Project -1

Part A : Theory & Definitions

(Q.1) Define the following with examples from the dataset

(A) Types of data : Numerical & Categorical

Ans. According to the dataset Numerical data's are as follows.

(1) Numerical Datas : Measurable quantities which are countable in numbers.

There are 2 types of Numerical Data :

- Discrete → Countable Numbers
- Continuous → Measurable quantities with decimals.

→ According to my dataset Numericals data's are as follows :

(1). Age of household Head → Discrete
(2). Household Income → Discrete
(3). Family Size → Discrete
(4) Height → Continuous

(2). Categorical Data → Descritive labels or names

There are 2 types of Categorical Data.

- Nominal → No specific order.
- Ordinal → Ordered Categories.

As per my dataset some example of Categorical data are:

(1). Household_ID
(2). Education_Level
(3). Owns_House
(4). Urban_Rural

(B). Types of Statistics : Descritive vs. Inferential

(1). Descriptive Statistics → It describes about the data, it summarises the data for eg: mean, median, mode
→ Measure of central tendency
Measure of Dispersion.

(2). Inferential Statistics → It provides us the conclusions from the data and we can all use it for making predictions.
eg. of Inferential are :- Linear Regression
logistic Regression etc.

(C). What is Descriptive Statistics ?

→ ~~It provides~~ Presenting the data in a summarizing organising and presenting raw data in a meaningful ways is known as descriptive Statistics.

(Q) Explain the difference between :

(A) Mean, Median, Mode

Mean = The mean is the sum of all values divided by the number of values.

Median = The median formula - $\frac{\text{Sum of all values}}{\text{Numbers of values}}$

Median → The middle value of a sorted list

Mode → The mode is the value that appears most frequently.

(B). Range, Variance, Standard Deviation

Range → The difference between the maximum and minimum values is known as Range. (Range = maximum − minimum).

Variance → It shows how far each values in the dataset is far from the mean.

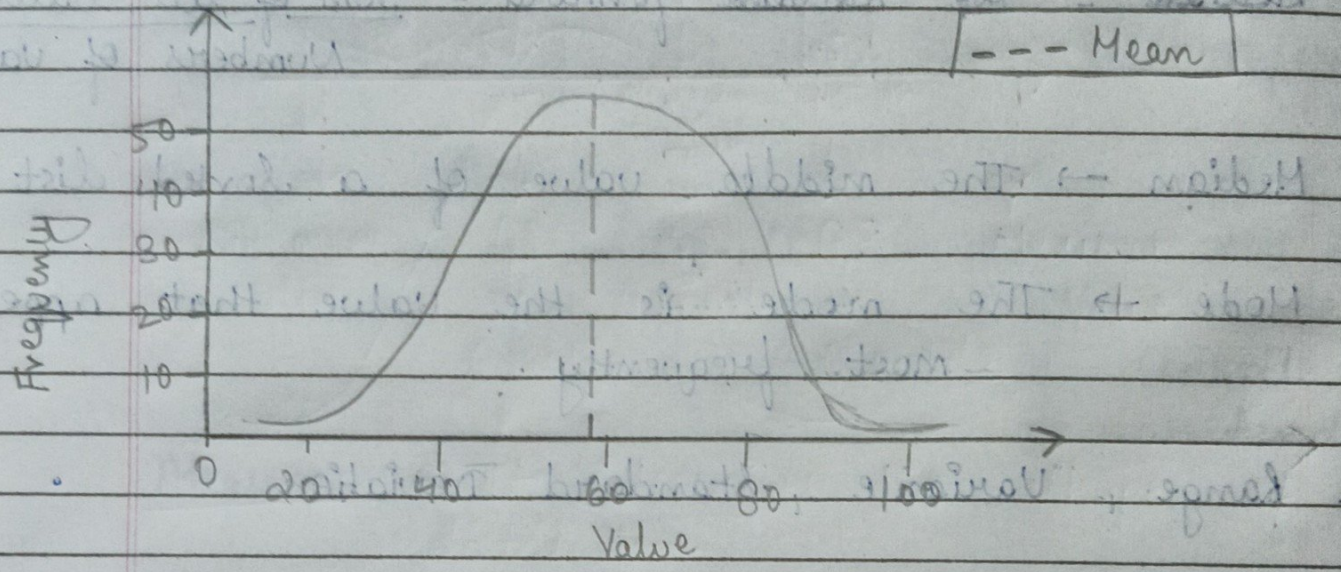
formula → Variance = $\frac{\sum (x_i - \bar{x})^2}{n}$

Standard Deviation → It shows how much on average the values deviate from the mean.

