Tarea 11

CIMAT

Nombre:	**Jairo Saul Diaz Soto / José Luis de León**
Maestría:	Ciencias Computacionales
Modulo:	Añhalisis de Datos
Instructor:	Dr. Johan Van Horebeek
Fecha de entrega:	2023 - 11 - 14

1. Considera el siguiente modelo:

$$Y = 1 + \beta log(x) + \epsilon, \quad \epsilon \sim \mathcal{N}(0, 0.1x), \quad x > 1$$

Calcule a mano el estimador de máxima verosimilitud de β usando la muestra $\{(y_i, x_i)\}$. ¿Qué distribución tiene?

Suponga que $n = 30$, y $x_i = i, 2 \leq i \leq n+1$; genera diferentes muestras del modelo, dibuja el modelo estimado, visualiza la distribución empírica de $\hat{Y}(10)$. Compáralo con lo que da el método de *1-KNN* ¿Qué observa?

R: Partiendo entonces de la definición de verosimilitud se tiene que es posible escribir la ecuación como

$$\epsilon = Y - 1 - \beta log(x)$$

y con ello la verosimilitud tiene la siguiente forma

$$L(\beta) = \prod_i P(\epsilon = y_i - 1 - \beta log(x_i))$$

dado que $\epsilon \sim \mathcal{N}(0, 0.1x_i)$, entonces

$$L(\beta) = \prod_i \frac{1}{\sqrt{2\pi \cdot 0.1x_i}} e^{-\frac{(y_i - \beta log(x_i) - 1)^2}{2 \cdot 0.1x_i}}$$

si se tiene ahora la log verosismilitud, se obtiene que

$$\ell(\beta) = -\sum_i \left(\frac{(y_i - \beta log(x_i) - 1)^2}{2 \cdot 0.1x_i} + \frac{1}{2} log(2\pi \cdot 0.1x_i) \right)$$

a lo cual, entonces obetenemos la derivada e igualando a 0

$$\frac{\partial \ell(\beta)}{\partial \beta} = -2 \sum_i \frac{y_i - \beta log(x_i) - 1}{2 \cdot 0.1x_i} \cdot (-log(x_i))$$

desarrollando los terminas e igualando a 0.

$$\sum_i \left(\frac{(y_i - 1)log(x_i)}{0.1x_i} - \frac{\beta log^2(x_i)}{0..1x_i} \right) = 0$$

despejando para β

$$\hat{\beta} = \frac{\sum_i \frac{(y_i - 1)log(x_i)}{x_i}}{\sum_i \frac{log^2(x_i)}{x_i}}$$

Ahora, para poder determinar la distribución de esta expresión, vamos a tomar y sustituir lo siguiente $y_i = 1 + \beta log(x_i) + \epsilon$, lo que deja lo siguiente

$$\hat{\beta} = \frac{\sum_i \frac{(\beta log(x_i) + \epsilon)log(x_i)}{x_i}}{\sum_i \frac{log^2(x_i)}{x_i}}$$

$$\hat{\beta} = \beta + \epsilon \frac{\sum_i \frac{\log(x_i)}{x_i}}{\sum_i \frac{log^2(x_i)}{x_i}}$$

de lo anterior, entonces es posible concluir que

$$\hat{\beta} \sim \mathcal{N}\left(\beta,\ 0.1 x_j \frac{\sum_i \frac{\log(x_i)}{x_i}}{\sum_i \frac{log^2(x_i)}{x_i}}\right)$$

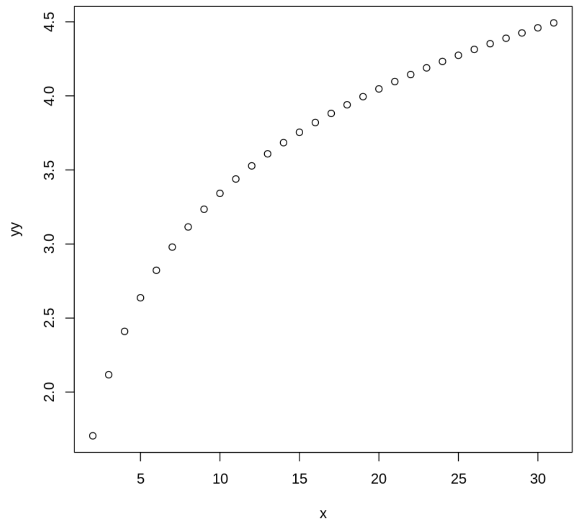

Figura 1: Dibujo de uno de los modelos estimados.

