VISUALIZING ANOMALIES IN CIRCULAR DATA

Davide Buttarazzi [1] and Giovanni C Porzio[2]

[1] Sample Survey and Data Analysis Lab, University of Cassino and Southern Lazio,
(e-mail: `davidebuttarazzi@outlook.com`)

[2] Department of Economics and Law, University of Cassino and Southern Lazio,
(e-mail: `porzio@unicas.it`)

ABSTRACT: Anomaly detection has a long history in Statistics, with one of the most effective approaches being robustness. First, a model describing the majority of the data is assumed. Second, its parameters are robustly estimated. Then, the distance of all the points from such a model is evaluated. Eventually, extremely far (i.e., unlikely) observations are flagged as outliers. Visually, this procedure is well described by the well-worn Tukey's box-and-whisker plot. Thanks to its robustness properties, it is probably the graphical tool mostly adopted to highlight anomalies in univariate data sets.

This work aims at investigating if the same strategy can be exploited in circular data analysis, i.e., for data lying on the boundary of the unit circle. For this kind of data, a specific boxplot has been designed. However, its first formulation did not focus on anomaly detection. It was rather conceived as an exploratory tool to display the main features of a circular data set. Reliyng on a non-robust estimate of the data dispersion, it will be simply misleading if used to visualize anomalies. A robust circular boxplot is then introduced. It will be able to correctly identify circular outliers under a specific parametric model.

KEYWORDS: Circular boxplot, directional statistics, von Mises distribution.