

MAE 5776

ANÁLISE MULTIVARIADA

Júlia M Pavan Soler

pavan@ime.usp.br

1º Sem/2020 - IME

Análise Multivariada

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

Já vimos 😊

- Estatísticas descritivas multivariadas, Episódios de Concentração, Boxplot Bivariado
- Distribuição N_p , Distribuições Amostrais (T^2 e W_p)
- $N_p(\mu_g; \Sigma_g)$: Inferências sobre μ_g (T^2 , MANOVA, ICS, Correções para Múltiplos testes)

Decomposições: SS_T e $Y_{n \times p}$



Técnicas Multivariadas:

✓ Análise de Componentes Principais

- Escalonamento Multidimensional
- Análise de Correspondência

- Análise Fatorial
- Análise Discriminante (MANOVA)
- Análise de Agrupamento
- Análise de Correlação Canônica

Dados dos cães

	Y1	Y2	Y3	Y4	Y5	Y6
[1,]	9.7	21.0	19.4	7.7	32.0	36.5
[2,]	8.1	16.7	18.3	7.0	30.3	32.9
[3,]	13.5	27.3	26.8	10.6	41.9	48.1
[4,]	11.5	24.3	24.5	9.3	40.0	44.6
[5,]	10.7	23.5	21.4	8.5	28.8	37.6
[6,]	9.6	22.6	21.1	8.3	34.4	43.1
[7,]	10.3	22.1	19.1	8.1	32.2	35.0

Revisando 😊

Componentes Principais

$$Y_{n \times p} \rightarrow Z_{n \times p} = YV; \quad Z_{ik} = V_k' Y_i$$

$$\Sigma_{p \times p} = V \Lambda V' \rightarrow Cov(Z) = \Lambda = Diag(\lambda_j)$$

$$tr(\Sigma) = tr(\Lambda) = \sum_{j=1}^p \lambda_j$$

Matriz de Covariância S:

	Y1	Y2	Y3	Y4	Y5	Y6
Y1	2.88	5.25	4.85	1.93	6.53	7.74
Y2	5.25	10.56	8.90	3.59	11.46	15.58
Y3	4.85	8.90	9.61	3.51	13.43	16.31
Y4	1.93	3.59	3.51	1.36	4.86	5.92
Y5	6.53	11.46	13.43	4.86	24.36	24.68
Y6	7.74	15.58	16.31	5.92	24.68	31.52

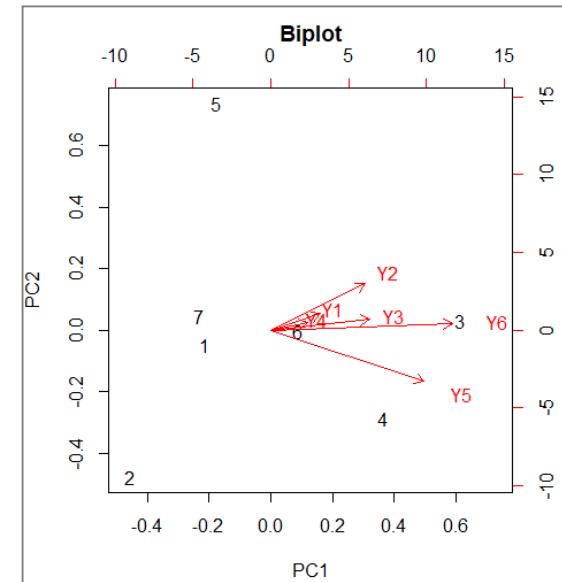
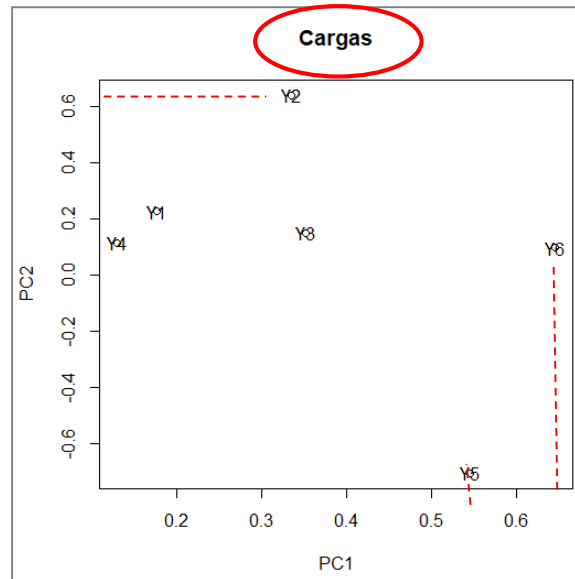
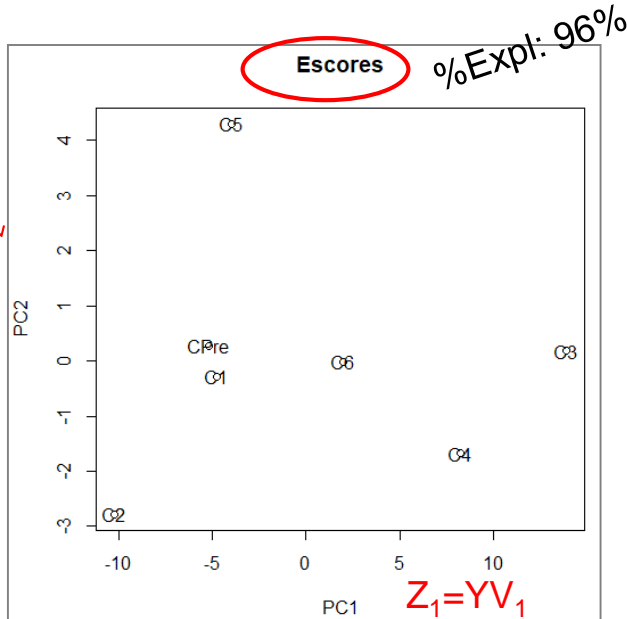
Autovalores de S (λ_j):

72.61 4.86 2.13 0.66 0.02 0.01

Autovetores de S (V_j):

	V1	V2	V3	V4	V5	V6
Y1	0.18	0.23	-0.41	0.10	0.65	-0.56
Y2	0.34	0.64	-0.33	-0.47	-0.37	0.09
Y3	0.35	0.15	-0.15	0.84	-0.36	0.01
Y4	0.13	0.11	-0.15	0.11	0.52	0.82
Y5	0.55	-0.70	-0.39	-0.21	-0.09	0.03
Y6	0.65	0.10	0.72	-0.08	0.18	-0.09

