

## Índices de similaridade ou dissimilaridade – Produzem matrizes, tipo triangulares, cheia de valores

NOME	Compara. ..	Admite quais tipos de dados?	O que está medindo	Para pensar!	Geralmente, valores maiores indicam maior...
Distância euclidiana	Casos	numérico	distâncias entre os casos	Diferentes escalas de medida vão causar impactos	dissimilaridade
Correspondência simples	Casos	presença/ausência	combinações de presença/ausência	Uma combinação ausência/ausência é tão importante quanto presença/presença	similaridade
Jaccard	Casos	presença/ausência	combinações de presença/ausência, excluindo a de ausência/ausência		similaridade
Dice	Casos	presença/ausência	combinações de presença/ausência, dando duplo peso presença/presença e excluindo a de ausência/ausência		similaridade
Gower	Casos	numérico, ordinal, presença/ausência, categoria	métodos distintos conforme os tipos de dados (combo de distância para numérico e ordinal; Jaccard/Dice para presença/ausência e categorias)	Contribuição de peso igual para cada variável. Continua usando presença/ausência para categorias mas não as tornam muitas variáveis distintas	similaridade
Variância-covariância	Variáveis	numérico, ordinal, presença/ausência	variação compartilhada	Diferentes escalas de medida vão causar impactos	
Correlação	Variáveis	numérico, ordinal, presença/ausência	coeficientes de correlação	Valores vão ser de 0 a 1 (efetivamente é a padronização de dados de covariância)	similaridade
Resultados de $X^2$ ( $V$ , $\phi$ ou $\phi^2$ )	Variáveis	categorias		Só vai ser usada em técnicas de ordenação como correspondência	

Visualizando relações entre presença e ausência

		CASO 1	
		Presença	Ausência
CASO 2	Presença	A	B
	Ausência	C	D

Correspondência simples

		CASO 1	
		Presença	Ausência
CASO 2	Presença	A	B
	Ausência	C	D

Jaccard

		CASO 1	
		Presença	Ausência
CASO 2	Presença	A	B
	Ausência	C	

Dice

		CASO 1	
		Presença	Ausência
CASO 2	Presença	2 x A	B
	Ausência	C	

## Análises multivariadas

Compara	O que produz?	NOME	O que faz?	O que saber?	Detalhes
Casos	Agrupamentos	K means cluster	Cada caso será dado identidade em um grupo	Não é hierárquica; você decide quantos grupos tem	A visualização de clusters é um passo separado. O resultado do cluster pode ser colocado em nova coluna como identidade para agrupar casos.
Casos	Arvores (uniões)	Cluster hierárquica	Decide em os pontos, sequencialmente, à unir os casos	O tipo de união/linkage escolhida é muito importante	Nearest-neighbors - união por menor distância entre grupos. Unweighted pair-group average (UPGMA) - união por distância média entre todos os membros dos dois grupos. Ward's method - união para minimizar aumentos na variância dentro dos grupos existentes.
Casos	Arvore tipo filogenética	Neighbor-joining	Encontra similaridade		Indicado para alguns dados ecológicos além da filogenia.
Casos	Mapeamento em 2D / 3D	Análise de coordenados principais (PCoA)	Mapear os casos em função das suas similaridades/ diferenças para melhor representar todos os casos	Serve para visualizar a estrutura dos dados, incluindo clusters	É exploratória. Produz um serie de coordenados. Eigenvalues (autovalores) estão iguais a variância explicada naquela dimensão.
Frequências	X <sup>2</sup> distâncias	Correspondência	Representação gráfica das categorias	Já foi bastante utilizada em serriação	Pode fazer de metodo "canonical" que inclui variáveis ambientais como temperatura como eixo que deve influenciar os dados. Análise pode ter fileiras 'misteriosos' em que variável ambiental não está conhecida
Variáveis	Variáveis melhores	Análise de componentes principais (PCA)	Captura o máximo da variação nos dados no menor número de novos variáveis, tornando as variáveis possivelmente correlacionadas (seus) em variáveis não correlacionadas	Visualização está em relação as novas variáveis ou componentes	Eigenvalues (autovalores) estão iguais a variância explicada naquele componente. Análise fatorial é muito, muito parecida.
Variáveis	Maior separação entre grupos	Análise discriminante (Canonical variates analysis)	Separação máxima entre os grupos que você especificou através de eixos que são combinações lineares das variáveis	Pressupõe que existem variáveis dependentes. Agrupamentos já indicados.	Pode incluir casos para os quais não se sabe o grupo, e esses serão "classificadas"