

Introducción a la regresión lineal

Ramon Ceballos

6/2/2021

CONCEPTOS BÁSICOS DE LA REGRESIÓN LINEAL

1. Introducción

Seguramente, en algún momento de vuestra vida ya sea hojeando un libro de matemáticas, curioseando artículos científicos... habréis visto una línea recta o algún otro tipo de curva en un gráfico que se ajusta a las observaciones representadas por medio de puntos en el plano.

En general, la situación es la siguiente: supongamos que tenemos una serie de puntos en el plano cartesiano \mathbb{R}^2 , de la forma:

$$(x_1, y_1), \dots, (x_n, y_n)$$

Esta serie de puntos representan las observaciones de dos variables numéricas. Digamos que x es la edad e y el peso de n estudiantes.

Nuestro objetivo: describir la relación entre la **variable independiente**, x , y la **variable dependiente**, y , a partir de estas observaciones.

Para ello, lo que haremos será buscar una función $y = f(x)$ cuya gráfica se aproxime lo máximo posible a nuestros pares ordenados $(x_i, y_i)_{i=1, \dots, n}$.

Esta función nos dará un modelo matemático de cómo se comportan estas observaciones, lo cual nos permitirá entender mejor los mecanismos que relacionan las variables estudiadas o incluso, nos dará la oportunidad de hacer predicciones sobre futuras observaciones.

La primera opción es la más fácil. Consiste en estudiar si los puntos $(x_i, y_i)_{i=1, \dots, n}$ satisfacen una relación lineal de la forma:

$$y = ax + b$$

con $a, b \in \mathbb{R}$.

En este caso, se busca la recta $y = ax + b$ que mejor aproxime los puntos dados imponiendo que la suma de los cuadrados de las diferencias entre los valores y_i y sus aproximaciones $\tilde{y}_i = ax_i + b$ sea mínima (**minimización de los errores cuadráticos**). Es decir, que: $\sum_{i=1}^n (y_i - \tilde{y}_i)^2$ sea mínima.

1.1 Objetivos de esta sección

El objetivo de este tema no es otro más que enseñaros como hacer uso de R para obtener esta recta de regresión.

Veremos también cómo se puede evaluar numéricamente si esta recta se ajusta bien a las observaciones dadas.

Para ello, introduciremos algunas funciones de R y haremos uso de transformaciones logarítmicas para tratar casos en los que los puntos dados se aproximen mejor mediante una función exponencial o potencial.