

BUSINESS TOOLS FOR DECISION MAKING

BY

A. VICTORIA SUBHA & D. KOWSALYA

GUEST LECTURER IN COMMERCE

K.N.GOV'T. ARTS COLLEGE (W) AUTONOMOUS, THANJAVUR

REFERENCE : STATISTICAL METHODS,

S.P.GUPTHA,SULTANCHAND&SONS,NEW DELHI.

UNIT – 1

STATISTICS:

**MEANING-DEFINITION-OBJECTIVES-CLASSIFICATION- TABULATION-
DIGRAMS-GRAPHS-MEASURES OF CENTRAL TENDANCY- ARITHMETIV MEAN
– MODE – GEOMETRIC MEAN –HARMONIC MEAN.**

Meaning:

“Statistics”, that a word is often used, has been derived from the Latin word ‘Status’ that means a group of numbers or figures; those represent some information of our human interest.

We find statistics in everyday life, such as in books or other information papers or TV or newspapers.

Although, in the beginning it was used by Kings only for collecting information about states and other information which was needed about their people, their number, revenue of the state etc.

But now-a-days due to its pervading nature, its scope has increased and widened. It is now used in almost in all the fields of human knowledge and skills like Business, Commerce, Economics, Social Sciences, Politics, Planning, Medicine and other sciences, Physical as well as Natural.

Definition:

“Statistics has two meanings, as in plural sense and in singular sense”. —Oxford Dictionary

In plural sense, it means a systematic collection of numerical facts and in singular sense; it is the science of collecting, classifying and using statistics.

“Statistics are numerical statements of facts in any department of enquiry placed in relation to each other.” —**A.L. Bowley**

“By statistics we mean aggregates of facts affected to a marked extent by multiplicity of causes, numerically expressed, enumerated or estimated according to reasonable standard of accuracy, collected in a systematic manner for a predetermined purpose, and placed in relation to each other.”—**Horace Sacrist**

“Statistics refers to the body of technique or methodology, which has been developed for the collection, presentation and analysis of quantitative data and for the use of such data in decision making.” —**Nettor and Washerman**

“Statistics may be defined as the collection, presentation, analysis, and interpretation of numerical data.” —**Croxton and Cowden**

Types of Statistical Method:

Statistical methods are of two types:

- **Descriptive Method:** This method uses graphs and numerical summaries.
- **Inferential Method:** This method uses confidence interval and significance test which are part of applied statistics.

Basic Terms of Statistics:

Population:

Population refers to the large group of individuals or objects from where a researcher or investigator starts his/her research or investigation.

Sample:

Sample refers to a randomly selected set of individuals or objects from the population.

Parameters and Statistics:

1. An unknown numerical briefing of the population is a parameter.
- 2..A known numerical briefing of the sample, used in making inference about the parameter is called statistics.
3. A statistic illustrates a sample whereas a parameter illustrates the population from where the sample was selected.

Basics of Descriptive Statistics:

- **Mean:**

The mean also referred to as average is calculated by dividing the total value or sum of the examined values by the number of the observations.

For example: 20, 30,60, 80, 90

Mean/Average= $(20+30+60+80+90)/5=280/5=56$

- **Standard Deviation:**

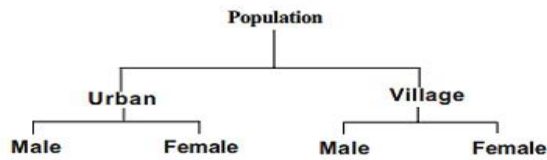
The standard deviation provides a concept of the closeness of the whole set of data to the mean or the average value. A small standard deviation denotes that the data sets are tightly grouped. A large standard deviation denotes that the entire data sets are scattered to a wide range of values.

- **Variance:**

The variance calculates the extent to which a set of data is spread out. A zero variance denotes that all the values are similar. Variance is the squared value of the standard deviation, so it can never be negative. A small variance denotes that the data points are very close to the average value or the mean. On the other hand, a large variance denotes that the data points are very widely spread around the average value or the mean.

Classification of Data

Classification is the process of arranging the collected data into classes and to subclasses according to their common characteristics. Classification is the grouping of related facts into classes. E.g. sorting of letters in post office



E.g. People under study can be classified as follows.

States	Production of food grains (in '000 tons)
Tamil Nadu	4500
Karnataka	4200
Andhra Pradesh	3600

Year	Profits (in 000 Rupees)
2001	72
2002	85
2003	92
2004	96
2005	95

Example: Profits of a company from 2001 to 2005.

\Types of classification

There are four types of classification. They are

Geographical classification

Chronological classification

Qualitative classification

Quantitative classification

(i) Geographical classification

When data are classified on the basis of location or areas, it is called geographical classification

Example: Classification of production of food grains in different states in India.

States Production of food grains
(in '000 tons)

Tamil Nadu 4500

Karnataka 4200

Andhra Pradesh 3600

(ii) Chronological classification

Chronological classification means classification on the basis of time, like months, years etc.

Year Profits (in 000 Rupees)

2001 72

2002 85

2003 92

2004 96

2005 95

Example: Profits of a company from 2001 to 2005.

Profits of a company from 2001 to 2005

(iii) Qualitative classification

In Qualitative classification, data are classified on the basis of some attributes or quality such as sex, colour of hair, literacy and religion. In this type of classification, the attribute under study cannot be measured. It can only be found out whether it is present or absent in the units of study.

(iv) Quantitative classification

Quantitative classification refers to the classification of data according to some characteristics, which can be measured such as height, weight, income, profits etc.

Example: The students of a school may be classified according to the weight as follows

Weight (in kgs)	No of Studentms
40-50	50
50-60	200
60-70	300
70-80	100
80-90	30
90-100	20
Total	700

Frequency Distribution of a Discrete Variable

There are two types of quantitative classification of data. They are

Discrete frequency distribution

Continuous frequency distribution

Since, a discrete variable can take some or discrete values within its range of variation, it will be natural to take a separate class for each distinct value of the discrete variable as shown in the following example relating to the daily number of car accidents during 30 days of a month.

3 4 4 5 5 3

4 3 5 7 6 4

4 3 4 5 5 5

5 5 3 5 6 4

5 4 4 6 5 6

Table No. 2: Showing frequency distribution for daily number of car accidents during a month.

Number of car accidents	Tallys	Frequency
3	///	5
4	/// III	9

5	II I	11
6	III	4
7	I	1
Total		30

Frequency Distribution of a Continuous Variable

For a continuous variable if we take a class for each distinct value of the variable, the number of classes will become unduly large, thus defeating the purpose of tabulation. In fact, since a continuous variable can assume an infinite number of values within its range of variation, the classification or sub-division of such data is necessarily artificial. Some guidelines that should be followed while dividing continuous data into classes are as follows:

1. The classes should be mutually exclusive, i.e., non-overlapping. No two classes should contain the same interval of values of the variable.
2. The classes should be exhaustive, i.e., they must cover the entire range of the data.
3. The number of classes and the width of each class should neither be too small nor too large. In other words, there should be relatively fewer classes if the difference between the least value of the variable and its highest value is small and relatively more classes if the same difference is large. This difference between the least value of the variable and the greatest value of the variable is called the range of the variable or the data set.
4. The classes should, preferably, be of equal width.

Let us consider the following example regarding daily maximum temperatures in $^{\circ}\text{C}$ in a city for 50 days.

28 28 31 29 35 33 28 31 34 29
 25 27 29 33 30 31 32 26 26 21
 21 20 22 24 28 30 34 33 35 29
 23 21 20 19 19 18 19 17 20 19
 18 18 19 27 17 18 20 21 18 19

Minimum Value= 17

Maximum Value=35

Range=35-17=18

Number of classes=5 (say)

\therefore width of each class=4

Table No. 3: Showing frequency distribution of temperature in a city for 50 days.

Class Intervals (Temperatures in $^{\circ}\text{C}$)	Tallys	Frequency
17-20	II	17
21-24	II	7
25-28	II	10

29-32	≡ III	9
33-36	≡ II	7
Total		50

In this type of classification there are two elements (i) variable (ii) frequency

Variable

Variable refers to the characteristic that varies in magnitude or quantity. E.g. weight of the students. A variable may be discrete or continuous.

Discrete variable

A discrete variable can take only certain specific values that are whole numbers (integers). E.g. Number of children in a family or Number of class rooms in a school.

Continuous variable

A Continuous variable can take any numerical value within a specific interval.

Example: the average weight of a particular class student is between 60 and 80 kgs.

Frequency

Frequency refers to the number of times each variable gets repeated.

For example there are 50 students having weight of 60 kgs. Here 50 students is the frequency.

Frequency distribution

Frequency distribution refers to data classified on the basis of some variable that can be measured such as prices, weight, height, wages etc.

The following are the two examples of discrete and continuous frequency distribution

The following technical terms are important when a continuous frequency distribution is formed

Class limits: Class limits are the lowest and highest values that can be included in a class. For example take the class 40-50. The lowest value of the class is 40 and the highest value is 50. In this class there can be no value lesser than 40 or more than 50. 40 is the lower class limit and 50 is the upper class limit.

Class interval: The difference between the upper and lower limit of a class is known as class interval of that class. Example in the class 40-50 the class interval is 10 (i.e. 50 minus 40).

Class frequency: The number of observations corresponding to a particular class is known as the frequency of that class

Example:

Income (Rs)	No. of persons
--------------------	-----------------------

1000 - 2000	50
-------------	----

In the above example, 50 is the class frequency. This means that 50 persons earn an income between Rs.1, 000 and Rs.2, 000.

(iv) **Class mid-point:** Midpoint of a class is formed out as follows.

Tabulation of Data

A table is a systematic arrangement of statistical data in columns and rows. Rows are horizontal arrangements whereas the columns are vertical ones.

Difference between classification and Tabulation

Classification is the process of grouping the data on the basis of some common characteristics.

Tabulation is the process of placing the classified data in columns and rows.

Classification is the first step in tabulation. Tabulation process starts only after the classification is completed.

Classification is done for the purpose of tabulation whereas tabulation is done for the purpose of analysis of data.

Rules for tabulation

The important rules for tabulation are given below:

- Number: Table must be arranged with number in order to identify the table
- Title: Table must have a title. The title should be clear, brief and self-explanatory. It should convey the content and purpose of the table.
- Stubs and Captions: Stubs are the row headings and Captions are column headings. Stubs and Captions should also be clear and brief.
- Body of the table: The body of the table contains the numerical information. This is the most important part of the table and should contain only relevant information. Table should not be overloaded with details.
- Spacing: Proper space should be provided in between various data presented in the table. It helps for easy identification of data. Proper spacing gives good appearance for the table.
- Total: Total must be given for each column and row.
- Head note: It is a brief explanation of the information given in the table. It is placed below the title in brackets
- Footnotes: If any item presented in the table needs further explanation, it should be given at the end of the table.
- Source: Source is an important requirement for the table. Source indicates the place or person from whom the information is collected. It gives more authenticity for the table.
- **Diagrammatic representation of data**
- Diagrams are based on scale but are not confined to points or lines. They are more attractive and easier to understand than graphs.

