Enhancement of boxplot chrarcters for model diagnostic of block maximum extremal events

ABSTRACT

A boxplot is an exploratory data analysis (EDA) tool for a compact visual display of a distributional summary of a univariate data set. It is designed to capture all typical observations and displays the location, spread, skewness and the tail of the data. The precision of some of this functionality is considered to be more reliable for symmetric data type and thus less appropriate for skewed data such as the extreme data. Many observations from extreme data were mistakenly marked as outliers by the Tukey's standard boxplot. A new boxplot implementation is presented which adopts a fence definition using the extent of skewness and enhances the plot with additional features such as a quantile region for the parameters of generalized extreme value (GEV) distribution in fitting an extreme data set. The advantage of the new superimposed region was illustrated in term of batch comparison of extreme samples and an EDA tool to determine search region or direction as contained in the optimisation routines of a maximum likelihood parameter estimation of GEV model. A simulated and real-life data were used to justify the advantages of the boxplot enhancement.

Keyword: Boxplot; Enhancement; Generalized extreme value