

Modelo normal linear multivariado

Prof. Caio Azevedo

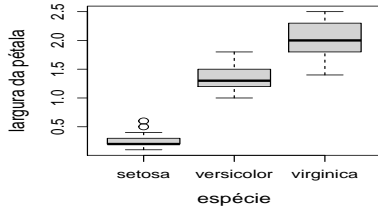
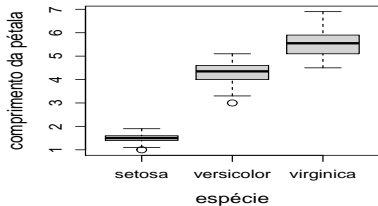
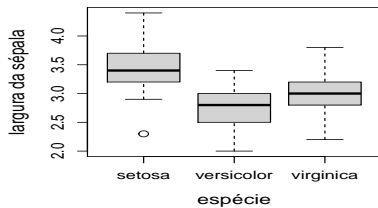
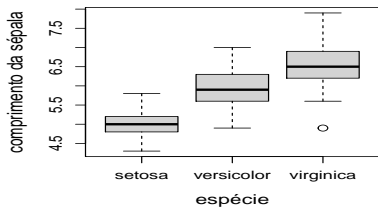
Recapitulando: dados das Iris de Fisher (veja)

- Os dados consistem de 50 unidades amostrais de três espécies (setosa, virginica, versicolor) de íris (uma espécie de planta), ou seja, temos um total de 150 unidades amostrais.
- De cada uma delas mediu-se quatro variáveis: comprimento e largura da sépala (CS, LS) e comprimento e largura da pétala (CP,LP).
- Objetivo original: quantificar a variação morfológica em relação à essas espécies com bases nas quatro variáveis de interesse.

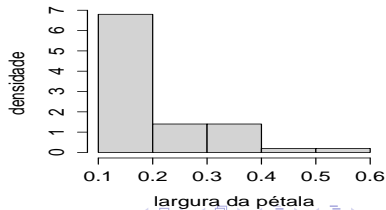
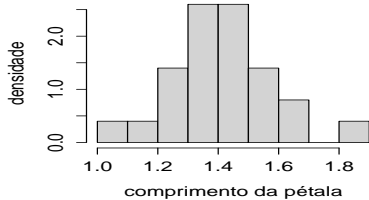
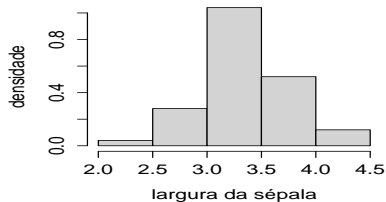
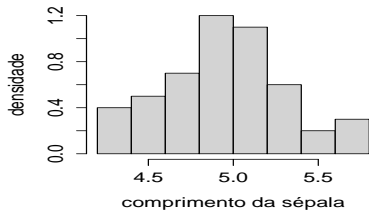
Modelagem dos dados

- Seja Y_{ijk} : o valor da k -ésima variável ($k=1,2,3,4$), para a j -ésimo unidade experimental (planta) ($j=1,\dots,50$) do i -ésimo grupo ($i = 1, 2, 3$).
- Suposição $\mathbf{Y}_{ij} = (Y_{ij1}, Y_{ij2}, Y_{ij3}, Y_{ij4}) \stackrel{ind.}{\sim} N_4(\boldsymbol{\mu}_i, \boldsymbol{\Sigma})$.
- Tem-se, portanto, dados balanceados (em relação aos grupos).

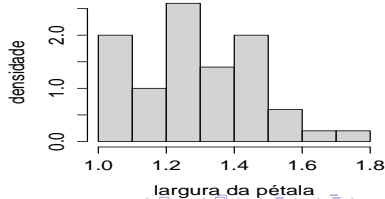
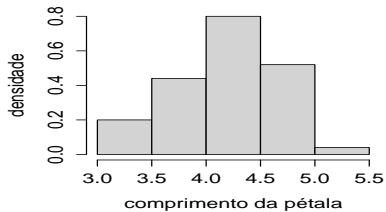
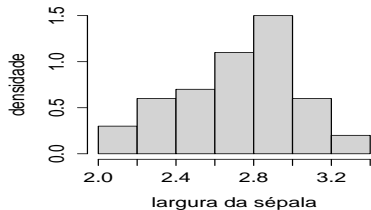
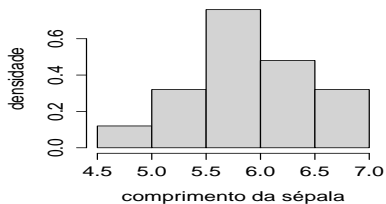
Boxplot das variáveis por grupo



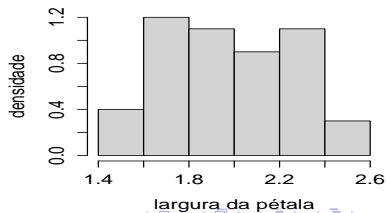
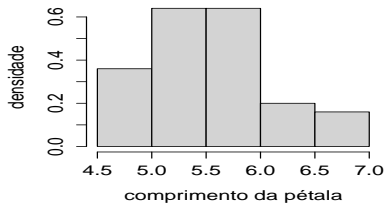
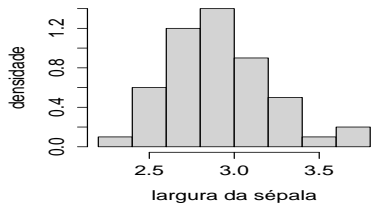
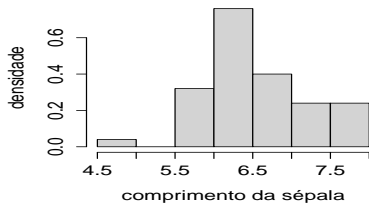
Histograma das variáveis: iris setosa



