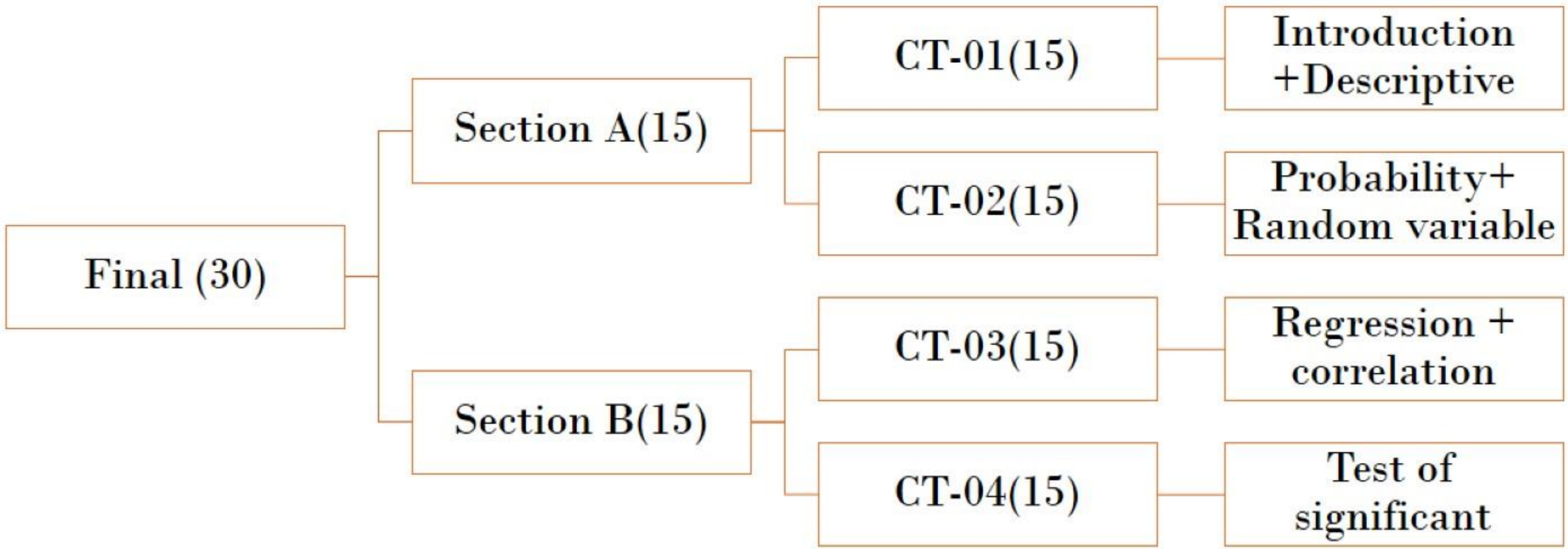
0542 02 Stat 3151
Statistics For Engineers

Credit: 03

Contents

Continuous Assessment outline



Reference book

- M. Sullivan III, Fundamentals of Statistics, 5th Edition, Pearson, 2017.
- W. Navidi, Statistics for Engineers and Scientists, 3rd Edition, McGraw Hill, 2010.

