Medidas de Semelhança

Índices de Semelhança

Grandezas numéricas que quantificam o grau de associação entre um par de objetos ou de descritores.

Como escolher um Índice?

- O objetivo da análise é associar objetos (modo Q) ou descritores (modo R)?
- Os descritores são espécies?
  - (planilha homogênea, descritores com a mesma unidade)
- Ambientes?
  - (planilha heterogênea, descritores físicos, químicos etc.)
- Os dados são binários (presença ou ausência); quantitativos (discretos ou contínuos) ou semiquantitativos (codificados)?
- Os dados são brutos ou transformados?
- O coeficiente de associação deve dar mais importância a espécies raras ou igual importância?

Objetos e Descritores

- Exemplo: uma árvore
  - Descritor da área onde ela foi encontrada
  - Objeto quando forem realizadas medições nela (biometria, pesagem, composição química/anatômica).
- Existem métodos exclusivos para análise comparativa dos objetos (análise em modo Q) e outros para análise comparativa dos descritores (análise em modo R).

Estudos em modo Q
(Associação de objetos)

Coeficientes de similaridade
associação de objetos – modo Q

- Para dados binários
  - 1 = espécie presente; 0 = espécie ausente

		Objeto A: 1	Objeto A: 0	
Objeto B: 1		a	b	a + b
Objeto B: 0		c	d	c + d
		a + c	b + d	p

onde:

a = número de espécies comuns aos dois objetos (1 – 1);

b = número de presença-ausência (1 – 0);

c = número de ausência-presença (0 – 1);

d = número de dupla-ausência (0 – 0);

p = número total de espécies (= a + b + c + d).

Coeficientes binários

associação de objetos – modo Q

Tabela 6 Coeficientes de similaridade para dados binários, incluindo a dupla-ausência (baseado em ROMESBURG, 1984)

Nome	Expressão	Intervalo de variação
Coincidência simples	$S1 = \frac{a+d}{a+b+c+d}$	0–1
Sokal & Sneath	$S2 = \frac{2a+2d}{a+2b+2c+d}$	0–1
Russel & Rao	$S3 = \frac{a}{a+b+c+d}$	0–1
ϕ (phi) de Pearson	$S4 = \frac{ad-bc}{[(a+b)(a+c)(b+d)(c+d)]^{0.5}}$	–1–1
Ochiai	$S5 = \frac{ad}{[(a+b)(a+c)(b+d)(c+d)]^{0.5}}$	0–1

Coeficientes binários, excluindo a dupla-ausência

Tabela 7 Coeficientes de similaridade para dados binários, excluindo a dupla-ausência (baseado em ROMESBURG, 1984)

Nome	Expressão	Intervalo de variação
Jaccard	$S6 = \frac{a}{a+b+c}$	0–1
Sorensen	$S7 = \frac{2a}{2a+b+c}$	0–1

Exercício 1

Exemplo 1: sejam duas amostras A e B com a presença (1) ou ausência (0) de nove espécies.

ESPÉCIE	A	B	C	D	E	F	G	H	I
Amostra A	1	1	0	1	1	0	0	1	1
Amostra B	1	1	0	1	0	1	1	1	0

- Calcular os índices binários de similaridade entre as amostras e checar as diferenças.

Coeficientes quantitativos

The coefficient is then presented as:

$$S_S = \frac{2\sum_{i=1}^{m} \min(X_i, Y_i)}{\sum_{i=1}^{m} X_i + \sum_{i=1}^{m} Y_i} \qquad (4.7)$$

where X_i and Y_i are the abundances of species i, min (X_i, Y_i) is the sum of the lesser scores of species i where it occurs in both quadrats, and m is the number of species. The coefficient values range from 0 (complete dissimilarity) to 1 (total similarity).

Using the same quadrats (11 and 12) from the New Jersey data but with the quantitative values (Table 4.1), the Steinhaus (Sørensen/Czekanowski) coefficient is calculated below. The lower values of joint occurrences are underlined, and when summed give a total of 20.

Species	Quadrat 11	Quadrat 12
Atriplex patula	5	2
Distichlis spicata	20	0
Iva frutescens	20	10
Juncus gerardii	0	0
Phragmites communis	5	30
Salicornia europaea	0	0
Salicornia virginica	0	0
Scirpus olneyi	1	0
Solidago sempervirens	1	2
Spartina alterniflora	0	0
Spartina patens	2	5
Sueda maritima	0	0
Total cover	54	49
Sum of the lower scores of species common to both quadrats (underlined values)		20

Thus

Czekanowski

Steinhaus: matemático polonês

The coefficient is then presented as:

$$S_S = \frac{2\sum_{i=1}^{m} \min(X_i, Y_i)}{\sum_{i=1}^{m} X_i + \sum_{i=1}^{m} Y_i} \tag{4.7}$$

where X_i and Y_i are the abundances of species i, min (X_i, Y_i) is the sum of the lesser scores of species i where it occurs in both quadrats, and m is the number of species. The coefficient values range from 0 (complete dissimilarity) to 1 (total similarity).