Histograma das variáveis: iris versicolor



Histograma das variáveis: iris virginica



Matriz de dispersão: ■ - S, ● - Ver, ▲ - Vir

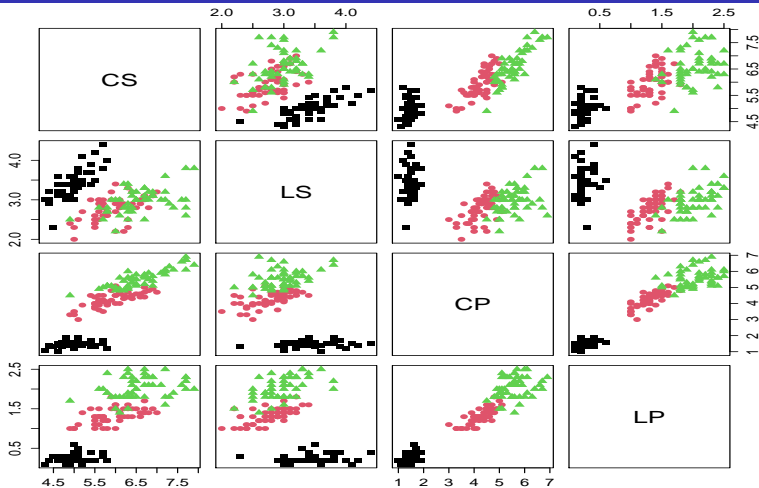


Diagrama de dispersão: setosa

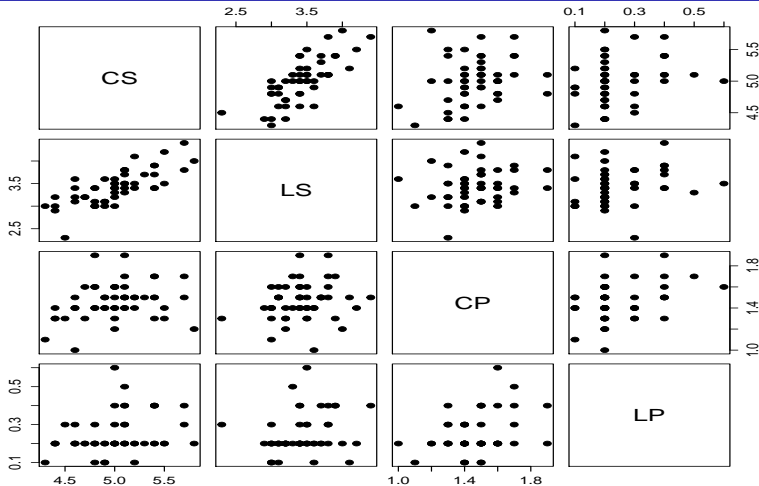


Diagrama de dispersão: versicolor

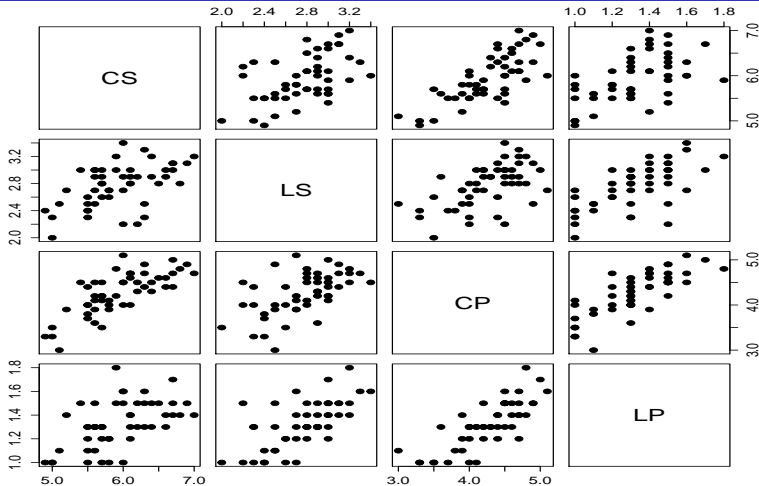
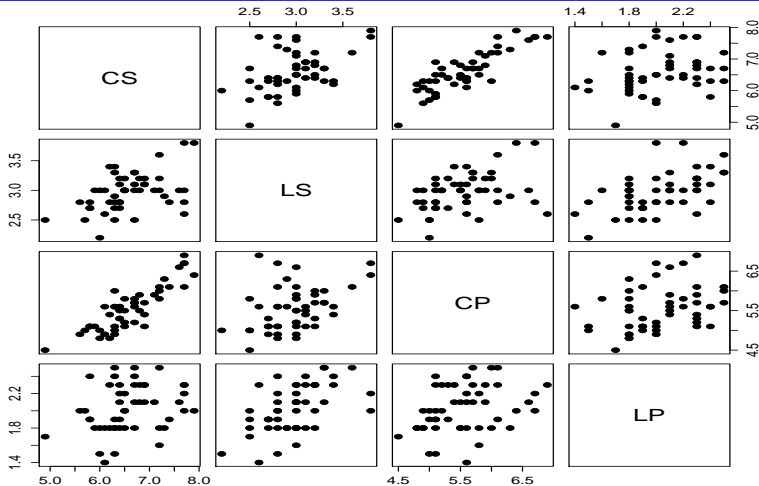


Diagrama de dispersão: vrginica



Variâncias na diagonal principal (dp), covariâncias abaixo e correlações acima (da dp)

$$\tilde{\Psi}_{\text{setosa}} = \begin{bmatrix} CS & 0,12 & 0,74 & 0,27 & 0,28 \\ LS & 0,10 & 0,14 & 0,18 & 0,23 \\ CP & 0,02 & 0,01 & 0,03 & 0,33 \\ LP & 0,01 & 0,01 & 0,01 & 0,01 \end{bmatrix}$$

$$\tilde{\Psi}_{\text{versicolor}} = \begin{bmatrix} CS & 0,27 & 0,53 & 0,75 & 0,55 \\ LS & 0,08 & 0,10 & 0,56 & 0,66 \\ CP & 0,18 & 0,08 & 0,22 & 0,79 \\ LP & 0,06 & 0,04 & 0,07 & 0,04 \end{bmatrix}$$

$$\tilde{\Psi}_{\text{virginica}} = \begin{bmatrix} CS & 0,40 & 0,46 & 0,86 & 0,28 \\ LS & 0,09 & 0,10 & 0,40 & 0,54 \\ CP & 0,30 & 0,07 & 0,30 & 0,32 \\ LP & 0,05 & 0,05 & 0,05 & 0,07 \end{bmatrix}$$

Medidas resumo: comprimento da sépala

Espécie	Média	DP	Var	CV(%)	Mínimo	Mediana	Máximo	n
setosa	5,01	0,35	0,12	7,04	4,30	5,00	5,80	50
versicolor	5,94	0,52	0,27	8,70	4,90	5,90	7,00	50
virginica	6,59	0,64	0,40	9,65	4,90	6,50	7,90	50

Medidas resumo: largura da sépala

Espécie	Média	DP	Var	CV(%)	Mínimo	Mediana	Máximo	n
setosa	3,43	0,38	0,14	11,06	2,30	3,40	4,40	50
versicolor	2,77	0,31	0,10	11,33	2,00	2,80	3,40	50
virginica	2,97	0,32	0,10	10,84	2,20	3,00	3,80	50

Medidas resumo: comprimento da pétala

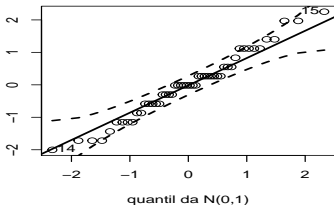
Espécie	Média	DP	Var.	CV(%)	Mínimo	Mediana	Máximo	n
setosa	1,46	0,17	0,03	11,88	1,00	1,50	1,90	50
versicolor	4,26	0,47	0,22	11,03	3,00	4,35	5,10	50
virginica	5,55	0,55	0,30	9,94	4,50	5,55	6,90	50

Medidas resumo: largura da pétala

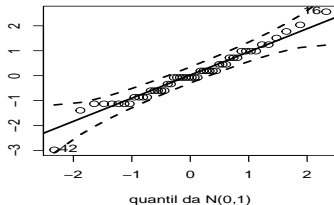
Espécie	Média	DP	Var.	CV(%)	Mínimo	Mediana	Máximo	n
setosa	0,25	0,11	0,01	42,84	0,10	0,20	0,60	50
versicolor	1,33	0,20	0,04	14,91	1,00	1,30	1,80	50
virginica	2,03	0,27	0,08	13,56	1,40	2,00	2,50	50

Gráficos de quantis-quantis com envelopes: setosa

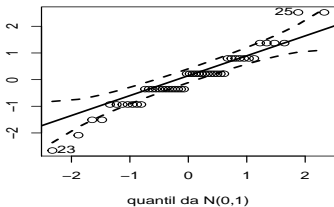
quantil da distribuição da comprimento da sépala



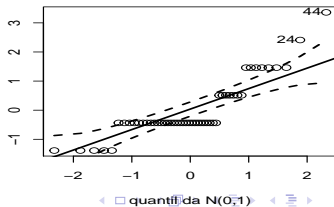
quantil da distribuição da largura da sépala



quantil da distribuição da comprimento da pétala

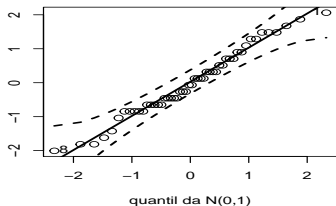


quantil da distribuição da largura da pétala

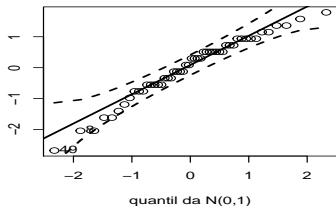


Gráficos de quantis-quantis com envelopes: versicolor

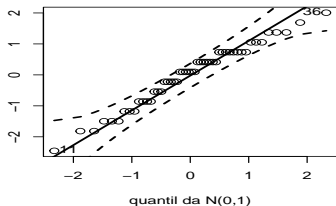
quantil da distribuição da comprimento da sépala



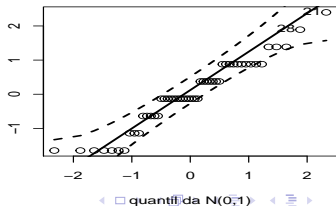
quantil da distribuição da comprimento da pétala



