## Lab-I

## Aim : To study measures of central tendency for grouped and ungrouped data

## Introduction

Statistics has been defined differently by different authors and each author has assigned new limits to the field which should be included in its scope. We can do no better than give selected definitions of statistics by some authors and then come to the conclusion about the scope of the subject.

*A.L. Bowley* defines, “Statistics may be called the science of counting”. At another place he defines, “Statistics may be called the science of averages”. Both these definitions are narrow and throw light only on one aspect of Statistic

According to King, “The science of statistics is the method of judging collective, natural or social, phenomenon from the results obtained from the analysis or enumeration or collection of estimates.

The word **statistics has two meanings:**

• In the more common usage, *statistics refers to numerical* facts. The numbers that represent the income of a family, the age of a students……

• The second meaning of *statistics refers to the field or* discipline of study.

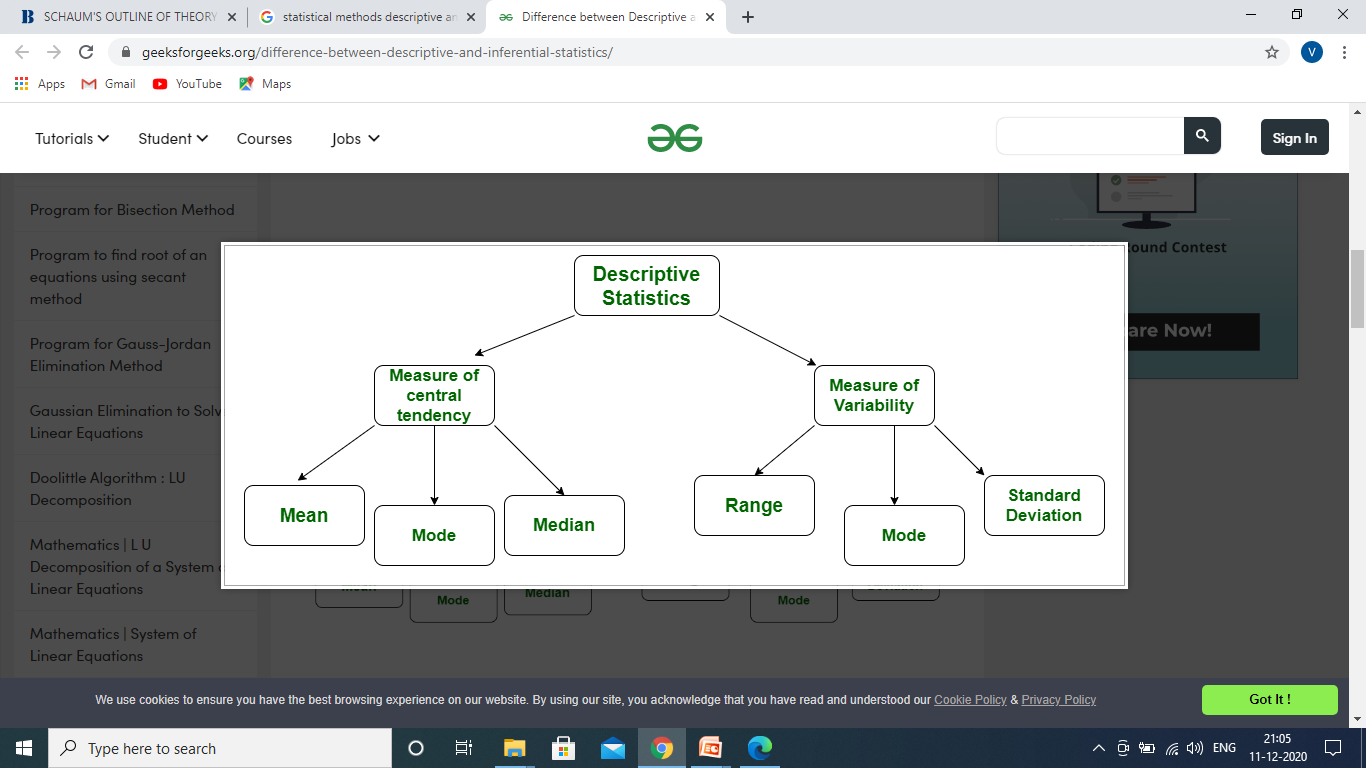
Usually these decisions are made under conditions of uncertainty. Decisions made by using

statistical methods are called *educated guesses.* Decisions made without using statistical

(or scientific) methods are *pure guesses and, hence, may prove to be* unreliable.

