1

Regresión lineal simple

1.1 Introducción

Fue introducido por Francis Galton (1908). El modelo de regresión lineal simple está formado típicamente por:

$$y = \beta_0 + \beta_1 x + \epsilon.$$

Donde:

$$\begin{aligned} y &= \text{variable dependiente o variable de respuesta.} \\ x &= \text{variable independiente o explicativo o predictor.} \\ \beta_0 &= \text{intercepto } y. \\ \beta_1 &= \text{pendiente.} \\ \epsilon &= \text{error aleatorio.} \end{aligned}$$

Una presentación más general de un modelo de regresión sería:

$$y = E(y) + \epsilon,$$

Donde: $E(y)$ es la esperanza matemática de la variable respuesta.Cuando $E(y)$ es una combinación lineal de las variables explicativas $x_1, x_2, \ldots, x_k$ la regresión es una regresión lineal. Con $E(\epsilon_i) = 0$ y $Var(\epsilon_i) = \sigma^2$. Todos los ϵ_i son independientes.

Ahora debemos hallar buenos estimadores para β_0 y β_1.

1.2 Estimaciones por mínimos cuadrados

El principal objetivo de los mínimos cuadrados para un modelo de regresión lineal simple es hallar los estimadores b_0 y b_1 tales que la suma de la distancia al cuadrados de la respuesta real y_i y las respuesta de las pronosticadas $\hat{y}_i = \beta_0 + \beta_1 x_i$ alcanza el mínimo entre todas las opciones posibles de coeficientes de regresión β_0 y β_1. Es decir,

$$(b_0, b_1) = \arg \min_{(\beta_0, \beta_1)} \sum_{i=1}^{n} [\beta_0 + \beta_1 x_i]^2 .$$

Matemáticamente, las estimaciones de mínimos cuadrados de la regresión lineal simple se obtienen resolviendo el siguiente sistema:

$$\frac{\partial}{\partial \beta_0} \sum_{i=1}^{n} [y_i - (\beta_0 + \beta_i x_i)]^2 = 0 \tag{1.1}$$

$$\frac{\partial}{\partial \beta_1} \sum_{i=1}^{n} [y_i - (\beta_0 + \beta_i x_i)]^2 = 0 \tag{1.2}$$