sum Q_t^{*2} = 1593,6446$

3. Estima la ecuación (2) por el método de mínimos cuadrados ordinarios.
4. En la ecuación (2)
   a) Calcula el valor del residuo correspondiente al año 1979.
   b) ¿Cuánto vale $\sum_{t=1}^{T} \hat{u}_t$? ¿Por qué?
5. Calcula una medida de la bondad del ajuste e interprétala,
6. Contrasta la significatividad individual y conjunta de las variables explicativas.
7. ¿Hay evidencia de algún problema muestral? Interpreta los resultados de los contrastes anteriores
8. ¿Tiene la empresa rendimientos constantes a escala?($H_0 : \beta_2 + \beta_3 = 1$)
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?

LE-2005.2 (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,

11. ¿Cómo estimarías el modelo (3) si resultara que $2K_t^* = T_t^*$? ¿qué propiedades tiene tu estimador?

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?

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)?

LE-2005.3 (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,

2. ¿Es la experiencia un factor determinante para 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.

4. Interpreta los resultados obtenidos en los apartados anteriores, ¿Qué puedes decir respecto de la relación entre experiencia y salarios en esa empresa?

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:

$$\underset{}{\widehat{Y}_i} = \underset{(16{,}576)}{0{,}0309} - \underset{(-3{,}553)}{0{,}025}\ X_i + \underset{(30{,}050)}{0{,}154}\ X_i^2$$

$$N = 50 \qquad R^2 = 0{,}9937 \qquad \widehat{\sigma}^2 = 0{,}0000277 \qquad \hat{u}'\hat{u} = 0{,}001303$$

$$(X'X) = \begin{pmatrix} 50 & 23{,}08 & 16{,}545 \\ & 16{,}545 & 15{,}298 \\ & & 16{,}481 \end{pmatrix} \qquad (X'X)^{-1} = \begin{pmatrix} 0{,}125 & -0{,}413 & 0{,}257 \\ & 1{,}789 & -1{,}245 \\ & & 0{,}957 \end{pmatrix}$$

$$X'Y = \begin{pmatrix} 3{,}532 \\ 2{,}669 \\ 2{,}681 \end{pmatrix}$$

1) Interpretar la medida de bondad del ajuste.

2) ¿Es posible que un estado con una renta per capita igual a 8267$ realice un gasto por alumno de 440$?

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.

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?

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.

2. ¿Es el sexo es una variable relevante a la hora de explicar la nota obtenida?

   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.
4. ¿Es relevante la nueva variable añadida?
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.
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

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.

   (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.

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?

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).

iv) ¿Cuál de los tres modelos es el apropiado?, ¿cuáles son las propiedades de los estimadores del modelo elegido?

LADE-2005.1 (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.

b) Interpreta los coeficientes β_2 y β_3 del modelo.

c) Estima el modelo por el método de mínimos cuadrados ordinarios. ¿Te parecen los signos razonables?, ¿por qué?

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

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.

f) Estima el modelo restringido correspondiente a la hipótesis nula anterior.

g) Teniendo en cuenta los resultados del contraste del apartado e), ¿qué propiedades tiene el estimador empleado en el apartado c)?

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:

$$\hat{Y}_t = \underset{(1{,}1901)}{4,0518} + \underset{(28{,}6606)}{13,8581}\,X_{2t} - \underset{(-11{,}8459)}{7,5349}\,X_{3t} + \underset{(0{,}0364)}{0,2922}\,X_{4t} \qquad (2)$$

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.

LADE-2005.2 (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:

$$\hat{Y}_i = \underset{(14{,}06)}{1{,}7932} + \underset{(9{,}60)}{0{,}0709}\,X_i \qquad (3)$$

$$R^2 = 0{,}8602 \qquad \sum_i \hat{u}_i^2 = 1{,}2086$$

$$\hat{Y}_i = \underset{(10{,}84)}{2{,}6733}\,H_i + \underset{(11{,}45)}{2{,}9950}\,M_i \qquad (4)$$

$$R^2 = 0{,}0506 \qquad \sum_i \hat{u}_i^2 = 8{,}2100$$

$$\hat{Y}_i = \underset{(12{,}17)}{1{,}7618}\,H_i + \underset{(11{,}52)}{1{,}8556}\,M_i + \underset{(9{,}16)}{0{,}0701}\,X_i \qquad (5)$$

$$R^2 = 0{,}8644 \qquad \sum_i \hat{u}_i^2 = 1{,}1725$$

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.

b) ¿Qué problema se habría producido en los contrastes anteriores si la muestra hubiera estado formada por 17 hombres?

c) Otro estudiante le sugiere que no ha tenido en cuenta la siguiente estimación:

$$\hat{Y}_i = \underset{(11{,}52)}{1{,}8556} \underset{(-0{,}65)}{-0{,}0937} H_i + \underset{(9{,}16)}{0{,}0701} X_i \tag{6}$$

$$R^2 = 0{,}8644 \qquad \textstyle\sum_i \hat{u}_i^2 = 1{,}1725$$

   i) Explica detalladamente a este segundo estudiante que el primero **sí ha tenido en cuenta** dicha estimación.

   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).

LADE-2005.3 (Jun-2005)

El gerente de una estación invernal de esquí del Pirineo, con datos anuales desde 1991 hasta 2003 de las variables:

E_t = Número de esquiadores por año, en miles, en el año t
P_t = Precio del forfait en € constantes de 1991
F_t = Precio medio del forfait en estaciones del Pirineo francés en € constantes de 1991
PIB_t = Tasa de crecimiento del PIB en el año t respecto al año anterior
S_t = % de días soleados en la temporada de esquí del año t

ha obtenido las siguientes estimaciones MCO:

$$\widehat{E}_t = \underset{(106{,}47)}{1029{,}57} + \underset{(2{,}22)}{0{,}91}\, P_t + \underset{(1{,}50)}{2{,}41}\, PIB_t - \underset{(-3{,}55)}{0{,}64}\, S_t \tag{7}$$

$$R^2 = 0{,}680$$

$$\widehat{E}_t = \underset{(99{,}05)}{1026{,}22} + \underset{(2{,}36)}{0{,}97}\, F_t + \underset{(1{,}47)}{2{,}30}\, PIB_t - \underset{(-3{,}55)}{0{,}64}\, S_t \tag{8}$$

$$R^2 = 0{,}695$$

$$\widehat{E}_t = \underset{(68{,}44)}{1018{,}38} - \underset{(-0{,}75)}{2{,}94}\, P_t + \underset{(0{,}99)}{3{,}99}\, F_t + \underset{(1{,}18)}{1{,}96}\, PIB_t - \underset{(-3{,}44)}{0{,}62}\, S_t \tag{9}$$

$$R^2 = 0{,}715$$

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?

b) Si ahora además se tiene en cuenta el siguiente modelo estimado:

$$\widehat{E}_t = \underset{(86,32)}{984{,}10} + \underset{(0,43)}{1{,}01}\, P_t - \underset{(-0,22)}{0{,}55}\, F_t + \underset{(2,39)}{2{,}20}\, PIB_t + \underset{(3,26)}{1{,}96}\, S_t - \underset{(-4,36)}{0{,}03}\, S_t^2 \tag{10}$$

$$R^2 = 0{,}9233$$

¿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.