

## Estadística II — Ejercicios del Tema 5, 2010/11

**Ejercicio 1.** Considera los cuatro conjuntos de datos dados en las transparencias del Tema 5 (sección 5.1)

- Comprueba que los cuatro conjuntos de datos dan lugar a la misma recta de regresión.
- Aplica los métodos de diagnóstico comentados en clase al conjunto de datos # 1, y comenta los resultados.
- Aplica los métodos de diagnóstico comentados en clase al conjunto de datos # 2, y comenta los resultados.
- Aplica los métodos de diagnóstico comentados en clase al conjunto de datos # 3, y comenta los resultados.
- En el conjunto de datos # 3, identifica el dato atípico. Obtén la recta de regresión tras eliminar este dato atípico, y comenta el resultado.

**Ejercicio 2.** A partir de una muestra de 30 observaciones, se estimó el modelo de regresión lineal simple  $\hat{y}_i = \hat{\beta}_0 + \hat{\beta}_1 x_i$ , con  $\hat{\beta}_0 = 10,1$  y  $\hat{\beta}_1 = 8,4$ . La variación cuadrática de la respuesta debida al modelo es  $\sum_i (\hat{y}_i - \bar{y})^2 = 128$ , mientras que la variación cuadrática residual de la respuesta es  $\sum_i e_i^2 = 286$ .

- Calcula e interpreta el coeficiente de determinación.
- ¿Qué puedes decir sobre el coeficiente de correlación entre las  $x_i$  y las  $y_i$ ?
- Construye la tabla ANOVA correspondiente a partir de estos datos.
- Contrasta a un nivel de significación del 5% la hipótesis de que la respuesta  $y$  no depende de  $x$ . Repite el contraste a un nivel de significación del 1%.
- Da un estimador insesgado de la varianza del error.

**Ejercicio 3.** El gestor de un concesionario de automóviles está interesado en la relación entre el número de vendedores que trabajan en fin de semana y el número de coches vendidos. Se obtuvieron los siguientes datos para seis fines de semana consecutivos:

$x_i$ (# de vendedores)	$y_i$ (# de coches vendidos)	
1	5	22
2	7	20
3	4	15
4	2	9
5	4	17
6	8	25

- Determina la recta de regresión de  $y$  (# de coches vendidos) sobre  $x$  (# de vendedores).
- Construye la tabla ANOVA y comprueba la validez de la descomposición ANOVA  $SCT = SCM + SCR$ .
- Calcula e interpreta el coeficiente de determinación.
- Utiliza la tabla ANOVA para contrastar, al 1% y al 5% de significación, la hipótesis de que el número de vendedores no influye en las ventas.
- Realiza los contrastes del apartado (d) mediante el método visto en el Tema 4. Comprueba que el estadístico  $T$  de tal contraste y el estadístico  $F$  del contraste del apartado (d) cumplen la relación  $F = T^2$ .

**Ejercicio 4.** Linealiza las siguientes relaciones no lineales, aplicando las transformaciones vistas en clase:

(a)  $y = \ln(5\sqrt{x})$ .

(b)  $y = \frac{2}{3}8^x$ .

(c)  $y = 1/(4-x)$ .

(d)  $y = \frac{5}{4}\sqrt{x}$ .

**Ejercicio 5.** Supongamos que se han obtenido las siguientes observaciones para una variable respuesta  $y$  en función de la variable explicativa  $x$ :

$x_i$	$y_i$
1	5,47
2	7,54
3	9,13
4	10,47
5	11,65
6	12,72

- (a) Dibuja el gráfico de puntos  $(x_i, y_i)$ . ¿Parece adecuada una recta para describir la relación entre los datos?
- (b) Suponiendo que el modelo correcto sea de la forma  $y = ax^b + u$ , lleva a cabo las transformaciones adecuadas de las variables  $x$  e  $y$ , y estima los parámetros  $a$  y  $b$  a partir de una regresión lineal en las variables transformadas.
- (c) Construye la tabla ANOVA para las variables transformadas, y calcula e interpreta el coeficiente de determinación.