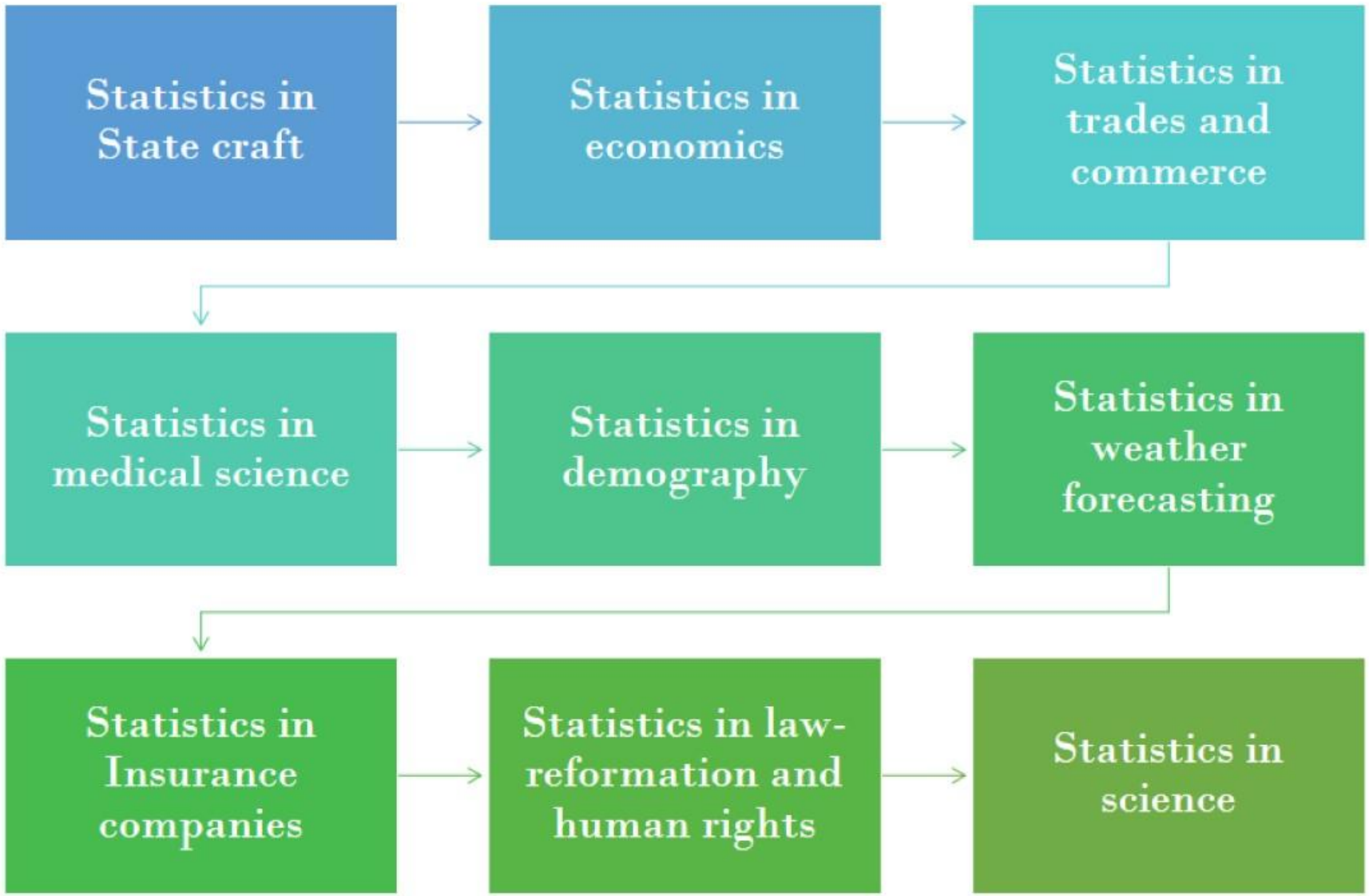
Introduction

Statistics is the science of collecting, organizing, presenting, analyzing and interpretation of the sample data to draw a meaningful conclusion about the population.

According to R.A Fisher, "The science of statistics is essentially a branch of applied mathematics and may be regarded as mathematics applied to observational data"