Merits

1. Most of the people are attracted by diagrams.
2. Technical Knowledge or education is not necessary.
3. Time and effort required are less.
4. Diagrams show the data in proper perspective.
5. Diagrams leave a lasting impression.
6. Language is not a barrier.
7. Widely used tool.

Demerits (or) limitations

1. Diagrams are approximations.
2. Minute differences in values cannot be represented properly in diagrams.
3. Large differences in values spoil the look of the diagram.
4. Some of the diagrams can be drawn by experts only. eg. Pie chart.
5. Different scales portray different pictures to laymen.

Types of Diagrams

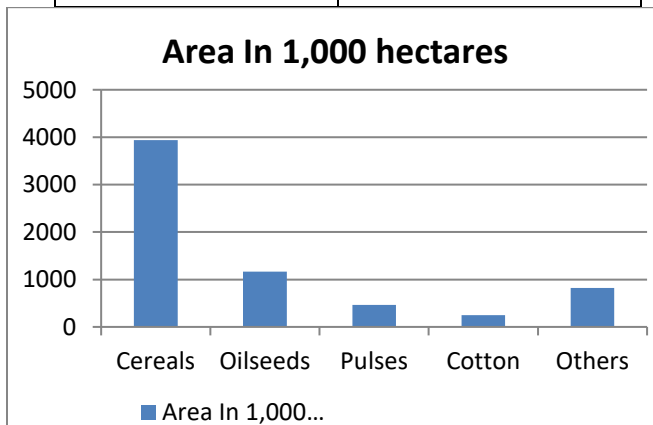
1. Simple Bar diagram.
2. Multiple Bar diagram.
3. Component Bar diagram.
4. Percentage Bar diagram.
5. Pie chart
6. Pictogram
7. Statistical maps or cartograms.

In all the diagrams and graphs, the groups or classes are represented on the x-axis and the volumes or frequencies are represented in the y-axis. Simple Bar diagram If the classification is based on attributes and if the attributes are to be compared with respect to a single character we use simple bar diagram.

Simple bar diagrams Consists of vertical bars of equal width. The heights of these bars are proportional to the volume or magnitude of the attribute. All bars stand on the same baseline. The bars are separated from each others by equal intervals. The bars may be coloured or marked. Example The cropping pattern in Tamil Nadu in the year 1974-75 was as follows.

The simple bar diagram for this data is given below

Crops	Area In 1,000 hectares
Cereals	3940
Oilseeds	1165
Pulses	464
Cotton	249
Others	822



Multiple bar diagram

If the data is classified by attributes and if two or more characters or groups are to be compared within each attribute we use multiple bar diagrams. If only two characters are to be compared within each attribute, then the resultant bar diagram used is known as double bar diagram.

The multiple bar diagram is simply the extension of simple bar diagram. For each attribute two or more bars representing separate characters or groups are to be placed side by side. Each bar within an attribute will be marked or coloured differently in order to distinguish them. Same type of marking or colouring should be done under each attribute. A footnote has to be given explaining the markings or colourings. Example Draw a multiple bar diagram for the following data which represented agricultural production for the period from 2000-2004

Component bar diagram

This is also called sub – divided bar diagram. Instead of placing the bars for each component side by side we may place these one on top of the other. This will result in a component bar diagram.

Percentage bar diagram

Sometimes when the volumes of different attributes may be greatly different for making meaningful comparisons, the attributes are reduced to percentages. In that case each attribute will have 100 as its maximum volume. This sort of component bar chart is known as percentage bar diagram.

$$\text{Percentage} = \frac{\text{Actual value}}{\text{Total of the actual value}} \times 100$$

Pie chart / Pie Diagram

Pie diagram is a circular diagram. It may be used in place of bar diagrams. It consists of one or more circles which are divided into a number of sectors. In the construction of pie diagram the following steps are involved.

Step 1: Whenever one set of actual value or percentage are given, find the corresponding angles in degrees using the following formula

$$\text{Angle} = \frac{\text{Actual value}}{\text{Total of the actual value}} \times 360^\circ$$

Step 2: Find the radius using the area of the circle πr^2 where value of π is $22/7$ or 3.14

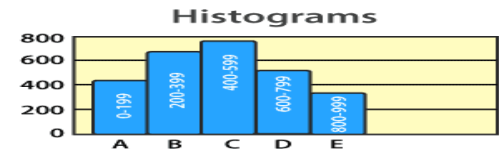
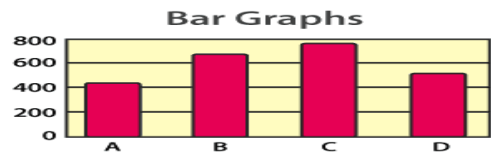
Graphical Representation

Graphical Representation is a way of analysing numerical data. It exhibits the relation between data, ideas, information and concepts in a diagram. It is easy to understand and it is one of the most important learning strategies. It always depends on the type of information in a particular domain. There are different types of graphical representation. Some of them are as follows:

- **Line Graphs** – Line graph or the linear graph is used to display the continuous data and it is useful for predicting future events over time.
- **Bar Graphs** – Bar Graph is used to display the category of data and it compares the data using solid bars to represent the quantities.

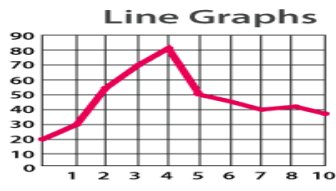
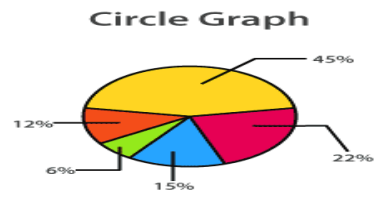
- **Histograms** – The graph that uses bars to represent the frequency of numerical data that are organised into intervals. Since all the intervals are equal and continuous, all the bars have the same width.
- **Line Plot** – It shows the frequency of data on a given number line. ‘ x ‘ is placed above a number line each time when that data occurs again.
- **Frequency Table** – The table shows the number of pieces of data that falls within the given interval.
- **Circle Graph** – Also known as the pie chart that shows the relationships of the parts of the whole. The circle is considered with 100% and the categories occupied is represented with that specific percentage like 15%, 56%, etc.
- **Stem and Leaf Plot** – In the stem and leaf plot, the data are organised from least value to the greatest value. The digits of the least place values from the leaves and the next place value digit forms the stems.
- **Box and Whisker Plot** – The plot diagram summarises the data by dividing into four parts. Box and whisker show the range (spread) and the middle (median) of the data.

TYPES OF GRAPHICAL REPRESENTATION



Frequency Table

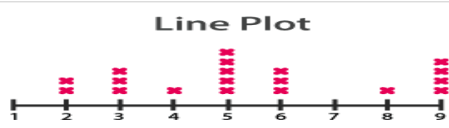
Rulers of France		
Reign (Years)	Tally	Frequency
1-15		18
16-30		11
31-45		6
46-60		4
61-75		1



Stem and Leaf Plot

Stem	Leaf
0	1, 1, 2, 2, 3, 4, 4, 4, 4, 5, 8
1	0, 0, 0, 1, 1, 3, 7, 9
2	5, 5, 7, 7, 8, 8, 9, 9
3	0, 1, 1, 1, 2, 2, 2, 4, 5
4	0, 4, 8, 9
5	2, 6, 7, 7, 8
6	3, 6

Key : 6 | 3 = 63 Year



© Byjus.com

Measure of central value
Average defined

“Average is an attempt to find one single figure to describe whole of figures”—clark

Types of average

1. Arithmetic Mean

Mean, often referred to as the arithmetic average or arithmetic mean, is calculated by adding all the numbers in a given set and then dividing by the total number of items within that set.

Let's discuss an example where we find use of arithmetic mean.

The arithmetic mean maintains its place in calculating a stock's average closing price during a particular month.

1. Simple arithmetic mean
2. Weight arithmetic mean

Properties of Arithmetic Mean

Some important properties of the arithmetic mean are as follows:

- The sum of deviations of the items from their arithmetic mean is always zero, i.e. $\sum(x - X) = 0$.
- The sum of the squared deviations of the items from Arithmetic Mean (A.M) is minimum, which is less than the sum of the squared deviations of the items from any other values.
- If each item in the arithmetic series is substituted by the mean, then the sum of these replacements will be equal to the sum of the specific items.

Merits of Arithmetic Mean

- The arithmetic mean is simple to understand and easy to calculate.
- It is influenced by the value of every item in the series.
- A.M is rigidly defined.
- It has the capability of further algebraic treatment.
- It is a measured value and not based on the position in the series.

Demerits of Arithmetic Mean

- It is changed by extreme items such as very small and very large items.
- It can rarely be identified by inspection.
- In some cases, A.M. does not represent the original item. For example, average patients admitted to a hospital are 10.7 per day.
- The arithmetic mean is not suitable in extremely asymmetrical distributions.

2. Median

The *median* is another form of an average. It usually represents the middle number in a given sequence of numbers when it's ordered by rank. When the quiz scores are listed from lowest to highest: 30, 56, 65, 70, 84, 90, 90, 91, 92, or highest to lowest: 92, 91, 90, 90, 84, 70, 65, 56, 30, we can see that the **median**, or middle, score is 84.

3. Mode

The *mode* is the most frequent value in a set of data. For our test takers, the **mode**, or most common, When the quiz scores are listed from lowest to highest: 30, 56, 65, 70, 84, 90, 90, 91, 92, or highest to lowest: 92, 91, 90, 90, 84, 70, 65, 56, 30, the mode value is 90.

4. Geometric mean

The geometric mean is a type of average, usually used for growth rates, like population growth or interest rates. While the arithmetic mean adds items, the geometric mean multiplies items. Also, you can only get the geometric mean for positive numbers.

Like most things in math, there's an easy explanation, and there's a more, ahem, mathematical way of stating the same thing. Formally, the geometric mean is defined as "...the nth root of the product of n numbers.

5. Harmonic mean

The harmonic mean is a type of numerical average. It is calculated by dividing the number of observations by the reciprocal of each number in the series. Thus, the harmonic mean is the reciprocal of the arithmetic mean of the reciprocals.

1. Arithmetic Mean

The most popular and widely used measures of representing the entire data by one value is what most laymen call on average and what the statistics call the arithmetic mean.

I. Calculation of Arithmetic mean – Individual series

1. The following tables given the monthly income of 10 employees in an office.
Income (RS): 1,780 1,760 1,690 1,750 1,840 1,920 1,100 1,810 1,050 1,950.

Calculate the arithmetic mean of incomes.

Let income be denoted by the symbol X.

CALCULATION OF ARITHMETIC MEAN

Employees	Monthly Income	Employees	Monthly Income
1	1,780	6	1,920
2	1,760	7	1,100
3	1,690	8	1,810
4	1,750	9	1,050
5	1,840	10	1,950
		N=10	∑x=16,650

$$\bar{X} = \frac{\sum X}{N}, \sum X = 16,650, N = 10$$

$$\bar{X} = \frac{16,650}{10} = 1,665$$

RS.1,665.

