- Assim, o MRLM (1) em termos matriciais é

$$\mathbf{Y}_{n\times 1} = \mathbf{X}_{n\times p}\boldsymbol{\beta}_{p\times 1} + \boldsymbol{\varepsilon}_{n\times 1}$$

onde $\boldsymbol{\varepsilon}$ é um vetor de variáveis aleatórias normais independentes com:

$$E(\boldsymbol{\varepsilon}) = \mathbf{0}$$
$$V(\boldsymbol{\varepsilon}) = \sigma^2\mathbf{I}$$

- ***Conseqüências:***

$$E(\mathbf{Y}) = E(\mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}) = \mathbf{X}\boldsymbol{\beta} + E(\boldsymbol{\varepsilon}) = \mathbf{X}\boldsymbol{\beta}$$

$$V(\mathbf{Y}) = V(\mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}) = \mathbf{0} + V(\boldsymbol{\varepsilon}) = \sigma^2\mathbf{I}$$

$$\hat{\mathbf{Y}}_{n\times 1} = \mathbf{H}_{n\times n}\mathbf{Y}_{n\times 1}$$

onde:

$$\mathbf{H}_{n\times n} = \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'$$

Similarmente, o vetor de resíduos pode ser expresso como:

$$\mathbf{e}_{n\times 1} = (\mathbf{I} - \mathbf{H})\mathbf{Y}$$

e a matriz de covariâncias do resíduo é dada por:

$$V(\mathbf{e})_{n\times n} = \sigma^2(\mathbf{I} - \mathbf{H})$$

- O ***coeficiente de determinação*** é definido como:

$$R^2 = \frac{SSR}{SSTO} = 1 - \frac{SSE}{SSTO}$$

- ***Coeficiente de Determinação Ajustado:***

$$R_a^2 = 1 - \frac{\frac{SSE}{n-p}}{\frac{SSTO}{n-1}} = 1 - \left(\frac{n-1}{n-p}\right)\frac{SSE}{SSTO}$$

- ***Coeficiente de Correlação Múltipla***

$$R = \sqrt{R^2}$$

- **Hipóteses:**

$$\begin{cases} H_0)\beta_k = 0 \\ H_1)\beta_k \neq 0 \quad (ou\ \beta_k < 0\ ou\ \beta_k > 0) \end{cases}$$

- **Estatística do Teste:**

$$t = \frac{\hat{\beta}_k}{S\{\hat{\beta}_k\}} \sim Student\ com\ (n-p)\ g.l. \quad \forall k = 1, 2, \ldots, p-1$$

$$\hat{Y}_0 = \mathbf{X}_0'\hat{\boldsymbol{\beta}}$$

O estimador $\hat{Y}_0$ é não tendencioso:

$$E(\hat{Y}_0) = \mathbf{X}_0'\boldsymbol{\beta} = E(Y_0)$$

$$\left(\hat{Y}_h \mp W\, S\{\hat{Y}_h\}\right)$$

onde:

$$W^2 = p\, F_{1-\alpha:(p;n-p)}$$

ou ainda:

$$\left(\hat{Y}_h \mp W\sqrt{MSE\left(\mathbf{X}_h'(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}_h\right)}\right)$$

$$\hat{\boldsymbol{\beta}} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{Y}$$

Em notação matricial, tem-se então:

$$\hat{\mathbf{Y}}_{n\times 1} = \mathbf{X}_{n\times 2}\hat{\boldsymbol{\beta}}_{2\times 1}$$

e o resíduo:

$$\underbrace{\mathbf{e}}_{n\times 1} = \underbrace{\mathbf{Y}}_{n\times 1} - \underbrace{\hat{\mathbf{Y}}}_{n\times 1} = \underbrace{\mathbf{Y}}_{n\times 1} - \underbrace{\mathbf{X}\hat{\boldsymbol{\beta}}}_{}$$

Teste de Ausência de Regressão

19 • **O TESTE:**

$$\begin{cases} H_0)\text{Ausência de regressão} \\ H_1)\text{Existência de regressão} \end{cases} \Leftrightarrow \begin{cases} H_0)\beta_1 = \beta_2 = \ldots = \beta_{p-1} = 0 \\ H_1)\text{Existe pelo menos um } \beta_j \neq 0 \end{cases}$$

