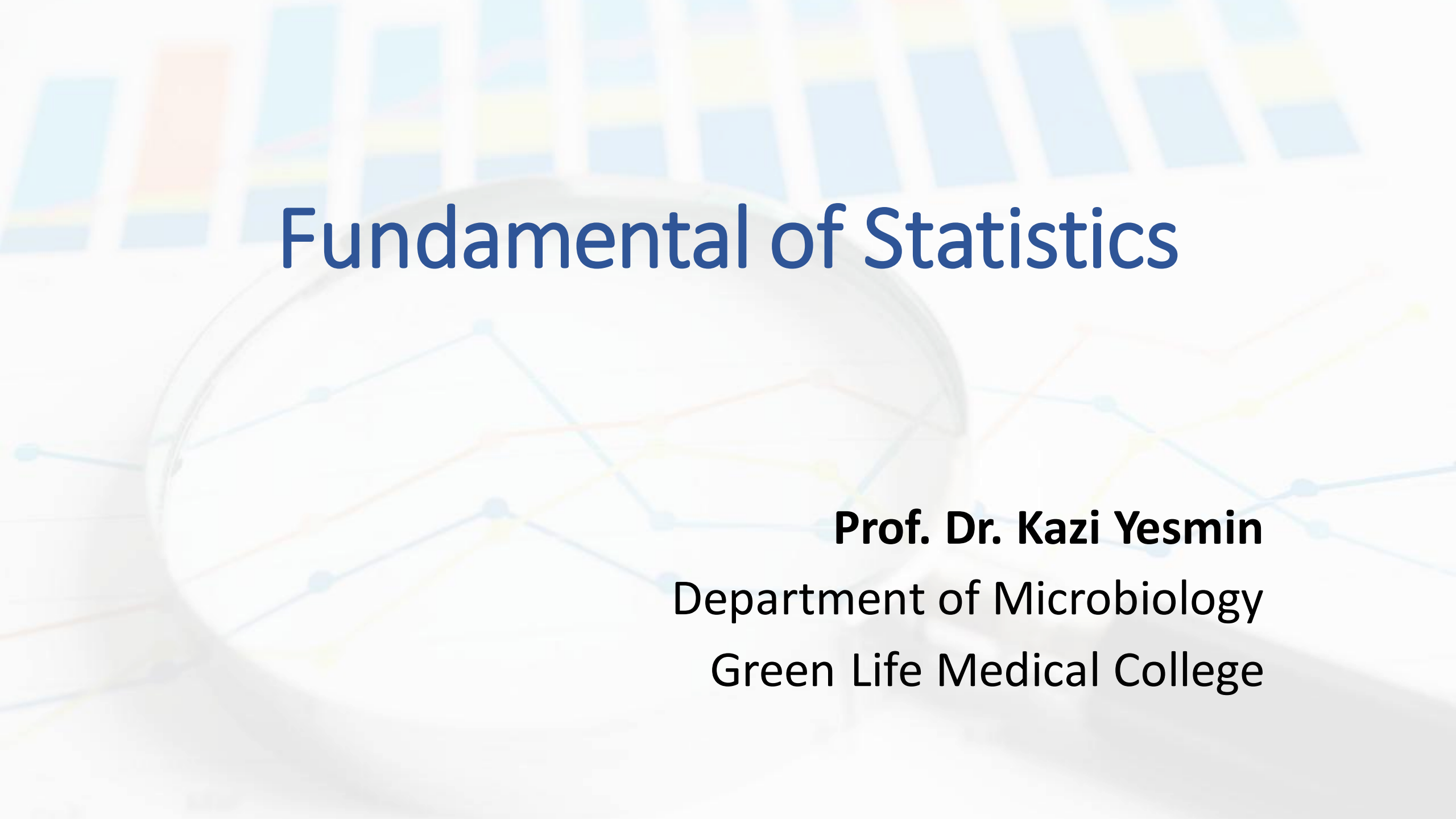


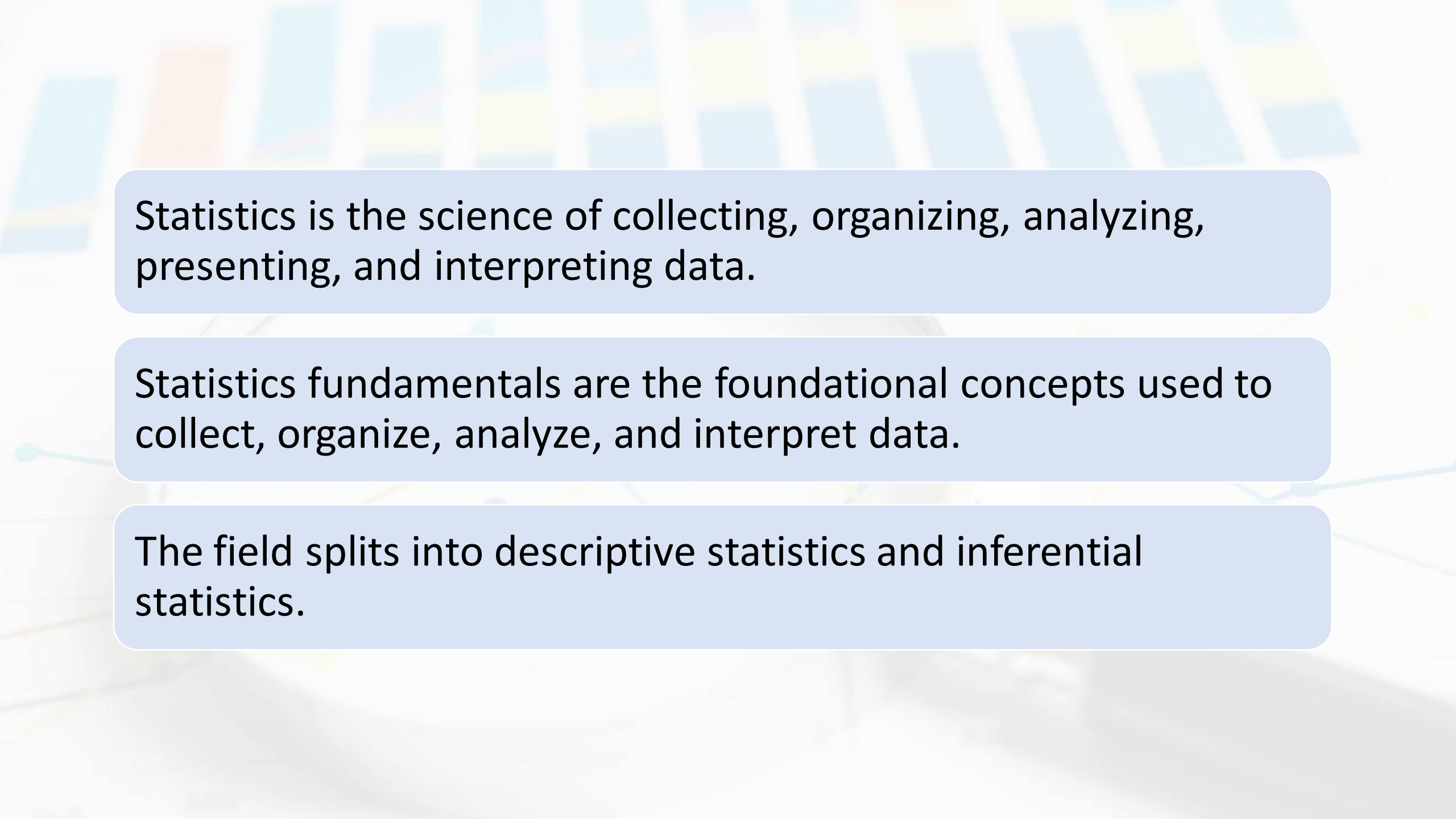
The background features a stylized, semi-transparent graphic. It includes a bar chart at the top with several vertical bars in shades of blue, orange, and yellow. Below the bars is a line graph with multiple colored lines (blue, orange, yellow, green) connecting various data points. The entire graphic is set against a light gray background.

Fundamental of Statistics

Prof. Dr. Kazi Yesmin

Department of Microbiology

Green Life Medical College



Statistics is the science of collecting, organizing, analyzing, presenting, and interpreting data.

Statistics fundamentals are the foundational concepts used to collect, organize, analyze, and interpret data.

The field splits into descriptive statistics and inferential statistics.