Using the same quadrats (11 and 12) from the New Jersey data but with the quantitative values (Table 4.1), the Steinhaus (Sørensen/Czekanowski) coefficient is calculated below. The lower values of joint occurrences are underlined, and when summed give a total of 20.

Species	Quadrat 11	Quadrat 12
Atriplex patula	5	2
Distichlis spicata	20	0
Iva frutescens	20	10
Juncus gerardii	0	0
Phragmites communis	5	30
Salicornia europaea	0	0
Salicornia virginica	0	0
Scirpus olneyi	1	0
Solidago sempervirens	1	2
Spartina alterniflora	0	0
Spartina patens	2	5
Sueda maritima	0	0
Total cover	54	49
Sum of the lower scores of species common to both quadrats (underlined values)		20

Thus

$$S_S = \frac{2 \times 20}{54 + 49} = 0.39(39\%)$$

Coeficientes quantitativos

- Índice de Kulczynski

$$S = \frac{1}{2}\left(\frac{W}{A} + \frac{W}{B}\right)$$

W = Soma das abundâncias mínimas entre as espécies

A e B = soma das abundâncias nas amostras A e B

Coeficientes quantitativos

- Índice de Steinhaus

$$S = \frac{2W}{A+B}$$

W = Soma das abundâncias mínimas entre as espécies

A e B = soma das abundâncias nas amostras A e B

Exercício 2

	Espécies									Soma
Amostra *A*	5	2	0	3	1	0	0	4	3	18
Amostra *B*	3	1	0	2	0	2	3	3	0	14
Mínimo *W*	3	1	0	2	0	0	0	3	0	9

- Calcular os coeficientes de similaridade quantitativos entre as amostras e checar as diferenças.

Coeficientes de distância

- Quanto mais próximas forem as amostras, i.é., quanto menor a distância métrica entre os pontos representativos dessas duas amostras, maior será a similaridade entre elas.
  - Um índice de distância corresponde a uma dissimilaridade;
  - A similaridade (S) para uma distância (D) será dada por [S = 1 – D].

Coeficientes de distância

- Distância Euclidiana
  - Mais difundida.
  - Baseada no teorema de Pitágoras sobre a hipotenusa do triângulo retângulo.
  - Não tem limite superior. Aumenta na medida em que aumenta o número de descritores. Além disso, depende da escala de valores de cada descritor. Correções:
    - Padronização dos dados (centralizar e reduzir)
    - Distância Euclidiana média: D/n (n é o número de descritores).

