Introduction to Statistics

Defn: Statistics is the science of collecting, organizing and analyzing data

⇓

Decision Making Process

Data: "facts or pieces of information"

Eg: Height of students in classroom

{175cm, 18 0cm, 190cm, 160cm ...}

IQ

{100, 90, 95, 80 ...}

Data is a collection of raw, unorganized facts and details like text, observations, figures, symbols, and descriptions of things etc.

In other words, data does not carry any specific purpose and has no significance by itself. Moreover, data is measured in terms of bits and bytes – which are basic units of information in the context of computer storage and processing.

Data can be recorded and doesn't have any meaning unless processed.

② Types of Statistics

① Descriptive Statistics

Defn: It consists of organizing and summarizing data

① Measure of Central Tendency [Mean, Median, Mode]

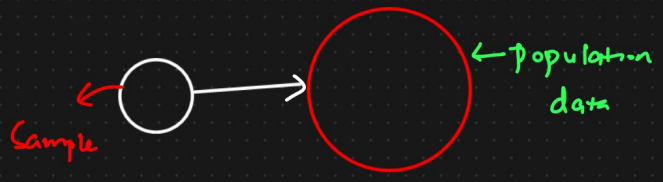
② Measure of Dispersion [Variance, Std]

③ Different type of Distribution of data.

Eg: Histogram, pdf, pmf

② Inferential Statistics

Defn: It consists of using data you have measured to form conclusion



① Z-test
② t-test
③ CHI SQUARE

} Hypothesis Testing Ho, H, p value, Significance value

Eg: Lets say there are 20 statistics classes at your college. and you have collected the heights of students in the class.

Heights are recorded [175cm, 180cm, 140, 140, 135, 160, 135, 190cm]

$$\frac{175+180+140+140+135+160+135+190}{8} = \text{Avg Height}$$

Descriptive Question

"What is the average height of the entire classroom"

Inferential Question

Sample

"Are the height of the students in classroom similar to what you expect in the entire college"

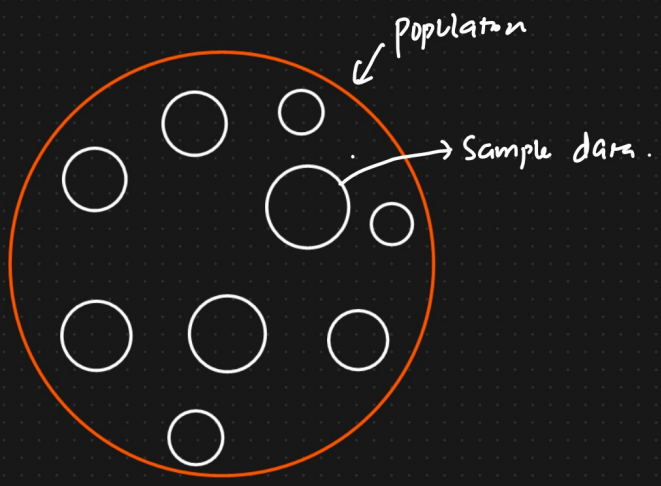
↳ population data

Population And Sample Data

Population: The group you are interested in studying

Sample : a subset of population

Eg: Exit Poll



Population:

Definition: In statistics, a population refers to the entire set of individuals, items, or data points that are the subject of interest for a particular study. It includes all possible elements that share a common characteristic.

Example: If you're conducting a survey to determine the average income of residents in a city, the population would consist of all the residents of that city.

Use case: Populations are often used in research studies where the goal is to make generalizations about a specific group or phenomenon. They provide a comprehensive understanding of the characteristics of the entire group under study.

Sample:

Definition: A sample is a subset of the population that is selected to represent the larger group. It is used to make inferences or draw conclusions about the population as a whole.

Example: In the survey mentioned earlier, instead of surveying every resident in the city, you might select a random sample of 500 residents and collect data on their incomes. This subset of 500 residents represents the larger population of the city.

