Modified boxplot for extreme data

ABSTRACT

A boxplot is an exploratory data analysis (EDA) tool for a compact distributional summary of a data set. It is designed to captures 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 erroneously marked as outliers by the Tukeys standard boxplot. We proposed a modified boxplot fence adjustment using the Bowley coefficient, a robust skewness measure. The adjustment will enable us to detect inconsistent observations without any parametric assumption about the distribution of the data. The new boxplot is capable of displaying some additional features such as the location parameter region of the Gumbel fitted extreme data. A simulated and real life data were used to show the advantages of this development over those found in the literature.

Keyword: Boxplot; Extreme data