Summary of Graphical Tests for Power Comparison of Competing Designs

Aarathy Babu

Exploratory graphics are quick to produce and flexible , however it is unlikely that the discovered information is conveyed to the audience efficiently. This is where visual inference comes to play.

How does the line up work?

Line up protocol is similar to the police line up. Data plots are considered to be test statistics and compared to the data plots generated from the null hypothesis using a line up. The null hypothesis emphasizes there is no structure visible among the plots or rather there is nothing that stands out.

If null hypothesis is true, then the randomly placed plot of data will not stand out from the null data plot. If the viewer identifies the data plot, it lends statistical significance to the conclusion that the data is not consistent with the null data. . If an observer is able to identify the data plot from the lineup, we reject the null hypothesis.

What is the power of a test?

When comparing different tests of the same quantity, we consider that test better if it has greater power. . The power of a test is the probability to reject the null hypothesis, irrespective of whether it is true or false – in a lineup this is the probability that an observer identifies the plot of the real data.

n = number of independent observers.

m = number of plots

xi = the number of observers who picked plot i from m plots

The power of a lineup can therefore be estimated as the ratio of correct identifications x out of n viewings.

Comparisons of Designs

- Create lineup data : assuming two variables X and Y. Data for lineup of size m is created by m-1 permutations of Y or drawing m-1 samples of size n from the null distribution. The original data is added to the line up randomly. nullabor package can be used for creating line up data.
- Create lineup from competing designs
- Evaluate Lineups : Present the lineups to independent observers and assess the signal strength and time taken by observers to come to a decision.
- Evaluate Competing designs : differences in strength and time taken is due to the differences in the design.

Comparing power of competing designs therefore involves comparing percentages of correct responses

Lineups provide a powerful tool for evaluating different designs in the framework of the data. In order to accomodate individual differences in visual ability, a generalized linear mixed model for response values. Lineups can be investigated for evaluating competing plot designs based on their power modeled by a logistic regression with subject-specific random intercepts.