

# Summary of Graphical inference for Infovis

Aarathy Babu

## What is inference?

The goal of statistical methods is to draw conclusions about the population that the data sample came from. Graphical inference helps us to see if what we see from the conclusions drawn, truly present. It allows us to discover new findings while controlling the tendency to see pattern in noise.

Traditional statistical tests do not cover all of the complexities that arise when exploring data. The benefit of visual inference is that it can be used in complex data analysis settings that do not have corresponding numerical tests.

## Type I and II errors

The two types of mistakes we can make in our decision are,

- type I error : falsely convict an innocent dataset (false positive)
- type II error : Acquitt a guilty dataset (false negative)

## Protocols for Graphical Inference

- Rorschach protocol

We use this protocol to calibrate our vision to the natural variability in plots in which the data is generated from scenarios consistent with the null hypothesis.

To keep the analyst on their toes and avoid the complacency that may arise if they know all plots are null plots the administrator might slip in a plot of the real data.

- Line Up protocol

A set of null data sets is created and plots of these decoys along with the plot of true data is randomly positioned and shown to an impartial observer. The observer is then asked to spot the real data. If the observer can pick out a plot as being noticeably different then we could reject the null hypothesis. If the observer cannot do so, rather than accepting the alternative we fail to reject the null.

## How to generate null datasets from the null hypothesis

- Resampling
- Simulation

This method is used when we are interested in a more specific set of hypothesis. For example, if the factors in question have a linear or exponential relationship. In those cases we have a probabilistic model and we can generate null data sets by sampling from the distribution implied by the model.

## Use

- Nullabor The user specifies how many decoy plots to create, and a mechanism (permutation ,simulation) to generate null datasets. For the line-up, nullabor generates the decoys, appends them to the real data set, and randomly chooses a position for the accused.

Graphical inference is important because it helps us to avoid (or at least calibrate the rate of) false convictions