

Statistical scaling of categorical data

Henning Lauter, Ayad Ramadan

University of Potsdam, Institute of Mathematics, Germany

Am Neuen Palais 10, D-14469 Potsdam

e-mail: laeuter@uni-potsdam.de, ayad_math@yahoo.com

Key words: Multivariate scaling, discrimination, power of multivariate tests

Estimation and testing of distributions in metric spaces are well known. R.A. Fisher, J. Neyman, W. Cochran and M. Bartlett achieved essential results to the statistical analysis of categorical data. In the last 40 years many other statisticians found important results to this topic.

Often data sets contain categorical data, e.g. levels of factors or names. There does not exist any ordering or any distance between these categories. But for any level there are measured some metric or categorical values. We define empirical probabilities for the original observations and find a class of distributions in a metric space where these empirical probabilities can be found as approximations for equivalently defined probabilities. With this method we identify probabilities connected with the categorical data and probabilities in metric spaces. Here we get a mapping from the levels of factors or names into points of a metric space. This mapping yields the scale for the categorical data.

From the statistical point of view we use multivariate statistical methods, we calculate maximum likelihood estimations and compare different approaches for scaling.