2. The following tables given the monthly income of 10 employees in an office.
Income (RS): 1,780 1,760 1,690 1,750 1,840 1,920 1,100 1,810 1,050 1,950.

Calculate the arithmetic mean by taking 1,800 as the assumed mean.

CALCULATION OF ARITHMETIC MEAN

Employee	Income	(X -1800)
1	1,780	-20
2	1,760	-40
3	1,690	-110
4	1,750	-50
5	1,840	+40
6	1,920	+120
7	1,100	-700
8	1,810	+10
9	1,050	-750
10	1,950	+150

N=10		$\sum d = -1350$
------	--	------------------

$$\bar{X} = A + \frac{\sum d}{N}, \quad A = 1800, \quad \sum d = -1350, \quad N = 10$$

$$\bar{X} = 1800 - \frac{1350}{10} = 1,800 - 135 = 1,665$$

RS.1,665.

II.CALCULATION OF ARITHMETIC MEAN - DISCRETE SERIES

1.From the following data of the marks obtained by 60 students of a class. Calculate the arithmetic mean.

Marks	No.of.students	Marks	No.of.students
20	8	50	10
30	12	60	6
40	20	70	4

Let the marks be denoted by X and the number of students by f

CALCULATION OF ARITHMETIC MEAN

Marks	No.of.students f	Fx
20	8	160
30	12	360
40	20	800
50	10	500
60	6	360
70	4	280
	N=60	$\sum fx = 2,460$

$$\bar{X} = \frac{\sum fx}{N} = \frac{2460}{60} = 41$$

2. .From the following data of the marks obtained by 60 students of a class. calculate the arithmetic mean by the short cut method.

Marks	No.of.students	Marks	No.of.students
20	8	50	10
30	12	60	6
40	20	70	4

Let the marks be denoted by X and the number of students by f.

CALCULATION OF ARITHMETIC MEAN

Marks X	No.of.students f	d=(X - 40)	Fd
20	8	-20	-160
30	12	-10	-120
40	20	0	0
50	10	+10	+100
60	6	+20	+120
70	4	+30	+120

	N =60		$\sum fd =60$
--	-------	--	---------------

$$\bar{X} = A + \frac{\sum fd}{N}$$

$$= 40 + \frac{60}{60} = 40 + 1 = 41$$

III. Calculation of Arithmetic mean - Continuous series

1. From the following data compute arithmetic mean by direct method

Marks : 0 -10 10 -20 20-30 30-40 40 -50 50-60

No.of.students : 5 10 25 30 20 10

CALCULATION OF ARITHMETIC MEAN

Marks	Mid- point m	No.of.Students f	Fm
0-10	5	5	25
10-20	15	10	150
20-30	25	25	625
30-40	35	30	1,050
40-50	45	20	900
50-60	55	10	550
		N= 100	$\sum fm =3,300$

$$\bar{X} = \frac{\sum fm}{N}$$

$$= \frac{3,300}{100} = 33$$

2. From the following data compute arithmetic mean by short cut method

Marks : 0 -10 10 -20 20-30 30-40 40 -50 50-60

No.of.students : 5 10 25 30 20 10

CALCULATION OF ARITHMETIC MEAN

Marks	Mid- point m	No.of.Students f	(m -35) d	Fd
0-10	5	5	-30	-150

10-20	15	10	-20	-200
20-30	25	25	-10	-250
30-40	35	30	0	0
40-50	45	20	+10	+200
50-60	55	10	+20	+200
		N= 100		$\sum fd = -200$

$$\bar{X} = A + \frac{\sum fd}{N}$$

$$= 35 - \frac{200}{100} = 35 - 2 = 33$$

3. From the following data compute arithmetic mean

Marks : 0 -10 10 -30 30-60 60 -100

No.of.students : 5 12 25 8

CALCULATION OF ARITHMETIC MEAN

Marks	Mid- point m	No.of.Students f	(m -45) d	Fd
0-10	5	5	-8	-40
10-30	20	12	-5	-60
30-60	45	25	0	0
60-100	80	8	+7	+56
		N= 50		$\sum fd = -44$

$$\bar{X} = A + \frac{\sum fd}{N} \times C$$

$$= 45 - \frac{44}{50} \times 5 = 45 - 4.4 = 40.6$$

Weighted Arithmetic Mean

Arithmetic mean discussed is that it equal importance to all the items .But there are cases where the relative importance of the different items is not the same. When this is so we compute weighted arithmetic mean. The term weight stands for the relative importance of the different items.

Weighted Arithmetic Mean

1. Comment on the performance of the students of the three universities given below using simple and weighted average.

University Course study	Pass %	Bombay No.of.students (in hundreds)	Pass %	Calcutta No.of.students (in hundreds)	Pass %	Madras No.of.students (in hundreds)
M.A	71	3	82	2	81	2
M.COM	83	4	76	3	76	3.5
B.A	73	5	73	6	74	4.5
B.COM	74	2	76	7	58	2
B.SC	65	3	65	3	70	7
M.SC	66	3	60	7	73	2

Calculation of Simple And Weighted Arithmetic Mean

University Course	Pass % X	W	Bombay No.of. students (inhundreds) WX	Pass % X	W	Calcutta No.of. students (inhundreds) WX	Pass % x	W	Madras No.of. students (inhundreds)
M.A	71	3	213	82	2	164	81	2.0	162
M.COM	83	4	332	76	3	228	76	3.5	266
B.A	73	5	365	73	6	438	74	4.5	333
B.COM	74	2	148	76	7	532	58	2.0	116
B.SC	65	3	195	65	3	195	70	7.0	490
M.SC	66	3	198	60	7	420	73	2.0	146
	$\Sigma X = 432$	$\Sigma W = 20$	$\Sigma WX = 1,451$	$\Sigma X = 432$	$\Sigma W = 28$	$\Sigma WX = 1,977$	$\Sigma X = 432$	$\Sigma W = 21$	$\Sigma WX = 1,513$

Simple and Weighted Arithmetic Mean

Bombay $\bar{X} = \frac{\Sigma X}{N} = \frac{432}{6} = 72$ $X = \frac{\Sigma WX}{\Sigma W} = \frac{1,451}{20} = 72.55$

Calcutta $\bar{X} = \frac{\sum X}{N} = \frac{432}{6} = 72$ $X = \frac{\sum WX}{\sum W} = \frac{1,977}{28} = 70.61$

Madras $\bar{X} = \frac{\sum X}{N} = \frac{432}{6} = 72$ $X = \frac{\sum WX}{\sum W} = \frac{1,513}{21} = 72.05$

II. Calculation of Median – Individual series

1. From the following data of the wages of 7 workers compute the median wage;

Wages (in RS) 1100 1150 1080 1120 1200 1160 1400

Calculation of median

Sl.No	Wages arranged in ascending order	Sl.No	Wages arranged in ascending order
1	1080	5	1160
2	1100	6	1200
3	1120	7	1400
4	1150		

Median = Size of $\frac{N+1}{2}$ th item = $\frac{7+1}{2} = 4$ th item

Size of 4th item = 1,150.

2. Obtain the value of median from the following data

Wages (in RS): 391 384 591 407 672 522 777 753 2,488 1,490

Calculation of median

Sl.No	Wages arranged in ascending order	Sl.No	Wages arranged in ascending order
1	384	6	672
2	391	7	753
3	405	8	777
4	522	9	1,490
5	591	10	2,488

Median = Size of $\frac{N+1}{2}$ th item = $\frac{11}{2} = 5.5$ th item

Size of 5.5th item = $\frac{5th\ Item + 6th\ Item}{2} = \frac{591+672}{2} = \frac{1,263}{2} = 631.5$

II. CALCULATION OF MEDIAN - DISCRETE SERIES

1.From the following data find the value of median

Income (RS) : 4,000 4,500 5,800 5,060 6,600 5,380

No.ofperson : 24 26 16 20 6 30

CALCULATION OF MEDIAN

Wages arranged in ascending order	No.of. person f	c.f	Wages arranged in ascending order	No.of. person f	c.f
4000	24	24	5380	30	100
4500	26	50	5800	16	116
5060	20	70	6600	6	122

$$\text{Median} = \text{Size of } \frac{N+1}{2} \text{ th item} = \frac{122+1}{2} = 61.5\text{th item}$$

Size of 61.5th item =Rs.5,060.

III. Calculation of Arithmetic mean - Continuous series

1.Calculate the median for the following frequency distribution

Marks	No.of. students	Marks	No.of. Students
45 - 50	10	20 - 25	31
40 - 45	15	15 - 20	24
35 - 40	26	10 - 15	15
30 - 35	30	5 - 10	7
25 - 30	42		

CALCULATION OF MEDIAN

Marks	No.of. Students f	c.f	Marks	No.of. Students f s	c.f
5 - 10	7	7	30- 35	30	149
10 -15	15	22	35 - 40	26	175
15 - 20	24	46	40 - 45	15	190
20 - 25	31	77	45 - 50	10	200
25 - 30	42	119			

Median = Size of $\frac{N}{2}$ th item = $\frac{200}{2} = 100$ th item

Median lies in the class 25 - 30

$$\text{Median} = L + \frac{\frac{N}{2} - c.f}{f} * i$$

$$L=25, \frac{N}{2}=100, c.f = 77, f = 42, i=5$$

$$= 25 + \frac{100-77}{42} * 5 = 25 + 2.74 = 27.74$$

I. Calculation of Mode – Individual series

1. Calculate the mode from the following data

Sl.No : 1 2 3 4 5 6 7 8 9 10

Marks obtained : 10 27 24 12 27 27 20 18 15 30

Calculation of Mode

Size of item	No.of. Times it occurs	Size of item	No.of. Times it occurs
10	1	20	1
12	1	24	1
15	1	27	3
18	1	30	1
20	1		

Since the item 27 occurs the maximum number of times ie,3 hence the modal marks are 27.