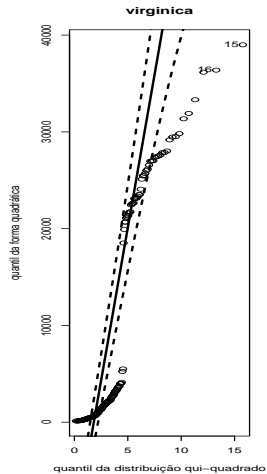
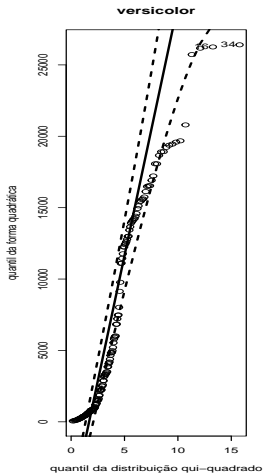
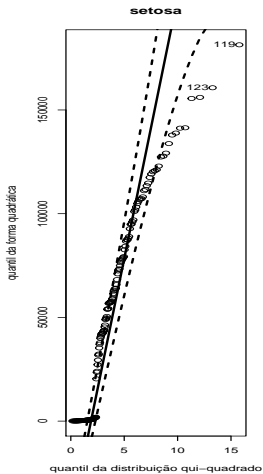
quantil da distribuição da largura da sépala



quantil da distribuição da largura da pétala



Gráficos de Q-Q com envelopes: forma quadrática



Matriz de dados: iris de Fisher

$$Y_{(n \times p)} = \begin{bmatrix} Y_{111} & Y_{112} & Y_{113} & Y_{114} \\ Y_{121} & Y_{122} & Y_{123} & Y_{124} \\ \vdots & \vdots & \ddots & \vdots \\ Y_{1(50)1} & Y_{1(50)2} & Y_{1(50)3} & Y_{1(50)4} \\ \text{---} & \text{---} & \text{---} & \text{---} \\ Y_{211} & Y_{212} & Y_{213} & Y_{214} \\ Y_{221} & Y_{222} & Y_{223} & Y_{224} \\ \vdots & \vdots & \ddots & \vdots \\ Y_{2(50)1} & Y_{2(50)2} & Y_{2(50)3} & Y_{2(50)4} \\ \text{---} & \text{---} & \text{---} & \text{---} \\ Y_{311} & Y_{312} & Y_{313} & Y_{314} \\ Y_{321} & Y_{322} & Y_{323} & Y_{324} \\ \vdots & \vdots & \ddots & \vdots \\ Y_{3(50)1} & Y_{3(50)2} & Y_{3(50)3} & Y_{3(50)4} \end{bmatrix}$$

Suposições

- Suponha um conjunto de G populações independentes da qual retiramos G amostras de tamanho n_i , $i = 1, \dots, G$,
- Por suposição, temos que $\mathbf{Y}_{ij} \sim N_p(\boldsymbol{\mu}_i, \boldsymbol{\Sigma})$, em que $i = 1, 2, \dots, G$ (grupo) e $j = 1, 2, \dots, n_i$ (indivíduo). Notação: Y_{ijk} observação referente à variável k do indivíduo j do grupo i .
- Homocedasticidades: $\boldsymbol{\Sigma}_1 = \boldsymbol{\Sigma}_2 = \dots = \boldsymbol{\Sigma}_G = \boldsymbol{\Sigma}$.
- Assim, temos a seguinte matriz de dados ($n = \sum_{i=1}^G n_i$):

$$\mathbf{Y}_{(n \times p)} = \begin{bmatrix}
 Y_{111} & Y_{112} & \dots & Y_{11p} \\
 Y_{121} & Y_{122} & \dots & Y_{12p} \\
 \vdots & \vdots & \ddots & \vdots \\
 Y_{1n_1 1} & Y_{1n_1 2} & \dots & Y_{1n_1 p} \\
 \hline
 Y_{211} & Y_{212} & \dots & Y_{21p} \\
 Y_{221} & Y_{222} & \dots & Y_{22p} \\
 \vdots & \vdots & \ddots & \vdots \\
 Y_{2n_2 1} & Y_{2n_2 2} & \dots & Y_{2n_2 p} \\
 \hline
 \vdots & \vdots & \ddots & \vdots \\
 \hline
 Y_{G11} & Y_{G12} & \dots & Y_{G1p} \\
 Y_{G21} & Y_{G22} & \dots & Y_{G2p} \\
 \vdots & \vdots & \ddots & \vdots \\
 Y_{Gn_G 1} & Y_{Gn_G 2} & \dots & Y_{Gn_G p}
 \end{bmatrix}$$

Hipóteses de interesse

- Queremos testar $H_0 : \mu_1 = \mu_2 = \dots = \mu_G$ vs H_1 : há pelo menos uma diferença.
- Uma abordagem: análise de variância multivariada (MANOVA).
- Comparar médias através do estudo da decomposição da matriz de variâncias-covariâncias total.
- Como resumir a informação das matrizes de covariâncias de interesse? Uma possibilidade: variâncias generalizadas.
- À rigor, a metodologia MANOVA está acomodada na classe de modelos de regressão normais lineares multivariados, a qual é uma extensão dos respectivos modelos univariados ([link](#)).

Modelo de regressão normal linear multivariado (MRNLM)

$$\mathbf{Y}_{(n \times p)} = \mathbf{X}_{(n \times q)} \mathbf{B}_{(q \times p)} + \boldsymbol{\xi}_{(n \times p)}$$

- $\mathbf{Y}_{(n \times p)}$: matriz de dados
- $\mathbf{X}_{(n \times q)}$: matriz de planejamento, conhecida e não-aleatória.
- $\mathbf{B}_{(q \times p)}$: parâmetros de regressão, desconhecido e não aleatório.
- $\boldsymbol{\xi}_{(n \times p)}$: matriz de erros, sendo que cada uma das suas linhas corresponde a uma vetor aleatório tal que $N_p(0, \boldsymbol{\Sigma})$, mutuamente independentes.
- Tal estrutura se aplica a diferentes situações, não somente em situação como a dos dados da iris de Fisher.

MRNLM (cont.)

- As suposições do MRNLM são:
 - Em relação aos erros associados a cada unidade amostral/experimental (linhas da matriz de dados):
 - Normalidade multivariada.
 - Matriz de variâncias e covariâncias constantes, conhecida como homocedasticidade.
 - Independência entre elas.
 - Linearidade entre a média $\mathcal{E}(\mathbf{Y})$ e as covariáveis \mathbf{X} , ou seja $\mathcal{E}(\mathbf{Y}) = \mathbf{XB}$.
- Em geral o MRNLM não é robusto a violação de pelo menos uma das suposições descritas acima ([artigo 1](#), [artigo 2](#))

Matriz de planejamento de matriz de parâm. de regressão

$$\mathbf{X}_{(n \times q)} = \begin{bmatrix} X_{111} & X_{112} & \dots & X_{11G} \\ X_{121} & X_{122} & \dots & X_{12G} \\ \vdots & \vdots & \vdots & \vdots \\ X_{1n_1 1} & X_{1n_1 2} & \dots & X_{1n_1 G} \\ \text{---} & \text{---} & \text{---} & \text{---} \\ \vdots & \vdots & \vdots & \vdots \\ \text{---} & \text{---} & \text{---} & \text{---} \\ X_{G11} & X_{G12} & \dots & X_{G1G} \\ X_{G21} & X_{G22} & \dots & X_{G2G} \\ \vdots & \vdots & \vdots & \vdots \\ X_{Gn_G 1} & X_{Gn_G 2} & \dots & X_{Gn_G G} \end{bmatrix}; \mathbf{B}_{(q \times p)} = \begin{bmatrix} \mu_{11} & \mu_{12} & \dots & \mu_{1p} \\ \mu_{21} & \mu_{22} & \dots & \mu_{2p} \\ \vdots & \vdots & \vdots & \vdots \\ \mu_{G1} & \mu_{G2} & \dots & \mu_{Gp} \end{bmatrix}$$

$$\xi_{(n \times p)} = \begin{bmatrix} \xi_{111} & \xi_{112} & \dots & \xi_{11p} \\ \xi_{121} & \xi_{122} & \dots & \xi_{12p} \\ \vdots & \vdots & \ddots & \vdots \\ \xi_{1n_11} & \xi_{1n_12} & \dots & \xi_{1n_1p} \\ \hline \xi_{211} & \xi_{212} & \dots & \xi_{21p} \\ \xi_{221} & \xi_{222} & \dots & \xi_{22p} \\ \vdots & \vdots & \ddots & \vdots \\ \xi_{2n_21} & \xi_{2n_22} & \dots & \xi_{2n_2p} \\ \hline \vdots & \vdots & \ddots & \vdots \\ \hline \xi_{G11} & \xi_{G12} & \dots & \xi_{G1p} \\ \xi_{G21} & \xi_{G22} & \dots & \xi_{G2p} \\ \vdots & \vdots & \ddots & \vdots \\ \xi_{Gn_G1} & \xi_{Gn_G2} & \dots & \xi_{Gn_Gp} \end{bmatrix}$$

Voltando aos dados da Iris de Fisher:

- $G = 3$, $p = 4$ e $n_i = 50$, $i = 1, 2, 3$ (3 grupos, 4 variáveis e 50 indivíduos por grupo).
- Objetivo: modelar o vetor de médias de cada grupo e compará-los entre si.
- $\mathbf{Y}_{(150 \times 4)}$.

