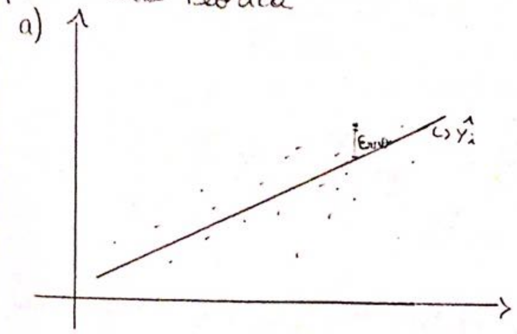
Etapa 2 - Bruno Leite e Elisa Malzoni

↳ Parte teórica

a)



$$\hat{y}_i = \hat{\beta}_0 + \hat{\beta}_1 x_i$$

$$\hat{y}_i = \hat{\beta}_0 + \hat{\beta}_1 \cdot x_i + \underbrace{\varepsilon_i}$$

$$\downarrow$$

$$\varepsilon_i = y_i - \hat{y}_i$$

Soma dos erros quadráticos:

$$\text{Desvio} = D = \sum_{i=1}^{n} (\varepsilon_i)^2 = \sum_{i=1}^{n} (y_i - \hat{y}_i)^2 = \sum_{i=1}^{n} (y_i - \hat{\beta}_0 - \hat{\beta}_1 x_i)^2$$

→ P/ minimizar D deve-se derivar em relação a β_0 e β_1

$\therefore$

$$\boxed{\frac{dD}{d\beta_0} = \sum_{i=1}^{n} -2(y_i - \hat{\beta}_0 - \hat{\beta}_1 x_i) = 0} \quad \text{(I)}$$

$$\boxed{\frac{dD}{d\beta_1} = \sum_{i=1}^{n} -2x_i (y_i - \hat{\beta}_0 - \hat{\beta}_1 x_i) = 0} \quad \text{(II)}$$

$$\begin{cases} \text{(I)} \; -2\sum_{i=1}^{n}(y_i - \hat{\beta}_0 - \hat{\beta}_1 \cdot x_i) = 0 \xrightarrow{\text{distribuindo}} -2 \cdot \left(\sum_{i=1}^{n} y_i - n \cdot \hat{\beta}_0 - \hat{\beta}_1 \sum_{i=1}^{n} x_i\right) = 0 \\ \text{(II)} \; -2\sum_{i=1}^{n}(x_i \cdot y_i - x_i \cdot \hat{\beta}_0 - \hat{\beta}_1 \cdot x_i^2) = 0 \Rightarrow -2\left(\sum_{i=1}^{n} x_i \cdot y_i - \hat{\beta}_0 \sum_{i=1}^{n} x_i - \hat{\beta}_1 \sum_{i=1}^{n} x_i^2\right) = 0 \end{cases}$$

$$\begin{cases} \text{I} & \hat{\beta}_0 = \dfrac{\left(\sum y_i - \hat{\beta}_1 \sum x_i\right)}{n} \\ \text{II} & \hat{\beta}_0 = \dfrac{\left(\sum x_i y_i - \hat{\beta}_1 \sum x_i^2\right)}{\sum x_i} \end{cases}$$

@...

$$\frac{\sum y_i - \hat{\beta}_1 \sum x_i}{n} = \frac{\sum x_i y_i - \hat{\beta}_1 \sum x_i^2}{\sum x_i}$$

$$\sum y_i \sum x_i - \hat{\beta}_1 \left(\sum x_i\right)^2 = n \sum x_i y_i - n \hat{\beta}_1 \sum x_i^2$$

$$n \hat{\beta}_1 \sum x_i^2 - \hat{\beta}_1 \left(\sum x_i\right)^2 = n \sum x_i y_i - \sum y_i \sum x_i$$

$$\hat{\beta}_1 \left[n \sum x_i^2 - \left(\sum x_i\right)^2 \right] = n \sum x_i y_i - \sum y_i \sum x_i$$

$$\boxed{\hat{\beta}_1 = \frac{n \sum x_i y_i - \sum y_i \sum x_i}{n \cdot \sum x_i^2 - \left(\sum x_i\right)^2}}$$

$$\text{I}: \quad \boxed{\hat{\beta}_0 = \frac{\sum y_i - \hat{\beta}_1 \cdot \sum x_i}{n}}$$

b) Suposições sobre os erros:

· Distribuição: aproxima-se de uma normal, uma vez que - após construída a reta - o desejado é que todos pontos sejam equidistantes da reta.

· Valor esperado / Variância: $E(\varepsilon_i)=0$ e $Var(\varepsilon_i)=\sigma^2$, já que $\varepsilon_i \sim N(0,\sigma^2)$, ou seja, o desejado é que o erro se aproxime de 0 e que a variância seja sempre a mesma.

→ Pode-se checar a adequação de tais suposições pelo gráfico.

c) $H_0: \hat{\beta}_1 = 0 \Rightarrow \nexists$ relação entre x e Y

→ De acordo com a equação $\hat{Y}_i = \hat{\beta}_0 + \hat{\beta}_1 \cdot x_i + \varepsilon_i$, se $\hat{\beta}_1$ for 0, o $\hat{Y}_i$ só dependerá de $\hat{\beta}_0$ e ε_i, portanto não haverá relação entre as variáveis.

$H_1: \hat{\beta}_1 \neq 0 \Rightarrow \exists$ relação entre x e Y

→ De acordo com a equação $\hat{Y}_i = \hat{\beta}_0 + \hat{\beta}_1 \cdot x_i + \varepsilon_i$, só há relação entre as variáveis se $\hat{\beta}_1$ for diferente de 0

d) Sim, é possível fazer uma regressão com mais do que uma variável. Neste caso ficaria da seguinte maneira a:

- equação:

$$\hat{Y} = \hat{\beta}_0 + \hat{\beta}_1 x_{i1} + \hat{\beta}_2 x_{i2} + \varepsilon_i$$

- distribuição: permanecerá uma normal, uma vez que — da mesma maneira — o desejado é que todos os pontos estejam equidistantes da reta.

- valor esperado/variância: $E(\varepsilon_i) = 0$ e $Var(\varepsilon_i) = \sigma^2$

Da mesma maneira, o valor esperado da normal será 0 uma vez que o desejado é que o erro se aproxime de tal valor. Já a variância, segue sendo sempre a mesma, porém não é igual à variância da regressão com 1 variável.

- teste de hipóteses:

$H_0: \hat{\beta}_1 = \hat{\beta}_2 = 0 \Rightarrow \nexists$ relação entre x_1, x_2 e Y

↳ Da mesma maneira que na regressão com 1 variável, se $\hat{\beta}_1$ e $\hat{\beta}_2$ forem iguais a 0, Y não dependerá nem de x_1 nem de x_2.

$H_1: \hat{\beta}_1 \neq 0 \vee \hat{\beta} \neq 0 \Rightarrow \exists$ relação entre x_1 e Y e/ou x_2 e Y

↳ A partir da mesma fórmula, a única maneira de x_1 e/ou x_2 se relacionarem à Y é com $\hat{\beta}_1$ e/ou $\hat{\beta}_2$ sendo diferente de 0