

PART A: Theory & Definitions

1. Define the following with examples from the dataset:

- Types of Data: Numerical & Categorical

Ans. • Numerical → means a countable numbers.

- Types of Numerical: Data

1. Discrete.

2. Continuous

According, to our dataset numerical columns are:-

1. Household_ID

2. Age_of_Household_Head

3. Household_Income

4. Family_size

- Categorical → means a data which contains labels or names.

- Types of Categorical data

1. Nominal

2. Cardinal

According, to our Dataset categorical columns are:-

1. Education_Level

2. Own_House

3. Urban_Rural

- Types of Statistics: Descriptive vs. Inferential

Ans. • Descriptive → statistics means the statistic which shows the summarization and main features of the dataset.

Types of Descriptive Statistics:-

1. Measure of Central Tendency
[Mean, Median, Mode]
2. Measure of dispersion
[Range, Variance, Standard deviation]
3. Shape of distribution:
[Skewness, Kurtosis]
4. Visualization
[Bar, Histogram, Pie, Box, Line, etc.]
- Inferential \rightarrow Sta means to get the conclusions or to make predictions on the dataset.

Types of Inferential Statistics:

1. Hypothesis testing
[Null Hypothesis, Alternative Hypothesis]
2. Confidence Interval
3. Correlation & covariance
4. Regression Analysis

• What is Descriptive Statistics?

Ans. To get the Summarization and main features of the dataset.

2. Explain the difference between:

- Mean, Median, Mode

Ans. Mean means the average

Median means the middle value

Mode means the most frequent value

- Range, Variance, Standard Deviation

Ans. Range = Max - Min

Variance means that how the data is spread out

Standard Deviation means Average deviation from the mean.

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

- Gaussian Distribution

- Log Normal Distribution

- 3-Sigma Rule or Empirical Rule

- Percentiles

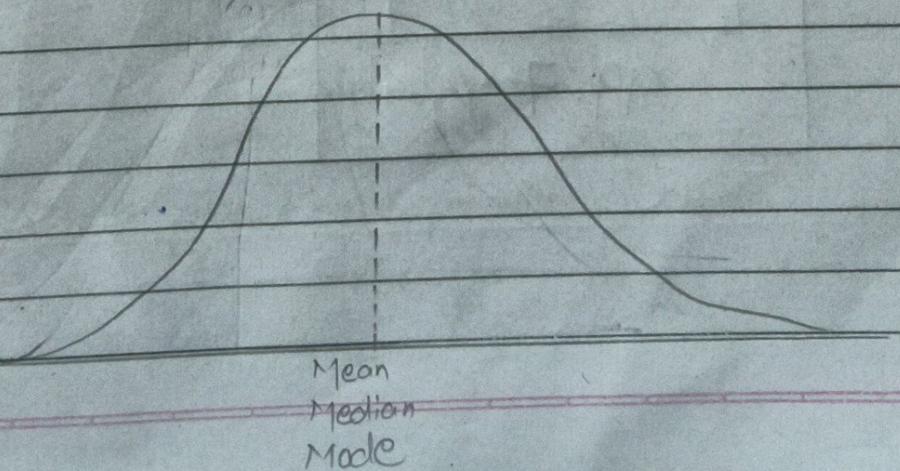
- Quartiles

- Five Number Summary

- Skewness

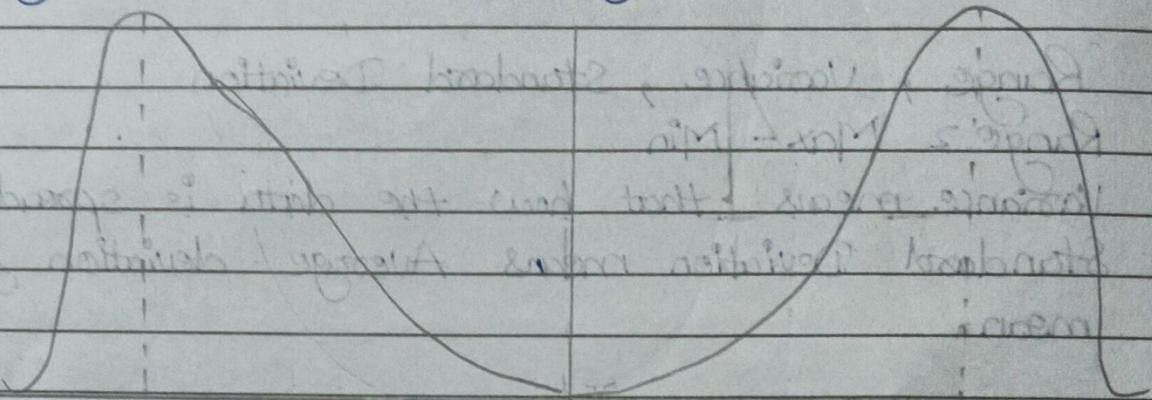
- Kurtosis

Ans. Gaussian distribution is the bell-shaped curve that shows how the data is distributed.



