

Análise de dados Multivariados

Introdução

Fabio Cop Ferreira
fabiocferreira@gmail.com

Exemplos de dados Multivariados: pardais de Hermon Bumpus



Em **fevereiro de 1898**, após uma forte e incomum tempestade com nevasca vários pardais moribundos foram levados ao laboratório de Hemon Bumpus (Brown University, Providence, Rhode Island, EUA). Cerca de metade destes pardais morreu logo, em consequência do stress. Bumpus viu neste fato uma possibilidade de encontrar suporte para a seleção natural. Foram tomadas várias medidas morfológicas de cada pássaro, anotado o sexo e se o pássaro sobreviveu ou não à tempestade.

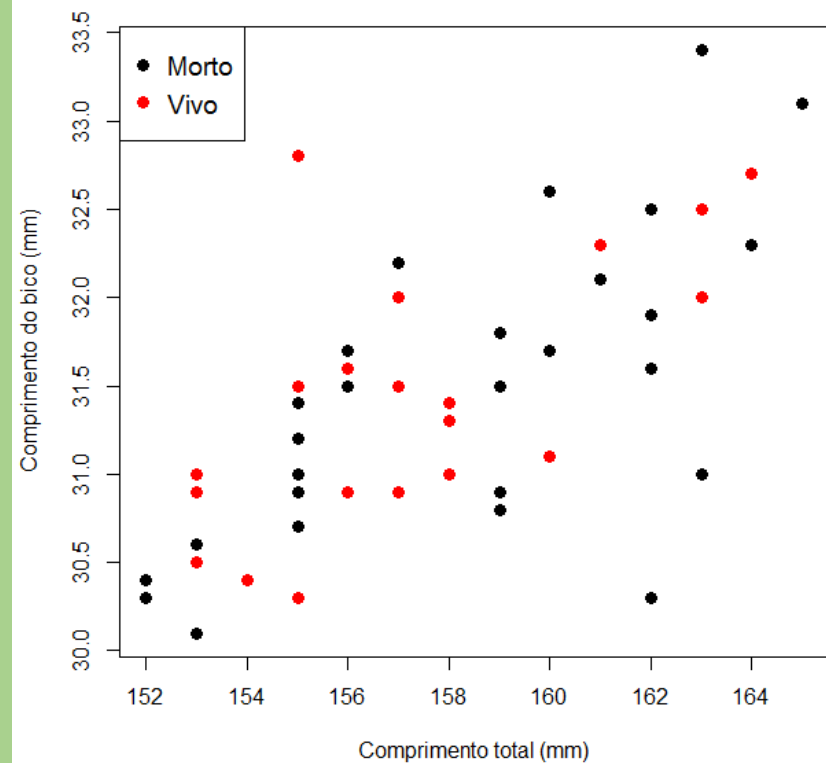
Condição	CT (mm)	Extensão Alar	Bico	Umero	Esterno
Vivo	157	235	31,5	18,1	19,8
Vivo	153	236	30,9	17,7	20,2
Vivo	155	236	30,3	18,5	20,1
Vivo	159	236	31,5	18	21,5
Vivo	156	237	30,9	18	20,3
:	:	:	:	:	:
Morto	165	245	33,1	19,8	22,7
Morto	159	245	31,8	18,5	21,7
Morto	161	245	32,1	19,1	20,8
Morto	162	245	32,5	18,5	21,1
Morto	159	247	30,9	18,1	19
Morto	162	247	31,9	19,1	20,4
Morto	164	248	32,3	18,8	20,9
Morto	163	249	33,4	19,5	22,8
Morto	160	250	31,7	18,8	22,5
Morto	162	252	31,9	19,1	22,2

Exemplos de dados Multivariados: pardais de Hermon Bumpus



Questões:

1. Qual a relação entre as variáveis? Ex. Valores grandes de uma estão associados a valores grandes de outra?



Exemplos de dados Multivariados: pardais de Hermon Bumpus



Questões:

1. Qual a relação entre as variáveis? Ex. Valores grandes de uma estão associados a valores grandes de outra?

2. Os sobreviventes e os não sobreviventes diferem quanto aos valores médios e de variabilidades nestas medidas?

Condição	CT (mm)	Extensão Alar	Bico	Umero	Esterno
Vivo	157	235	31,5	18,1	19,8
Vivo	153	236	30,9	17,7	20,2
Vivo	155	236	30,3	18,5	20,1
Vivo	159	236	31,5	18	21,5
Vivo	156	237	30,9	18	20,3
:	:	:	:	:	:
Morto	165	245	33,1	19,8	22,7
Morto	159	245	31,8	18,5	21,7
Morto	161	245	32,1	19,1	20,8
Morto	162	245	32,5	18,5	21,1
Morto	159	247	30,9	18,1	19
Morto	162	247	31,9	19,1	20,4
Morto	164	248	32,3	18,8	20,9
Morto	163	249	33,4	19,5	22,8
Morto	160	250	31,7	18,8	22,5
Morto	162	252	31,9	19,1	22,2
$\bar{X}_{Vivos}$	157,38	241,00	31,43	18,50	20,81
$\bar{X}_{Mortos}$	158,43	241,57	31,48	18,45	20,84
s^2_{Vivos}	11,05	17,50	0,53	0,18	0,57
s^2_{Mortos}	15,07	32,55	0,73	0,43	1,32

Exemplos de dados Multivariados: pardais de Hermon Bumpus



É possível criar uma função matemática que diferencie os sobreviventes dos não sobreviventes em função de sua morfometria?

Condição	CT (mm)	Extensão Alar	Bico	Umero	Esterno
Vivo	157	235	31,5	18,1	19,8
Vivo	153	236	30,9	17,7	20,2
Vivo	155	236	30,3	18,5	20,1
Vivo	159	236	31,5	18	21,5
Vivo	156	237	30,9	18	20,3
:	:	:	:	:	:
Morto	165	245	33,1	19,8	22,7
Morto	159	245	31,8	18,5	21,7
Morto	161	245	32,1	19,1	20,8
Morto	162	245	32,5	18,5	21,1
Morto	159	247	30,9	18,1	19
Morto	162	247	31,9	19,1	20,4
Morto	164	248	32,3	18,8	20,9
Morto	163	249	33,4	19,5	22,8
Morto	160	250	31,7	18,8	22,5
Morto	162	252	31,9	19,1	22,2

Ecomorfologia em peixes: descrevendo o padrão morfológico de 8 espécies

Espécies	Familia	CI	RD	IVF	RAC	REP	MO
<i>Bryconamericus stramineus</i>	Characidae	16,7	2,1	5,7	1,4	7,0	12,6
<i>Bryconamericus</i> sp.	Characidae	17,1	2,6	5,3	1,3	6,8	17,4
<i>Characidium fasciatum</i>	Crenuchidae	14,1	2,1	5,1	1,4	7,9	30,0
<i>Characidium zebra</i>	Crenuchidae	14,7	2,2	4,9	1,1	8,0	19,8
<i>Cetopsorhamdia iheringi</i>	Heptapteridae	9,7	1,7	5,9	2,2	9,1	20,6
<i>Imparfinis minutus</i>	Heptapteridae	6,9	1,2	5,5	2,6	7,6	22,4
<i>Hisonotus</i> sp.	Loridariidae	7,3	1,6	3,9	1,9	7,1	31,4
<i>Hypostomus garmani</i>	Loridariidae	6,6	1,9	3,5	3,0	8,5	31,4
Médias		11,64	1,92	4,98	1,86	7,75	23,20
Variâncias		20,19	0,18	0,73	0,47	0,63	49,38

CI: Índice de compressão lateral;

RD: Altura relativa;

IVF: Índice de achatamento ventral;

RAC: Área relativa da nadadeira caudal;

REP: Posição relativa do olho;

MO: orientação da boca.

Qual a semelhança morfológica entre as espécies com base nestas 6 medidas?



