

Histogram and skewness

A histogram is a graphical representation of the distribution of the numerical data. It is an estimate of the probability distribution of a continuous variable and is used to visualize the shape, central tendency, and variability of a data set.

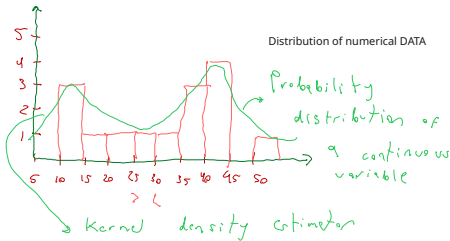
ages = {11, 12, 14, 18, 24, 26, 30, 35, 36, 37, 40, 41, 42, 43, 50} \Rightarrow Histogram

0-50

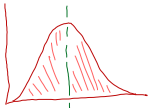
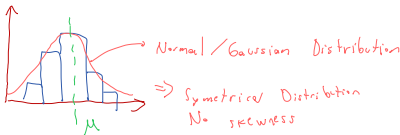
1) No of bins = 10

$50 / 10 = 5 \rightarrow$ bin size

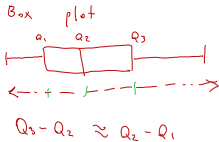
Bins \rightarrow [0-5, 5-10, 10-15, 15-20, 20-25, 25-30, ..., 40-45, 45-50]



Skewness

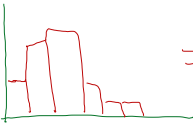


The mean, median, and mode are all perfectly at the center

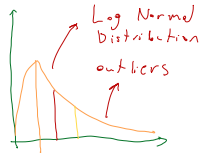


Mean = Median = Mode

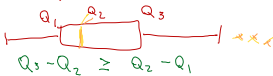
Right Skewed



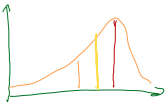
\Rightarrow Positive skewed \Rightarrow



Box plot



Left Skewed distribution



\Rightarrow Negative skewed

Relationship

$mean \leq median \leq mode$

Box plot

