Introduction to Data Analytics

Lecture: Descriptive Statistics: Summary Statistics: Measures of Dispersion

NPTEL MOOC

By

Prof. Nandan Sudarsanam, DoMS, IIT-M and

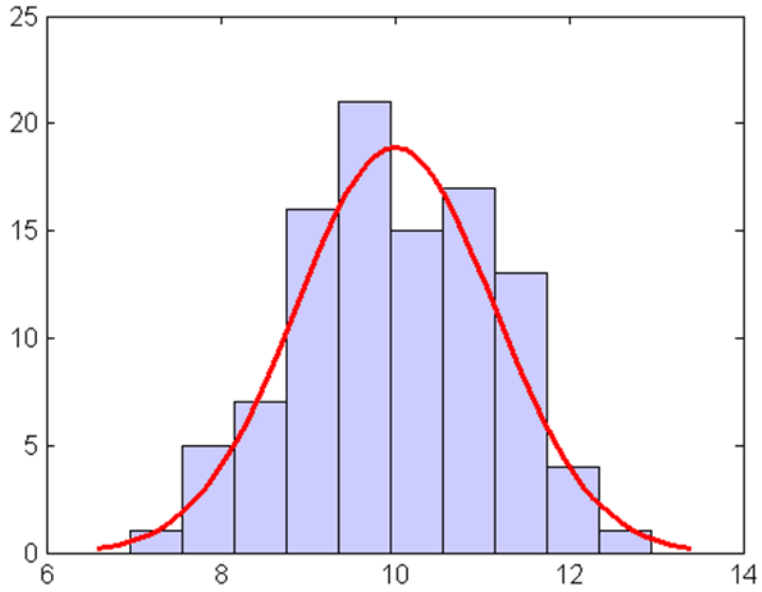
Prof. B. Ravindran, CS&E, IIT-M

Summarizing Data through numbers

- Measures of Dispersion

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 Dispersion

- Data set: 3,4,3,1,2,3,9,5,6,7,4,8
- Range (Max-Min) (9-1 = 8)
- Inter Quartile Range: 3rd quartile -1st quartile (75th Percentile – 25th Percentile) (6.5 – 3 = 3.5)
- Sample Standard deviation

$$\sqrt{\frac{1}{N-1}\sum_{i=1}^{N}(x_i-\bar{x})^2} = \frac{1}{12-1}\sum((3-4.58)^2+(4-4.58)^2\dots)$$

Measures of Dispersion

- Questions that go with Standard deviation
  - Why do we use the square function on the deviations? What are its implications?
  - Why do we work on standard deviation and not the variance?
  - Why do we average by dividing by N-1 and not N?

- Mean absolute Deviation and its variants
  - Use $|x_i - \bar{x}|$ instead of $(x_i - \bar{x})^2$