LADE/LADE-DER
CURSO 2008/2009
EXAMEN DE ESTADÍSTICA II
9 de septiembre de 2009
PROBLEMAS (tiempo: 1 hora 40 minutos)

PROBLEMA 1: Con el objeto de analizar el efecto del gasto en publicidad (x_1) y el gasto en personal (x_2) sobre las ventas de un determinado producto (y), se ha obtenido una muestra de 30 marcas comerciales que fabrican dicho producto. Se ha planteado estimar un modelo de regresión donde todas las variables están medidas en miles de euros. Los resultados obtenidos para el modelo estimado en base a la muestra son los siguientes

$$\widehat{y} = 52.98 + 4.114x_1 + 4.739x_2$$

$$\overline{R}^2 = 0.4879;\ \ \widehat{S}_R = 7.1831;\ \ \widehat{S}_R^2(X'X)^{-1} = \begin{pmatrix} 208.466 & 23.2200 & -6.001 \\ 23.220 & 2.8912 & -0.215 \\ -6.001 & -0.2150 & 1.044 \end{pmatrix}$$

a. Calcula el efecto que tendría sobre las ventas un aumento de 1000 euros en el gasto de publicidad. (0.75 puntos).

b. Señala qué variables son individualmente significativas al 5 %. (0.25 puntos).

c. Una nueva marca sale al mercado con un gasto en publicidad de 3000 euros y gasto en personal de 15000. Calcule un intervalo al 95 % para el valor medio de las ventas de la nueva marca. (0.75 puntos).

d. Construye la tabla ADEVA y efectua el contraste de regresión al 5 %, interpretando el resultado. (0.75 puntos).

Solución

a)$\widehat{y}_i = 52.98 + 4.114x_1 + 4.739x_2$ Cuando el gasto en publicidad aumenta en 1000 euros, manteniendo constante el gasto en personal el volumen de ventas aumenta en media 4114 euros.

b)

$$H_0 = \beta_i = 0; H_1 = \beta_i \neq 0.$$

$$t_i = \frac{\hat{\beta}_i}{\sqrt{\hat{S}^2_R (X`X)^{-1}_{i+1,i+1}}}$$

$$t_1 = 2.41; t_2 = 4.63$$

Los dos coeficientes son significativas al 5 %.

c) $\widehat{y}_i = 52.98 + 4.114 * 3000 + 4.739 * 15000 = 83480$

Intervalo al 95 % para el valor medio de las ventas

$$\hat{y} \pm t_{27}^{0.025} \sqrt{\hat{S}^2_R x'_h (X'X)^{-1} x_h} = 83480 \pm 2.05 * 15541$$

$$[51620, 115339]$$

$$\hat{S}^2_R x'(X'X)^{-1} = [\ 1 \quad 3000 \quad 15000\ ] \begin{pmatrix} 208.466 & 23.22 & -6.001 \\ 23.22 & 2.8912 & -0.215 \\ -6.001 & -0.215 & 1.044 \end{pmatrix} [\ 1 \quad 3000 \quad 15000\ ]'$$

$$\sqrt{\hat{S}^2_R x'(X'X)^{-1}} = 15541; t^0.0025_{27} = 2.05;$$

d)

Fuente de variación	Suma de cuadrados	Grados de libertad	Varianzas	Estadístico F
V. Explicada	1528.46	2	764.23	14.81
V. No explicada	1393.2	27	51.60	
V. Total	2921.66	29		

PROBLEMA 2: Se les pidió a cuatro analistas financieros predecir el aumento en los ingresos, en el siguiente año para tres compañías petroleras. Sus predic-

ciones, como incremento porcentual en los ingresos se dan en la siguiente tabla, junto con información de utilidad:

	Analista A	Analista B	Analista C	Analista D	Suma Filas
Compañía 1	9.0	8.7	9.3	8.5	35.5
Compañía 2	8.6	7.9	8.7	8.3	33.5
Compañía 3	7.6	7.4	7.7	7.3	30.0
Suma Columnas	25.2	24.0	25.7	24.1	99

a. Escribe una ecuación de un modelo adecuado para estudiar el efecto "compañia" y el efecto "analista". Estime los 7 efectos. (0.9 puntos).

b. Copia la siguiente tabla en su hoja de respuestas y rellene las partes faltantes marcadas con "...". Note que la suma de cuadrados total está dada. (1 punto).