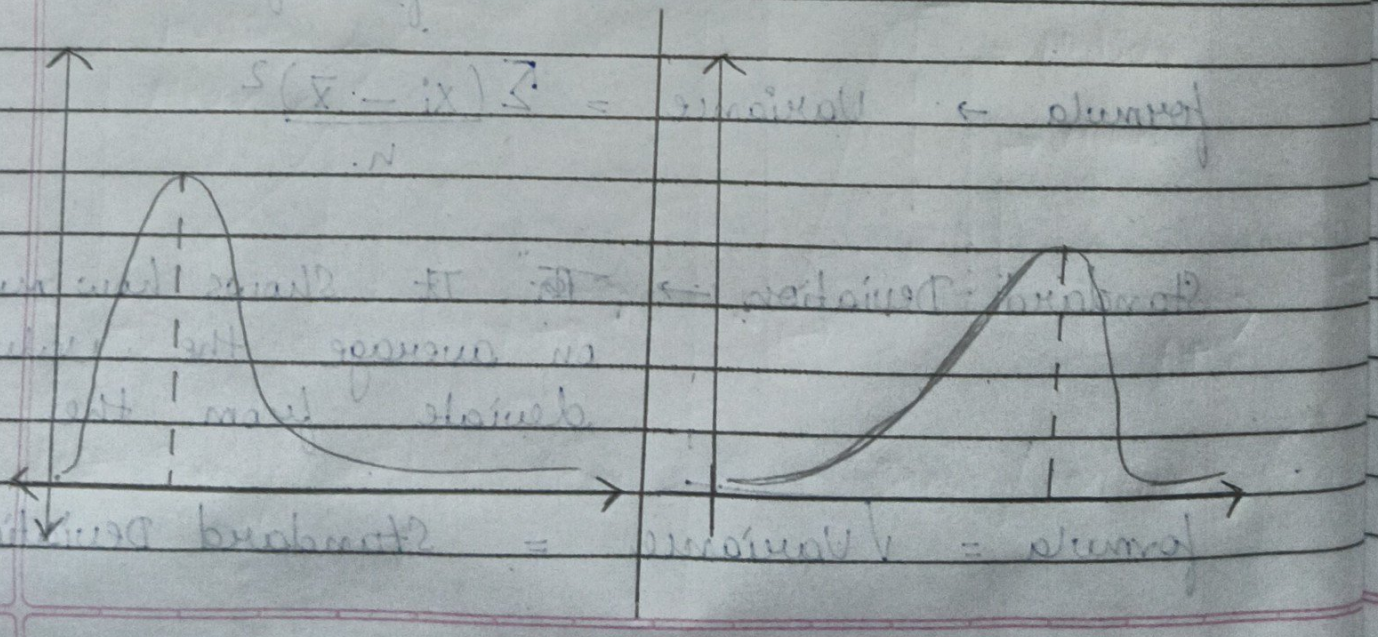
formula = $\sqrt{\text{Variance}}$ = Standard Deviation.

(Q.3). Explain the following term with neat and clean diagram along with its formula:

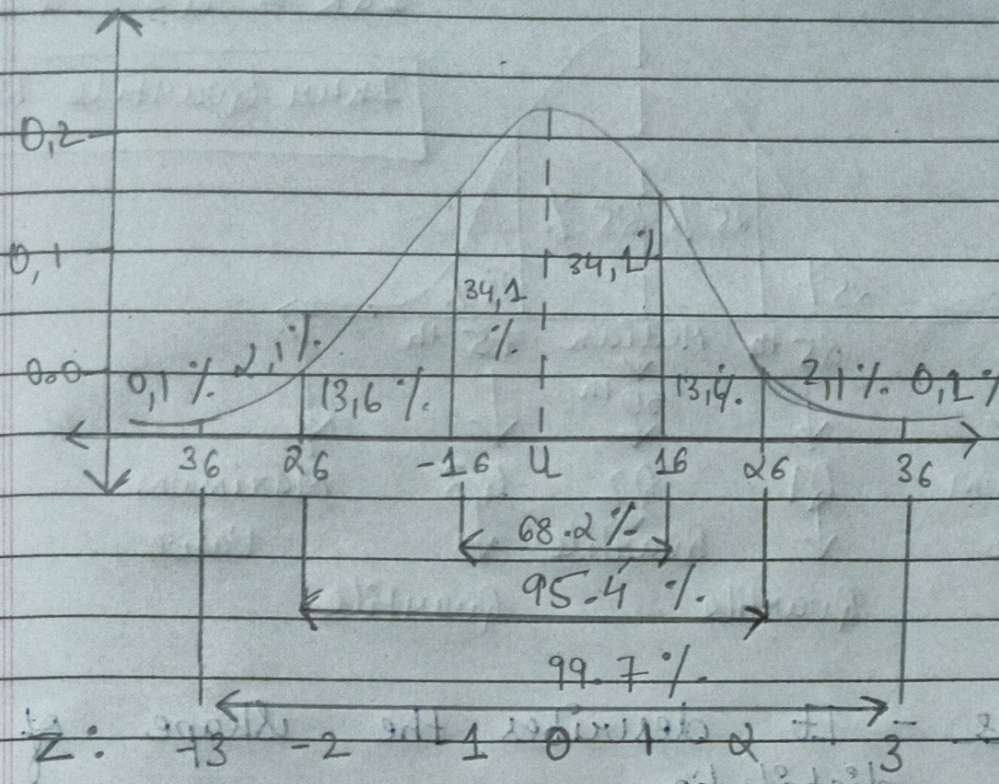
- Gaussian Distribution - Gaussian Distribution is a bell shaped curve that shows how data is distributed for eg. Height of people Marks of students, Blood pressure.



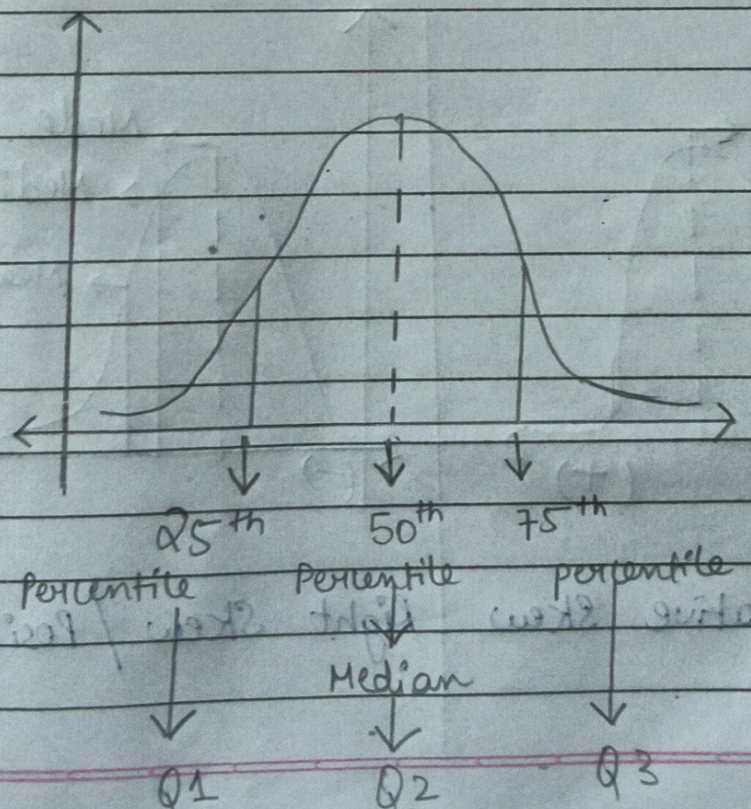
- Log Normal Distribution - A unevenly distributed data is known as log normal data. It may be right or left skewed.