Cargas



Escalonamento Multidimensional ou Análise de Coordenadas Principais

$Y_{n \times p}$: não é conhecida
Somente D (ou C) é conhecida

Dados Multivariados

$$D = \begin{pmatrix} 0 & & & \\ d_{21} & 0 & & \\ \dots & \dots & \dots & \\ d_{n1} & d_{n2} & \dots & 0 \end{pmatrix}_{n \times n}$$

D: Matriz de Distâncias entre indivíduos

$$C = \begin{pmatrix} 1 & & & \\ r_{21} & 1 & & \\ \dots & \dots & \dots & \\ r_{n1} & r_{n2} & \dots & 1 \end{pmatrix}_{n \times n}$$

C: Matriz de Similaridades entre indivíduos

Objetivos:

- A partir de matrizes de distância D (ou similaridade C) entre n objetos obter uma representação das correspondentes observações $Y_{n \times p}$ que geraram D (ou C);
- A partir de D (ou C) obter Eixos Principais (Coordenadas Principais) \Rightarrow Identificar a dimensão das observações multivariadas que geraram D



Escalonamento Multidimensional

Análise baseada no espaço linha da matriz de dados

Escalonamento Multidimensional

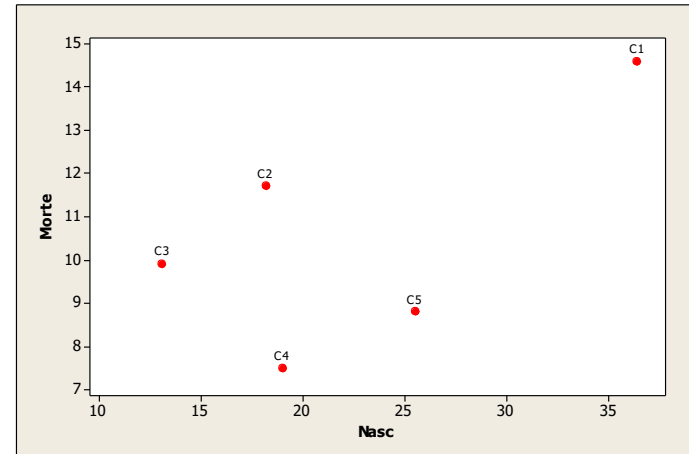
Motivação: Representação gráfica de observações

$Y_{5 \times 2}$: Matriz conhecida

Cidade	Nascimento	Morte
C1	36,4	14,6
C2	18,2	11,7
C3	13,1	9,9
C4	19	7,5
C5	25,5	8,8



Representação das cidades



$$d_{12}; d_{12}^2 = (36,4 - 18,2)^2 + (14,6 - 11,7)^2$$

	C1	C2	C3	C4	C5
C1	0				
C2	18,43	0			
C3	23,76	5,41	0		
C4	18,79	4,28	6,37	0	
C5	12,34	7,85	12,45	6,63	0

Suponha que o único dado disponível corresponde à Matriz de Distância Euclidiana D entre as observações.

Como representar as observações (obter Y)?

Escalonamento Multidimensional

Matriz de Distância Euclidiana entre os 7 cães (p=6 variáveis)

Matriz de Distância Euclidiana

	1	2	3	4	5	6	7
1	0						
2	6.21	0					
3	18.70	24.34	0				
4	13.13	18.55	5.99	0			
5	4.83	9.44	18.38	13.64	0		
6	7.43	12.94	12.50	7.26	7.98	0	
7	2.03	6.62	19.20	13.78	5.09	8.67	0

Como representar os 7 pontos em um gráfico?

$$D_{7 \times 7} \xrightarrow{?} Y_{7 \times m}; m = 2$$

Escalonamento Multidimensional

Matriz de Distância* (“postos”) entre 6 Docerias

	A	B	C	D	E	F
A	-					
B	2	-				
C	13	12	-			
D	4	6	9	-		
E	3	5	10	1	-	
F	8	7	11	14	15	-

D não é matriz de distância Euclidiana mas uma distância empírica!

*1: é o par mais similar 15: é o par menos similar $(6(6-1)/2)$

Como representar os 6 pontos em um gráfico?

$$D_{6 \times 6} \xrightarrow{?} Y_{6 \times m}; \quad m = 2$$

Escalonamento Multidimensional

Distâncias (em km) entre 12 cidades \Rightarrow matriz de “distância” empírica

	1	2	3	4	5	6	7	8	9	10	11	12
1	0											
2	244	0										
3	218	350	0									
4	284	77	369	0								
5	197	167	347	242	0							
6	312	444	94	463	441	0						
7	215	221	150	236	279	245	0					
8	469	583	251	598	598	169	380	0				
9	166	242	116	257	269	210	55	349	0			
10	212	53	298	72	170	392	168	531	190	0		
11	253	325	57	340	359	143	117	264	91	273	0	
12	270	168	284	164	277	378	143	514	174	111	256	0

D não é matriz de distância Euclidiana mas uma distância empírica!

Como representar os 12 pontos em um gráfico?

$$D_{12 \times 12} \xrightarrow{?} Y_{12 \times m}; \quad m = 2$$

Escalonamento Multidimensional

Notação:

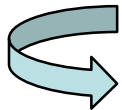
Dada uma matriz de distâncias D ,

$$D = (d_{ij})_{n \times n}$$

O objetivo do Escalonamento Multidimensional é predizer pontos, $\hat{Y}_i \in \mathcal{R}^m$

$i=1, \dots, n$, tal que, se \hat{d}_{ik} é a distância Euclidiana entre \hat{Y}_i e \hat{Y}_k , então

$\hat{D} = (\hat{d}_{ik})$ é uma “boa aproximação” para D .



Solução:

- **Métodos métricos** \Rightarrow baseados na teoria da decomposição espectral
- **Métodos não métricos** \Rightarrow baseados na minimização de funções objetivo como o “stress”.

