Análise Discriminante: parte 2

Prof. Caio Azevedo

Comentários

- Um método de classificação entre duas populações sob heterocedasticidade ($\boldsymbol{\Sigma}_1 \neq \boldsymbol{\Sigma}_2$) pode ser encontrada em [Johnson and Wichern(2007)].
- A regra de classificação apresentada no supramencionado livro pode gerar resultados indesejados (altas taxas de erro), principalmente quando temos mais de duas variáveis e quando a suposição de normalidade multivariada não é válida.

Comentários

- Não discutiremos a AD sob heterocedasticidade. Contudo, apresentaremos a situação com vários grupos, sob homocedasticidade.
- Algumas alternativas nesse caso modelos de regressão para dados binários (dois grupos) e modelos de regressão para dados politômicos (três ou mais grupos).

Classificação com várias populações, baseada no CECE

- Suponha g populações de sorte que cada unidade amostral (experimental) pertença a uma e somente uma população. Defina (analogamente ao caso de duas populações):
  - Suporte da distribuição: $A = \{\boldsymbol{x} \in \mathcal{R}^p, f(\boldsymbol{x}) > 0\}$ e a respectiva partição $A = \dot{\bigcup}_{i=1}^{g} A_i$.
  - p_i : probabilidade (a priori) de um determinado indivíduo pertencer à população i.
  - $c(k|i)$: custo de classificar o indivíduo na população k dado que ele pertence à população i, naturalmente $c(i|i) = 0, \forall i$.
  - $P(k|i) = \int_{A_k} f_i(\boldsymbol{x}) d\boldsymbol{x}$: probabilidade de classificar um indivíduo na população k dado que ele pertence à população i. Além disso $P(i|i) = 1 - \sum_{k=1, k\neq i}^{g} P(k|i)$.

Classificação com várias populações, baseada no CECE

- O custo esperado condicional (esperança condicional) de classificar equivocadamente uma unidade pertencente à população 1, em uma outra população, (2,3,..,g) é dada por:

$$\begin{aligned} ECE(1) &= P(2|1)c(2|1) + P(3|1)c(3|1) + ... + P(g|1)c(g|1) \\ &= \sum_{k=2}^{g} P(k|1)c(k|1) = \sum_{k=1}^{g} P(k|1)c(k|1) \end{aligned} \quad (1)$$

- Assumi-se que este valor esperado condicional ocorre com probabilidade p_1 (a priori).
- Analogamente podemos definir $ECE(2), ..., ECE(g)$ e, além disso, cada uma delas ocorre com probabilidade $p_k, k = 2, ..., g$

Classificação com várias populações, baseada no CECE

- Assim o custo (global) esperado de classificação errada é dado por:

$$\begin{aligned} CECE &= p_1 ECE(1) + p_2 ECE(2) + ... + p_g ECE(g) \\ &= \sum_{i=1}^{g} p_i \sum_{k=1, k \neq i}^{g} P(k|i) c(k|i) \end{aligned} \tag{2}$$

Classificação com várias populações, baseada no CECE

- A regra de classificação que minimiza o CECE (equação (2)) consiste em classificar uma dada observação ($\boldsymbol{x}_0$), associada à uma determinada unidade, na população $k, k = 1, 2, ..., g$, para o qual

$$\sum_{i=1, i\neq k}^{g} p_i f_i(\boldsymbol{x}_0) c(k|i) = \sum_{i=1}^{g} p_i f_i(\boldsymbol{x}_0) c(k|i)$$

é mínimo. Para a demonstração veja [Anderson(2003)].

- Note que, neste caso, é preciso supor alguma forma para a distribuição multivariada ($f_i(.)$). A suposição de homocedasticidade também é importante (no mesmo sentido apresentado para o caso de duas populações).

Classificação com várias populações baseado no CECE com custos iguais ($c(.|.)$)

- Nesse caso a regra passa a ser: aloca-se a unidade amostral na população k, com base em um vetor de observações $\boldsymbol{x}_0$ se

$$p_k f_k(\boldsymbol{x}_0) > p_i f_i(\boldsymbol{x}_0), \forall i \neq k \tag{3}$$

ou de modo equivalente, se

$$\ln(p_k f_k(\boldsymbol{x}_0)) > \ln(p_i f_i(\boldsymbol{x}_0)), \forall i \neq k$$

Classificação com várias populações baseada no CECE com custos iguais ($c(.|.)$)

- Um aspecto interessante é que a regra de clasificação (3) equivale àquela que maximiza a probabilidade à posteriori $P(k|\boldsymbol{x}_0) =$ ($\boldsymbol{x}_0$ é oriundo da população k, dado que $\boldsymbol{x}_0$ foi observado), a qual é dada por:

$$P(k|\boldsymbol{x}_0) = \frac{p_k f_k(\boldsymbol{x}_0)}{\sum_{i=1}^{g} p_i f_i(\boldsymbol{x}_0)} = \frac{\text{priori} \times \text{verossimilhança}}{\sum[\text{priori} \times \text{verossimilhança}]}$$

- Que é, essencialmente, um resultado do Teorema de Bayes.