Characidae



Crenuchidae



Heptapteridae

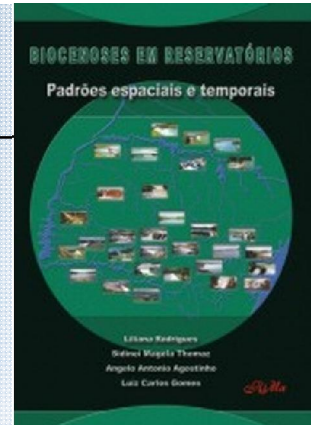


Loridariidae

Ecologia de comunidades: peixes em reservatórios

Rodrigues, Thomaz, Agostinho & Gomes, 2005. Biocenoses em reservatórios: padrões espaciais e temporais.

30 reservatórios – PR: *tamanho/idade/sub-bacia dos reservatórios, comunidades de peixes e invertebrados, variáveis físicas e químicas da água.*



1. Qual a relação entre as variáveis físicas e químicas?
2. Como as espécies de peixes estão associadas? Quais tendem a ocorrer juntas ou separadas?
3. Quais reservatórios têm faunas similares? Existem grupos bem definidos, ex. dentro de sub-bacias?
4. Quais reservatórios são mais similares quanto às variáveis físicas e químicas?
5. Como as variáveis ambientais estão correlacionadas à composição de espécies?

Representação matricial de dados multivariados: descritores e objetos

		Descritores						
Objetos		V_1	V_2	V_3	$\cdots$	V_j	$\cdots$	V_n
	Obj_1	y_{11}	y_{12}	y_{13}	$\cdots$	y_{1j}	$\cdots$	y_{1n}
	Obj_2	y_{21}	y_{22}	y_{23}	$\cdots$	y_{2j}	$\cdots$	y_{2n}
	Obj_3	y_{31}	y_{32}	y_{33}	$\cdots$	y_{3j}	$\cdots$	y_{3n}
	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\cdots$	$\vdots$	$\cdots$	$\vdots$
	Obj_i	y_{i1}	y_{i2}	y_{i3}	$\cdots$	y_{ij}	$\cdots$	y_{in}
	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\cdots$	$\vdots$	$\cdots$	$\vdots$
	Obj_m	y_{m1}	y_{m2}	y_{m3}	$\cdots$	y_{mj}	$\cdots$	y_{mn}

- Qual a relação entre os objetivos quando caracterizados pelas variáveis descritoras?
- Qual a relação entre as variáveis descritoras, de acordo com os valores que estas assumem nos objetos?

Representação matricial de dados multivariados: descritores e objetos

Objetos	Descritores							
		pH	O_2	$T^{\circ}C$	Prof	Sal	$Cond.$	$Turbidez$
	$Riacho_1$	6.5	8.0	22.0	0.75	0.110	0.025	20
	$Riacho_2$	4.3	7.5	21.1	0.54	0.120	0.025	33
	$Riacho_3$	6.8	7.1	19.6	0.20	0.320	0.027	72
	$Riacho_4$	3.0	2.0	23.2	0.15	0.090	0.020	30
	$Riacho_5$	5.2	6.0	20.4	1.10	0.091	0.030	10
	$Riacho_6$	6.9	8.2	20.1	1.09	0.095	0.022	49
	$Riacho_7$	4.2	5.1	20.9	0.70	0.130	0.021	60

- Qual a relação entre os objetivos quando caracterizados pelas variáveis descritoras?
- Qual a relação entre as variáveis descritoras, de acordo com os valores que estas assumem nos objetos?

Representação matricial de dados multivariados: descritores e objetos

		Descritores						
Objetos		Sp_1	Sp_2	Sp_3	Sp_4	Sp_5	Sp_6	Sp_7
	$Riacho_1$	4	33	6	12	11	3	0
	$Riacho_2$	10	2	78	90	0	0	0
	$Riacho_3$	70	0	0	1	85	0	2
	$Riacho_4$	0	17	0	2	76	0	0
	$Riacho_5$	2	300	0	1	93	0	1
	$Riacho_6$	6	142	0	1	187	0	9
	$Riacho_7$	0	12	0	0	91	1	6

- Qual a relação entre os objetivos quando caracterizados pelas variáveis descritoras?
- Qual a relação entre as variáveis descritoras, de acordo com os valores que estas assumem nos objetos?

Representação matricial de dados multivariados: descritores e objetos

	Sp_1	Sp_2	Sp_3	Sp_4	Sp_5	Sp_6	Sp_7
<i>Riacho</i> ₁	4	33	6	12	11	3	0
<i>Riacho</i> ₂	10	2	78	90	0	0	0
<i>Riacho</i> ₃	70	0	0	1	85	0	2
<i>Riacho</i> ₄	0	17	0	2	76	0	0
<i>Riacho</i> ₅	2	300	0	1	93	0	1
<i>Riacho</i> ₆	6	142	0	1	187	0	9
<i>Riacho</i> ₇	0	12	0	0	91	1	6



	pH	O_2	$T^{\circ}C$	Prof	Sal	Cond .	Turbidez
<i>Riacho</i> ₁	6.5	8.0	22.0	0.75	0.110	0.025	20
<i>Riacho</i> ₂	4.3	7.5	21.1	0.54	0.120	0.025	33
<i>Riacho</i> ₃	6.8	7.1	19.6	0.20	0.320	0.027	72
<i>Riacho</i> ₄	3.0	2.0	23.2	0.15	0.090	0.020	30
<i>Riacho</i> ₅	5.2	6.0	20.4	1.10	0.091	0.030	10
<i>Riacho</i> ₆	6.9	8.2	20.1	1.09	0.095	0.022	49
<i>Riacho</i> ₇	4.2	5.1	20.9	0.70	0.130	0.021	60

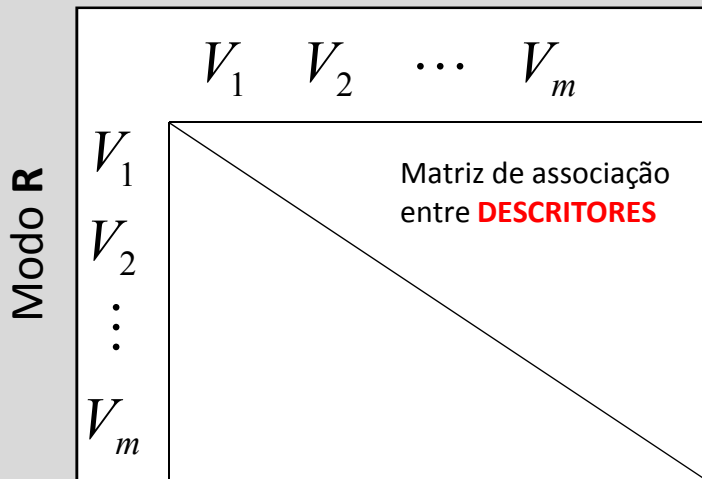
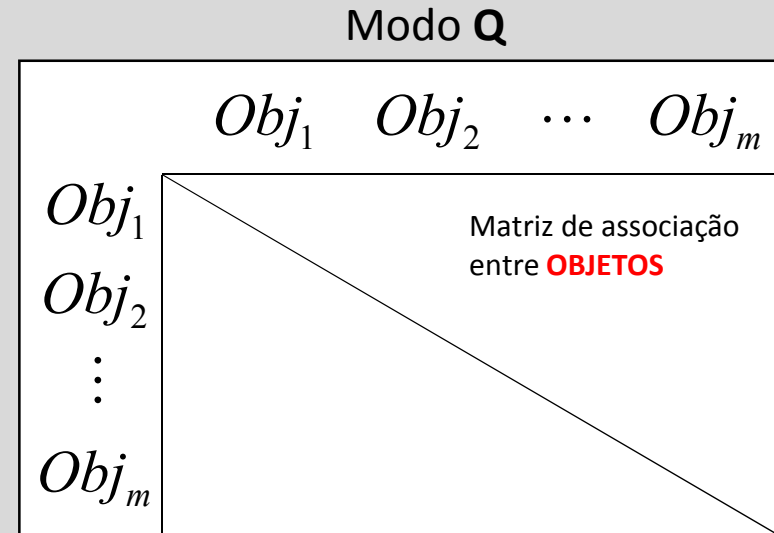


Associação de matrizes

Como a distribuição dos organismos dependem
das características ambientais?

