Fabricio Camacho
4.822.364-8

Primera entrega de ejercicios – Regresión Lineal Simple.

EJERCICIO 2

Considere el siguiente modelo lineal:

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

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.

1)

El modelo por considerar no presenta variables explicativas x_i por lo que la variable y solo está representada por β_0 y ε_i.

La expresión matricial del estimador $\hat{\beta}_0$ por MCO entonces está dada por:

$$Y = \hat{\beta}_0 + \varepsilon$$

Donde: Y es un vector de dimensión n × 1

$\hat{\beta}_0$ es un escalar

ε es un vector de dimensión n × 1

Por Teorema 2.4.1 de Francesc Carmona:

Toda estimación MCO de β es solución de la ecuación $\boldsymbol{X'X\beta} = \boldsymbol{X'Y}$

Si $\boldsymbol{X}$ es de rango máximo entonces $\boldsymbol{X'X}$ tiene inversa y el estimador es

$$\hat{\boldsymbol{\beta}} = (\boldsymbol{X'X})^{-1}\boldsymbol{X'Y}$$

Por la particularidad de nuestro modelo la matriz X corresponde a una columna de unos, por lo que el estimador $\hat{\beta}_0$ es $\bar{Y}$.

2)

La varianza del modelo lineal es la varianza de los errores del mismo modelo.

$$\sigma^2 = var(\varepsilon_i) = var(y_i) \qquad i = 1, \dots . n$$

Así mismo, los ε_i son las únicas variables aleatorias del modelo, ya que los valores de y_i son determinísticos, y esta varianza es constante.

De forma matricial se puede expresar de la manera:

$$\boldsymbol{var(\varepsilon) = I_n \sigma^2}$$

Sin embargo, la varianza del modelo es desconocida, y por lo tanto debe ser estimada:

Por Teorema 2.4.3 y 2.5.1 de Francesc Carmona:

Suma de cuadrados residual (SCR) = $(\boldsymbol{Y - X\widehat{\beta}})'(\boldsymbol{Y - X\widehat{\beta}})$

$$\widehat{\sigma}^2 = \frac{SCR}{1 - Rg(X)}$$

De esta forma, $var(\widehat{\beta}) = \sigma^2 (X^{'}X)^{-1}$ y $(\widehat{var(\widehat{\beta})} = \widehat{\sigma}^2 (X^{'}X)^{-1}$ donde una vez más al ser un vector de unos nuestra matriz X, la estimación de la varianza del modelo es $\frac{\widehat{\sigma}^2}{n}$.

Este caso tiene similitudes con los de inferencia 1 debido a que en dicho curso las variables aleatorias que teníamos eran de la misma naturaleza, una variable Y tomaba distintos valores dependiendo de un parámetro β y tenía una varianza σ^2. No se da el caso donde otra variable X aporte información sobre el comportamiento de Y.

Entonces de esta variable Y obteníamos su información por medio de la media muestral $\bar{Y}$ y de su varianza muestral $\frac{\widehat{\sigma}^2}{n}$.

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.

1)

El modelo de regresión quedaría de la siguiente forma:

$$\text{"Altura del hijo"} = \hat{\beta}_0 + \hat{\beta}_1 \text{Altura promedio de los padres"} + \varepsilon$$

2)

Con el método de mínimos cuadrados lo que se busca es estimar valores de β_j que minimicen los residuos. La idea es que sean mínimos:

$$S = \sum_{i=1}^{n} \varepsilon_i^2 = \sum_{i=1}^{n} (y_i - (\beta_0 + \beta_1 x_{i1} + \cdots + \beta_k x_{ik}))^2$$

Entonces derivando e igualando a 0 esta expresión se obtienen los estimadores mínimos cuadrados

- $\hat{\beta}_0 = \bar{y} - \hat{\beta}_1 \bar{x}$
- $\hat{\beta}_1 = \dfrac{\sum_{i=1}^{n} (y_i - \bar{y})(x_{i-\bar{x}})}{\sum_{i=1}^{n} (x_{i-\bar{x}})^2}$

Sustituyendo con los datos de la letra del ejercicio:

$$\bar{x} = \frac{\sum_{i=1}^{n} x_i}{898} = \frac{1466{,}556}{898} = 1{,}633136$$

$$\bar{y} = \frac{\sum_{i=1}^{n} y_i}{898} = \frac{1468{,}802}{898} = 1{,}635637$$

$$\widehat{\beta}_1 = \frac{\sum_{i=1}^{n}(y_i - \bar{y})(x_i - \bar{x})}{\sum_{i=1}^{n}(x_i - \bar{x})^2} = \frac{\sum_{i=1}^{n}(y_i x_i - y_i\bar{x} - \bar{y}x_i + \bar{y}\bar{x})}{\sum_{i=1}^{n}(x_i^2 - 2x_i\bar{x} + \bar{x}^2)} =$$

$$= \frac{\sum_{i=1}^{n} y_i x_i - \sum_{i=1}^{n} y_i\bar{x} - \sum_{i=1}^{n} \bar{y}x_i + \sum_{i=1}^{n} \bar{y}\bar{x}}{\sum_{i=1}^{n} x_i^2 - \sum_{i=1}^{n} 2x_i\bar{x} + \sum_{i=1}^{n} \bar{x}^2}$$

$$= \frac{\sum_{i=1}^{n} y_i x_i - \bar{x}\sum_{i=1}^{n} y_i - \bar{y}\sum_{i=1}^{n} x_i + \sum_{i=1}^{n} \bar{y}\bar{x}}{\sum_{i=1}^{n} x_i^2 - 2\bar{x}\sum_{i=1}^{n} x_i + \sum_{i=1}^{n} \bar{x}^2}$$

$$= \frac{(2399{,}859 - 1{,}633136 * 1468{,}802 - 1{,}635637 * 1466{,}556 + 898 * 1{,}635637 * 1{,}633136)}{(2396{,}738 - 2 * 1{,}633136 * 1466{,}556 + 898 * 1{,}633136^2)}$$

$$= \frac{1{,}1058}{1{,}6528}$$

$$= 0{,}6690 \qquad \blacksquare$$

$$\widehat{\beta}_0 = \bar{y} - \widehat{\beta}_1\bar{x}$$

$$= \frac{1468{,}802}{898} - (0{,}6690) * \frac{1466{,}556}{898}$$

$$= 0{,}5430 \qquad \blacksquare$$

3) Para un valor de “Altura promedio de los padres de $x = 1{,}68$ según nuestro modelo $y_{(\text{"Altura del hijo"})} = \hat{\beta}_0 + \hat{\beta}_1 x + \varepsilon$ un niño sería de la altura 1,667. $\blacksquare$