

MAE 5776

ANÁLISE MULTIVARIADA

Júlia M Pavan Soler

pavan@ime.usp.br

1º Sem/2019 - IME

Análise Multivariada

Já vimos ☺

$$Y_{n \times p} = (Y_{ij}) \in \mathcal{R}^{n \times p}$$

- Revisão de Metodologias Clássicas: Foco na obtenção de vetores reducionistas

$$\mathcal{R}^p \rightarrow \mathcal{R}^m$$

- Componentes Principais ($m \leq \min(n, p)$)
- Análise Fatorial Exploratória (via CP: $m \leq \min(n, p)$)
- Coordenadas Principais – Escalonamento Multidimensional ($m \leq \text{posto}(D_{n \times n})$)
- Análise de Correspondência
- Análise Discriminante ($m \leq \min(n, p, G-1)$)
- Correlação Canônica ($m \leq \min(n, p, q)$)

Combinações
lineares de Y

$n > p$
Observações
Independentes

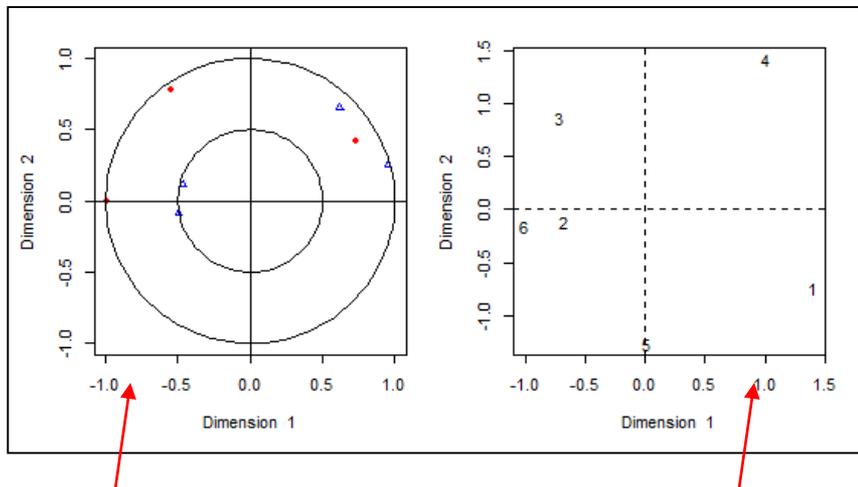


- ⇒ Problema Dual de redução de dimensionalidade ($\mathcal{R}^{n \times p}$, $\mathcal{R}^{p \times p}$, $\mathcal{R}^{n \times n}$)
- ⇒ Representações Bi-Plot
- ⇒ Problema $n \ll p$ (*Big-n*)

TABLE 1.
Wine tasting data from Abdi and Valentin (2007).

Wine	Oak-type	Expert 1			Expert 2				Expert 3		
		Fruity	Woody	Coffee	Red fruit	Roasted	Vanillin	Woody	Fruity	Butter	Woody
1	1	1	6	7	2	5	7	6	3	6	7
2	2	5	3	2	4	4	4	2	4	4	3
3	2	6	1	1	5	2	1	1	7	1	1
4	2	7	1	2	7	2	1	2	2	2	2
5	1	2	5	4	3	5	6	5	2	6	6
6	1	3	4	4	3	5	4	5	1	7	5

Análise dos dados padronizados:



Mostra as correlações dos dois primeiros Escores com as correspondentes variáveis do grupo

Dispersão dos dois primeiros Escores (default é do grupo de variáveis Y1). Mostra possíveis agrupamentos das observações (vinhos).

Correlações canônicas

[1,] 1.0000000 1.0000000 0.4656637

Expert1_Cargas

```

[ ,1]      [ ,2]      [ ,3]
[1,] -0.84138034 -0.1112332 1.2243924
[2,]  0.16237785  0.5759744 1.1835037
[3,] -0.08821111  0.9690145 -0.6429336

```

Expert2_Cargas

```

[ ,1]      [ ,2]      [ ,3]
[1,] -0.29972189  0.8818167  0.5114628
[2,] -0.40100915  1.6609644 -0.1364524
[3,]  1.15168106 -0.8371717 -0.2248502
[4,] -0.01455843 -0.3661562 -1.0847665

```

Expert1_Escores Canônicos

```

[ ,1]      [ ,2]      [ ,3]
[1,]  1.405758474 -0.7488203 -0.2014498
[2,] -0.685505624 -0.1203654 -1.1972958
[3,] -0.720252850  0.8691857  1.3987455
[4,]  1.006112507  1.4287984 -0.2650015
[5,]  0.005258323 -1.2653525  0.9880921
[6,] -1.011370830 -0.1634459 -0.7230906

```

Expert2_Escores Canônicos

```

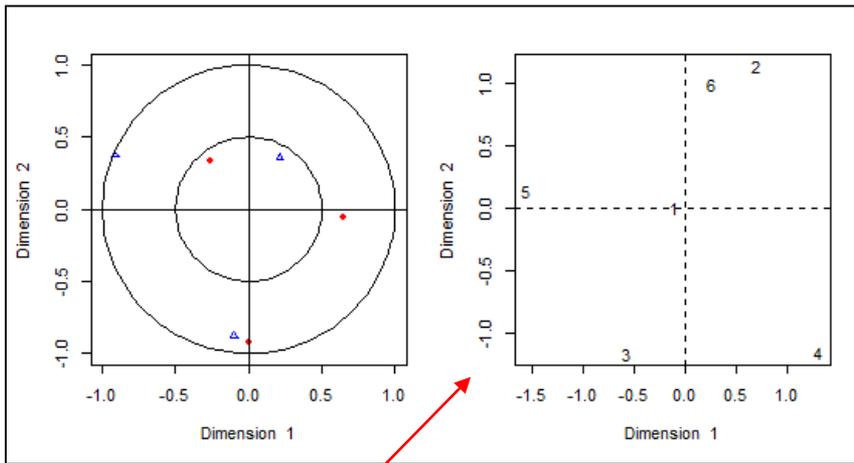
[ ,1]      [ ,2]      [ ,3]
[1,]  1.405758474 -0.7488203  0.33065100
[2,] -0.685505624 -0.1203654  0.60908538
[3,] -0.720252850  0.8691857  1.27371420
[4,]  1.006112507  1.4287984 -0.69629315
[5,]  0.005258323 -1.2653525  0.03565979
[6,] -1.011370830 -0.1634459 -1.55281721

```

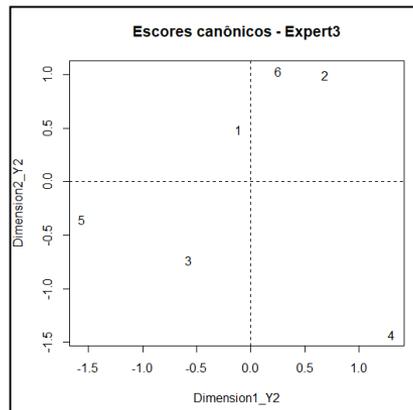
TABLE 1.
Wine tasting data from Abdi and Valentin (2007).

Wine	Oak-type	Expert 1			Expert 2				Expert 3		
		Fruity	Woody	Coffee	Red fruit	Roasted	Vanillin	Woody	Fruity	Butter	Woody
1	1	1	6	7	2	5	7	6	3	6	7
2	2	5	3	2	4	4	4	2	4	4	3
3	2	6	1	1	5	2	1	1	7	1	1
4	2	7	1	2	7	2	1	2	2	2	2
5	1	2	5	4	3	5	6	5	2	6	6
6	1	3	4	4	3	5	4	5	1	7	5

Análise dos dados padronizados:



Default: Expert1



Correlações Canônicas

[1] 1.0000000 0.9230238 0.7291302

Expert1_cargas

	[,1]	[,2]	[,3]
[1,]	-1.1113912	-0.3775500	0.9174400
[2,]	-0.4630267	-1.2050024	0.3039234
[3,]	1.0955141	-0.1469552	0.3720039

Expert3_cargas

	[,1]	[,2]	[,3]
[1,]	-0.40730814	-0.8634015	-0.4745978
[2,]	0.08401328	0.2882993	-0.9711619
[3,]	-1.03220452	0.1613599	-0.1769375

Expert1_Escore Canônicos

	[,1]	[,2]	[,3]
[1,]	-0.1127343	0.005748101	-1.6014770
[2,]	0.6866542	1.150428201	0.3506336
[3,]	-0.5739199	-1.156176303	1.2508434
[4,]	1.3015684	-1.145180622	-0.3441426
[5,]	-1.5577827	0.140976892	-0.3618140
[6,]	0.2562143	1.004203730	0.7059566