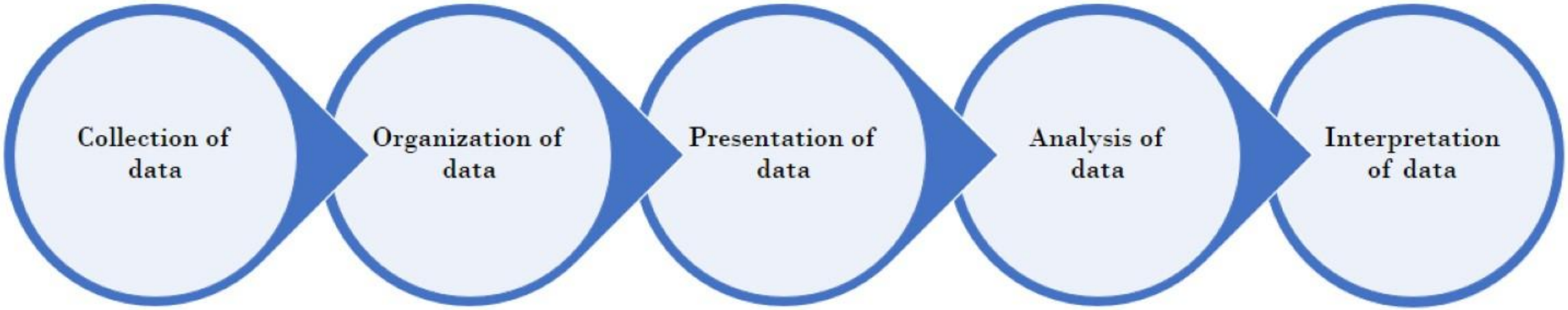
Introduction

Importance and scope of Statistics



Introduction

Function of Statistics



Introduction

Collection of data

- First step
- Faulty data, unreliable conclusion
- Collection of data-census, sampling, primary, secondary, inquiry

Organization of data

- Published source – organized form.
- Data must be edited to remove omissions, inconsistency, irreverent answer.

Presentation of data

- Tabulated
- Classified

Analysis of data

- Central tendency, dispersion, skew, kurtosis, correlation, regression, etc.

Introduction

Interpretation of data

- Last stage
- Valid conclusion
- Making wise decision

Example: App Crash Analysis

A mobile app development team wants to reduce the number of app crashes experienced by users.

Using Statistics:

Collecting: Crash reports are collected from 10,000 users, including device type, OS version, and crash time.

Organizing: The data is sorted into categories (e.g., Android vs. iOS, version numbers).

Presenting: A bar chart shows crash frequency by OS version.

Analyzing:

- The team finds that 60% of crashes occur on a specific OS version.
- They calculate the mean number of crashes per user and use standard deviation to measure variability.

Interpreting:
They discover that older devices running an outdated OS are the most crash-prone. The team then prioritizes a software patch for those versions.

Introduction

Objectives of Statistics

- **To Collect Data Systematically:**
  Gather accurate and relevant data in an organized manner.
- **To Organize and Summarize Data:**
  Present data in tables, charts, or graphs for easy understanding.
- **To Analyze Data:**
  Use statistical methods to examine data patterns, relationships, and trends.
- **To Interpret Results:**
  Draw meaningful conclusions from data analysis to aid decision-making.
- **To Make Predictions:**
  Use past data to forecast future events or trends.

Introduction

- **To Support Decision-Making:**
  Provide a scientific basis for decisions in business, science, and government.
- **To Identify Relationships and Correlations:**
  Understand how variables are related or influence each other.
- **To Estimate Population Parameters:**
  Make inferences about a whole population from sample data.

Why an engineer should learn about Statistics?

- **Data-Driven Decision Making:**
  Engineers often work with data from experiments, sensors, or production lines. Statistics helps them analyze this data to make informed and reliable decisions.
- **Quality Control and Improvement:**
  Statistical methods like control charts and hypothesis testing are essential to monitor and improve product quality, reducing defects and waste.
- **Design and Analysis of Experiments:**
  Engineers use statistics to design experiments efficiently, test variables, and optimize processes or product designs.