Descriptive and Inferential Statistics

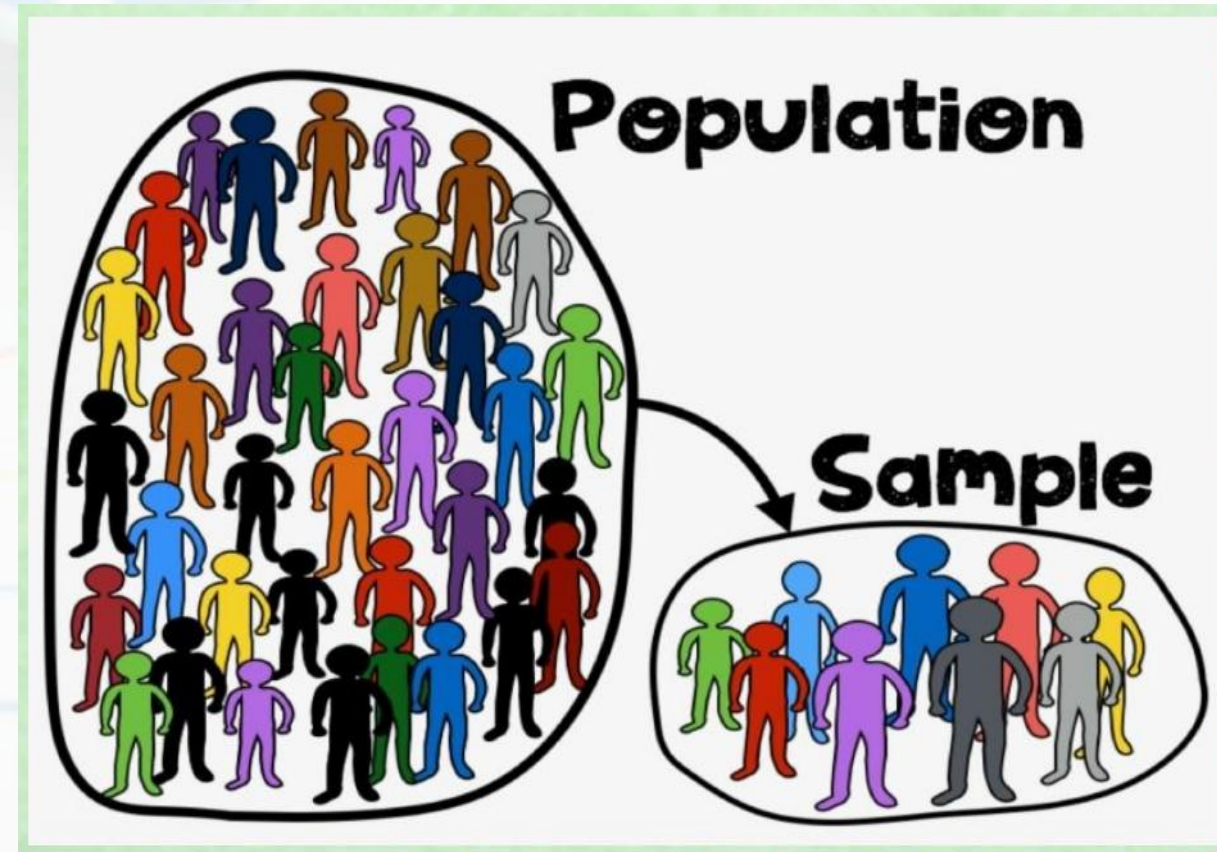
- **Descriptive statistics** consist of methods for organizing and summarizing information (which summarize features of a data set).
- **Inferential statistics** consist of methods for drawing and measuring the reliability of conclusions about population (which make predictions about a larger population using sample data).

Objectives

- Population & Sample
- Parameters & Statistic
- Data & Variables
- Descriptive statistics
- Measures of Central Tendency
- Measures of Dispersion
- Inferential Statistics
- Hypothesis
- Probability

Population and Sample

- A population is the set of all measurements of interest. Typically, the population is not observed, but we wish to make statements or inferences about it.
- Population is a complete set of items or subjects which can be studied.



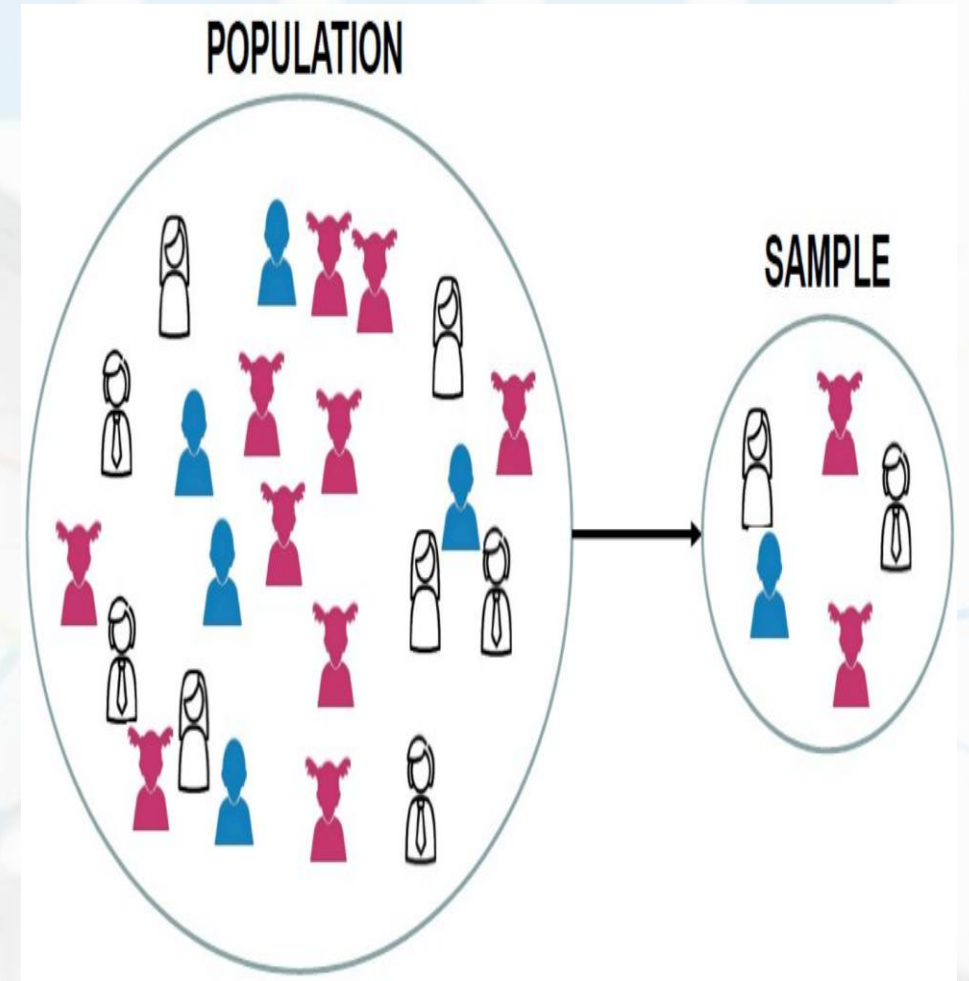
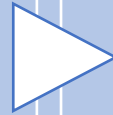
Population and Sample

Study population

The specific population from which data are collected

Sample

A subset of population/
A small part of that population



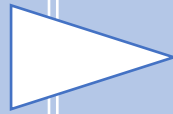
Parameters & Statistics

Parameters

A parameters is an unknown numerical summary of the population

Statistics

It is a known numerical summary of the sample which can be used to make inference about parameters

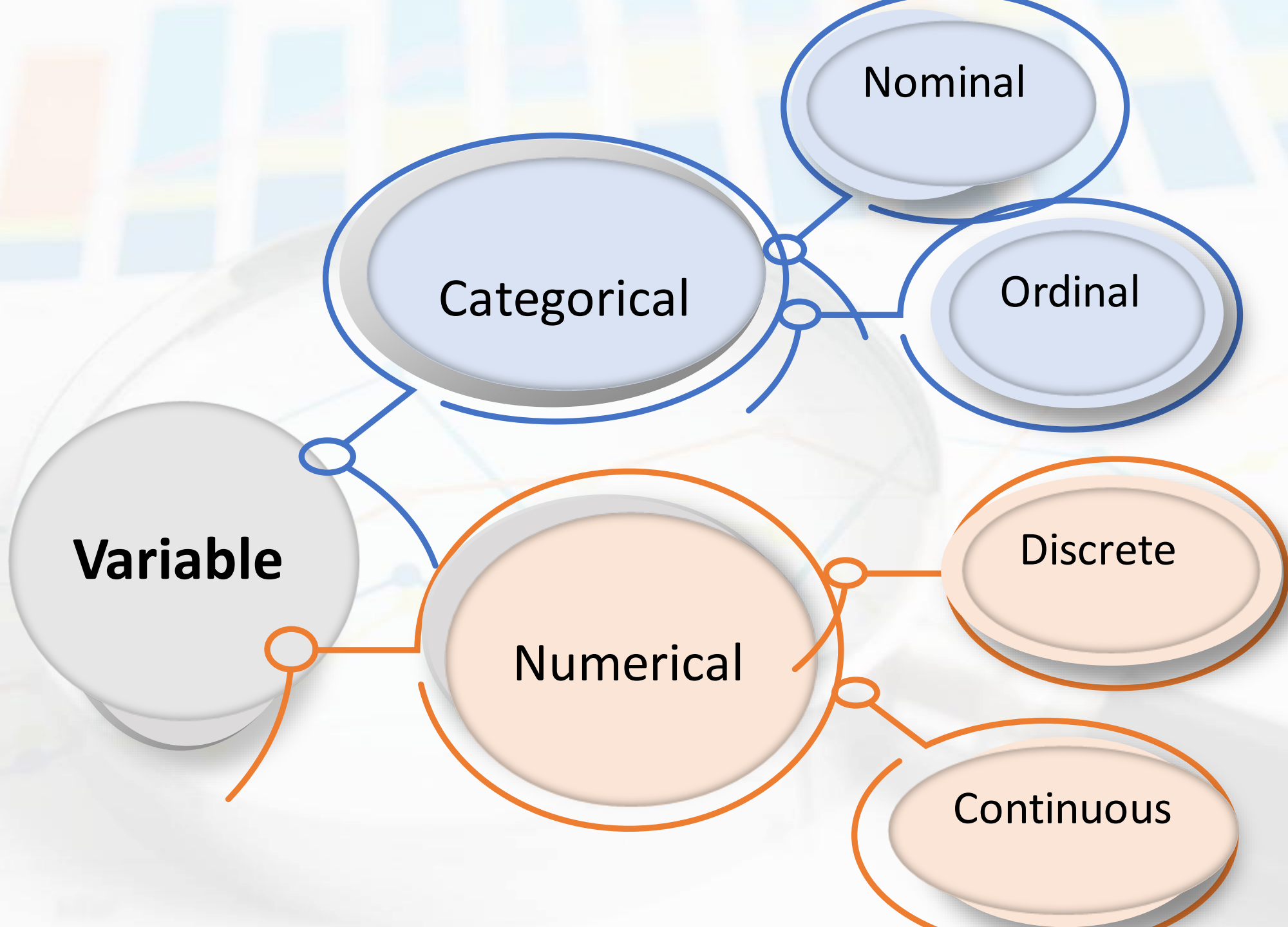


Variable & Data

A variable is a characteristic whose value varies from person to person, object to object or from phenomenon to phenomenon.

