Correspondence analysis related biplot visualisations to aid analyses of categorical data containing missing values

Sugnet Lubbe; Niël le Roux; Johané Nienkemper-Swanepoel*; Carel van der Merwe

Department of Statistics and Actuarial Science, Stellenbosch University, South Africa

slubbe@sun.ac.za; njlr@sun.ac.za; nienkemperj@sun.ac.za*; cjvdmerwe@sun.ac.za

Abstract:

Visualising incomplete data enables the recognition of response patterns and the evaluation of the effect of the unobserved information on the interpretable information. Multiple imputation is the preferred method for imputing missing values but this poses a real challenge for visualisation of multiple completed data sets. The GPAbin procedure has been developed to combine multiple correspondence analysis (MCA) biplots, Procrustes analysis and Rubin's rules in creating a single display for nominal categorical data with missing values. Instead of imputation, subset MCA can be used to visualise the observed and missing values together or separately. In addition, the procedure also allows for investigating the underlying missing data mechanism. In this presentation these ideas will be explored and illustrated with several novel enhancements to R plotting functions.

Keywords:
biplots; GPAbin; missing data; multiple correspondence analysis; Shiny.