

1) Types of Data

→ Numerical →

There are two type of Numerical Data Discrete and Continuous

* Discrete :- This is type of Data where Data is in integer form

for example :- 1, 2, 3, etc (in counting)

* Continuous :- This is type of Data where Data is in continuous form

for example :- 1.1, 1.2, 1.3
used in measuring

• Categorical →

There are two type of categorical Data Cardinal & ordinal

* Cardinal :-

This is a type of Data where Data has no order

for example :- Man, Woman, child

* ordinal:-

This is a type of Data where Data has order

for example:- low, medium, high

• Type of Statistics

→ * Descriptive Statistics:-

Used to summarise and Describe data in form of Average, mean, chart, etc

* Inferential Statistics:-

Used to make prediction about a larger population using sample dataset

• What is Descriptive Statistics?

→ Descriptive statistics involves summarizing and Describing data and organizing data using

Mean, Median, Mode
Range, Variance, Std. Dev
Table, charts, Graphs

2] Explain the difference between:

- Mean :- Average Value of all data
- Median :- Middle Value of all data
- Mode :- Most frequent Value ~~in~~ all data
- Range :- Max Value - Min Value in dataset
- Variance :- Average of squared deviation from mean

for example $\rightarrow 3 + 1 + 5 + 7 + 9 + 10 = 35$

Dota	$x - \text{mean}$	$(x - \text{mean})^2$
1	-6	36
5	-2	4
7	0	0
9	2	4
10	3	9
3	-4	16

$$\frac{36 + 4 + 0 + 4 + 9 + 16}{6} = 11.5$$

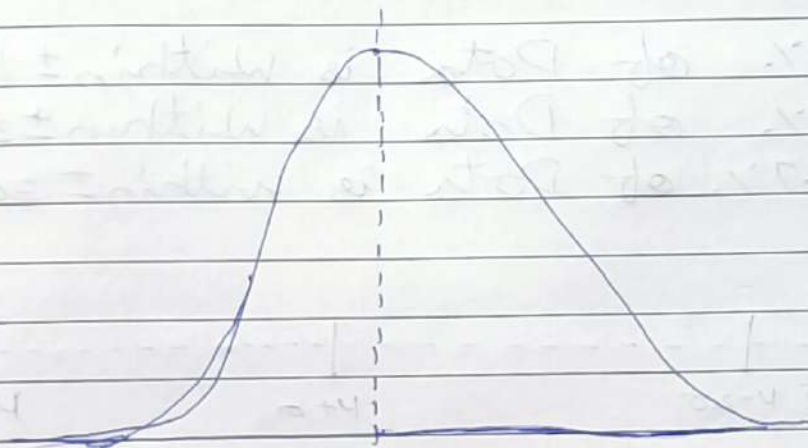
$$V_{\text{variance}} = 11.9$$

Standard Deviation $SD = \sqrt{\text{Variance}}$

3] Explain the following term with neat and clear diagram along with its formula:

o Gaussian Distribution:-

It is a symmetric, bell shape curve where mean, median, mode are equal to each other.



formula

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}$$

o Log Normal Distribution:-

It is a method to correct skewness

formula :-

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(\ln x - \mu)^2}{2\sigma^2}}$$

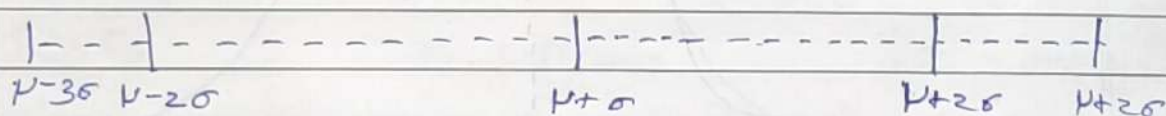
o 3-Sigma Rule or Empirical Rule

Its a Distribution rule

68% of Data is within $\pm 1\sigma$

95% of Data is within $\pm 2\sigma$

99.7% of Data is within $\pm 3\sigma$



o Percentiles

Percentiles is the thing that tell you at what percentage your ~~data~~ data is in rank in your Dataset

o quartiles

quartiles divide data into 3 parts

1st quartile = 0 to 25 Percentile

2nd quartile = 25 to 75 Percentile

3rd quartile = 75 to 100 Percentile

o Five number Summary

Minimum - The smallest Value

Q1 - 25% of data lies below this

Median - The middle Value (50% Percentile)

Q3 - 75% of data lies below this

Maximum - the largest Value

$$\text{Min} | \text{---} [Q1] === | \text{Median} | === [Q3] \text{---} | \text{Max}$$

o Skewness

It measures asymmetry of distribution

* Right Skew

$$\text{mean} > \text{median}$$

* Left skew

$$\text{mean} < \text{median}$$

o Kurtosis

At height of Peak KDE Plot

* High Kurtosis

* Normal Kurtosis

* Low Kurtosis