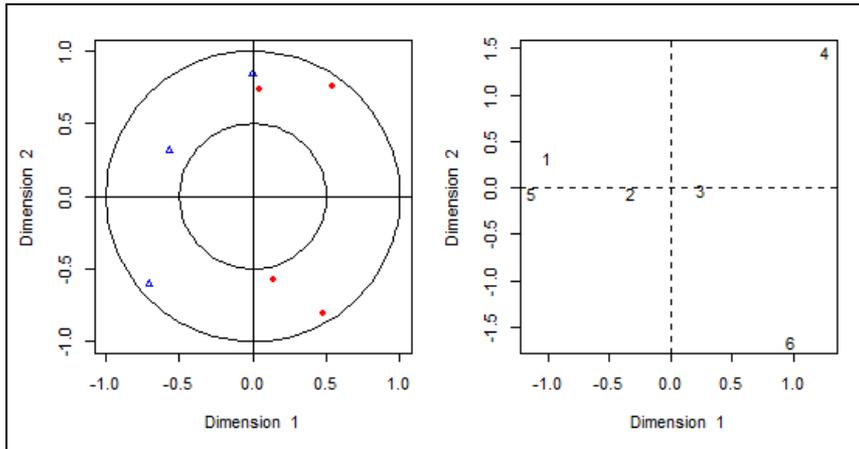
Expert3_Escore Canônicos

	[,1]	[,2]	[,3]
[1,]	-0.1127343	0.4903040	-1.17575646
[2,]	0.6866542	0.9951620	-0.76373492
[3,]	-0.5739199	-0.7293023	0.63427093
[4,]	1.3015684	-1.4316855	-0.04058934
[5,]	-1.5577827	-0.3548733	-0.25573837
[6,]	0.2562143	1.0303951	1.60154816

TABLE 1.
Wine tasting data from Abdi and Valentin (2007).

Wine	Oak-type	Expert 1			Expert 2				Expert 3		
		Fruity	Woody	Coffee	Red fruit	Roasted	Vanillin	Woody	Fruity	Butter	Woody
1	1	1	6	7	2	5	7	6	3	6	7
2	2	5	3	2	4	4	4	2	4	4	3
3	2	6	1	1	5	2	1	1	7	1	1
4	2	7	1	2	7	2	1	2	2	2	2
5	1	2	5	4	3	5	6	5	2	6	6
6	1	3	4	4	3	5	4	5	1	7	5

Análise dos dados padronizados:



Correlações Canônicas

```
[1] 1.0000000 1.0000000 0.218548
```

Expert2_Cargas

```

          [,1]          [,2]          [,3]
[1,]  0.5689416  0.01326276  0.2489435
[2,]  1.4853794  0.56264379  0.7512700
[3,] -0.7561736  0.06717931 -1.4568672
[4,]  0.3038776 -0.65411725 -0.8028081

```

Expert3_Cargas

```

          [,1]          [,2]          [,3]
[1,] -0.3591130  0.8007657  0.6053241
[2,] -0.6791661  0.4018158 -0.6407867
[3,] -0.8690189 -0.3220997  0.5136613

```

Expert2_Escores Canônicos

```

          [,1]          [,2]          [,3]
[1,] -1.0078586  0.31794546 -1.5019667
[2,] -0.3238446 -0.06296618  1.1048929
[3,]  0.2394681 -0.02033852  0.9597462
[4,]  1.2585796  1.47017986 -0.2992835
[5,] -1.1377202 -0.05244255  0.3571499
[6,]  0.9713758 -1.65237806 -0.6205388

```

Expert3_Escores Canônicos

```

          [,1]          [,2]          [,3]
[1,] -1.0078586  0.31794546 -7.201680e-01
[2,] -0.3238446 -0.06296618 -1.391507e+00
[3,]  0.2394681 -0.02033852  1.098092e+00
[4,]  1.2585796  1.47017986 -8.880114e-05
[5,] -1.1377202 -0.05244255  1.149290e+00
[6,]  0.9713758 -1.65237806 -1.356179e-01

```

Redução de Dimensionalidade

Obtenção de Vetores Reducionistas

$$Y_{n \times p} = (Y_{ij}) \in \mathfrak{R}^{n \times p}$$

- Componentes Principais: $f(\Sigma; a) = \frac{a' \Sigma a}{a' a}, \quad a' a = 1 \Rightarrow Z_{ki} = a_k' Y_i \quad \text{Cov}(Y) = \Sigma$
- Análise Fatorial Exploratória (via CP): $\Rightarrow F_{ki} = \lambda^{-1/2} Z_{ki} \quad \Sigma = \Phi \Phi' + \text{Diag}(\Psi_{jj})$
- Análise Discriminante (Linear de Fisher): $f(\Sigma_w^{-1} \Sigma_b; a) = \frac{a' \Sigma_b a}{a' \Sigma_w a}, \quad a' \Sigma_w a = 1 \quad \Sigma = \Sigma_b + \Sigma_w$
 $Y_{n \times p}; n = \sum n_g$
- Análise de Correlação Canônica:

$$Y_{i(p+q) \times 1} = \begin{pmatrix} Y_{1i p \times 1} \\ Y_{2i q \times 1} \end{pmatrix} \quad \Sigma = \begin{pmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{pmatrix}$$

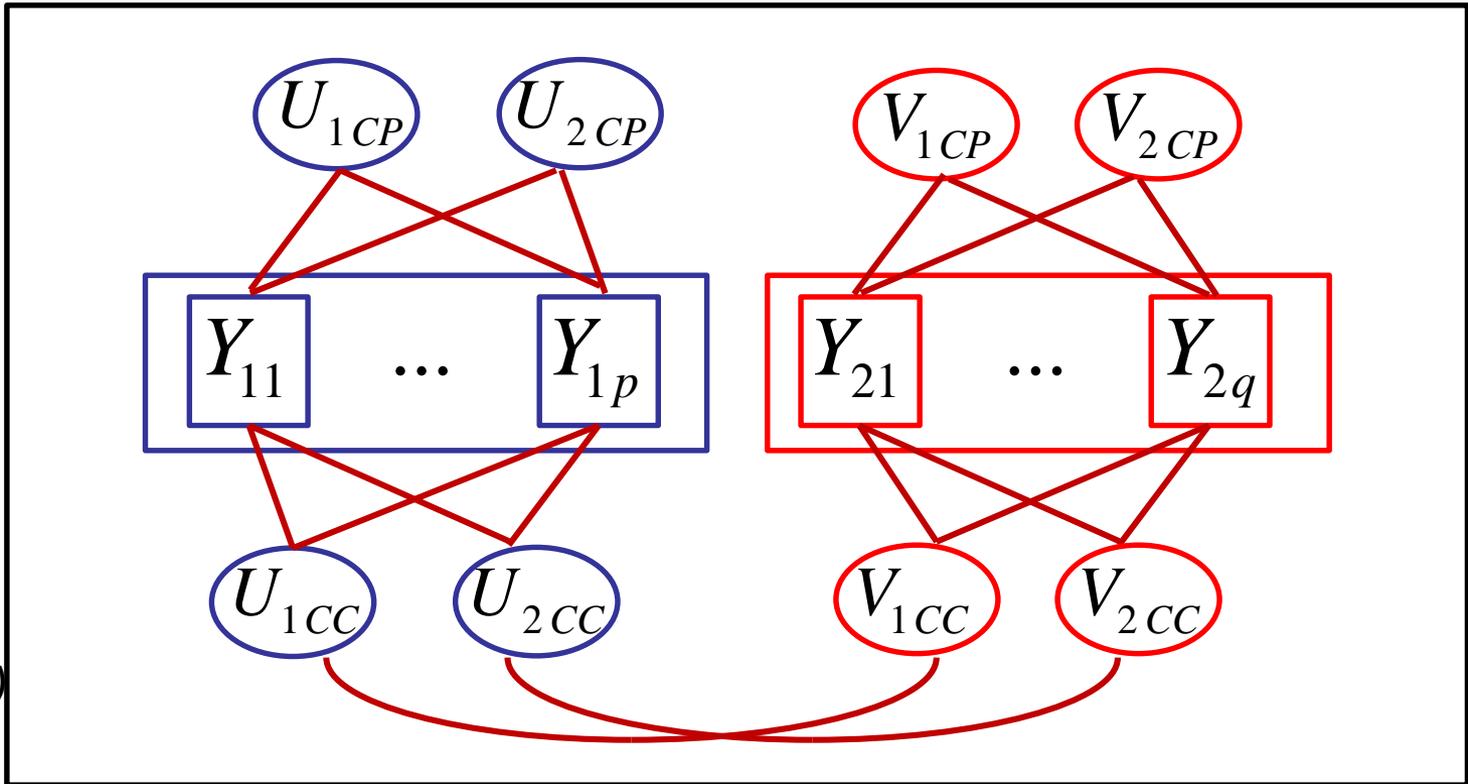
$$f_1(\Sigma_{11}^{-1} \Sigma_{12} \Sigma_{22}^{-1} \Sigma_{21}; a) = \frac{a' \Sigma_{12} \Sigma_{22}^{-1} \Sigma_{21} a}{a' \Sigma_{11} a}, \quad a' \Sigma_{11} a = 1$$

$$f_2(\Sigma_{22}^{-1} \Sigma_{21} \Sigma_{11}^{-1} \Sigma_{12}; b) = \frac{b' \Sigma_{21} \Sigma_{11}^{-1} \Sigma_{12} b}{b' \Sigma_{22} b}, \quad b' \Sigma_{22} b = 1$$