Cont.: Parametrização de médias

- $\mathbf{X} = \mathbf{I}_{(3 \times 3)} \otimes \mathbf{1}_{(50 \times 1)}$ (\otimes denota o produto de Kronecker à

esquerda). $\mathbf{X} = \begin{bmatrix} \mathbf{1}_{(50 \times 1)} & \mathbf{0}_{(50 \times 1)} & \mathbf{0}_{(50 \times 1)} \\ \mathbf{0}_{(50 \times 1)} & \mathbf{1}_{(50 \times 1)} & \mathbf{0}_{(50 \times 1)} \\ \mathbf{0}_{(50 \times 1)} & \mathbf{0}_{(50 \times 1)} & \mathbf{1}_{(50 \times 1)} \end{bmatrix}$

- $\mathbf{B} = \begin{bmatrix} \mu_{11} & \mu_{12} & \mu_{13} & \mu_{14} \\ \mu_{21} & \mu_{22} & \mu_{23} & \mu_{24} \\ \mu_{31} & \mu_{32} & \mu_{33} & \mu_{34} \end{bmatrix}$

- μ_{ik} : média da variável k do grupo i .

Cont.: Parametrização casela de referência

$$\blacksquare \mathbf{X} = \begin{bmatrix} \mathbf{1}_{(50 \times 1)} & \mathbf{0}_{(50 \times 1)} & \mathbf{0}_{(50 \times 1)} \\ \mathbf{1}_{(50 \times 1)} & \mathbf{1}_{(50 \times 1)} & \mathbf{0}_{(50 \times 1)} \\ \mathbf{1}_{(50 \times 1)} & \mathbf{0}_{(50 \times 1)} & \mathbf{1}_{(50 \times 1)} \end{bmatrix}$$

$$\blacksquare \mathbf{B} = \begin{bmatrix} \mu_{11} & \mu_{12} & \mu_{13} & \mu_{14} \\ \alpha_{21} & \alpha_{22} & \alpha_{23} & \alpha_{24} \\ \alpha_{31} & \alpha_{32} & \alpha_{33} & \alpha_{34} \end{bmatrix}$$

- μ_{1k} : média da variável k do grupo 1 (grupo de referência - setosa).
- $\alpha_{ik} = \mu_{ik} - \mu_{1k}$, $i = 2, 3$: é a diferença entre a média do grupo i e a média do grupo 1, em relação a variável k .

Decomposição da matriz de covariâncias total

- Pode-se demonstrar, para o exemplo em questão, que:

$$\underbrace{\sum_{i=1}^G \sum_{j=1}^{n_i} (\mathbf{Y}_{ij} - \bar{\mathbf{Y}}) (\mathbf{Y}_{ij} - \bar{\mathbf{Y}})'}_{\text{Matriz de SQ Total}} = \underbrace{\sum_{i=1}^G n_i (\bar{\mathbf{Y}}_i - \bar{\mathbf{Y}}) (\bar{\mathbf{Y}}_i - \bar{\mathbf{Y}})'}_{\text{Matriz de SQ do Modelo}} + \underbrace{\sum_{i=1}^G \sum_{j=1}^{n_i} (\mathbf{Y}_{ij} - \bar{\mathbf{Y}}_i) (\mathbf{Y}_{ij} - \bar{\mathbf{Y}}_i)'}_{\text{Matriz de SQ do Resíduo}}$$
$$\mathbf{T} = \mathbf{M} + \mathbf{E}$$

- De uma forma geral: $\mathbf{T} = \mathbf{Y}'(\mathbf{I} - n^{-1}\mathbf{J})\mathbf{Y}$; $\mathbf{M} = \mathbf{Y}'(\mathbf{H} - n^{-1}\mathbf{J})\mathbf{Y}$; $\mathbf{E} = \mathbf{Y}'(\mathbf{I} - \mathbf{H})\mathbf{Y}$, $\mathbf{H} = \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'$, $\mathbf{J} = \mathbf{1}\mathbf{1}'$, \mathbf{I} é uma matriz identidade de ordem n e $\mathbf{1}$ é um vetor coluna unitário de dimensão n .



Variância generalizada

- Seja $\Sigma_{(p \times p)}$ uma matriz de covariâncias.
- Variância generalizada $|\Sigma|$ (resume a informação contida em Σ).
- Suponha $p = 2$.
- Assim $|\Sigma| = \sigma_1^2 \sigma_2^2 - \sigma_{11}^2 = \sigma_1^2 \sigma_2^2 - \sigma_1^2 \sigma_2^2 \rho^2 = \sigma_1^2 \sigma_2^2 (1 - \rho^2)$.
- Estamos supondo que $\Sigma_1 = \Sigma_2 = \dots = \Sigma_G = \Sigma$ (teste de Box para igualdade de matrizes de covariâncias).

As quatro estatísticas “tradicionais”

- Lembre-se de que $\mathbf{T} = \mathbf{M} + \mathbf{E}$.
- Sejam $\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_s$ os autovalores diferentes de zero da matriz $\mathbf{E}^{-1}\mathbf{M}$, em que $s = \min(p, G - 1)$.
- **Lambda de Wilks:**
$$\prod_{i=1}^s \frac{1}{1 + \lambda_i} = \frac{|\mathbf{E}|}{|\mathbf{E} + \mathbf{M}|}.$$
- **Traço de Pillai:**
$$\sum_{i=1}^s \frac{\lambda_i}{1 + \lambda_i} = \text{tr}[\mathbf{M}(\mathbf{M} + \mathbf{E})^{-1}].$$
- **Traço de Lawley-Hotelling:**
$$\sum_{i=1}^s \lambda_i^{-1} = \text{tr}[\mathbf{M}\mathbf{E}^{-1}].$$
- **Máxima raiz de Roy:**
$$\frac{\lambda_1}{1 + \lambda_1}.$$

Cont.

- Quanto menor o valor das estatística de Wilks e maior os valores das estatísticas de Pillai, Lawley-Hotelling e de Roy, mais evidências tem-se contra H_0 .
- Existem aproximações pela distribuição F, para cada uma destas estatísticas ([link manual do SAS](#), livro: [Multivariate Statistics: High-Dimensional and Large-Sample Approximations](#)).

Aplicação

- Dados da iris: as quatro variáveis apresentadas e os três grupos.
- O teste de Box indica a rejeição da igualdade entre as matrizes de covariâncias (estatística = 140,94 ; $p\text{valor} < 0,0001$).
- Além da heterocedasticidade, a suposição de normalidade não parece ser razoável. Assim, os resultados obtidos podem não ser confiáveis.
- Utilização do pacote *manova* ([manual](#)) implementado na linguagem R (próximos slides).

Aplicação

- Outras opções de ajuste de modelos lineares multivariados no R:
 - Função apresentada no [Apêndice do Livro: An R Companion to Applied Regression](#).
 - Outras metodologias (funções) apresentadas em [livro: An R Companion to Applied Regression](#).
 - Funções do pacote [car](#).
 - Veja também o livro: [Modern Multivariate Statistical Techniques Regression, Classification, and Manifold Learning](#).

Estatísticas calculadas e p-valores

Estatística	Valor	Aproxim. pela dist. F.	p-valor
Wilks	0,02	119,15	< 0,0001
Pillai	1,19	53,46	< 0,0001
Hotelling-Lawley	32,47	580,53	< 0,0001
Roy	32,19	1167,00	< 0,0001

A igualdade simultânea dos vetores de médias é rejeitada, de acordo com cada uma das estatísticas. Como realizar outras comparações de interesse?

Anovas univariadas

CS

FV	g.l.	SQ	QM	Estat. F	p-valor
Espécie	2	63,21	31,61	119,26	< 0,0001
Resíduos	147	38,96	0,27		

LS

FV	g.l.	SQ	QM	Estat. F	p-valor
Espécie	2	11,34	5,67	49,16	< 0,0001
Resíduos	147	16,96	0,12		

Anovas univariadas

CP

FV	g.l.	SQ	QM	Estat. F	p-valor
Espécie	2	437,10	218,55	1180,16	< 0,0001
Resíduos	147	27,22	0,19		

LP

FV	g.l.	SQ	QM	Estat. F	p-valor
Espécie	2	80,41	40,21	960,01	< 0,0001
Resíduos	147	6,16	0,04		

Comentários

- Realizar anovas univariadas, desconsidera a estruturas de correlação existente.
- Além disso, não permitem testar hipóteses envolvendo mais de uma variável, de modo simultâneo.
- Uma vez que se decide realizar, desde o início, análises multivariadas, é melhor considerar o MRNLM.