II. CALCULATION OF MODE - DISCRETE SERIES

1. Calculate the value of mode for the following data.

Marks : 10 15 20 25 30 35 40

Frequency : 8 12 36 35 28 18 9

CALCULATION OF MODE

GROUPING TABLE

X	I(f)	II	III	IV	V	VI
----------	-------------	-----------	------------	-----------	----------	-----------

10	8				
		20			
15	12	(8+12)		56	
			48	(8+12+36)	83
20	36		(12+36)		(12+36+35)
		71			
25	35				(36+35+28)
			63	81	
30	28				
		46			56
35	18				
			27		
40	9				

Col.No	20	25	30
I	1		
II	1	1	
III		1	1
IV		1	1
V	1	1	
VI	1	1	1

Mode

I. In a moderately asymmetrical distribution, the mode and mean are 32.1 and 35.4 respectively. Find out the value of median

Solution:**Mode**

$$\text{Mode} = 3 \text{ median} - 2 \text{ mean}$$

$$\text{Given mean} = 35.4 \text{ mode} = 32.1$$

$$32.1 = 3 \text{ median} - 2 * 35.4$$

$$3 \text{ median} = 32.1 + 70.8 = 102.9 \text{ or median} = 102.9 \div 3 = 34.3$$

$$\text{Given median} = 20.6 \text{ mode} = 26$$

Find Mean

$$\text{Mode} = 3 \text{ median} - 2 \text{ mean}$$

$$\text{Given means} = 20.6, \text{ mode} = 26$$

$$26 = 3 * 20.6 - 2 \text{ mean}$$

$$26 = 61.8 - 2 \text{ mean}$$

$$2 \text{ means} = 61.8 - 26 = 35.8 \text{ or mean} = 17.9$$

Calculate arithmetic mean, median and mode from the following frequency distribution

variable	frequency	Variable	frequency
10-13	8	25-28	54
13-16	15	28-31	36
16-19	27	31-34	18
19-22	51	34-37	9
22-25	75	37-40	7

Calculation of mean, median and mode

variable	m.p m	F	m-23.5\3 d	fd	c.f.
10-13	11.5	8	-4	-32	8
13-16	14.5	15	-3	-45	23
16-19	17.5	27	-2	-54	50
19-22	20.5	51	-1	-51	101
22-25	23.5	75	0	0	176
25-28	26.5	54	+1	+54	230
28-31	29.5	36	+2	+72	266
31-34	32.5	18	+3	+54	284
34-37	35.5	9	+4	+36	293

37-40	38.5	7	+5	+35	300
		N=300		fd= 69	

$$\bar{X} = A + \frac{\sum fd}{N} * i$$

$$= 23.5 + \frac{69}{300} * 3 = 24.19$$

Median = size of $\frac{N}{2}$ th item = $\frac{300}{2} = 150^{\text{th}}$ item

Median lies in the class 22- 25

$$\text{Median} = L + \frac{\frac{N}{2} - c.f}{f} * i$$

$$L=22, \frac{N}{2}=150, c.f = 101, f = 75, i=3$$

$$= 22 + \frac{150-101}{75} * 3 = 22+1.96 = 23.96$$

Mode lies in the class 22-25

$$\text{Mode} = L + \frac{\Delta_1}{\Delta_1 + \Delta_2} * i$$

$$L = 22, \Delta_1 = (75 - 51) = 24, \Delta_2 = (75 - 54) = 21, i = 3$$

$$22 + \frac{24}{24+21} * 3 = 22+1.6 = 23.6$$

GEOMETRIC MEAN

Calculation of Geometric Mean – Individual series

1. Daily income of 10 families of a particular place is given below. Find out G.M.

85 70 15 75 500 8 45 250 40 36

Calculation of Geometric Mean

X	Log X	X	Log X
85	1.9294	8	0.9031
70	1.8451	45	1.6532
15	1.1761	250	2.3979
75	1.8751	40	1.6021
500	2.6990	36	1.5563
			$\sum \log X =$ 17.6373

$$\text{Geometric Mean} = \text{AL} \left(\frac{\sum \log X}{N} \right)$$

$$\text{AL} (17.6373) = \text{AL}(1.7637) = 58.03$$

Calculation of Geometric Mean – Continuous series

1. Find the Geometric mean for the data given below

Marks	Frequency	Marks	Frequency
4-8	6	24-28	12
8-12	10	28-32	10
12-16	18	32-36	6
16-20	30	36-40	2
20-24	15		

Calculation of Geometric Mean

Marks	m.p	F	Log m	F*log m
4-8	6	6	0.7782	4.6692
8-12	10	10	1.0000	10.0000
12-16	14	18	1.1461	20.6298
16-20	18	30	1.2553	37.6590
20-24	22	15	1.3424	20.1360
24-28	26	12	1.4150	16.9800
28-32	30	10	1.4771	14.7710
32-36	34	6	1.5315	9.1890
36-40	38	2	1.5798	3.1596
		N=109		$\sum f \cdot \log m = 137.1936$

$$\text{G.M} = \text{A.L} \left(\frac{\sum f \log X}{N} \right)$$

$$\text{AL} = \left(\frac{137.1936}{109} \right) = \text{AL} = 1.2587 = 18.14$$

HARMONIC MEAN

I. Calculation of harmonic mean – Individual series

(I) find the harmonic mean from the following

2574 475 75 5 0.8 0.08 0.005 0.0009

Calculation of harmonic mean

X	(1/X)	X	(1/X)
2574	0.0004	0.8	1.2500

475	0.0021	0.08	12.5000
75	0.0133	0.005	200.0000
5	0.2000	0.0009	1111.111
			$\Sigma(1/X)=1325.0769$

$$H.M = \frac{N}{\Sigma(1/X)} = \frac{8}{1325.0769} = 0.006$$

II. Calculation of harmonic mean – Discrete series

(I) find the harmonic mean from the following

Mark 10 20 25 40 50

No. of students 20 30 50 15 5

Calculation of harmonic mean

Marks x	f	(f/x)
10	20	2000
20	30	1500
25	50	2000
40	15	0.375
50	5	0.100
	N=120	$\Sigma(f/X)=5.975$

$$H.M = \frac{N}{\Sigma(f/X)} = \frac{120}{5.975} = 20.08$$

I. Calculation of harmonic mean – continuous series

From the following data compute the value of harmonic mean

Class interval 10-20 20-30 30-40 40-50 50-60

Frequency 4 6 10 7 3

Calculation of harmonic mean

Class intervals	Mid-points m	Frequency f	f/m
10-20	15	4	0.267
20-30	25	6	0.240
30-40	35	10	0.286
40-50	45	7	0.156
50-60	55	3	0.055
		N=30	f/m= 1.004

$$\text{H.M} = \frac{N}{\sum(f \backslash X)} = \frac{30}{1.004} = 29.88$$

BUSINESS TOOLS FOR DECISION MAKING

UNIT – II

MEASURES OF DISPERSION:

RANGE – QUARTILES – DECILES – PERCENTILES – QUARTILE DEVIATION – MEAN DEVIATION - STANDARD DEVIATION – CO-EFFICIENT OF VARIATION.

The measure of dispersion indicates the scattering of data. It explains the disparity of data from one another, delivering a precise view of the distribution of data. The measure of dispersion displays and gives us an idea about the variation and central value of an individual item.

In other words, Dispersion is the extent to which values in a distribution differ from the average of the distribution. It gives us an idea about the extent to which individual items vary from one another and from the central value.

Characteristics of a good measure of dispersion

An ideal measure of dispersion is expected to possess the following properties

1. It should be rigidly defined
2. It should be based on all the items.
3. It should not be unduly affected by extreme items.
4. It should lend itself for algebraic manipulation.
5. It should be simple to understand and easy to calculate

The variation can be measured in different numerical measures, namely:

- (i) **Range** – It is the simplest method of measurement of dispersion and defines the difference between the largest and the smallest item in a given distribution.

Merits of Range

1. It is simple to understand.
2. It is easy to calculate.
3. In certain types of problems like quality control, weather forecasts, share price analysis, etc., range is most widely used.

Demerits of Range

1. It is very much affected by the extreme items.
2. It is based on only two extreme observations.
3. It cannot be calculated from open-end class intervals.
4. It is not suitable for mathematical treatment.
5. It is a very rarely used measure

Range = L – S; where L = largest item and S = smallest item

The relative measure is given by co-efficient of Range

$$\text{Co-efficient of range} = \frac{L-s}{L+s}$$

(ii) **Quartile Deviation** – It is known as Semi-Inter-Quartile Range, i.e. half of the difference between the upper quartile and lower quartile. The first quartile is derived as (Q), the middle digit (Q1) connects the least number with the median of the data. The median of a data set is the (Q2) second quartile. Lastly, the number connecting the largest number and the median is the third quartile (Q3). Quartile deviation can be calculated by

$$Q = \frac{1}{2} \times (Q3 - Q1)$$

(iii) **Mean Deviation**-Mean deviation is the arithmetic mean (average) of deviations $|D|$ of observations from a central value {Mean or Median}.

Mean deviation can be evaluated by using the formula: $A = \frac{1}{n} [\sum i |x_i - A|]$

(iv) **Standard Deviation**- Standard deviation is the Square Root of the Arithmetic Average of the squared of the deviations measured from the mean. The standard deviation is given as

$$\sigma = [(\sum i (y_i - \bar{y})^2 / n)]^{1/2} = [(\sum i y_i^2 / n) - \bar{y}^2]^{1/2}$$

Apart from a numerical value, graphics method are also applied for estimating dispersion.

Types of Measures of Dispersion

(1) Absolute Measures

- Absolute measures of dispersion are expressed in the unit of Variable itself. Like, Kilograms, Rupees, Centimeters, Marks etc.

(2) Relative Measures

- Relative measures of dispersion are obtained as ratios or percentages of the average.
- These are also known as ‘Coefficient of dispersion.’
- These are pure numbers or percentages totally independent of the units of measurements.

Various ‘Absolute Measures’ of Dispersion

Following Are the Different ‘absolute Measures’ of Dispersion:

(1) Range

- It is the simplest method of measurement of dispersion.
- It is defined as the difference between the largest and the smallest item in a given distribution.
- **Range = Largest item (L) – Smallest item (S)**

(2) Interquartile Range

- It is defined as the difference between the Upper Quartile and Lower Quartile of a given distribution.

Interquartile Range = Upper Quartile (Q₃)–Lower Quartile (Q₁)

(3) Quartile Deviation

- It is known as Semi-Inter-Quartile Range, i.e. half of the difference between the upper quartile and lower quartile.
- **Quartile Deviation** = $\frac{\text{Upper Quartile}(Q_3) - \text{Lower Quartile}(Q_1)}{2}$

(4) Mean Deviation

- Mean deviation is the arithmetic mean (average) of deviations $|D|$ of observations from a central value {Mean or Median}.

(5) Standard Deviation

- Standard deviation is the Square Root of the Arithmetic Average of the squared of the deviations measured from the mean.

Merits of Standard Deviation

1. It is rigidly defined and its value is always definite and based on all the observations and the actual signs of deviations are used.
2. As it is based on arithmetic mean, it has all the merits of arithmetic mean.
3. It is the most important and widely used measure of dispersion.
4. It is possible for further algebraic treatment.
5. It is less affected by the fluctuations of sampling and hence stable.
6. It is the basis for measuring the coefficient of correlation and sampling.

Demerits

1. It is not easy to understand and it is difficult to calculate.
2. It gives more weight to extreme values because the values are squared up.
3. As it is an absolute measure of variability, it cannot be used for the purpose of comparison.

(6) Lorenz Curve

- The Lorenz Curve is a graphic method of measuring estimated dispersion.
- This curve is often used to measure the inequalities of income or wealth in a society.