Escalonamento Multidimensional

Solução Clássica (Métrica) na dimensão m

(Mardia, 1979)

Dado D , matriz de distância Euclidiana \Leftrightarrow Existe $Y_{n \times p}$ matriz de dados tal que:

$$D = (d_{ik})_{n \times n}; \quad d_{ik}^2 = \sum_{j=1}^p (y_{ij} - y_{kj})^2$$

d_{ik} : conhecido

y_{ik} : desconhecido

Logo, existe: $B_{n \times n} = Y_{n \times p} Y'_{p \times n} \Rightarrow b_{ik} = \sum_{j=1}^p y_{ij} y_{kj}$

Matriz de Produto Interno entre as linhas de Y

$$d_{ik}^2 = b_{ii} + b_{kk} - 2b_{ik}$$

$$b_{ik} = -\frac{1}{2} (d_{ik}^2 - d_{i.}^2 - d_{.k}^2 + d_{..}^2)$$

B: calculada de **D**

Escalonamento Multidimensional

Solução Clássica na dimensão m

$$D = (d_{ik})_{n \times n} \Leftrightarrow Y_{n \times p} \quad ? \quad d_{ik}^2 = \sum_{j=1}^p (y_{ij} - y_{kj})^2$$

$$B = YY' \Rightarrow B = \left(b_{ik} = -\frac{1}{2} (d_{ik}^2 - d_{i.}^2 - d_{.k}^2 + d_{..}^2) \right)$$

Logo, temos:

$B_{n \times n} = Y Y'$ Matriz p.s.d. (sob $n > p$) e sua **Decomposição Espectral** é:

$$= U \Lambda U' = U \Lambda^{1/2} \Lambda^{1/2} U' = (U \Lambda^{1/2}) (U \Lambda^{1/2})' \Rightarrow \hat{Y} = (U \Lambda^{1/2})$$

- Quando $n > p$, o posto de D é p . Logo, há $(n-p)$ autovalores nulos.
 - Podemos escolher uma representação para Y em uma dimensão m ($m < p$).
- $$\hat{Y} = U \Lambda^{1/2} = \begin{pmatrix} u_{11} & u_{12} & \dots & u_{1n} \\ u_{21} & u_{22} & \dots & u_{2n} \\ \dots & \dots & \dots & \dots \\ u_{n1} & u_{n1} & \dots & u_{nn} \end{pmatrix} \begin{pmatrix} \sqrt{\lambda_1} & 0 & \dots & 0 \\ 0 & \sqrt{\lambda_2} & \dots & 0 \\ \dots & \dots & \dots & \dots \\ 0 & 0 & \dots & \sqrt{\lambda_n} \end{pmatrix}$$
- $n > p \Rightarrow \lambda_{(p+1)} = \dots = \lambda_n = 0$

Escalonamento Multidimensional

Solução Clássica na dimensão m

$$\begin{array}{l} \text{Matriz de distâncias} \\ D = (d_{ik})_{n \times n} \end{array} \left\{ \begin{array}{l} \Leftrightarrow Y_{n \times p} ? \\ B_{n \times n} = \left(b_{ik} = -\frac{1}{2} (d_{ik}^2 - d_{i.}^2 - d_{.k}^2 + d_{..}^2) \right) \end{array} \right.$$

Obter os “ m ” primeiros componentes da decomposição espectral de B :

Autovalores: $\lambda_1 > \lambda_2 > \dots > \lambda_m > \lambda_{m+1} > \dots > \lambda_n > 0$

Autovetores normalizados: $U = (U_1, U_2, \dots, U_m, U_{m+1}, \dots, U_p, U_{p+1}, \dots, U_n)$

\Rightarrow As coordenadas do vetor Y_i são obtidas a partir da i -ésima linha da matriz $U=(u_{ij})$

$$\hat{Y}_i = (U_1 \dots U_m)_i \Lambda_m^{1/2} = (u_{i1} \sqrt{\lambda_1}, u_{i2} \sqrt{\lambda_2}, \dots, u_{im} \sqrt{\lambda_m})$$

Escalonamento Multidimensional

Matriz de Distância Euclidiana entre os 7 cães ($p=6$)

Matriz de distância Euclidiana entre as observações

	1	2	3	4	5	6	7
1	0.00	6.21	18.70	13.13	4.83	7.43	2.03
2	6.21	0.00	24.34	18.55	9.44	12.94	6.62
3	18.70	24.34	0.00	5.99	18.38	12.50	19.20
4	13.13	18.55	5.99	0.00	13.64	7.26	13.78
5	4.83	9.44	18.38	13.64	0.00	7.98	5.09
6	7.43	12.94	12.50	7.26	7.98	0.00	8.67
7	2.03	6.62	19.20	13.78	5.09	8.67	0.00

Matriz B

	[,1]	[,2]	[,3]	[,4]	[,5]	[,6]	[,7]
[1,]	23.09	48.94	-66.33	-38.97	17.32	-9.68	25.62
[2,]	48.94	113.39	-142.41	-79.68	29.54	-20.68	50.89
[3,]	-66.33	-142.41	194.13	114.78	-54.35	25.24	-71.07
[4,]	-38.97	-79.68	114.78	71.28	-39.90	15.65	-43.17
[5,]	17.32	29.54	-54.35	-39.90	34.88	-8.09	20.62
[6,]	-9.68	-20.68	25.24	15.65	-8.09	12.69	-15.14
[7,]	25.62	50.89	-71.07	-43.17	20.62	-15.14	32.25

Escalonamento Multidimensional

Dados dos Cães (n=7; p=6)

Matriz B

	[,1]	[,2]	[,3]	[,4]	[,5]	[,6]	[,7]
[1,]	23.09	48.94	-66.33	-38.97	17.32	-9.68	25.62
[2,]	48.94	113.39	-142.41	-79.68	29.54	-20.68	50.89
[3,]	-66.33	-142.41	194.13	114.78	-54.35	25.24	-71.07
[4,]	-38.97	-79.68	114.78	71.28	-39.90	15.65	-43.17
[5,]	17.32	29.54	-54.35	-39.90	34.88	-8.09	20.62
[6,]	-9.68	-20.68	25.24	15.65	-8.09	12.69	-15.14
[7,]	25.62	50.89	-71.07	-43.17	20.62	-15.14	32.25

Decomposição Espectral de B

Autovalores

435.64 29.17 12.78 3.95 0.15 0.03 0.00

Autovetores

	U1	U2	U3	U4	U5	U6	U7
[1,]	-0.23	-0.05	0.05	0.25	-0.02	0.86	-0.38
[2,]	-0.49	-0.52	-0.06	-0.52	-0.22	-0.18	-0.38
[3,]	0.67	0.03	0.23	-0.19	-0.57	0.04	-0.38
[4,]	0.40	-0.31	0.07	-0.11	0.77	-0.03	-0.38
[5,]	-0.19	0.80	-0.05	-0.39	0.18	-0.05	-0.38
[6,]	0.10	0.00	-0.79	0.40	-0.11	-0.21	-0.38
[7,]	-0.25	0.05	0.55	0.55	-0.03	-0.43	-0.38

