

Regresión Lineal Simple

Regresión lineal simple

Podemos decir que la relación más simple entre dos variables x e y es la relación lineal entre ellas, de la forma:

$$y = m \times x + b$$

Donde:

x es la variable independiente

Para x fija, y va a ser la variable dependiente o respuesta.

m es la pendiente de la recta y b el intercepto (punto donde corta al eje Y).

La pendiente indica el cambio promedio en la variable de respuesta cuando la variable predictora aumenta en una unidad; y el intercepto indica el valor promedio de la variable de respuesta cuando la variable predictora vale 0.

El modelo de regresión lineal simple:

$$y_i = \beta_1 + \beta_2 \times X_i + e_i, i = 1, 2, \dots, n$$

La interpretación de los parámetros es la misma, pero se agrega el término de error o residuos. Éstos deberían tener una distribución normal y media igual a cero.

Para poder proponer un modelo de regresión lineal, debe haber una relación lineal entre las variables (prueba de correlación) y las observaciones deben ser independientes entre sí.

Estimación con R

Para encontrar modelos de regresión lineal, tanto simple como múltiple, vamos a utilizar la función de la instalación base de R: *lm()* para ajustar modelos lineales.

Los parámetros que vamos a utilizar son:

formula: vamos a indicar que variable va a ser la dependiente y cuáles las explicativas, con el operador `~`.

data: el data.frame del que se especifican las variables (opcional).

Para utilizar el modelo con datos nuevos y predecir el valor de la variable de respuesta, vamos a utilizar la función *predict()*.

Los parámetros que vamos a utilizar son:

object: objeto tipo modelo "lm".

newdata: el data.frame con los datos nuevos (opcional).

Ejercicios

Ejercicio 1 – Utilice los datos del ejercicio 1 de la guía anterior y encuentre un modelo lineal para explicar el peso en función de la altura de los individuos.

- Grafique la recta que ajusta el modelo e indique cuál es su expresión matemática.
- ¿Cómo son los residuos del modelo encontrado?
- Utilice el modelo para predecir el peso de un individuo de 177cm de altura.

Ejercicio 2 – En el archivo de texto “datos.txt” se encuentran los datos de peso y perímetro de cintura de doce individuos.

- Encuentre un modelo de regresión que permita predecir el perímetro de la cintura. El modelo es significativo?
- ¿Cuál es la expresión del modelo?
- Grafique los datos y la recta que ajusta el modelo.
- ¿Cómo cambia la variable de respuesta respecto a la explicativa?

Ejercicio 3 – El conjunto de datos precargado en R “faithful” contiene la duración de la erupción del geyser Old Faithful del parque nacional de Yellowstone, en EEUU; y el tiempo de espera entre erupciones.

- Encuentre un modelo para predecir la duración de la próxima erupción del geyser según el tiempo de espera.
- Grafique los datos y la recta que ajusta el modelo.
- Use el modelo obtenido para predecir la duración de la erupción, si el tiempo de espera fue:

Tiempo de espera [min]
58
67,2
75
92

- ¿Cómo describiría el comportamiento de las erupciones del geyser de acuerdo con el tiempo de espera?

Ejercicio 4 – En el archivo “supermercados.csv” se encuentran los datos de ventas de un producto en 200 supermercados de distintas ciudades y los gastos de publicidad que se realizaron en cada uno de ellos para ese producto.

Se quiere conocer la relación entre las ventas del producto y los gastos de publicidad.

- a. Explore los datos.
- b. Encontrar un modelo de regresión lineal simple para predecir las ventas según el tipo de publicidad utilizada.
- c. Grafique los datos y los modelos obtenidos.
- d. ¿Cuál de los tres modelos es el más adecuado para predecir las ventas?
- e. Prediga las ventas para los siguientes valores de publicidad:

TV	radios	diarios
125.55	23.40	89.70
234.00	85.50	65.00

Referencias

1. “Estadística y Machine Learning con R”. Disponible en:
<https://bookdown.org/content/2274/portada.html>