PROXIMITY OR DISTANCE OR DISSIMILARITY MEASURES

(1) BINARY DATA

		Object j	
		+	-
Object i	+	a	b
	-	c	d

Jaccard coefficient	Dissimilarity (1–S)
$S_J = \frac{a}{a+b+c}$	$D_J = \frac{b+c}{a+b+c}$

Simple matching coefficient

$$S_{SMC} = \frac{a+d}{a+b+c+d} \qquad D_{SMC} = \frac{b+c}{a+b+c+d}$$

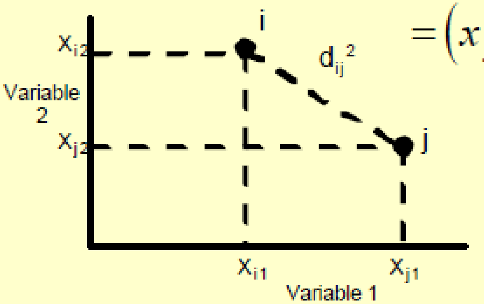
Baroni-Urbani & Buser

$$\frac{\sqrt{ad}+a}{\sqrt{ad}+a+b+c}$$

Syst. Zool. 1976 25, 251-9

Hubalek	1982	Biol. Rev. 97, 669-689
Gower & Legendre	1986	J. Classific. 3, 5-48
Archer & Maples	1987	Palaois 2, 609-617
Maples & Archer	1988	Palaois 3, 95-103
Legendre & Legendre	1998	Numerical ecology. Chapter 7

(2) QUANTITATIVE DATA



$$= \left(x_{ji} - x_{i1}\right)^2 + \left(x_{j2} - x_{i2}\right)^2$$

Euclidean distance

$$d_{ij} = \sqrt{\sum_{k=1}^{m} \left(x_{ik} - x_{jk}\right)^2}$$

dominated by large values

Manhattan or city-block metric

$$d_{ij} = \sum_{k=1}^{m} \left| x_{ik} - x_{jk} \right|$$

less dominated by large values

Bray & Curtis (percentage similarity)

$$d_{ij} = \frac{\sum \left| x_{ik} - x_{jk} \right|}{\sum \left(x_{ik} + x_{jk}\right)}$$

sensitive to extreme values

relates minima to average values and represents the relative influence of abundant and uncommon variables

(2) Quantitative data (cont)

Similarity ratio or Steinhaus-Marczewski coefficient (≡ Jaccard)

$$d_{ij} = \frac{\sum x_{ik} x_{jk}}{\left(\sum x_{ik}^2 + \sum x_{jk}^2 - \sum x_{ik} x_{jk}\right)^2}$$

less dominated by extremes

Chord distance for % data

$$d_{ij} = \left[\sum_{k=1}^{m} \left(\sqrt{p_{ik}} - \sqrt{p_{jk}}\right)^2\right]^{1/2}$$

"signal to noise"

(3) PERCENTAGE DATA (E.G. POLLEN, DIATOMS)

Standardised Euclidean distance	- gives all variables 'equal' weight, increases noise in data
Euclidean distance	- dominated by large values, rare variables almost no influence
Chord distance (= Euclidean distance of square-root transformed data)	- good compromise, maximises signal to noise ratio

(4) TRANSFORMATIONS OF ECOLOGICAL DATA

Normalise samples	- 'equal' weight	$\sqrt{\sum_{k=1}^{m} (x_{ik})^2}$
Normalise variables	- 'equal' weght, rare species inflated	
No transformation	- quantitiy dominated	
Double transformation	- equalise both, compromise	

Noy-Meir et al., 1975 J. Ecology 63, 779-800