Data are the facts and figures collected, analyzed, and summarized for presentation and interpretation.

All the data collected in a particular study are referred to as the data set for the study.



Categorical Data

Nominal

Gender of study participant
Religion
Resident

Ordinal

Education level
Income level
Satisfaction ranking

Numerical Data

Discrete

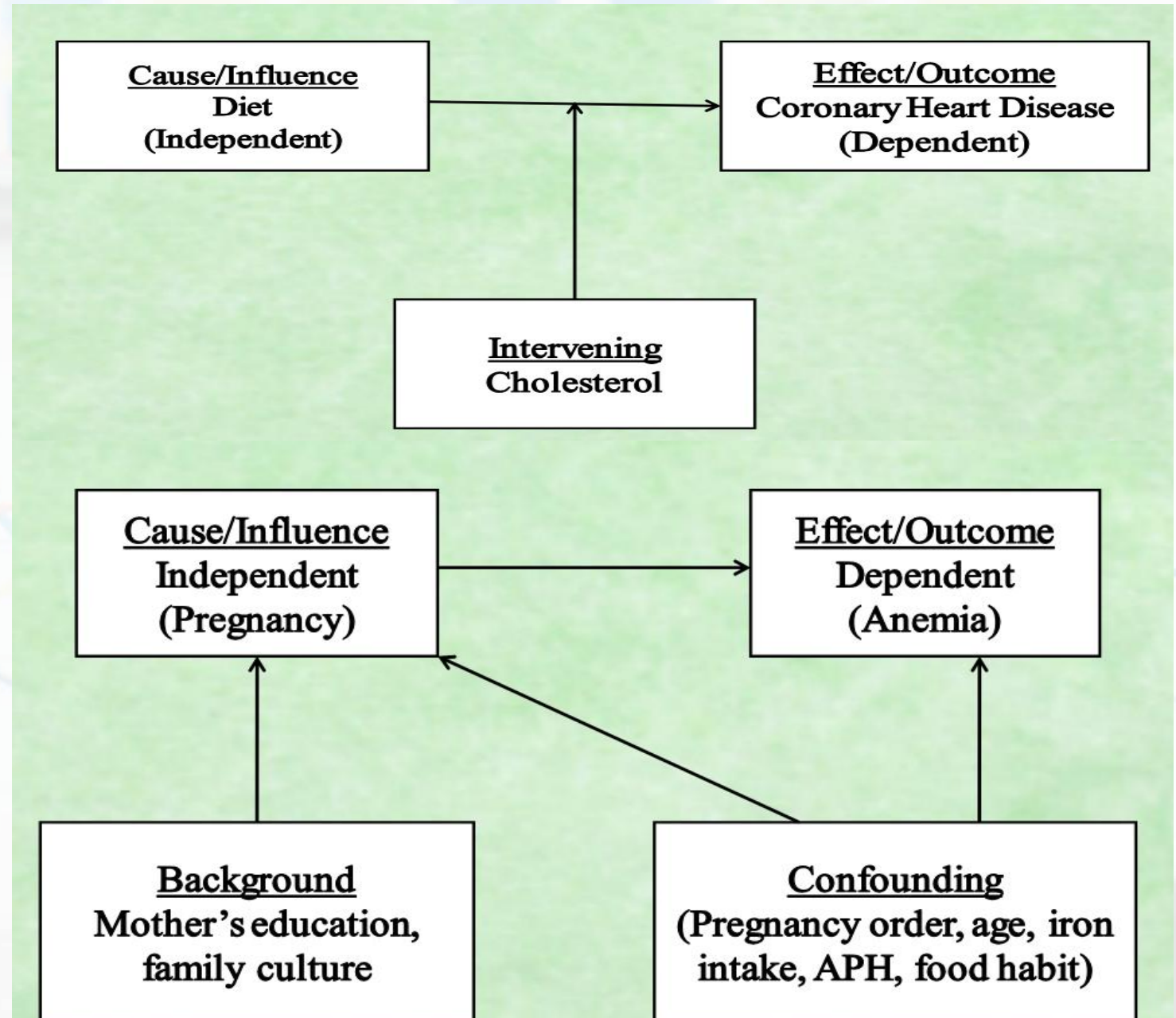
No. of children
No. of hospital beds
No. of family member

Continuous

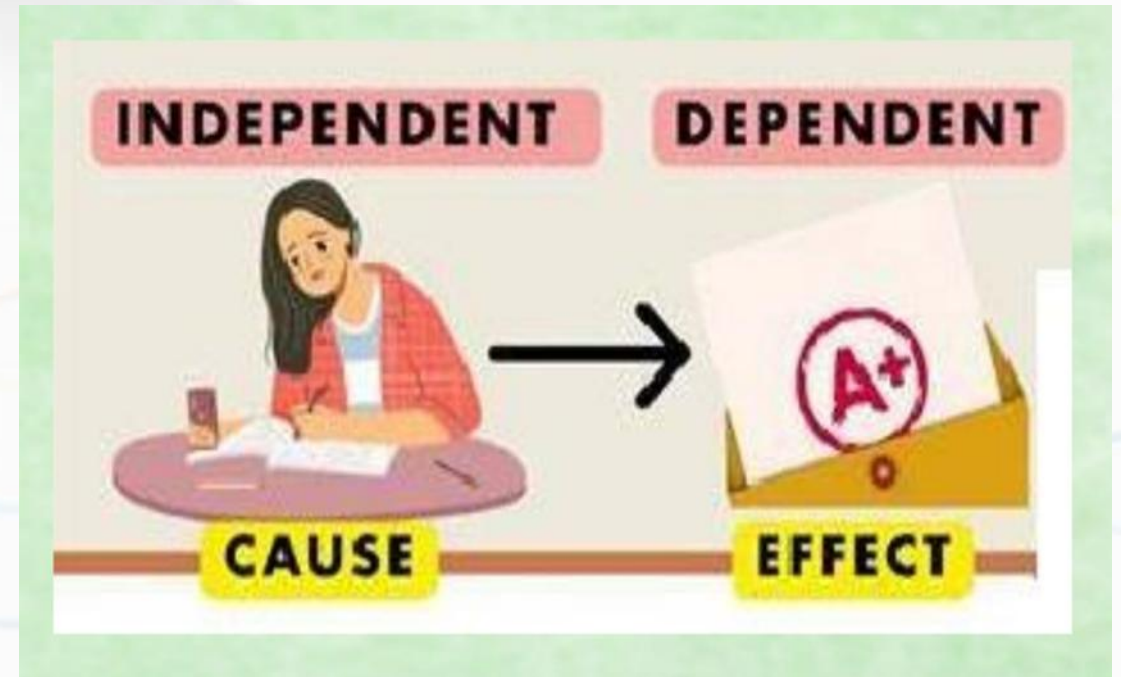
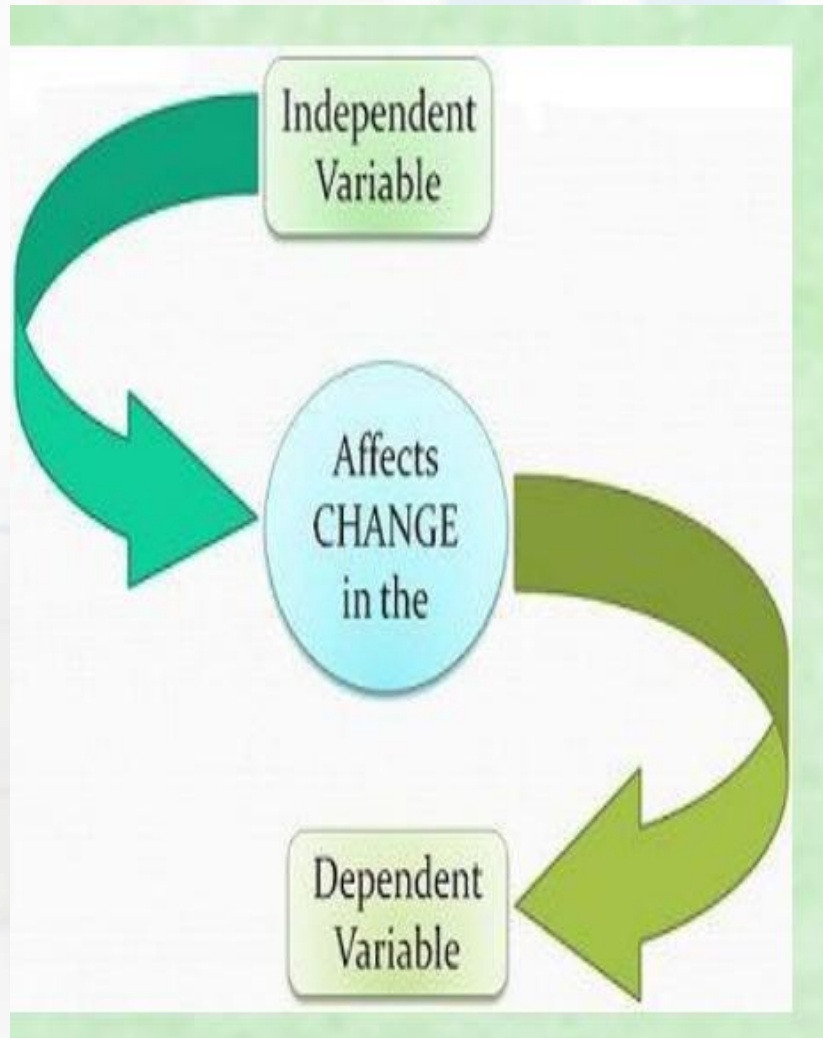
Age, height, weight, blood sugar etc.

