# Teste t na regressão

## Variâncias e covariâncias

Lembrando que:

$S_{xx}$ é a variação total elevada ao quadrado de $x$ em relação à média $\bar{x}$

$$S_{xx} = \sum_{i=1}^{n} (x_i - \bar{x})^2$$

$S_{yy}$ é a variação total elevada ao quadrado de $y$ em relação à média $\bar{y}$

$$S_{yy} = \sum_{i=1}^{n} (y_i - \bar{y})^2$$

Temos também que as variância $\sigma_X^2$ e $\sigma_Y^2$ são:

$$\sigma_X^2 = \frac{S_{xx}}{n}$$

$$\sigma_Y^2 = \frac{S_{yy}}{n}$$

$S_{xy}$ é o produto da variação total de cada variável em relação à sua média $\bar{x}$ e $\bar{y}$:

$$SS_{xy} = \sum_{i=1}^{n} (x_i - \bar{x})(y_i - \bar{y})$$

A covariância $Cov(X, Y)$ é:

$$Cov(X, Y) = \frac{S_{xy}}{n}$$

## Regressão simples

Conforme foi demonstrado na entrega 2 do projeto, os resultados para regressão de mínimos quadrados são:

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

e

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

Lembrando que $\hat{\beta}_0$ e $\hat{\beta}_1$ são os estimadores encontrados a partir dos dados para os parâmetros $\beta_0$ e $\beta_1$ do modelo de regressão.

## Erros na regressão

### Soma dos quadrados dos resíduos

A soma dos quadrados dos resíduos é o quadrado da variação encontrada nos dados que **não é explicada** pelo modelo de regressão. Ou seja, é a diferença entre $y_i$ que está presente nos dados e o valor $\hat{y}_i$ que a reta dá para o $x_i$ correspondente.

Este valor costuma ser chamado de soma dos quadrados dos resíduos ($SQRes$) ou também *error sum of squares* ou $SS_E$

$$SQRes = SS_E = \sum_{i=1}^{n}(y_i - \hat{y}_i)^2 = \sum_{i=1}^{n}\epsilon_i^2$$

### Soma dos quadrados da regressão

É a variabilidade que é explicada pela regressão. Tipicamente é chamada de SQR ou $SS_R$

$$SQReg = SS_R = (\hat{y}_i - \bar{y})^2$$

### Soma dos quadrados totais

É a soma da variabilidade total presente no modelo. Costuma ser chamado de SQT ou de $SS_T$ .

$$SQT = SS_T = \sum_{i=1}^{n}(y_i - \bar{y})^2$$

A soma dos quadrados totais é a soma da porção explicada pela regressão com a parte que não é explicada.

$$SS_T = \sum_{i=1}^{n}(y_i - \bar{y})^2 = \sum_{i=1}^{n}(\hat{y}_i - \bar{y}_i)^2 + \sum_{i=1}^{n}(y_i - \hat{y}_i)^2 = SS_R + SS_E = SQReg + SQRes$$

**Variância dos erros do modelo de regressão**

Para o método dos mínimos quadrados funcionar, foi assumido que os resíduos ou erros têm média zero e distribuição normal.

Uma estimativa para a variância $\sigma^2$ dos resíduos é:

$$\hat{\sigma^2} = \frac{SS_E}{n-2}$$

## Testes dos coeficientes $beta_0$ e $beta_1$ na regressão

Vamos testar que a inclinação $\beta_1$ do modelo de regressão é igual a algum valor constante $c$

As hipóteses ficarão:

$$H_0 : \beta_1 = c$$
$$H_1 : \beta_1 \neq c$$

A estatística de teste ficará então:

$$T_0 = \frac{\hat{\beta}_1 - c}{\sqrt{\hat{\sigma}^2/S_{xx}}}$$

Esta estatística segue uma distribuição $t$ com $n-2$ graus de liberdade sob a hipótese nula $H_0 : beta_1 = c$. Esta hipótese nula será rejeitada se o valor da estatística de teste for:

$$|t_0| = t_{\alpha/2,n-2}$$

Para o teste do coeficiente $\beta_0$ pode-se proceder da forma a seguir, em que $d$ é uma constante qualquer.

Hipóteses:

$$H_0 : \beta_0 = d$$
$$H_1 : \beta_0 \neq d$$

A estatística de teste fica:

$$T_0 = \frac{\hat{\beta}_0 - d}{\sqrt{\hat{\sigma}^2 \left[\frac{1}{n} + \frac{\bar{x}^2}{S_{XX}}\right]}}$$

Portanto a hipótese nula deve ser rejeitada se:

$$|t_0| = t_{\alpha/2,n-2}$$

## Coeficiente de determinação $R^2$

É uma medida de quão bem uma regressão descreve os dados.

$$R^2 = 1 - \frac{SS_E}{SS_T}$$