Below are equations for the association metrics discussed in class. D indicates a distance and S indicates a similarity. The subscripts are those used in Legendre and Legendre 1998.

Q mode

Quantitative Data

Symmetrical

-Euclidean distance $$D_1 = \sqrt{\sum_{j=1}^{p} \left(y_{1,j} - y_{2,j}\right)^2}$$

Asymmetrical

-Bray-Curtis $$D_{14} = \frac{\sum_{j=1}^{p} |y_{1,j} - y_{2,j}|}{\sum_{j=1}^{p} (y_{1,j} + y_{2,j})}$$

-Chord distance $$D_3 = \sqrt{2\left(1 - \frac{\sum_{j=1}^{p} y_{1,j} y_{2,j}}{\sqrt{\sum_{j=1}^{p} y_{1,j}^2 \sum_{j=1}^{p} y_{2,j}^2}}\right)}$$

-Hellinger distance $$D_{17} = \sqrt{\sum_{j=1}^{p} \left(\sqrt{\frac{y_{1,j}}{y_{1+}}} - \sqrt{\frac{y_{2,j}}{y_{2+}}}\right)^2},$$

where y_{i+} is the sum of all descriptors for the ith object.

Binary Data

When conducting pairwise comparisons with binary data each comparison of a descriptor in two objects has four potential outcomes (*a*, *b*, *c*, or *d* in the table below). Binary similarity coefficients are calculated from the distribution of descriptors in each of these four outcomes.

	Object 2		
		1	0
Object 1	1	*a*	*b*
	0	*c*	*d*

, where $p = a + b + c + d$

Symmetrical

-Simple matching coefficient $$S_1 = \frac{a+d}{p}$$

Asymmetrical

-Jaccard's $$S_7 = \frac{a}{a+b+c}$$

-Sørensen's $$S_8 = \frac{2a}{2a+b+c}$$

Qualitative Data

-Gower's coefficient $S_{15} = \frac{1}{p}\sum_{j=1}^{p} s_{12j}$,

where s_{12j} =1 (agreement) or 0 (disagreement) for binary and qualitative descriptors. For quantitative descriptors, s_{12j} is:

$$s_{12j} = 1 - \frac{|y_{1,j} - y_{2,j}|}{R_j}$$

where R_j is the largest difference ound for this descriptor across all objects in the study.

R mode

Quantitative & Qualitative Data

Symmetrical

-covariance $s_{jk} = \frac{1}{n}\sum_{i=1}^{n}(y_{i,j} - \bar{y}_j)(y_{i,k} - \bar{y}_k)$

-Pearson's correlation (covariance of standardized variables)

$$r_{jk} = \frac{\sum_{i=1}^{n}(y_{i,j} - \bar{y}_j)(y_{i,k} - \bar{y}_k)}{\sqrt{\sum_{i=1}^{n}(y_{i,j} - \bar{y}_j)^2 \sum_{i=1}^{n}(y_{i,k} - \bar{y}_k)^2}}$$

-Chi-square distance $D_{16} = \sqrt{y_{++}}\sqrt{\sum_{i=1}^{n}\frac{1}{y_{i+}}\left(\frac{y_{i,j}}{y_{j+}} - \frac{y_{i,k}}{y_{k+}}\right)^2}$

Binary – Jaccard and Sørensen's