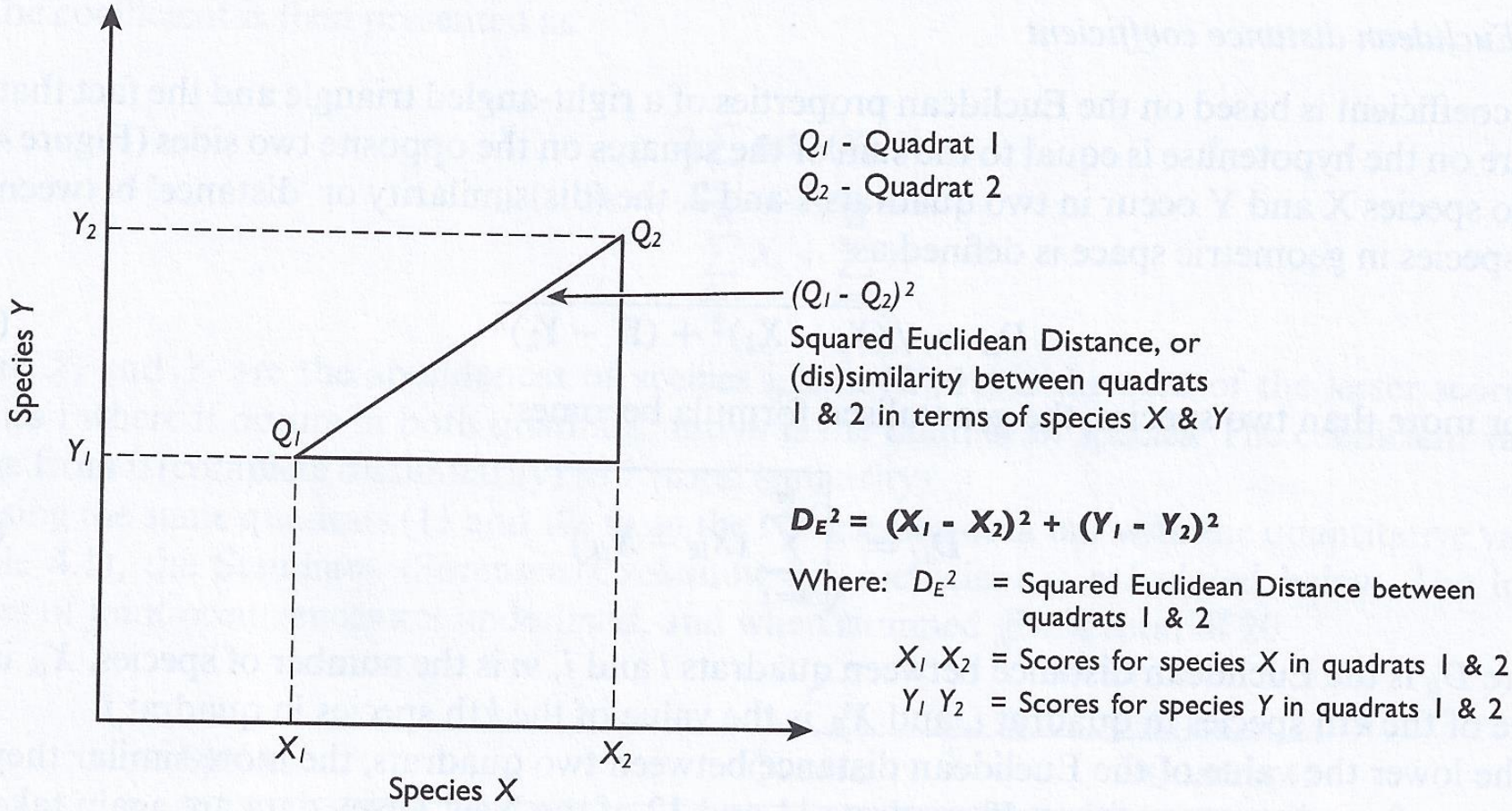

Figure 4.5 Euclidean distance between two quadrats/samples (1 and 2) in terms of two species (X and Y). Reproduced with permission from John Wiley & Sons, Inc.

Distância Euclidiana

$$D_{A-B} = \sqrt{\sum_{i=1}^{p} (x_{Ai} - x_{Bi})^2}$$

Distância Euclidiana entre as amostras A e B, em razão da abundância *x* de *p* espécies, *i* variando de 1 a *p*

Coeficientes de distância

- Distância Bray-Curtis
  - Varia de zero (similaridade) a 1 (dissimilaridade)
  - Não considera as duplas-ausências e é fortemente influenciado pelas espécies dominantes.
  - O cálculo é baseado nas diferenças absolutas e nas somas das abundâncias de cada espécie (*i*) nas duas amostras (A e B).
  - Ao fazer 1-D obtém-se o índice de Steinhaus.

$$D_{A-B} = \frac{\sum |x_{Ai} - x_{Bi}|}{\sum (x_{Ai} + x_{Bi})}$$

Coeficientes de distância

- Distância Camberra
  - Não tem limite superior. Vale zero quando todos os valores são iguais, e aumenta à medida que o número de espécies (p) aumenta.
  - A similaridade é obtida fazendo-se 1-(1/p). D variando entre 0 e 1.
  - Deve-se eliminar as duplas ausências.
  - Cada diferença é dividida pelo total da espécie

Distância de Camberra

$$D3_{A-B} = \sum_{i=1}^{p} \left[\frac{|x_{Ai} - x_{Bi}|}{(x_{Ai} + x_{Bi})} \right]$$

Coeficientes de distância

- Distância de Manhattan
  - Corresponde à soma das diferenças absolutas entre as abundâncias das espécies nas duas amostras A e B.

Distância de Manhattan

$$D4_{A-B} = \sum_{i=1}^{p} |x_{Ai} - x_{Bi}|$$

Coeficientes de distância

- Porcentagem de divergência
  - Deve ser aplicada de preferência a dados binários.
  - Compara o número de divergências (b + c) ao total de dados, com dupla-ausência (D5) ou sem dupla-ausência (D6)

$$D5_{A-B} = \frac{b+c}{a+b+c+d}$$

$$D6_{A-B} = \frac{b+c}{a+b+c}$$

Estudos em modo R
(Associação de descritores)

Coeficientes de dependência

- Esses coeficientes podem ser submetidos a testes estatísticos de significância, devendo, para isso, respeitar certas exigências, como a normalidade dos dados, no caso do coeficiente de correlação linear de Pearson.

