Statistics

Definition -

A branch of applied maths that involves the collection description, presentation, analysis and Interepretation of numerical data. is called statistics.

Type of Data -

① structured data

② unstructured data

Stages of stats -

① collection of data

② Organizing of data

③ Presentation of data

④ Analysis of data

⑤ Interepretation of data

Type of stats -

① Descriptive stats

Available data sample or population on it we perform action like analysing. discribe, summerize. it called descriptive stats.

② Inferential stats -

On describe data we perform interepretation like hypothesis testing on the data for example Z-test, T-test, F-test, chi-square test. is called inferential statistics

① Descriptive stats.

① univariate Des. stats.

② Bivariate Des. stats.

③ multi variate Des. stats.

Descriptive stats -

1. measure of Center tendency
2. Measure of Dispersion or variation
3. measure of position
4. measure of shape

Measure of Center Tendency

1. mean -

[2, 3, 4, 5, 6]

$$\text{mean} = \frac{2+3+4+5+6}{5}$$

$$= 4$$

population mean = μ

Sample mean = $\bar{x}$

② median –

I-case $[6,2,4,5,3]$

data sort

$$[2,3,\underset{\uparrow}{4},5,6]$$

median = 4

II-case

$$[2,3,5,\underbrace{7,8},9,11,13]$$

$$\text{median} = \frac{7+8}{2} \Rightarrow 7.5$$

③ mode –

① uni-modal

(ii) Bi-modal

(iii) multi-modal

① [2, 3, 4, 5, 5, 6]

mode = 5

② [2, 2, 4, 5, 6, 6]

mode = 2, 6

③ [2, 2, 4, 4, 5, 5, 6, 7, 8]