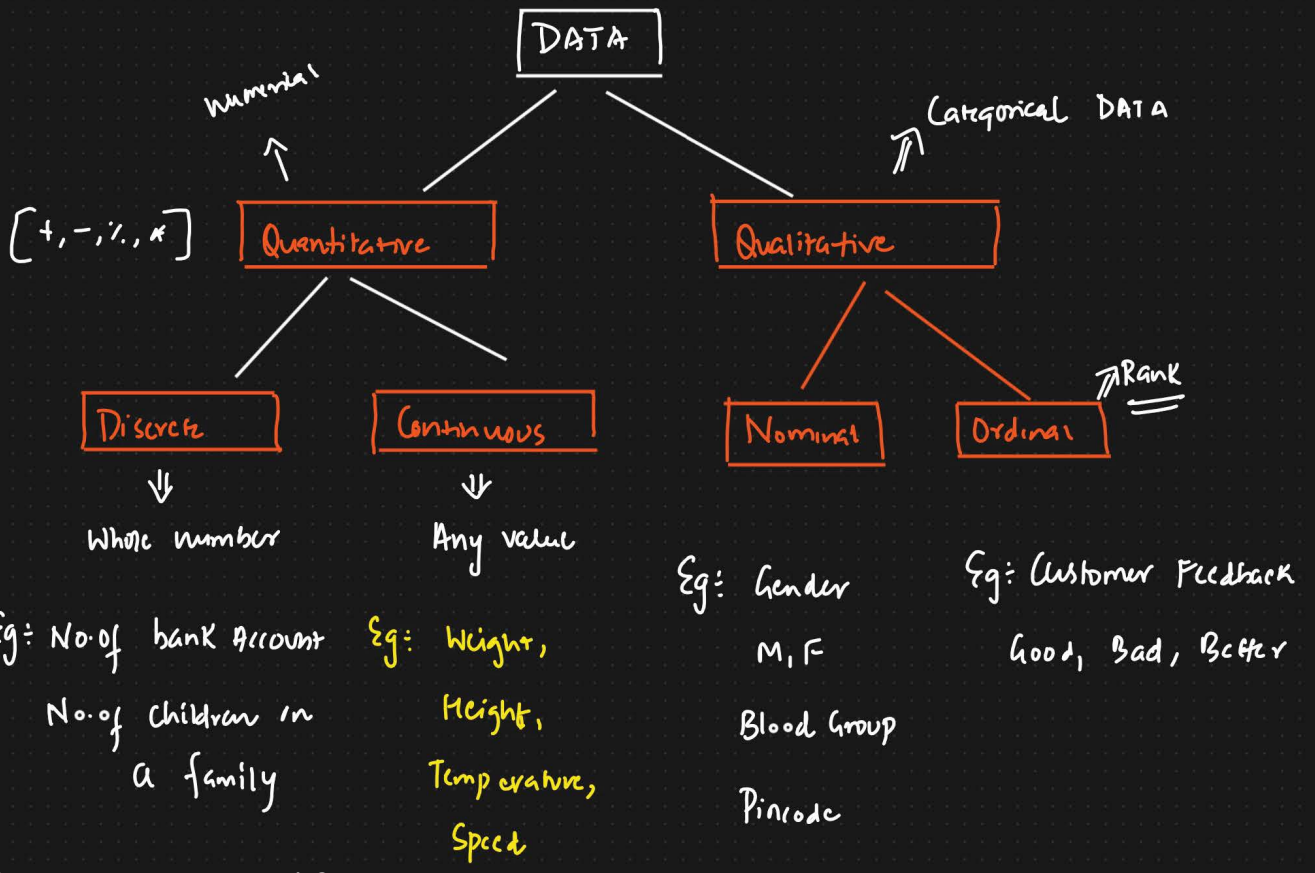
Use case: Samples are used when it is impractical or impossible to collect data from an entire population. By studying a smaller, representative subset, researchers can make educated estimates or predictions about the population parameters.

Key Differences:

Size: The population includes all individuals or elements, while the sample is a smaller subset of the population.
Representation: A well-chosen sample should accurately represent the population characteristics from which it is drawn.
Cost and Time: Collecting data from an entire population can be time-consuming and expensive, whereas sampling is often more efficient and cost-effective.

③ Typs of Data

DMC	DC	ISI	BUI	FWI	Classes	Region
3.4	7.6	1.3	3.4	0.5	not fire	0
4.1	7.6	1	3.9	0.4	not fire	0
2.5	7.1	0.3	2.7	0.1	not fire	0
1.3	6.9	0	1.7	0	not fire	0
3	14.2	1.2	3.9	0.5	not fire	0
5.8	22.2	3.1	7	2.5	fire	0
9.9	30.5	6.4	10.9	7.2	fire	0
12.1	38.3	5.6	13.5	7.1	fire	0



What is a measurement scale?
Scales of measurement refer to ways in which variables/numbers are defined and categorized. Each scale of measurement has certain properties which in turn determines the appropriateness for use of certain statistical analyses.

④ Scale of Measurement

1. Nominal Scale Data
2. Ordinal Scale Data
3. Interval Scale Data
4. Ratio. Scale Data.

① Nominal Scale Data

1. Qualitative / Categorical
2. Eg: Gender, Colors, Fabrics
3. Order does not matter

Eg: Favorite Color

Red → 5 → 50%

Blue → 3 → 30%

Orange → 2 → 20%

10

Gender

M

F

② Ordinal Scale Data

1. Ranking is important
2. Order matter
3. Difference cannot be measured

Eg:

1 → Best

2 → Good

3 → Bad

Working Professional

M

T

W

Th

← F

Race:

1st

2nd

3rd

4:20

5:30

6:00

OF

③ Interval Scale Data

① The order matter

② Difference can be measured

③ Ratio cannot be measured ✓

④ No "0" Starting Point

Eg: Temperature Variable.

-30° F

-15° F

30 F

60 F

90 F

120 F

60:30 = 2:1 →

90 - 60 = 30 F

120 - 90 = 30 F

④ Ratio Scale Data

① The order matter ✓

② Differences are measurable (including ratio)

③ Contains a "0" Starting point.

Eg: Students marks in a class

0, 90, 60, 30, 75, 40, 50

Asc = 30, 40, 50, 60, 75, 90

40 - 30 = 10

50 - 30 = 20

Ratio = $\frac{90}{30}$ = 3 : 1

Example

1. Length of Different Rivers In the World ?
2. Favorite food based on Gender ?
3. Marital Status ?
4. IQ measurement ?