Matriz de variâncias-covariâncias

- A covariância é a medida de dispersão conjunta de duas variáveis Y_k e Y_j em torno das suas médias m_k e m_j.
- Considerando-se *n* amostras onde foram medidas *p* variáveis Y, ou seja, o vetor $[Y_1, Y_2, ..., Y_p]$ observado *n* vezes. Pode ser construída a matriz de covariância ***S*** para as *p* variáveis tomadas duas a duas:

$$S_{k,j}^2 = \frac{1}{v}\sum_{i=1}^{n}(Y_{ki} - m_k)(Y_{ji} - m_j)$$

Onde *v* é o grau de liberdade (*n-1*) em relação a *n*.

- A matriz ***S*** será:

$$S_{pp}\begin{bmatrix} S_{11} & S_{12} & \cdots & S_{1p} \\ S_{21} & S_{22} & \cdots & S_{2p} \\ \cdots & \cdots & \cdots & \cdots \\ S_{p1} & S_{p2} & \cdots & S_{pp} \end{bmatrix}$$

- Ela é simétrica: os elementos da diagonal principal são as variâncias e os outros elementos as covariâncias.
- **S** pode ser obtida multiplicando-se a matriz dos dados centralizados **Y**=[y-m_y] pela sua transposta **Y'**:

$$\boldsymbol{S} = \frac{1}{v}\mathbf{Y'.Y}$$

- A matriz ***S*** será:

$$S_{pp}\begin{bmatrix} S_{11} & S_{12} & \cdots & S_{1p} \\ S_{21} & S_{22} & \cdots & S_{2p} \\ \cdots & \cdots & \cdots & \cdots \\ S_{p1} & S_{p2} & \cdots & S_{pp} \end{bmatrix}$$

- Ela é simétrica: os elementos da diagonal principal são as variâncias e os outros elementos as covariâncias.
- **S** pode ser obtida multiplicando-se a matriz dos dados centralizados **Y**=[y-m_y] pela sua transposta **Y'**:

$$\boldsymbol{S} = \frac{1}{v}\mathbf{Y'.Y}$$

Exemplo

- Calcular a matriz de variâncias e covariâncias entre quatro espécies (p=4) contadas em n = 5 estações de coleta.

		Dados brutos			
	Espécies	Y_1	Y_2	Y_3	Y_4
Amostras	(1)	1	5	2	6
	(2)	2	2	1	8
	(3)	3	1	3	4
	(4)	4	2	5	0
	(5)	5	5	4	2
	Médias	3	3	3	4

Exemplo

- Calcular a matriz de variâncias e covariâncias entre quatro espécies (p=4) contadas em n = 5 estações de coleta.

- **S** pode ser obtida multiplicando-se a matriz dos dados centrados **Y**=[y-m_y] pela sua transposta **Y'**:

$$\boldsymbol{S} = \frac{1}{v}\mathbf{Y'.Y}$$

Exemplo

- Calcular a matriz de variâncias e covariâncias entre quatro espécies (p=4) contadas em n = 5 estações de coleta.

- **S** pode ser obtida multiplicando-se a matriz dos dados centrados **Y**=[y-m_y] pela sua transposta **Y'**:

$$\boldsymbol{S} = \frac{1}{v}\mathbf{Y'.Y}$$

Passo 1: centrar os dados: Y=[y-m]

		Dados brutos			
	Espécies	Y_1	Y_2	Y_3	Y_4
Amostras	(1)	1	5	2	6
	(2)	2	2	1	8
	(3)	3	1	3	4
	(4)	4	2	5	0
	(5)	5	5	4	2
	Médias	3	3	3	4

Passo 1: centrar os dados: Y=[y-m]

		Dados centrados			
	Espécies	Y_1-3	Y_2-3	Y_3-3	Y_4-4
Amostras	(1)	–2	2	–1	2
	(2)	–1	–1	–2	4
	(3)	0	–2	0	0
	(4)	1	–1	2	–4
	(5)	2	2	1	–2
	Médias	0	0	0	0

Passo 2: fazer a multiplicação da matriz transposta pela matriz centrada

Passo 2: fazer a multiplicação da matriz transposta pela matriz centrada

O grau de liberdade é $\nu = n - 1 = 4$.