Dimensão: m=2 explica 96,49%

$$\Rightarrow \hat{Y} = (U \Lambda^{1/2})$$

Coordenadas Principais:

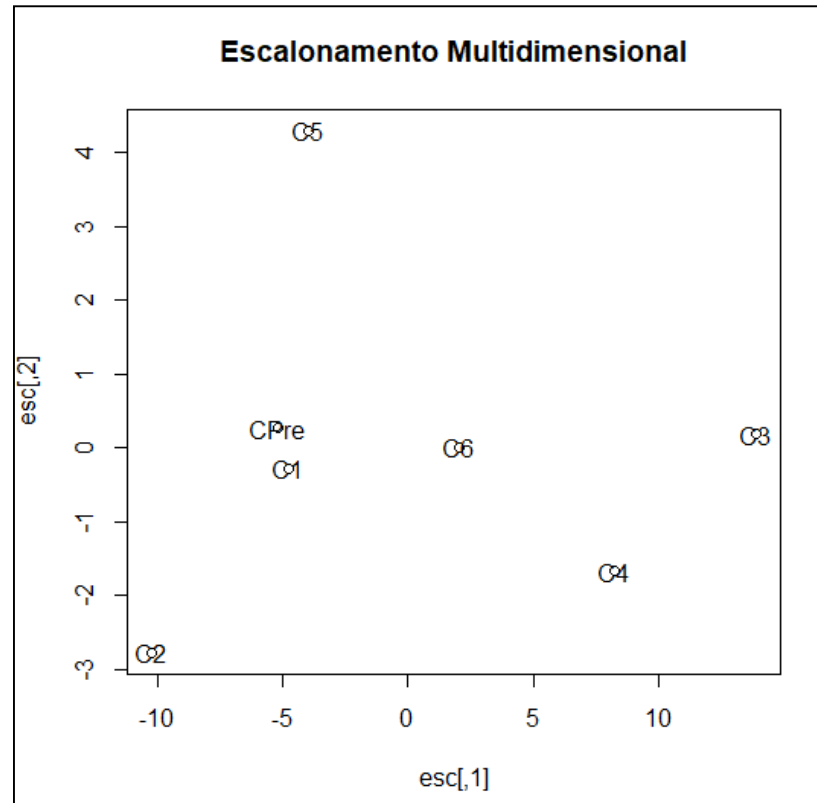
	\hat{Y}_1	\hat{Y}_2	
			$\sqrt{435.64} (-0.23)$
1	-4.76	-0.28	$\sqrt{29.17} (-0.05)$
2	-10.23	-2.78	
3	13.90	0.18	
4	8.26	-1.68	
5	-3.98	4.29	
6	2.01	0.00	
7	-5.21	0.26	

Escalonamento Multidimensional

Representação dos 7 cães obtida da matriz de distância Euclidiana (D)

Coordenadas Principais:

	\hat{Y}_1	\hat{Y}_2
1	-4.76	-0.28
2	-10.23	-2.78
3	13.90	0.18
4	8.26	-1.68
5	-3.98	4.29
6	2.01	0.00
7	-5.21	0.26



Escalonamento Multidimensional

Critério de otimalidade da solução Métrica: $\hat{D} \cong D$

Matrizes de Distância Euclidiana (observada e predita) para os dados dos cães (n=7).

Triangular superior: Matriz de distância Euclidiana observada (D)

Triangular inferior: Matriz de distância Euclidiana predita das Coordenadas Principais (\hat{D})

	1	2	3	4	5	6	7	
1	0	6.21	18.70	13.13	4.83	7.43	2.03	
2	6.01	0	24.34	18.55	9.44	12.94	6.62	→ = D
3	18.67	24.31	0	5.99	18.38	12.50	19.20	
4	13.10	18.52	5.94	0	13.64	7.26	13.78	
5	4.64	9.44	18.35	13.62	0	7.98	5.09	
6	6.78	12.55	11.89	6.48	7.36	0	8.67	
7	0.70	5.87	19.11	13.61	4.21	7.22	0	

↙ = \hat{D}

Usar os desvios para avaliar a qualidade da representação das observações em \mathfrak{R}^2

Análise de Componentes Principais e Coordenadas Principais

Resultado importante!

Análise de CP
Decomposição
espectral de Σ

$$Y_{n \times p}; \left[\Sigma_{p \times p} = V_{p \times p} \Lambda_p V'_{p \times p} \right] \Rightarrow \left[Z_{n \times p} = YV \right]$$

Decomposição
de Y em valores
singulares

$$\left[Y = U_{n \times n} \Lambda_n^{1/2} V'_{p \times p} \right]; \quad n \geq p \quad \Rightarrow \quad \lambda_{p+1} = \dots = \lambda_n = 0$$

$$\left[YV \right] = U \Lambda^{1/2} V' V = \left[U \Lambda^{1/2} \right]$$

Componentes principais

Coordenadas principais

Análise de EM
Decomposição
espectral de B

$$Y_{n \times p}; D_{n \times n} \rightarrow \left[B_{n \times n} = U_{n \times n} \Lambda_n U'_{n \times n} \right] \Rightarrow \left[Y_{n \times p} = U_{n \times p} \Lambda^{1/2} \right]$$

- Os **m primeiros Componentes Principais** são “ótimos” \Rightarrow a soma de suas variâncias é maior que qualquer outro conjunto de **m** combinações lineares não correlacionadas.
- as **m primeiras Coordenadas Principais** são “ótimas” \Rightarrow a projeção de Y no sub-espaço de dimensão **m** de \mathbb{R}^p é mais próxima (em distância Euclidiana) da configuração original do que qualquer outra ($\hat{D} \cong D$)

Componentes Principais – Coordenadas Principais

Solução via Espaços Duais

Resultado importante!