Redução de Dimensionalidade

Componente Principal
 $m \leq (n, "p")$

Correlação Canônica
 $m \leq \min(n, p, q)$



$$U_1 = a_1' Y_1 = e_1' \underbrace{\Sigma_{11}^{-1/2}}_{P_1 \Lambda^{-1/2} P_1'} Y_1 = e_1' P_1 \Lambda^{-1/2} \underbrace{P_1' Y_1}_{\text{CP de } Y_1}$$

Decomposição
 spectral de Σ_{11}

CP padronizado de Y_1
 (Fator Comum)

Redução de Dimensionalidade

Análise de Componentes Principais

$G=1, p=2, m \leq \min(n,p)$

$$Y_{n \times p} \Rightarrow Z_{n \times m}$$

$$\Sigma = V \Lambda V'$$

$$\Sigma V_j = \lambda_j V_j$$

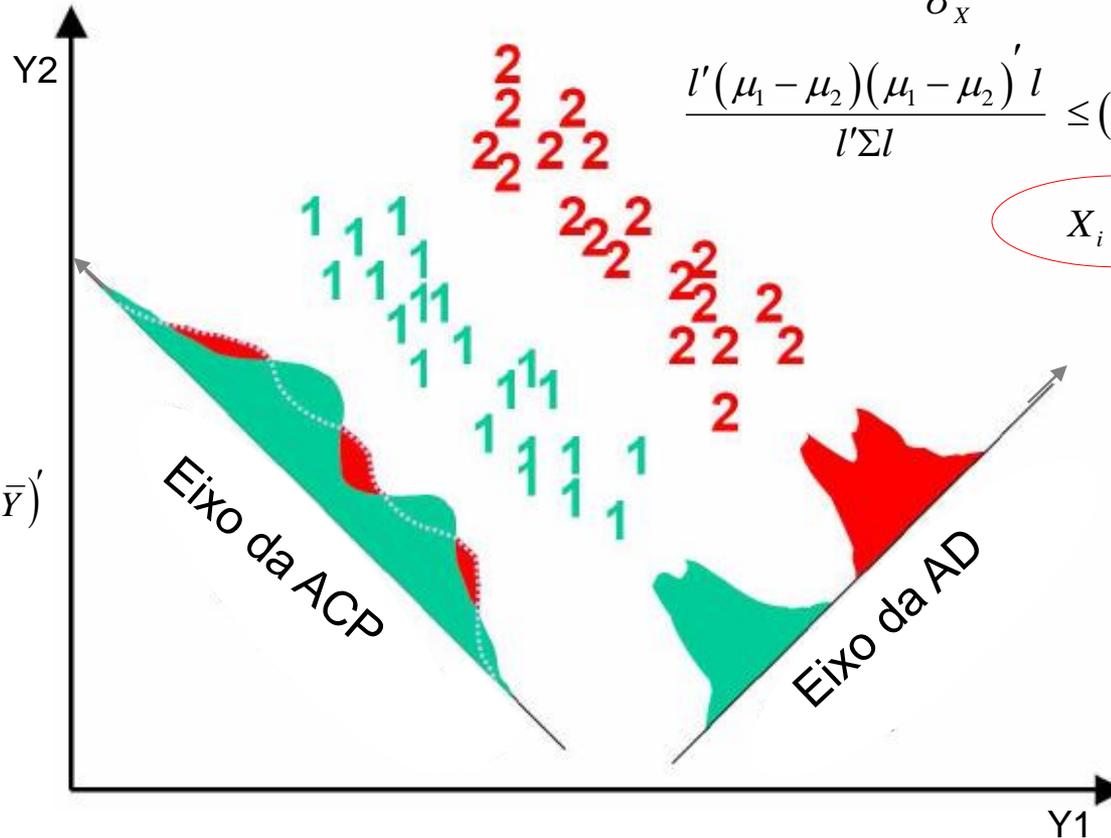
$$Z_{ji} = V_j' Y_i$$

$$V_j = a; \max_{\|a\|=1} \frac{a' \Sigma a}{a' a}$$

$$\hat{\Sigma} = \frac{1}{n-1} S_T$$

$$= \frac{1}{n-1} \sum_{i=1}^n (Y_i - \bar{Y})(Y_i - \bar{Y})'$$

$$Z_{n \times m} = Y_{n \times p} V_{p \times m}$$



Análise Discriminante

$G=2 (n=n_1+n_2), p=2, m \leq (n,p,G-1)$

$$Y_{n \times p}; n = \sum_{g=1}^G n_g; Y_{i \times p \times 1} \Rightarrow X_i = l' Y_i$$

$$\max_{l; X=l'Y} \frac{(\mu_{X_1} - \mu_{X_2})^2}{\sigma_X^2};$$

$$\frac{l'(\mu_1 - \mu_2)(\mu_1 - \mu_2)' l}{l' \Sigma l} \leq (\mu_1 - \mu_2)' \Sigma^{-1} (\mu_1 - \mu_2)$$

$$X_i = l' Y_i = (\bar{Y}_1 - \bar{Y}_2)' S_c^{-1} Y_i$$

$$S_T = S_B + S_W$$

$$S_c = \frac{1}{n-G} S_W$$

TABLE 1.
Wine tasting data from Abdi and Valentin (2007).

Wine	Oak-type	Expert 1			Expert 2				Expert 3		
		Fruity	Woody	Coffee	Red fruit	Roasted	Vanillin	Woody	Fruity	Butter	Woody
1	1	1	6	7	2	5	7	6	3	6	7
2	2	5	3	2	4	4	4	2	4	4	3
3	2	6	1	1	5	2	1	1	7	1	1
4	2	7	1	2	7	2	1	2	2	2	2
5	1	2	5	4	3	5	6	5	2	6	6
6	1	3	4	4	3	5	4	5	1	7	5

Redução via Componentes Principais

Análise dos dados padronizados:

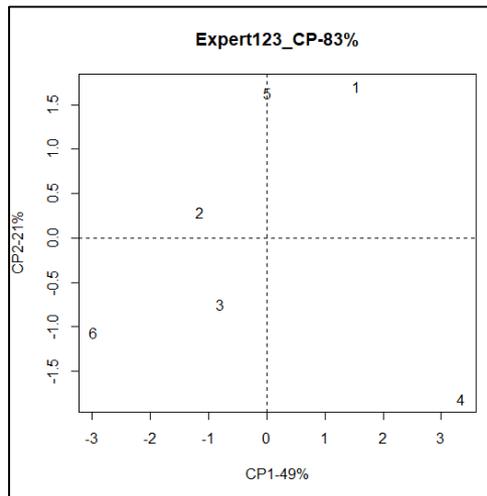
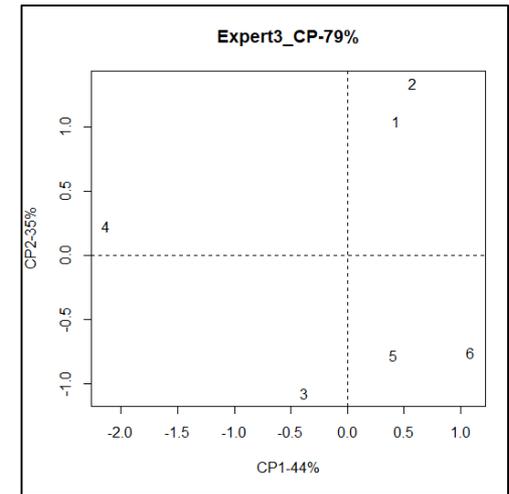
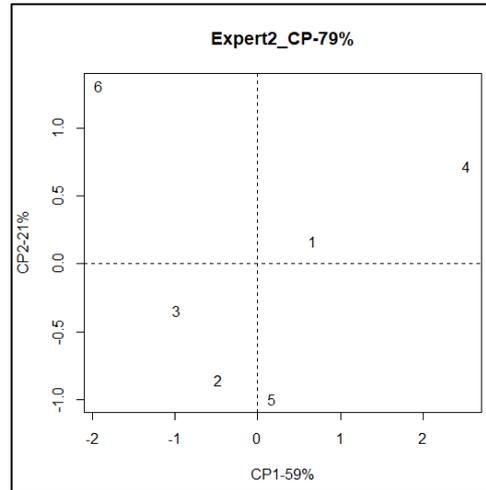
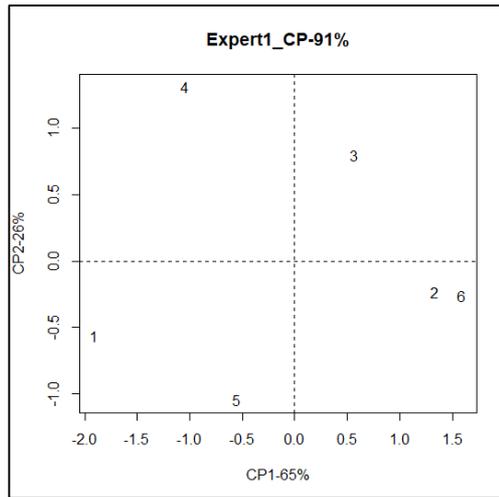
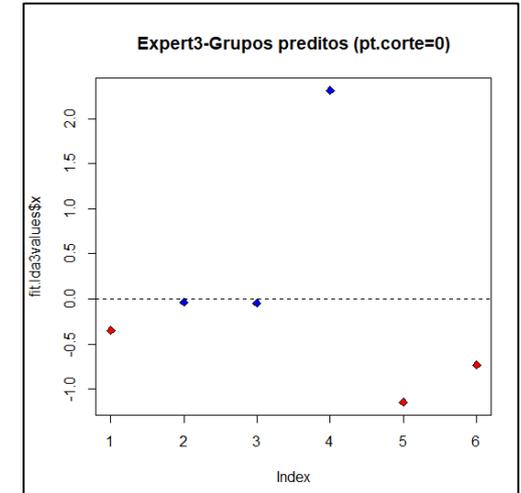
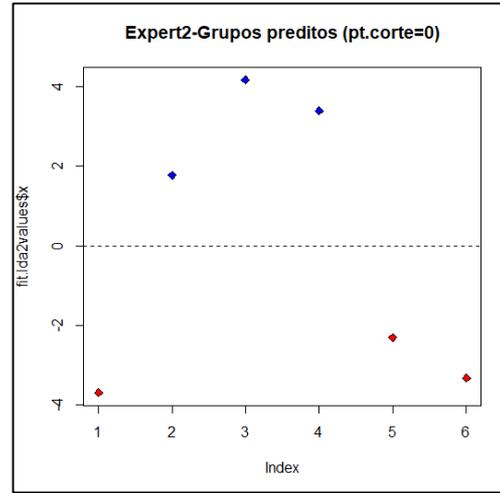
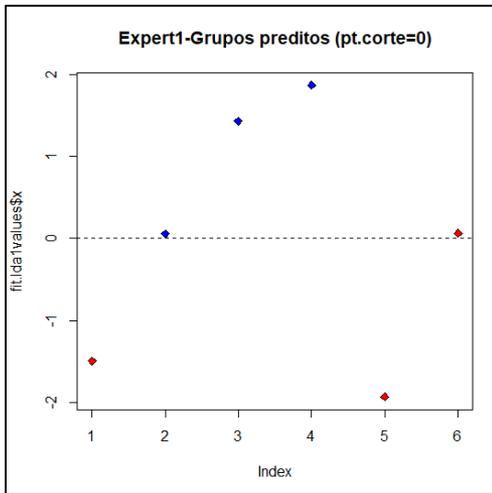


TABLE 1.
Wine tasting data from Abdi and Valentin (2007).

Wine	Oak-type	Expert 1			Expert 2				Expert 3		
		Fruity	Woody	Coffee	Red fruit	Roasted	Vanillin	Woody	Fruity	Butter	Woody
1	1	1	6	7	2	5	7	6	3	6	7
2	2	5	3	2	4	4	4	2	4	4	3
3	2	6	1	1	5	2	1	1	7	1	1
4	2	7	1	2	7	2	1	2	2	2	2
5	1	2	5	4	3	5	6	5	2	6	6
6	1	3	4	4	3	5	4	5	1	7	5

Redução via Análise Discriminante

Análise dos dados padronizados:



Gr	Pred1	Pred2
1	2	1
2	0	3

%ClassCorreta		
1	2	Total
0.67	1.00	0.83

Gr	Pred1	Pred2
1	3	0
2	0	3

%ClassCorreta		
1	2	Total
1.00	1.00	1.00

Gr	Pred1	Pred2
1	3	0
2	2	1

%ClassCorreta		
1	2	Total
1.00	0.33	0.67

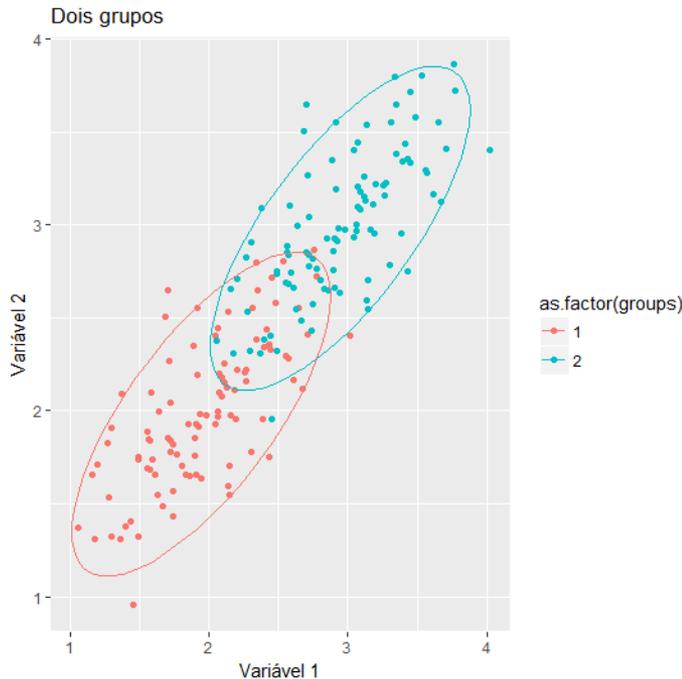
Onde estão os Vetores Reducionistas?

Um gráfico pode valer mais que mil palavras mas pode exigir milhares de palavras para construí-lo. Tukey

Obter a direção do CP e do Eixo Discriminante.

Observações independentes. Indicação da elipse de concentração (95%).