Fuente de Variación	SC	gl	Varianza
Compañia	...	...	...
Analista	...	...	...
Residual	...	...	...
Total	4.73	...	

c. Usa el apartado b) para analizar la hipótesis nula (al 5 %) de que las predicciones del crecimiento medio poblacional son las mismas para todas las compañías petroleras. (0.2 puntos).

d. ¿Hay efecto analista? Use el apartado b) para probar la hipótesis nula (al 5 %) de que las predicciones del crecimiento medio poblacional son las mismas para todos los analistas. (0.2 puntos).

e. En base en tu respuesta al apartado d), ¿será apropiado ignorar la variable bloque analista? Expliquelo brevemente. (0.2 puntos).

Solución

a) [0.9pts](0.1 por ecuación y 0.1 por cada uno de los 8 valores)

$$y = \mu + \alpha_i + \beta_j + \epsilon_{ij},$$

donde y = 'predicción', μ media, $\alpha_i = \mu_i - \mu$ (efecto fila), $\beta_j = \mu_j - \mu$ (efecto columna), $\sum_i \alpha_i = \sum_j \beta_j = 0$, ϵ_{ij} $iid \sim N(0, \sigma^2)$, $i = 1, 2, 3 = I$, $j = 1, 2, 3, 4 = J$, $n = 12$. La media estimada es $\bar{y}_{\bullet\bullet} = 99/12 = 8.25$

	A	B	C	D	Sumas Filas	Media por filas	$\hat{\alpha}$
compañia 1	9.0	8.7	9.3	8.5	35.5	8.875	0.625
compañia 2	8.6	7.9	8.7	8.3	33.5	8.375	0.125
compañia 3	7.6	7.4	7.7	7.3	30.0	7.500	-0.750
Sumas Columnas	25.2	24.0	25.7	24.1	99		
Media por columnas	8.400	8.000	8.567	8.033			
$\hat{\beta}$	0.150	-0.250	0.317	-0.217			

b) [1pts](0.05 por cada valor en las columnas 'gl' y 'Varianza'; 0.3 por cada uno de los dos valores en la columna 'SC' and 0.05 para el resto)

$$\begin{aligned} SS_\alpha &= J\sum_{i=1}^{I}\hat{\alpha}_i^2 \\ &= 4(0.969) = 3.876 \\ SS_\beta &= I\sum_{j=1}^{J}\hat{\beta}_j^2 \\ &= 3(0.233) = 0.699 \\ SS_R &= SS_T - (SS_\alpha + SS_\beta) \\ &= 0.155 \end{aligned}$$

Fuente de Variación	SC	gl	Varianza
Compañia	3.876	2	1.938
Analista	0.699	3	0.233
Residual	0.155	6	0.026
Total	4.730	11	

c) [0.2pts](0.05 por H_0 y H_1, 0.05 por F y 0.05 por decision y 0.05 por conclusion)

$$H_0:\ \alpha_1 = \alpha_2 = \alpha_3 = 0 \quad \text{vs} \quad H_1:\ \text{at least one } \alpha_i \neq 0$$

$$F = \frac{1.938}{0.026} = 74.538 > F_{2,6;0.05} = 5.1433$$

Por lo tanto, se rechaza la hipótesis nula y se concluye que al menos una de las compañias de petroleo tiene una predicción de la media diferente a las otras.

d) [0.2pts](como en c))

$$H_0:\ \beta_1 = \ldots = \beta_4 = 0 \quad \text{vs} \quad H_1:\ \text{at least one } \beta_j \neq 0$$

$$F = \frac{0.233}{0.026} = 8.962 > F_{3,6;0.05} = 4.7571$$

Por lo tanto, se rechaza la hipótesis nula y se concluye que al menos uno de los analistas da, en promedio, predicciones diferentes

e) [0.2pts] No, porque la variable bloque es significativa, nos permite comparar grupos mas homogéneos y reducir la variabilidad no explicada.