Método de Fisher para AD com várias populações

- Fisher também desenvolveu uma metodologia de AD considerando várias populações.
- A ideia é semelhante ao caso de duas populações, no sentido que ele buscou definir funções univariadas (função discriminante) com base nas observações multivariadas.
- Neste caso, também, a metodologia de Fisher é equivalente à regra do mínimo CECE, sob normalidade multivariada, homocedasticidade, custos de classificação errada e probabilidades à priori iguais (veja [Johnson and Wichern(2007)]).

Método de Fisher para AD com várias populações

- Características importantes:
  - Permite trabalhar com representações univariadas de dados multivariados.
  - Permite análises gráficas que permitem analisar de modo simples o comportamento das populações de interesse.
  - Não assume normalidade dos dados.
  - Considera homocedasticidade.
- Para a AD de Fisher para duas ou mais populações, veja Fisher (1936) e Fisher (1938).

Método de Fisher para AD com várias populações

- A ideia é trabalhar com combinações lineares das observações $\boldsymbol{x}$ ($\boldsymbol{x}_0$), ou seja, $Y = \boldsymbol{a}'\boldsymbol{X}$. Assuma que $\mathcal{E}(\boldsymbol{X}|i) = \boldsymbol{\mu}_i$ e $Cov(\boldsymbol{X}|i) = \boldsymbol{\Sigma}$.
- Assim $\mathcal{E}(Y) = \boldsymbol{a}'\boldsymbol{\mu}_i$ (população i) e $Cov(Y) = \boldsymbol{a}'\boldsymbol{\Sigma}\boldsymbol{a}$ (para todas as populações).
- Defina $\left(\text{considerando: } \overline{\boldsymbol{\mu}} = \frac{1}{g}\sum_{i=1}^{g}\boldsymbol{\mu}_i\right)$

$$
\begin{aligned}
V &= \frac{\boldsymbol{a}'\left(\sum_{i=1}^{g}\left(\boldsymbol{\mu}_i - \overline{\boldsymbol{\mu}}\right)\left(\boldsymbol{\mu}_i - \overline{\boldsymbol{\mu}}\right)'\right)\boldsymbol{a}}{\boldsymbol{a}'\boldsymbol{\Sigma}\boldsymbol{a}} \\
&= \frac{\text{variabilidade entre grupos}}{\text{variabilidade dentro de cada grupo}}
\end{aligned}
$$

Método de Fisher para AD com várias populações

- A idéia de Fisher foi tentar "separar" ao máximo as populações em relação à medida V, ou seja, ele buscou maximizá-la.
- Resultado: defina $\boldsymbol{w} = \sum_{i=1}^{g}(n_i - 1)\boldsymbol{s}_i^2$ (variabilidade dentro de cada população), $\boldsymbol{b} = \sum_{i=1}^{g}(\overline{\boldsymbol{x}}_i - \overline{\boldsymbol{x}})(\overline{\boldsymbol{x}}_i - \overline{\boldsymbol{x}})'$, em que $\overline{\boldsymbol{x}}_i$ é a média amostral, $\boldsymbol{s}_i^2$ é a matriz de covariâncias amostral, ambas relativas a população i e $\overline{\boldsymbol{x}} = \frac{1}{g}\sum_{i=1}^{g}\overline{\boldsymbol{x}}_i$.
- Sejam $\widetilde{\lambda}_1, \widetilde{\lambda}_2, ..., \widetilde{\lambda}_s > 0$, em que $s \leq \min(g-1, p)$ os autovalores diferentes de zero de $\boldsymbol{w}^{-1}\boldsymbol{b}$ e $\widetilde{\boldsymbol{e}}_1, \widetilde{\boldsymbol{e}}_2, ..., \widetilde{\boldsymbol{e}}_s$ os respectivos autovetores, devidamente reescalonados (veja [Johnson and Wichern(2007)]).