Other variables

- Independent variables
/Predictor variables
- Dependent variables
/Outcome variables
- Confounding variables
- Background variables
- Intervening variables



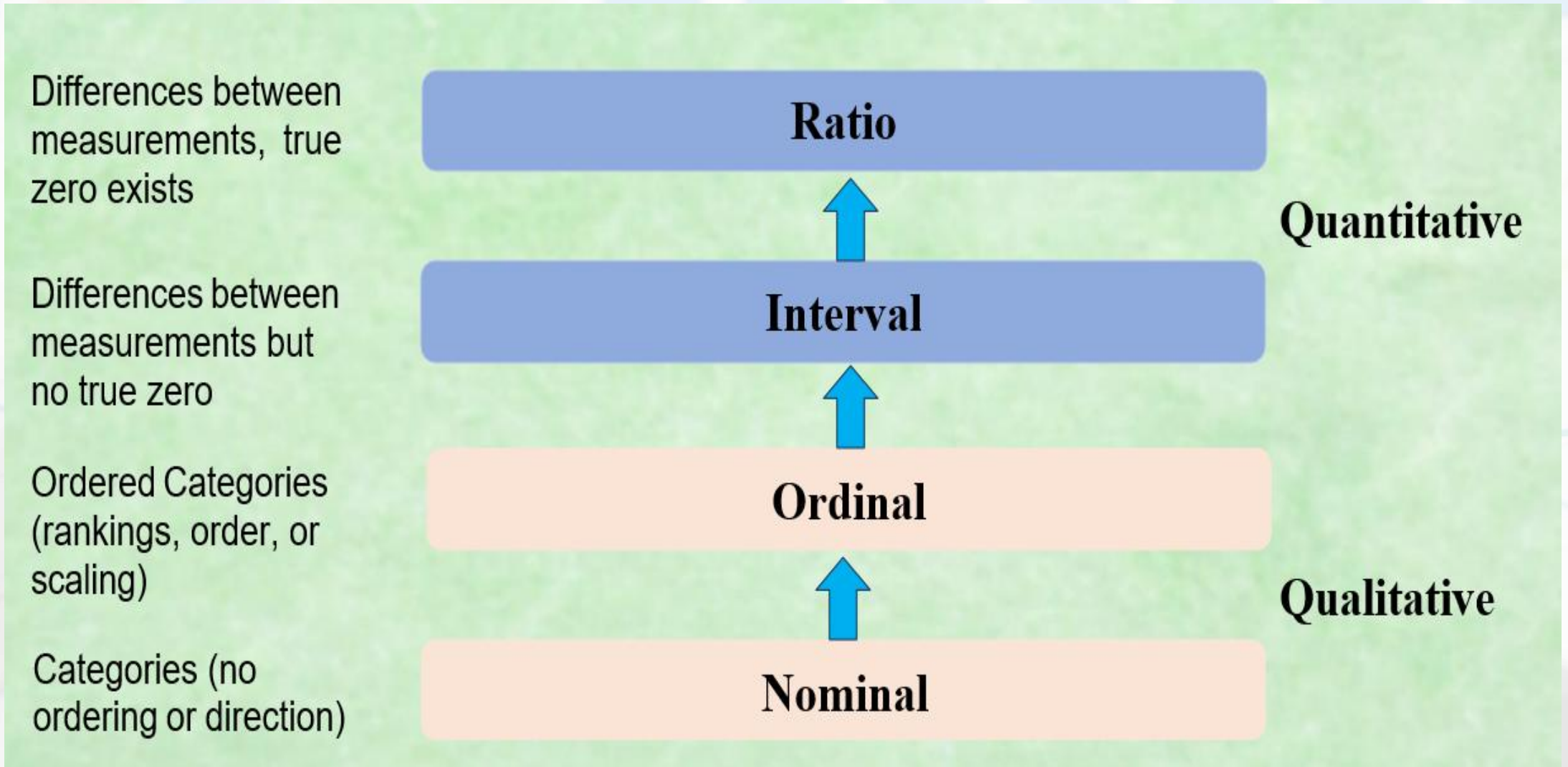
Independent and dependent variable



Measurement

- Measurement may be defined as the assignment of numbers to objects or events according to a set of rules.
- The various measurement scales result from the fact that measurement may be carried out under different sets of rules.
- **There are four types of scales of measurement:-**
 1. Nominal scale Qualitative data
 2. Ordinal
 3. Interval Quantitative data
 4. Ratio

Measurement scales

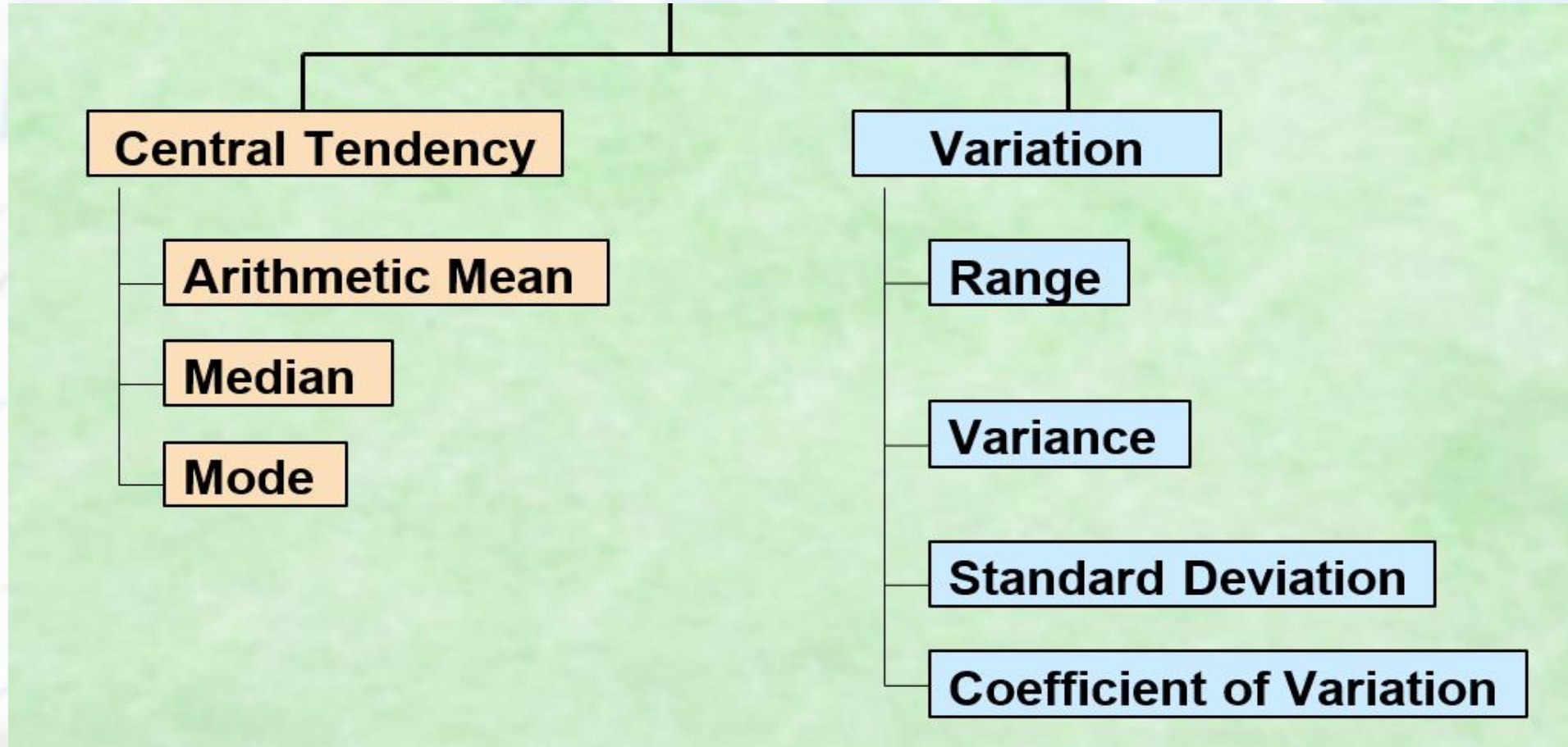


File Edit View Data Transform Analyze Graphs Utilities Extensions Window Help											
              											
	Name	Type	Width	Decimals	Label	Values	Missing	Columns	Align	Measure	Scale
7	S2.Q2.4_rel...	Numeric	8	0	religion of the s...	{1, hindu}...	None	8	Right	Nominal	Ordinal
8	S2.Q2.5_inc...	Numeric	8	0	Monthly income	None	None	8	Right	Scale	Nominal
9	S2.Q25.2_l...	Numeric	8	0	level of socioec...	{1, low inco...	None	8	Right	Ordinal	
0	S2.Q2.6.1_...	Numeric	8	0	where do you r...	{1, GLMC}...	None	8	Right	Nominal	
1	S2.Q2.6.2_r...	Numeric	8	0	what do you re...	{1, MBBS}...	None	8	Right	Nominal	
2	S2.Q2.6.3_...	Numeric	8	0	read in which y...	{1, 3rd year}...	None	8	Right	Ordinal	
3	S3.Q3.1_b...	Numeric	8	0	do you know th...	{1, yes}...	None	8	Right	Nominal	
4	S3.Q3.2_learn	Numeric	8	0	where did you l...	{1, teaching...	None	8	Right	Nominal	