Análise em modo R e Q

		Descritores						
Objetos		V_1	V_2	V_3	$\dots$	V_j	$\dots$	V_n
	Obj_1	y_{11}	y_{12}	y_{13}	$\cdots$	y_{1j}	$\cdots$	y_{1n}
	Obj_2	y_{21}	y_{22}	y_{23}	$\cdots$	y_{2j}	$\cdots$	y_{2n}
	Obj_3	y_{31}	y_{32}	y_{33}	$\cdots$	y_{3j}	$\cdots$	y_{3n}
	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\dots$	$\vdots$	$\dots$	$\vdots$
	Obj_i	y_{i1}	y_{i2}	y_{i3}	$\cdots$	y_{ij}	$\cdots$	y_{in}
	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\dots$	$\vdots$	$\dots$	$\vdots$
	Obj_m	y_{m1}	y_{m2}	y_{m3}	$\cdots$	y_{mj}	$\cdots$	y_{mn}

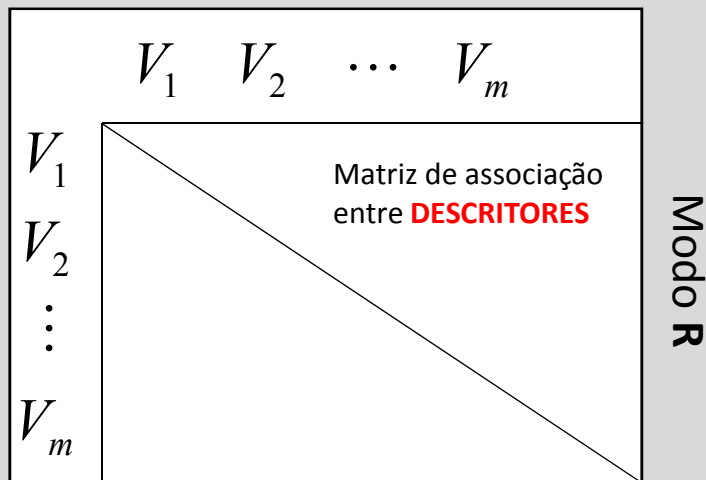


Medem a relação entre cada par de
Objetos (modo Q) ou Descritores
(Modo R).

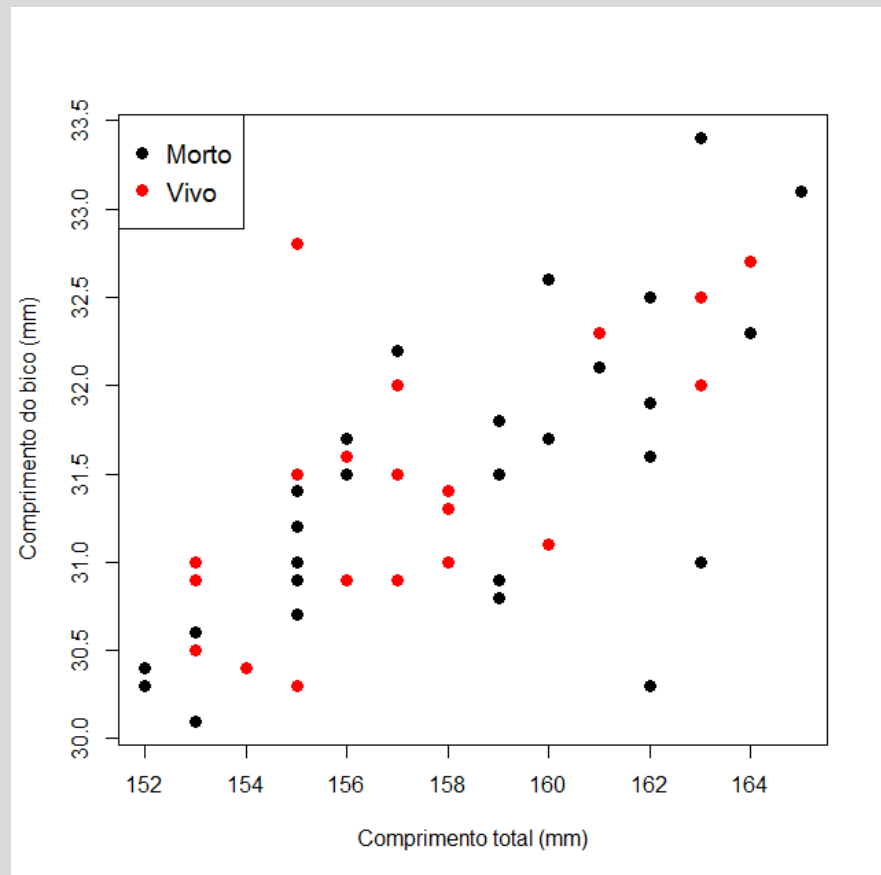
As métricas utilizadas são distintas
para da tipo de análise.

