

## Clase 16 de abril

---

### 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:

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

- ¿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?
- 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?
- ¿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?
- ¿Cómo se calcula la matriz de correlación de los estimadores?
- ¿Qué supuestos necesita realizar para que las conclusiones anteriores sean válidas?

### 2. Data: `mtcars`, en R (no requiere ningún paquete adicional)

Del set de datos trabajaremos con las variables:

$Y$  = mpg (millas por galón)

$X_1$  = disp ( desplazamiento de los cilindros).

$X_2$  = wt ( peso del motor).

$X_3$  = cyl ( cantidad de cilindros).

$X_4$  = vs ( forma del motor).

- Consideremos sólo las variables  $Y$  y  $X_2$  para poder verlo graficamente:
  - Calcular, para cada punto del diseño, el intervalo de confianza de nivel 0.95 para la respuesta.
  - Calcular, para cada punto del diseño, el intervalo de predicción de nivel 0.95 para la respuesta.

- 3) Realizar en un mismo gráfico los pares de puntos  $(x,y)$ , la recta de mínimos cuadrados y los límites de los intervalos obtenidos en a) y b) para cada punto del diseño.
- 4) Calcular, para cada punto del diseño, el intervalo de confianza para la respuesta de manera que el nivel global de los 32 intervalos obtenidos sea 0.95.
- 5) Agregar al gráfico obtenido en c las bandas de confianza calculadas en d.