Método de Fisher para AD com várias populações

- Então o vetor de coeficientes, digamos $\widetilde{\boldsymbol{a}}$, que maximiza a razão

$$\frac{\widetilde{\boldsymbol{a}}' \boldsymbol{b} \widetilde{\boldsymbol{a}}}{\widetilde{\boldsymbol{a}}' \boldsymbol{w} \widetilde{\boldsymbol{a}}},$$

  é dado por $\widetilde{\boldsymbol{a}}_1 = \widetilde{\boldsymbol{e}}_1$.

- Então, a combinação linear $\widetilde{y}_1 = \widetilde{\boldsymbol{a}}_1 \boldsymbol{x}$ é denominada de primeira função discriminante (amostral).

- Analogamente, a combinação linear $\widetilde{y}_k = \widetilde{\boldsymbol{a}}_k \boldsymbol{x}$ é denominada k-ésima função discriminante (amostral).

Método de Fisher para AD com várias populações

- Regra de classificação: aloca-se a observação $\boldsymbol{x}_0$ (ou seja, a unidade amostral), à população k, com base em s funções discriminantes, de sorte que:

$$\sum_{j=1}^{s}(\widetilde{y}_j - \overline{y}_{kj})^2 = \sum_{j=1}^{s}[\widetilde{\boldsymbol{a}}_j'(\boldsymbol{x}_0 - \overline{\boldsymbol{x}}_k)]^2 \leq \sum_{j=1}^{s}[\widetilde{\boldsymbol{a}}_j'(\boldsymbol{x}_0 - \overline{\boldsymbol{x}}_i)]^2, \forall i \neq k$$

- Nota: assim como no caso de duas populações, a regra de classificação de Fisher coincide com a regra do CECE (sob normalidade e homocedasticidade), sob custos iguais e probabilidades de classificação iguais.

Voltando ao Exemplo 1 (considerando os três grupos)

- Resultados da classificação:

	S	VE	VI
S	25	0	0
VE	0	25	0
VI	0	0	25

- TEA (%) = 0,00 .

Medidas resumo

Função discriminante 1

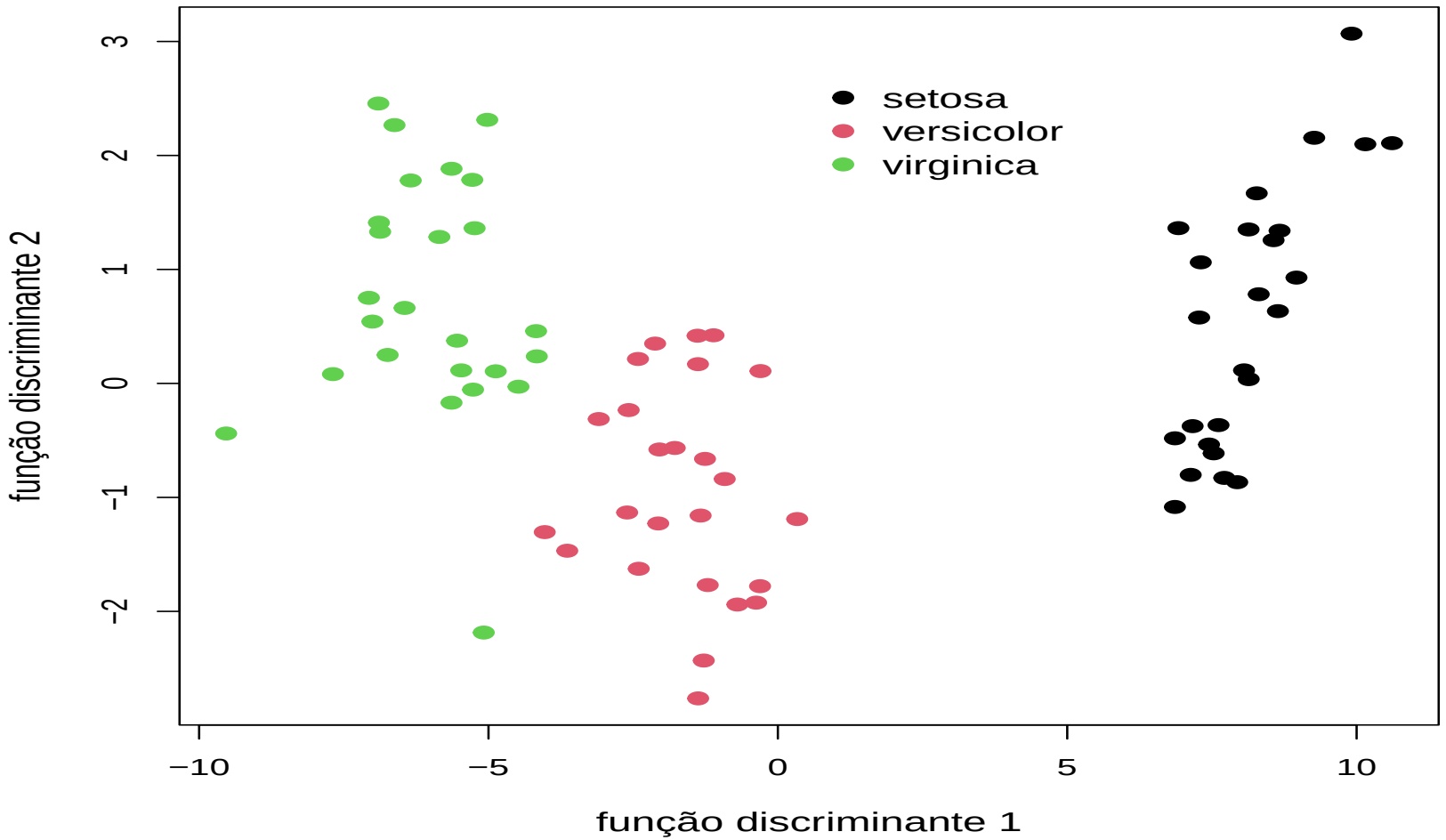
Grupo	Média	DP	Var.	Min.	Med.	Máx.	CA	Curt.	n
Set.	8,14	1,03	1,06	6,86	8,06	10,61	0,78	-0,28	25
Vers.	-1,66	1,06	1,12	-4,03	-1,38	0,33	-0,34	-0,45	25
Virg.	-5,99	1,22	1,48	-9,53	-5,64	-4,17	-0,76	0,67	25