Análise em modo R: medidas de associação

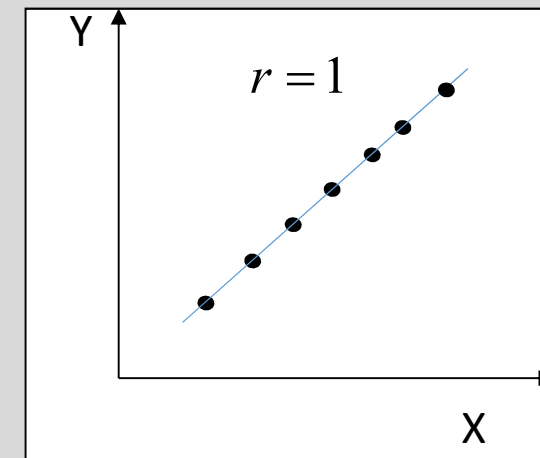
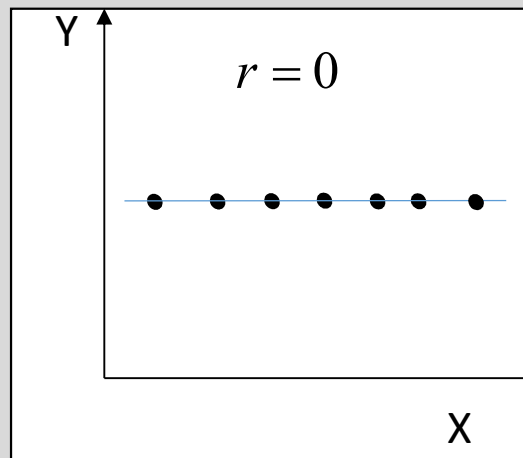
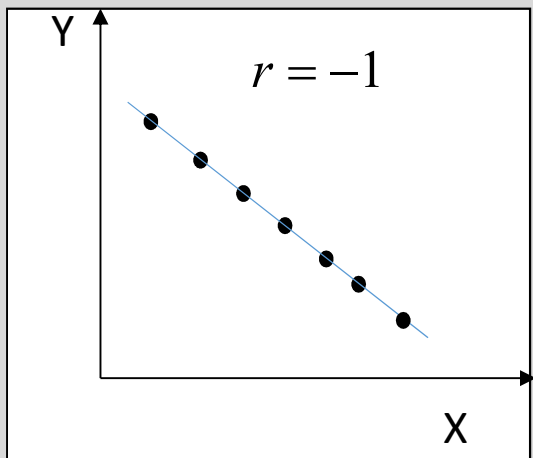
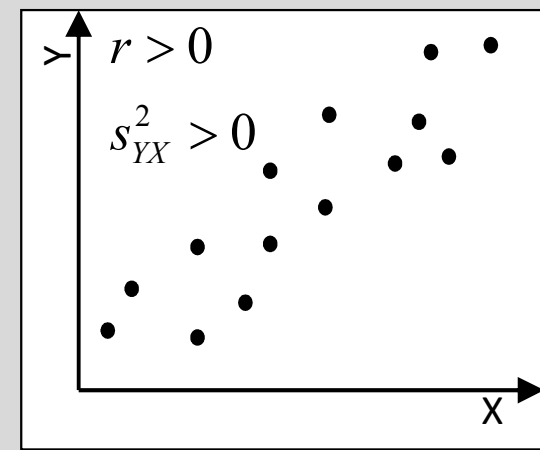
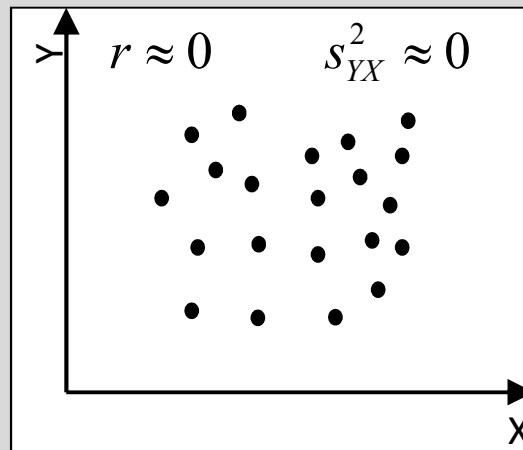
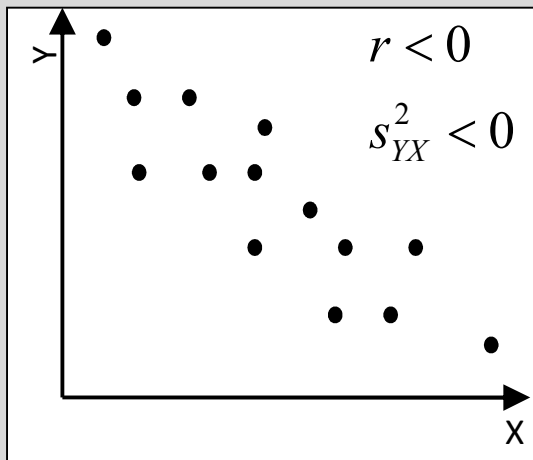
		Descritores						
Objetos		V_1	V_2	V_3	$\cdots$	V_j	$\cdots$	V_n
	Obj_1	y_{11}	y_{12}	y_{13}	$\cdots$	y_{1j}	$\cdots$	y_{1n}
	Obj_2	y_{21}	y_{22}	y_{23}	$\cdots$	y_{2j}	$\cdots$	y_{2n}
	Obj_3	y_{31}	y_{32}	y_{33}	$\cdots$	y_{3j}	$\cdots$	y_{3n}
	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\cdots$	$\vdots$	$\cdots$	$\vdots$
	Obj_i	y_{i1}	y_{i2}	y_{i3}	$\cdots$	y_{ij}	$\cdots$	y_{in}
	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\cdots$	$\vdots$	$\cdots$	$\vdots$
	Obj_m	y_{m1}	y_{m2}	y_{m3}	$\cdots$	y_{mj}	$\cdots$	y_{mn}



- Coeficiente de correlação de Pearson (r).
- Covariância entre X e Y (s_{XY}^2).

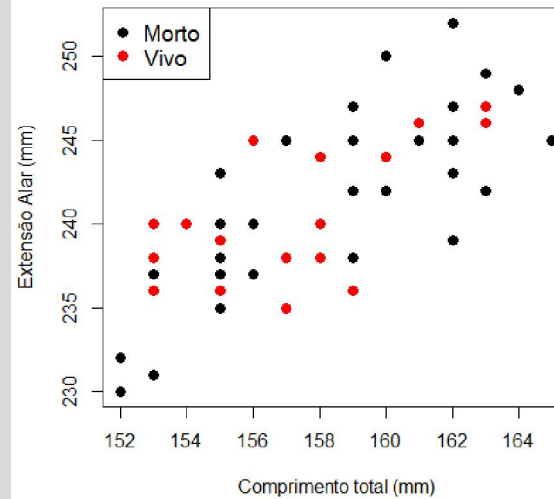


Correlação (r) e covariância (s_{XY}^2)



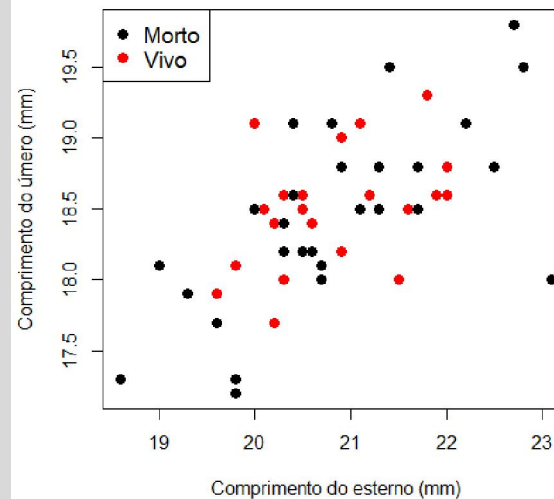
Correlação (r) e covariância (s_{XY}^2)

Condição	CT (mm)	Extensão Alar	Bico	Umero	Esterno
Vivo	157	235	31,5	18,1	19,8
Vivo	153	236	30,9	17,7	20,2
Vivo	155	236	30,3	18,5	20,1
Vivo	159	236	31,5	18	21,5
Vivo	156	237	30,9	18	20,3
:	:	:	:	:	:
Morto	165	245	33,1	19,8	22,7
Morto	159	245	31,8	18,5	21,7
Morto	161	245	32,1	19,1	20,8
Morto	162	245	32,5	18,5	21,1
Morto	159	247	30,9	18,1	19
Morto	162	247	31,9	19,1	20,4
Morto	164	248	32,3	18,8	20,9
Morto	163	249	33,4	19,5	22,8
Morto	160	250	31,7	18,8	22,5
Morto	162	252	31,9	19,1	22,2



$$r_{BicoVSEalar} = 0,73$$

$$s_{BicoVSEalar}^2 = 13,61$$



$$r_{UmeroVSEsterno} = 0,61$$

$$s_{UmeroVSEsterno}^2 = 0,34$$

Matrizes de correlação (r) e covariância (s_{XY}^2)

Condição	CT (mm)	Extensão Alar	Bico	Umero	Esterno
Vivo	157	235	31,5	18,1	19,8
Vivo	153	236	30,9	17,7	20,2
Vivo	155	236	30,3	18,5	20,1
Vivo	159	236	31,5	18	21,5
Vivo	156	237	30,9	18	20,3
:	:	:	:	:	:
Morto	165	245	33,1	19,8	22,7
Morto	159	245	31,8	18,5	21,7
Morto	161	245	32,1	19,1	20,8
Morto	162	245	32,5	18,5	21,1
Morto	159	247	30,9	18,1	19
Morto	162	247	31,9	19,1	20,4
Morto	164	248	32,3	18,8	20,9
Morto	163	249	33,4	19,5	22,8
Morto	160	250	31,7	18,8	22,5
Morto	162	252	31,9	19,1	22,2

Correlação (r)

	Ctotal	Ealar	Bico	Umero	Esterno
Ctotal	1,00	0,73	0,66	0,65	0,61
Ealar	0,73	1,00	0,67	0,77	0,53
Bico	0,66	0,67	1,00	0,76	0,53
Umero	0,65	0,77	0,76	1,00	0,61
Esterno	0,61	0,53	0,53	0,61	1,00

Covariância (s_{XY}^2)

	Ctotal	Ealar	Bico	Umero	Esterno
Ctotal	13,35	13,61	1,92	1,33	2,19
Ealar	13,61	25,68	2,71	2,20	2,66
Bico	1,92	2,71	0,63	0,34	0,41
Umero	1,33	2,20	0,34	0,32	0,34
Esterno	2,19	2,66	0,41	0,34	0,98

Matrizes de correlação (r) e covariância (s_{XY}^2)

