FORMULÁRIO DE ESTATÍSTICA II

- ## MODELO REGRESSÃO LINEAR (MRL)

$$y_i = \beta_0 + \beta_1 x_{i1} + \ldots + \beta_k x_{ik} + u_i\ ,\ i = 1,2,\ldots,n \Leftrightarrow y_i = \mathbf{x}_i \boldsymbol{\beta} + u_i\ ,\ i = 1,2,\ldots,n \quad \Leftrightarrow \quad \mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{u}$$

- **OLS (estimadores dos mínimos quadrados)**

Caso geral	**Caso particular:** $y_i = \beta_0 + \beta_1 x_i + u_i$
$\hat{\boldsymbol{\beta}} = (\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T\mathbf{y}$ $\hat{y}_i = \hat{\beta}_0 + \hat{\beta}_1 x_{i1} + \ldots + \hat{\beta}_k x_{ik}$ $\hat{u}_i = y_i - \hat{y}_i$ $\hat{\sigma}^2 = \dfrac{\sum \hat{u}_i^2}{(n-k-1)}$ $\hat{V}ar(\hat{\beta} \mid \mathbf{X}) = \hat{\sigma}^2 (\mathbf{X}^T\mathbf{X})^{-1}$ $\hat{V}ar(\hat{\beta}_j \mid \mathbf{X}) = \dfrac{\hat{\sigma}^2}{SST_j(1-R_j^2)}$	$\hat{\beta}_0 = \bar{y} - \hat{\beta}_1 \bar{x}$ $\hat{V}ar(\hat{\beta}_0 \mid X) = \dfrac{\hat{\sigma}^2 \sum x_i^2}{n\sum (x_i - \bar{x})^2}$ $\hat{\beta}_1 = \dfrac{\sum (x_i - \bar{x}) y_i}{\sum (x_i - \bar{x})^2}$ $\hat{V}ar(\hat{\beta}_1 \mid \mathbf{X}) = \dfrac{\hat{\sigma}^2}{\sum (x_i - \bar{x})^2}$ $\hat{\sigma}^2 = \dfrac{\sum \hat{u}_i^2}{n-2}$

<u>Nota</u>: R_j^2 - R^2 da regressão de x_j sobre todos os outros regressores; $SST_j = \sum_{i=1}^n (x_{ij} - \bar{x}_j)^2$.

Propriedades:

$\sum_{i=1}^n \hat{u}_i = 0$ (com termo independente); $\sum_{i=1}^n x_{ij}\hat{u}_i = 0$ $(j = 1,2,\ldots,k)$; $\sum_{i=1}^n \hat{y}_i \hat{u}_i = 0$;

$\sum_{i=1}^n \left(y_i - \bar{y}\right)^2 = \sum_{i=1}^n \left(\hat{y}_i - \bar{y}\right)^2 + \sum_{i=1}^n \hat{u}_i^2$ (com termo independente)

Coeficiente de Determinação:

$$R^2 = \frac{\left(\sum_i (y_i - \bar{y})(\hat{y}_i - \bar{\hat{y}})\right)^2}{\sum_i (y_i - \bar{y})^2 \sum_i (\hat{y}_i - \bar{\hat{y}})^2} = r_{y,\hat{y}}^2$$ com $r_{y,\hat{y}}$ o coeficiente de correlação entre y e $\hat{y}$

Coeficiente de Determinação no modelo com termo independente:

$$R^2 = \frac{SSE}{SST} = 1 - \frac{SSR}{SST}\ ;\ SST = \sum_{i=1}^n (y_i - \bar{y})^2\ ; SSE = \sum_{i=1}^n (\hat{y}_i - \bar{y})^2\ ;\ SSR = \sum_{i=1}^n \hat{u}_i^2$$

Coeficiente de Determinação ajustado:

$$\bar{R}^2 = 1 - \frac{SSR/(n-k-1)}{SST/(n-1)} = 1 - (1-R^2)\frac{n-1}{n-k-1}\ .$$

- **INFERÊNCIA ESTATÍSTICA DO MRL**:

  - $y_i \sim N(\mathbf{x}_i \boldsymbol{\beta}, \sigma^2)$; $u \sim N(0,\sigma^2)$; $\hat{\beta}_j \sim N\left(\beta_j, Var(\hat{\beta}_j)\right)$; $\dfrac{\sum_{i=1}^n \hat{u}_i^2}{\sigma^2} = \dfrac{(n-k-1)\hat{\sigma}^2}{\sigma^2} \sim \chi^2(n-k-1)$
  - $t_j = \dfrac{\hat{\beta}_j - \beta_j}{se(\hat{\beta}_j)} \sim t(n-k-1)$ ou $F_j = t_j^2 = \dfrac{\left(\hat{\beta}_j - \beta_j\right)^2}{se(\hat{\beta}_j)^2} \sim F(1, n-k-1)$

Casos particulares:

