

Ejercicio (Peña, vol II, pág. 358)

La tabla siguiente proporciona la latitud en grados (L), la altura en metros (A) y la temperatura media anual (T) de seis capitales marítimas españolas (datos del INE):

| | L | A | T |
|-----------|------|----|------|
| Gijón | 43.4 | 22 | 13.9 |
| Vigo | 43.2 | 45 | 14.9 |
| Barcelona | 41.3 | 45 | 16.9 |
| Valencia | 39.5 | 24 | 17.2 |
| Almería | 36.8 | 7 | 18 |
| Cádiz | 36.5 | 30 | 18 |

Se pide:

- Construir e interpretar un modelo de regresión para explicar la temperatura en función de estas dos variables.
- Calcular R^2 y la varianza residual.
- Prever la temperatura media de Tortosa sabiendo que la latitud es 40.5 y la altitud 50 m.

Solución:

$$\text{Modelo: } T = \beta_0 + \beta_1 L + \beta_2 A + U$$

Hiperplano de regresión: $\hat{Y} = X\hat{\beta}$, donde $\hat{\beta} = (\hat{\beta}_0 \quad \hat{\beta}_1 \quad \hat{\beta}_2)'$, $Y = T$, $X = [1 \quad L \quad A]$.

$$\text{a) } Y = \begin{pmatrix} 13.9 \\ 14.9 \\ 16.4 \\ 17.2 \\ 18 \\ 18 \end{pmatrix}, \quad X = \begin{pmatrix} 1 & 43.4 & 22 \\ 1 & 43.2 & 45 \\ 1 & 41.3 & 45 \\ 1 & 39.5 & 24 \\ 1 & 36.8 & 7 \\ 1 & 36.5 & 30 \end{pmatrix}, \quad (X'X)^{-1} = \begin{pmatrix} 38.36167 & -0.97849 & 0.02849 \\ -0.97849 & 0.02526 & -0.00094 \\ 0.02849 & -0.00094 & 0.00024 \end{pmatrix}$$

$$X'Y = \begin{pmatrix} 98.4 \\ 3923.06 \\ 3613.1 \end{pmatrix}, \quad \hat{\beta} = (X'X)^{-1}X'Y = \begin{pmatrix} 39.0384 \\ -0.5755 \\ 0.0121 \end{pmatrix}$$

$$\hat{T} = 39.0384 - 0.5755 L + 0.0121 A$$

- Una ciudad situada 1 grado más hacia el Norte tiene una **disminución** de la temperatura media de aproximadamente 0.58 °C (para una misma altura).
- Un incremento de 10 m en la altura supone un **incremento** de 0.12 °C en la temperatura media (para una misma latitud).

b) Coeficiente de determinación y varianza residual:

$$n = 6, \quad k = 2, \quad Y'Y = 1628.02, \quad \hat{\beta}'X'Y = 1627.28, \quad \bar{y} = \frac{98.4}{6} = 16.4$$

$$R^2 = 1 - \frac{Y'Y - \hat{\beta}'X'Y}{Y'Y - n\bar{y}^2} = 1 - \frac{0.74}{14.26} = 0.9481$$

$$s_R^2 = \frac{1}{n - k - 1} (Y'Y - \hat{\beta}'X'Y) = \frac{1}{3} 0.74 = 0.2464$$

c) Predicción para la temperatura media de Tortosa (L= 40.5, A= 50):

¿Peligro de extrapolación oculta? $v_{hh} = x_h'(X'X)^{-1}x_h \leq v_{max} = \max(\text{diag}(H))$

$$x_h = (1 \quad 40.5 \quad 50)', \quad v_{hh} = x_h'(X'X)^{-1}x_h = 0.2014$$

$$v_{max} = \max(\text{diag}(H)) = \max(0.5884, 0.3766, 0.8911, 0.2034, 0.4795, 0.4611) = 0.8911$$

Predicción puntual para la temperatura media de Tortosa: $\hat{y}_h = x_h'\hat{\beta} = 16.3346$

Intervalo de predicción o de pronóstico para la temperatura media de Tortosa:

$$y_h \in \left[\hat{y}_h \pm t_{1-\alpha/2} s_R \sqrt{1 + \frac{1}{\hat{n}_h}} \right], \quad \text{donde } \hat{n}_h = \frac{1}{v_{hh}}$$

para $\alpha = 0.05$, $t_{1-\alpha/2} = t_{0.975}(3) = 3.18$, y por tanto:

$$y_h \in [16.3346 \pm 3.18 \sqrt{0.2464} \sqrt{1 + 0.2014}] = [14.6044, 18.0648]$$

Según el INE la temperatura media de Tortosa fue de 15.9°C en el período 1931-1980.