mode = 2, 4, 5

→ main application

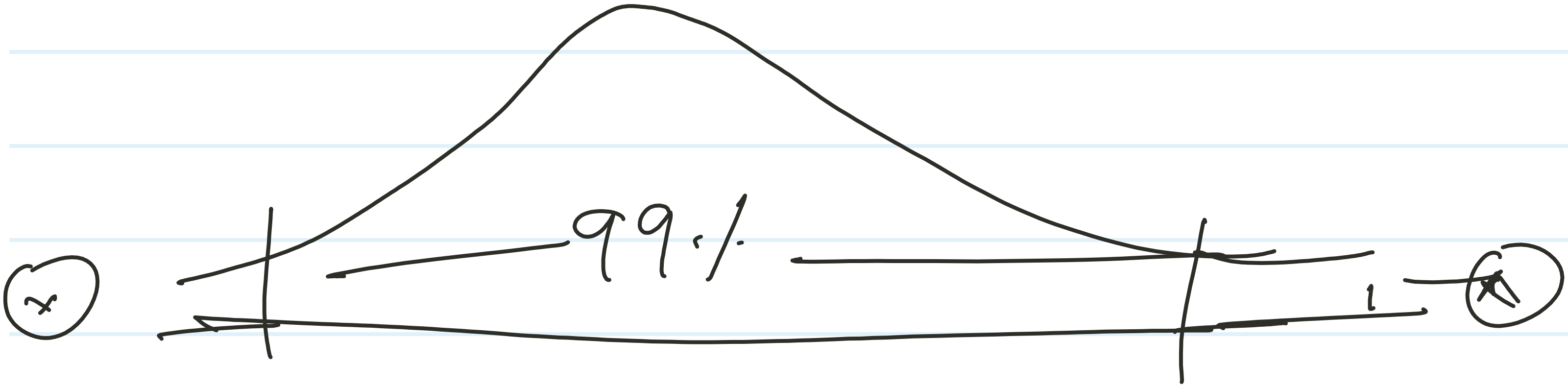
→ outlier handeling

→ missing valu handel

X	Y
2	2
3	3
✓	4
7	45✓
✓	7
11	5
13	

$$2,3,\underbrace{4,5}_{4.5},7,45$$

$$\underline{2,3,4,5,7,4.5}$$



Result

P

F

P

✓

F

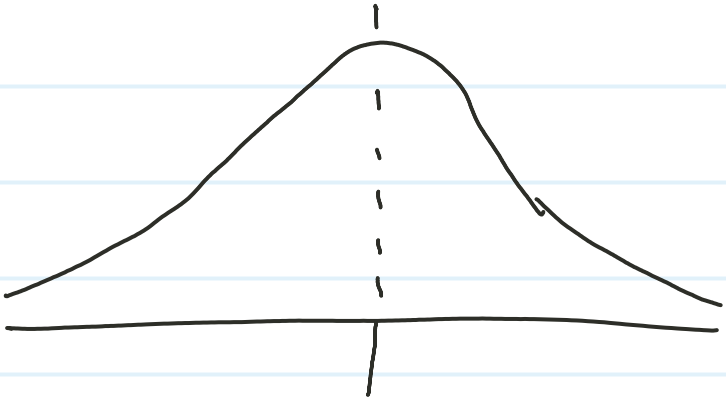
P

✓

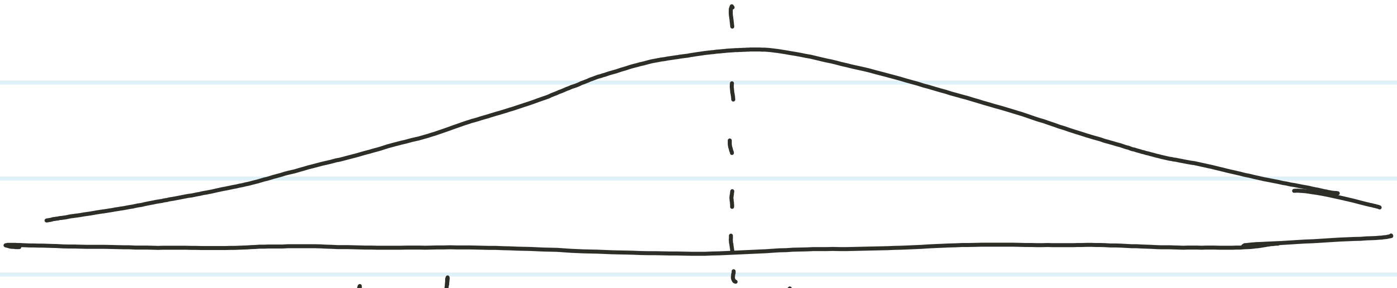
F

P

measure of Dispersion or variance



— low variability



High variability

① mean absolute deviation-

The mean absolute devi. of a dataset is the avg. distance b/w each data point and the mean.

$$\Rightarrow \frac{1}{N} \sum_{i=1}^{n} |x_i - \bar{x}|$$

Data - [10, 15, 15, 17, 18, 21]

$$\text{mean} = \frac{96}{6} = 16$$

$$|10-16| = |-6| = 6$$
$$|15-16| = 1$$
$$|15-16| = 1$$
$$|17-16| = 1$$
$$|18-16| = 2$$
$$|21-16| = 5$$
$$\overline{16}$$

$$\Rightarrow \frac{16}{6} \Rightarrow 2.67$$

② Variance -

It tells the degree of spread in dataset.

High variability - datapoint spread widely low variability - data point close to mean.

population $\sigma^2 = \frac{1}{N}\sum_{i=1}^{n}(x_i-\mu)^2$

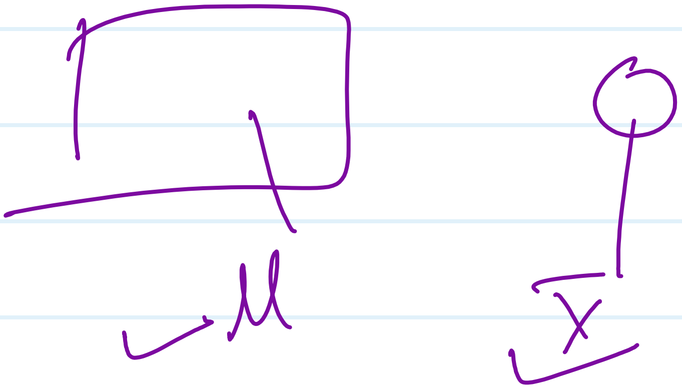
Sample $s^2 = \frac{1}{n-1}\sum_{i=1}^{n}(x_i-\bar{x})^2$

Note- n-1 is a degree of freedom

OR

Besils correction

For keep away result from biased.