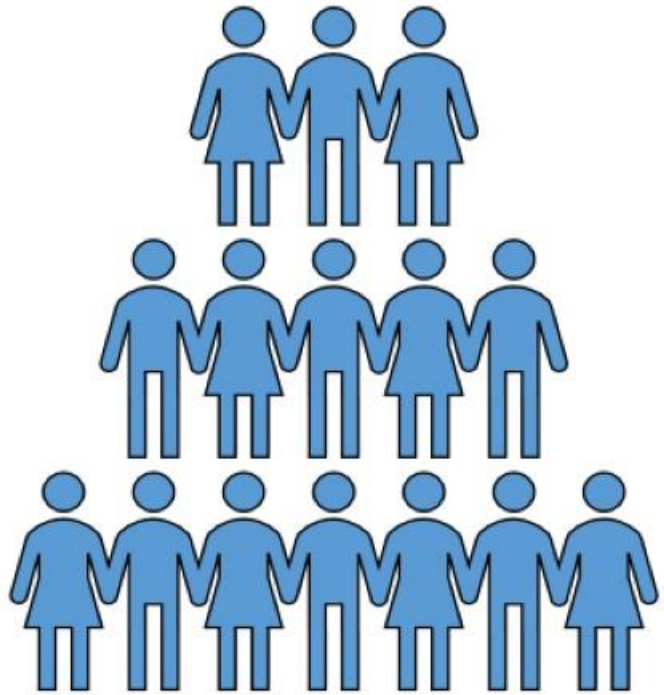
Introduction

- **Risk Assessment and Management:**
  Understanding probabilities helps engineers evaluate risks, predict failures, and develop safer systems.
- **Modeling and Simulation:**
  Many engineering problems involve uncertainties. Statistics provides tools to model variability and make predictions under uncertainty.
- **Problem-Solving Skills:**
  Statistical thinking helps engineers critically evaluate data, spot trends, and identify root causes of problems.
- **Communication of Results:**
  Engineers need to present findings clearly to stakeholders. Statistics aids in summarizing and visualizing complex data effectively.

Introduction

Population

Sample



Introduction

Introduction

Introduction

Definition

Population

-is a collection of all possible objects having certain characteristic under a study.

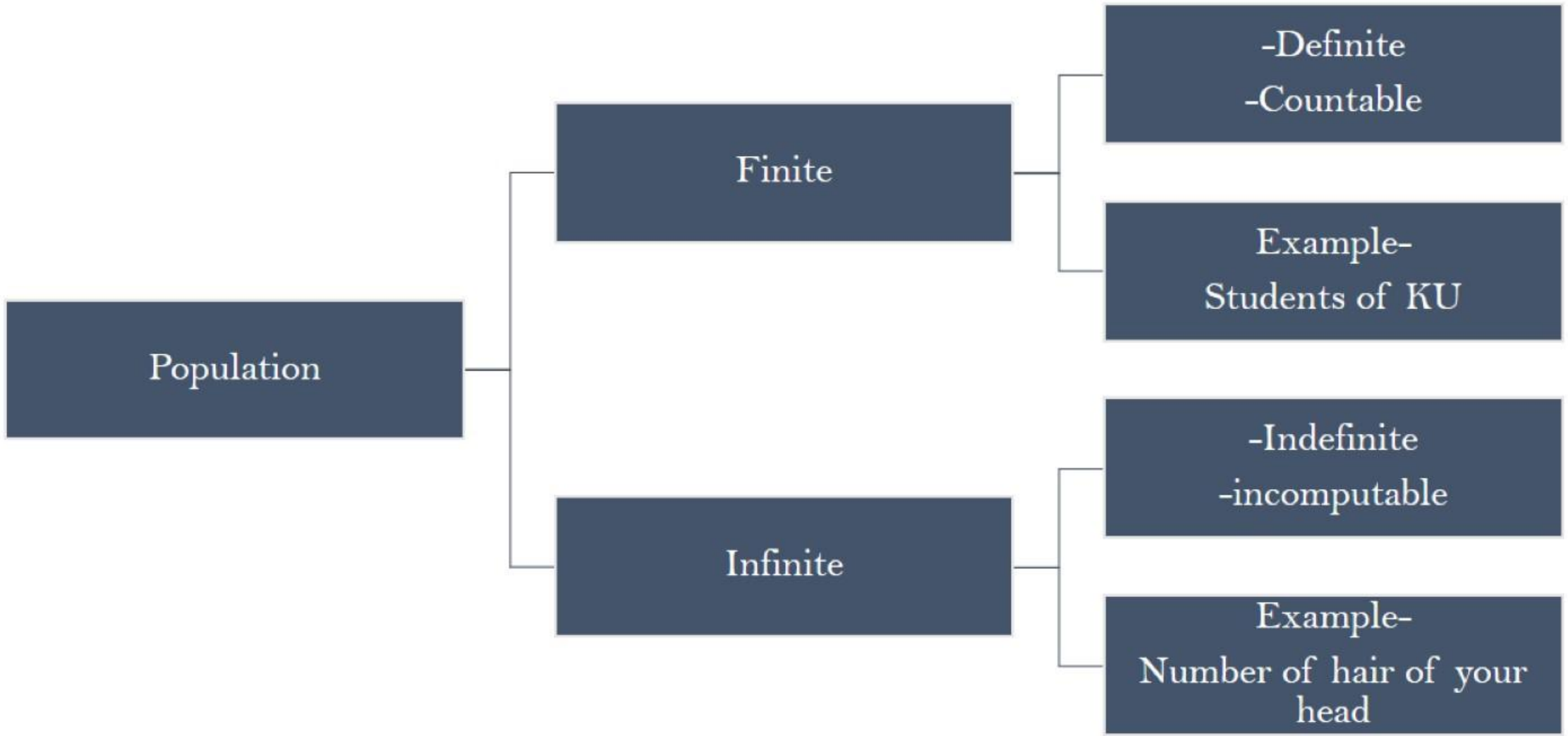
Example- All students in Khulna university.