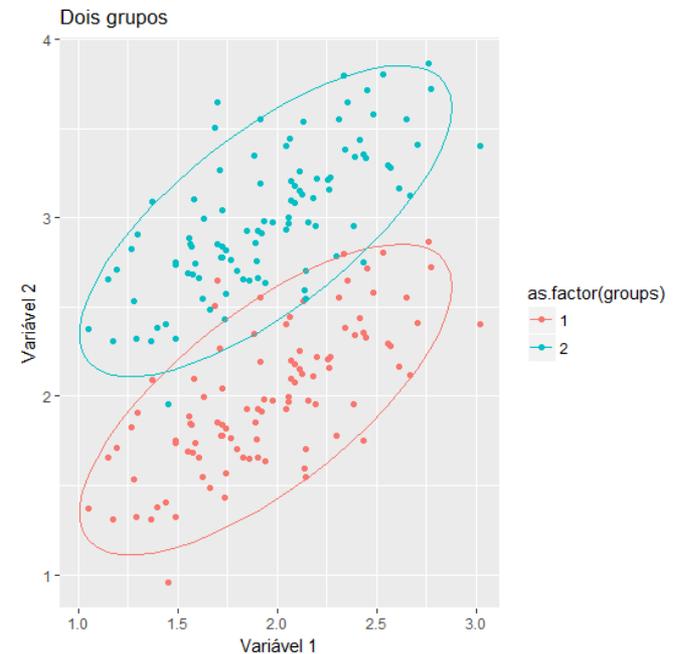
Exemplo 1



$$n = 200, p = 2, \mu_1 = (2,2), \mu_2 = (3,3)$$

$$R_{2 \times 2} = \begin{pmatrix} 1 & 0.7 \\ 0.7 & 1 \end{pmatrix}; \sigma = (0.4 \quad 0.4)$$

Exemplo 2



$$n = 200, p = 2, \mu_1 = (2,2), \mu_2 = (2,3)$$

$$R_{2 \times 2} = \begin{pmatrix} 1 & 0.7 \\ 0.7 & 1 \end{pmatrix}; \sigma = (0.4 \quad 0.4)$$

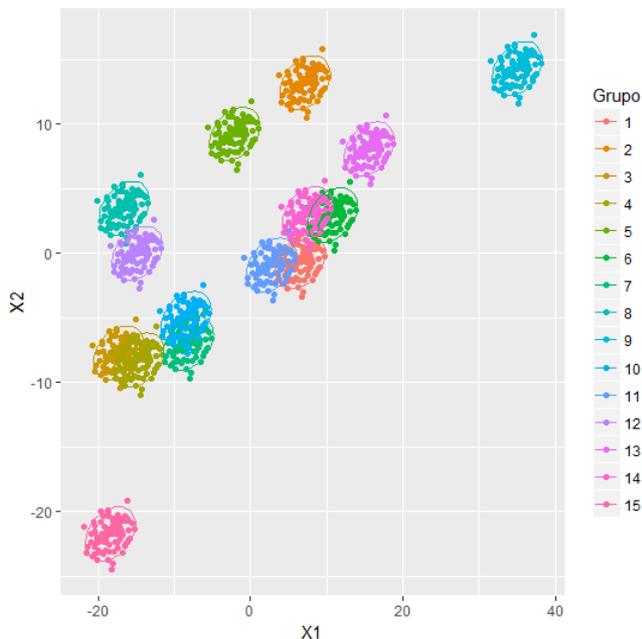
Onde estão os Vetores Reducionistas?

Obter a direção do Eixo Discriminante.

Observações independentes ENTRE e DENTRO de grupos.

Exemplo 3: “Sinais Iguais”

$$T = B + W$$

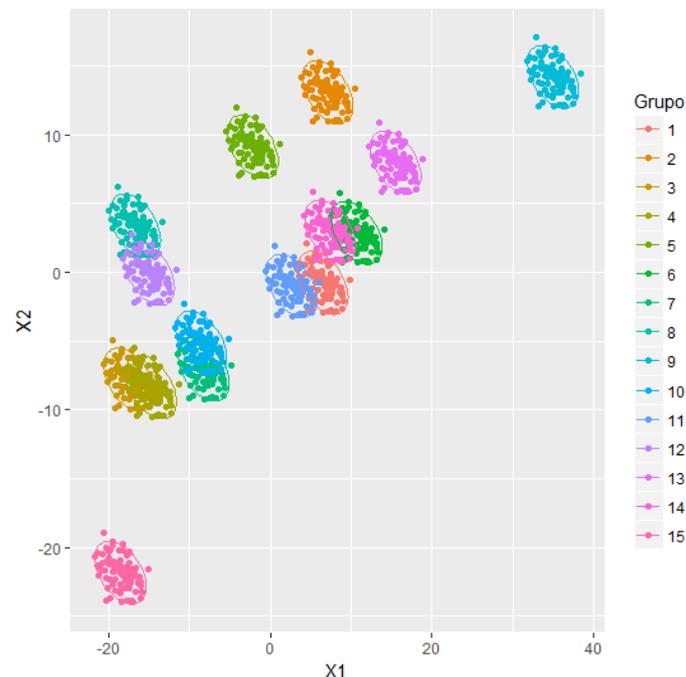


$$G = 15, n_g = 100, \mu = (0, 0)$$

$$S_b = \begin{pmatrix} 150 & 100 \\ 100 & 150 \end{pmatrix}, S_w = \begin{pmatrix} 2 & 0.5 \\ 0.5 & 1 \end{pmatrix}$$

Exemplo 4: “Sinais Opostos”

$$T = B + W$$



$$G = 15, n_g = 100, \mu = (0, 0)$$

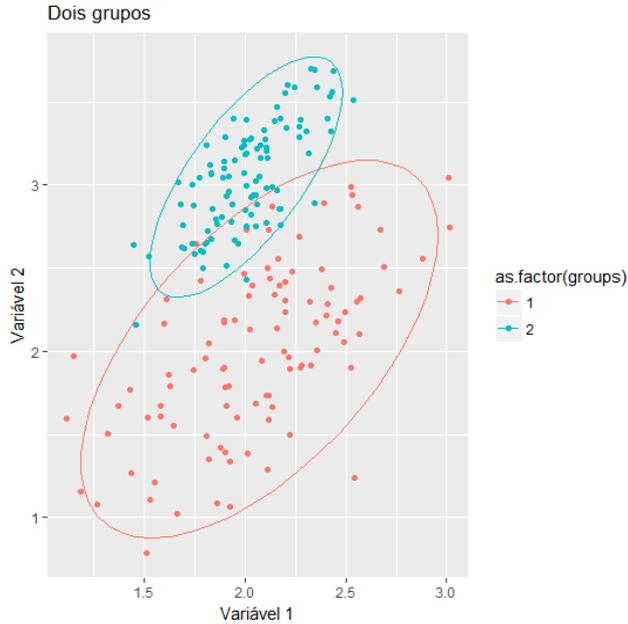
$$S_b = \begin{pmatrix} 150 & 100 \\ 100 & 150 \end{pmatrix}, S_w = \begin{pmatrix} 2 & -0.5 \\ -0.5 & 1 \end{pmatrix}$$

Onde estão os Vetores Reducionistas?

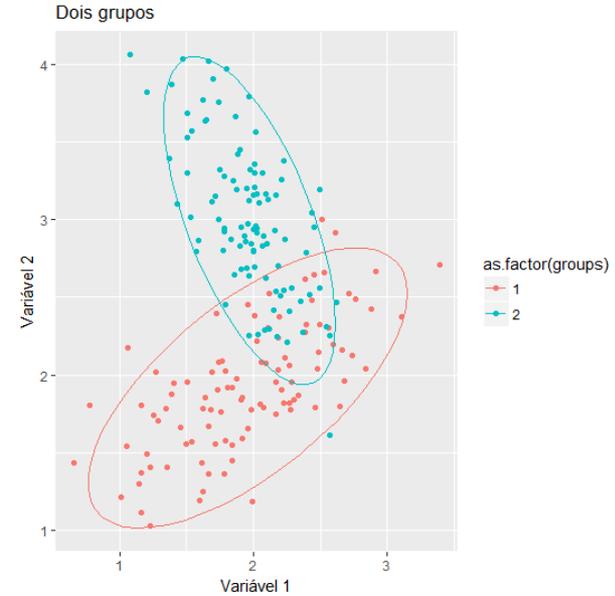
Obter a direção da Variável Canônica.

Observações independentes avaliadas em \mathfrak{R}^{p+q} .