Level of Measurement

Level of Measurement	Brief Description	Example
Ratio	There is a natural zero starting point and ratios make sense.	Heights, lengths, distances, volumes
Interval	Differences are meaningful, but there is no natural zero starting point and ratios are meaningless.	Body temperatures in degrees Fahrenheit or Celsius
Ordinal	Data can be arranged in order, but differences either can't be found or are meaningless.	Ranks of colleges in <i>U.S. News & World Report</i>
Nominal	Categories only. Data cannot be arranged in order.	Eye colors

Describing data



Mean

- The most common measure of central tendency
- The mean (or arithmetic mean) of a set of data is the measure of center found by adding all of the data values and dividing the total by the number of data values.
- Mean = sum of values divided by the number of values
- Affected by extreme values (outliers)

Mean

- For a population of N values:

$$\mu = \frac{\sum_{i=1}^N x_i}{N} = \frac{x_1 + x_2 + \dots + x_N}{N}$$

Population values

Population size

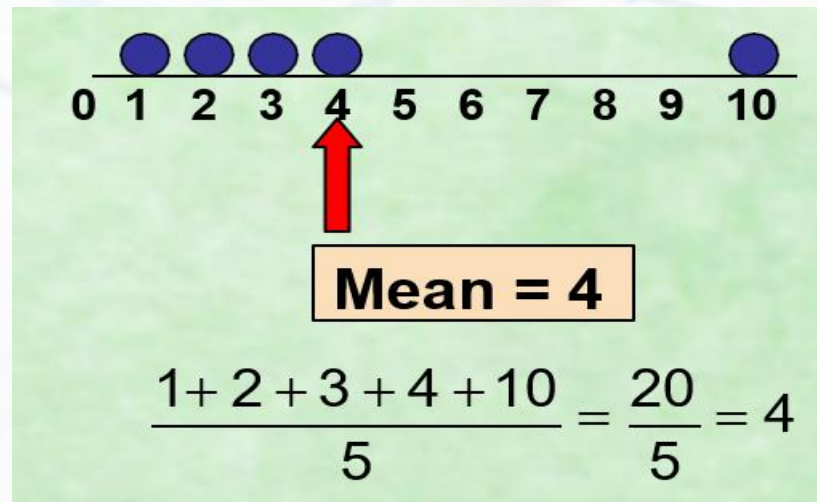
- For a sample of size n :

$$\bar{x} = \frac{\sum_{i=1}^n x_i}{n} = \frac{x_1 + x_2 + \dots + x_n}{n}$$

Observed values

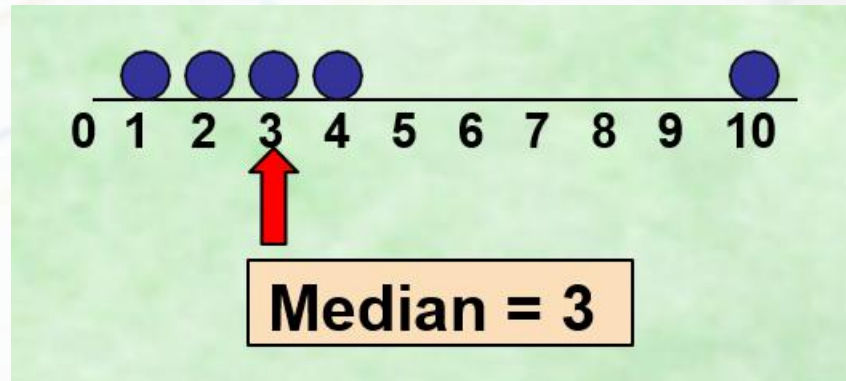
Sample size

- Example:



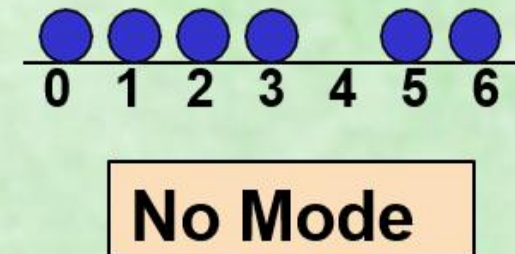
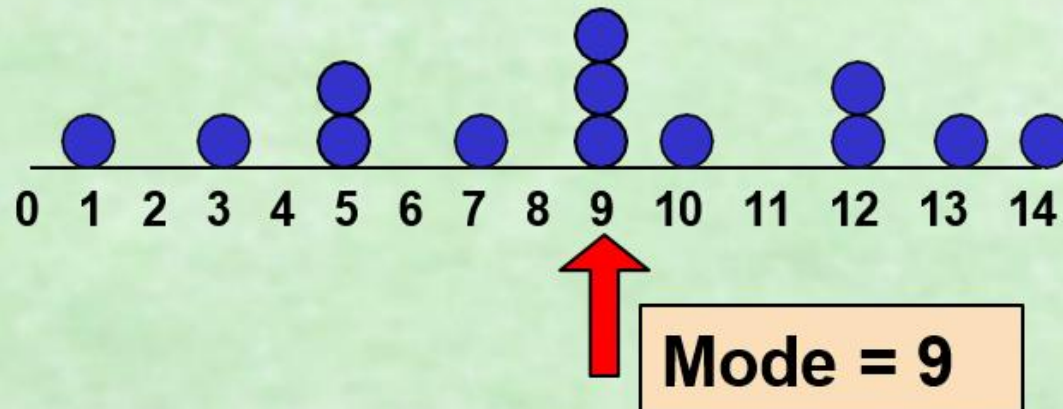
Median

- The median of a data set is the measure of center that is the middle value when the original data values are arranged in order of increasing (or decreasing) magnitude.
- In an ordered list, the median is the “middle” number (50% above, 50% below)



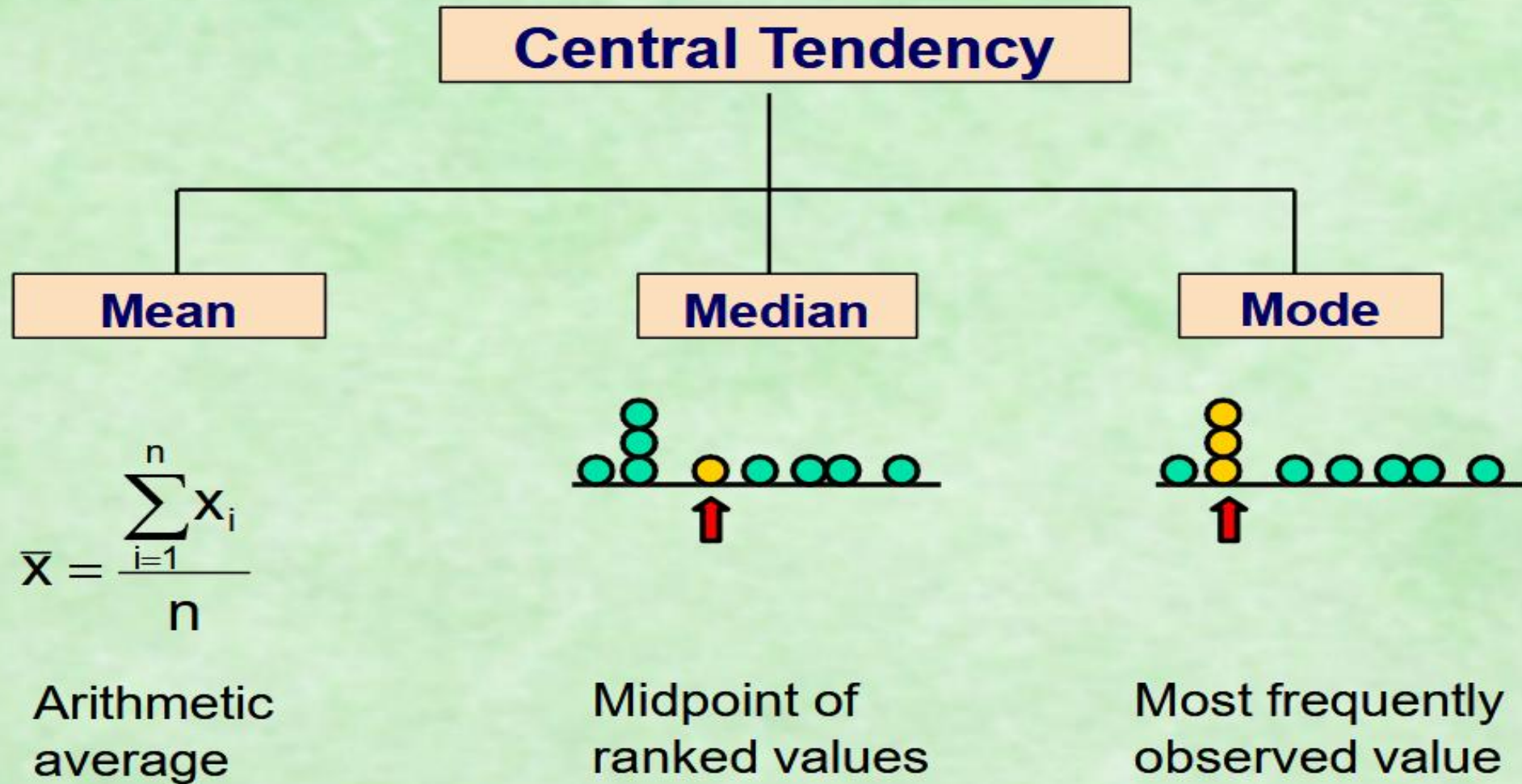
Mode

- The mode isn't used much with quantitative data, but it's the only measure of center that can be used with qualitative data (consisting of names, labels, or categories only).
- The mode of a data set is the value(s) that occurs with the greatest frequency.
- Finding the Mode: A data set can have one mode, more than one mode (bimodal, multimodal), or no mode.