Sample

-a part of the population selected for study.

Example- A housewife tests a small quantity of rice to see that it has been cooked or not. This sample quantity of rice is a sample and represents the entire quantity of rice cooked.

Introduction



Introduction

Distinguish between population and sample

Population	Points of difference	Sample
Population refers the totality of all the items or individuals having some specific characteristics.	**Definition**	A representative and considerably small part of a population is known as a sample of that population.
The number of elements of population can be finite or infinite.	**Number of elements**	The number of elements of sample can be finite.
Any statistical measure computed from population data is known as parameter.	**Statistical measure**	Any statistical measure computed from sample data is known as statistic.
Population can be hypothesis as the universal set.	**Set**	Sample can be hypothesis as a subset of population.
All the students of Khulna university constitute a population.	**Example**	A group of 30 students from 5000 students of Khulna university constitute a sample

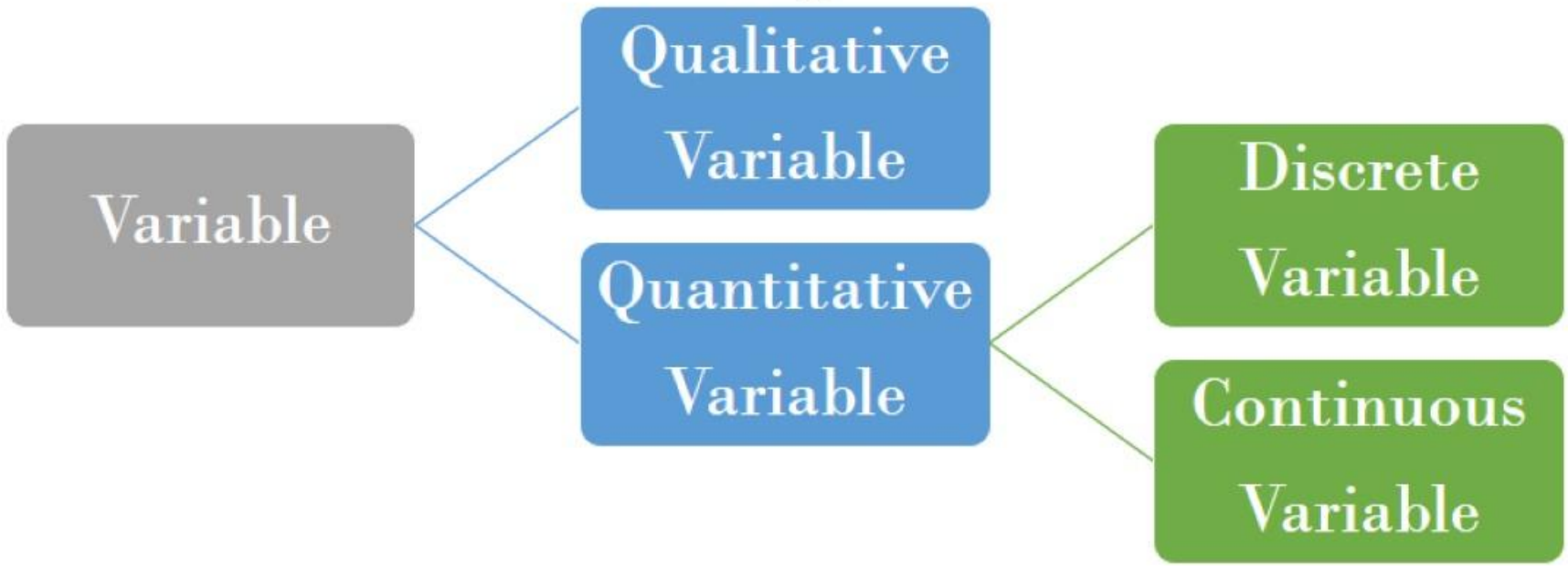
Introduction

Variable

The characteristic which varies over the units.