Matriz centralizada

$$\begin{bmatrix} -2 & 2 & -1 & 2 \\ -1 & -1 & -2 & 4 \\ 0 & -2 & 0 & 0 \\ 1 & -1 & 2 & -4 \\ 2 & 2 & 1 & -2 \end{bmatrix}$$

Passo 2: fazer a multiplicação da matriz transposta pela matriz centrada

O grau de liberdade é $\nu = n - 1 = 4$.

$$S = \frac{1}{4} \cdot \begin{bmatrix} -2 & -1 & 0 & 1 & 2 \\ 2 & -1 & -2 & -1 & 2 \\ -1 & -2 & 0 & 2 & 1 \\ 2 & 4 & 0 & -4 & -2 \end{bmatrix} \cdot \begin{bmatrix} -2 & 2 & -1 & 2 \\ -1 & -1 & -2 & 4 \\ 0 & -2 & 0 & 0 \\ 1 & -1 & 2 & -4 \\ 2 & 2 & 1 & -2 \end{bmatrix}$$

Matrizes transposta e centralizada

Passo 2: fazer a multiplicação da matriz transposta pela matriz centrada

O grau de liberdade é $\nu = n - 1 = 4$.

$$\mathbf{S} = \frac{1}{4} \cdot \begin{bmatrix} -2 & -1 & 0 & 1 & 2 \\ 2 & -1 & -2 & -1 & 2 \\ -1 & -2 & 0 & 2 & 1 \\ 2 & 4 & 0 & -4 & -2 \end{bmatrix} \cdot \begin{bmatrix} -2 & 2 & -1 & 2 \\ -1 & -1 & -2 & 4 \\ 0 & -2 & 0 & 0 \\ 1 & -1 & 2 & -4 \\ 2 & 2 & 1 & -2 \end{bmatrix}$$

Efetuar a multiplicação

Passo 2: fazer a multiplicação da matriz transposta pela matriz centrada

O grau de liberdade é $\nu = n - 1 = 4$.

$$\mathbf{S} = \frac{1}{4} \cdot \begin{bmatrix} -2 & -1 & 0 & 1 & 2 \\ 2 & -1 & -2 & -1 & 2 \\ -1 & -2 & 0 & 2 & 1 \\ 2 & 4 & 0 & -4 & -2 \end{bmatrix} \cdot \begin{bmatrix} -2 & 2 & -1 & 2 \\ -1 & -1 & -2 & 4 \\ 0 & -2 & 0 & 0 \\ 1 & -1 & 2 & -4 \\ 2 & 2 & 1 & -2 \end{bmatrix}$$

$$\mathbf{S} = \begin{bmatrix} 2{,}5 & 0 & 2 & -4 \\ 0 & 3{,}5 & 0 & 0 \\ 2 & 0 & 2{,}5 & -5 \\ -4 & 0 & -5 & 10 \end{bmatrix}$$

Passo 2: fazer a multiplicação da matriz transposta pela matriz centrada

O grau de liberdade é $\nu = n - 1 = 4$.

$$\mathbf{S} = \frac{1}{4} \cdot \begin{bmatrix} -2 & -1 & 0 & 1 & 2 \\ 2 & -1 & -2 & -1 & 2 \\ -1 & -2 & 0 & 2 & 1 \\ 2 & 4 & 0 & -4 & -2 \end{bmatrix} \cdot \begin{bmatrix} -2 & 2 & -1 & 2 \\ -1 & -1 & -2 & 4 \\ 0 & -2 & 0 & 0 \\ 1 & -1 & 2 & -4 \\ 2 & 2 & 1 & -2 \end{bmatrix}$$

Variâncias e Covariâncias

$$\mathbf{S} = \begin{bmatrix} 2,5 & 0 & 2 & -4 \\ 0 & 3,5 & 0 & 0 \\ 2 & 0 & 2,5 & -5 \\ -4 & 0 & -5 & 10 \end{bmatrix}$$

Quais são as variâncias?

		Dados brutos			
	Espécies	Y_1	Y_2	Y_3	Y_4
Amostras	(1)	1	5	2	6
	(2)	2	2	1	8
	(3)	3	1	3	4
	(4)	4	2	5	0
	(5)	5	5	4	2
	Médias	3	3	3	4

Matriz de correlação

- É a matriz **R** de dispersão conjunta entre duas variáveis centralizadas e reduzidas:

$$z = \frac{y - m_y}{s_y}$$

- O coeficiente calculado é o coeficiente de correlação linear *r* de Pearson.
- A matriz ***R*** é calculada da mesma forma que a matriz ***S***, multiplicando a matriz de dados centralizados e reduzidos ***Z*** pela sua transposta ***Z'***.

Usando os mesmos dados do exemplo da matriz de var-covar

- Divide-se cada elemento da matriz de dados centralizados pelo desvio-padrão (*s)* da respectiva variável. Os valores de *s* encontram-se na diagonal de *S*.

Usando os mesmos dados do exemplo da matriz de var-covar

- Divide-se cada elemento da matriz de dados centralizados pelo desvio-padrão (*s)* da respectiva variável. Os valores de *s* encontram-se na diagonal de *S*.