Medidas resumo

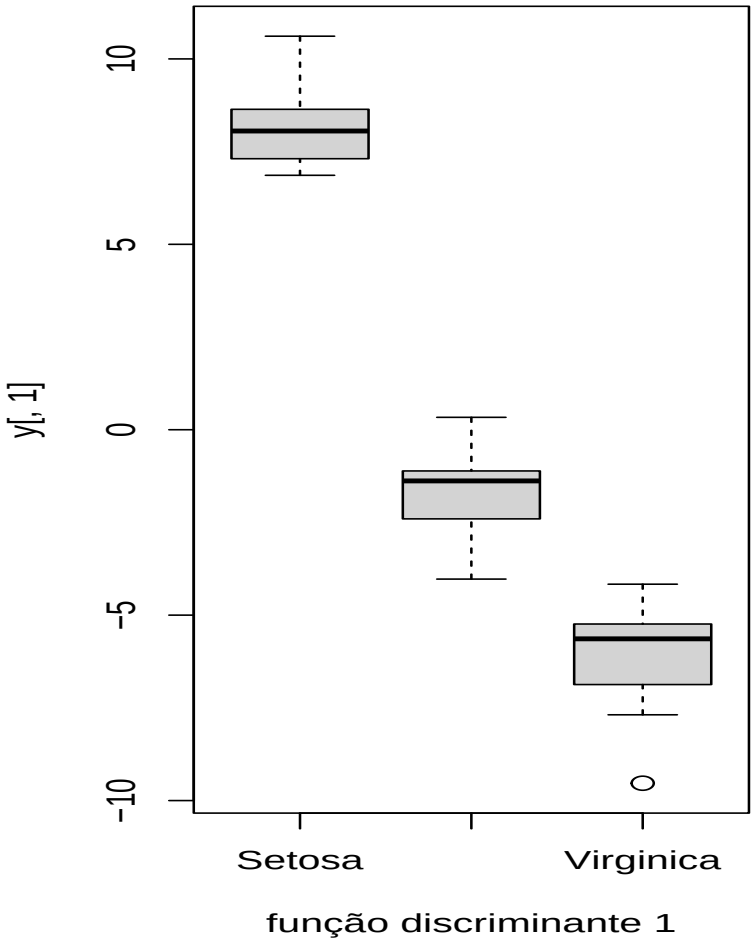
Função discriminante 2

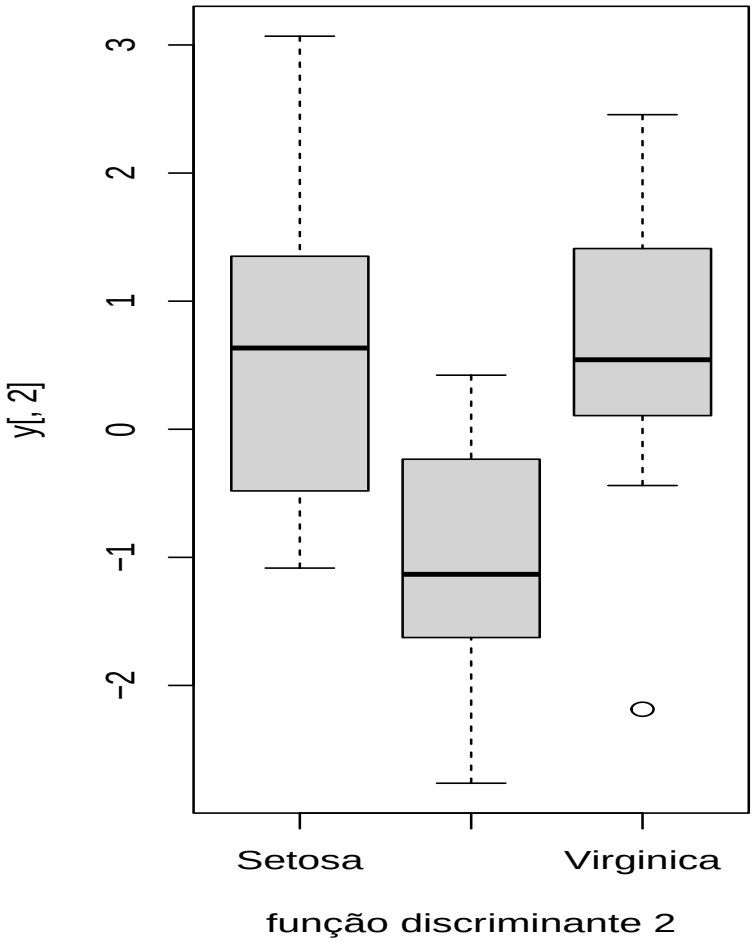
Grupo	Média	DP	Var.	Min.	Med.	Máx.	CA	Curt.	n
Set.	0,58	1,15	1,33	-1,08	0,63	3,07	0,78	-0,28	25
Vers.	-0,93	0,92	0,85	-2,76	-1,13	0,42	-0,34	-0,45	25
Virg.	0,74	1,05	1,11	-2,19	0,54	2,46	-0,76	0,67	25