Formula:-

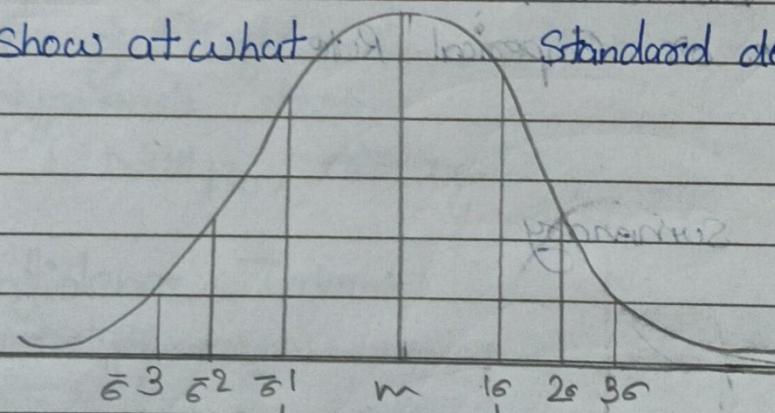
- Log-Normal Distribution states whether the right skewed is happening or left skewed.



and here Right Skewed parallel Left Skewed

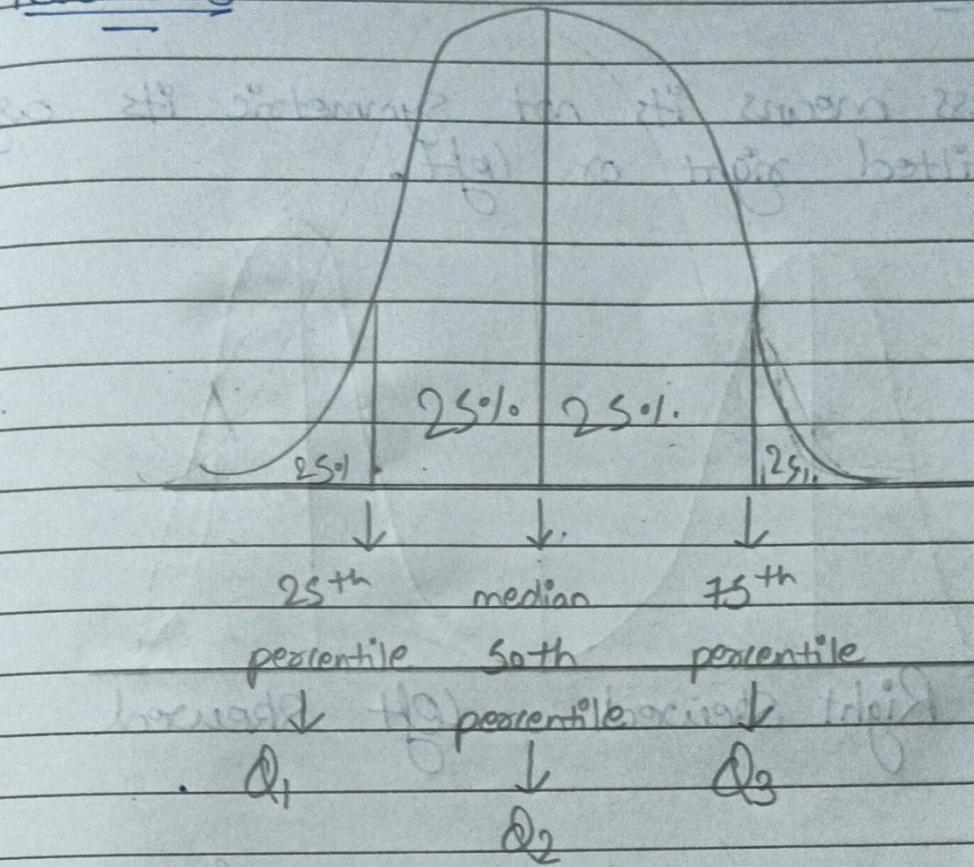
- 3-Sigma Rule or Empirical Rule

It shows at what standard deviation our data lies



68.2%	95.4%	99.7%
-------	-------	-------

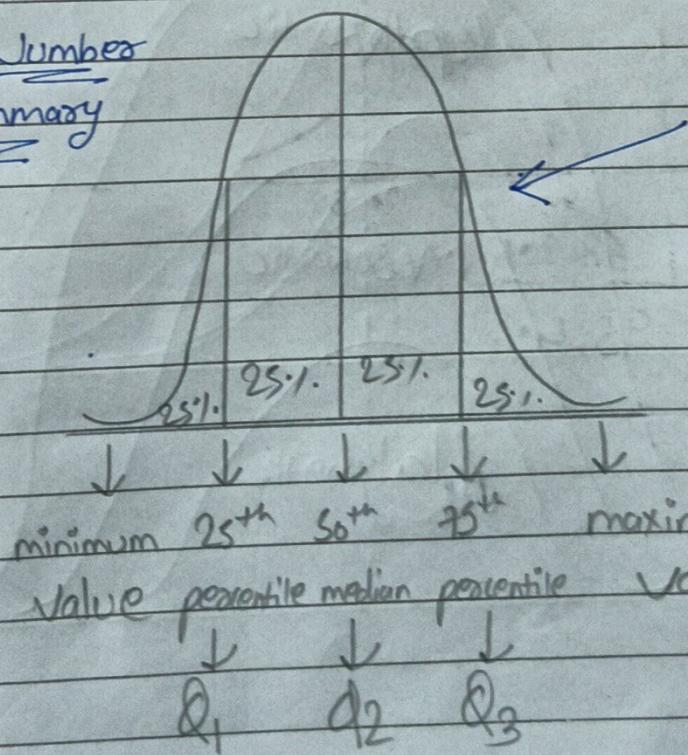
