Introduction to Data Analytics

Lecture: Descriptive Statistics: Summary Statistics: Measures of Central Tendency

NPTEL MOOC

By

Prof. Nandan Sudarsanam, DoMS, IIT-M and

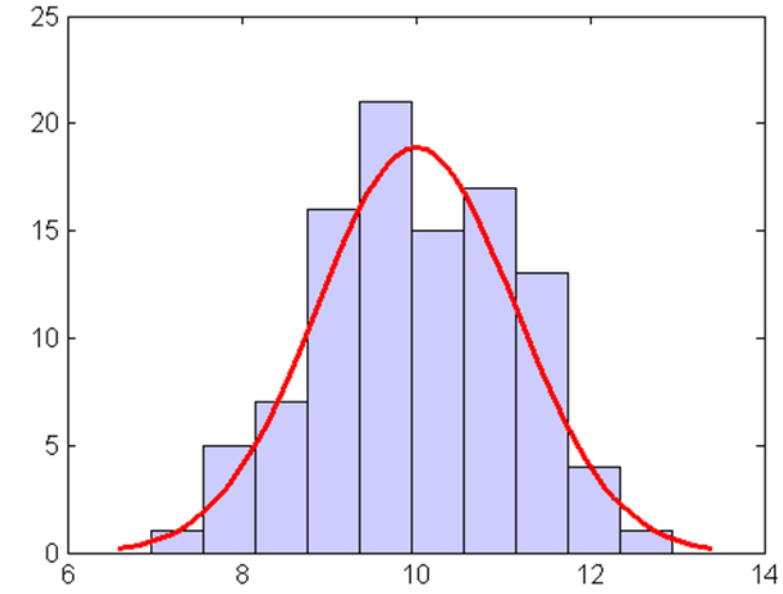
Prof. B. Ravindran, CS&E, IIT-M

Summarizing Data through numbers

- Measures of Central Tendency
- Dispersion
- Skew and Kurtosis

Data Set
10.04
9.31
11.15
11.22
10.19
10.49
8.38
10.32
8.14
7.89
10.07
10.42
11.55
9.63
9.05
8.96
12.57
.
.
.

Histogram



Measures of Central Tendency

- Data Set: 3,4,3,1,2,3,9,5,6,7,4,8
- Mean

$$\frac{3+4+3+1+2+3+9+5+6+7+8+4}{12} = 4.583 \quad \frac{x_1+x_2+x_3+\cdots}{n} \quad \text{or} \quad \frac{\sum_{i=1}^{n} x_i}{n}$$

- Median

1,2,3,3,3,4,4,5,6,7,8,9 Hence Answer = 4

- Mode

The value 3 appears 3 times, and 4 appears 2 times and all other values appear once. Hence 3 is the mode

Measures of Central Tendency

- Where do we want to use Mean, Median and Mode
- Choosing between mean and median
  - Bad outliers
    - Errors
    - Do not provide a realistic picture of the story
  - Good outliers
    - The story is in the outliers
- Mode
  - Useful with nominal variables
  - Multi modal distributions