Various 'Relative Measures' of Dispersion

Following Are the Relative Measure of Dispersion:

(1) Coefficient of Range

It refers to the ratio of the difference between two extreme items of the distribution to their sum.

$$\text{Coefficient of Range} == \frac{L-s}{L+s}$$

(2) Coefficient of Quartile Deviation

It refers to the ratio of the difference between Upper Quartile and Lower Quartile of a distribution to their sum.

$$\text{Coefficient of Quartile Deviation} = \frac{Q3-Q1}{Q3+Q1}$$

(3) Coefficient of Mean Deviation

- Mean deviation is an absolute measure of dispersion.
- In order to transform it into a relative measure, it is divided by the particular average, from which it has been calculated.
- It is then known as the Coefficient of Mean Deviation.
- Coefficient of Mean Deviation from Mean (\bar{x}) = $\frac{MD\bar{x}}{\bar{x}}$
- Coefficient of Mean Deviation from Median (ME) = $\frac{MDMe}{Me}$

(4) Coefficient of Standard Deviation

- It is calculated by dividing the standard deviation (σ) by the mean (\bar{x}) of the data.
- Coefficient of Standard Deviation = $\frac{\sigma}{\bar{x}}$

(5) Coefficient of Variation

- It is used to compare two data with respect to stability (or uniformity or consistency or homogeneity).
- It indicates the relationship between the standard deviation and the arithmetic mean expressed in terms of percentage.
- Coefficient of Variation (C.V.) = $\frac{\sigma}{\bar{x}} \times 100$

Where, C.V. = Coefficient of Variation; σ = Standard Deviation;
 \bar{X} = Arithmetic Mean

Problems:

Individual observation :

1. The following re the prices of shares of AB co. limited from Monday to Saturday

Days	Monday	Tuesday	Wednesday	Thursday	Friday	Saturday
Prices(x)	200	210	208	160	220	250

Calculate Range & its Co-efficient.

Range = L – S (L= largest item , S = smallest item)

$$\text{Range} = L - S = 250 - 160 = 90$$

$$\text{Co-efficient of range} = \frac{L-S}{L+S} ; \frac{250-160}{250+160} = \frac{90}{410} = 0.219$$

2. Find the Range & its Co-efficient. X : 8 10 5 9 12 11

$$\text{Range} = L - S = 12 - 5 = 7$$

$$\text{Co-efficient of range} = \frac{L-S}{L+S} ; \frac{12-5}{12+5} = \frac{7}{17} = 0.411$$

Continuous series : In the frequency distribution Range is calculated between lower limit of the lowest class interval and upper limit of the highest class interval

3. Calculate the range and co-efficient of the range from the following data:

Marks(x)	10-20	20-30	30-40	40-50	50-60
No. of students (f)	8	10	12	8	4

$$\text{Range} = L - S = 60 - 10 = 50$$

$$\text{Co-efficient of range} = \frac{L-S}{L+S} ; \frac{60-10}{60+10} = \frac{50}{70} = 0.714$$

QUARTILES :

Quartiles divides the distribution into four equal parts and we have three quartiles i.e, Q₁, Q₂, Q₃

Individual observation :

$$Q_1 = \text{Size of } \left(\frac{N+1}{4}\right)^{\text{th}} \text{ Item}$$

$$Q_3 = \text{Size of } 3\left(\frac{N+1}{4}\right)^{\text{th}} \text{ Item}$$

Continuous series :

$$Q_1 = L + \left(\frac{\frac{N}{4} - cf}{f}\right) X C$$

$$Q_2 = L + \left(\frac{\frac{2N}{4} - cf}{f}\right) X C$$

$$Q_3 = L + \left(\frac{\frac{3N}{4} - cf}{f}\right) X C$$

DECILES :

Deciles divides the distribution into ten equal parts . And hence we have 9 Deciles i.e., D₁, D₂, D₃,.....D₉

Individual observation :

$$D_1 = \text{Size of } \left(\frac{N+1}{10}\right)^{\text{th}} \text{ Item}$$

$$D_2 = \text{Size of } 2\left(\frac{N+1}{10}\right)^{\text{th}} \text{ Item}$$

$$D_9 = \text{Size of } 9\left(\frac{N+1}{10}\right)^{\text{th}} \text{ Item}$$

Continuous Series :

$$D_1 = L + \left(\frac{\frac{N}{10} - cf}{f}\right) X C$$

$$D_2 = L + \left(\frac{\frac{2N}{10} - cf}{f}\right) X C$$

$$D_5 = L + \left(\frac{\frac{5N}{10} - cf}{f}\right) X C$$

$$D_9 = L + \left(\frac{\frac{9N}{10} - cf}{f} \right) X C$$

PERCENTILES:

Percentiles divides the distribution into 100 equal parts and hence we have 99 percentiles i.e., P₁, P₂, P₃, P₉₉

Individual observation :

Continuous series :

$$P_1 = \text{Size of } \left(\frac{N+1}{100} \right)^{\text{th}} \text{ Item}$$

$$P_1 = L + \left(\frac{\frac{N}{100} - cf}{f} \right) X C$$

$$P_2 = \text{Size of } 2 \left(\frac{N+1}{100} \right)^{\text{th}} \text{ Item}$$

$$P_2 = L + \left(\frac{\frac{2N}{100} - cf}{f} \right) X C$$

$$P_{99} = \text{Size of } 99 \left(\frac{N+1}{100} \right)^{\text{th}} \text{ Item}$$

$$P_{99} = L + \left(\frac{\frac{99N}{100} - cf}{f} \right) X C$$

$$P_{99} = L + \left(\frac{\frac{99N}{100} - cf}{f} \right) X C$$

Problems:

1. Find D₁, D₉ , Q₁ , Q₃ , P₁₆, P₇₀ from the following data.

X	282	754	125	765	875	645	985
	235	175	895	905	112	155	

Arrange the X value into Ascending order :

112 125 155 175 235 282 645 754 765 875 895 905 985

1. D₁ = Size of $\left(\frac{N+1}{10}\right)^{\text{th}}$ Item, D₁ = Size of $\left(\frac{13+1}{10}\right)^{\text{th}}$ Item, D₁ = Size of $\left(\frac{14}{10}\right)^{\text{th}}$ Item,

$$D_1 = \text{Size of } (1.4)^{\text{th}} \text{ Item, } D_1 = \frac{112+125}{2}; \mathbf{D_1 = 118.5}$$

2. D₉ = Size of $9\left(\frac{N+1}{10}\right)^{\text{th}}$ Item , D₉ = Size of $9\left(\frac{13+1}{10}\right)^{\text{th}}$ Item; D₉ = Size of $9\left(\frac{14}{10}\right)^{\text{th}}$ Item

$$D_9 = \text{Size of } 9(1.4)^{\text{th}} \text{ Item; } D_9 = \text{Size of } 12.6^{\text{th}} \text{ Item; } D_9 = \frac{905+985}{2}; \mathbf{D_9 = 945}$$

3. Q₁ = Size of $\left(\frac{N+1}{4}\right)^{\text{th}}$ Item; Q₁ = Size of $\left(\frac{13+1}{4}\right)^{\text{th}}$ Item; Q₁ = Size of $\left(\frac{14}{4}\right)^{\text{th}}$ Items

$$Q_1 = \text{Size of } (3.5)^{\text{th}} \text{ Item; } Q_1 = \frac{155+175}{2}; Q_1 = \frac{330}{2}; \mathbf{Q_1 = 165}$$

4. Q₃ = Size of $3\left(\frac{N+1}{4}\right)^{\text{th}}$ Item ; Q₃ = Size of $3\left(\frac{13+1}{4}\right)^{\text{th}}$ Item; Q₃ = Size of $3\left(\frac{14}{4}\right)^{\text{th}}$ Item

$$Q_3 = \text{Size of } 3(3.5)^{\text{th}} \text{ Item; } Q_3 = \text{Size of } 10.5^{\text{th}} \text{ Item; } Q_3 = \frac{875+895}{2} \quad Q_3 = \frac{1330}{2}; \mathbf{Q_3 = 885}$$

5. $P_{16} = \text{Size of } 16 \left(\frac{N+1}{100}\right)^{\text{th}} \text{ Item}$; $P_{16} = \text{Size of } 16 \left(\frac{13+1}{100}\right)^{\text{th}} \text{ Item}$;

$P_{16} = \text{Size of } 16 \left(\frac{14}{100}\right)^{\text{th}} \text{ Item}$; $P_{16} = \text{Size of } 16 (0.14)^{\text{th}} \text{ Item}$;

$P_{16} = \text{Size of } (2.24)^{\text{th}} \text{ Item}$; $P_{16} = \left(\frac{125+155}{2}\right)$; $P_{16} = \frac{280}{2}$; **$P_{16} = 140$**

6. $P_{70} = \text{Size of } 70 \left(\frac{N+1}{100}\right)^{\text{th}} \text{ Item}$; $P_{70} = \text{Size of } 70 \left(\frac{13+1}{100}\right)^{\text{th}} \text{ Item}$;

$P_{16} = \text{Size of } 70 \left(\frac{14}{100}\right)^{\text{th}} \text{ Item}$; $P_{70} = \text{Size of } 70 (0.14)^{\text{th}} \text{ Item}$;

$P_{16} = \text{Size of } (9.8)^{\text{th}} \text{ Item}$; $P_{70} = \left(\frac{765+875}{2}\right)$; $P_{70} = \frac{1640}{2}$; **$P_{70} = 820$**

2. Calculate $Q_3, D_6, D_9, P_{24}, P_{90}$ from the following data (Discrete series)

X	1	3	6	9	12	15	2	4	7	5	8
F	7	16	31	1	3	10	11	17	11	26	1
X	10	11	13	14							
F	9	7	6	4							

X	F	CF	X	F	CF
1	7	7	8	1	120
2	11	18	9	1	121
3	16	34	10	9	130
4	17	51	11	7	137
5	26	77	12	3	140
6	31	108	13	6	146
7	11	119	14	4	150
			15	10	160
N = 160					

1. $Q_3 = \text{Size of } 3 \left(\frac{N+1}{4}\right)^{\text{th}} \text{ Item}$; $Q_3 = \text{Size of } 3 \left(\frac{160+1}{4}\right)^{\text{th}} \text{ Item}$; $Q_3 = \text{Size of } 3 \left(\frac{161}{4}\right)^{\text{th}} \text{ Item}$

$Q_1 = \text{Size of } 3(40.25)^{\text{th}} \text{ Item}$; $Q_1 = \text{Size of } 120.75^{\text{th}} \text{ Item}$; **$Q_1 = 885$**

2. $D_6 = \text{Size of } 6 \left(\frac{N+1}{10}\right)^{\text{th}} \text{ Item}$, $D_6 = \text{Size of } 6 \left(\frac{160+1}{10}\right)^{\text{th}} \text{ Item}$; $D_6 = \text{Size of } 6 \left(\frac{161}{10}\right)^{\text{th}} \text{ Item}$

$D_6 = \text{Size of } 6(16.1)^{\text{th}} \text{ Item}$; $D_9 = \text{Size of } 96.6^{\text{th}} \text{ Item}$; **$D_6 = 6$**