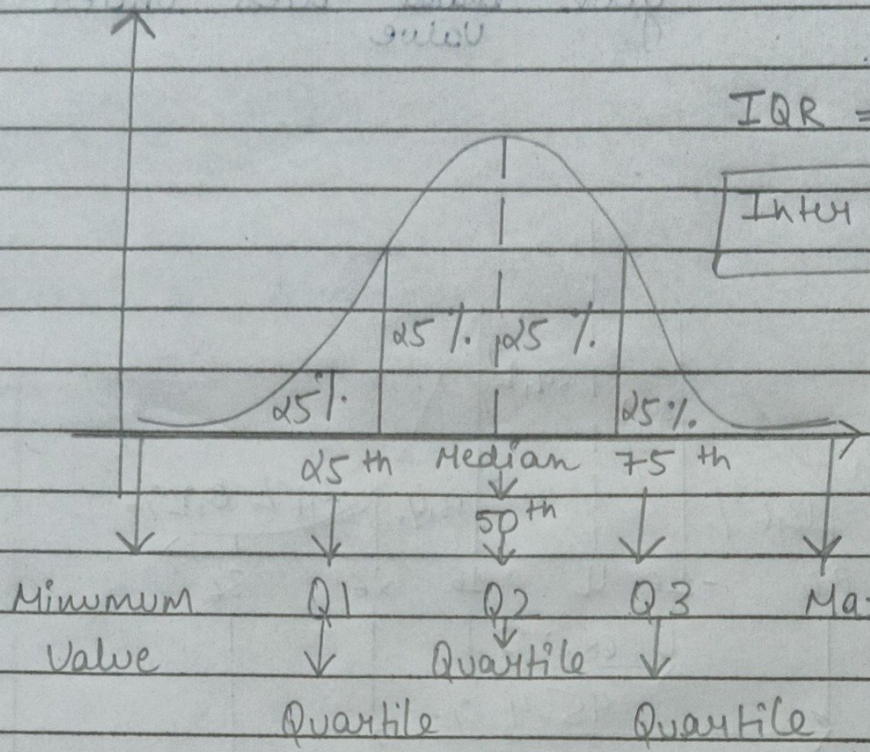
- 3 Sigma rule or Empirical Rule: It shows where your ~~data~~ value lies under the given percentage.



- Percentiles: A percentile indicates the value below which a given observation lies/fall.

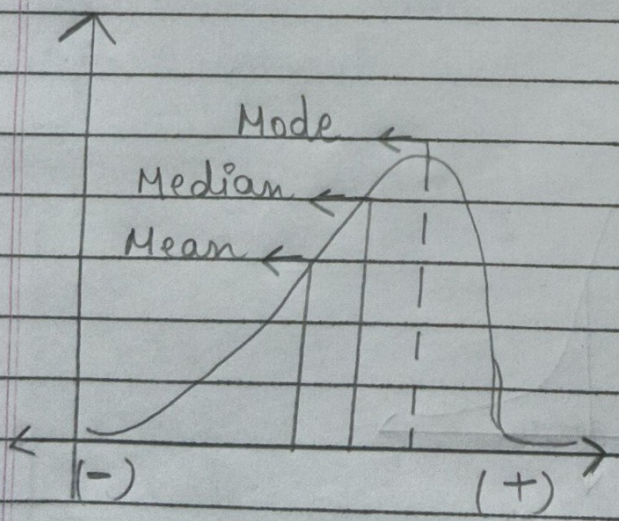


- **Quartiles & Five number summary:**

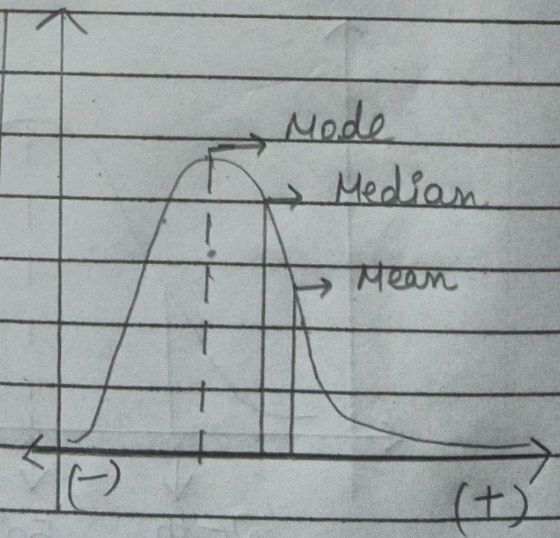


- **Skewness** - It describes the shape of the distribution.

There are 2 types of skewness:



Left Skew / Negative Skew



Right Skew / Positive Skew

PAGE NO.
DATE / /

Kurtosis