Exemplo 5: Correlações de mesmo sinal



Exemplo 6: Correlações de sinal diferentes



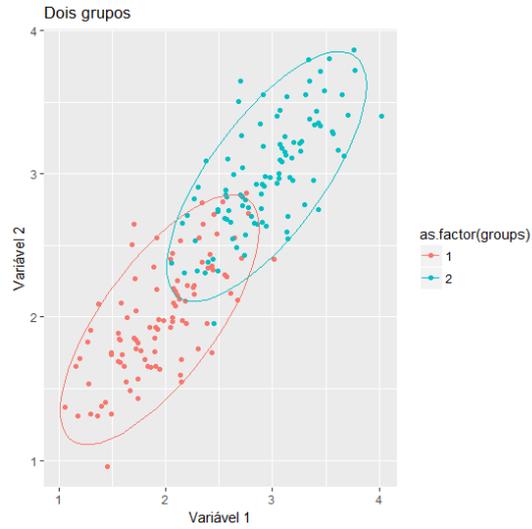
$$R = \begin{pmatrix} 1 & 0.65 & 0.2 & 0.5 \\ 0.65 & 1 & 0.1 & 0.4 \\ 0.2 & 0.1 & 1 & 0.7 \\ 0.5 & 0.4 & 0.7 & 1 \end{pmatrix}$$

$$\sigma_1 = (0.4 \quad 0.5); \sigma_2 = (0.2 \quad 0.3); \mu_1 = (2,2); \mu_2 = (3,3)$$

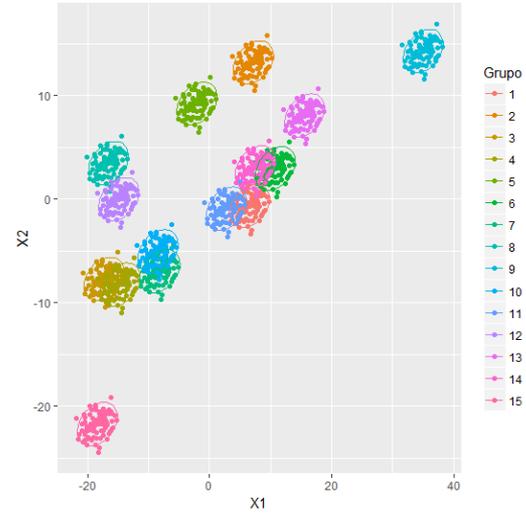
$$R = \begin{pmatrix} 1 & 0.7 & 0.2 & 0.5 \\ 0.7 & 1 & 0.3 & 0.4 \\ 0.2 & 0.3 & 1 & -0.7 \\ 0.5 & 0.4 & -0.7 & 1 \end{pmatrix}$$

$$\sigma_1 = (0.5 \quad 0.4); \sigma_2 = (0.3 \quad 0.5); \mu_1 = (2,2); \mu_2 = (3,3)$$