3. $D_9 = \text{Size of } 9 \left(\frac{N+1}{10}\right)^{\text{th}} \text{ Item}$, $D_9 = \text{Size of } 9 \left(\frac{160+1}{10}\right)^{\text{th}} \text{ Item}$; $D_9 = \text{Size of } 9 \left(\frac{161}{10}\right)^{\text{th}} \text{ Item}$

$D_9 = \text{Size of } 9(16.1)^{\text{th}} \text{ Item}$; $D_9 = \text{Size of } 144.9^{\text{th}} \text{ Item}$; **$D_9 = 13$**

4. $P_{24} = \text{Size of } 24 \left(\frac{N+1}{100}\right)^{\text{th}} \text{ Item}$; $P_{24} = \text{Size of } 24 \left(\frac{160+1}{100}\right)^{\text{th}} \text{ Item}$;

$P_{24} = \text{Size of } 24 \left(\frac{161}{100}\right)^{\text{th}} \text{ Item}$; $P_{24} = \text{Size of } 24 (1.61)^{\text{th}} \text{ Item}$;

$P_{24} = \text{Size of } (38.64)^{\text{th}} \text{ Item}$; **$P_{24} = 4$**

5. $P_{90} = \text{Size of } 70 \left(\frac{N+1}{100}\right)^{\text{th}} \text{ Item}$; $P_{90} = \text{Size of } 90 \left(\frac{160+1}{100}\right)^{\text{th}} \text{ Item}$;

$P_{90} = \text{Size of } 90 \left(\frac{161}{100}\right)^{\text{th}} \text{ Item}$; $P_{90} = \text{Size of } 90 (1.61)^{\text{th}} \text{ Item}$;

$P_{90} = \text{Size of } (144.9)^{\text{th}} \text{ Item}$; $P_{90} = 13$

3. Calculate $D_2, D_4, Q_1, Q_3, P_{57}, P_{75}$ for the following data :

Marks (x): 0-10 10-20 20-30 30-40 40-50 50-60 60-70 70-80 80-90

f : 33 53 108 221 153 322 439 526 495

X	F	Cf
0-10	33	33
10-20	53	86
20-30	108	194
30-40	221	415
40-50	153	568
50-60	322	890
60-70	439	1329
70-80	526	1855
80-90	495	2350
N = 2,350		

1. $D_2 = L + \left(\frac{\frac{2N}{10} - cf}{f}\right) X C$;

$D_2 = \text{Size of } \left(\frac{2N}{10}\right)^{\text{th}} \text{ Item}$; $D_2 = \text{Size of } \left(\frac{2(2350)}{10}\right)^{\text{th}} \text{ Item}$; $D_2 = \text{Size of } \left(\frac{4700}{10}\right)^{\text{th}} \text{ Item}$;

$D_2 = \text{Size of } 470^{\text{th}} \text{ Item}$.

$D_2 - \text{CI} = 40-50$; $L = 40$; $\frac{2N}{10} = 470$; $cf = 415$; $f = 153$; $c = 10$

$D_2 = 40 + \left(\frac{470-415}{153}\right) X 10$; $D_2 = 40 + \left(\frac{55}{153}\right) X 10$; $D_2 = 40 + (0.359) X 10$;

$D_2 = 40 + 3.59$; $D_2 = 43.59$

2. $D_4 = L + \left(\frac{\frac{4N}{10} - cf}{f}\right) X C$;

$D_4 = \text{Size of } \left(\frac{4N}{10}\right)^{\text{th}} \text{ Item}$; $D_4 = \text{Size of } \left(\frac{4(2350)}{10}\right)^{\text{th}} \text{ Item}$; $D_4 = \text{Size of } \left(\frac{9400}{10}\right)^{\text{th}} \text{ Item}$;

$D_4 = \text{Size of } 940^{\text{th}} \text{ Item}$.

$$D_2 - CI = 60-70; L = 60; \frac{2N}{10} = 940; cf = 890; f = 439; c = 10$$

$$D_2 = 60 + \left(\frac{940-890}{439} \right) X 10 ; D_2 = 60 + \left(\frac{50}{439} \right) X 10 ; D_2 = 60 + (0.113) X 10$$

$$D_2 = 60 + 1.13; \mathbf{D_2 = 61.13}$$

$$3. Q_1 = L + \left(\frac{\frac{N}{4} - cf}{f} \right) X C ;$$

$$CI = 50 - 60; L = 50; \frac{N}{4} = \frac{2350}{4} = 587.5 \text{ cf} = 568; f = 322; c = 10$$

$$Q_1 = 50 + \left(\frac{587.5 - 568}{322} \right) X 10 ; Q_1 = 50 + \left(\frac{19.5}{322} \right) X 10 ; Q_1 = 50 + (.060) X 10 ;$$

$$Q_1 = 50 + 0.605 ; \mathbf{Q_1 = 50.60}$$

$$4. Q_3 = L + \left(\frac{\frac{3N}{4} - cf}{f} \right) X C$$

$$\frac{3N}{4} = \frac{3(2350)}{4} = \frac{7050}{4} = 1762.5 \text{ CI} = 70 - 80; L = 70; \text{ cf} = 1329; f = 526; c = 10$$

$$Q_3 = 70 + \left(\frac{1762.5 - 1329}{526} \right) X 10 ; Q_3 = 70 + \left(\frac{433.5}{526} \right) X 10 ; Q_3 = 70 + (0.824) X 10 ;$$

$$Q_3 = 70 + 8.24 ; \mathbf{Q_3 = 78.24}$$

$$5. P_{57} = L + \left(\frac{\frac{57N}{100} - cf}{f} \right) X C$$

$$\frac{57N}{100} = \frac{357(2350)}{100} = \frac{133950}{100} = 1339.5 \text{ CI} = 70 - 80; L = 70; \text{ cf} = 1329; f = 526; c = 10$$

$$P_{57} = 70 + \left(\frac{1339.5 - 1329}{526} \right) X 10 ; P_{57} = 70 + \left(\frac{10.5}{526} \right) X 10 ; P_{57} = 70 + (0.019) X 10$$

$$P_{57} = 70 + 0.19 ; \mathbf{P_{57} = 70.19}$$

$$6. P_{75} = L + \left(\frac{\frac{75N}{100} - cf}{f} \right) X C$$

$$\frac{75N}{100} = \frac{75(2350)}{100} = \frac{176250}{100} = 1762.5 \text{ CI} = 70 - 80; L = 70; \text{ cf} = 1329; f = 526; c = 10$$

$$P_{75} = 70 + \left(\frac{1762.5 - 1329}{526} \right) X 10 ; P_{75} = 70 + \left(\frac{433.5}{526} \right) X 10 ; P_{75} = 70 + (0.824) X 10$$

$$P_{75} = 70 + 8.24 ; \mathbf{P_{75} = 78.24}$$

QUARTILE DEVIATION

This measure of dispersion is based on the first quartile and third quartile. It is given by the formula:

$$\mathbf{Absolute\ measure - Q. D = \frac{Q_3 - Q_1}{2}}$$

$$\text{Relative measure - Co-efficient of Q.D} = \frac{\frac{Q_3 - Q_1}{2}}{\frac{Q_3 + Q_1}{2}} = \frac{Q_3 - Q_1}{2} \times \frac{2}{Q_3 + Q_1}$$

$$\text{Hence, Q.D} = \frac{Q_3 - Q_1}{Q_3 + Q_1}$$

1. find the value of Q.D & co-efficient of QD of the following(individual observation)

X 20 23 40 12 30 15 50

Arrange in ascending order : X 12 15 20 23 30 40 50

$$Q_3 = \text{Size of } 3\left(\frac{N+1}{4}\right)^{\text{th}} \text{ Item ; } Q_3 = \text{Size of } 3\left(\frac{7+1}{4}\right)^{\text{th}} \text{ Item ; } Q_3 = \text{Size of } 3\left(\frac{8}{4}\right)^{\text{th}} \text{ Item ;}$$

$$Q_3 = \text{Size of } 3(2)^{\text{th}} \text{ Item ; } Q_3 = \text{Size of } 6^{\text{th}} \text{ Item ; } Q_3 = 40$$

$$Q_1 = \text{Size of } \left(\frac{N+1}{4}\right)^{\text{th}} \text{ Item ; } Q_1 = \text{Size of } \left(\frac{7+1}{4}\right)^{\text{th}} \text{ Item ; } Q_1 = \text{Size of } \left(\frac{8}{4}\right)^{\text{th}} \text{ Item}$$

$$Q_1 = \text{Size of } 2^{\text{th}} \text{ Item ; } Q_1 = 15$$

$$Q. D = \frac{Q_3 - Q_1}{2} ; Q. D = \frac{40 - 15}{2} ; Q. D = \frac{25}{2} ; \mathbf{Q. D = 12.5}$$

$$\text{Co- efficient of Q.D} = \frac{Q_3 - Q_1}{Q_3 + Q_1} ; Q.D = \frac{40 - 15}{40 + 15} ; Q.D = \frac{25}{55} ; \mathbf{Q.D = 0.4545}$$

2. Compute QD and its co-efficient from the data (Discrete series):

Marks (x): 10 20 30 40 50

No. of students (f): 4 7 15 8 7

X	F	Cf
10	4	4
20	7	11
30	15	26
40	8	34
50	7	41
N = 41		

$$Q_3 = \text{Size of } 3\left(\frac{N+1}{4}\right)^{\text{th}} \text{ Item ; } Q_3 = \text{Size of } 3\left(\frac{41+1}{4}\right)^{\text{th}} \text{ Item ; } Q_3 = \text{Size of } 3\left(\frac{42}{4}\right)^{\text{th}} \text{ Item ;}$$

$$Q_3 = \text{Size of } 3(10.5)^{\text{th}} \text{ Item ; } Q_3 = \text{Size of } 31.5^{\text{th}} \text{ Item ; } Q_3 = 40$$

$$Q_1 = \text{Size of } \left(\frac{N+1}{4}\right)^{\text{th}} \text{ Item ; } Q_1 = \text{Size of } \left(\frac{41+1}{4}\right)^{\text{th}} \text{ Item ; } Q_1 = \text{Size of } \left(\frac{42}{4}\right)^{\text{th}} \text{ Item}$$

$$Q_1 = \text{Size of } 10.5^{\text{th}} \text{ Item ; } Q_1 = 20$$

$$Q. D = \frac{Q_3 - Q_1}{2} ; Q. D = \frac{40 - 20}{2} ; Q. D = \frac{20}{2} ; \mathbf{Q. D = 10}$$

$$\text{Co- efficient of Q.D} = \frac{Q_3 - Q_1}{Q_3 + Q_1} ; Q.D = \frac{40 - 20}{40 + 20} ; Q.D = \frac{20}{60} ; \mathbf{Q.D = 0.3333}$$