Example: Name, Age, Gender

Classification of variable according to characteristics



Introduction

Qualitative variable:

- Not expressed in numerical form or in numbers.
- Example: color, gender, occupation

Quantitative variable:

- Expressed in numerical form or in numbers.
- Example: Monthly income, age, height

Discrete variable

- Possesses isolated or integral value
- Example: Family size, number of road accident

Continuous variable:

- Takes value within a range or limit
- Example: Height, weight.

Introduction

Constant

A numerical characteristic which does never change or vary it's value.

Example: $\pi=3.1416$

Introduction

Data source

Primary data

- Direct observation
- Collected for the first time
- Original in character
- Not well organized

Secondary data

- Already obtained by some other persons or organization
- Already publish
- Well organized

Descriptive Statistics

Representation of data

(1) Tabular representation
(2) Stem and leaf display
(3) Graphical representation

(1) Tabular representation

- Data is an orderly and logical listing of a quantitative data.
- Arranged in rows(horizontal) and columns(vertical)

Frequency

The repeated time of a value of the variable

Frequency distribution

Listing of a data set which divides the data in different classes;
Gives a count number of observations in each class.

Descriptive Statistics

- Types of frequency distribution

(1) Discrete(ungrouped) frequency distribution

(2) Continuous(grouped) frequency distribution

Construction of discrete frequency distribution

Step-1: Succession of data

Step-2: Tally Marks

Step-3: Frequency

Descriptive Statistics

Construction of continuous frequency distribution

Step-1: Determine the range

Step-2:Determination of the number of class

Step-3:Determination of the class interval

Step-4: Determination of the class limit

Step-5: Tally and frequency

Descriptive Statistics

(2)Stem and Leaf display

- Represent quantitative data
- Examine shape of the frequency distribution

Descriptive Statistics

(3)Graphical representation

(i) Histogram

(ii) Frequency polygon

(iii) Frequency curve

(iv) Cumulative frequency curve or ogive curve

(v) Pie chart

Descriptive Statistics

Location (Central Tendency)