		Dados centrados e reduzidos			
	Espécies	Z1	Z2	Z3	Z4
Amostras	(1)	$-2/\sqrt{2,5}$	$-2/\sqrt{3,5}$	$-1/\sqrt{2,5}$	$-2/\sqrt{10}$
	(2)	$-1/\sqrt{2,5}$	−0,53	−1,27	1,27
	(3)	0	−1,07	0	0
	(4)	$-1/\sqrt{2,5}$	−0,53	1,27	−1,27
	(5)	$-2/\sqrt{2,5}$	1,07	0,63	−0,63
Média		0	0	0	0
Variância		1	1	1	1

Usando os mesmos dados do exemplo da matriz de var-covar

- Divide-se cada elemento da matriz de dados centrados pelo desvio-padrão (*s)* da respectiva variável. Os valores de *s* encontram-se na diagonal de *S*.

		Dados centrados e reduzidos			
	Espécies	Z1	Z2	Z3	Z4
Amostras	(1)	$-2/\sqrt{2,5}$	$-2/\sqrt{3,5}$	$-1/\sqrt{2,5}$	$-2/\sqrt{10}$
	(2)	$-1/\sqrt{2,5}$	−0,53	−1,27	1,27
	(3)	0	−1,07	0	0
	(4)	$-1/\sqrt{2,5}$	−0,53	1,27	−1,27
	(5)	$-2/\sqrt{2,5}$	1,07	0,63	−0,63
Média		0	0	0	0
Variância		1	1	1	1

Checar os sinais negativos e positivos dos valores:
alguns estão trocados em relação à matriz reduzida

Matriz de correlação

$$\mathbf{R} = \frac{1}{\nu} \cdot \mathbf{Z}' \cdot \mathbf{Z}$$

Matriz de correlação

$$\mathbf{R} = \frac{1}{\nu} \cdot \mathbf{Z}' \cdot \mathbf{Z} \,.$$

$$\mathbf{R} = \begin{bmatrix} 1 & 0 & 0,8 & -0,8 \\ 0 & 1 & 0 & 0 \\ 0,8 & 0 & 1 & -1 \\ -0,8 & 0 & -1 & 1 \end{bmatrix}$$

Coeficiente de correlação linear **r** de Pearson

- Expressa exclusivamente a intensidade da relação linear entre duas variáveis.
- Pode ser submetido a testes estatísticos para verificar se seu valor é significativamente diferente de zero
- A ausência de significância (baixo valor de r) revela somente ausência de relação linear entre duas variáveis, podendo haver relação não-linear. Neste caso, pode-se fazer uma transformação linearizante dos dados ou usar coeficientes não paramétricos.

Situações de dependências entre variáveis (*r*)

independência

Relação linear positiva

Relação linear negativa

Relação não linear

Bimodal, duas populações

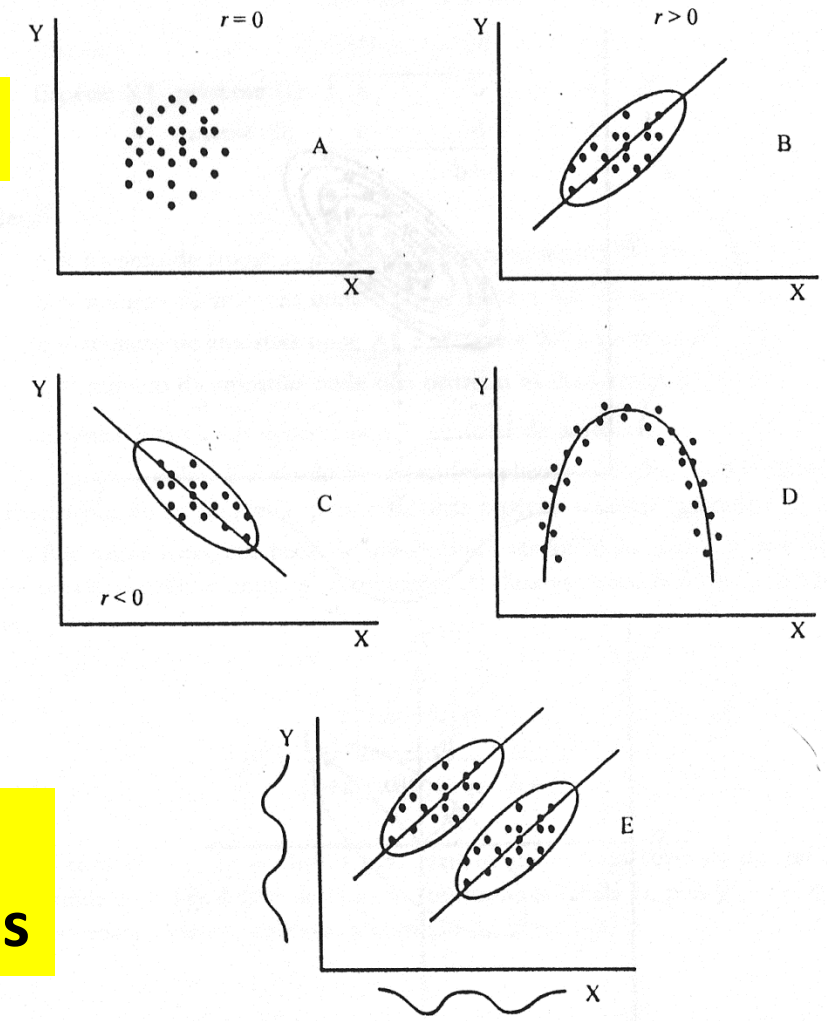

Figura 7 – Exemplos de relações lineares e não-lineares entre duas variáveis.

