Lista 1 B

Davi Wentrick Feijó

2023-04-21

Em um exercício de simulação considerando um modelo de regressão linear simples com parâmetros $\beta_0 = 100$, $\beta_1 = 20$ e $\sigma^2 = 25$, será gerada uma observação de Y para $X = 5$.

a) É possível determinar a probabilidade exata de que será um valor entre 195 e 205? Justifique sua resposta. Nao pois nao sabemos qual é a distribuicao do erro do modelo

b) Se for utilizado um modelo de regressão com erro normal, é possível determinar a probabilidade acima? Caso seja possível, determine seu valor. Justifique sua resposta. Assumindo normalidade nos obtemos a seguinte distribuicao dos erros:

$$\epsilon \sim N(\mu, \sigma^2)$$

onde:

$$\mu = \hat{Y} = E(\hat{\beta}_0 + \hat{\beta}_1 \cdot X)$$

Substituindo para $X = 5$ obtemos:

$$\mu = \hat{Y} = E(100 + 20 \cdot 5) = 200$$

Como no exercicio foi dado que $\sigma^2 = 25$ chegamos na seguinte distribuicao dado $X = 5$:

$$\epsilon \sim N(200, 25)$$

Agora podemos calcular a probabilidade de $X \in [195, 205]$:

Manulamente teriamos que resolver a seguinte conta $P(195 < x < 205)$

$$P(195 < x < 205) = P(x < 205) - (1 - P(x > 195)) = P(x < 205) - P(x < 195)$$

Em seguida teriamos que passar para a normal padrao $N(0, 1)$ e calcular a probabilidade, mas vamos fazer computacionalmente por meio desse codigo:

```r
beta_0 = 100
beta_1 = 20
x_escolhido = 5
sigma_quadrado = 25 #desvio padrao
y_new_hat <- beta_0 + beta_1*x_escolhido #media

#se assumirmos que os erros seguem normalidade podemos
#calcular a partir de uma normal de media 200 e desvio padrao 25

n_value_lower <- pnorm(195, mean = y_new_hat,sd = 25) #P(x<205)
n_value_upper <- pnorm(205, mean = y_new_hat,sd = 25) #P(x<195)

prob = n_value_upper-n_value_lower
```

```
## O valor estimado Y quando X= 5 é: 200
```

```
## A probabilidade de Y estar entre 195 e 200 é: 0.1585194
```

Considere um modelo regressivo linear simples. com erro normal. Suponha que os valores dos parâmetros são $\beta_0 = 200$, $\beta_1 = 5$ e $\sigma = 4$

$$Y = 200 + 5X$$

a) Qual a distribuição de probabilidade de Y para $X = 10$? E para $X = 25$? Para $X = 10$:

$$Y = 200 + 5 \cdot 10 = 250$$

$$\hat{Y}|X = 10 \sim N(250, 16)$$

Para $X = 25$:

$$Y = 200 + 5 \cdot 10 = 325$$

$$\hat{Y}|X = 25 \sim N(325, 16)$$

b) Explique o significado dos parâmetros β_0 e β_1. Suponha que o alcance do modelo inclui $X = 0$. O paramentro β_0 nos indica onde que a reta toca o eixo Y, já os β_1 indica a taxa de aumento de Y em relacao a X. O ponto de $X = 0$ é igual ao β_0

Seja o MRLS $Y_i = \beta_0 + \beta_1 + X_i + \epsilon_i$, $\forall i = 1, n$ e β_0 e β_1 Seus estimadores:

Prove que:

$\sum_{i=1}^{n} Y_i = \sum_{i=1}^{n} \hat{Y}_i$ Vale relembrar que a $E(C) = C$ caso C seja uma constante

$$\sum_{i=1}^{n} Y_i = \sum_{i=1}^{n} \hat{Y}_i$$

$$\sum_{i=1}^{n} Y_i = \sum_{i=1}^{n} E(Y_i)$$

$$\sum_{i=1}^{n} Y_i = \sum_{i=1}^{n} E(\hat{\beta}_0 + \hat{\beta}_1 \cdot X_i)$$

$$\sum_{i=1}^{n} Y_i = \sum_{i=1}^{n} E(\hat{\beta}_0) + E(\hat{\beta}_1 \cdot X_i)$$

$$\sum_{i=1}^{n} Y_i = \sum_{i=1}^{n} \hat{\beta}_0 + \hat{\beta}_1 \cdot E(X_i)$$

$$\sum_{i=1}^{n} Y_i = \sum_{i=1}^{n} \hat{\beta}_0 + \hat{\beta}_1 \cdot X_i$$

$\sum_{i=1}^{n} e_i = 0$ **onde** $e_i = Y_i - \hat{Y}_i$ Utilizando do resultado que encontramos anteriormente temos:

$$\sum_{i=1}^{n} e_i = \sum_{i=1}^{n}(Y_i - \hat{\beta}_0 - \hat{\beta}_1 X_i)$$

$$\sum_{i=1}^{n}(Y_i - \hat{\beta}_0 - \hat{\beta}_1 X_i) = \sum_{i=1}^{n} Y_i - n\hat{\beta}_0 - \hat{\beta}_1 \sum_{i=1}^{n} X_i$$