Condição	CT (mm)	Extensão Alar	Bico	Umero	Esterno
Vivo	157	235	31,5	18,1	19,8
Vivo	153	236	30,9	17,7	20,2
Vivo	155	236	30,3	18,5	20,1
Vivo	159	236	31,5	18	21,5
Vivo	156	237	30,9	18	20,3
:	:	:	:	:	:
Morto	165	245	33,1	19,8	22,7
Morto	159	245	31,8	18,5	21,7
Morto	161	245	32,1	19,1	20,8
Morto	162	245	32,5	18,5	21,1
Morto	159	247	30,9	18,1	19
Morto	162	247	31,9	19,1	20,4
Morto	164	248	32,3	18,8	20,9
Morto	163	249	33,4	19,5	22,8
Morto	160	250	31,7	18,8	22,5
Morto	162	252	31,9	19,1	22,2

Correlação (r)

	Ctotal	Ealar	Bico	Umero	Esterno
Ctotal	1,00	0,73	0,66	0,65	0,61
Ealar	0,73	1,00	0,67	0,77	0,53
Bico	0,66	0,67	1,00	0,76	0,53
Umero	0,65	0,77	0,76	1,00	0,61
Esterno	0,61	0,53	0,53	0,61	1,00

Covariância (s_{XY}^2)

	Ctotal	Ealar	Bico	Umero	Esterno
Ctotal	13,35	13,61	1,92	1,33	2,19
Ealar	13,61	25,68	2,71	2,20	2,66
Bico	1,92	2,71	0,63	0,34	0,41
Umero	1,33	2,20	0,34	0,32	0,34
Esterno	2,19	2,66	0,41	0,34	0,98

Matrizes de correlação (r) e covariância (s_{XY}^2)

Condição	CT (mm)	Extensão Alar	Bico	Umero	Esterno
Vivo	157	235	31,5	18,1	19,8
Vivo	153	236	30,9	17,7	20,2
Vivo	155	236	30,3	18,5	20,1
Vivo	159	236	31,5	18	21,5
Vivo	156	237	30,9	18	20,3
:	:	:	:	:	:
Morto	165	245	33,1	19,8	22,7
Morto	159	245	31,8	18,5	21,7
Morto	161	245	32,1	19,1	20,8
Morto	162	245	32,5	18,5	21,1
Morto	159	247	30,9	18,1	19
Morto	162	247	31,9	19,1	20,4
Morto	164	248	32,3	18,8	20,9
Morto	163	249	33,4	19,5	22,8
Morto	160	250	31,7	18,8	22,5
Morto	162	252	31,9	19,1	22,2

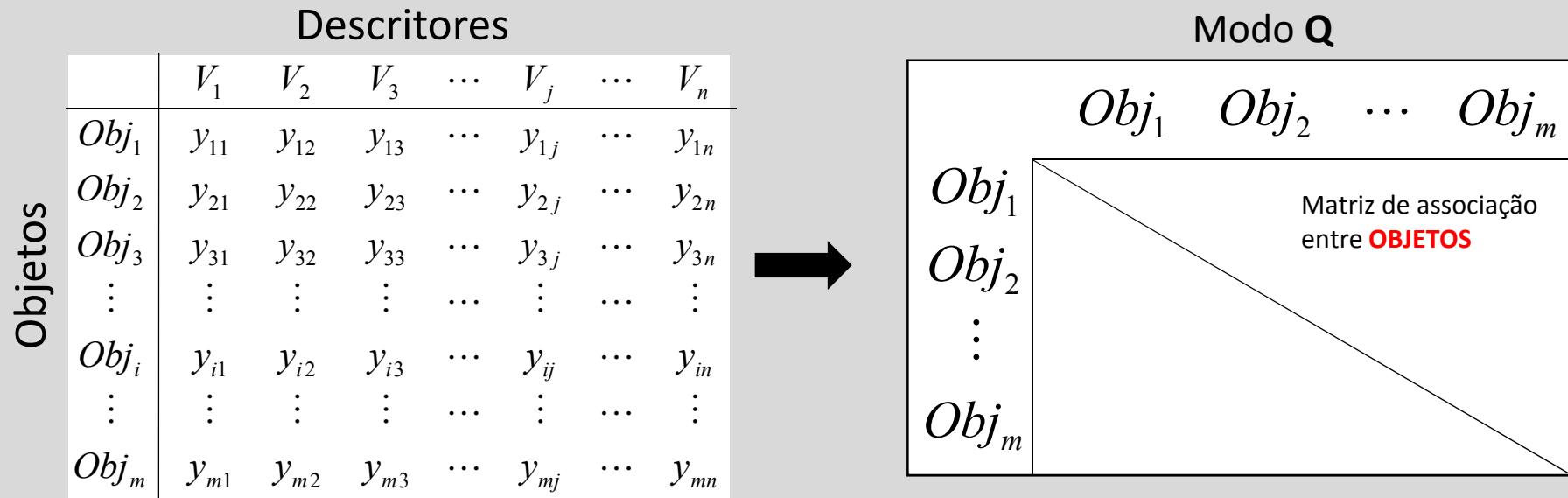
Correlação (r)

	Ctotal	Ealar	Bico	Umero	Esterno
Ctotal	1,00	0,73	0,66	0,65	0,61
Ealar	0,73	1,00	0,67	0,77	0,53
Bico	0,66	0,67	1,00	0,76	0,53
Umero	0,65	0,77	0,76	1,00	0,61
Esterno	0,61	0,53	0,53	0,61	1,00

Covariância (s_{XY}^2)

	Ctotal	Ealar	Bico	Umero	Esterno
Ctotal	13,35	13,61	1,92	1,33	2,19
Ealar	13,61	25,68	2,71	2,20	2,66
Bico	1,92	2,71	0,63	0,34	0,41
Umero	1,33	2,20	0,34	0,32	0,34
Esterno	2,19	2,66	0,41	0,34	0,98

Análise em modo Q: medidas de associação



- Índices de similaridade e Índices de distância. Medem quão similar ou diferentes são dois objetos.
- Similaridade: 0 (diferentes) $\sim$ 1 (iguais).
- Distância: 0 (iguais) $\sim$ > 0 (diferentes).
- É possível transformar índices de similaridade (S) em índices de distância (D).

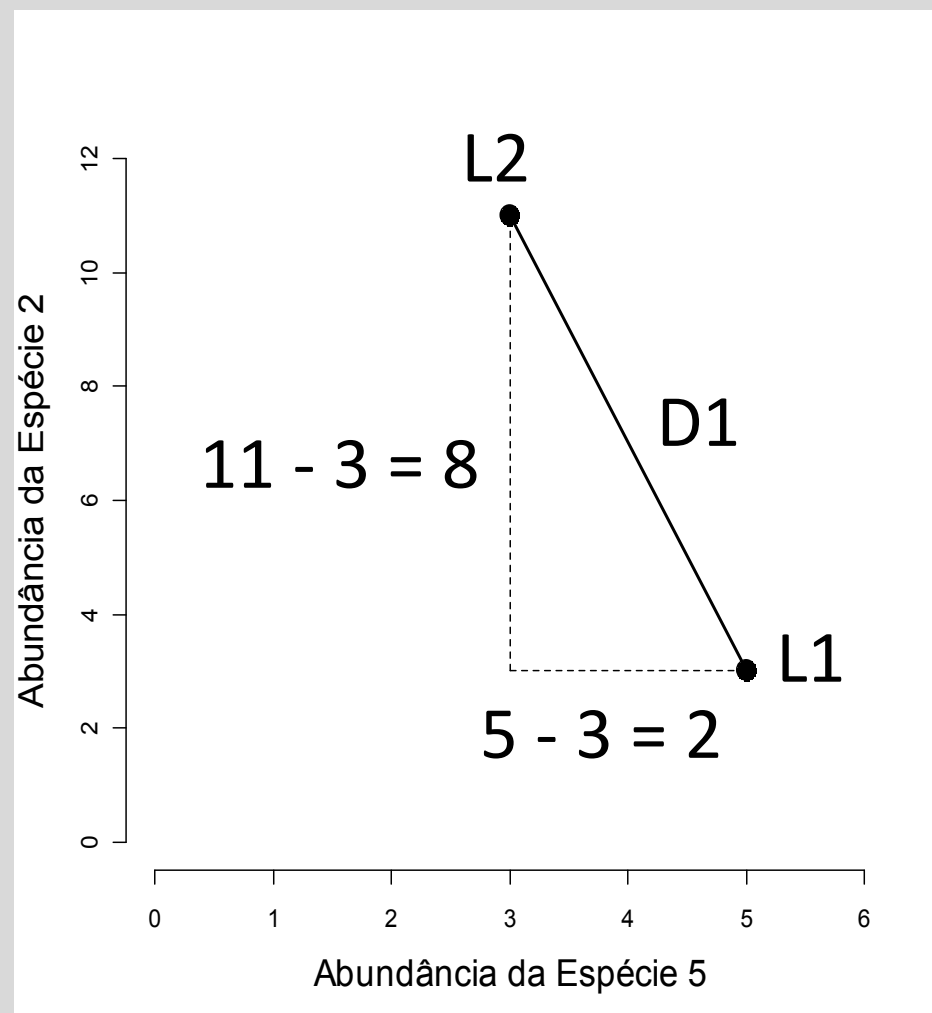
$$D = 1 - S \qquad D = \sqrt{1 - S} \qquad D = \sqrt{1 - S^2}$$

