

Visualizing the Effects of Data Transformations on Errors

Robert M Flight and Hunter NB Moseley

Department of Molecular and Cellular Biochemistry

Markey Cancer Center

Resource Center for Stable Isotope-Resolved Metabolomics

University of Kentucky

In many omics analyses, a primary step in the data analysis is the application of a transformation to the data. Transformations are generally employed to convert proportional error (variance) to additive error, which most statistical methods appropriately handle. However, omics data frequently contain error sources that result in both additive and proportional errors. To our knowledge, there has not been a systematic study on detecting the presence of proportional error in omics data, or the effect of transformations on the error structure. In this work, we demonstrate a set of three simple graphs which facilitate the detection of proportional and mixed error in omics data when multiple replicates are available. The three graphs illustrate proportional and mixed error in a visually compelling manner that is both straight-forward to recognize and to communicate. The graphs plot the 1) absolute range, 2) standard deviation and 3) relative standard deviation against the mean signal across replicates. In addition to showing the presence of different types of error, these graphs readily demonstrate the effect of various transformations on the error structure as well. Using these graphical summaries we find that the log-transform is the most effective method of the common methods employed for removing proportional error.