③ standard deviation -

The square root of variance is called st. dev. .

The farthes the data points from the higher the deviation

Pop $\sigma = \sqrt{\frac{1}{N}\sum_{i=1}(x_i-\mu)^2}$

Sample $S = \sqrt{\frac{1}{n-1}\sum_{i=1}(X_i-\bar{X})^2}$

③ Range -

[1, 2, 5, 6, 11, 15, 19, 25, 30]

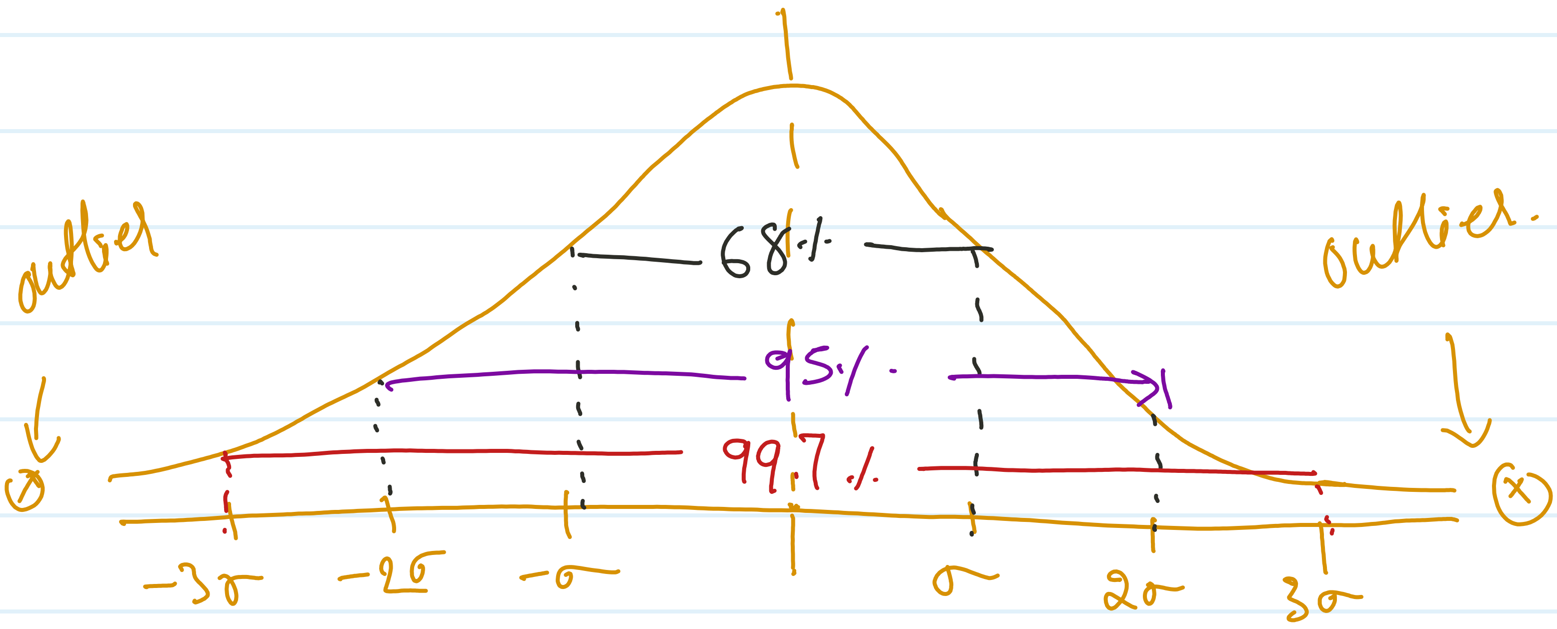
max - 30

min - 1

Range = max - min

= 30 - 1 = 29

Emperical Rule



m = m = m