As Associações de Espécies

- As associações biológicas podem ser definidas a partir de co-ocorrências entre espécies, em vez de correlações entre abundâncias.
- Uma das formas mais simples de analisar a vegetação é observando o nível de associação entre espécies e o nível de similaridade entre parcelas, amostras ou áreas.

		Espécie X1		
		presente (1)	**ausente (0)**	
Espécie X2	presente (1)	a	b	$a + b$
	ausente (0)	c	d	$c + d$
		$a + c$	$b + d$	n

onde:

a = número de amostras possuindo as duas espécies (1-1);

b = número de amostras onde $X1$ é presente e $X2$ é ausente (1–0);

c = número de amostras onde $X1$ é ausente e $X2$ é presente (0–1);

d = número de amostras onde não ocorrem as duas espécies (0–0).

A soma $n = a + b + c + d$ corresponde ao total de amostras.

Qui-quadrado (χ^2) como medida de associação entre espécies

- Se o χ^2 for significativo a um determinado nível de probabilidade, há associação entre as duas espécies.
- Se a frequência observada para a ocorrência conjunta das espécies for maior do que o valor da frequência esperada, então as duas espécies são positivamente associadas.
- Associação positiva: duas espécies são encontradas juntas mais frequentemente do que seria esperado pelo acaso.

Qui-quadrado (χ^2) como medida de associação entre espécies

Inga cyli				
Myra urun		+	-	
	+	a	c	a+c
	-	b	d	b+d
		a+b	c+d	

Qui-quadrado (χ^2) como medida de associação entre espécies

Inga cyli				
Myra urun		+	-	
	+	112	15	127
	-	25	28	53
		137	43	180

- Calcula-se as frequências esperadas e o valor de χ^2

$$\chi^2 = \sum_{i=1}^{K} \frac{(f_0 - f_e)^2}{f_e} \qquad f_e = \frac{(total_da_linha)(total_da_coluna)}{N}$$

Qui-quadrado (χ^2) como medida de associação entre espécies

Inga cyli				
Myra urun		+	-	
	+	112 (96)	15 (30)	127
	-	25 (40)	28 (12)	53
		137	43	180

- Calcula-se as frequências esperadas e o valor de χ^2

$$f_e = \frac{(total_da_linha)(total_da_coluna)}{N}$$

Qui-quadrado (χ^2) como medida de associação entre espécies

Inga cyli				
Myra urun		+	-	
	+	112 (96)	15 (30)	127
	-	25 (40)	28 (12)	53
		137	43	180