Example:- opening a large store in an area with or without assessing the need for it may affect its success.

There are two types of statistics (i) Descriptive statistics: *Descriptive**statistics consists of methods for* organizing, displaying, and describing data by using tables, graphs, and summary measures



## (ii) Inferential Statistics: A major portion of statistics deals with making decisions, inferences, predictions, and forecasts about populations based on results obtained from samples.

## For example: we may make some decisions about the political views of all college and university students based on the political views of 1000 students selected from a few colleges and universities.

The collection of data that are relevant to the problem being studied is commonly the most difficult, expensive, and time-consuming part of the entire research project.Statistical data are usually obtained by counting or measuring items.

* + **Primary data** are collected specifically for the analysis desired
  + **Secondary data** have already been compiled and are available for statistical analysis

A **variable** is an item of interest that can take on many different numerical values. A variable is often represented by letters like A,B,C,....

Statistical data are usually obtained by counting or measuring items. Most data can be put into the following categories:

**Qualitative -** data are measurements that each fail into one of several categories. (hair color, ethnic groups and other attributes of the population)

**Quantitative** - data are observations that are measured on a numerical scale (distance traveled to college, number of children in a family, etc.)

There are several techniques for describing a data graphically are

* Frequency distribution
* Histogram
* Bar graphs
* Pie Chart

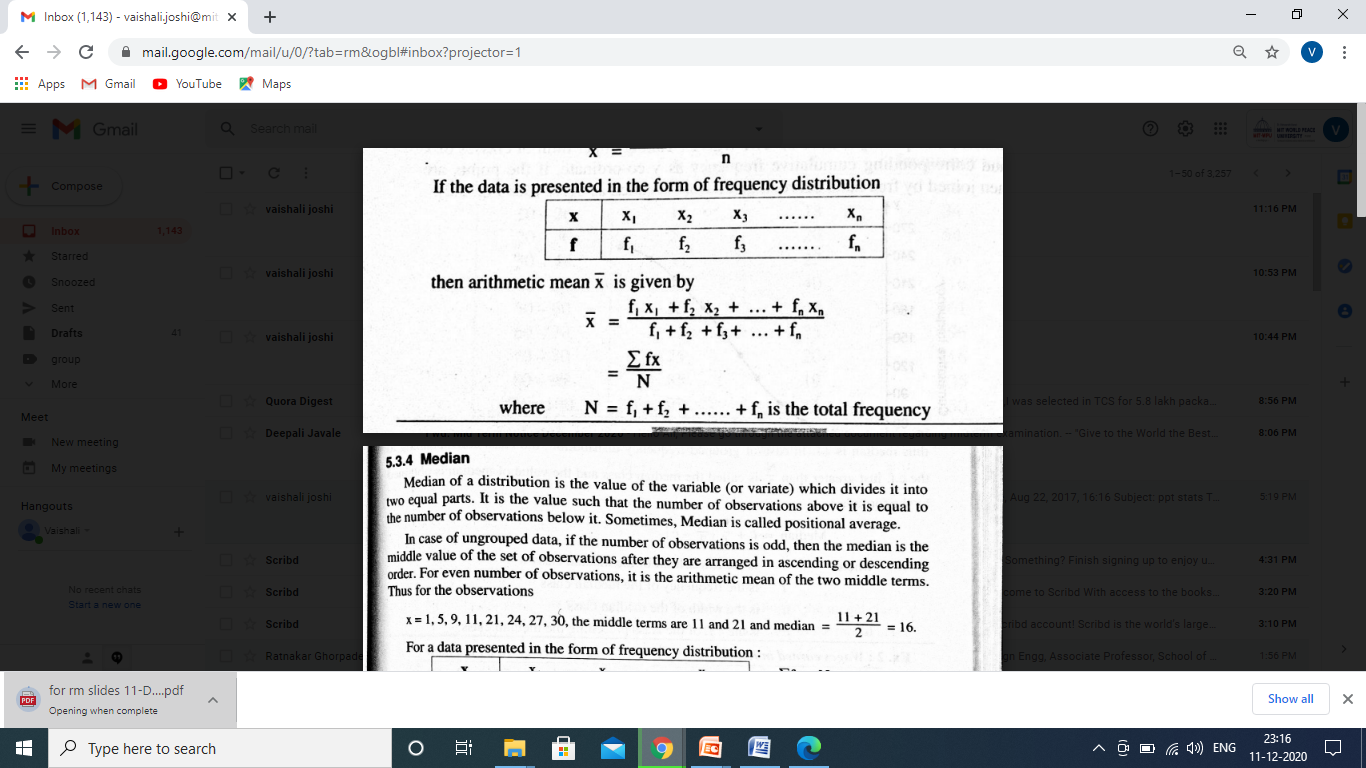
Measures of Central tendency: Average is the value which is the representative of a set of data.