Fontes de Variação	SS	GL	MS	F
Regressão	$SSR = \hat{\boldsymbol{\beta}}'\mathbf{X}'\mathbf{Y} - \frac{1}{n}\mathbf{Y}'\mathbf{J}\mathbf{Y}$	*p-1*	$MSR = \frac{SSR}{p-1}$	$F_c = \frac{MSR}{MSE}$
Resíduo	$SSE = \mathbf{Y}'\mathbf{Y} - \hat{\boldsymbol{\beta}}'\mathbf{X}'\mathbf{Y}$	*n-p*	$MSE = \frac{SSE}{n-p}$	
TOTAL	$SSTO = \mathbf{Y}'\mathbf{Y} - \frac{1}{n}\mathbf{Y}'\mathbf{J}\mathbf{Y}$	*n-1*		

$$\hat{\sigma}^2 = \frac{\mathbf{Y}'\mathbf{Y} - \hat{\boldsymbol{\beta}}'\mathbf{X}'\mathbf{Y}}{n-p}$$

$$\beta_k \in \left(\hat{\beta}_k \mp t_{1-\alpha/2}S\{\hat{\beta}_k\}\right) \quad \forall k = 1, 2, \ldots, p-1$$

$$onde: \ S\{\hat{\beta}_k\} = \sqrt{V(\hat{\beta}_k)}$$

$$\hat{\boldsymbol{\beta}} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{Y} = \mathbf{C}\mathbf{Y}$$

$$\therefore E(\hat{\boldsymbol{\beta}}) = \boldsymbol{\beta}$$

$$\therefore V(\hat{\boldsymbol{\beta}}) = \sigma^2(\mathbf{X}'\mathbf{X})^{-1}$$

$$\beta_k \in \left(\hat{\beta}_k \mp B\, S\{\hat{\beta}_k\}\right) \quad \forall k = 1, 2, \ldots, p-1$$

$$onde: \quad B = t_{1-\alpha/2g}$$

$$V(\hat{Y}_0) = \mathbf{X}_0'\mathbf{V}(\hat{\boldsymbol{\beta}})\mathbf{X}_0$$

$$S^2\{\hat{Y}_0\} = MSE\ \mathbf{X}_0'(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}_0 = \mathbf{X}_0'\hat{V}(\hat{\boldsymbol{\beta}})\mathbf{X}_0$$

$$\left(\hat{Y}_0 \mp t_{\alpha/2}\sqrt{MSE\ (\mathbf{X}_0'(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}_0)}\right)$$

1. Usar a ***Região de Confiança*** $(1-\alpha)$ Working-Hotteling para a ***superfície de regressão:***

$$\left(\hat{Y}_h \mp W\, S\{\hat{Y}_h\}\right)$$

onde:

$$W^2 = p\, F_{1-\alpha:(p;n-p)}$$

2. Intervalos de Confiança simultâneos de Bonferroni. Para *g* intervalos simultâneos:

$$\left(\hat{Y}_h \mp B\, S\{\hat{Y}_h\}\right)$$

onde:

$$B = t_{1-\alpha/2g\,:\,(n-p)}$$

$$\left(\hat{Y}_h \mp t_{1-\alpha/2}\ s\{media\ pred\}\right)$$

onde:

$$s^2\{media\ pred\} = \frac{MSE}{m} + s^2\{\hat{Y}_h\} = MSE\left(\frac{1}{m} + \mathbf{X}_h'(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}_h\right)$$

1. Usar limites de predição de Schffé para **g** novas observações em **g** diferentes níveis $\mathbf{X}_h$

$$\left(\hat{Y}_h \mp S\ \sqrt{MSE\left(1+\mathbf{X}_h'(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}_h\right)}\right)$$

onde:

$$S = g\,F_{(1-\alpha\,;\,g,\,n-p)}$$

2. Intervalos de Confiança simultâneos de Bonferroni. Para *g* intervalos simultâneos:

$$\left(\hat{Y}_h \mp B\ \sqrt{MSE\left(1+\mathbf{X}_h'(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}_h\right)}\right)$$

onde:

$$B = t_{\ (1-\alpha/2g\,;\,n-p)}$$

43

$$SSR(X_2|X_1) = SSE(X_1) - SSE(X_1, X_2)$$

ou

$$SSR(X_2|X_1) = SSR(X_1, X_2) - SSR(X_1)$$

- Assim:

$$SSR(X_1,X_2,X_3) = SSR(X_1) + SSR(X_2|X_1) + SSR(X_3|X_1,X_2)$$

- Ou em outra ordem:

