# Correlation between variables

One context in which resemblances between variables is often of primary interest is in dealing with environmental variables, biomarkers, morphology etc. Concepts of ignoring joint absences do not apply – in fact zero no longer necessarily means absence (e.g. $0^\circ$C), particularly after normalisation (see Section 4). Variables are usually on different measurement scales (or are non-comparable on the same units), so correlation is a natural choice, with its built-in normalisation. The final option in Measure•Others is ✓Correlation, with seven variations of a correlation coefficient, $\rho$, namely

$\rho ^P = \frac{ \sum_j \left( y_{1j}-y_{1\bullet} \right)\left( y_{2j}-y_{2\bullet} \right) }{ \sqrt{ \sum_j \left( y_{1j}-y_{1\bullet} \right)^2 \sum_j \left( y_{2j}-y_{2\bullet} \right)^2}} \text{ \hspace{14mm} Pearson (product-moment) correlation,} $

where $y_{1\bullet} = \left( \sum_j y_{1j} \right)/n$ is the average of the $n$ sample readings for variable 1, etc., and two non-parametric choices, based only on rank values ($r_{ij}$), the numbers 1, 2, 3, .., *n* across samples $j$, for each variable $i$. Spearman is simply Pearson correlation calculated on the ranks, reducing to:

$\rho^S = 1 - \frac{6}{n \left(n^2-1 \right) } \sum_j \left( r_{1j} - r_{2j} \right)^2 \text{ \hspace{10mm} Spearman rank correlation,} $

and Kendall's $\tau$ is an alternative (Kendall MG 1970, *Rank correlation methods*, Griffin, London), which in practice tracks Spearman closely, but with lower absolute values. These three coefficients are then given as absolute values, $| \rho |$, to cater for situations where it is not especially meaningful to distinguish between positive and negative correlations (e.g. some biomarkers increase under impact and some decrease, so an absolute $\rho$ is often a better description of their inter-relationship). A final weighted form of Spearman gives more emphasis to small ranks (high variable values):

$\rho^W = 1 - \frac{6}{n \left(n-1 \right) } \sum_j \frac{ \left( r_{1j} - r_{2j} \right)^2 }{ r_{1j} + r_{2j} } \text{ \hspace{16mm} Weighted Spearman rank correlation,} $

but this really only makes sense in an asymmetric context, such as correlating the entries of two resemblance matrices, thus emphasising matching pairs of high similarities – see the discussion of equation (11.4) in CiMC.

---

Revision #17
Created 11 June 2024 23:27:45 by Arden
Updated 21 January 2025 21:37:37 by Abby Miller