Análise em modo Q: índices distância ou similaridade

1. Distância Euclideana (DE)

	sp2	sp5
L1	3	5
L2	11	3

$$DE = \sqrt{(3 - 11)^2 + (5 - 3)^2} = 8,25$$



Análise em modo Q: índices distância ou similaridade

1. Distância Euclideana (DE)

		Descritores			
Objetos		Sp1	Sp2	Sp3	Sp4
	A	2	0	2	3
	B	5	3	0	6
	C	5	4	1	2
	D	3	2	0	0
	E	0	1	2	5

$$DE_{A-B} = \sqrt{\sum (Y_A - Y_B)^2}$$

Onde:

A = Valor da variável **Y** no Local A;

B = Valor da variável **Y** no Local B.

$$DE_{A-B} = \sqrt{(2-5)^2 + (0-3)^2 + (2-0)^2 + (3-6)^2}$$

$$DE_{A-B} = 5,57$$

Matrizes de distâncias: calculando as distâncias entre TODOS os pares de objetos

1. Distância Euclideana (DE)

Objetos

Descritores				
	Sp1	Sp2	Sp3	Sp4
A	2	0	2	3
B	5	3	0	6
C	5	4	1	2
D	3	2	0	0
E	0	1	2	5

Matriz de Distância					
	A	B	C	D	E
A	0				
B	5,57	0			
C	5,20	4,24	0		
D	4,24	6,40	3,61	0	
E	3,00	5,83	6,63	6,24	0

Distância Euclideana

$$DE_{A-B} = \sqrt{\sum (Y_A - Y_B)^2}$$

Matrizes de distâncias: calculando as distâncias entre TODOS os pares de objetos

1. Distância Euclideana (DE)

Objetos	Descritores				Matriz de Distância					
	Sp1	Sp2	Sp3	Sp4	A	B	C	D	E	
	A	2	0	2	3	0	5,57	5,20	4,24	3,00
	B	5	3	0	6	5,57	0	4,24	6,40	5,83
	C	5	4	1	2	5,20	4,24	0	3,61	6,63
	D	3	2	0	0	4,24	6,40	3,61	0	6,24
	E	0	1	2	5	3,00	5,83	6,63	6,24	0

Distância Euclideana

$$DE_{A-B} = \sqrt{\sum (Y_A - Y_B)^2}$$

Matrizes de distâncias: calculando as distâncias entre TODOS os pares de objetos

2. Índice de Jaccard

	Sp1	Sp2	Sp3	Sp4
A	1	0	1	1
B	1	1	0	1
C	1	1	1	1
D	1	1	0	1
E	0	1	1	1

	B	
A		
	1	0
	1	a
	0	c
		d

$$S = \frac{a}{a + b + c}$$

Matrizes de distâncias: calculando as distâncias entre TODOS os pares de objetos

2. Índice de Jaccard

	Sp1	Sp2	Sp3	Sp4
A	1	0	1	1
B	1	1	0	1
C	1	1	1	1
D	0	0	0	1
E	0	1	1	1

	B	
A	1	0
	1	b
	0	c

$$S = \frac{2}{2 + b + c}$$

Matrizes de distâncias: calculando as distâncias entre TODOS os pares de objetos

2. Índice de Jaccard

	Sp1	Sp2	Sp3	Sp4
A	1	0	1	1
B	1	1	0	1
C	1	1	1	1
D	0	0	0	1
E	0	1	1	1

	B	
A		
	1	0
	1	2
	0	c
		d

$$S = \frac{2}{2 + 1 + c}$$

Matrizes de distâncias: calculando as distâncias entre TODOS os pares de objetos

2. Índice de Jaccard

	Sp1	Sp2	Sp3	Sp4
A	1	0	1	1
B	1	1	0	1
C	1	1	1	1
D	0	0	0	1
E	0	1	1	1

	B	
A		
	1	0
	1	2
	0	1
		d

$$S = \frac{2}{2+1+1}$$

Matrizes de distâncias: calculando as distâncias entre TODOS os pares de objetos

2. Índice de Jaccard

	Sp1	Sp2	Sp3	Sp4
A	1	0	1	1
B	1	1	0	1
C	1	1	1	1
D	0	0	0	1
E	0	1	1	1

	B	
	1	0
A	1	1
	0	0

$$S = \frac{2}{2+1+1}$$

Matrizes de distâncias: calculando as distâncias entre TODOS os pares de objetos

2. Índice de Jaccard

	Sp1	Sp2	Sp3	Sp4
A	1	0	1	1
B	1	1	0	1
C	1	1	1	1
D	0	0	0	1
E	0	1	1	1

B		
	1	0
A	1	2
	0	1

$$S_{AB} = \frac{2}{2+1+1} = 0,5$$

Matrizes de distâncias: calculando as distâncias entre TODOS os pares de objetos

2. Índice de Jaccard

	Sp1	Sp2	Sp3	Sp4
A	1	0	1	1
B	1	1	0	1
C	1	1	1	1
D	0	0	0	1
E	0	1	1	1



Matriz de Similaridade					
	A	B	C	D	D
A	1				
B	0.50	1			
C	0.75	0.75	1		
D	0.33	0.33	0.25	1	
E	0.50	0.50	0.75	0.33	1

Índice de Jaccard

$$S = \frac{a}{a + b + c}$$

Matrizes de distâncias: calculando as distâncias entre TODOS os pares de objetos

Matriz de Distância						Matriz de Similaridade					
	A	B	C	D	D		A	B	C	D	D
A	0					A	1				
B	5.57	0				B	0.50	1			
C	5.20	4.24	0			C	0.75	0.75	1		
D	4.24	6.40	3.61	0		D	0.33	0.33	0.25	1	
E	3.00	5.83	6.63	6.24	0	E	0.50	0.50	0.75	0.33	1

1. São matrizes simétricas. Têm o mesmo número de linhas e colunas.

Matrizes de distâncias: calculando as distâncias entre TODOS os pares de objetos

Matriz de Distância						Matriz de Similaridade					
	A	B	C	D	D		A	B	C	D	D
A	0					A	1				
B	5.57	0				B	0.50	1			
C	5.20	4.24	0			C	0.75	0.75	1		
D	4.24	6.40	3.61	0		D	0.33	0.33	0.25	1	
E	3.00	5.83	6.63	6.24	0	E	0.50	0.50	0.75	0.33	1

1. São matrizes simétricas. Têm o mesmo número de linhas e colunas.
2. A diagonal principal mostra o grau de associação de um objeto com ele mesmo.