Forma vetorial (MLNM)

- Considere novamente o modelo: $\mathbf{Y}_{(n \times p)} = \mathbf{X}_{(n \times q)}\mathbf{B}_{(q \times p)} + \boldsymbol{\xi}_{(n \times p)}$.
Portanto, temos que ([propriedades de matrizes](#)):

$$\begin{aligned}\mathbf{Y}'_{(p \times n)} &= \mathbf{B}'_{(p \times q)}\mathbf{X}'_{(q \times n)} + \boldsymbol{\xi}'_{(p \times n)} \\ \text{vec}(\mathbf{Y}') &= (\mathbf{X}_{(n \times q)} \otimes \mathbf{I}_p) \text{vec}(\mathbf{B}') + \text{vec}(\boldsymbol{\xi}') \\ \mathbf{Y}^*_{(np \times 1)} &= \mathbf{X}^*_{(np \times pq)}\boldsymbol{\beta}_{(pq \times 1)} + \boldsymbol{\xi}^*_{(np \times 1)}\end{aligned}$$

pois $\text{vec}(\mathbf{ABC}) = (\mathbf{C}' \otimes \mathbf{A})\text{vec}(\mathbf{B})$.

- Note, assim, que as observações dos indivíduos foram concatenadas (uma abaixo da outra), nos vetores \mathbf{Y}^* e $\boldsymbol{\xi}^*$.

Cont.

- Portanto, temos que $\mathbf{Y}^* \sim N_{pn}(\mathbf{X}^*\boldsymbol{\beta}, \boldsymbol{\Sigma}^*)$, em que
$$\boldsymbol{\Sigma}^* = \mathbf{I}_n \otimes \boldsymbol{\Sigma}_{(p \times p)}$$
- O estimador de **mínimos quadrados generalizados** de $\boldsymbol{\beta}$ é obtido minimizando-se

$$(\mathbf{Y}^* - \mathbf{X}^*\boldsymbol{\beta})' \boldsymbol{\Sigma}^{*-1} (\mathbf{Y}^* - \mathbf{X}^*\boldsymbol{\beta}).$$

- O que implica que $\hat{\boldsymbol{\beta}} = \left(\mathbf{X}^{*'} \boldsymbol{\Sigma}^{*-1} \mathbf{X}^* \right)^{-1} \mathbf{X}^{*'} \boldsymbol{\Sigma}^{*-1} \mathbf{Y}^*$.
- Exercício: obtenha o estimador de MQO de \mathbf{B} .

Cont.

- Note que

$$\begin{aligned}\hat{\beta} &= \left[(\mathbf{X} \otimes \mathbf{I})' (\mathbf{I} \otimes \Sigma)^{-1} (\mathbf{X} \otimes \mathbf{I}) \right]^{-1} (\mathbf{X} \otimes \mathbf{I})' (\mathbf{I} \otimes \Sigma)^{-1} \mathbf{Y}^* \\ &= \left[(\mathbf{X}'\mathbf{X} \otimes \Sigma^{-1}) \right]^{-1} [\mathbf{X}' \otimes \Sigma^{-1}] \mathbf{Y}^* \\ &= \left[(\mathbf{X}'\mathbf{X})^{-1} \mathbf{X}' \otimes \mathbf{I} \right] \mathbf{Y}^* \\ &= \mathbf{A} \mathbf{Y}^*\end{aligned}\tag{1}$$

- Por outro lado, temos que

$$\mathcal{E}(\hat{\beta}) = \left[(\mathbf{X}'\mathbf{X})^{-1} \mathbf{X}' \otimes \mathbf{I} \right] [\mathbf{X} \otimes \mathbf{I}] \beta = \beta.$$

- Além disso,

$$\text{Cov}(\hat{\beta}) = \left[(\mathbf{X}'\mathbf{X})^{-1} \mathbf{X}' \otimes \mathbf{I} \right] [\mathbf{I} \otimes \Sigma] \left[\mathbf{X} (\mathbf{X}'\mathbf{X})^{-1} \otimes \mathbf{I} \right] = (\mathbf{X}'\mathbf{X})^{-1} \otimes \Sigma$$

Cont.

- Logo, $\widehat{\beta} \sim N_{pq}(\beta, \Sigma_\beta)$, em que $\Sigma_\beta = (\mathbf{X}'\mathbf{X})^{-1} \otimes \Sigma$.
- Portanto, se Σ for conhecido (e conseqüentemente, o for Σ_β), então

$$(\widehat{\beta} - \beta)' (\Sigma_\beta)^{-1} (\widehat{\beta} - \beta) \sim \chi_{(qp)}^2$$

- Por outro lado, **de modo semelhante ao caso univariado**, a grande maioria das hipóteses de interesse, podem ser escritas na forma

$$H_0 : \mathbf{C}_{(r \times q)} \mathbf{B}_{(q \times p)} \mathbf{U}_{(p \times s)} = \mathbf{M}_{(r \times s)} \text{ vs } H_1 : \mathbf{CBU} \neq \mathbf{M}$$

em que $r \leq q$ e $s \leq p$.

Cont.

- Note agora que $H_0 : \mathbf{U}'\mathbf{B}'\mathbf{C}' = \mathbf{M}'$ e, assim, temos que:

$$\text{vec}(\mathbf{U}'\mathbf{B}'\mathbf{C}') - \underbrace{\text{vec}(\mathbf{M}')}_{\mathbf{M}^*} = \underbrace{(\mathbf{C} \otimes \mathbf{U}')}_{\mathbf{C}^*} \boldsymbol{\beta} - \mathbf{M}^* = \mathbf{C}^* \boldsymbol{\beta} - \mathbf{M}^*.$$

- Logo, podemos escrever as hipóteses anteriores como:

$$H_0 : \mathbf{C}^* \boldsymbol{\beta} - \mathbf{M}^* = 0 \text{ vs } H_1 : \mathbf{C}^* \boldsymbol{\beta} - \mathbf{M}^* \neq 0.$$

- Além disso, $\hat{\boldsymbol{\theta}} = \mathbf{C}^* \hat{\boldsymbol{\beta}} - \mathbf{M}^* \sim N_{rs} \left(\mathbf{C}^* \boldsymbol{\beta} - \mathbf{M}^*, \mathbf{C}^* \boldsymbol{\Sigma}_\beta \mathbf{C}^{*'} \right)$.

- Se $\boldsymbol{\Sigma}_\beta$ for conhecido, temos que, sob H_0

$$\hat{\boldsymbol{\theta}}' \left(\mathbf{C}^* \boldsymbol{\Sigma}_\beta \mathbf{C}^{*'} \right)^{-1} \hat{\boldsymbol{\theta}} \sim \chi^2_{(rs)}$$

Cont.

- Seja, então, $\widehat{\Sigma}$, um estimador consistente de Σ ($\widehat{\Sigma} \xrightarrow[n \rightarrow \infty]{P} \Sigma$).
Logo $\widehat{\Sigma}_\beta = (\mathbf{X}'\mathbf{X})^{-1} \otimes \widehat{\Sigma}$ o será para Σ_β .
- $\widehat{\Sigma} = \frac{1}{n-q} (\mathbf{Y} - \mathbf{X}\widehat{\mathbf{B}})' (\mathbf{Y} - \mathbf{X}\widehat{\mathbf{B}})$, em que $\widehat{\mathbf{B}} = (\mathbf{X}'\mathbf{X})^{-1} \mathbf{X}'\mathbf{Y}$.
- Assim, por Slutsky,

$$Q = (\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta})' (\mathbf{C}^* \widehat{\Sigma}_\beta \mathbf{C}^{*'})^{-1} (\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta})' \xrightarrow[n \rightarrow \infty]{D} \chi^2_{(rs, \delta)},$$

em que

$$\delta = (\mathbf{C}^* \boldsymbol{\beta} - \mathbf{M}^*)' (\mathbf{C}^* \Sigma_\beta \mathbf{C}^{*'})^{-1} (\mathbf{C}^* \boldsymbol{\beta} - \mathbf{M}^*).$$

- Sob H_0 tem-se que $\delta = 0$.

Cont.

- Nível descritivo: $P(Q > q_{calc} | H_0)$, em que $Q \approx \chi^2_{(rs)}$, para n suficientemente grande e q_{calc} é o valor calculado da estatística Q .
- Função poder: $P(Q > q_c | H_1, \alpha)$, em que $Q \approx \chi^2_{(rs)}$ para n suficientemente grande e q_c é o valor crítico para um dado α (nível de significância).
- Assim, o poder estimado do teste é dado por: $P(\tilde{Q} > q_c | H_1, \alpha)$, em que $\tilde{Q} \approx \chi^2_{(rs, \tilde{\delta})}$ para n suficientemente grande, q_c é o valor crítico para um dado α (nível de significância),
$$\tilde{\delta} = \left(\mathbf{C}^* \tilde{\beta} - \mathbf{M}^* \right)' \left(\mathbf{C}^* \tilde{\Sigma}_{\beta} \mathbf{C}^{*'} \right)^{-1} \left(\mathbf{C}^* \tilde{\beta} - \mathbf{M}^* \right)$$
 e “ \sim ” representa a respectiva estimativa.

