Projeto 3 - 2ª etapa

a)

$$SQE = \sum_{i=1}^{n} \varepsilon_i^2 = \sum_{i=1}^{n} (y_i - \hat{y}_i)^2 = \sum_{i=1}^{n} (y_i - \beta_0 - \beta_1 x_i)^2$$

β_0)

$$\frac{SQE}{\beta_0} = \sum_{i=1}^{n} 2(y_i - \beta_0 - \beta_1 x_i).(-1) = \sum_{i=1}^{n} -2(y_i - \beta_0 - \beta_1 x_i) = 0$$

$$-2\sum_{i=1}^{n}(y_i - \beta_0 - \beta_1 x_i) = 0$$

$$\frac{-\cancel{2}\sum_{i=1}^{n} y_i}{\cancel{2}n} + \frac{\cancel{2}\sum_{i=1}^{n} \beta_0}{\cancel{2}n} + \frac{\cancel{2}\sum_{i=1}^{n} \beta_1 x_i}{2n} = 0$$

$$-y_i + \beta_0 + \beta_1 \bar{x} = 0$$

$$\boxed{\beta_0 = -\beta_1 \bar{x} + \bar{y}}$$

β_1) $$\frac{SQE}{\beta_1} = \sum_{i=1}^{n} 2(y_i - \beta_0 - \beta_1 x_i).(-x_i) = \sum_{i=1}^{n} -2x_i(y_i - \beta_0 - \beta_1 x_i)$$

$$\frac{-\cancel{2}\sum_{i=1}^{n} x_i(y_i + \beta_1\bar{x} + \bar{y} - \beta_1 x_i) = 0}{-\cancel{2}}$$

$$\sum_{i=1}^{n} x_i(y_i - \bar{y}) + \beta_1 \sum_{i=1}^{n} x_i(\bar{x} - x_i) = 0$$

$$\boxed{\beta_1 = \frac{\sum_{i=1}^{n} x_i(y_i - \bar{y})}{\sum_{i=1}^{n} x_i(x_i - \bar{x})}}$$

b) A distribuição de erros segue o padrão Normal (X ~ N), sendo seu valor esperado (E(x)) igual a 0 e sua variância (Var(x)) constante. Pode-se checar sua adequação através de análise visual.

c) $\begin{cases} H_0: \beta_1 = 0 \\ H_A: \beta_1 \neq 0 \end{cases}$ Na hipótese nula, não há correlação entre as variáveis X e Y, enquanto na hipótese alternativa há.

d) Há sim a possibilidade de fazer uma regressão múltipla. O que muda é que ao invés de ter apenas o β_0 e o β_1, como ocorre na regressão simples, a regressão com mais de uma variável explicativa possui um número maior de Betas (β) na função, número esse que é proporcional a quantidade de variáveis na regressão. Portanto, a função passa a ser

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \ldots + \beta_n x_n$$, sendo "n" o número de variáveis.

Além disso, ao analisar a suposição do modelo em questão pode-se perceber que ela permanece igual tanto na regressão simples como na regressão múltipla. Quanto ao teste de hipótese a diferença entre ambas as regressões é que na múltipla o número de testes aumenta de acordo com a quantidade de variáveis.