$Y_{n \times p}$: Matriz de dados (“padronizados”) multivariados de posto $r = \min(n, p)$

Análise no espaço das variáveis: $\mathfrak{R}^{p \times p}$

$$Y'Y = \Sigma_{p \times p} = V_{p \times p} \begin{pmatrix} \Lambda_r & 0 \\ 0 & 0 \end{pmatrix} V_{p \times p}' \Rightarrow Y_{n \times p} V_{p \times r}$$

r Componentes Principais

Análise no espaço dos indivíduos: $\mathfrak{R}^{n \times n}$

$$D_{n \times n} \Rightarrow B_{n \times n} = YY' = U_{n \times n} \begin{pmatrix} \Lambda_r & 0 \\ 0 & 0 \end{pmatrix} U_{n \times n}' \Rightarrow U_{n \times n} \Lambda_r^{1/2}$$

Escalonamento Multidimensional:
r Coordenadas Principais
obtidas da Matriz de Distâncias

Análise no espaço $\mathfrak{R}^{n \times p}$

$$Y_{n \times p} = U_{n \times n} \begin{pmatrix} \Lambda_r^{1/2} & 0 \\ 0 & 0 \end{pmatrix} V_{p \times p}' \Rightarrow Y_{n \times p} V_{p \times r} = U_{n \times n} \Lambda_r^{1/2}$$

Equivalência entre os Componentes Principais e as Coordenadas Principais

$n \ll p$: Componentes Principais de Y podem ser obtidos da decomposição espectral da matriz de distâncias D (n x n), de dimensão muito menor que Σ (p x p)

Componentes Principais – Coordenadas Principais

Equivalência das Soluções em Espaços Duais

Resultado importante!

$Y_{n \times p}$: Matriz de dados (“**originais**”) de posto $r = \min(n, p)$

HY: linhas de Y centradas

$$(HY)_{n \times p} = U_{n \times n} \begin{pmatrix} \Lambda_r^{1/2} & 0 \\ 0 & 0 \end{pmatrix} V'_{p \times p}$$

$B = HYY'H$

Análise em $\mathfrak{R}^{n \times n}$

Análise em $\mathfrak{R}^{p \times p}$

$(n-1)S_u = Y'HY$

$$HYY'H = U \Lambda U'$$

$$Y'HY = V \Lambda V'$$

$$U_{n \times n} \Lambda_m^{1/2} = (HY)_{n \times p} V_{p \times m} \quad m \leq r$$

Coordenadas Principais

Componentes Principais (das linhas de HY)

Coordenadas Principais

Dados Cães: n=7; p=6 variáveis

	X1	X2	X3	X4	X5	X6
[1,]	9.7	21.0	19.4	7.7	32.0	36.5
[2,]	8.1	16.7	18.3	7.0	30.3	32.9
[3,]	13.5	27.3	26.8	10.6	41.9	48.1
[4,]	11.5	24.3	24.5	9.3	40.0	44.6
[5,]	10.7	23.5	21.4	8.5	28.8	37.6
[6,]	9.6	22.6	21.1	8.3	34.4	43.1
[7,]	10.3	22.1	19.1	8.1	32.2	35.0

Matriz de Distância Euclidiana (D)

	1	2	3	4	5	6	7
1	0.00	6.21	18.70	13.13	4.83	7.43	2.03
2	6.21	0.00	24.34	18.55	9.44	12.94	6.62
3	18.70	24.34	0.00	5.99	18.38	12.50	19.20
4	13.13	18.55	5.99	0.00	13.64	7.26	13.78
5	4.83	9.44	18.38	13.64	0.00	7.98	5.09
6	7.43	12.94	12.50	7.26	7.98	0.00	8.67
7	2.03	6.62	19.20	13.78	5.09	8.67	0.00

Coordenadas Principais

	\hat{Y}_1	\hat{Y}_2
1	-4.76	-0.28
2	-10.23	-2.78
3	13.90	0.18
4	8.26	-1.68
5	-3.98	4.29
6	2.01	0.00
7	-5.21	0.26