3. Compute Quartile deviation and co-efficient of QD from the following data (continuous series)

X : 10-20 20-30 30-40 40-50 50-60 60-70 70-80

F : 12 19 5 10 9 6 6

X	F	Cf
10-20	12	12
20-30	19	31
30-40	5	36
40-50	10	46
50-60	9	55
60-70	6	61
70-80	6	67
N = 67		

$$Q_3 = \text{Size of } 3\left(\frac{N}{4}\right)^{\text{th}} \text{ Item; } Q_3 = \text{Size of } 3\left(\frac{67}{4}\right)^{\text{th}} \text{ Item;}$$

$$Q_3 = \text{Size of } 3(16.75)^{\text{th}} \text{ Item; } Q_3 = \text{Size of } 50.25^{\text{th}} \text{ Item;}$$

$$Q_3 - \text{CI} = 50-60 ; L = 50; \frac{3N}{4} = 50.25 ; f = 9; cf = 46; c = 10$$

$$Q_3 = L + \left(\frac{\frac{3N}{4} - cf}{f}\right) X C ; Q_3 = 50 + \left(\frac{50.25 - 46}{9}\right) X 10 ; Q_3 = 50 + \left(\frac{4.25}{9}\right) X 10 ;$$

$$Q_3 = 50 + (0.472) X 10 ; Q_3 = 50 + 4.72 ; \mathbf{Q_3 = 54.72}$$

$$Q_1 = \text{Size of } \left(\frac{N}{4}\right)^{\text{th}} \text{ Item; } Q_1 = \text{Size of } \left(\frac{67}{4}\right)^{\text{th}} \text{ Item; } Q_1 = \text{Size of } 16.75^{\text{th}} \text{ Item;}$$

$$Q_1 - \text{CI} = 20-30 ; L = 20; \frac{N}{4} = 16.75 ; f = 19; cf = 12 c = 10$$

$$Q_1 = L + \left(\frac{\frac{N}{4} - cf}{f}\right) X C ; Q_1 = 20 + \left(\frac{16.75 - 12}{19}\right) X 10 ; Q_1 = 20 + \left(\frac{4.75}{19}\right) X 10 ;$$

$$Q_1 = 20 + (0.25) X 10 ; Q_1 = 20 + 2.5 ; \mathbf{Q_1 = 22.5}$$

$$Q. D = \frac{Q_3 - Q_1}{2} ; Q. D = \frac{54.72 - 22.5}{2} ; Q. D = \frac{32.22}{2} ; \mathbf{Q. D = 16.11}$$

$$\text{Co-efficient of Q.D} = \frac{Q_3 - Q_1}{Q_3 + Q_1} ; Q.D = \frac{54.72 - 22.5}{54.72 + 22.5} ; Q.D = \frac{32.22}{77.22} ; \mathbf{Q.D = 0.417}$$

Mean Deviation

The mean deviation is defined as a statistical measure which is used to calculate the average deviation from the mean value of the given data set.

Mean deviation are two types :

1. Mean deviation about Mean
2. Mean deviation about Median

Individual observation :

1. Calculate Mean deviation about Mean and Median of the Income groups of 5 members given below

Group 1(x) : 4000 4200 4400 4600 4800

M.D about Mean = $\frac{\sum |D|}{n}$ or = $\frac{\sum |d|}{n}$; Where $|D|$ or $|d| = x - \text{mean}$; $n = \text{no. of observations}$.

M. D about Median = $\frac{\sum |D|}{n}$; Where $|D| = x - \text{median}$; $n = \text{no. of observations}$.

Note : where $|d|$ within parallel lines denotes deviations ignoring ('+' '-') signs.

Mean = $\frac{\sum x}{n}$; $\sum x = 4000 + 4200 + 4400 + 4600 + 4800 = 22000$; $n = 5$

Mean = $\frac{22000}{5} = 4400$; $|D| = x - \text{mean}$; $|D| = x - 4400$

X	Mean	$ D = (x - 4400)$
4000	4400	400
4200	4400	200
4400	4400	0
4600	4400	200
4800	4400	400
$\sum D = 1200$		

M.D about Mean = $\frac{\sum |D|}{n} = \frac{1200}{5} = \mathbf{240}$

Co-efficient of M.D about Mean = $\frac{M.D}{\text{Mean}} = \frac{240}{4400} = \mathbf{0.05}$

M. D about Median = $\frac{\sum |D|}{n}$; $|D| = x - \text{Median}$

Median = size of $\left(\frac{N+1}{2}\right)^{\text{th}}$ item; Median = size of $\left(\frac{5+1}{2}\right)^{\text{th}}$ item;

Median = size of $\left(\frac{6}{2}\right)^{\text{th}}$ item; Median = size of 3rd item; Median = 4400

$|D| = x - \text{Median}$; $|D| = x - 4400$

X	Mean	$ D = (x - 4400)$
4000	4400	400
4200	4400	200
4400	4400	0
4600	4400	200

4800	4400	400
$\Sigma D = 1200$		

$$\text{M.D about Median} = \frac{\Sigma|D|}{n} = \frac{1200}{5} = \mathbf{240}$$

$$\text{Co-efficient of M.D about Median} = \frac{M.D}{\text{Mean}} = \frac{240}{4400} = \mathbf{0.05}$$

Discrete series:

2. Calculate Mean deviation about Mean and Median from the following data

X	12	10	13	11	14
F	18	3	12	12	3

M.D about Mean = $\frac{\Sigma f|D|}{n}$ or = $\frac{\Sigma f|d|}{n}$; Where $|D|$ or $|d|$ = x -mean ; n = Total of frequencies or Σf

X	f	Fx	$ D = x$ -mean	f D
12	18	216	0	0
10	3	30	2	6
13	12	156	1	12
11	12	132	1	12
14	3	42	2	6
	N = 48	$\Sigma fx = 576$		$\Sigma f D = 36$

$$\text{Mean} = \frac{\Sigma fx}{n} = \frac{576}{48} = 12$$

$$\text{M.D about Mean} = \frac{\Sigma f|D|}{N} = \frac{36}{48} = \mathbf{0.75}$$

$$\text{Co-efficient of M.D about Mean} = \frac{M.D}{\text{Mean}} = \frac{0.75}{12} = \mathbf{0.06}$$

X	f	CF	$ D = x$ -median	f D
10	3	3	2	6
11	12	15	1	12
12	18	33	0	0
13	12	45	1	12
14	3	48	2	6
	N = 48			$\Sigma f D = 36$

M. D about Median = $\frac{\Sigma f|D|}{n}$; Where $|D|$ = x -median ; n = Total of frequencies or Σf

Median = size of $\left(\frac{N+1}{2}\right)$ th item; Median= size of $\left(\frac{48+1}{2}\right)$ th item;

Median= size of $\left(\frac{49}{2}\right)$ th item; Median= size of 24.5 th item;

$$M. D \text{ about Median} = \frac{\Sigma f|D|}{n} = \frac{36}{48} = \mathbf{0.75}$$

$$\text{Co-efficient of M.D about Median} = \frac{M.D}{Median} = \frac{0.75}{12} = \mathbf{0.06}$$

Continuous series:

Calculate Mean deviation about Mean and Median from the following data

X 10-20 20-30 30-40 40-50 50-60

F 30 20 15 10 5

X	f	M	$d = \frac{M-A}{c}$	Fd ¹	D = x-mean	f D
10-20	30	15	-2	-60	+12.5	375
20-30	20	25	-1	-20	+2.5	0.50
30-40	15	35	0	0	7.5	112.5
40-50	10	45	1	10	17.5	175
50-60	5	55	2	10	27.5	137.5
	N =80			$\Sigma Fd^1 = -60$		$\Sigma f D =850$

$$M.D \text{ about Mean} = \frac{\Sigma f|D|}{n} \text{ or } = \frac{\Sigma f|d|}{n};$$

Where |D| or |d| = m-mean ; n = Total of frequencies or Σf ; m = midpoint if class intervals

$$\text{Mean} = A + \frac{\Sigma Fd^1}{N} \times c ; A = 35; \Sigma Fd^1 = -60; N =80; c =10$$

$$\text{Mean} = 35 + \frac{-60}{80} \times 10 ; \text{Mean} = 35 + (-0.75) \times 10; \text{Mean} = 35 + (-7.5) ; \text{Mean} = 27.5$$

$$M.D \text{ about Mean} = \frac{\Sigma f|D|}{n} = \frac{850}{80} = \mathbf{10.625}$$

$$\text{Co-efficient of M.D about Mean} = \frac{M.D}{Mean} = \frac{10.63}{27.5} = \mathbf{0.38}$$

X	f	M	cf	D = m-median	f D
10-20	30	15	30	10	300
20-30	20	25	50	0	0
30-40	15	35	65	10	150
40-50	10	45	75	20	200
50-60	5	55	80	30	150
	N =80				$\Sigma f D =800$

$$M. D \text{ about Median} = \frac{\Sigma f|D|}{n};$$

Where $|D|$ or $|d| = m - \text{median}$; $n = \text{Total of frequencies or } \Sigma f$; $m = \text{midpoint of class intervals}$.

$$\text{Median} = \text{size of } \left(\frac{N}{2}\right)^{\text{th}} \text{ item} = \text{size of } \left(\frac{80}{2}\right)^{\text{th}} \text{ item} = \text{size of } 40^{\text{th}} \text{ item}$$

$$\text{Median CI} = 20-30; L = 20; \frac{N}{2} = 40; cf = 30 ; f = 20; c = 10$$

$$\text{Median} = L + \left(\frac{\frac{N}{2} - cf}{f}\right) \times C = 20 + \left(\frac{40 - 30}{20}\right) \times 10 = 20 + \frac{10}{2} \times 10 = 20 + (0.5) \times 10 = 20 + 5 = 25$$

$$M. D \text{ about Median} = \frac{\Sigma f|D|}{n} = \frac{800}{80} = \mathbf{10}$$

$$\text{Co-efficient of M. D about Median} = \frac{M.D}{\text{Median}} = \frac{10}{25} = \mathbf{0.4}$$

STANDARD DEVIATION

Standard deviation is the square root of the means of the squared deviations from Arithmetic mean. It is also known as “Root Mean Square Deviation”. It is represented by the symbols “ σ ”. It is given by the following formula

Individual observation

1. Calculate SD from the following observation of marks of 5 students of a tutorial group marks out of 25.

X 8 12 30 15 22

X	$X = X - \bar{X}$	X^2
8	-6	36
12	-2	4
30	-1	1
15	1	1
22	8	64
N = 70		ΣX^2 = 106

$$\bar{X} = \frac{\Sigma X}{N} = \frac{70}{5} = 14;$$

$$\sigma = \sqrt{\frac{\Sigma x^2}{n}}; \text{ where } x = X - \bar{X}; n = \text{no. of observation}$$

$$\sigma = \sqrt{\frac{106}{14}} = \sqrt{21.5} = 4.60$$

Discrete series

2. Calculate SD from the data.

X	0	1	2	3	4	5	6	7
F	12	27	29	19	8	4	1	0

X	F	fx	X = X - \bar{x}	X ²	fX ²
0	12	0	-2	4	48
1	27	27	-1	1	27
2	29	58	0	0	0
3	19	57	1	1	19
4	8	32	2	4	32
5	4	20	3	9	36
6	1	6	4	16	16
7	0	0	5	25	0
	N = 100	Fx = 200			$\Sigma fX^2 = 178$

$$\bar{x} = \frac{\Sigma fX}{N} = \frac{200}{100} = 2 ;$$

$$\sigma = \sqrt{\frac{\Sigma fx^2}{N}} \text{ where } x = x - \bar{x} ; n = \text{total of frequencies}$$

$$\sigma = \sqrt{\frac{\Sigma fx^2}{N}} = \sqrt{\frac{178}{100}} = \sqrt{1.78} = 1.33$$