$$\boxed{SSR(X_1,X_2,X_3) = SSR(X_3) + SSR(X_1|X_3) + SSR(X_2|X_1,X_3)}$$

Fontes de Variação	SS	GL	MS
Regressão	$SSR(X_1,X_2,X_3)$	*3*	$MSR(X_1,X_2,X_3)$
X_1	$SSR(X_1)$	*1*	$MSR(X_1)$
$X_2\|X_1$	$SSR(X_2\|X_1)$	*1*	$MSR(X_2\|X_1)$
$X_3\|X_1,X_2$	$SSR(X_3\|X_1,X_2)$	*1*	$MSR(X_3\|X_1,X_2)$
Resíduo	$SSE(X_1,X_2,X_3)$	*n-4*	$MSE(X_1,X_2,X_3)$
TOTAL	$SSTO$	*n-1*	

hipóteses:

$$\begin{cases} (H_0)\,\beta_k = 0 \\ (H_1)\,\beta_k \neq 0 \end{cases}$$

- E a estatística do teste usada é:

$$t = \frac{\hat{\beta}_k}{s\{\hat{\beta}_k\}}$$

$$R^2_{Y1|23} = \frac{SSR(X_1 \mid X_2, X_3)}{SSE(X_2, X_3)}$$

$$\boxed{\frac{Y_i - \bar{Y}}{S_Y}, \quad onde \quad S_Y = \sqrt{\frac{\sum_{i=1}^{n}(Y_i - \bar{Y})^2}{n-1}}}$$

$$\boxed{\frac{X_{ik} - \bar{X}_k}{S_k} \quad (k = 1,2,\ldots,p-1) \qquad onde \quad S_Y = \sqrt{\frac{\sum_{i=1}^{n}(X_{ik} - \bar{X}_k)^2}{n-1}}}$$

- A ***transformação correlação*** é dada por:

$$\boxed{Y_i^* = \frac{1}{\sqrt{n-1}}\left(\frac{Y_i - \bar{Y}}{S_Y}\right)}$$

$$\boxed{X_{ik}^* = \frac{1}{\sqrt{n-1}}\left(\frac{X_{ik} - \bar{X}_k}{S_k}\right) \quad (k = 1,2,\ldots,p-1)}$$

$$Y_i^* = \beta_1^* X_{i1}^* + \beta_2^* X_{i2}^* + \ldots + \beta_{p-1}^* X_{i,p-1}^* + \varepsilon_i^*$$

$$\beta_k = \left(\frac{S_Y}{S_k}\right)\beta_k^* \quad k = 1,2,\ldots,p-1$$

$$\beta_0 = \bar{Y} - \beta_1\bar{X}_1 - \beta_2\bar{X}_2 - \ldots - \beta_{p-1}\bar{X}_{p-1}$$

$$\mathbf{X}'\mathbf{X} = \mathbf{r}_{\mathbf{XX}} = \begin{bmatrix} 1 & r_{12} & \cdots & r_{1,p-1} \\ r_{21} & 1 & \cdots & r_{2,p-1} \\ \vdots & \vdots & & \vdots \\ r_{p-1,1} & r_{p-1,2} & \cdots & 1 \end{bmatrix}$$

$$\mathbf{r}_{\mathbf{XX}}\mathbf{b} = \mathbf{r}_{\mathbf{YX}} \quad \Rightarrow \mathbf{b} = \mathbf{r}_{\mathbf{XX}}^{-1}\mathbf{r}_{\mathbf{YX}}$$

$$\boxed{R_p^2 = 1 - \frac{SSE_p}{SSTO}}$$

$$\boxed{R_{a,p}^2 = 1 - \left(\frac{n-1}{n-p}\right)\frac{SSE_p}{SSTO} = 1 - \frac{MSE_p}{\frac{SSTO}{n-1}}}$$

$$C_p = \frac{SSE_p}{MSE} - (n-2p)$$

- ***Critério da Informação de Akaike*** (AIC_p):

$$AIC_p = n \ln SSE_p - n \ln n + 2p$$

$$Y_i - \hat{Y}_{i(i)}$$

E

$$PRESS_p = \sum_{i=1}^{n}\left(Y_i - \hat{Y}_{i(i)}\right)^2$$

- ***Critério Bayesiano de Schwarz*** (SBC_p):

$$SBC_p = n \ln SSE_p - n \ln n + (\ln n)p$$

□ Teste de Falta de Ajustamento

- ***Hipótese:***

$$\begin{cases} H_o)E(Y) = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \ldots + \beta_{p-1} X_{p-1} \\ H_1)E(Y) \neq \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \ldots + \beta_{p-1} X_{p-1} \end{cases}$$

