Measures of multivariate association using distances

Carles M. Cuadras[1] **and Daniel Cuadras**[2]

[1] *University of Barcelona, Spain*
[2] *Catalan Oncology Institute, Spain*

Abstract

Several coefficients have been proposed in measuring the multivariate association between two data sets taken on the same individuals. Ecology is a clear example, where environmental data is related to species. In genomic data we may seek relations between genotype (e.g., DNA data) and phenotypes of interest. There are also examples in biometry and psychology. Often, the data sets are represented by two quantitative matrices, where the rows are multivariate observations taken on the same individuals. Then some dependence measures (Wilks, Hotelling, Pillai, Cramer-Nicewander, etc.) based on canonical correlations can be used. However, if the two data sets are non quantitative (binary, categorical, nominal), the information can alternatively be given by a similarity or distance matrix. Then we must relate two distance matrices, and some coefficients are proposed by correlating principal coordinates. When the first data set is quantitative and univariate, these measures reduce to the squared multiple correlation coefficient.

Keywords

Multivariate statistics, Dependence measures, Canonical correlation, Distance analysis, Multidimensional scaling.