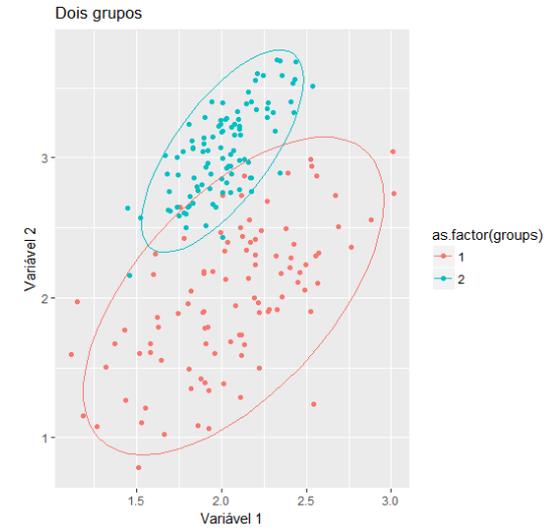
Exemplo 1



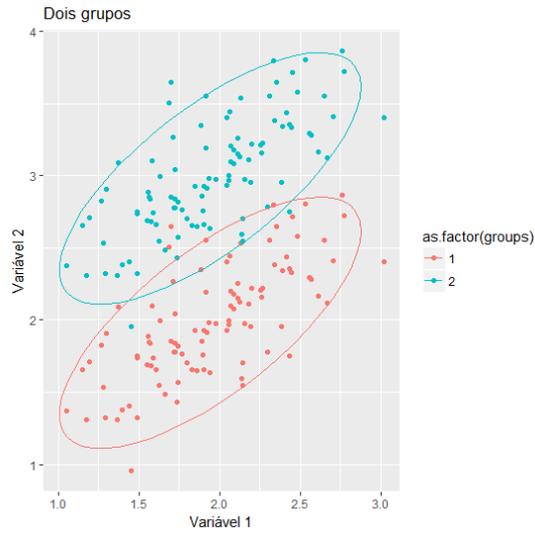
Exemplo 3



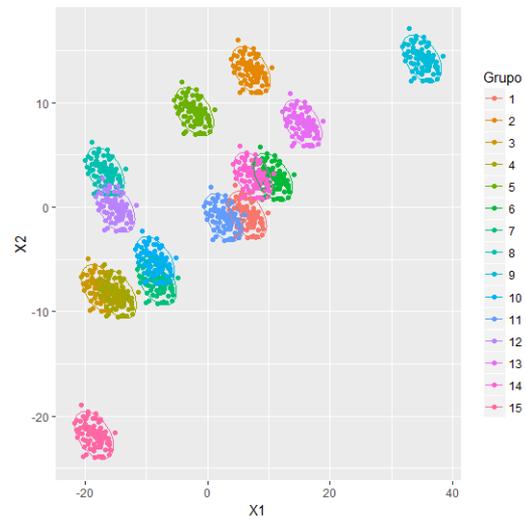
Exemplo 5



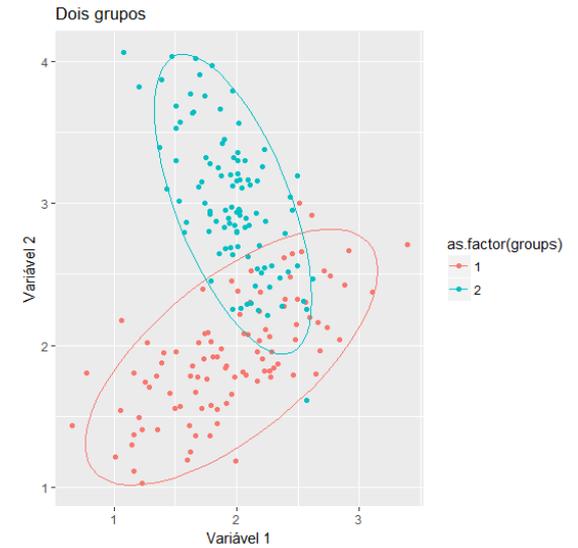
Exemplo 2



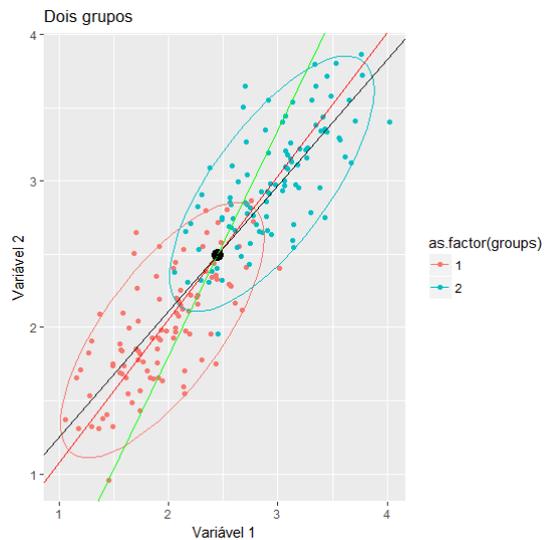
Exemplo 4



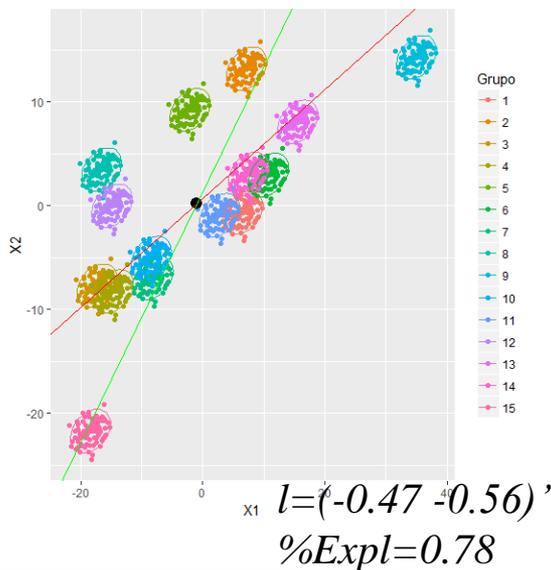
Exemplo 6



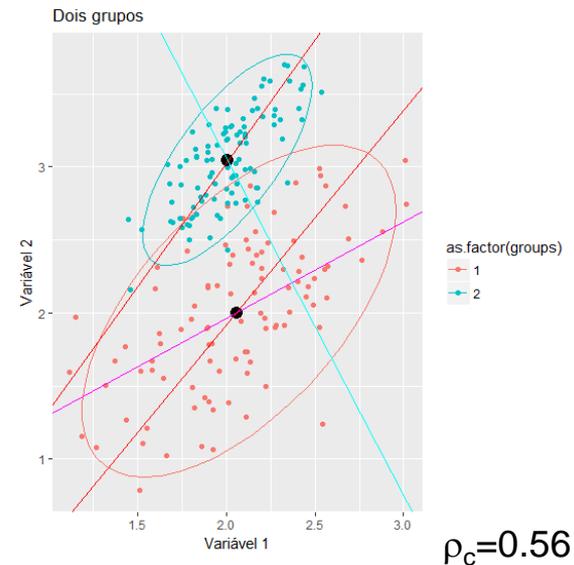
Exemplo 1



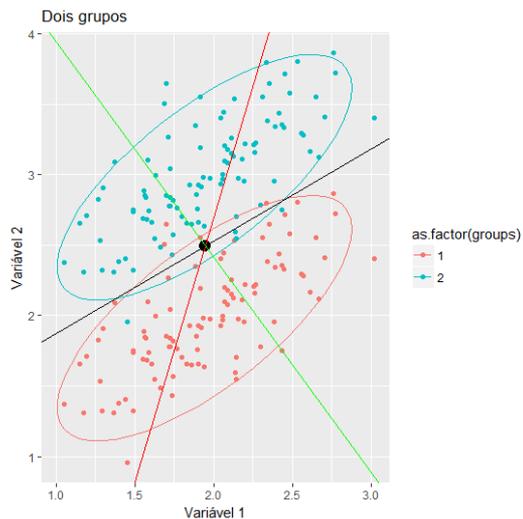
Exemplo 3



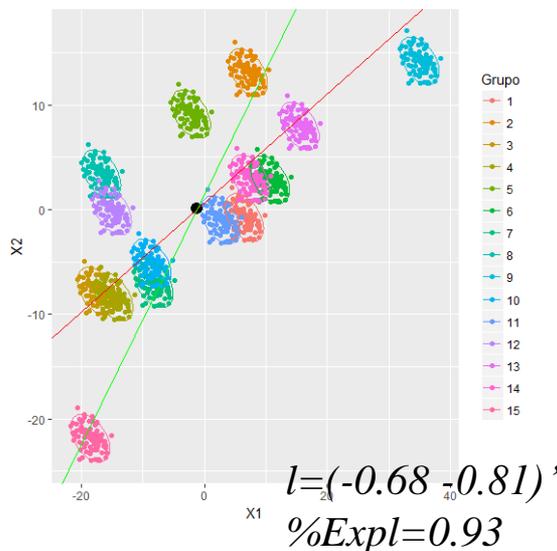
Exemplo 5



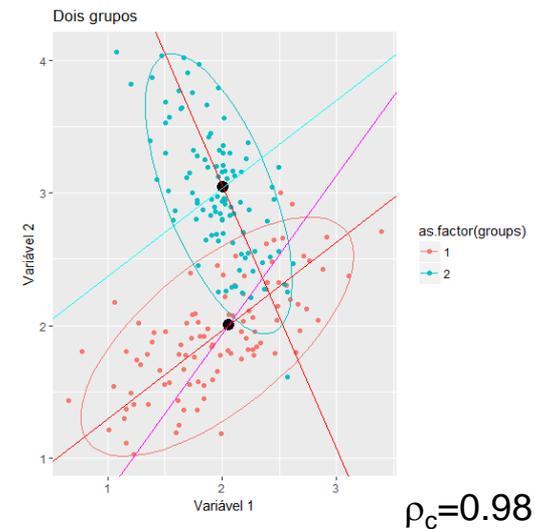
Exemplo 2



Exemplo 4



Exemplo 6



Preto:reta de MQ Vermelho:vetor de CP Verde:vetor discriminante Azul e rosa:variáveis canônicas

Análise Multivariada

Já vimos 😊

- Metodologias Clássicas: Foco na obtenção de vetores reducionistas

$$Y_{n \times p} = (Y_{ij}) \in \mathfrak{R}^{n \times p} \quad \mathfrak{R}^p \rightarrow \mathfrak{R}^m$$

- ✓ Componentes Principais - CP ($m \leq \min(n, p)$)
- ✓ Análise Fatorial Exploratória - AF (via CP: $m \leq \min(n, p)$)
- ✓ Coordenadas Principais (CoP) – Escalonamento Multidimensional ($m \leq \text{posto}(D_{n \times n})$)
- ✓ Análise de Correspondência – AC ($m \leq \min(I-1, J-1)$)
- ✓ Análise Discriminante - DA ($m \leq \min(n, p, G-1)$)
- ✓ Análise de Agrupamento
- ✓ Correlação Canônica - CCA ($m \leq \min(n, p, q)$)

Combinações
lineares de Y

$n > p$
Observações
Independentes

- ✓ Problema Dual de redução de dimensionalidade ($\mathfrak{R}^{n \times p}$, $\mathfrak{R}^{p \times p}$, $\mathfrak{R}^{n \times n}$)
- ✓ Representações Bi-Plot



PLS: Mínimos Quadrados Parciais
PLS – CCA – DA

Componentes Principais via Modelos de Regressão

Revisando!

Como obter as variáveis originais a partir das componentes principais?

$$Y \in \mathbb{R}^p \Rightarrow Z = V' Y \in \mathbb{R}^p; \quad Z_j = a_{1j}Y_1 + a_{2j}Y_2 + \dots + a_{pj}Y_p \quad V = (a_{ij}); \quad V' = (a_{ji})$$

$$Z = V' Y \Leftrightarrow V Z = V V' Y = Y \Rightarrow Y = V Z$$

$$Y_j = a_{j1}Z_1 + a_{j2}Z_2 + \dots + a_{jp}Z_p$$

$$\text{Var}(Y_j) = \sum_{k=1}^p a_{jk}^2 \text{Var}(Z_k) = \sum_{k=1}^p a_{jk}^2 \lambda_k \Rightarrow R_{jk}^2 = \frac{a_{jk}^2 \lambda_k}{\sigma_{jj}}$$

é o quadrado da correlação entre Y_j e Z_k

Proporção da $\text{Var}(Y_j)$ explicada por Z_k . É o coeficiente de determinação da regressão de Y_j em Z_k .

Aplicação de CP em Modelos de Regressão

Considere a seguinte situação da Regressão de Y em p Variáveis Predictoras:

$$\text{Matriz de dados: } \begin{pmatrix} Y_{n \times 1}, X_{1n \times 1}, \dots, X_{pn \times 1} \end{pmatrix}_{n \times (p+1)}$$

$Y_i \in \mathbb{R}$

Objetivo: Modelo de Regressão de Y nos Componentes Principais das p Variáveis Predictoras (problemas de Multicolinearidade)

Componentes Principais e Modelos de Regressão

Seleção de Variáveis Predictoras

Redução de Dimensionalidade em Modelos de Regressão

$$Y_i = \beta_0 + \sum_{j=1}^p \beta_j X_{ij} + e_i; \quad Y_{n \times 1} = 1_n \beta_0 + X_{n \times p} \beta_{p \times 1} + e_{n \times 1}$$

Seleção de variáveis predictoras por **PCR** (Componentes Principais de Var. Predictoras)

$$X = U \Lambda V'; \quad Z = XV, \quad X = ZV'$$

$$\Rightarrow Y_{n \times 1} = 1_n \beta_0 + Z_{n \times p} V'_{p \times p} \beta_{p \times 1} + e_{n \times 1} = \alpha_0 + Z_{n \times p} \alpha_{p \times 1} + e_{n \times 1}; \quad \alpha = V' \beta$$

$$\begin{array}{ccc} X_1, \dots, X_p & \rightarrow & Z_1, \dots, Z_m; \\ \text{Estimação em } \mathfrak{R}^{(p+1)} & & \text{Estimação em } \mathfrak{R}^{(m+1)} \end{array}; \quad Z_{ki} = \sum_{j=1}^p a_{jk} X_{ij}, \quad k = 1, \dots, m; \quad m < p$$



$$Y_i = \alpha_0 + \sum_{k=1}^m \alpha_k Z_{ki} + e_i$$

Variáveis Z são não correlacionadas \Rightarrow Útil na seleção de variáveis predictoras.

Restrições impostas pela redução de dimensionalidade

$$\sum_{k=1}^m \alpha_k Z_{ki} = \sum_{k=1}^m \alpha_k \sum_{j=1}^p a_{jk} X_{ij} = \sum_{j=1}^p \left[\sum_{k=1}^m \alpha_k a_{jk} \right] X_{ij} = \sum_{j=1}^p \beta_j X_{ij}$$

Exemplo: Gerar $n=50$ observações da N_1 com $p=45$ variáveis predictoras.