Cont.

- Voltando ao exemplo da íris, se quisermos testar se as médias entre os grupos, em relação à variável CS (que é a variável 1), são simultaneamente iguais, teríamos (parametrização de médias):

$$\mathbf{C} = \begin{bmatrix} 1 & -1 & 0 \\ 1 & 0 & -1 \end{bmatrix}$$
$$\mathbf{U} = \begin{bmatrix} 1 \\ 0 \\ 0 \\ 0 \end{bmatrix}; \mathbf{M} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}; \mathbf{CBU} = \begin{bmatrix} \mu_{11} - \mu_{21} \\ \mu_{11} - \mu_{31} \end{bmatrix}$$

Cont.

- Voltando ao exemplo da íris, se quisermos testar se as médias entre os grupos, em relação à variável CS (que é a variável 1), são simultaneamente iguais, teríamos (parametrização casela de referências):

$$\mathbf{C} = \begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}$$
$$\mathbf{U} = \begin{bmatrix} 1 \\ 0 \\ 0 \\ 0 \end{bmatrix}; \mathbf{M} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}; \mathbf{CBU} = \begin{bmatrix} \alpha_{21} \\ \alpha_{31} \end{bmatrix}$$

Comparações múltiplas

- Comparações de interesse (para $k = 1, 2, 3, 4$)

$$H_0 : \mu_{1k} = \mu_{2k} \text{ vs } H_1 : \mu_{1k} \neq \mu_{2k}$$

$$H_0 : \mu_{1k} = \mu_{3k} \text{ vs } H_1 : \mu_{1k} \neq \mu_{3k}$$

$$H_0 : \mu_{2k} = \mu_{3k} \text{ vs } H_1 : \mu_{2k} \neq \mu_{3k}$$

Resultados

- Estatística do teste e p-valor entre parênteses.
- Variável CS:
 - Setosa x Versicolor: 81,59 ($< 0,0001$).
 - Setosa x Virginica: 236,10 ($< 0,0001$).
 - Versicolor x Virgínica: 40,10 ($< 0,0001$).
- Variável LS:
 - Setosa x Versicolor: 93,81 ($< 0,0001$).
 - Setosa x Virginica: 44,60 ($< 0,0001$).
 - Versicolor x Virgínica: 9,02 ($0,0027$).

Resultados

- Estatística do teste e p-valor entre parênteses.
- Variável CP:
 - Setosa x Versicolor: 1056,87 ($< 0,0001$).
 - Setosa x Virginica: 2258,26 ($< 0,0001$).
 - Versicolor x Virgínica: 225,35 ($< 0,0001$).
- Variável LP:
 - Setosa x Versicolor: 696,25 ($< 0,0001$).
 - Setosa x Virginica: 1891,28 ($< 0,0001$).
 - Versicolor x Virgínica: 292,49 ($< 0,0001$).

Análise de resíduos

- Veja, para o caso univariado: [link](#).
- Resíduo ordinário: $\mathbf{R}^* = \mathbf{Y} - \mathbf{X}\hat{\mathbf{B}}$. (distribuição de referência: $N(0,a)$, exata, em que “a” depende de Σ e \mathbf{X}).
- Resíduo padronizado: $\mathbf{R} = \mathbf{R}^* \mathbf{D}^{-1}$, em que $\mathbf{D} = \sqrt{(\text{diag}(\hat{\Sigma}))}$. (distribuição de referência: $N(0,1)$, assintótica)
- Resíduo studentizado (RS): $\mathbf{R}_s = \mathbf{R} ./ \sqrt{(1_{(n \times p)} - \mathbf{h}_{(n \times p)})}$, em que $./$ denota a divisão elemento por elemento, \mathbf{h} é uma matriz em que cada coluna corresponde à diagonal principal da matriz $\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'$. (distribuição de referência: $N(0,1)$, assintótica)

Análise de resíduos

- Resíduo studentizado (multivariado) (RSM) :

$\mathbf{R}_s^{(m)} = \mathbf{R}^*(chol(\widehat{\Sigma}))^{-1}$. (distribuição de referência: $N(0,1)$, assintótica)

- Resíduo baseado na distância de Mahalanobis (via RSM):

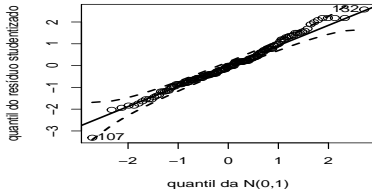
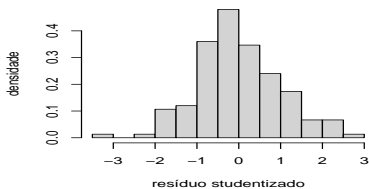
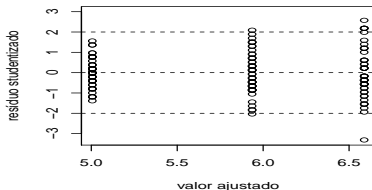
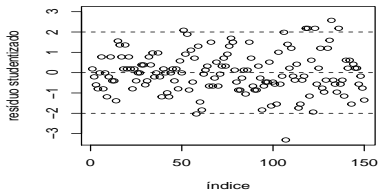
$(\mathbf{R}_i^*)'(\mathbf{A}_i\widehat{\Sigma}\mathbf{A}_i)^{-1}\mathbf{R}_i^*$, em que $\mathbf{R}_i = \mathbf{Y}_i - \mathbf{X}_i\widehat{\mathbf{B}}$, em que \mathbf{A}_i é a i -ésima linha da matriz $\mathbf{A} = \mathbf{I}_n - \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'$. (distribuição de referência: χ_p^2 , assintótica)

- Podemos também considerar uma combinação do RS e do RSM, ou

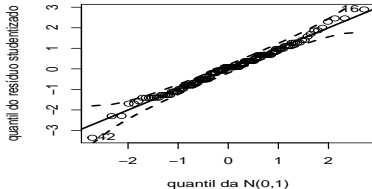
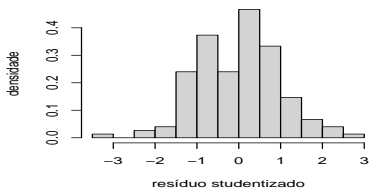
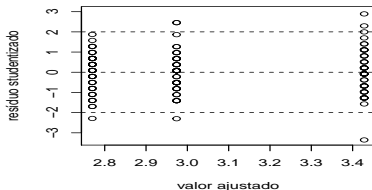
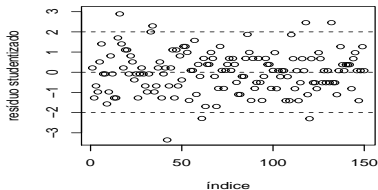
seja $\mathbf{R}_s^{*(m)} = \mathbf{R}_s^{(m)} ./ \sqrt{(1_{(n \times p)} - \mathbf{h}_{(n \times p)})}$.

- Exercício: escrever os valores individuais, para cada um dos resíduos acima .

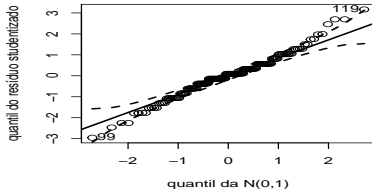
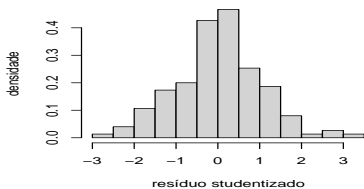
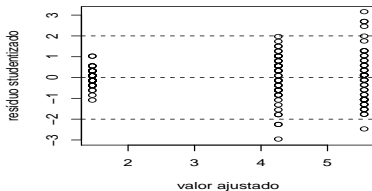
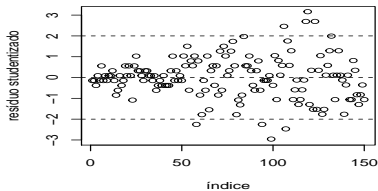
Análise de resíduos (RS) - comprimento da sépala



