Práctico 1
Regresión Lineal Simple

Modelos Lineales - 2024

EJERCICIO 1

Un equipo de investigadores se propuso evaluar la hipétesis que dice que hogares con salarios más altos poseen, en promedio, un ahorro más alto. Indique:

1. ¿Cuáles deberían ser las unidades de análisis a relevar en la muestra?
2. ¿Qué variables deberían evaluarse en las unidades de análisis del punto anterior?
3. ¿Cuál sería un modelo lineal apropiado para estudiar esta hipótesis?
4. ¿Podría forular una prueba de hipótesis sobre alguno de los parámetros del modelo que de respuesta a la interrogante de los investigadores?

EJERCICIO 2

Considere el siguiente modelo lineal:

$$y_i = \beta_0 + \epsilon_i \qquad \forall_i = 1, \dots, n$$

Emplee la expresión matricial del estimador de MCO para obtener:

1. La expresión de $\hat{\beta}_0$.
2. La expresión de $\widehat{\mathbb{V}ar(\hat{\beta}_0)}$.

Relacione los resultados obtenidos con los obtenidos en el curso de Inferencia I.

EJERCICIO 3

Considere el modelo de regresión lineal simple dado por la ecuación:

$$y_i = \beta_0 + \beta_1 x_i + \epsilon_i \qquad \forall_i = 1, \dots, n$$

A partir del estimador $\hat{\beta}_{MCO} = \left(X^{'}X\right)^{-1} X^{'}Y$ demuestre que:

1. El primer elemento del vector $\hat{\beta}_{MCO}$ es $\bar{y} - \hat{\beta}_1 \bar{x}$.
2. El segundo elemento del vector $\hat{\beta}_{MCO}$ es $\frac{\sum_i (y_i - \bar{y})(x_i - \bar{x})}{\sum_i (x_i - \bar{x})^2}$
3. $\sum_i y_i = \sum_i \hat{y}_i$
4. Definiendo los residuos $\hat{\epsilon}_i = y_i - \hat{y}_i$, compruebe que $\sum_i \hat{\epsilon}_i = 0$.

5. De manera más general, comprueba que el vector de residuos es ortogonal a todas las columnas de la matriz X.

Nota: Para la resolución del último punto, pueden resultarle conveniente las ecuaciones normales.

EJERCICIO 4

Considere el modelo de regresión lineal simple dado por la ecuación:

$$y_i = \beta_0 + \beta_1 x_i + \epsilon_i \qquad \forall_i = 1, \ldots, n$$

A partir de la matriz de varianzas y covarianzas del estimador $\hat{\beta}_{MCO}$ dada por la expresión:

$$\mathbb{V}ar(\hat{\beta}) = \sigma^2 \left(X^{'}X\right)^{-1}$$

Demuestre que:

1. $\mathbb{V}ar(\hat{\beta}_0) = \frac{\sigma^2}{1/n - \bar{x}^2 / \sum_i (x_i - \bar{x})^2}$

2. $\mathbb{V}ar(\hat{\beta}_1) = \frac{\sigma^2}{\sum_i (x_i - \bar{x})^2}$

Nota: Se sugiere expresar los términos de la matrix $X^{'}X$ empleando sumatorias y luego utilizar el hecho de que:

$$\begin{bmatrix} a & b \\ c & d \end{bmatrix}^{-1} = \frac{1}{ad - bc} \begin{bmatrix} d & -b \\ -c & a \end{bmatrix}$$

EJERCICIO 5

La siguiente tabla contiene los mejores tiempos conseguidos en algunas pruebas de velocidad en atletismo en los Juegos Olímipicos de Atlanta 96. Las distancias están expresadas en metros y los tiempos en segundos.

Distancia	Tiempo (hombres)	Tiempo (mujeres)
100	9,84	10,84
200	19,32	22,12
400	43,19	48,25
800	102,58	117,73
1500	215,78	240,83
5000	787,96	899,88
10000	1627,34	1861,63
42195	7956,00	8765,00

1. Ajuste dos modelos de regresión lineal simple. Uno que relacione los tiempos de los hombres con las distancias y otro que relacione los tiempos de las mujeres con las distancias. Cuando obtenga las estimaciones de los coeficientes grafique las rectas de regresión.

2. A partir de las estimaciones calculadas en el punto anterior, obtenga la predicción del mejor tiempo para hombres y mujeres en una hipotética carrera de 3000 metros.

3. Calcule el valor del coeficiente de determinación R^2 en ambos casos e interprete su valor.

EJERCICIO 6

A partir de los datos del estudio de Francis Galton sobre la heredabilidad de la altura de los padres y los hijos se obtuvieron los siguientes estadísticos:

- $n = 898$.
- $\sum_i x_i = 1466,556$.
- $\sum_i y_i = 1468,802$.
- $\sum_i x_i y_i = 2399,859$.
- $\sum_i x_i^2 = 2396,738$.

Siendo X la variable "Altura promedio de los padres" y Y la variable "Altura del hijo", ambas expresadas en metros. A partir de estos datos:

1. Plantee el modelo de regresión lineal simple en el que se pretende explicar la altura de los niños en función de la altura de los padres.
2. Estime el valor de los parámetros.
3. Obtenga la predicción de la altura de un niño cuyos padres tienen una altura promedio de 1,68m e interprete dicha predicción.