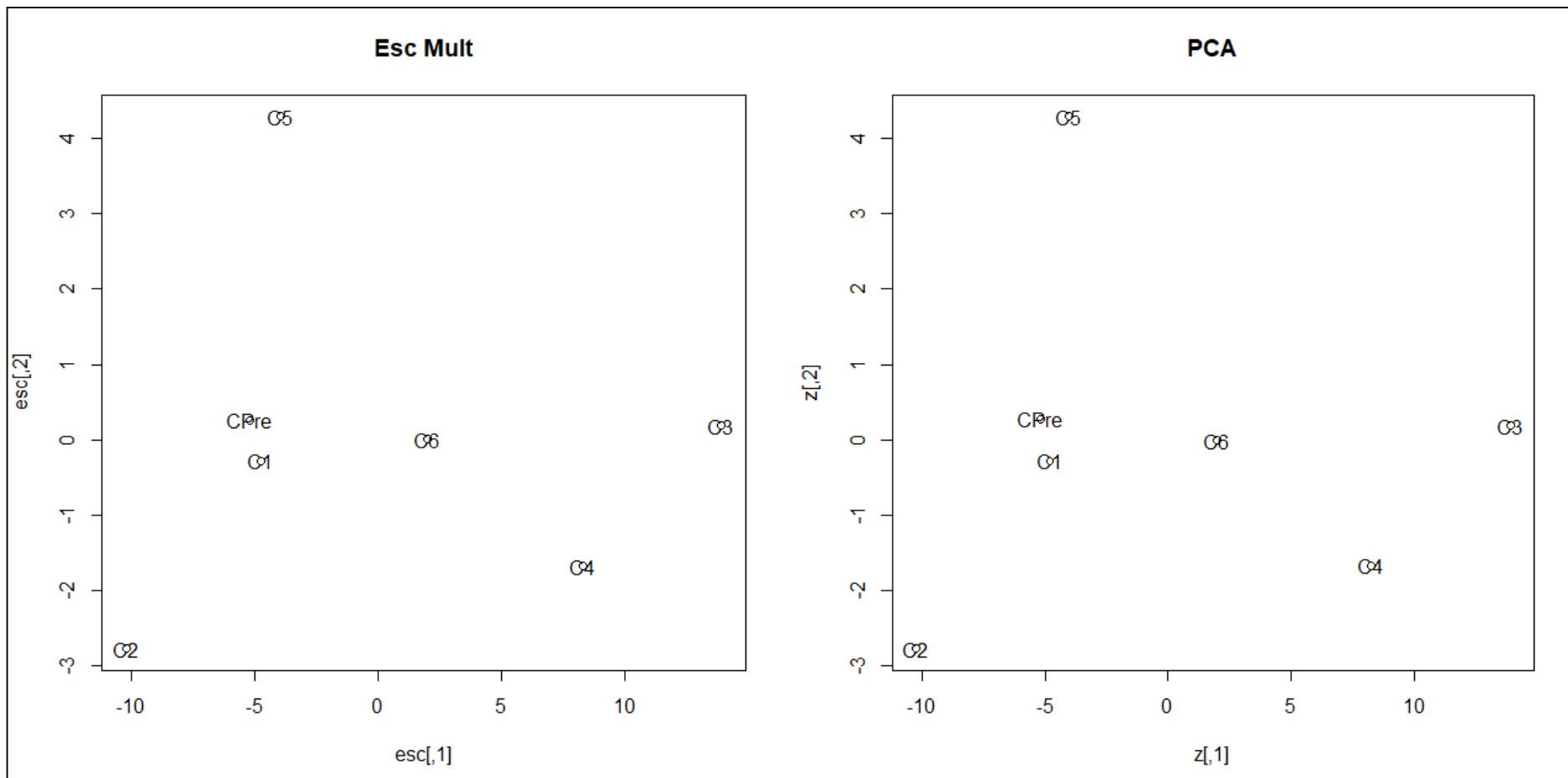
**Escalonamento
Multidimensional**

As Coordenadas Principais representam uma escala (dados) construída a partir da informação das distâncias (D)

Sendo D distância Euclidiana as CoP são os Escores dos Componentes Principais

Componentes Principais e Coordenadas Principais Equivalência

Coordenadas Principais obtidas de $D_{n \times n} \Rightarrow$ Representação (em \mathbb{R}^2) equivalente aos Componentes Principais obtidos de $S_{p \times p}$



Escalonamento Multidimensional

A análise de **Coordenadas Principais** (CoP ou Escalonamento Multidimensional) é baseada em uma matriz de Distâncias ($n \times n$) entre observações enquanto a análise de **Componentes Principais** (CP) é baseada em uma matriz de covariâncias ($p \times p$) entre variáveis.

Equivalências entre CoP e CP:

1. A análise de Coordenadas Principais da matriz de distâncias Euclidianas é equivalente à análise de Componentes Principais da matriz de covariâncias.
2. A análise de Coordenadas Principais da matriz de distâncias de Penrose (ou Pearson) é equivalente à análise de Componentes Principais da matriz de correlação.

Na redução de dimensionalidade, a análise de Coordenadas Principais pode ser aplicada de maneira mais geral, para diferentes escolhas de matriz de distâncias entre observações (Mahalanobis, Manhattan, ou até mesmo distâncias empíricas). Neste caso, quando a matriz D não Euclidiana, NÃO está garantida a equivalência entre as duas analyses, CoP e CP.

Escalonamento Multidimensional

Métodos Não-Métricos

- \mathbf{D} é considerada uma matriz de “dissimilaridade” geral (não precisa ser de distância Euclidiana)
- Os elementos de \mathbf{D} podem ser ordenados

$$d_{ik}^{(1)} \leq d_{ik}^{(2)} \leq \dots \leq d_{ik}^{(q)}; \quad q = n(n-1)/2$$

- Seja \hat{D} , tal que os elementos \hat{d}_{ik} estão monotonicamente relacionados aos elementos d_{ik}

$$d_{ik} < d_{rs} \Rightarrow \hat{d}_{ik} \leq \hat{d}_{rs} ; \quad i < k, \quad r < s$$

- Seja $Y_{n \times m}$ uma configuração em \mathfrak{R}^m com distâncias \hat{d}_{ik} . $Y_{n \times m}$ é ótima no sentido de minimizar seu “stress”, definido como:

$$S^2(Y) = \frac{\sum_{i < k} (d_{ik} - \hat{d}_{ik})^2}{\sum_{i < k} (d_{ik} - \bar{d})^2}$$

Medida de stress de Y: mede quanto da variância de d_{ik} NÃO é explicada pelas m coordenadas principais

Escalonamento Multidimensional

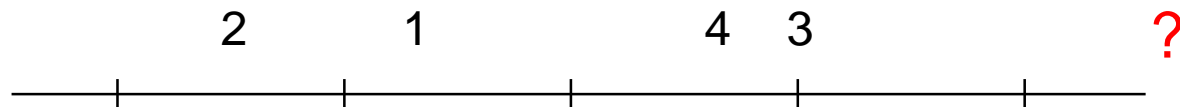
Distância Euclidiana entre os 7 cães (considerando as p=6 variáveis)

Matriz de Distância Euclidiana

	1	2	3	4	5	6	7
1	0						
2	6.21	0					
3	18.70	24.34	0				
4	13.13	18.55	5.99	0			
5	4.83	9.44	18.38	13.64	0		
6	7.43	12.94	12.50	7.26	7.98	0	
7	2.03	6.62	19.20	13.78	5.09	8.67	0

$d_{43} < d_{21} < d_{41} < d_{42} < d_{31} < d_{32}$

Tente localizar os cães 1, 2, 3 e 4 em uma única dimensão (m=1):



