

Ejemplo de regresión lineal multivariado

1. Analizaremos el set de datos “Boston Houssing” en R (del paquete mlbench)

El set de datos cuenta con 14 variables, de las cuales estudiaremos las siguientes:

X_1 : (nox) Concentración de óxido de nitrógeno (partes por 10 millones)

X_2 : (rm) Promedio de cantidad de habitaciones por vivienda

X_3 : (lstat) Status mínimo de la población (porcentaje)

Y : (medv) Valor mediano de las casas ocupadas por dueños en miles de dólares

Se ajustó el modelo:

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + \epsilon$$

- a) ¿Qué puede decir a partir de la tabla de correlaciones entre las variables? Si tuviera que elegir una sola variable para plantear el modelo, ¿Cuál elegiría?
- b) A partir de la tabla de coeficientes estimados, ¿Qué variables son significativas? ¿A qué nivel? ¿Cuál es el valor de s? Especificar las hipótesis nulas y alternativas de cada uno de los test t reportados en la tabla. ¿Cómo se calculan el valor p de este test?
- c) ¿Es la regresión significativa? Especificar las hipótesis nula y alternativa de este test. ¿Cómo se calcula el p-valor en este caso? ¿Rechazaría a un nivel de significación de 0.05?
- d) ¿Cómo se calcula la matriz de correlación de los estimadores?
- e) ¿Qué supuestos necesita realizar para que las conclusiones anteriores sean válidas?