Obtenção dos estimadores

Obtenha os estimadores de β_0 e β_1 a partir do Método dos Mínimos Quadrados, cujo objetivo é encontrar a reta que passa mais próxima ao mesmo tempo de todos os pontos. Neste caso, encontre os estimadores $\hat{\beta}_0$ e $\hat{\beta}_1$ que minimizam a soma dos erros ao quadrado.

Temos que o $\hat{y}_i$ estimado é dado pela relação: $\hat{y}_i = \hat{\beta}_0 + \hat{\beta}_1 x_i$

E o erro é: $e_i = y_i - \hat{y}_i$

$$y_i = \hat{\beta}_0 + \hat{\beta}_1 x_i + e_i$$

$$S(\hat{\beta}_0, \hat{\beta}_1) = \sum_{i=1}^{n} \left(y_i - \hat{\beta}_0 - \hat{\beta}_1 x_i \right)^2$$

Aplicando a regra da cadeia para a derivada parcial sobre os eixos:

$$\frac{\partial S}{\partial \hat{\beta}_0} = \frac{\partial S}{\partial \left[\sum_{i=1}^{n} \left(y_i - \hat{\beta}_0 - \hat{\beta}_1 x_i \right) \right]} \cdot \frac{\partial \left[\left(y_i - \hat{\beta}_0 - \hat{\beta}_1 x_i \right) \right]}{\partial \hat{\beta}_0}$$

Temos que:

$$\frac{\partial S}{\partial \left[\sum_{i=1}^{n} \left(y_i - \hat{\beta}_0 - \hat{\beta}_1 x_i \right) \right]} = 2 \sum_{i=1}^{n} \left(y_i - \hat{\beta}_0 - \hat{\beta}_1 x_i \right)$$

e

$$\frac{\partial \left[\sum_{i=1}^{n} \left(y_i - \hat{\beta}_0 - \hat{\beta}_1 x_i \right) \right]}{\partial \hat{\beta}_0} = -1$$

Impondo a condição de valor mínimo do parabolóide para $\hat{\beta}_0$:

$$\frac{\partial S}{\partial \hat{\beta}_0} = 2 \sum_{i=1}^{n} \left(y_i - \hat{\beta}_0 - \hat{\beta}_1 x_i \right) (-1) = 0$$

$$\frac{\partial S}{\partial \hat{\beta}_0} = -2 \sum_{i=1}^{n} \left(y_i - \hat{\beta}_0 - \hat{\beta}_1 x_i \right) = 0$$

E para $\hat{\beta}_1$:

$$\frac{\partial S}{\partial \hat{\beta}_1} = \frac{\partial S}{\partial \big[\sum_{i=1}^{n} \Big(y_i - \hat{\beta}_0 - \hat{\beta}_1 x_i \Big) \big]} \cdot \frac{\partial \big[\Big(y_i - \hat{\beta}_0 - \hat{\beta}_1 x_i \Big) \big]}{\partial \hat{\beta}_1}$$

$$\frac{\partial S}{\partial \hat{\beta}_1} = -2 \sum_{i=1}^{n} x_i \Big(y_i - \hat{\beta}_0 - \hat{\beta}_1 x_i \Big) = 0$$

Dividindo por $2n$ e distribuindo

$$\frac{-2 \sum_{i=1}^{n} y_i}{2n} + \frac{2 \sum_{i=1}^{n} \hat{\beta}_0}{2n} + \frac{2 \sum_{i=1}^{n} \hat{\beta}_1 x_i}{2n} = \frac{0}{2n}$$

$$\frac{- \sum_{i=1}^{n} y_i}{n} + \frac{\sum_{i=1}^{n} \hat{\beta}_0}{n} + \frac{\hat{\beta}_1 \sum_{i=1}^{n} x_i}{n} = 0$$

$$-\bar{y} + \hat{\beta}_0 + \hat{\beta}_1 \bar{x} = 0$$

Chegamos à expressão para $\hat{\beta}_0$:

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

Na expressão acima $\bar{y}$ é a Média Amostral de y e $\bar{x}$ é a Média Amostral de x