Solução: obter uma escala (Y) que minimize o *stress*

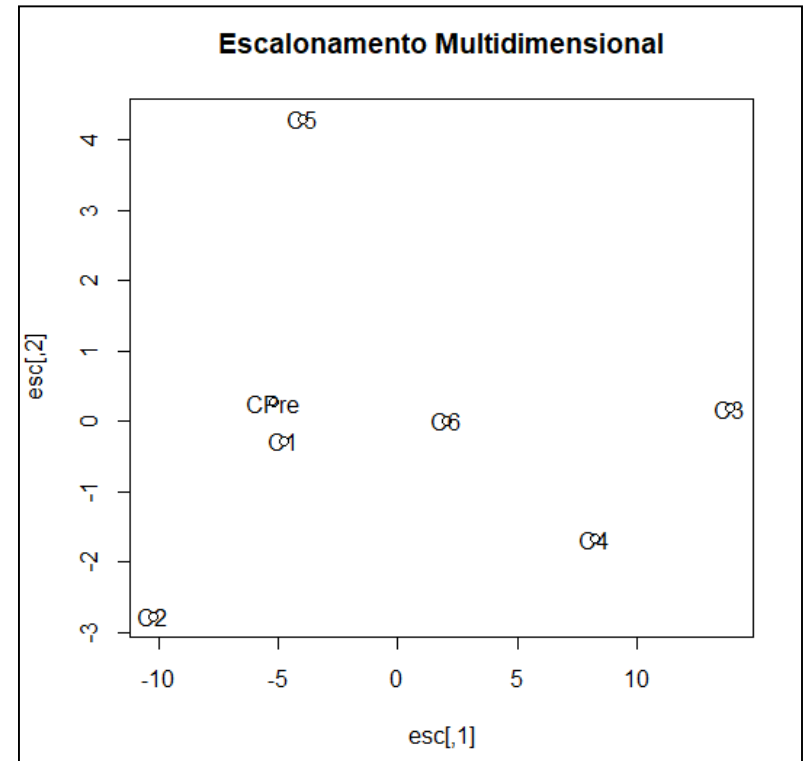
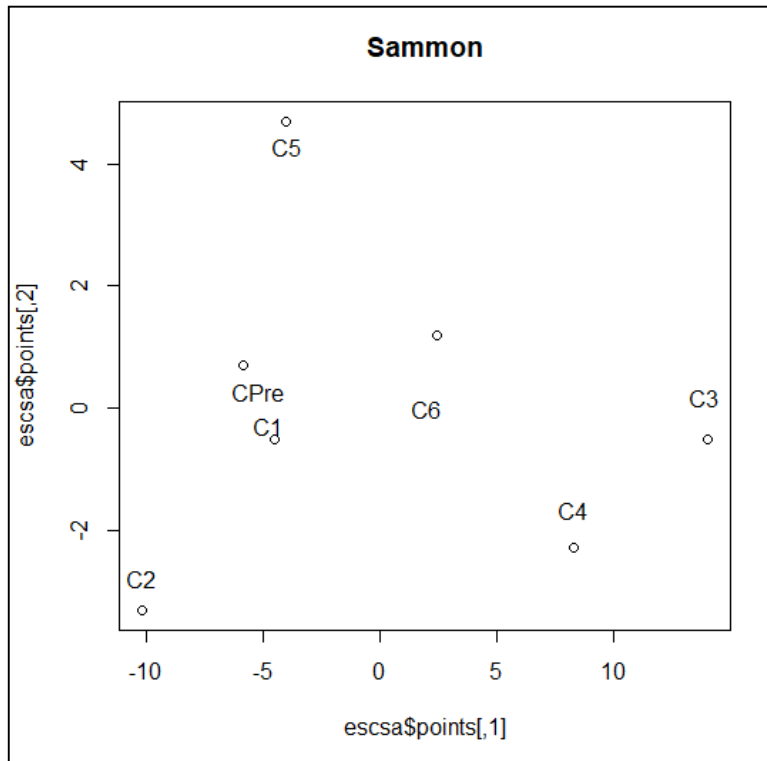
Escalonamento Multidimensional

Solução Não Métrica: Sammon

	[,1]	[,2]
1	-4.53	-0.51
2	-10.20	-3.31
3	14.02	-0.51
4	8.27	-2.28
5	-4.07	4.70
6	2.40	1.20
7	-5.90	0.71

Solução Métrica

	[,1]	[,2]
1	-4.76	-0.28
2	-10.23	-2.78
3	13.90	0.18
4	8.26	-1.68
5	-3.98	4.29
6	2.01	0.00
7	-5.21	0.26



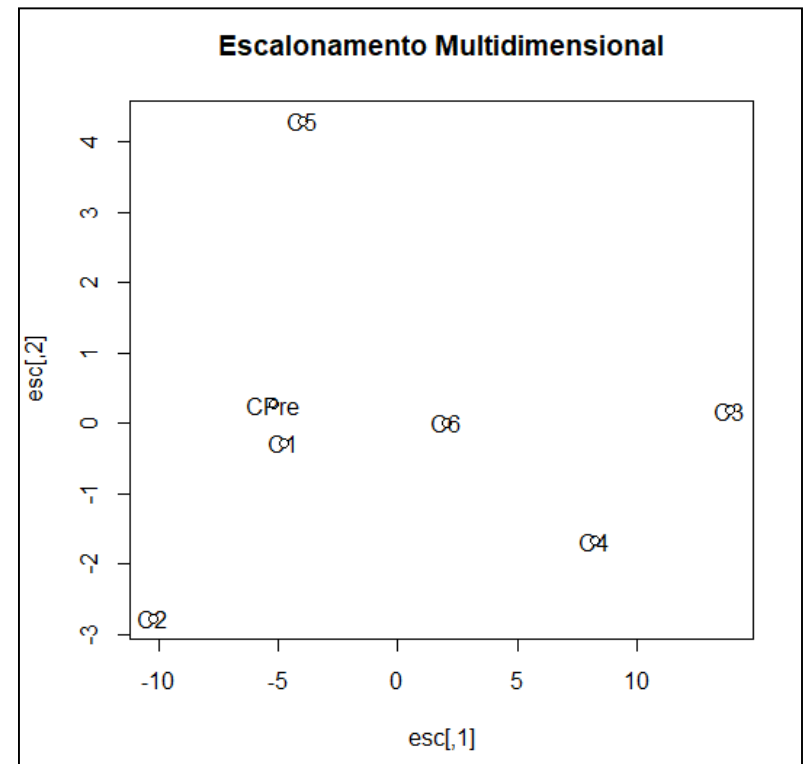
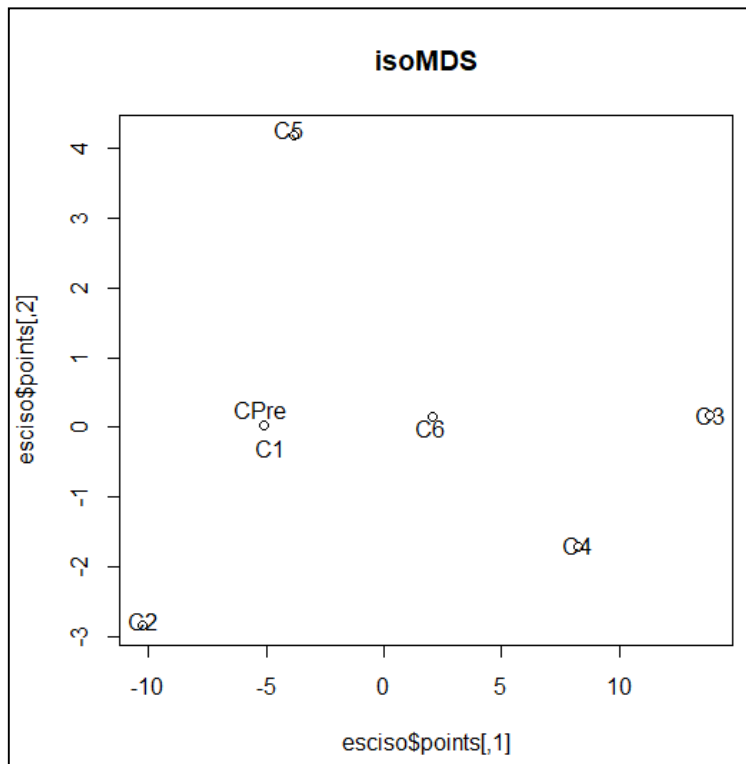
Escalonamento Multidimensional