Measures of Central Tendency

Overview



Standard deviation

The standard deviation of a set of sample values, is a measure of how much data values deviate away from the mean.

It measures how spread out the values in a dataset are from the mean (average) value.

Low SD

- The data points tend to be close to the mean

High SD

- The data points are spread out over a wider range of values

- Most commonly used measure of variation
- Shows variation about the mean
- Has the same units as the original data

■ Sample standard deviation:

$$S = \sqrt{\frac{\sum_{i=1}^n (x_i - \bar{x})^2}{n-1}}$$

Calculation Example: Sample Standard Déviation

Sample

Data (x_i):

10 12 14 15 17 18 18 24

$n = 8$

Mean = $\bar{x} = 16$

$$s = \sqrt{\frac{(10 - \bar{x})^2 + (12 - \bar{x})^2 + (14 - \bar{x})^2 + \square + (24 - \bar{x})^2}{n - 1}}$$

$$= \sqrt{\frac{(10 - 16)^2 + (12 - 16)^2 + (14 - 16)^2 + \square + (24 - 16)^2}{8 - 1}}$$

$$= \sqrt{\frac{126}{7}} = \underline{4.2426}$$



A measure of the “average”
scatter around the mean

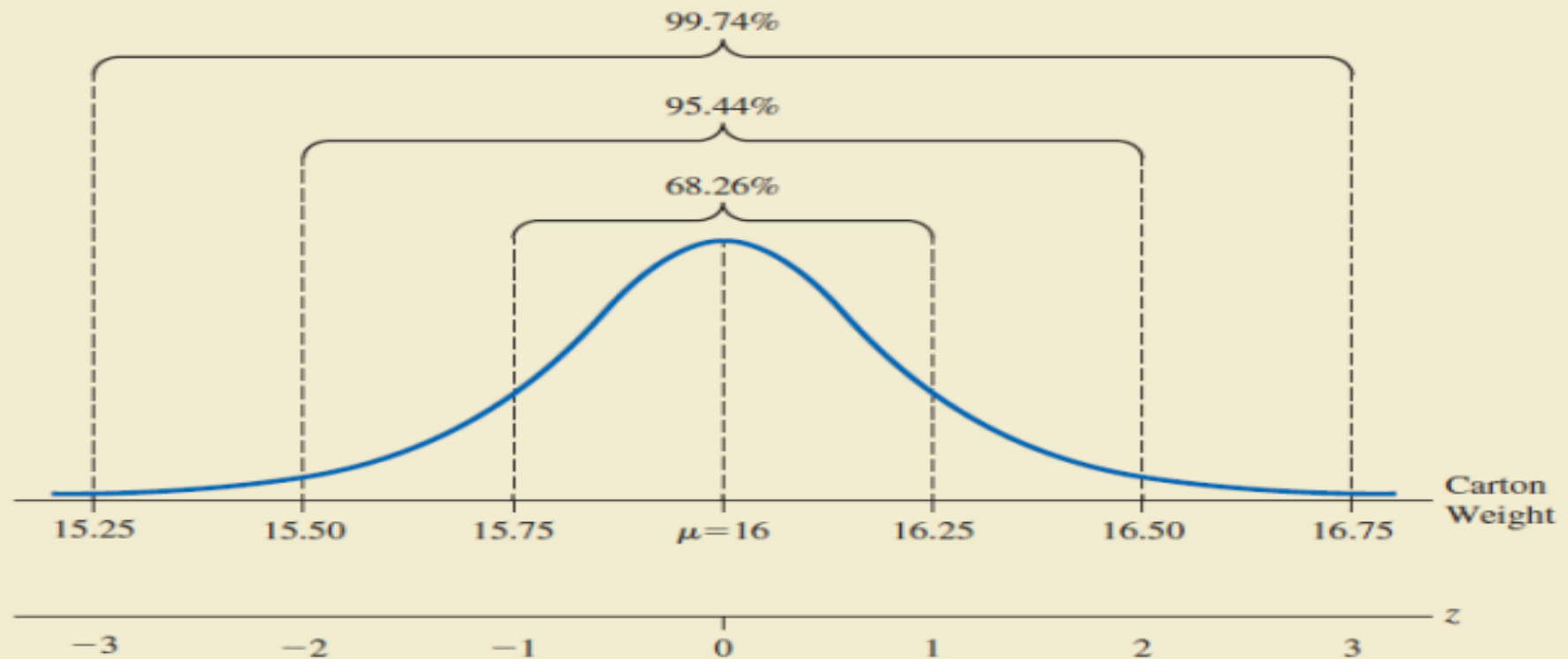
Normal distribution curve

- Data is symmetrically distributed with no skew.
- When plotted on a graph the data follows a bell shape with most values clustering around a central region.
- It is a probability distribution.
- The mean, median and mode are all equal.

Normal distribution curve

FIGURE 3.5 A BELL-SHAPED DISTRIBUTION OF DETERGENT CARTON WEIGHTS WITH PERCENTAGE OF DATA VALUES WITHIN 1, 2, AND 3 STANDARD DEVIATIONS

FIGURE 6.3



Inferential analysis

Hypothesis testing

- Deciding about the value of a parameter based on preconceived hypothesis.
- **Parametric tests**
- **Non-parametric test**

Estimation

- Estimating the value of the parameter
- **Point estimation**
- **Interval estimation**

Hypothesis

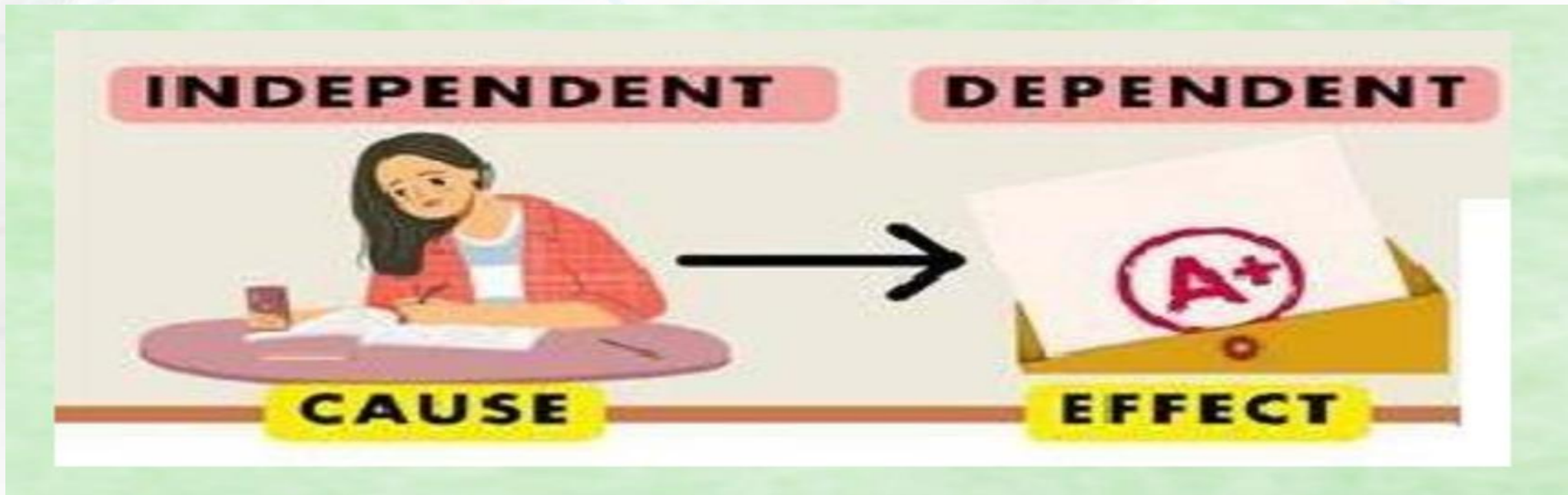
- A hypothesis is a tentative statement about the relationship between two or more variables.
- What the researchers predict the relationship between two or more variables?
- **Two types of hypothesis-**
 1. Null hypothesis
 2. Alternate / Research hypothesis

Null hypothesis

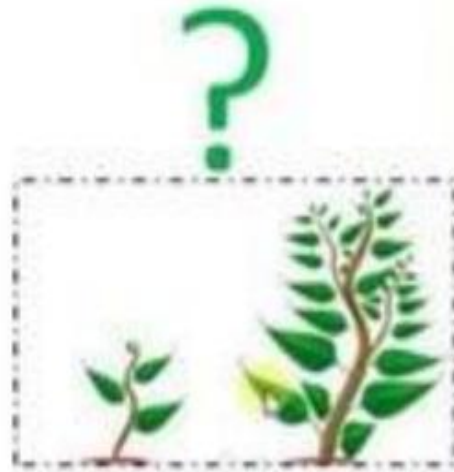
- There is no association/ relationship between two or more variables
- It is denoted by H_0

Alternate hypothesis

- There is association/ relationship between two or more variables
- It is denoted by H_a / H_1



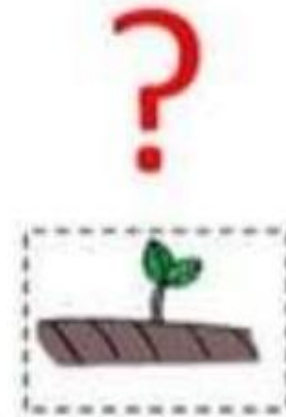
Hypothesis



H_1 : Application of bio-fertilizer 'x' increase plant growth.