**Ejercicio 6.** Para el conjunto de datos # 1 de los considerados en el ejercicio 1, calcula los estimadores de mínimos cuadrados de los coeficientes de regresión lineal empleando la formulación matricial.

**Ejercicio 7.** Un análisis de regresión lineal múltiple a partir de  $n = 34$  observaciones proporciona el modelo estimado  $\hat{y} = 2,50 + 6,8x_1 + 6,9x_2 - 7,2x_3$ . Los errores estándar de los coeficientes estimados de las variables explicativas son  $s(\hat{\beta}_1) = 3,1$ ,  $s(\hat{\beta}_2) = 3,7$  y  $s(\hat{\beta}_3) = 3,2$ . El coeficiente de determinación obtenido es  $R^2 = 0,85$ .

- (a) Calcula intervalos de confianza al 95 % para los coeficientes de las variables explicativas.
- (b) Para cada variable explicativa, contrasta al 5 % de significación la hipótesis de que la respuesta no depende de dicha variable.
- (c) Para cada variable explicativa, ¿existe evidencia significativa al 1 % de que el coeficiente correspondiente es positivo?

**Ejercicio 8.** Supongamos que has estimado los coeficientes de un modelo de regresión lineal múltiple  $y_i = \beta_0 + \beta_1x_1 + \dots + \beta_kx_k + u_i$ . Contrasta al 5 % de significación la hipótesis de que la respuesta no depende de las variables explicativas a partir de las siguientes tablas ANOVA parciales:

	Fuente de variación	SC	G.L.	Media	Cociente F
(a)	Modelo	4500	3		
	Residuos/Errores	500	26		
	Total				

	Fuente de variación	SC	G.L.	Media	Cociente F
(b)	Modelo	9780	6		
	Residuos/Errores	2100	32		
	Total				

	Fuente de variación	SC	G.L.	Media	Cociente F
(c)	Modelo	460000	8		
	Residuos/Errores	25000	27		
	Total				

**Ejercicio 9.** Tenemos los siguientes datos de 10 viviendas unifamiliares, para las que se ha registrado el precio (en M€), la superficie (en  $m^2$ ), la superficie del terreno (en Has.), y el número de cuartos de baño:

	precio (M€)	superficie ( $m^2$ )	superf. terreno (Has.)	# baños
	170	120,90	0,10	1
	177	134,85	0,12	1,5
	191	148,80	0,12	2
	194	172,05	0,18	2
	202	195,30	0,16	2
	210	186,00	0,16	2,5
	214	195,30	0,20	2
	228	223,20	0,20	2,5
	240	251,10	0,20	2,5
	252	241,80	0,28	3

Más abajo se dan los resultados (obtenidos con Statgraphics) de un análisis de regresión lineal múltiple de  $y$  (precio) sobre  $x_1$  (superficie),  $x_2$  (superficie del terreno), y  $x_3$  (# de baños).

- Calcula intervalos de confianza al 95 % para los coeficientes del modelo de regresión  $y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + u$ .
- Contrasta al 5 % y al 10 % de significación la hipótesis de que la respuesta no depende de la variable  $x_j$ , para  $j = 1, 2, 3$ .
- Calcula e interpreta el coeficiente de determinación múltiple. Estima la desviación típica del error.

#### Multiple Regression Analysis

Dependent variable: precio

Parameter	Estimate	Standard Error	T Statistic	P-Value
CONSTANT	100,985	7,86246	12,844	0,0000
superfi	0,354243	0,0975193	3,63255	0,0109
superfTerreno	109,115	73,4594	1,48537	0,1880
WCs	10,3945	6,86311	1,51454	0,1807

#### Analysis of Variance

Source	Sum of Squares	Df	Mean Square	F-Ratio	P-Value
--------	----------------	----	-------------	---------	---------

Model	6158,96	3	2052,99	73,92	0,0000
Residual	166,635	6	27,7726		
-----					
Total (Corr.)	6325,6	9			