Solução Não Métrica: IsoMDS

	[,1]	[,2]
1	-5.08	0.03
2	-10.26	-2.84
3	13.88	0.17
4	8.30	-1.72
5	-3.82	4.19
6	2.06	0.15
7	-5.08	0.02

Solução Métrica

	[,1]	[,2]
1	-4.76	-0.28
2	-10.23	-2.78
3	13.90	0.18
4	8.26	-1.68
5	-3.98	4.29
6	2.01	0.00
7	-5.21	0.26



Espaços Duais

Dados dos cães: **Yn_{xp}**

	Y1	Y2	Y3	Y4	Y5	Y6
[1,]	9.7	21.0	19.4	7.7	32.0	36.5
[2,]	8.1	16.7	18.3	7.0	30.3	32.9
[3,]	13.5	27.3	26.8	10.6	41.9	48.1
[4,]	11.5	24.3	24.5	9.3	40.0	44.6
[5,]	10.7	23.5	21.4	8.5	28.8	37.6
[6,]	9.6	22.6	21.1	8.3	34.4	43.1
[7,]	10.3	22.1	19.1	8.1	32.2	35.0

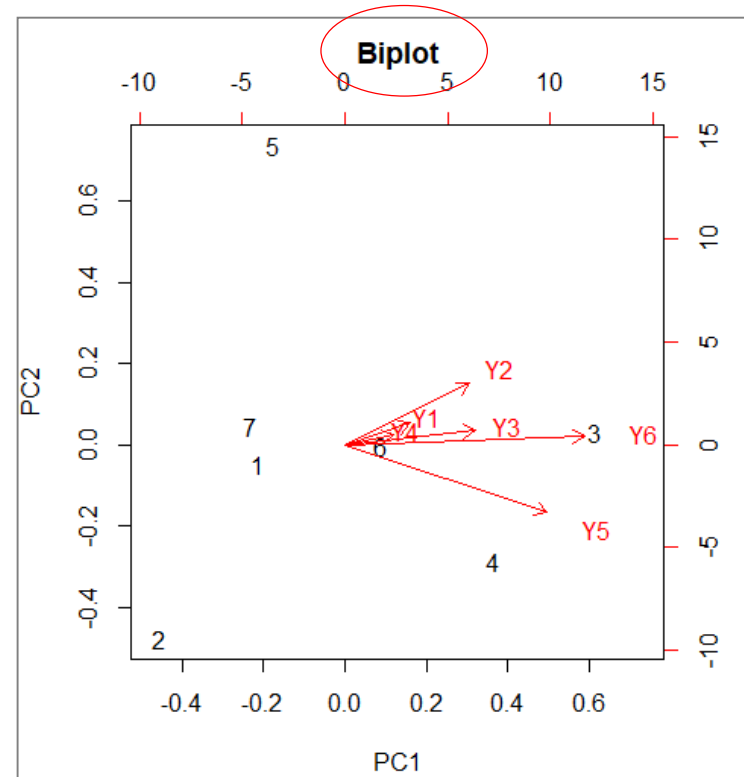
Matriz de Covariância: **S_{xp}**

	Y1	Y2	Y3	Y4	Y5	Y6
Y1	2.88	5.25	4.85	1.93	6.53	7.74
Y2	5.25	10.56	8.90	3.59	11.46	15.58
Y3	4.85	8.90	9.61	3.51	13.43	16.31
Y4	1.93	3.59	3.51	1.36	4.86	5.92
Y5	6.53	11.46	13.43	4.86	24.36	24.68
Y6	7.74	15.58	16.31	5.92	24.68	31.52

Matriz de Distância Euclidiana: **D_{n_{xn}}**

	1	2	3	4	5	6	7
1	0.00	6.21	18.70	13.13	4.83	7.43	2.03
2	6.21	0.00	24.34	18.55	9.44	12.94	6.62
3	18.70	24.34	0.00	5.99	18.38	12.50	19.20
4	13.13	18.55	5.99	0.00	13.64	7.26	13.78
5	4.83	9.44	18.38	13.64	0.00	7.98	5.09
6	7.43	12.94	12.50	7.26	7.98	0.00	8.67
7	2.03	6.62	19.20	13.78	5.09	8.67	0.00

Biplot: representação gráfica simultânea de n observações e p variáveis em \mathbb{R}^2



Biplots

Biplot: representação gráfica simultânea de n observações e p variáveis em \mathcal{R}^2

$$Y_{n \times p} = U_{n \times n} \Lambda^{1/2} V'_{p \times p}$$

Matriz de "escores dos CP" Matriz de "cargas"

$$\left\{ \begin{array}{l} YY'_{n \times n} = U \Lambda U' \quad \text{Análise em } \mathcal{R}^{n \times n} \\ Y'Y_{p \times p} = V \Lambda V' \quad \text{Análise em } \mathcal{R}^{p \times p} \end{array} \right.$$

$\mathcal{R}^p \rightarrow \mathcal{R}^m$
 $m=2$

$$Y_{n \times p}; \quad Y \approx [U_1 \ U_2]_{n \times 2} \Lambda_2^{1/2} [V_1 \ V_2]'_{2 \times n} = [U_1 \ U_2] \Lambda_2^{1/2 - c/2 + c/2} [V_1 \ V_2]'$$

$$Y \approx \left(U_1 \lambda_1^{1/2 - c/2} \quad U_2 \lambda_2^{1/2 - c/2} \right) \left(\lambda_1^{c/2} V_1 \quad \lambda_2^{c/2} V_2 \right)'$$

Análises das Linhas e das colunas sob os "mesmos" autovalores

$$U_1 \lambda_1^{1/2 - c/2} \quad \times \quad U_2 \lambda_2^{1/2 - c/2} \quad \text{n escores}$$

$$\lambda_1^{c/2} V_1 \quad \times \quad \lambda_2^{c/2} V_2 \quad \text{p cargas}$$

- $c=0$: linhas em coordenadas principais e colunas em coordenadas padronizadas
- $c=1$: linhas em coordenadas padronizadas e colunas em coordenadas principais
- $c=1/2$: representação interativa