Dispersões entre as funções discriminantes

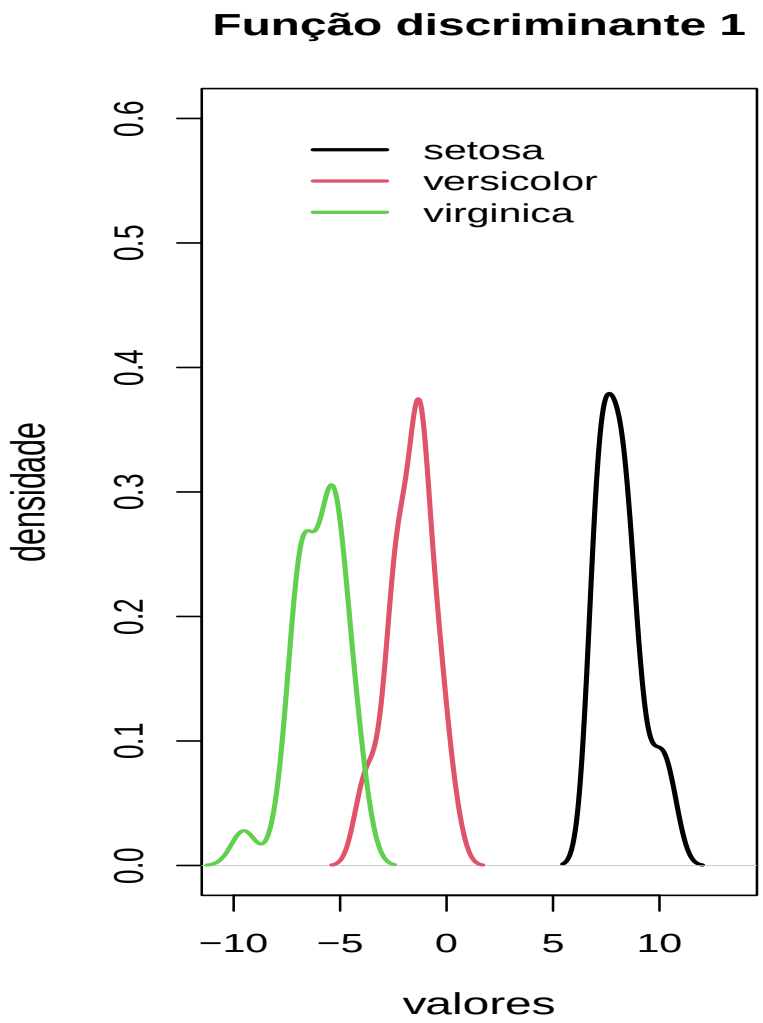


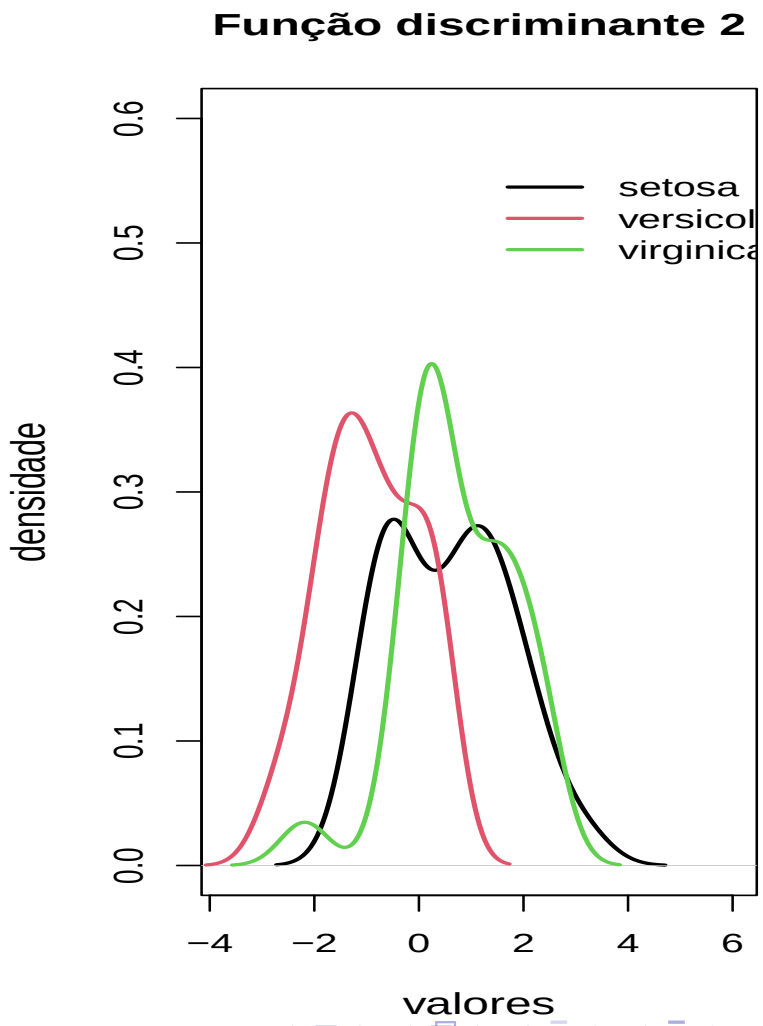
Ex. 1: boxplots da função discriminante





Ex. 1: densidade estimada da função discriminante





Comentários

- A função discriminante 1, como esperado, consegue distinguir melhor os grupos.
- Os grupos versicolor e virgínica são bem diferentes do grupo setosa e menos diferentes entre si.
- Em relação a função discriminante 1, o grupo setosa apresenta uma distribuição essencialmente positiva, enquanto que os dois outros grupos, negativa.
- A suposição de normalidade não parece ser razoável, para nenhuma das funções discriminantes e nenhum dos grupos.
- É possível construir gráficos do tipo "bi-plot", usando as funções discriminantes, de modo semelhante ao que fora feito na ACP, AF e ACC (para fins semelhantes).

Referências I

T. W. Anderson.

An introduction to multivariate statistical analysis.

John Wiley, New York, 2003.

R. A. Johnson and D. W. Wichern.

Applied multivariate statistical analysis.

Prentice Hall, Upper Saddle River, 6 edition, 2007.