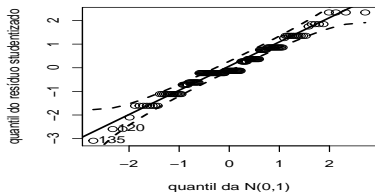
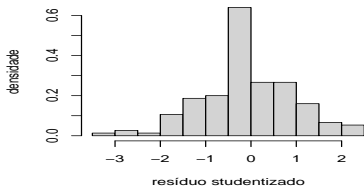
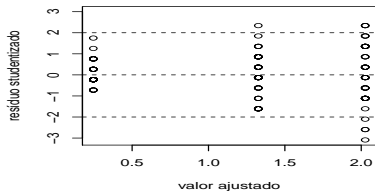
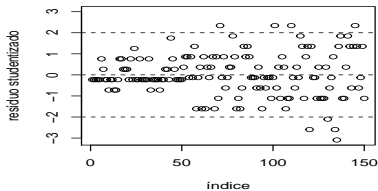
Análise de resíduos (RS) - largura da sépala



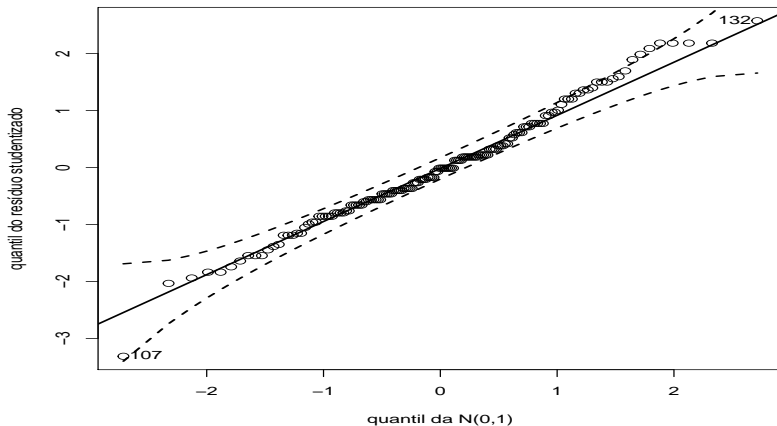
Análise de resíduos (RS) - comprimento da pétala



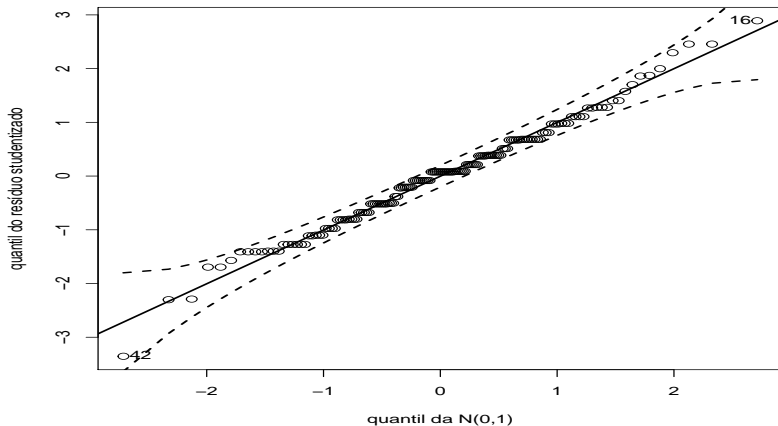
Análise de resíduos (RS) - largura da pétala



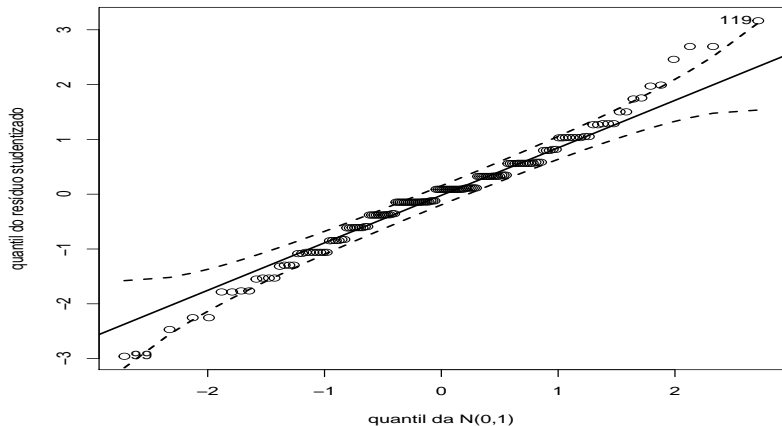
Envelopes (RS) - comprimento da sépala



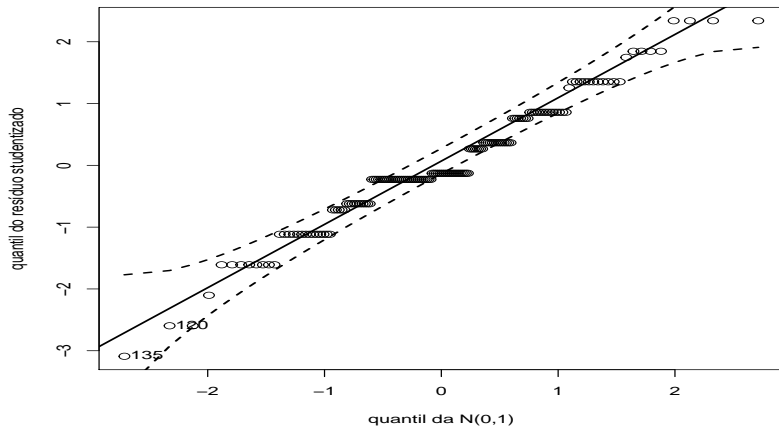
Envelopes (RS) - largura da sépala



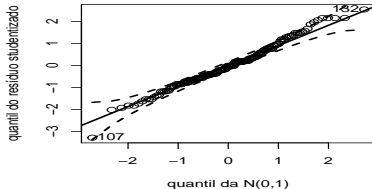
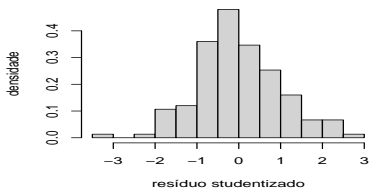
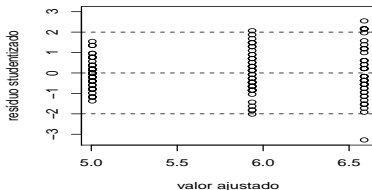
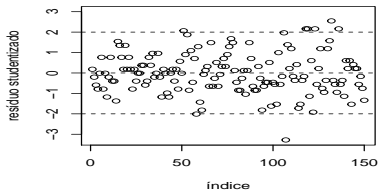
Envelopes (RS) - comprimento da pétala



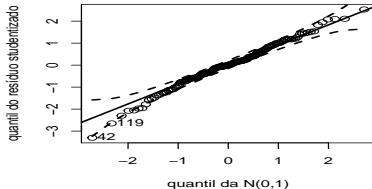
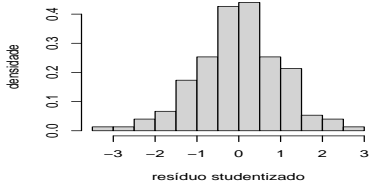
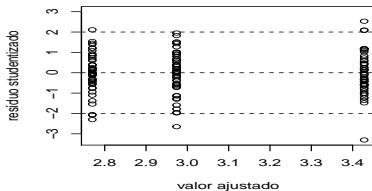
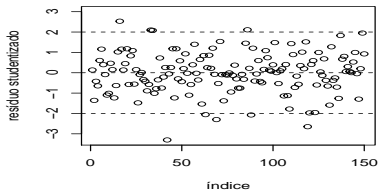
Envelopes (RS) - largura da pétala



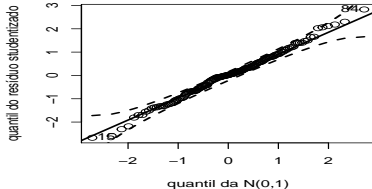
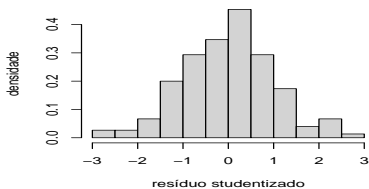
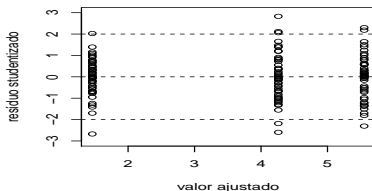
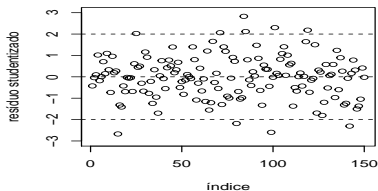
Análise de resíduos (RSM) - comprimento da sépala



