Modelos No Paramétricos y de Regresión

Semestre 2022-1

Tarea #2: Primeros ajustes del modelo de Regresión

1.

Una empresa dedicada a la venta de equipo deportivo desea modelar su ventas mensuales para pronosticar las ventas del siguiente mes. (La base de sus ventas se puede descargar desde este link)

- Ajuste un modelo lineal de la forma

$$Y = \beta_0 + \beta_1 X + \varepsilon$$

  y calcule la suma de los cuadrados de los errores.

- Se sabe que la venta de artículos deportivos tiene una periodicidad de 2 años (48 meses) por lo que se propone ajustar un modelo de la forma de la forma:

$$Y = \beta_0 + \beta_1 sin\left(\frac{2\pi}{48}X\right) + \beta_2 cos\left(\frac{2\pi}{48}X\right) + \varepsilon$$

  Ajuste el modelo por medio de mínimos cuadrados haciendo uso de la función *nlm()*. Una vez ajustado calcule la suma de los cuadrados de los errores

- Suponga ahora que no sabemos la periodicidad de los datos exactamente, solo sabemos que es aproximadamente cada 48 meses, ajuste el modelo "no" lineal haciendo uso de la función *nlm()* :

$$Y = \beta_0 + \beta_1 sin\left(\frac{2\pi}{\beta_3}X\right) + \beta_2 cos\left(\frac{2\pi}{\beta_4}X\right) + \varepsilon$$

  Una vez ajustado calcule la suma de los cuadrados de los errores y compare los tres modelos. (Hint: Comienze la búsqueda de los parámetros β_3 y β_4 con el valor inicial de 40)

- De los tres modelos cuál eligiría y cuál sería su pronóstico de la venta del siguiente mes.

NOTA:

La base se anexa en el classroom

2.

(2 punto) Se tiene el modelo de regresión lineal simple $y = \beta_0 + \beta_1 x + \varepsilon$, con $\mathbb{E}(\varepsilon_i) = 0$, $\mathrm{Var}(\varepsilon_i) = \sigma^2$ y tal que $cov(\varepsilon_i, \varepsilon_j) = 0$ con $i \neq j$ Demostrar que:

$$\sum_{i=1}^{n} y_i = \sum_{i=1}^{n} \hat{y}_i$$

$$\sum_{i=1}^{n} x_i e_i = 0 = \sum_{i=1}^{n} \hat{y}_i e_i$$

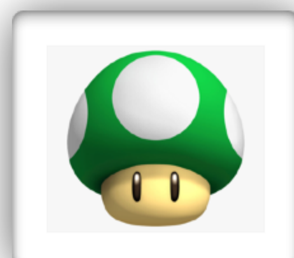
$$Cov(\hat{\beta}_0, \hat{\beta}_1) = \frac{-\overline{x}\sigma^2}{S_{xx}}$$

$$Cov(\overline{y}, \hat{\beta}_1) = 0$$

$$\sum_{i=1}^{n} (\hat{y}_i - \overline{y})^2 = \hat{\beta}_1^2 S_{xx}$$

$$\mathbb{E}\left(\sum_{i=1}^{n} (\hat{y}_i - \overline{y})^2\right) = \sigma^2 + \beta_1^2 S_{xx}$$

Donde $\hat{y}_i = \hat{\beta}_0 + \hat{\beta}_1 x_i$ y e_i es el residual i, es decir $e_i = y_i - \hat{y}_i$



Bonus point

Recordemos que DEFINIMOS el estimador para la varianza como sigue:

$$\sigma_M^2 := \frac{\sum_{i=1}^{n} (y_i - \hat{y}_1)^2}{2}$$

Demuestra que dicho estimador es sesgado.