AN IMPROVED BOXPLOT FOR SKEWED DISTRIBUTIONS

Mia Hubert[1] and Ellen Vandervieren[2]

[1] **Katholieke Universiteit Leuven, Department of Mathematics, W. de Croylaan 54, B-3001 Leuven, Belgium**

[2] **University of Antwerp, Department of Mathematics and Computer Science, Middelheimlaan 1, B-2020 Antwerp, Belgium**

E-mail:

[1] mia.hubert@wis.kuleuven.ac.be

[2] ellen.vandervieren@ua.ac.be

The boxplot is a very popular graphical tool to visualize the distribution of continuous univariate data. First of all, it shows information about the location and the spread of the data by means of the median and the interquartile range. The length of the whiskers on both sides of the box and the position of the median within the box are helpful to detect possible skewness in the data. Finally, observations that fall outside the whiskers are pinpointed as outliers, hence the boxplot also includes information from the tails. However, when the data are skewed, usually too many points are classified as outliers. This is because the outlier rule is solely based on measures of location and scale, and the cutoff values are derived from the normal distribution. We present a generalization of the boxplot that includes a robust measure of skewness in the determination of the whiskers. We show with several simulation results that this adjusted boxplot gives a more accurate representation of the data and of possible outliers.

References

[1] Brys G., Hubert M., Struyf A. (2003). *A robust measure of skewness*. Journal of Computational and Graphical Statistics, to appear.

[2] Hoaglin D.C., Mosteller F., Tukey J.W. (1983). *Understanding Robust and Exploratory Data Analysis*. Wiley, New York.

[3] Hoaglin D.C., Mosteller F., Tukey J.W. (1985). *Exploring Data Tables, Trends and Shapes*. Wiley, New York.

[4] Tukey, J.W. (1977). *Exploratory Data Analysis*. Reading, MA: Addison-Wesley.