Percentiles



Quartiles divide the data into 3 equal parts
→ each get 25%.

Q

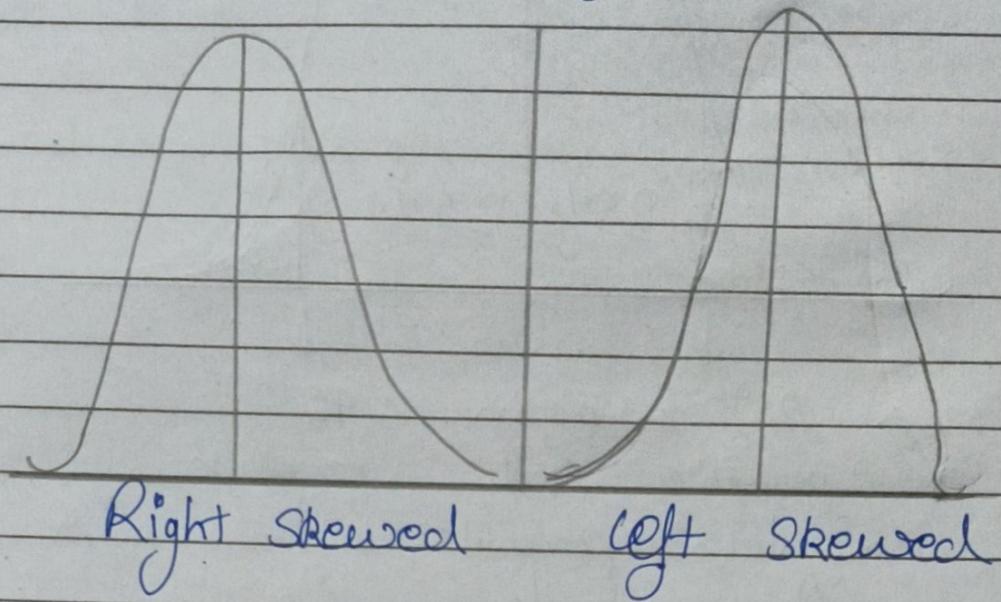
Five Number Summary



Five number summary means we just have to add minimum in the first and maximum at the last

Skewness

Skewness means it's not symmetric it's asymmetric.
It's tilted right or left.



Kurtosis tells us how peaked or flat the distribution is.

