Lista 1

Regressão

Caio Gomes Alves

0.1 Questão 5

Mostre que a reta de regressão obtida pelo método de mínimos quadrados passa pelo ponto $(\bar{x}, \bar{Y})$.

Proof. Como a reta de regressão é definida para todo $x \in \mathbb{R}$, e como sabemos que $\bar{x} \in \mathbb{R}$, o valor de $\bar{x}$ pertencerá à reta de regresão.

Para encontrar o $\hat{Y}$ associado a $\bar{x}$, substituemos na equação da reta estimada:

$$\begin{aligned}
\hat{Y} &= \hat{\beta}_0 + \hat{\beta}_1 \bar{x} \\
\hat{Y} &= \bar{Y} - \hat{\beta}_1 \bar{x} + \hat{\beta}_1 \bar{x} \\
\hat{Y} &= \bar{Y}
\end{aligned}$$

Assim, para o valor de $x = \bar{x}$, temos que o $\hat{Y}$ é igual a $\bar{Y}$. Assim, o ponto $(\bar{x}, \bar{Y})$ pertence à reta de regressão. □

0.2 Questão 8

Considere uma nova observação (x_{n+1}, Y_{n+1}), com $x_{n+1} = \bar{x}$. Sejam $\hat{\beta}_0^{n+1}$ e $\hat{\beta}_1^{n+1}$ os estimadores de mínimos quadrados obtidos utilizando todas as $n+1$ amostras. Mostre que

$$\hat{\beta}_1^{n+1} = \hat{\beta}_1$$

Interprete esse resultado e esboce um gráfico que represente essa propriedade.

Proof. Sabemos que o estimador de mínimos quadrados para β_1 é dado por:

$$\begin{aligned}
\hat{\beta}_1 &= \frac{\sum_{i=1}^{n}(y_i - \bar{y})(x_i - \bar{x})}{\sum_{i=1}^{n}(x_i - \bar{x})^2} \\
&= \frac{(y_1 - \bar{y})(x_1 - \bar{x}) + (y_2 - \bar{y})(x_2 - \bar{x}) + \ldots + (y_n - \bar{y})(x_n - \bar{x})}{(x_1 - \bar{x})^2 + (x_2 - \bar{x})^2 + \ldots + (x_n - \bar{x})^2}
\end{aligned}$$

Assim, o estimador para β_1 incluindo a observação $n+1$ é dada por:

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

$$= \frac{(y_1 - \bar{y})(x_1 - \bar{x}) + (y_2 - \bar{y})(x_2 - \bar{x}) + \ldots + (y_n - \bar{y})(x_n - \bar{x}) + (y_{n+1} - \bar{y})(\bar{x} - \bar{x})}{(x_1 - \bar{x})^2 + (x_2 - \bar{x})^2 + \ldots + (x_n - \bar{x})^2 + (\bar{x} - \bar{x})^2}$$

$$= \frac{(y_1 - \bar{y})(x_1 - \bar{x}) + (y_2 - \bar{y})(x_2 - \bar{x}) + \ldots + (y_n - \bar{y})(x_n - \bar{x}) + 0}{(x_1 - \bar{x})^2 + (x_2 - \bar{x})^2 + \ldots + (x_n - \bar{x})^2 + 0^2}$$

$$\hat{\beta}_1^{n+1} = \hat{\beta}_1$$

Assim, podemos ver que o estimador para o β_1 não será alterado, caso a nova observação tenha valor $x_{n+1} = \bar{x}$. Para visualizar esse efeito, consideremos o seguinte exemplo simples:

```r
# Seed para reprodutibilidade:
set.seed(42)

# Considere os valores de x vindos de uma normal N(10, 2):
df <- data.frame(
    x <- rnorm(10, 10, 10)
)

# Os valores de y são gerados a partir de 3*x + N(0, 1):
df$y <- (3 * x + rnorm(10, 0, 1.5))

# Valores de Beta_0 e Beta_1 gerados a partir de mínimos quadrados:
(lm(y ~ x, data = df))
```

```
##
## Call:
## lm(formula = y ~ x, data = df)
##
## Coefficients:
## (Intercept)            x
##       1.453        2.890
```

```r
# Valor da média de x:
(x_barra <- mean(df$x))
```

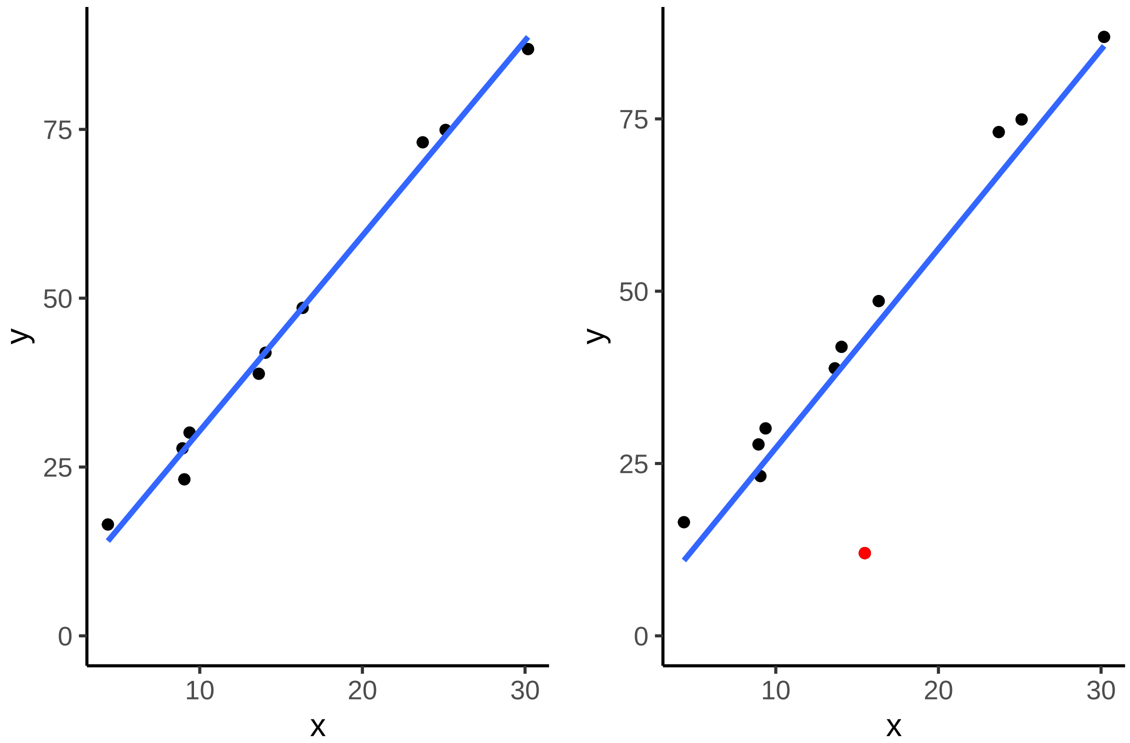
```
## [1] 15.47297
```

```r
# Considere agora uma nova observação, no ponto (x_barra, 17):
df2 <- data.frame(
    x = c(df$x, x_barra),
    y = c(df$y, 12)
)

# Verifique que a estimativa de Beta_0 foi alterada,
# enquanto que a de Beta_1 se manteve a mesma:
(lm(y ~ x, data = df2))
```

```
##
## Call:
## lm(formula = y ~ x, data = df2)
##
## Coefficients:
## (Intercept)            x
##      -1.653        2.890
```

```
# Podemos observar graficamente a mudança na estimativa de
# Beta_0, com os graficos a seguir:
```



□