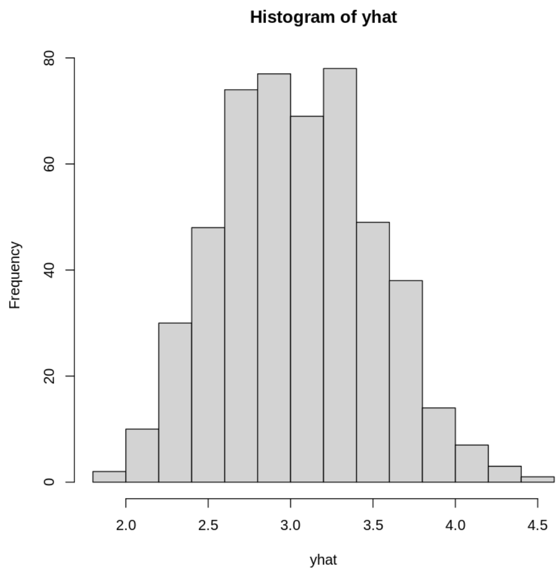

Figura 2: Distribución empírica para $\hat{Y}(10)$

Se puede observar que para ambos modelos, la estimación obtenida es bastante similar.

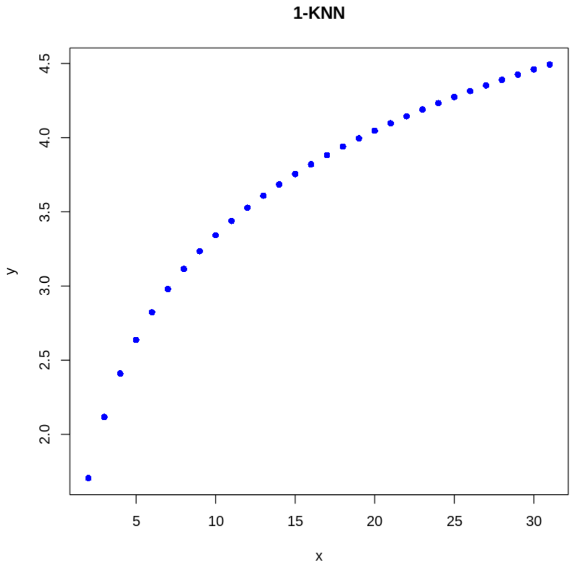

Figura 3: Modelo estimado con 1-KNN.

Se anexa el Notebook donde se resolvió dicho problema

2. Para una muestra de 14 muchachas elegidas al azar se tienen las mediciones del ancho de se cara a los 5 y 6 años. Haz una prueba de hipótesis para verificar si en promedio la diferencia entre las dos mediciones es de 0.2 pulgadas a $\alpha = 0.1$. Los datos son

5 Años	6 Años
7.33	7.53
7-49	7.70
7.27	7.46
7.93	8.21
7.56	7.81
7.81	8.01
7.46	7.72
6.94	7.13
7.49	7.68
7.44	7.66
7.95	8.11
7.04	7.20
7.10	7.25
7.64	7.79

R: Coomenzaremos creando el estadístico de prueba, en este caso, se n pide hallar la diferencia promedio, para cual si tomamos como X las medidas a los 5 años, mientras que Y la medida a los 6 años, entonces la diferencia promedio se definirá como

$$\mu_d = \frac{1}{n}\sum_{i}^{n}(y_i - x_i)$$

Es posible obtener la varianza para este valor, la cual se obtendrá como

$$\hat{\sigma}^2 = \frac{1}{n}\sum_{i}^{n}(y_i - x_i)^2 - \mu_d^2$$

entonces nuestro estadístico de prueba es el siguiente

$$T = \frac{\mu_d - \mu}{\frac{\hat{\sigma}}{\sqrt{n}}} \sim t_{n-1}$$

si se pide un nivel de significancia de $\alpha = 0.1$, entonces tendríamos que hallar lo siguiente

$$P(t > p) = 0.95, \qquad P(t < p) = 0.05$$

de lo que se obtiene entonces que

$$-1.7709 \leq \frac{\mu_d - \mu}{\frac{\hat{\sigma}}{\sqrt{n}}} \leq 1.7709$$

a lo cual se obtiene que

$$-1.7709\frac{\hat{\sigma}}{\sqrt{n}} + \mu \leq \mu_d \leq \mu + 1.7709\frac{\hat{\sigma}}{\sqrt{n}}$$

$$0.1814 \leq \mu_d \leq 0.2185$$

donde $\mu_d = 0.2007$ a lo que se apoya la hipotesis nula.