

Poeth Mongroliya

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 Describ^ee 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 :

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

$$\text{for example} = 3 + 1 + 5 + 7 + 9 + 10 = \underline{\underline{36}} \quad 6$$

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

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

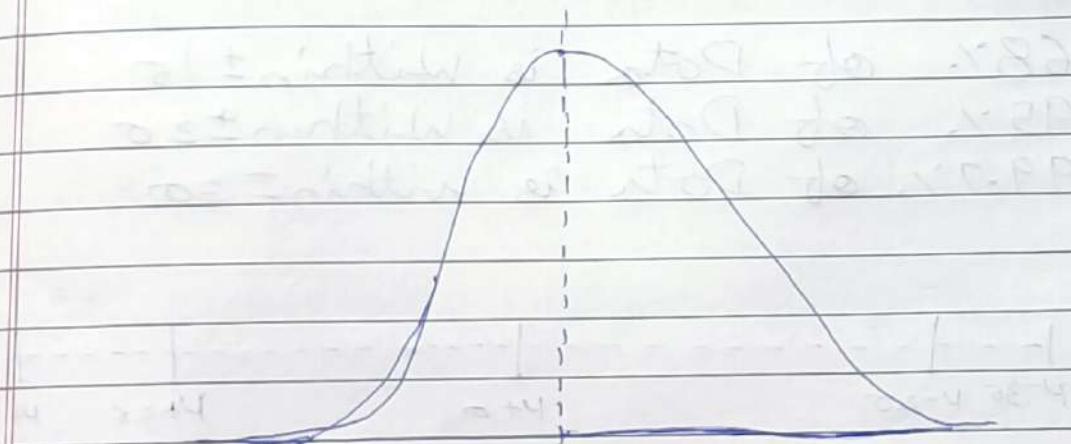
$$\text{Variance} = 11.5$$

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

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

- o Gaussian Distribution:-

It's 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 :-

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

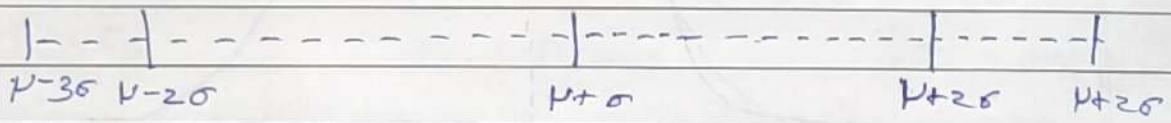
- o 3-Sigma Rule or Empirical Rule

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

$\text{MIN} \dots [Q1] == |\text{Median}| == [Q3] \dots \text{MAX}$

o Skewness

It measure asymmetry of distribution

* Right Skew

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

* Left Skew

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

- o Kurtosis

At right of Peak KDE Plot

- * High kurtosis
- * Normal kurtosis
- * Low kurtosis