$H_0 : \beta_j = 0$	$H_0 : \beta_j = \beta_j^0$
$t_j = \dfrac{\hat{\beta}_j}{se(\hat{\beta}_j)} \sim t(n-k-1)$	$t_j = \dfrac{\hat{\beta}_j - \beta_j^0}{se(\hat{\beta}_j)} \sim t(n-k-1)$

- Testes de q restrições lineares sobre os coeficientes de regressão, $H_0 : \mathbf{R\beta} = \mathbf{r}$,

$$F = \frac{SSR_r - SSR_{ur}}{SSR_{ur}} \times \frac{n-k-1}{q} \sim F(q, n-k-1)$$

$$F = \frac{R_{ur}^2 - R_r^2}{1 - R_{ur}^{\;2}} \times \frac{n-k-1}{q} \sim F(q, n-k-1) \text{ (restrições homogéneas)}$$

Nota: SSR_r = soma dos quadrados dos resíduos do modelo com as q restrições lineares;
SSR_{ur} = soma dos quadrados dos resíduos do modelo sem restrições.
R_r^2 = Coeficiente de determinação do modelo com as q restrições
R_{ur}^2 = Coeficiente de determinação do modelo sem restrições

Casos particulares

- Uma única restrição (q=1) $H_0 : \mathbf{R\beta} = r \Leftrightarrow H_0 : \theta = 0$ com $\theta = \mathbf{R\beta} - r$,

$$t = \frac{R\hat{\boldsymbol\beta} - r}{\sqrt{Var\left(R\hat{\boldsymbol\beta}\right)}} \sim t(n-k-1) \quad \text{ou} \quad t = \frac{\hat\theta}{se(\hat\theta)} \sim t(n-k-1) \text{ com } \hat\theta = \mathbf{R}\hat{\boldsymbol\beta} - r$$

- $H_0 : \beta_1 = \beta_2 = ... = \beta_k = 0$

$$F = \frac{R^2}{1-R^2} \times \frac{n-k-1}{k} \sim F(k, n-k-1)$$

- **HETEROCEDASTICIDADE**:

$$Var(\hat{\boldsymbol\beta} \mid \mathbf{X}) = (\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T Var(\mathbf{u} \mid \mathbf{X})\mathbf{X}(\mathbf{X}^T\mathbf{X})^{-1} = (\mathbf{X}^T\mathbf{X})^{-1}\sum_{i=1}^{n}\sigma_i^2\mathbf{x}_i^T\mathbf{x}_i\,(\mathbf{X}^T\mathbf{X})^{-1}$$

  - **Estimador robusto de White** (estimador da variância consistente com heterocedasticidade)

$$\hat{V}ar(\hat{\boldsymbol\beta} \mid \mathbf{X}) = (\mathbf{X}^T\mathbf{X})^{-1}\sum_{i=1}^{n}\hat{u}_i^2\mathbf{x}_i^T\mathbf{x}_i\,(\mathbf{X}^T\mathbf{X})^{-1}.$$

  - **Inferência sobre** β_j:

$$t_j = \frac{\hat\beta_j - \beta_j}{se^*(\hat\beta_j)} \overset{a}{\sim} t(n-k-1) \quad \text{com } se^*(\hat\beta_j) = \text{erro-padrão consistente com heterocedasticidade}$$

  - **Testes de Heterocedasticidade:**

Estatística-teste LM: $\text{LM} = nR_{\hat u^2}^2 \overset{a}{\sim} \chi^2(p)$ com $R_{\hat u^2}^2$ e p respectivamente o coeficiente de determinação e o nº de regressores (excluindo o termo independente) da regressão auxiliar de teste.

- **PREVISÃO**

  - **Previsão em Média:** $E(y \mid x_1 = x_1^0, ..., x_k = x_k^0) = \beta_0 + \beta_1 x_1^0 + ... + \beta_k x_k^0 = \theta_0$; $\hat\theta_0 = \hat\beta_0 + \hat\beta_1 x_1^0 + ... + \hat\beta_k x_k^0$

$$t_{\hat\theta} = \frac{\hat\theta_0 - \theta_0}{se(\hat\theta_0)} \sim t(n-k-1)$$

  - **Previsão pontual:** $y^0 = \beta_0 + \beta_1 x_1^0 + ... + \beta_k x_k^0 + u^0$; $\hat{y}^0 = \hat\beta_0 + \hat\beta_1 x_1^0 + ... + \hat\beta_k x_k^0 = \hat\theta_0$

$$t = \frac{y^0 - \hat{y}^0}{\sqrt{se(\hat\theta_0)^2 + \hat\sigma^2}} \sim t(n-k-1)$$

- **TESTE RESET**

$y = \beta_0 + \beta_1 x_1 + ... + \beta_k x_k + \delta\hat{y}^2 + u$ - testar $H_0 : \delta = 0$

$y = \beta_0 + \beta_1 x_1 + ... + \beta_k x_k + \delta_1\hat{y}^2 + \delta_2\hat{y}^3 + u$ - testar $H_0 : \delta_1 = \delta_2 = 0$