Vale lembrar que $\bar{X} = \frac{\sum_{i=1}^{n} X_i}{n}$ que pode ser reescrito como: $\sum_{i=1}^{n} X_i = n\bar{X}$

$$\sum_{i=1}^{n} Y_i - n\hat{\beta}_0 - \hat{\beta}_1 \sum_{i=1}^{n} X_i = n\bar{Y} - n\hat{\beta}_0 - n\hat{\beta}_1 \bar{X}$$

$$n\bar{Y} - n\hat{\beta}_0 - n\hat{\beta}_1 \bar{X} = n(\bar{Y} - \hat{\beta}_0 - \hat{\beta}_1 \bar{X})$$

lembrando que $\hat{\beta}_0 = \bar{Y} - \hat{\beta}_1 \bar{X}$

$$n(\bar{Y} - \hat{\beta}_0 - \hat{\beta}_1 \bar{X}) = n(\bar{Y} - (\bar{Y} - \hat{\beta}_1 \bar{X}) - \hat{\beta}_1 \bar{X}) = 0$$

A reta de regressão estimada passa pelo centro de gravidade $(\bar{X}, \bar{Y})$ no diagrama de dispersão (X_i, Y_i).

$$\hat{\beta}_0 - \hat{\beta}_1 \bar{X} = (\bar{Y} - \hat{\beta}_1 \bar{X}) + \hat{\beta}_1 \bar{X} = \bar{Y}$$

$\sum_{i=1}^{n} X_i \cdot e_i = 0$

$$\sum_{i=1}^{n} X_i e_i = \sum_{i=1}^{n} X_i (Y_i - \hat{Y}_i) = \sum_{i=1}^{n} X_i (Y_i - \hat{\beta}_0 - \hat{\beta}_1 X_i)$$

$$= \sum_{i=1}^{n} X_i (Y_i - (\bar{Y} - \hat{\beta}_1 \bar{X}) - \hat{\beta}_1 X_i)$$

Reorganizando e tirando fator comum

$$= \sum_{i=1}^{n} X_i ((Y_i - \bar{Y}) - \hat{\beta}_1 (X_i - \bar{X}))$$

$$= \sum_{i=1}^{n} X_i ((Y_i - \bar{Y}) - \hat{\beta}_1 (X_i - \bar{X}))$$

$$= \sum_{i=1}^{n} X_i (Y_i - \bar{Y}) - \hat{\beta}_1 \sum_{i=1}^{n} X_i (X_i - \bar{X})$$

Vale notar de que no caso de centralizarmos (subtrair a media centralizando em torno do 0) X_i temos:

$$\hat{\beta}_1 = \frac{\sum_{i=1}^{n} (X_i - \bar{X})(Y_i - \bar{Y})}{\sum_{i=1}^{n} (X_i - \bar{X})^2} = \frac{\sum_{i=1}^{n} X_i (Y_i - \bar{Y})}{\sum_{i=1}^{n} (X_i - \bar{X})^2}$$

podemos chegar a seguinte formula e substituir na conta:

$$\sum_{i=1}^{n} X_i (Y_i - \bar{Y}) = \hat{\beta}_1 \sum_{i=1}^{n} (X_i - \bar{X})^2$$

$$= \hat{\beta}_1 \sum_{i=1}^{n} (X_i - \bar{X})^2 - \hat{\beta}_1 \sum_{i=1}^{n} X_i (X_i - \bar{X})$$

$$= \hat{\beta}_1 \sum_{i=1}^{n} [(X_i - \bar{X})^2 - X_i (X_i - \bar{X})]$$

$$= \hat{\beta}_1 \sum_{i=1}^{n} [X_i^2 - 2X_i \bar{X} + \bar{X^2} - X_i^2 + X_i \bar{X}]$$

$$= \hat{\beta}_1 \sum_{i=1}^{n} [\bar{X^2} - X_i \bar{X}]$$

$$= \bar{X} \hat{\beta}_1 \sum_{i=1}^{n} (\bar{X} - X_i) = 0$$

$\sum_{i=1}^{n} Y_i \cdot e_i = 0$

$$\sum_{i=1}^{n} Y_i \cdot e_i = 0$$

$$\sum_{i=1}^{n} Y_i \cdot (Y_i - \hat{Y}_i) = 0$$

$$\sum_{i=1}^{n} Y_i Y_i - Y_i \hat{Y}_i = 0$$

$$\sum_{i=1}^{n} Y_i^2 - Y_i \hat{Y}_i = 0$$

$$\sum_{i=1}^{n} Y_i^2 - (\beta_0 + \beta_1 \cdot X_i)(\hat{\beta}_0 + \hat{\beta}_1 \cdot X_i) = 0$$

$$\sum_{i=1}^{n} Y_i^2 - (\hat{\beta}_0 \beta_0 + \beta_0 \hat{\beta}_1 X_i + \hat{\beta}_0 \beta_1 X_i + \beta_1 X_i \hat{\beta}_1 X_i) = 0$$

Como $\hat{\beta}_0$ e $\hat{\beta}_1$ sao estimadores de β_0 e β_1 entao voce pode reorganizar a conta para ter os mesmos elementos de cada lado!

$$\sum_{i=1}^{n} Y_i^2 - \sum_{i=1}^{n} (\beta_0^2 + 2(\beta_0 \beta_1 X_i) + \beta_1^2 X_i^2) = 0$$

$$\sum_{i=1}^{n} Y_i^2 = \sum_{i=1}^{n} (\beta_0^2 + 2(\beta_0 \beta_1 X_i) + \beta_1^2 X_i^2)$$

Sabemos que $\beta_0^2 + 2(\beta_0 \beta_1 X_i) + \beta_1^2 X_i^2 = Y_i \hat{Y}_i$

$$\sum_{i=1}^{n} Y_i^2 = \sum_{i=1}^{n} Y_i \hat{Y}_i$$