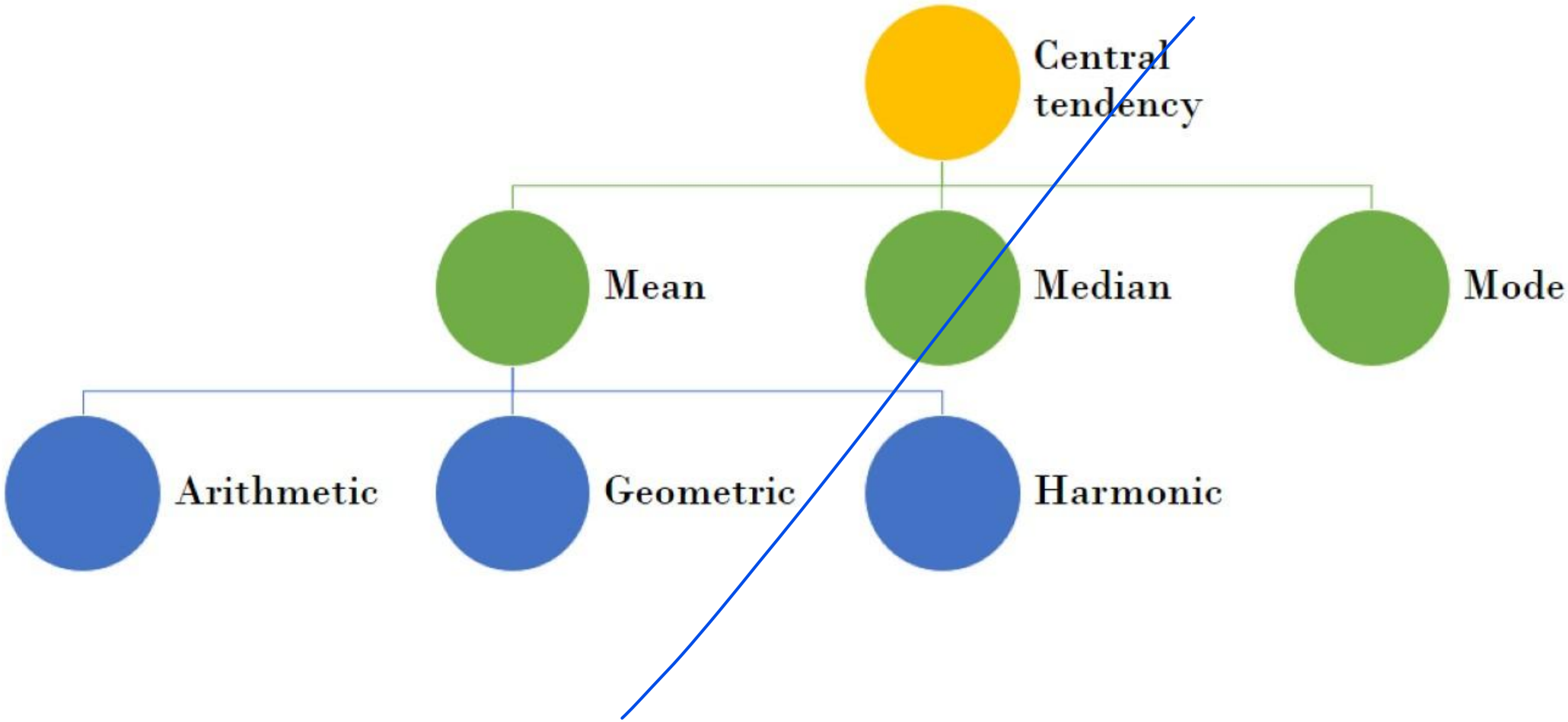
Measure of central tendency is a single value that summarizes a set of data. It locates the center of the value.

Example: It is hardly possible to memorize the height of B.Sc(Hons) students of Khulna university. But the average value in this case can express the overall height.

Purposes of central tendency

- Describing the position or location of a set of observations or values.
- Comparing two or more groups
- Picture of a complete universe

Descriptive Statistics



Descriptive Statistics

Arithmetic mean

(i)Ungroup data (When individual observations are given)

$$\mathrm{AM}(\bar{X}) = \frac{x_1+x_2+\cdots+x_n}{n} = \frac{\sum_{i=1}^{n} x_i}{n}$$

(ii) Group data (When individual observations with corresponding frequencies are given)

$$\mathrm{AM}(\bar{X}) = \frac{f_1x_1+f_2x_2+\cdots+f_nx_n}{N} = \frac{\sum_{i=1}^{n} f_1x_i}{n}$$

Weighted arithmetic mean or weighted mean

$$\bar{X}_w = \frac{w_1x_1 + w_2x_2 + \cdots + w_nx_n}{W_1 + W_2 + \cdots + W_n} = \frac{\sum_{i=1}^{n} w_ix_i}{\sum_{i=1}^{n} w_i}$$

Descriptive Statistics

Geometric mean

(i)Ungroup data

$$G = (x_1 \times x_2 \times \cdots \times x_n)^{1/n}$$

(ii)Group data

$$G = ({x_1}^{f_1} \times {x_1}^{f_2} \times \cdots \times {x_n}^{f_n})^{1/N}$$

Harmonic mean

(i) Group data

$$\mathrm{HM} = \frac{n}{\frac{1}{x_1} + \frac{1}{x_2} + \cdots + \frac{1}{x_n}}$$

(i) Ungroup data

$$\mathrm{HM} = \frac{n}{\frac{f_1}{x_1} + \frac{f_2}{x_2} + \cdots + \frac{f_n}{x_n}}$$

Descriptive Statistics

Median

(i) Ungroup data

When n is odd, $\text{Me} = \frac{n+1}{2}$

When n is even, $\text{Me} = \frac{\frac{n}{2}th\ item + \left(\frac{n}{2}+1\right)th\ item}{2}$