**(i) Arithmetic Mean**

**a. Discrete type or ungrouped type data set**



**b. Discrete frequency distribution type**



**C. Continuous frequency distribution type or Grouped type data set:** For the given continuous type of data set , first find the mid values of each class interval and then use the procedure done in (b)

(ii) Median:

1. **Discrete type data:** Median of a distribution is the value of the variate which divides the given data set into two equal parts .

If the number of observations n=odd , middle value of the variate is median

Ex:- 1 , 2, 3, 4 , 5

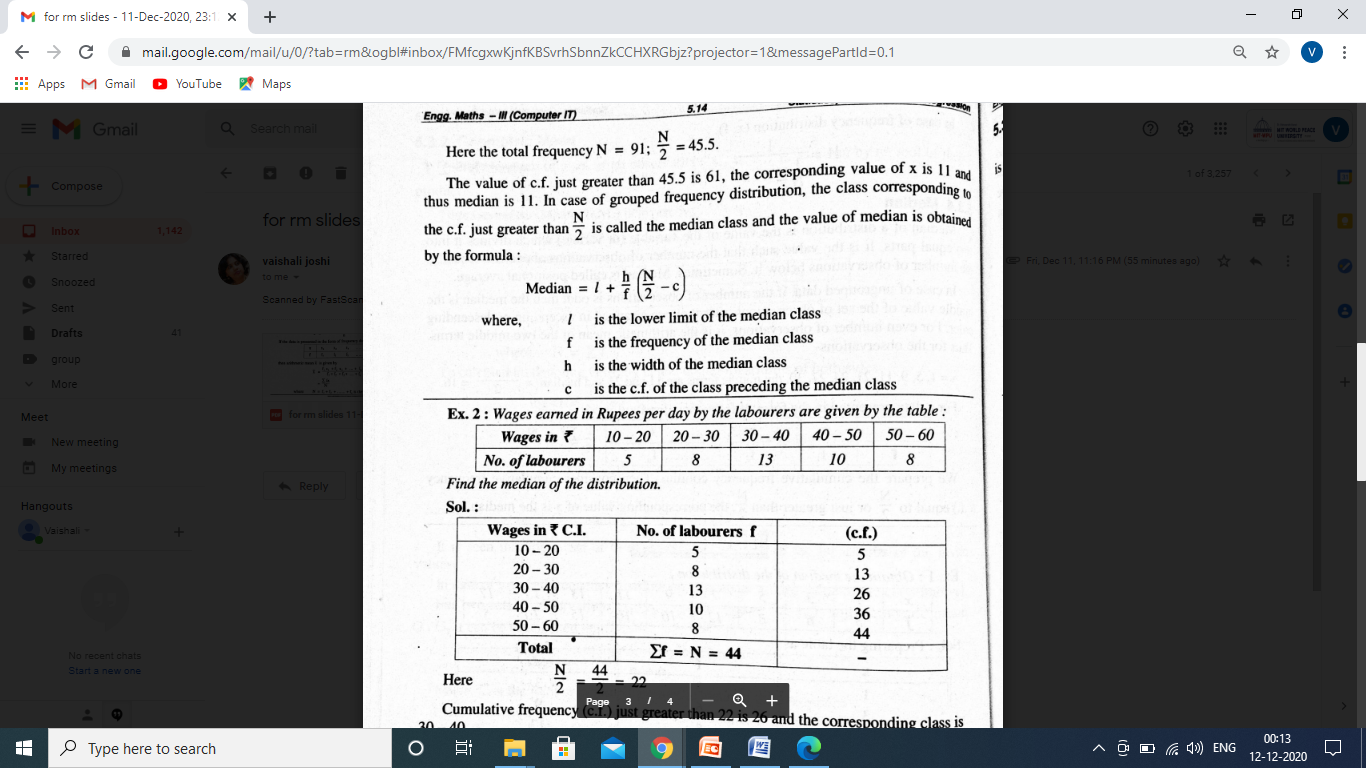
Median=3

If n=even, average of two middle terms is the median.

Ex:- 1 , 2, 3, 4 , 5, 6

Median=(3+4)/2=3.5

1. **Discrete frequency type:** For given frequency type distribution. First prepare cumulative frequency column. Find the value of N/2. Consider cumulative frequency equal to (N/2) or just greater than (N/2), the corresponding value of variate is Median.
2. **Continuous frequency distribution type or Grouped type data set:**



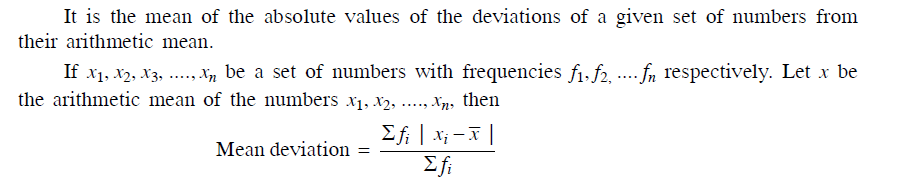
(iii) **Mode:** It is the value of the variate which occurs most frequently in a set of observations

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **x** | **1** | **6** | **7** | **9** | **11** | **15** |
