

1 Types of data:-

→ Numerical :- It measures the quantities from data.

Types of numeric data:-

→ Discrete :- It means countable numbers
Eg: No. of children

→ Continuous:- Measurable quantities expressed with decimals

Eg: Height, Weight

Eg: Age, temperature

→ Categorical :- It means descriptive label or names

Eg: Gender, color

Types:-

→ Nominal:- No specific order

Eg: Male / Female

→ Ordinal:- Ordered categories

Eg: Ratings, Ranking

2 Types of statistics:-

→ Descriptive:- To summarize and describe main features of a dataset.

→ Uses:- Estimation of objects with individuals

→ Measure of central tendency

→ Measure of dispersion

→ Shape of distribution

→ Visualizing tools

→ Inferential: To make predictions or inferences about a population based on a sample.

Uses:-

→ Hypothesis testing

→ Confidence interval

→ Correlation & causation

→ Regression analysis

3

Descriptive statistics

→ It is a branch of statistics that focuses on summarizing, organizing & presenting raw data in a meaningful way.

→ It does not make predictions or infer conclusions beyond data, it only describes what's in front of you.

② Ex difference

Mean	Median	Mode
------	--------	------

→ Arithmetic average	→ Data is symmetric & no outliers	→ Most frequent values
----------------------	-----------------------------------	------------------------

→ Middle value is sorted	→ Data has outliers or is skewed	→ Categorical or frequent value is needed
--------------------------	----------------------------------	---

Range	Variance	Standard deviation
→ Max-min values	→ Avg squared deviation from mean	→ Root of variance
→ quick idea of spread	→ gt measures mathematical dispersion in data.	→ Real world measure of spread.

③ Exp full terms:

- Gaussian distribution :- A gaussian distribution or normal distribution is a bell-shaped curve that shows how data is distributed
- Most of data points are concentrated around the mean

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

- Percentiles help in understanding the distribution of data.
- It indicates the values below which a given percentage of observation fall.

Eg:- The 80th percentile means 80% of data lies below that value

Formula: ... Eg. $P = (m/N) \times 100$

→ Quartiles:

→ Quartiles divide data into 3 parts:

$Q_1 = 1^{\text{st}}$ quartile = 25% of data

$Q_2 = 2^{\text{nd}}$ quartile = Median (50% of data)

$Q_3 = 3^{\text{rd}}$ quartile = 75% of data

→ Five number summary:

→ It provides a quick overview of spread & center of a dataset.

It includes:

→ Minimum

→ Q_1

→ Q_2

→ Q_3

→ Maximum

→ Skewness:-

→ It is a statistical measure that describes the asymmetry of a data distribution around its mean. It describes if data is symmetrically distributed or tilted towards one side.

Formula:- Skewness = $3(\text{Mean} - \text{Median}) / (\text{Standard deviation})$

→ Kurtosis :

→ It measures the "tailedness" of a distribution - how heavy or light the tails are compared to a normal distribution. It tells us how likely it is to get outliers.

formula: $\frac{E(X-\mu)^4}{\sigma^4}$