Projeto 3 Etapa 2:

a) Cada observação Y é estimada pela equação:

$$\boxed{Y = \beta_0 + \beta_1 x + \varepsilon}$$

$$\boxed{\hat{Y} = \beta_0 + \beta_1 \cdot x_i}$$

→ sendo i o número de amostras.

Variável

$$L = \sum_{i=1}^{m} \varepsilon_i^2 = \sum_{i=1}^{m} (Y - \hat{Y})^2 = \sum_{i=1}^{m} (Y_i \underbrace{- \beta_0 - \beta_1 x_i}_{-\hat{y}\ \to \text{ previsto pela reta}})^2$$

os estimadores dos mínimos quadrados de β_0 e β_1, dizem que suas aproximações devem satisfazer a seguinte equação:

$$\left.\frac{\partial L}{\partial \beta_0}\right|_{\hat{\beta}_0, \hat{\beta}_1} = -2 \sum_{i=1}^{m} (Y_i - \hat{\beta}_0 - \hat{\beta}_1 x_i) = 0$$

$$\left.\frac{\partial L}{\partial \beta_1}\right|_{\hat{\beta}_0, \hat{\beta}_1} -2 \sum_{i=1}^{m} (Y_i - \hat{\beta}_0 - \hat{\beta}_1 x_i) x_i = 0$$

Obs: A solução dessas equações resulta no estimador dos mínimos quadrados de $\hat{\beta}_0$ e $\hat{\beta}_1$.

simplificando as equações obtemos:

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

$$\hat{\beta}_1 = \frac{S_{xy}}{S_{xx}} = \frac{\sum_{i=1}^{m} (x_i - \bar{x}) \cdot (y_i - \bar{y})}{\sum_{i=1}^{m} (x_i - \bar{x})^2}$$

sendo $\bar{y} = (1/m) \sum_{i=1}^{m} y_i$ e

$\bar{x} = (1/m) \sum_{i=1}^{m} x_i$

B) A suposição feita é que podemos aproximar um conjunto de dados através de uma regressão linear, usando o método dos mínimos quadrados. Para verificar a adequação dessas suposições, basta observar se a reta de ajuste aproxima-se da maioria dos *data points.*

C) A hipótese nula constata que as variáveis utilizadas possuem uma correlação considerável entre si. Ao recusar H_0, estamos afirmando que o modelo não ajustou bem os dados, pois a variável regressora em questão não altera a variável em resposta.

D) É possível. A mudança seria na quantidade de variáveis regressoras e de suas respectivas derivadas parciais. A equação abaixo precisaria ser adaptada para incluir as demais variáveis, o que impactaria todo o processo matemático. Ao adicionar mais variáveis, conseguimos um modelo que se ajusta melhor aos dados. Em relação ao teste de hipóteses, acreditamos que quanto mais variáveis adiciona-se, maior é a chance de rejeição de H_0, já que maior será o erro acumulado das variáveis regressoras.

$$Y = \beta_0 + \beta_1 x_i + \beta_2 x_i + \cdots + \beta_n x_i$$