

## Parte Teórica

a)

Cálculos:

$$\left. \begin{aligned} a) \hat{y}_i &= \hat{\beta}_0 + \hat{\beta}_1 X_i \\ y_i &= \hat{\beta}_0 + \hat{\beta}_1 X_i + E_i \end{aligned} \right\} \therefore E_i = y_i - \hat{y}_i$$

Para estimar  $\beta_0$  e  $\beta_1$  será usado o método dos mínimos quadrados, onde que esses valores devem ser escolhidos de modo a minimizar a soma dos erros ao quadrado (SEQ).

$$SEQ = \sum_{i=1}^n E_i^2 = \sum_{i=1}^n (y_i - \hat{y}_i)^2$$

$$SEQ = \sum_{i=1}^n (y_i - \hat{\beta}_0 - \hat{\beta}_1 X_i)^2 \left\{ \begin{aligned} &\frac{dSEQ}{d\hat{\beta}_0} = 0 \quad (1) \\ &\frac{dSEQ}{d\hat{\beta}_1} = 0 \quad (2) \end{aligned} \right.$$

$$(1) \frac{dSEQ}{d\hat{\beta}_0} = 2(y_1 - \hat{\beta}_0 - \hat{\beta}_1 X_1)^{2-1} (-1) + \dots + 2(y_n - \hat{\beta}_0 - \hat{\beta}_1 X_n)^{2-1} (-1)$$

$$= \sum_{i=1}^n (-2) \cdot (y_i - \hat{\beta}_0 - \hat{\beta}_1 X_i) = -2 \sum_{i=1}^n (y_i - \hat{\beta}_0 - \hat{\beta}_1 X_i)$$

$$= (-2) \left\{ \sum_{i=1}^n y_i - n\hat{\beta}_0 - \hat{\beta}_1 \sum_{i=1}^n X_i \right\}$$

$$= (-2) \cdot n \cdot \left\{ \frac{\sum y_i}{n} - \hat{\beta}_0 - \hat{\beta}_1 \frac{\sum X_i}{n} \right\} = (-2)n$$

DOM SEG TER QUA QUI SEX SAB

$$= n(-2) \{ \bar{y} - \hat{\beta}_0 - \hat{\beta}_1 \cdot \bar{x} \}$$

$$\Rightarrow \hat{\beta}_0 = -\hat{\beta}_1 \bar{x} + \bar{y} \quad (I)$$

$$(2) \frac{dSEQ}{d\hat{\beta}_1} = -2 \sum_{i=1}^n x_i (y_i - \hat{\beta}_0 - \hat{\beta}_1 x_i) = 0$$

$$= -2 \{ \sum_{i=1}^n x_i y_i - \hat{\beta}_0 \sum_{i=1}^n x_i - \hat{\beta}_1 \sum_{i=1}^n x_i^2 \} = 0$$

Ao substituir a equação I na obtida acima obtemos:

$$\sum_{i=1}^n x_i y_i - (-\hat{\beta}_1 \bar{x} + \bar{y}) \sum_{i=1}^n x_i - \hat{\beta}_1 \sum_{i=1}^n x_i^2 = 0$$

$$\hat{\beta}_1 = \frac{\bar{y} \sum_{i=1}^n x_i - \sum_{i=1}^n x_i y_i}{\bar{x} \sum_{i=1}^n x_i - \sum_{i=1}^n x_i^2} = \frac{cov(X, Y)}{var(X)}$$

Obtemos então que  $\hat{\beta}_0 = -\hat{\beta}_1 \bar{x} + \bar{y}$  e que  $\hat{\beta}_1 = \frac{cov(X, Y)}{var(X)}$ .

**b)** Podemos supor que a distribuição esperada para o erro  $E$  é a normal, pois descreveria melhor o comportamento do erro, com a média  $\mu$  tendo seu valor esperado igual a 0 e a variância sendo desconhecida e constante ( $Var(E_i) = \sigma^2$ ), ou seja,  $E_i \sim (0, \sigma^2)$ .

A adequação dessas suposições pode ser checada na prática pela construção de gráficos e análise destes.

**c)**

$$H_0: \hat{\beta}_1 = 0$$

$$H_A: \hat{\beta}_1 \neq 0$$

A hipótese nula não ser rejeitada significa que não existe uma relação entre as variáveis X (variável independente ou explicativa) e Y (variável dependente ou resposta), pois faz com que  $Y = \hat{\beta}_0 + E$ .

**d)** Sim, é possível realizar regressão com mais de uma variável explicativa (regressão múltipla) e a diferença na fórmula principal é que agora, como existe mais de uma variável e cada uma delas possui seu próprio  $\beta$ , fica:  $Y = \hat{\beta}_0 + \hat{\beta}_1 X_1 + \dots + \hat{\beta}_w X_w + E$ , onde “w” é o número de variáveis.

Em termos de suposição do modelo, o da regressão múltipla não possui diferenças significativas em relação ao da regressão linear simples.

Já o teste de hipótese tem que ser realizado mais de uma vez, com a quantidade dependendo do número de variáveis que estão sendo analisadas (por exemplo, caso estejam sendo avaliadas cinco variáveis, terão que ser feitos cinco testes de hipótese).