Componentes Principais e Modelos de Regressão

PLS

PLS: Mínimos Quadrados Parciais

(Tibshirani et al., 2015; Abdi, 2010)

$$Y_i \in \mathfrak{R}$$

$$Y_i = \beta_0 + \sum_{j=1}^p \beta_j X_{ij} + e_i; \quad X_1, \dots, X_p \rightarrow Z_1, \dots, Z_m$$

Obter direções (variáveis Z) em X que melhor explicam Y.

Estimação em $\mathfrak{R}^{(p+1)}$

PCR: variáveis Z só explicavam X

$$Y_i = \varphi_{1j} X_{ij} + e_i, \quad j = 1, 2, \dots, p \xrightarrow{\hat{\varphi}_{1j} : \text{Solução de MQ}} Z_{1i} = \sum_{j=1}^p \hat{\varphi}_{1j} X_{ij}$$

$$X_{ij} = a_{0j} + b_j Z_{1i} + e_{(X_j|Z_1)i} \quad j = 1, \dots, p \Rightarrow \hat{e}_{(X_j|Z_1)i}$$

Algoritmo de obtenção dos escores

$$Y_i = \varphi_{2j} \hat{e}_{(X_j|Z_1)i} + e_i, \quad j = 1, 2, \dots, p \longrightarrow Z_{2i} = \sum_{j=1}^p \hat{\varphi}_{2j} X_{ij}$$

CP de X que melhor predizem Y

$$\Rightarrow Z_1, \dots, Z_m;$$

$$Y_i = \alpha_0 + \sum_{k=1}^m \alpha_k Z_{ki} + e_i$$

Estimação em $\mathfrak{R}^{(m+1)}$

Regressão PLS: Formulação Geral

(Tibshirani et al., 2015; Abdi, 2010)

$$Y_i \in \mathbb{R}^p$$

$$\begin{pmatrix} Y_{n \times p} \\ X_{n \times q} \end{pmatrix}_{n \times (p+q)}$$

Obter “Componentes Principais” de X que melhor predizem Y

$$X_{n \times q} = \underbrace{F_{X_{n \times m}}}_{\text{Matriz de Escores (CoP = CP)}} \underbrace{W'_{X_{m \times q}}}_{\text{Matriz de Cargas}}$$

Matriz de Escores (CoP = CP) Matriz de Cargas

$$Y_{n \times p} = \underbrace{F_{Y_{n \times m}}}_{\text{Matriz de Escores (CoP = CP)}} \underbrace{W'_{Y_{m \times p}}}_{\text{Matriz de Cargas}} = \underbrace{F_X}_{\text{Matriz de Escores (CoP = CP)}} \underbrace{B}_{\text{Matriz de Cargas}} \underbrace{W'_Y}_{\text{Matriz de Cargas}}; \quad B = \text{diag}(d_j)$$

Matriz de Escores (CoP = CP) Matriz de Cargas



Obter

$$F_{X_j}, F_{Y_j}; \max_{F_{X_j}, F_{Y_j}} F_{X_j}' F_{Y_j} = d_j$$

Maximizar o produto interno dos Escores

Considerando X e Y normalizadas, Passo1: obter as primeiras soluções

$$\begin{cases} (X'Y)_{q \times p} = W'_x B W'_Y \\ (X(X'Y)Y')_{n \times n} = F'_X F'_Y \\ (Y(Y'X)X')_{n \times n} = F'_Y F'_X \end{cases}$$

Algoritmo: obtenção dos escores

Atualizar X e Y, e usar o resíduo do Passo1, até obter m componentes.



$$\begin{cases} X^{(2)} = X - X^{(1)} = X - F_{X1} W_{X1}' \\ Y^{(2)} = Y - Y^{(2)} = Y - F_{Y1} W_{Y1}' d_1 \end{cases}$$

Essas direções não são iguais às da **Correlação Canônica** (que é uma solução simétrica, diferentemente de **PLS** que é assimétrica/direcionada)

Regressão PLS: Formulação Geral

Obter “Componentes Principais” de X que melhor predizem Y

$$\begin{pmatrix} Y_{n \times p} & X_{n \times q} \end{pmatrix}_{n \times (p+q)}$$

$$X_{n \times q} = F_{X n \times m} W'_{X m \times q}$$

$$Y_{n \times p} = F_{Y n \times m} W'_{Y m \times p} = F_X B W'_Y; \quad B = \text{diag}$$

$$(XY)_{q \times p} = W_X B W'_Y$$

$$(X(X'Y)Y')_{n \times n} = F_X F'_Y$$

$$(Y(Y'X)X')_{n \times n} = F_Y F'_X$$

$$X^{(2)} = X - F_{X1} W'_{X1}$$

$$Y^{(2)} = Y - F_{Y1} W'_{Y1} d_1$$

$$(F_{Xj}, F_{Yj}), j = 1, \dots, m$$

$$(F_X, F_Y)_{n \times m}$$

Algoritmo: obtenção dos escores

$$(F_X, F_Y)_{n \times m} \begin{cases} F_X = XW_X & F_Y = YW_Y \\ \hat{Y} = F_X B W'_Y = XW_X B W'_Y = X B_{PLS}; & B_{PLS} = W'_X B W_Y \end{cases}$$

Regressão PLS e Correlação Canônica

Integração de Bancos de Dados

Solução PLS (Partial Least Square):

$$a, b; \max_{a \in \mathbb{R}^p, b \in \mathbb{R}^q} \frac{[Cov(a'X; b'Y)]^2}{(a'a)(b'b)} \Rightarrow \Sigma_{XY} \Sigma_{YX} a = \lambda a, \quad b = \Sigma_{YX} a$$

Solução CCA (Canonical Correlation Analysis):

$$a, b; \max_{a \in \mathbb{R}^p, b \in \mathbb{R}^q} [Corr(a'X; b'Y)]^2$$



$$[Cov(a'X; b'Y)]^2 = Var(a'X) [Corr(a'X; b'Y)]^2 Var(b'Y)$$

PLS é a CCA com regularizações
definidas pelos PC de X e PC de Y

Regressão PLS , CCA e DA: Integração de BD

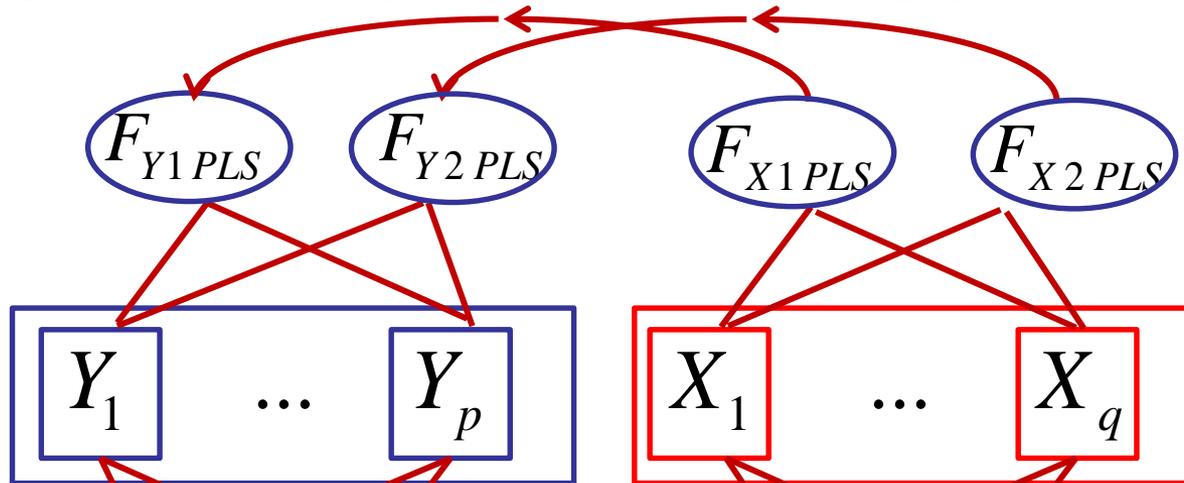
PLS $\left[Cov(a'X; b'Y) \right]^2 = Var(a'X) \left[Corr(a'X; b'Y) \right]^2 Var(b'Y)$ CCA

Bartlet (1938): estabelece conexões entre CCA e DA \Rightarrow CCA pode ser realizada com um conjunto X de treinamento e uma matriz Y de variáveis indicadoras da Doença (Matriz de var. dummy)



Barker and Raynes (2003) propõem PLS-DA: Assim, um procedimento formulado para problemas de regressão múltipla, passa a ser usado para prever estrutura de grupos.

PLS



CCA

DA

