

RELACIÓN DE PROBLEMAS: REGRESIÓN LINEAL SIMPLE Y MÚLTIPLE

ANÁLISIS ESTADÍSTICO MULTIVARIANTE

GRADO EN CIENCIA E INGENIERÍA DE DATOS

1. Dado el vector aleatorio (X, Y) con función de densidad

$$f(x, y) = \begin{cases} 2 & \text{si } 0 < x < 1, 0 < y < x \\ 0 & \text{en otro caso} \end{cases}$$

- a) Obtener la curva de regresión para predecir Y en función de valores de la variable X .
b) ¿Coincide con la recta de regresión?

2. Sea (X, Y) un vector aleatorio con función de densidad

$$f(x, y) = \begin{cases} \frac{3}{4} \left[xy + \frac{x^2}{2} \right] & \text{si } 0 < x < 1, 0 < y < 2 \\ 0 & \text{en otro caso} \end{cases}$$

A partir de la distribución de Y condicionada a $X = x$, obtener la curva de regresión para predecir Y a partir de valores de X y proporcionar una predicción para $X = 2/3$.

3. Sabiendo que el vector (X, Y) tiene una distribución normal con medias 1 y 2, varianzas 2 y correlación $-1/2$, calcular la curva de regresión para predecir Y a partir de valores de X y obtener una predicción para $X = 1.5$.
4. Encontrar la recta de regresión para el conjunto de datos:

$$\{(x_i, y_i)\} = \{(1, 4), (2, 2), (1, 5), (5, 3), (6, 2)\}$$

Estimar y para $x = 3$. ¿Será fiable esa aproximación?

5. El rendimiento de una reacción química depende de la concentración del reactivo y de la temperatura de la operación.

Rendimiento	Concentración	Temperatura
81	1.00	150
89	1.00	180
83	2.00	150
91	2.00	180
79	1.00	150
87	1.00	180
84	2.00	150
90	2.00	180

En el modelo de regresión del rendimiento sobre la temperatura y la concentración, los valores ajustados son

$$\hat{y} = (80.25, 87.75, 83.25, 90.75, 80.25, 87.75, 83.25, 90.75)'$$

Calcular el error cuadrático medio y el coeficiente de determinación e interpretar su valor.

6. Un modelo ajustado para predecir la extracción de manganeso en % (y) a partir del tamaño de partícula en mm (x_1), la cantidad de dióxido de azufre en múltiplos de la cantidad estequiométrica necesaria para la disolución de manganeso (x_2) y la duración de la filtración en minutos (x_3) están dadas como

$$y = 56.145 - 9.0469x_1 - 33.421x_2 + 0.243x_3 - 0.5963x_1x_2 \\ - 0.0394x_1x_3 + 0.60022x_2x_3 + 0.6901x_1^2 + 11.7244x_2^2 - 0.0097x_3^2$$

Se consideraron 27 observaciones, con $\sum_i^n (y_i - \hat{y}_i)^2 = 209.55$, $\sum_i^n (y_i - \bar{y})^2 = 6777.5$.

Se pide:

- Obtener una predicción para el porcentaje de extracción cuando el tamaño de partícula es 3 mm, la cantidad de dióxido de azufre 1.5 y la duración de la filtración es de 20 minutos.
 - ¿Es posible predecir el cambio en el porcentaje de extracción cuando la duración de la filtración aumenta en un minuto? Si la respuesta es afirmativa, encontrar el cambio pronosticado. Si la respuesta es negativa, ¿qué otra información se necesitaría para determinarlo?
 - Calcular el coeficiente de determinación R^2 . Interpretar su valor.
7. Se efectúa un estudio sobre el desgaste de un cojinete (y) y su relación con la viscosidad del aceite (x_1) y carga (x_2). Se obtienen los datos siguientes:

y	x_1	x_2
193	1.6	851
230	15.5	816
172	22.0	1058
91	43.0	1201
113	33.0	1357
125	40.0	1115

- a) El análisis de regresión lineal múltiple con los datos de la tabla anterior proporciona $\hat{\theta} = (350.994, -1.27999, -0.15390)'$ y los valores ajustados son

$$\hat{y} = (217.99, 205.69, 160.18, 111.46, 100.17, 128.51)'$$

Calcular una estimación de la varianza residual y el coeficiente de determinación R^2 .

- b) Utilizar el modelo para predecir el desgaste cuando $x_1 = 25$ y $x_2 = 1000$.
8. En un proceso industrial se sospecha que la dureza de las láminas de acero reducido en frío depende del contenido en cobre (x_1) y de la temperatura de recocido (x_2). Para comprobar esta suposición se mide en doce especímenes de láminas de acero la dureza para varios valores del contenido de cobre y de la temperatura, obteniéndose los siguientes resultados:

Dureza (Rockwell 30-T)	Contenido de cobre (%) (x_1)	Temperatura de recocido (? F) (x_2)
79.1	0.025	1000
65.3	0.025	1100
55.5	0.025	1200
56.6	0.025	1300
81.1	0.15	1000
69.9	0.15	1100
57.6	0.15	1200
55.6	0.15	1300
85.5	0.2	1000
72.0	0.2	1100
60.9	0.2	1200
59.1	0.2	1300

- Plantear el modelo de regresión lineal múltiple para explicar la dureza de la lámina en función del contenido de cobre y de la temperatura.
- Sabiendo que $\hat{\theta} = (161.404, 27.1923, -0.08547)'$, obtener una estimación para la varianza del error y el valor de R^2 . Comentar la bondad del ajuste.
- Se decide eliminar la variable x_1 del modelo y ajustar un modelo de regresión lineal simple para predecir la dureza en función de x_2 .
 - ¿Qué modelo se obtendría?
 - ¿Qué cambio se espera en la dureza promedio si la temperatura se incrementa en 100 grados Fahrenheit?