| f | 5 | 12 | 11 | 23 | 4 | 11 |

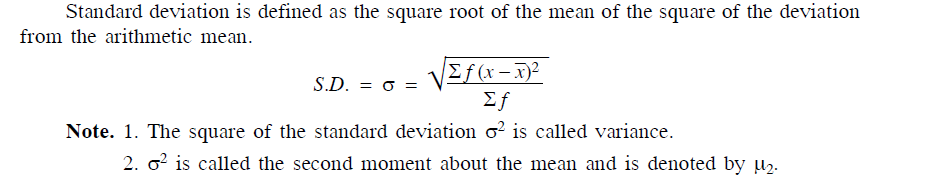
Here variate “9” occurs maximum number of times. Hence Mode = 9

**Dispersion**

**(i)Mean Deviation:**

****

**(ii) Standard Deviation:**

****

**Using R programming attempt the following**

1. Find averages for grouped and ungrouped type of data.
2. Plot Histogram, Frequency distribution curve.

Ex :- Wages per day in Rupees of workers are given below

|  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| X | 0-10 | 10-20 | 20-30 | 30-40 | 40-50 | 50-60 | 60-70 | 70-80 | 80-90 | 90-100 |
| F | 21 | 12 | 14 | 12 | 15 | 25 | 14 | 7 | 4 | 10 |

Find Arithmetic Mean, Median , Mode, Standard deviation . Also plot the frequency curve

1. Find standard deviation of given data points.
2. To check which data is more consistent (Coefficient of Variation)
3. Input number of observations n
4. Input data points for i=1,……,n
5. Use formula

Ex:- Reliance Ltd. Is actively considering two mutually exclusive projects

(Project A and Project B) for adoption.

|  |  |  |
| --- | --- | --- |
| Year | Project A | Project B |
| Cost Profit ( Rs. In Lakh.) | Cost Profit ( Rs. In Lakh.) |
| 1 | 10 | 5 |
| 2 | 5 | 25 |
| 3 | 20 | 45 |
| 4 | 40 | 30 |
| 5 | 60 | 30 |

Which of the two is more risky project (Use coefficient of variation)?