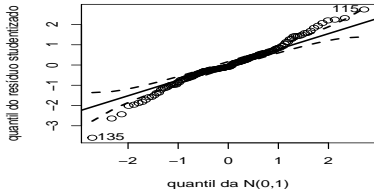
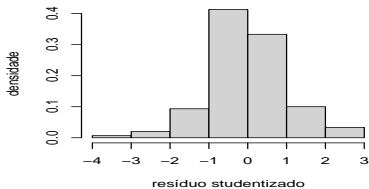
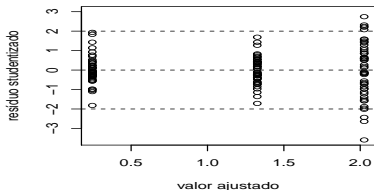
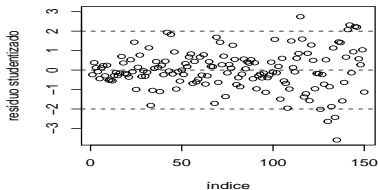
Análise de resíduos (RSM) - largura da sépala



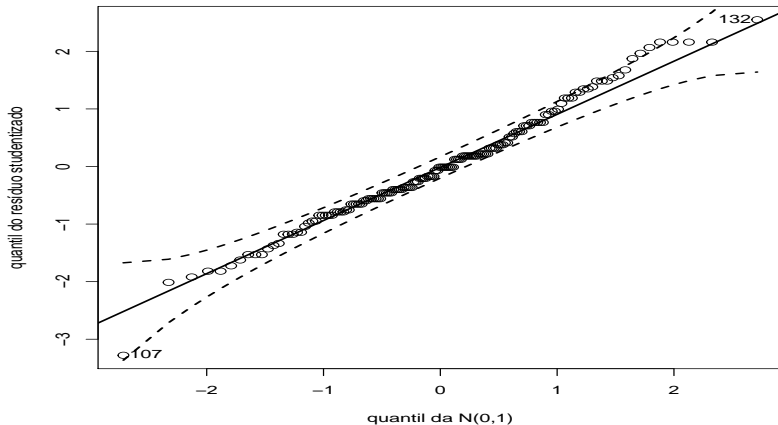
Análise de resíduos (RSM) - comprimento da pétala



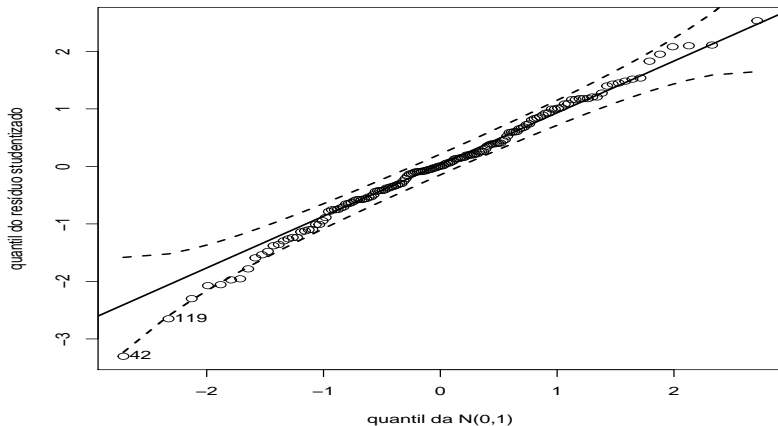
Análise de resíduos (RSM) - largura da pétala



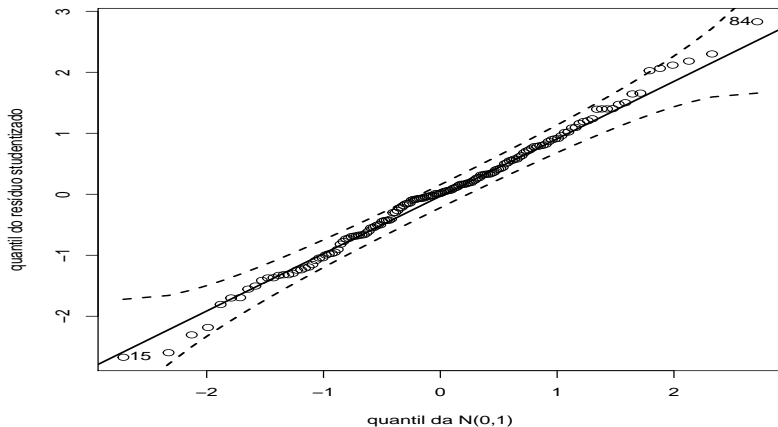
Envelopes (RSM) - comprimento da sépala



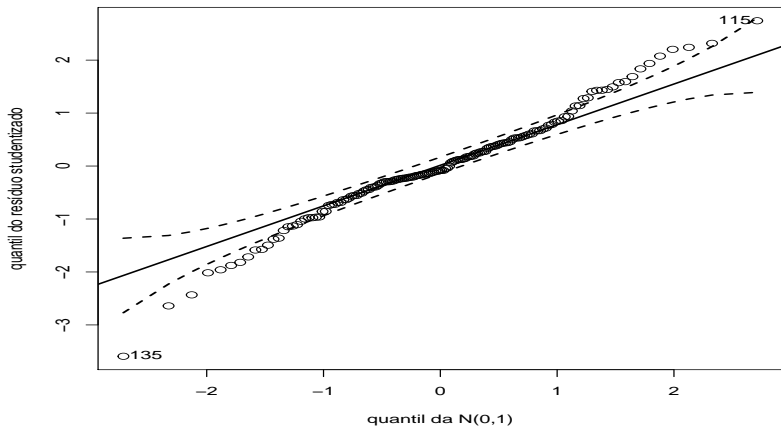
Envelopes (RSM) - largura da sépala



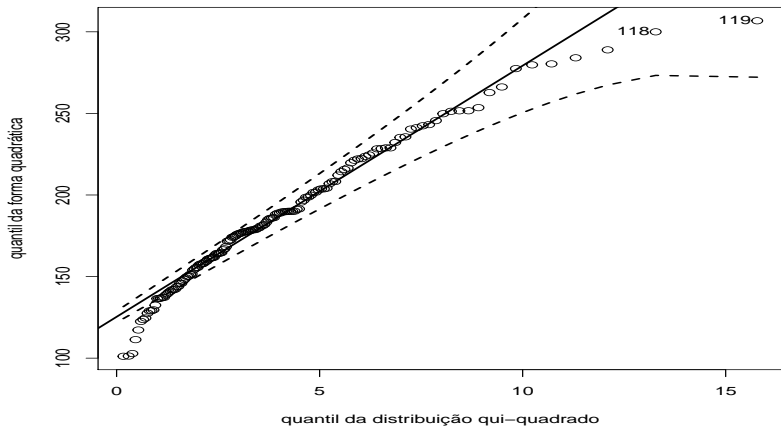
Envelopes (RSM) - comprimento da pétala



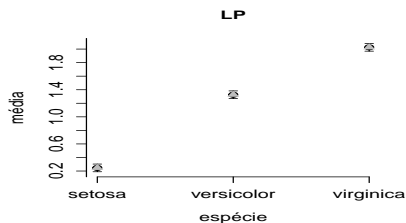
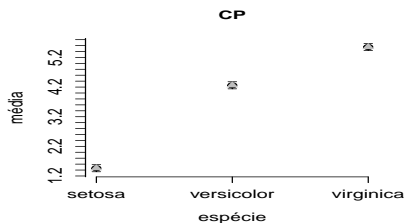
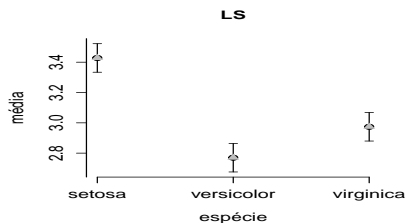
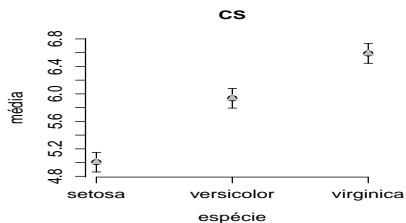
Envelopes (RSM) - largura da pétala



Envelopes (RSM) - distância de Mahalanobis



Médias observadas e previstas e respectivos IC(95%)



Comentários

- Os graficos de resíduos acusam:
 - Heterocedasticidade dos erros (para cada uma das variáveis).
 - Não normalidade dos erros (para cada uma das variáveis).
 - Independência entre as observações de diferentes flores (iris).
- Assim, podemos concluir que o modelo não está bem ajustado
- Lembrando que o MRNLM é, em geral, sensível (não robusto) à ausência das suposições sobre os erros, conclui-se que os resultados não são confiáveis.