

# Modelos no paramétricos y de regresión/Estadística II Semestre 2018-1

## Tarea 6

Fecha de entrega: 20 de octubre

1. En el modelo de regresión simple, mostrar que los elementos de la matriz  $\mathbf{H}$  son

$$h_{ij} = \frac{1}{n} + \frac{(x_i - \bar{x}_n)(x_j - \bar{x}_n)}{S_{xx}} \quad y \quad h_{ii} = \frac{1}{n} + \frac{(x_i - \bar{x}_n)^2}{S_{xx}}.$$

2. Mostrar que el coeficiente  $R^2$  es el cuadrado de la correlación entre  $y_1, \dots, y_n$  y  $\hat{y}_1, \dots, \hat{y}_n$ .
3. Mostrar que una forma equivalente de contrastar la hipótesis de significancia de la regresión

$$H_0 : \beta_1 = \dots = \beta_p = 0 \quad \text{vs.} \quad H_1 : \beta_j \neq 0, \text{ para algún } j, \quad j = 1, \dots, p.$$

a partir del coeficiente  $R^2$  es la siguiente: rechazar  $H_0$  si el estadístico

$$F_0 = \frac{R^2(n-p-1)}{p(1-R^2)}$$

es mayor que  $F_{p, n-p-1}^{(1-\alpha)}$ .

4. Suponer que se ha ajustado un modelo con  $p = 2$  variables explicativas, con  $n = 25$  observaciones y que  $R^2 = 0.35$ .
- a) Probar la significancia del modelo utilizando  $\alpha = 0.05$ .
- b) Encontrar el mínimo valor de  $R^2$  que llevaría a rechazar  $H_0$  con  $\alpha = 0.05$ .
5. Se tiene interés en explicar el Producto Interno Bruto per capita de las entidades del país a partir de algunas variables utilizadas en el Índice de Rezago Social del CONEVAL. Un análisis exploratorio de los datos sugiere utilizar el modelo

$$\log \text{PIBpp} = \beta_0 + \beta_1 \text{EBIN} + \beta_2 \text{NDRE} + \beta_3 \text{NLAV} + \epsilon \quad (1)$$

donde:

- $\log \text{PIBpp}$  es el logaritmo de PIB per capita.
- $\text{EBIN}$  es el porcentaje de población de 15 y más años con educación básica incompleta.
- $\text{NDRE}$  es el porcentaje de viviendas particulares habitadas que no disponen de drenaje.
- $\text{NLAV}$  es el porcentaje de viviendas particulares habitadas que no disponen de lavadora.

Utilizar los datos en el archivo [rezago.csv](#) para ajustar el modelo (1).

- a) Reportar las estimaciones de  $\beta$ ,  $\sigma^2$  y  $V(\hat{\beta})$ .
- b) Interpretar  $\hat{\beta}_1$ ,  $\hat{\beta}_2$  y  $\hat{\beta}_3$  en el contexto de los datos.
- c) Probar la significancia del modelo (1). Interpretar los resultados. Utilizar  $\alpha = 0.05$ .
- d) Construir intervalos de confianza 95 % para las componentes de  $\beta$ , individuales y simultáneos (Bonferroni y Hotelling-Scheffé). Comparar las longitudes de los intervalos.
- e) Probar la significancia del modelo (1) utilizando los intervalos de confianza simultáneos del inciso anterior y comparar los resultados con el inciso 3. Utilizar  $\alpha = 0.05$ .
- f) Calcular el  $R^2$  con y sin ajuste e interpretar.
- g) Contrastar si el efecto de  $\text{EBIN}$  y  $\text{NDRE}$  es el mismo, es decir, contrastar la hipótesis  $H_0 : \beta_1 = \beta_2$ . Interpretar los resultados en el contexto de los datos.