Podemos substituir $\hat{\beta}_0$ na relação original:

$$-2 \sum_{i=1}^{n} x_i \Big(y_i - \bar{y} + \hat{\beta}_1 \bar{x} - \hat{\beta}_1 x_i \Big) = 0$$

$$\sum_{i=1}^{n} \Big[x_i \left(y_i - \bar{y} \right) + x_i \hat{\beta}_1 \left(\bar{x} - x_i \right) \Big] = 0$$

$$\sum_{i=1}^{n} x_i \left(y_i - \bar{y} \right) + \hat{\beta}_1 \sum_{i=1}^{n} x_i \left(\bar{x} - x_i \right) = 0$$

Que nos dá, isolando $\beta 1$:

$$\hat{\beta}_1 = \frac{\sum_{i=1}^{n} x_i \left(y_i - \bar{y} \right)}{\sum_{i=1}^{n} x_i \left(\bar{x} - x_i \right)}$$

Reescrevendo levando em conta as relações

$$\sum_{i=1}^{n}(x_i - \overline{x})^2 = \sum_{i=0}^{n}(x_i^2 - x_i\overline{x})$$

e

$$\sum_{i=1}^{n}(x_i - \overline{x})(y_i - \overline{y}) = \sum_{i=0}^{n}(x_i y_i - y_i\overline{x})$$

desenvolvidas mais abaixo, temos:

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

Relações auxiliares

Temos que

$$\frac{1}{n}\sum_{i=1}^{n} x_i = \overline{x}$$

portanto

$$\sum_{i=1}^{n} x_i = n\overline{x}$$

e temos que

$$\frac{1}{n}\sum_{i=1}^{n} y_i = \overline{y}$$

portanto

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

Relação 1

A igualdade abaixo é importante para entendermos a fórmula dos β :

$$\sum_{i=1}^{n}(x_i - \overline{x})(y_i - \overline{y}) = \sum_{i=0}^{n}(x_i y_i - y_i\overline{x})$$

Vamos estudar como reescrever a relação:

$$\sum_{i=1}^{n}(x_i - \overline{x})(y_i - \overline{y})$$

Aplicando a propriedade distributiva:

$$\sum_{i=1}^{n}(x_i y_i - x_i\overline{y} - \overline{x}y_i + \overline{xy}) =$$

$$\sum_{i=1}^{n} x_i y_i - \overline{y}\sum_{i=1}^{n} x_i - \overline{x}\sum_{i=1}^{n} y_i + \overline{xy}\sum_{i=1}^{n} 1 =$$

$$\sum_{i=1}^{n} x_i y_i - \overline{y}n\overline{x} - \overline{x}n\overline{y} + \overline{xy}n =$$

$$\sum_{i=1}^{n} x_i y_i - n\overline{xy}$$

A relação acima pode ser escrita como:

$$\sum_{i=1}^{n} x_i y_i - n\overline{xy} = \sum_{i=1}^{n} x_i y_i - \overline{x}\sum_{i=0}^{n} y_i =$$

$$\sum_{i=0}^{n}(x_i y_i - y_i\overline{x})$$

Relação 2

Outra relação importante é a seguinte :

$$\sum_{i=1}^{n}(x_i - \overline{x})^2 = \sum_{i=0}^{n}(x_i^2 - x_i\overline{x})$$

Vamos estudar como reescrever a relação:

$$\sum_{i=1}^{n}(x_i - \overline{x})^2$$

Expandindo o quadrado:

$$\sum_{i=1}^{n}(x_i^2 - 2x_i\overline{x} + \overline{x}^2) =$$

$$\sum_{i=1}^{n} x_i^2 - 2n\overline{x}^2 + \sum_{i=1}^{n} x_i^2\overline{x}^2 =$$

$$\sum_{i=1}^{n} x_i^2 - 2n\overline{x}^2 + nx_i^2\overline{x}^2 =$$

$$\sum_{i=1}^{n} x_i^2 - n\overline{x}^2 =$$

$$\sum_{i=1}^{n} x_i^2 - \sum_{i=1}^{n} x_i\overline{x} = \sum_{i=1}^{n} \left(x_i^2 - x_i\overline{x}\right)$$