(ii) Group data

$$Me = L + \frac{\frac{N}{2} - F_c}{F_m} \times h$$

Descriptive Statistics

Mode

- The observation which occurs most frequently in a set of data.
- Mode is repeated maximum times
- Obtain the highest density in the data

(i) Ungroup data

observation which occurs most frequently in a set of data

(i) Group data

$$Mo = L + \frac{\Delta_1}{\Delta_1 + \Delta_2} \times h$$

Descriptive Statistics

Situation to use the measures of central tendency

Arithmetic mean

When there is no extreme values in any set of information

Geometric mean

Averaging ratios and percentages

Computing average rates of increase or decrease

Harmonic mean

Speed, price

Median

When extreme values in any data

Mode

Deal with qualitative measure rather than quantitative.

Descriptive Statistics

Theorem:

- Arithmetic mean effected by the extreme value of the data series.
- AM=GM=HM
- For any two non-zero positive numbers, $AM \times HM = GM^2$
- For two non-zero positive numbers, $AM \geq GM \geq HM$

Descriptive Statistics

Measure of Dispersion

Subject	Management	Marketing	Statistics	Geography	English	Accounting	Average
Student A	48	50	52	51	49	50	50
Student B	1	2	100	99	98	0	50

Dispersion:

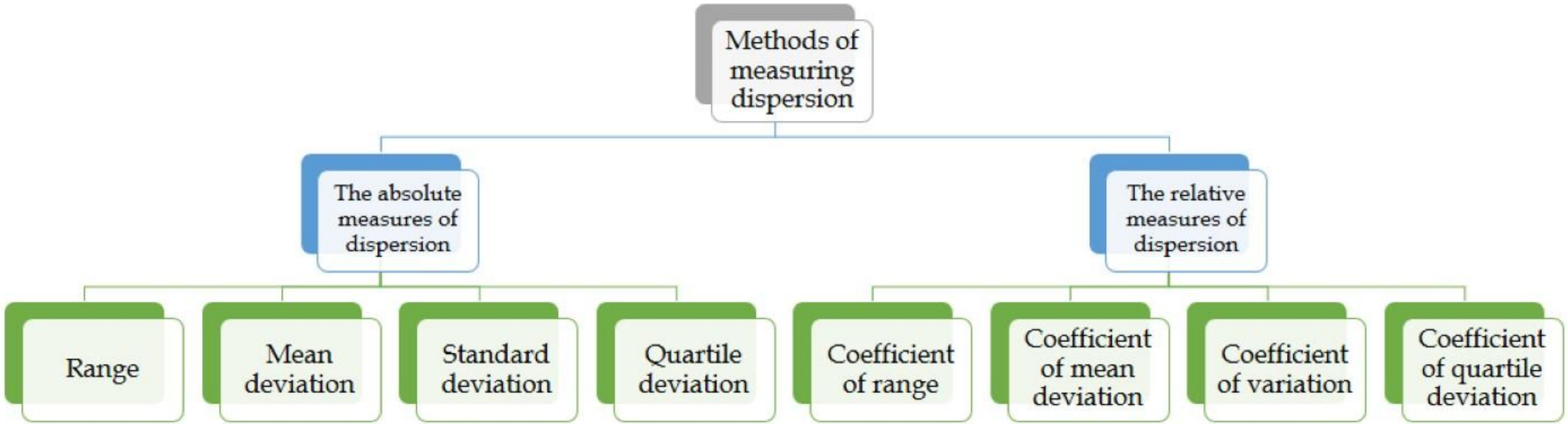
- The distance of different values from the central values.
- The variation of the items around an average.
- Central tendency does not give us how spreated the observations.
- Measure of the variation of the items.
- The degree to which numerical data tend to spread about an average value.
- The measurement of the scatterness of the mass of figures in a series about an average.
- Degree of scatter shown by observation.

Descriptive Statistics

Significance and necessity of measuring dispersion:

- To realize the reliability of the measures of central tendency
- To compare the variability of two or more sets of data
- To suggest various methods for controlling the variations in a set of observations.
- To facilitate as a basis for further statistical analysis.

Descriptive Statistics



Descriptive Statistics

Descriptive Statistics

Descriptive Statistics

Descriptive Statistics

Descriptive Statistics

Descriptive Statistics

Descriptive Statistics