

Visualizing characteristics of aggregated symbolic data

Junji Nakano¹; Yoshikazu Yamamoto²; Nobuo Shimizu³

- ¹ Chuo University, Japan
- ² Tokushima Bunri University, Japan
- ³ The Institute of Statistical Mathematics, Japan

Abstract:

We sometimes are interested in comparing groups of data, not in each individual observation, especially when we have huge amount of data. Aggregated symbolic data (ASD) expresses a group of observations which have continuous and categorical variables by using up to second moments of variables. ASD for a group of data is equivalent to the set of means, variances, and correlations for continuous variables, Burt matrix for categorical variables, and means of a continuous variable for one value of a categorical variable. As ASD for many categorical variables is still complicated, we define simpler measures of location and dispersion for a categorical variable to compare ASDs constructed by same variables, and a measure of correlation between two categorical and/or continuous variables. We use these measures to summarise each ASD and visualise them using an extended parallel coordinate plot in order to compare several ASDs at the same time.

Keywords:

Categorical variable; Continuous variable; Measure of correlation; Measure of dispersion; Measure of location