Matrizes de distâncias: calculando as distâncias entre TODOS os pares de objetos

Matriz de Distância						Matriz de Similaridade					
	A	B	C	D	D		A	B	C	D	D
A	0	5.57	5.20	4.24	3.00	A	1	0.50	0.75	0.33	0.50
B	5.57	0	4.24	6.40	5.83	B	0.50	1	0.75	0.33	0.50
C	5.20	4.24	0	3.61	6.63	C	0.75	0.75	1	0.25	0.75
D	4.24	6.40	3.61	0	6.24	D	0.33	0.33	0.25	1	0.33
E	3.00	5.83	6.63	6.24	0	E	0.50	0.50	0.75	0.33	1

1. São matrizes simétricas. Têm o mesmo número de linhas e colunas.
2. A diagonal principal mostra o grau de associação de um objeto com ele mesmo.
3. Os elementos acima (triangular superior) e abaixo (triangular inferior) da diagonal principal são idênticos

Agrupamento (Cluster): visualizando matrizes de distância

Matriz de Distância					
	A	B	C	D	D
A	0				
B	5.57	0			
C	5.20	4.24	0		
D	4.24	6.40	3.61	0	
E	3.00	5.83	6.63	6.24	0

Agrupamento (Cluster): visualizando matrizes de distância

Matriz de distância para a composição de espécies em 23 parcelas em um fragmento de cerrado

	L1	L2	L3	L4	L5	L6	L7	L8	L9	L10	L11	L12	L13	L14	L15	L16	L17	L18	L19	L20	L21	L22	L23
L1	0																						
L2	9,36	0																					
L3	20,02	29,38	0																				
L4	20,79	30,15	3,73	0																			
L5	23,44	32,8	6,38	2,65	0																		
L6	49,2	58,56	38,9	3,53	42,32	0																	
L7	66,25	75,61	55,95	20,58	59,37	21,91	0																
L8	65,96	75,32	55,66	20,29	59,08	21,62	9,13	0															
L9	67,49	76,85	57,19	21,82	60,61	23,15	10,66	1,53	0														
L10	61,63	70,99	51,33	15,96	54,75	17,29	17,66	17,37	18,9	0													
L11	62,99	72,35	52,69	17,32	56,11	18,65	19,02	18,73	20,26	6,22	0												
L12	66,7	76,06	56,4	21,03	59,82	22,36	22,73	22,44	23,97	5,07	11,29	0											
L13	67,58	76,94	57,28	21,91	60,7	23,24	23,61	23,32	24,85	5,95	12,17	3,9	0										
L14	66,6	75,96	56,3	20,93	59,72	22,26	22,63	22,34	23,87	4,97	11,19	2,92	0,98	0									
L15	65,29	74,65	54,99	19,62	58,41	20,95	21,32	21,03	22,56	3,66	9,88	1,61	2,29	1,31	0								
L16	67,13	76,49	56,83	21,46	60,25	22,79	23,16	22,87	24,4	5,5	11,72	0,43	4,33	3,35	2,04	0							
L17	68,92	78,28	58,62	23,25	62,04	24,58	24,95	24,66	26,19	7,29	13,51	5,24	1,34	2,32	3,83	5,67	0						
L18	45,68	55,04	35,38	36,15	38,8	50,32	67,37	67,08	68,61	62,75	64,11	67,82	68,7	67,72	66,41	68,25	70,04	0					
L19	45,75	55,11	35,45	36,22	38,87	50,39	67,44	67,15	68,68	62,82	64,18	67,89	68,77	67,79	66,48	68,32	70,11	0,07	0				
L20	47,14	56,5	36,84	37,61	40,26	51,78	68,83	68,54	70,07	64,21	65,57	69,28	70,16	69,18	67,87	69,71	71,5	1,46	1,39	0			
L21	47,27	56,63	36,97	37,74	40,39	51,91	68,96	68,67	70,2	64,34	65,7	69,41	70,29	69,31	68,00	69,84	71,63	1,59	1,52	0,49	0		
L22	39,36	48,72	29,06	29,83	32,48	44,00	61,05	60,76	62,29	56,43	57,79	61,5	62,38	61,4	60,09	61,93	63,72	10,2	10,27	11,66	11,79	0	
L23	38,52	47,88	28,22	28,99	31,64	43,16	60,21	59,92	61,45	55,59	56,95	60,66	61,54	60,56	59,25	61,09	62,88	9,36	9,43	10,82	10,95	0,84	0

$$\frac{23 \times (23 - 1)}{2} = 253 \text{ pares de valores}$$

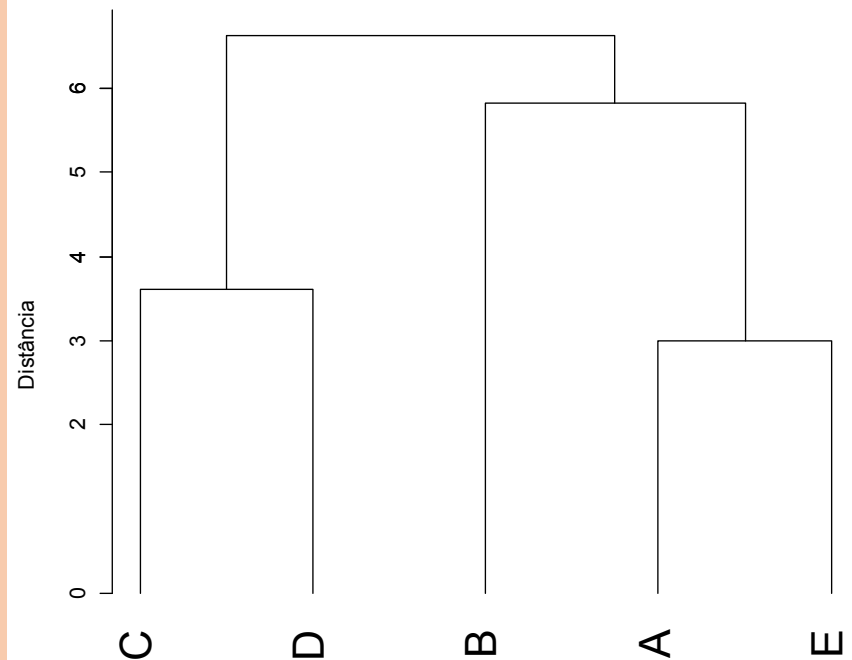
Agrupamento (Cluster): visualizando matrizes de distância