Alternative hypothesis

✓ The alternative hypothesis is a hypothesis which the researcher tries to prove.



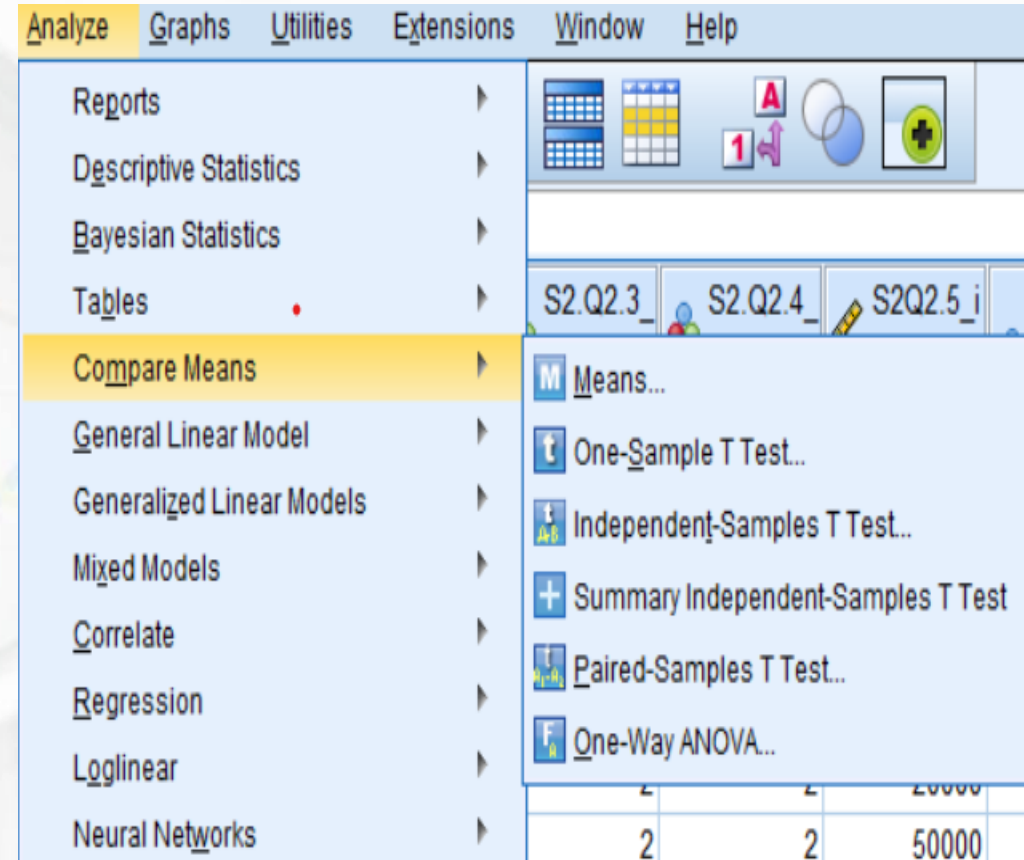
H_0 : Application of bio-fertilizer 'x' do not increase plant growth.

Null hypothesis

✓ The null hypothesis is a hypothesis which the researcher tries to disprove, or nullify.

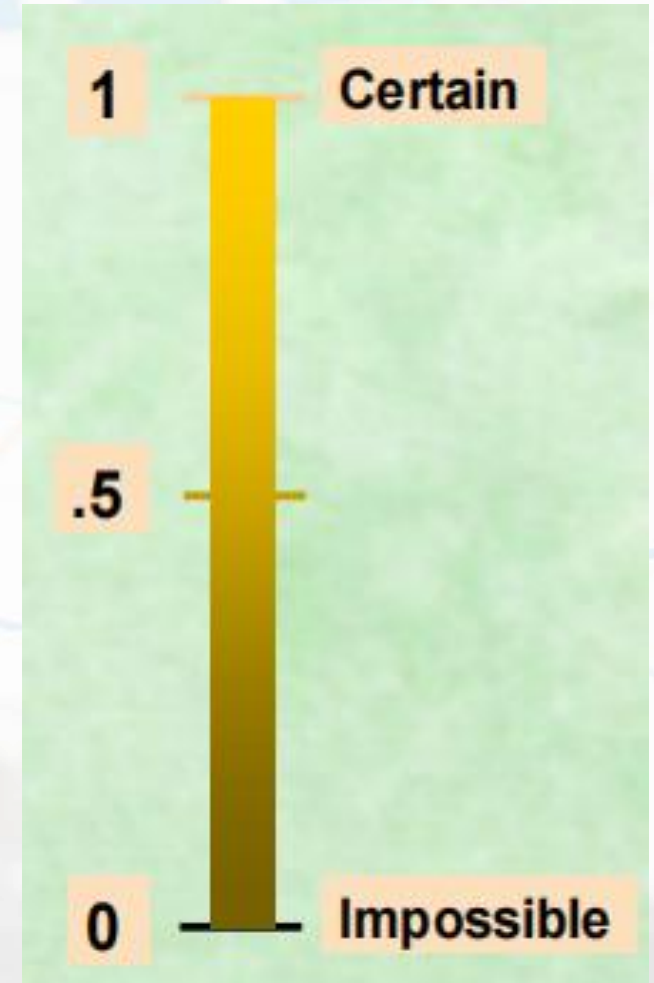
Interpretation of Hypothesis testing

- By p value
- By calculated test value such as t test, Z test, Chi-square test
- By critical value

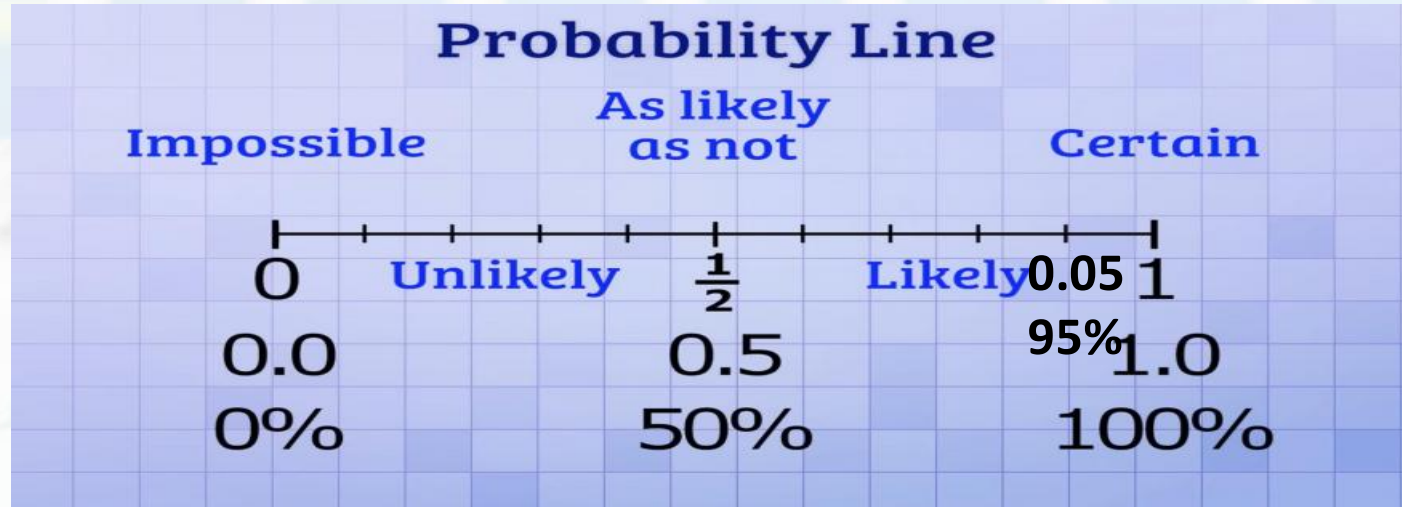


Probability

- Probability is a measure of the likelihood of an event to occur. Many events cannot be predicted with total certainty.
- We can predict only the chance of an event to occur.
- Probability can range from 0 to 1, where 0 means the event to be an impossible one and 1 indicates a certain event.



