Estadística Aplicada – Grado en Ingeniería Matemática – Curso 2020-21

Hoja de problemas 2

1. Construye un ejemplo de datos bidimensionales $(x_i, y_i) \in \mathbb{R}^2,\ i = 1,\dots,n,$ de modo que se cumpla que $y_i = x_i^2$ para todo i y para los que su correlación sea 0.

2. Obtén mediante el método de mínimos cuadrados la expresión general de los estimadores de los parámetros β_0 y β_1 de un modelo de regresión lineal simple.

3. Las notas obtenidas por un grupo de alumnos en las asignaturas de Informática y Estadística son las siguientes:

I	3	4	6	7	5	8	7	3	5	4	8	5	5	8	8	8	5
E	5	5	8	7	7	9	10	4	7	4	10	5	7	9	10	5	7

a) Obtener la distribución marginal de las notas en ambas materias, así como su media y su varianza.

b) Obtener la covarianza y la correlación de las calificaciones de ambas materias.

c) Calcular la recta de regresión para explicar las notas de Estadística en función de las de Informática.

4. Demuestra las 6 propiedades establecidas en la Proposición 1 de los apuntes del Tema 2.

5. Demostrar que en el contexto de la regresión lineal simple se cumple $Cov[\overline{y}, \hat{\beta}_1] = 0.$

6. Un psicólogo afirma que, en base a los datos observados, el número de respuestas incorrectas dadas por un niño en un estudio experimental decrecen al aumentar la edad del niño.

Edad	2	3	4	4	5	5	6	7	7	9	9	10	11	11	12
Respuestas	11	12	10	13	11	9	10	7	12	8	7	3	6	5	5

a) ¿Existe realmente evidencia que soporte esa afirmación?

b) Alberto, de diez años y medio, va a participar en el experimento. ¿Cuántas respuestas incorrectas es posible esperar que realice?

7. Demostrar las consecuencias del supuesto S3 expuestas en los apuntes del Tema 2.

8. Demostrar que $\hat{\beta}_1 \cdot \sum (x_i - \overline{x})(y_i - \overline{y}) = MCM$.

9. Se han recogido mediciones de la densidad de oxígeno (Y, en mg/l) a diferentes profundidades (X, en metros) en el lago Worther (Austria), obteniéndose la siguiente distribución:

X	15	20	30	40	50	60	70
Y	6.5	5.6	5.4	6.0	4.6	1.4	0.1

a) Ajustar una recta a los datos por medio del método de mínimos cuadrados.

b) Analizar la correlación entre ambas variables

c) ¿Qué densidad de oxígeno es esperable a una profundidad de 90 metros?

10. Considérese la siguiente distribución conjunta:

X	Y
1	10
2	100
3	1000
4	10000

a) Ajustar una curva $Y = a\ 10^{bX}$ a los datos mediante técnicas de regresión lineal.

b) ¿Cuál es el valor predicho para $X = 3.5$? ¿Y para $X = 5$?

11. Denotemos por ρ el coeficiente de correlación poblacional entre una variable respuesta Y y una variable explicativa X. Si $\rho \neq 0$, entonces X e Y están relacionados linealmente, y un test apropiado para contrastar las hipótesis H_0: $\rho = 0$ vs. H_1: $\rho \neq 0$ viene dado por

$$T = \frac{r_{XY}\sqrt{n-2}}{\sqrt{1-r_{XY}^2}}$$

donde r_{XY} es el coeficiente de correlación muestral entre X e Y, que es un estimador de ρ. Se pide:

a) Demostrar que este test es equivalente (de hecho algebraicamente equivalente) al contraste t para las hipótesis H_0: $\beta_1 = 0$ vs. H_1: $\beta_1 \neq 0$ en un análisis de regresión simple (una pista: usar que r_{XY}^2 es igual al coeficiente de determinación, que se define como…).

b) Dar la región crítica o de rechazo en términos de T.

c) Si se ha observado $r_{XY} = -0.45$ con $n = 30$, ¿qué conclusión se obtiene cuando $\alpha = 0.05$?

12. Hallar una expresión para los elementos de las matrices $\mathbf{X}^t\mathbf{X}$ y $(\mathbf{X}^t\mathbf{X})^{-1}$ en el contexto de la regresión lineal simple.

.