

REVISITING HISTOGRAM

Demudu Naganaidu (GS49320)
Proposal Seminar

15th May 2018

Abstract

Histogram is a key tool used in Exploratory Data Analysis and is one of the oldest tools in graphical displays in understanding the distribution of a raw data. The histogram is helpful in summarizing continuous data or at times discrete data by grouping the observations into bins. The shape of the histogram would be decided by the number of bins or the bin width. Several methods are available in deciding the number of bins. Several methods available are the Sturges rule, Scotts rule or the Doannes rule and Freedman and Diaconis. The result from each rule varies and none is convincing. This study aims to use this mean and variance of raw data as guide to evaluate the binning methods that suitable for normal and skewed data. The result of these studies used to select the appropriate rule for binning histogram for given set of data. Current method of histogram construction modified to detect outliers.

Keywords: Histogram, Binning, Exploratory data, Frequency tables, Grouped data