p value



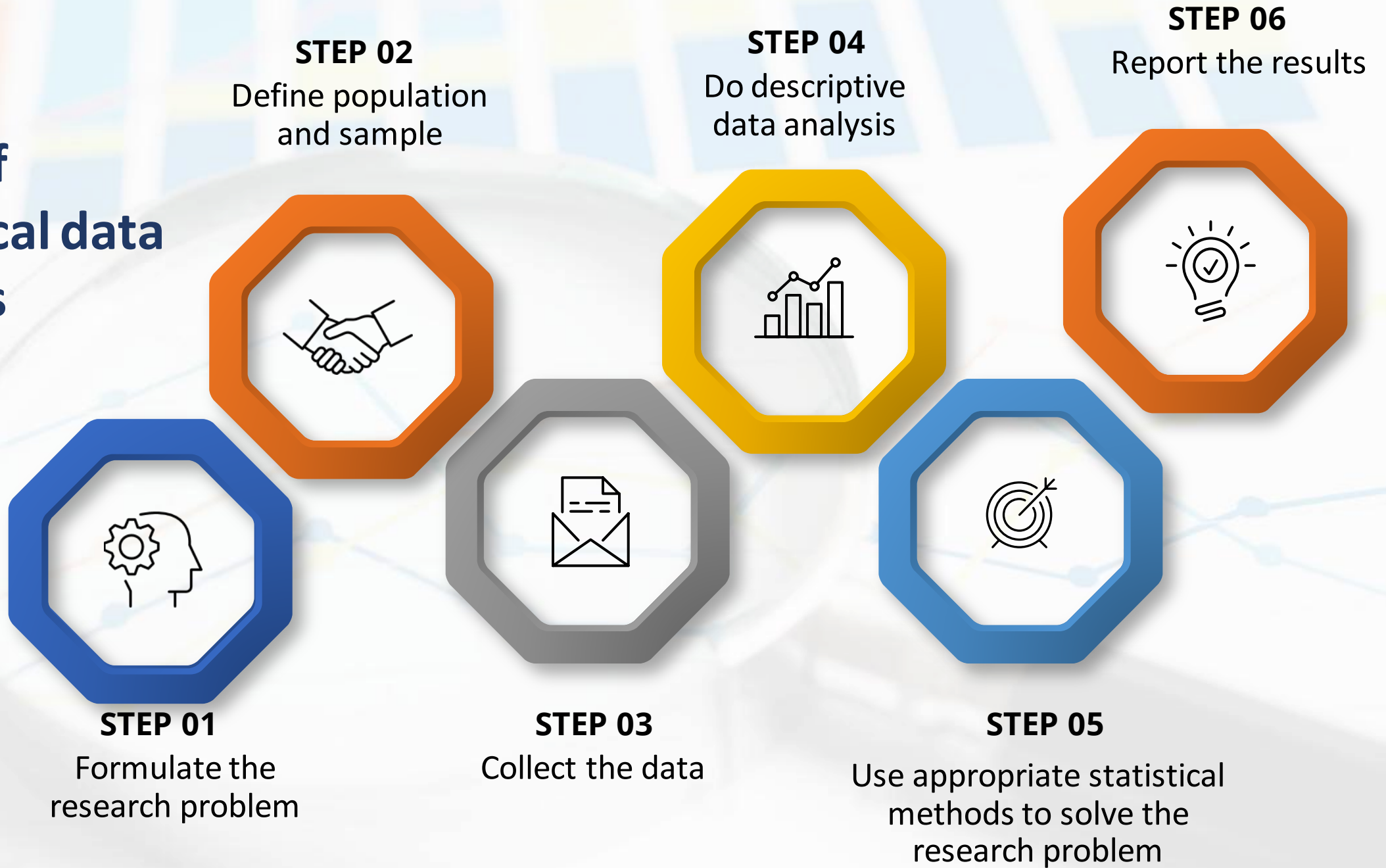
Small $p (\leq 0.05)$

- Reject the null hypothesis.
- Strong evidence.

Large $p (> 0.05)$

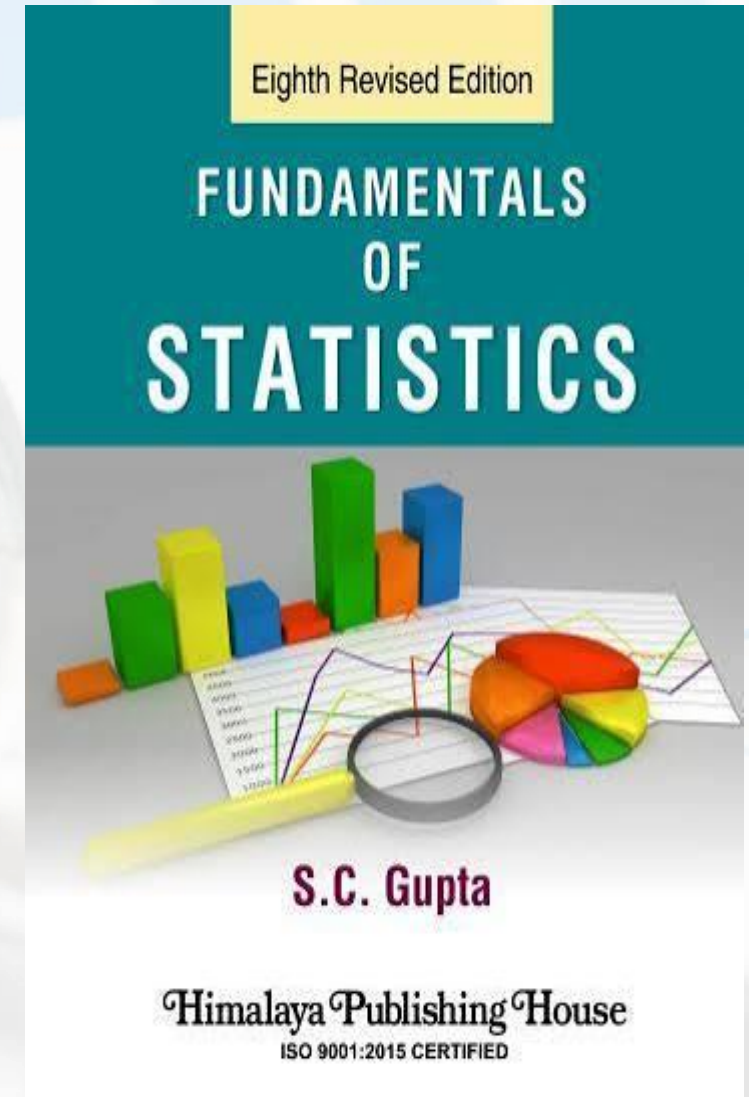
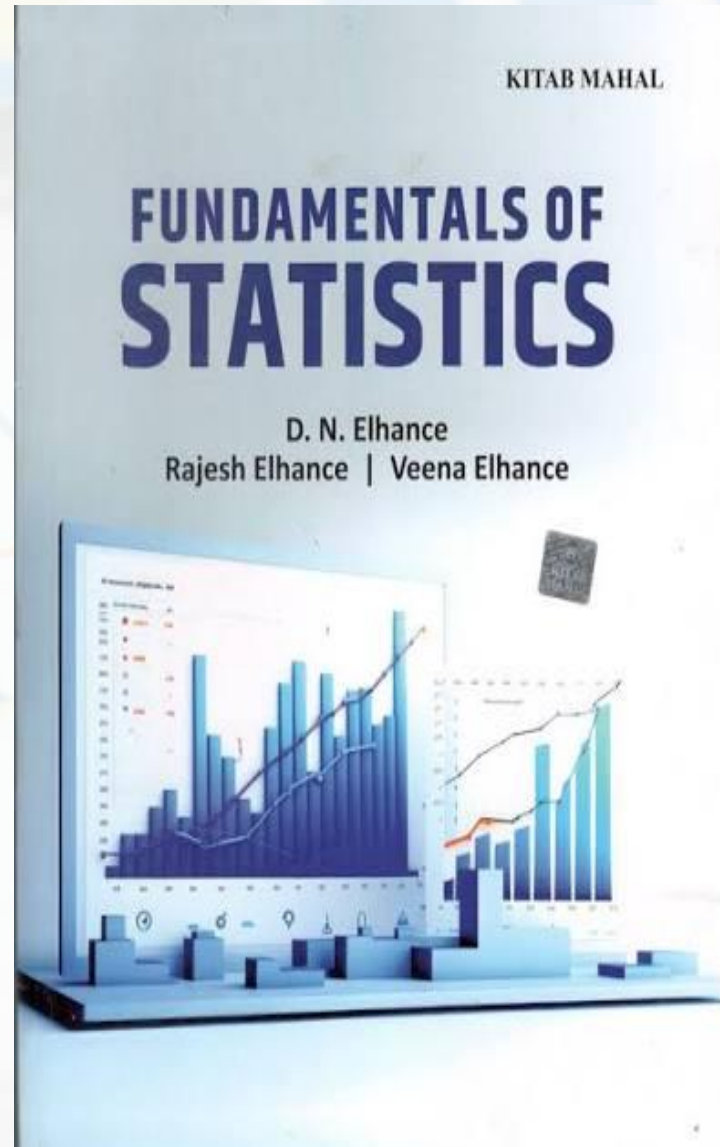
- Not reject the null hypothesis

Steps of Statistical data analysis



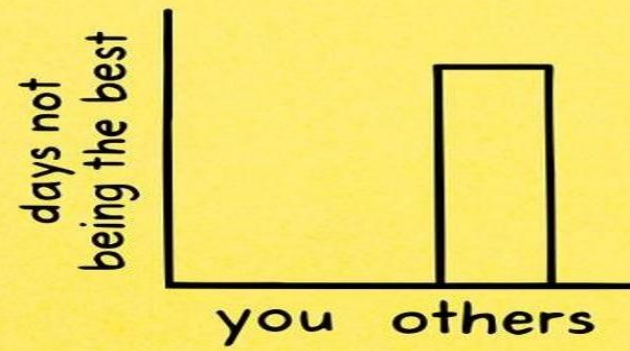
Take Home Messages

- Population: set of all measurements of interest
- Sample: Subset of population
- Mean: The mathematical average.
- Median: The exact middle value when ordered least to greatest.
- Mode: The most frequent value.
- Spread (Dispersion): How far apart data points lie.
- Variance & Standard Deviation: Average distance that data points fall from the mean.
- Probability: The chance an event happens, scored from 0 to 1.
- Distributions: The pattern of how data values scatter (like the bell-shaped normal distribution).
- Hypothesis: Tentative statement about the relationship between two or more variables

[illegible]

THANK YOU!





thank you for being
the best, bar none