

Introducción al análisis de datos con R

Introducción al análisis de regresión con R

Ejercicios - Sesión 4

5 - 9 junio de 2023

Ejercicio 1: regresión lineal simple

Se cuenta con un conjunto de datos que incluyen información sobre indicadores educativos y la renta media de cada barrio en la ciudad de Madrid. El objetivo es analizar si existe una relación entre la titularidad de los colegios (público vs privado-concertado) y la renta media del barrio en los indicadores de rendimiento educativo.

Información sobre la base de datos: https://github.com/marespadafor/panelmadrid/blob/main/panel_information.md

Pasos a seguir:

- Cargar los datos desde el archivo `madrid_schools.dta` y explorar su estructura y distribución. Para esto se pueden utilizar las funciones `read.dta()` y `summary()` en R.
- Crear un gráfico de dispersión para visualizar la relación entre las variables renta media del barrio y el porcentaje de alumnos que completan la ESO en cada colegio. Para ello, se puede utilizar la función `ggplot()` del paquete `ggplot2`.
- Ajustar un modelo de regresión lineal simple para predecir el porcentaje de alumnos que completan la ESO en función de la renta media del barrio donde se ubica cada colegio. Para esto se puede utilizar la función `lm()` en R.
- Evaluar el modelo de regresión lineal simple utilizando el coeficiente de determinación (R^2). Para esto, se puede utilizar la función `summary()` en R.
- Realizar una predicción del porcentaje de alumnos que completan la ESO en función de la renta media de los barrios. Para ello, se puede utilizar la función `ggeffect()` en R.

Ejercicio 2: regresión lineal multivariante

Basándonos en el mismo supuesto que en el ejercicio anterior, vamos a demostrar cómo la titularidad del colegio se relaciona con el porcentaje de estudiantes que completan la Educación Secundaria Obligatoria (ESO), en función de la renta media del barrio donde se encuentra ubicado el colegio.

- Ajustar un modelo de regresión lineal para predecir el porcentaje de alumnos que completan la ESO en función de la renta media del barrio donde se ubica cada colegio, añadiendo un término de interacción entre la renta del barrio y la titularidad del colegio. Esto permitirá evaluar si la relación entre la renta del barrio y el porcentaje de alumnos que completan la ESO difiere según la titularidad del colegio.

- Realizar una predicción del porcentaje de alumnos que completan la ESO según la titularidad del centro en función de la renta media de los barrios. Para ello, se puede utilizar la función `ggeffect()` en R, que permitirá visualizar los efectos de la renta del barrio y la titularidad del colegio en la predicción del porcentaje de alumnos que completan la ESO. Finalmente, se deberá interpretar el gráfico para entender cómo se relacionan las variables