Matriz de Distância

	A	B	C	D	D
A	0				
B	5.57	0			
C	5.20	4.24	0		
D	4.24	6.40	3.61	0	
E	3.00	5.83	6.63	6.24	0



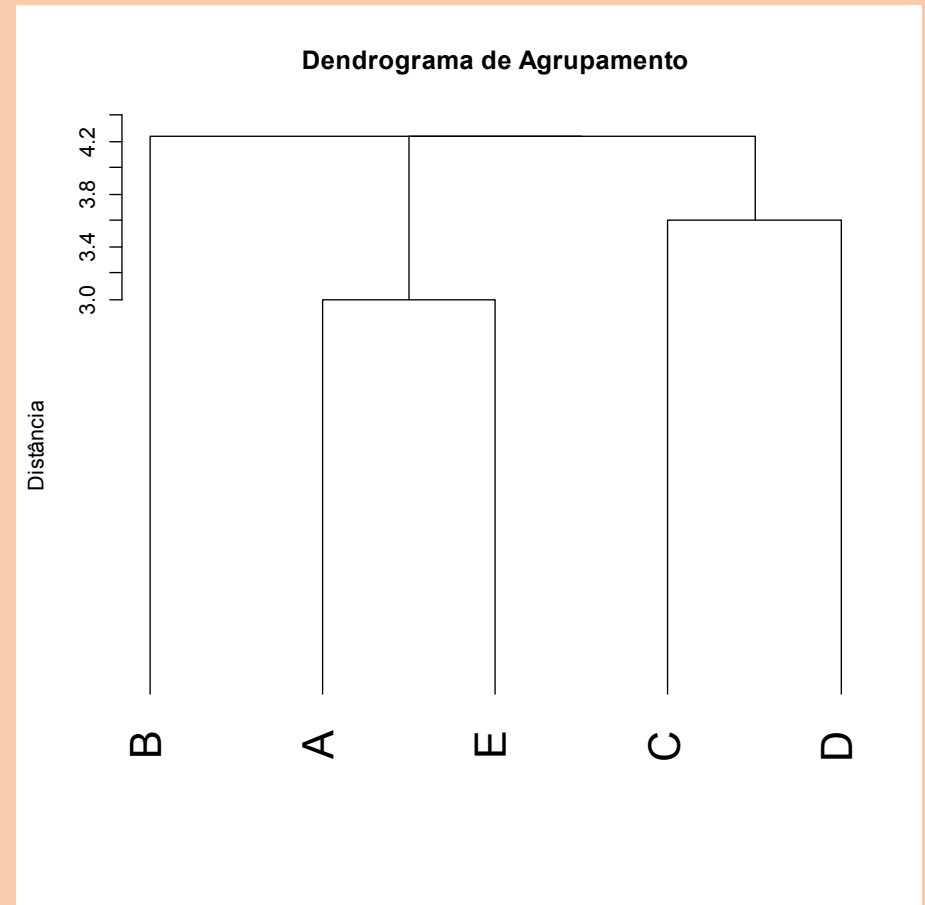
Dendrograma de Agrupamento



Agrupamento (Cluster): visualizando matrizes de distância

1. Método de ligação simples (vizinho mais próximo)

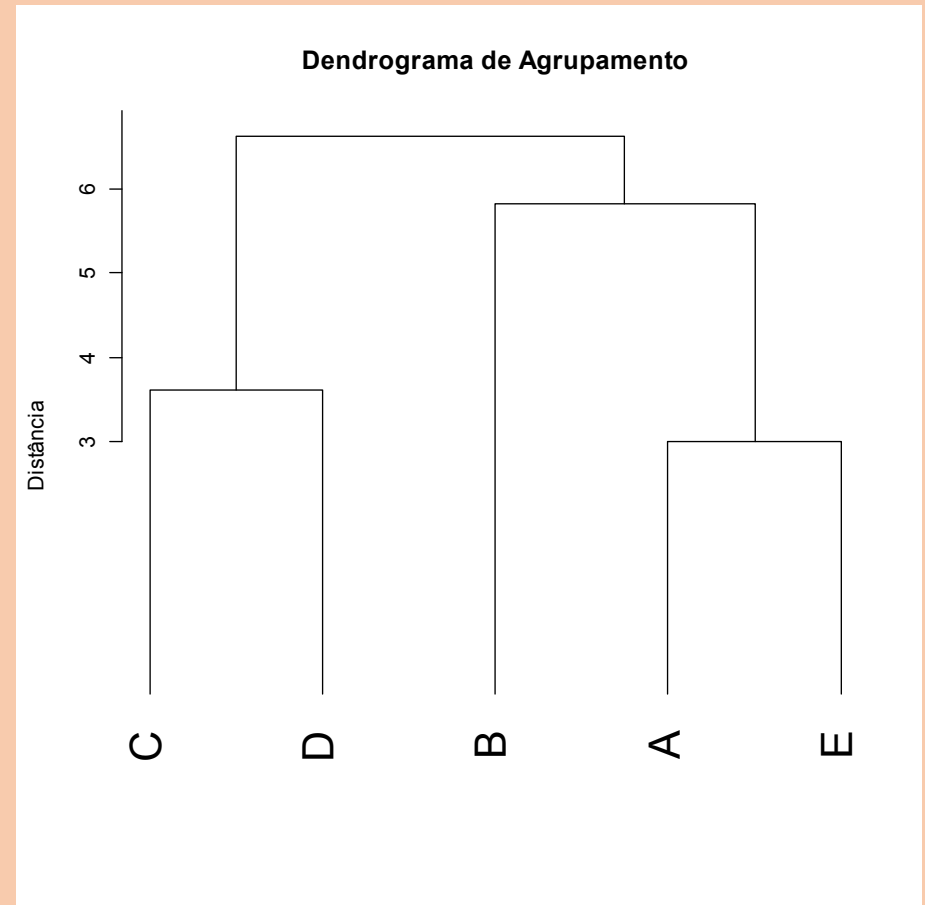
	A	B	C	D	E
A	0				
B	5.57	0			
C	5.20	4.24	0		
D	4.24	6.40	3.61	0	
E	3.00	5.83	6.63	6.24	0



Agrupamento (Cluster): visualizando matrizes de distância

2. Método de ligação simples (vizinho mais distante)

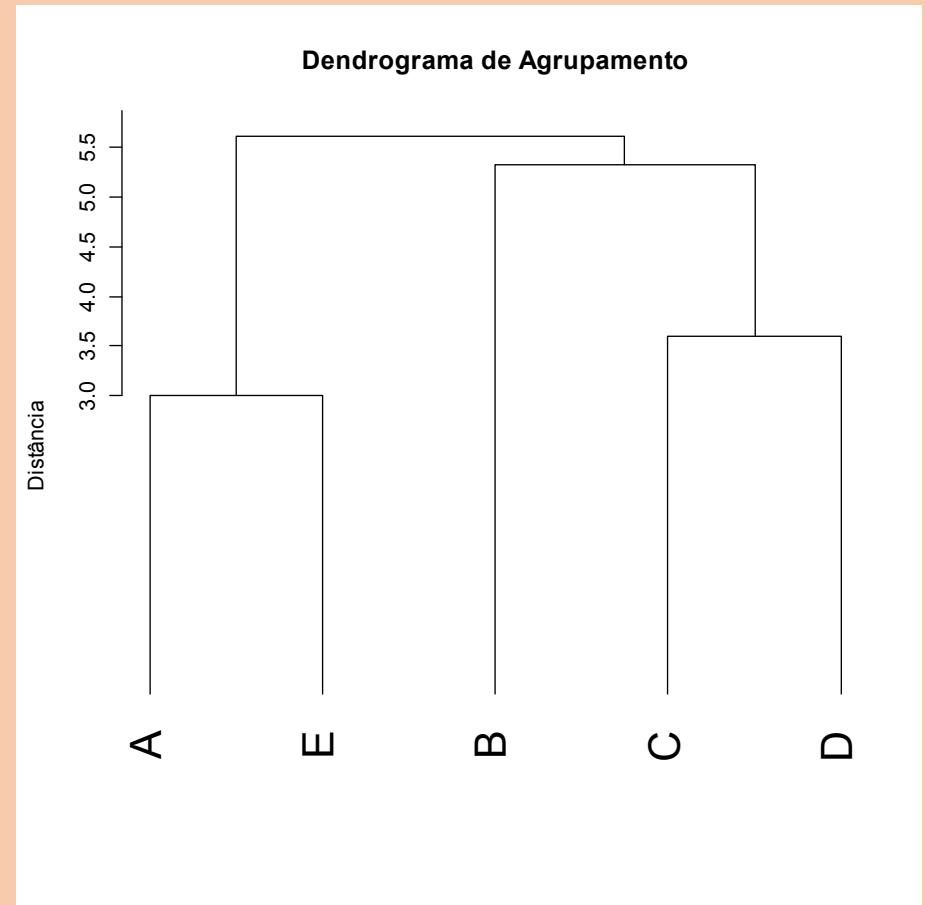
	A	B	C	D	E
A	0				
B	5.57	0			
C	5.20	4.24	0		
D	4.24	6.40	3.61	0	
E	3.00	5.83	6.63	6.24	0



Agrupamento (Cluster): visualizando matrizes de distância

3. Método de ligação pela média (UPGMA)

	A	B	C	D	E
A	0				
B	5.57	0			
C	5.20	4.24	0		
D	4.24	6.40	3.61	0	
E	3.00	5.83	6.63	6.24	0



Medindo a qualidade do agrupamento: correlação cofenética

