

Quantitative analysis, power transformations and categorical data visualization

Ting-Wu Wang

School of Information and Physical Sciences, University of Newcastle

Abstract

A common difficulty when analysing categorical variables of a contingency table is that cell frequencies are often prone to overdispersion as they are typically assumed to be Poisson random variables. Transformation of data has been a popular and successful technique in statistics to overcome the issues introduced due to the equivalence feature between the expected value and the variance. A power transformation of the frequencies of the contingency table is therefore a strategy that can be used to help overcome this issue in the application of correspondence analysis. However, currently there is little research on this matter. The traditional approach to correspondence analysis and the method of reciprocal averaging has close links in the analysis of the association between categorical variables. This talk illustrates the connections between correspondence analysis and the method of reciprocal averaging in the context of power transformations. Applications are also used to demonstrate the algorithms between how the two methods acquire solutions that are a weighted average of one another. In the calculation of the scores and the association between them, two alternative ways of choosing an appropriate power parameter are considered. One approach calculates the power parameter that explains the optimal association as a percentage of the phi-squared statistic from the first dimension of the solution. The second method involves determining a power transformation, such that the data becomes non-dispersed.

Keywords: Correspondence analysis, Reciprocal averaging, Categorical data visualisation