- ***Estatística do Teste***

$$F^* = \frac{MSLF}{MSPE}$$

- ***Resíduos semistudentizados***

$$e_i^* = \frac{e_i - \bar{e}}{\sqrt{MSE}} = \frac{e_i}{\sqrt{MSE}}$$

- ***Resíduos studentizados (ou resíduos studentizados internamente)***

$$r_i = \frac{e_i}{S\{e_i\}}$$

Resíduo excluído:

$$\begin{cases} S^2\{e_i\} = MSE(1-h_{ii}) \\ S\{e_i, e_j\} = -h_{ij} MSE \qquad i \neq j \end{cases}$$

$$d_i = \frac{e_i}{1-h_{ii}} \qquad S^2\{d_i\} = \frac{MSE_{(i)}}{1-h_{ii}}$$

- ***Resíduos ´excluído´ studentizado (ou resíduos studentizados externamente)***

$$t_i = \frac{d_i}{S\{d_i\}}$$

Ou quivalentemente:

:

$$t_i = \frac{e_i}{\sqrt{MSE_{(i)}(1-h_{ii})}} \quad ou \quad t_i = e_i\left[\frac{n-p-1}{\sqrt{SSE(1-h_{ii}) - e_i^2}}\right]^{1/2}$$

Teste de Valores Discrepantes

$$H_0)O \quad i-esimo\ residuo\ nao\ e\ um\ valor\ discrepante$$

$$H_1)O \quad i-esimo\ residuo\ e\ um\ valor\ discrepante$$

•**Estatística do Teste:**

$$t_i = e_i\left[\frac{n-p-1}{\sqrt{SSE(1-h_{ii}) - e_i^2}}\right]^{1/2} : \ Student\ (n-p-1)g.l.$$

•**Regra de decisão:**

$$Se\ |t_i| \geq t(1-\alpha/2n; n-p-1) \Rightarrow rejeitar\ H_0$$

$$(DFFITS)_i = \frac{\hat{Y}_i - \hat{Y}_{i(i)}}{\sqrt{MSE_{(i)} h_{ii}}} \qquad (DFFITS)_i = e_i\left[\frac{n-p-1}{SSE(1-h_{ii}) - e_i^2}\right]^{1/2}\left(\frac{h_{ii}}{1-h_{ii}}\right)^{1/2} = t_i\left(\frac{h_{ii}}{1-h_{ii}}\right)^{1/2}$$

Distância de cook

$$D_i = \frac{(\hat{\mathbf{Y}} - \hat{\mathbf{Y}}_{(i)})'(\hat{\mathbf{Y}} - \hat{\mathbf{Y}}_{(i)})}{pMSE} \qquad D_i = \frac{e_i^2}{p\ MSE}\left(\frac{h_{ii}}{(1-h_{ii})^2}\right)$$

$$(DFBETAS)_i = \frac{\hat{\beta}_k - \hat{\beta}_{k(i)}}{\sqrt{MSE_{(i)} c_{kk}}} \quad k = 0,1,2,\dots,p-1$$

Fator de inflação da variância (VIF)

$$\boxed{Y_i^* = \frac{1}{\sqrt{n-1}} \left(\frac{Y_i - \bar{Y}}{S_Y} \right)}$$

$$\boxed{X_{ik}^* = \frac{1}{\sqrt{n-1}} \left(\frac{X_{ik} - \bar{X}_k}{S_k} \right) \quad (k = 1, 2, \dots, p-1)}$$

$$\boxed{V\left(\hat{\beta}_k^*\right) = \frac{\left(\sigma^*\right)^2}{1 - R_k^2} \quad k = 1,2,\dots,p-1}$$

$$\boxed{E\left[\sum_{k=1}^{p-1}\left(\hat{\beta}_k^* - \beta_k^*\right)^2\right] = \left(\sigma^*\right)^2 \times \sum_{k=1}^{p-1}(VIF)_k \qquad (1)}$$

- A razão (1) / (2) nos dá:

$$\overline{(VIF)} = \frac{\sum_{k=1}^{p-1}(VIF)_k}{k}$$

$$\boxed{E\left[\sum_{k=1}^{p-1}\left(\hat{\beta}_k^* - \beta_k^*\right)^2\right] = \left(\sigma^*\right)^2 \times k \qquad (2)}$$

- Se $\overline{(VIF)} \gg 1$ ➡ ***MULTICOLINEARIDADE INFLUENCIANDO ESTIMATIVAS MQ***

- **Segunda Ordem:**

$$Y_i = \beta_0 + \beta_1 x_i + \beta_2 x_i^2 + \varepsilon_i$$

onde: $x_i = X_i - \bar{X}$ **Modelo d**

- **Terceira Ordem:**

$$Y_i = \beta_0 + \beta_1 x_i + \beta_{11} x_i^2 + \beta_{111} x_i^3 + \varepsilon_i$$

onde: $x_i = X_i - \bar{X}$ **Modelo de Ter**