$$\chi^2 = \sum_{i=1}^{K} \frac{(f_0 - f_e)^2}{f_e}$$

χ^2=37,17

TABELA IV

Distribuição do Qui-Quadrado - χ^2_n

Os valores tabelados correspondem aos pontos x tais que: $P(\chi^2_n \leq x)$

	$P(\chi^2_n \leq x)$													
n	0,005	0,01	0,025	0,05	0,1	0,25	0,5	0,75	0,9	0,95	0,975	0,99	0,995	
1	3,93E-05	0,000157	0,000982	0,003932	0,016	0,102	0,455	1,323	2,706	3,841	5,024	6,635	7,879	1
2	0,010	0,020	0,051	0,103	0,211	0,575	1,386	2,773	4,605	5,991	7,378	9,210	10,597	2
3	0,072	0,115	0,216	0,352	0,584	1,213	2,366	4,108	6,251	7,815	9,348	11,345	12,838	3
4	0,207	0,297	0,484	0,711	1,064	1,923	3,357	5,385	7,779	9,488	11,143	13,277	14,860	4
5	0,412	0,554	0,831	1,145	1,610	2,675	4,351	6,626	9,236	11,070	12,832	15,086	16,750	5
6	0,676	0,872	1,237	1,635	2,204	3,455	5,348	7,841	10,645	12,592	14,449	16,812	18,548	6
7	0,989	1,239	1,690	2,167	2,833	4,255	6,346	9,037	12,017	14,067	16,013	18,475	20,278	7
8	1,344	1,647	2,180	2,733	3,490	5,071	7,344	10,219	13,362	15,507	17,535	20,090	21,955	8
9	1,735	2,088	2,700	3,325	4,168	5,899	8,343	11,389	14,684	16,919	19,023	21,666	23,589	9
10	2,156	2,558	3,247	3,940	4,865	6,737	9,342	12,549	15,987	18,307	20,483	23,209	25,188	10
11	2,603	3,053	3,816	4,575	5,578	7,584	10,341	13,701	17,275	19,675	21,920	24,725	26,757	11
12	3,074	3,571	4,404	5,226	6,304	8,438	11,340	14,845	18,549	21,026	23,337	26,217	28,300	12
13	3,565	4,107	5,009	5,892	7,041	9,299	12,340	15,984	19,812	22,362	24,736	27,688	29,819	13
14	4,075	4,660	5,629	6,571	7,790	10,165	13,339	17,117	21,064	23,685	26,119	29,141	31,319	14
15	4,601	5,229	6,262	7,261	8,547	11,037	14,339	18,245	22,307	24,996	27,488	30,578	32,801	15
16	5,142	5,812	6,908	7,962	9,312	11,912	15,338	19,369	23,542	26,296	28,845	32,000	34,267	16
17	5,697	6,408	7,564	8,672	10,085	12,792	16,338	20,489	24,769	27,587	30,191	33,409	35,718	17
18	6,265	7,015	8,231	9,390	10,865	13,675	17,338	21,605	25,989	28,869	31,526	34,805	37,156	18
19	6,844	7,633	8,907	10,117	11,651	14,562	18,338	22,718	27,204	30,144	32,852	36,191	38,582	19
20	7,434	8,260	9,591	10,851	12,443	15,452	19,337	23,828	28,412	31,410	34,170	37,566	39,997	20
21	8,034	8,897	10,283	11,591	13,240	16,344	20,337	24,935	29,615	32,671	35,479	38,932	41,401	21
22	8,643	9,542	10,982	12,338	14,041	17,240	21,337	26,039	30,813	33,924	36,781	40,289	42,796	22
23	9,260	10,196	11,689	13,091	14,848	18,137	22,337	27,141	32,007	35,172	38,076	41,638	44,181	23
24	9,886	10,856	12,401	13,848	15,659	19,037	23,337	28,241	33,196	36,415	39,364	42,980	45,558	24
25	10,520	11,524	13,120	14,611	16,473	19,939	24,337	29,339	34,382	37,652	40,646	44,314	46,928	25
26	11,160	12,198	13,844	15,379	17,292	20,843	25,336	30,435	35,563	38,885	41,923	45,642	48,290	26
27	11,808	12,878	14,573	16,151	18,114	21,749	26,336	31,528	36,741	40,113	43,195	46,963	49,645	27
28	12,461	13,565	15,308	16,928	18,939	22,657	27,336	32,620	37,916	41,337	44,461	48,278	50,994	28
29	13,121	14,256	16,047	17,708	19,768	23,567	28,336	33,711	39,087	42,557	45,722	49,588	52,335	29
30	13,787	14,953	16,791	18,493	20,599	24,478	29,336	34,800	40,256	43,773	46,979	50,892	53,672	30
40	20,707	22,164	24,433	26,509	29,051	33,660	39,335	45,616	51,805	55,758	59,342	63,691	66,766	40
50	27,991	29,707	32,357	34,764	37,689	42,942	49,335	56,334	63,167	67,505	71,420	76,154	79,490	50
60	35,534	37,485	40,482	43,188	46,459	52,294	59,335	66,981	74,397	79,082	83,298	88,379	91,952	60
70	43,275	45,442	48,758	51,739	55,329	61,698	69,334	77,577	85,527	90,531	95,023	100,425	104,215	70
80	51,172	53,540	57,153	60,391	64,278	71,145	79,334	88,130	96,578	101,879	106,629	112,329	116,321	80
90	59,196	61,754	65,647	69,126	73,291	80,625	89,334	98,650	107,565	113,145	118,136	124,116	128,299	90
100	67,328	70,065	74,222	77,929	82,358	90,133	99,334	109,141	118,498	124,342	129,561	135,807	140,170	100

Exercício

1. Calcular e apresentar a matriz de correlações para a distribuição das seguintes espécies em relação a quatro unidades amostrais. Use o PAST.

	espécies			
Amostras	1	2	3	4
A	2	5	32	4
B	4	10	16	10
C	6	20	8	18
D	8	40	4	6

Exercício II

2. Com os mesmos dados, apresentar duas matrizes de distâncias e duas matrizes de similaridades entre as amostras e interpretar as diferenças.

	espécies			
Amostras	1	2	3	4
A	2	5	32	4
B	4	10	16	10
C	6	20	8	18
D	8	40	4	6