Continuous series

3. Calculate SD from the following distribution giving 300 telephone calls according to the deviation in seconds.

Duration (x):	0-30	30-60	60-90	90-120	120-150	150-180	180-210
F	9	17	43	82	81	44	24

X	F	M	d'	fd'	d' ²	fd' ²
---	---	---	----	-----	-----------------	------------------

0-30	9	15	-3	-27	9	81
30-60	17	45	-2	-34	4	68
60-90	43	75	-1	-43	1	43
90-120	82	105	0	0	0	0
120-150	81	135	1	81	1	81
150-180	44	165	2	88	4	176
180-210	24	195	3	72	9	216
	N = 300			$\Sigma fd' = 137$		$\Sigma fd'^2 = 665$

$$\sigma = \sqrt{\frac{\Sigma fd'^2}{N} - \left(\frac{\Sigma fd'}{N}\right)^2} \times C; \text{where } d' = \frac{M-A}{C}; A = \text{Assumed Mean};$$

C = With of the class interval; N = Total of frequencies

$$\Sigma fd' = 137; \Sigma fd'^2 = 665; N = 300; c = 30; A = 105$$

$$\bar{x} = A + \frac{\Sigma fd'}{N} \times c = 105 + \frac{137}{300} \times 30 = 105 + 0.45 \times 30 = 105 + 13.7 = \mathbf{118.7}$$

$$\sigma = \sqrt{\frac{\Sigma fd'^2}{N} - \left(\frac{\Sigma fd'}{N}\right)^2} \times C = \sqrt{\frac{665}{300} - \left(\frac{137}{300}\right)^2} \times 30 = \sqrt{2.22 - (0.46)^2} \times 30$$

$$\sigma = \sqrt{2.01} \times 30 = 1.14 \times 30 = \mathbf{42.3}$$

CO – EFFICIENT OF VARIATION

Relative measures of standard deviation is known as co-efficient of variations and is given by the formula:

$$\text{Co-efficient of variation} = \frac{\sigma}{\bar{x}} \times 100$$

It is used in problems where we want to compare the variability or homogeneity or stability or consistency or uniformity of two of more than two series. If co-efficient of variations is less than the other series then it is said to be less variable, More homogenous, more consistence or more stable or more uniform .

$$\text{Co- efficient of standard deviation} = \frac{\sigma}{\bar{x}}$$

1. From the following table of marks obtained by 2 set of students A & B of 100 marks each. Find out who is the more intelligent and who is more consistence.

Marks – A : 25 50 45 30 70 42 36 48 34 60

Marks – B : 10 70 50 20 95 55 42 60 48 80

$$\text{Co-efficient of variation} = \frac{\sigma}{\bar{X}} \times 100$$

Marks – A			Marks - B		
X	X- \bar{x}	X- \bar{x}^2	X	X- \bar{x}	X- \bar{x}^2
25	19	361	10	-43	1849
50	6	36	70	17	289
45	1	1	50	-3	9
30	-14	196	20	-33	1089
70	26	676	95	42	1764
42	-2	4	55	2	4
36	-8	64	42	-11	121
48	4	16	60	7	49
34	-10	100	48	-5	25
60	16	256	80	27	729
440		1710	530		5928

$$\bar{x} = \frac{\Sigma X}{N} = \frac{440}{10} = 44 ;$$

$$\bar{x} = \frac{\Sigma X}{N} = \frac{530}{10} = 53$$

$$\sigma = \sqrt{\frac{\Sigma X^2}{N}} = \sqrt{\frac{1710}{44}}$$

$$\sigma = \sqrt{\frac{\Sigma X^2}{N}} = \sqrt{\frac{5928}{53}}$$

$$\sigma = \sqrt{38.86} = 6.23$$

$$\sigma = \sqrt{111.84} = 10.57$$

$$\text{C.V} = \frac{\sigma}{\bar{X}} \times 100$$

$$\text{C.V} = \frac{6.23}{44} \times 100$$

$$\text{C.V} = 0.1415 \times 100$$

$$\text{C.V} = 14.15$$

$$\text{C.V} = \frac{10.57}{53} \times 100$$

$$\text{C.V} = 0.1994 \times 100$$

$$\text{C.V} = 19.94$$

2. Calculate C. V from the following distribution giving 300 telephone calls according to the deviation in seconds.

Duration (x): 0-30 30-60 60-90 90-120 120-150 150-180 180-210

F : 9 17 43 82 81 44 24

X	F	M	d'	fd'	d' ²	fd' ²
---	---	---	----	-----	-----------------	------------------

0-30	9	15	-3	-27	9	81
30-60	17	45	-2	-34	4	68
60-90	43	75	-1	-43	1	43
90-120	82	105	0	0	0	0
120-150	81	135	1	81	1	81
150-180	44	165	2	88	4	176
180-210	24	195	3	72	9	216
	N = 300			$\Sigma fd' = 137$		$\Sigma fd'^2 = 665$

$$\sigma = \sqrt{\frac{\Sigma fd'^2}{N} - \left(\frac{\Sigma fd'}{N}\right)^2} \times C; \Sigma fd' = 137; \Sigma fd'^2 = 665; N = 300; c = 30; A = 105$$

$$\bar{x} = A + \frac{\Sigma fd'}{N} \times c = 105 + \frac{137}{300} \times 30 = 105 + 0.45 \times 30 = 105 + 13.7 = \mathbf{118.7}$$

$$\sigma = \sqrt{\frac{\Sigma fd'^2}{N} - \left(\frac{\Sigma fd'}{N}\right)^2} \times C = \sqrt{\frac{665}{300} - \left(\frac{137}{300}\right)^2} \times 30 = \sqrt{2.22 - (0.46)^2} \times 30$$

$$\sigma = \sqrt{2.01} \times 30 = 1.417 \times 30 = \mathbf{42.53}$$

$$C.V = \frac{\sigma}{\bar{x}} \times 100; C.V = \frac{42.53}{118.7} \times 100; C.V = 0.3582 \times 100 = \mathbf{35.82}$$

UNIT – III

CORRELATION :

SIMPLE CORRELATION – RANK CORRELATION – CONCURRENT DEVIATIONS METHODS.

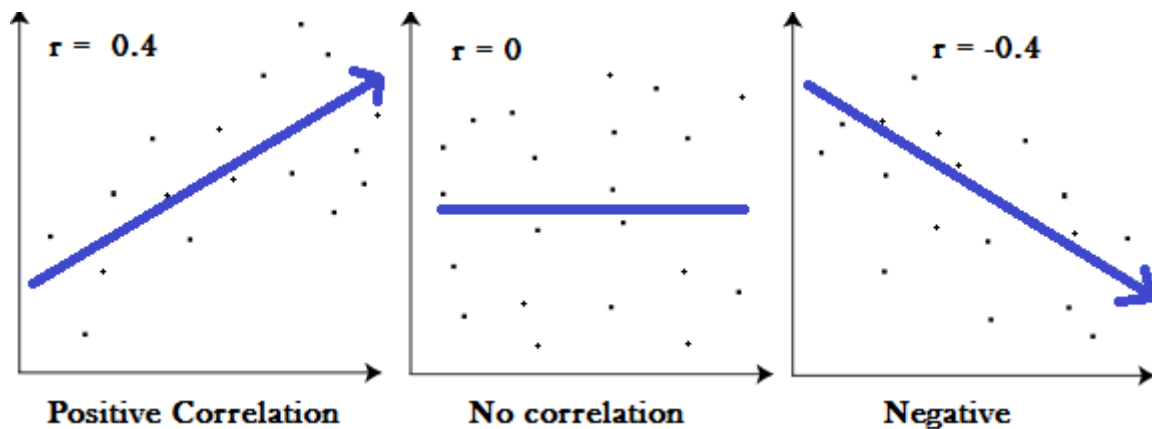
Correlation:

Correlation is used to test relationships between quantitative variables or categorical variables. In other words, it's a measure of how things are related. The study of how variables are correlated is called correlation analysis.

Correlation coefficients are used to measure how strong a relationship is between two variables. There are several types of correlation coefficient, but the most popular is Pearson's. Pearson's correlation (also called Pearson's R) is a correlation coefficient commonly used in linear regression.

Correlation coefficient formulas are used to find how strong a relationship is between data. The formulas return a value between -1 and 1, where:

- 1 indicates a strong positive relationship.
- -1 indicates a strong negative relationship.
- A result of zero indicates no relationship at all.



Graphs showing a correlation of -1, 0 and +1

Meaning

- A correlation coefficient of 1 means that for every positive increase in one variable, there is a positive increase of a fixed proportion in the other. For example, shoe sizes go up in (almost) perfect correlation with foot length.
- A correlation coefficient of -1 means that for every positive increase in one variable, there is a negative decrease of a fixed proportion in the other. For example, the amount of gas in a tank decreases in (almost) perfect correlation with speed.
- Zero means that for every increase, there isn't a positive or negative increase. The two just aren't related.

The absolute value of the correlation coefficient gives us the relationship strength. The larger the number, the stronger the relationship. For example, $|-0.75| = 0.75$, which has a stronger relationship than 0.65 .

Correlation Coefficient

The correlation coefficient, r , is a summary measure that describes the extent of the statistical relationship between two interval or ratio level variables. The correlation coefficient is scaled so that it is always between -1 and +1. When r is close to 0 this means that there is little relationship between the variables and the farther away from 0 r is, in either the positive or negative direction, the greater the relationship between the two variables.

The two variables are often given the symbols X and Y . In order to illustrate how the two variables are related, the values of X and Y are pictured by drawing the scatter diagram, graphing combinations of the two variables. The scatter diagram is given first, and then the method of determining Pearson's r is presented. From the following examples, relatively small sample sizes are given. Later, data from larger samples are given.

Scatter Diagram

A scatter diagram is a diagram that shows the values of two variables X and Y, along with the way in which these two variables relate to each other. The values of variable X are given along the horizontal axis, with the values of the variable Y given on the vertical axis.

Later, when the regression model is used, one of the variables is defined as an independent variable, and the other is defined as a dependent variable. In regression, the independent variable X is considered to have some effect or influence on the dependent variable Y. Correlation methods are symmetric with respect to the two variables, with no indication of causation or direction of influence being part of the statistical consideration. A scatter diagram is given in the following example. The same example is later used to determine the correlation coefficient.

Types of Correlation

The scatter plot explains the correlation between the two attributes or variables. It represents how closely the two variables are connected. There can be three such situations to see the relation between the two variables –

- Positive Correlation – when the value of one variable increases with respect to another.
- Negative Correlation – when the value of one variable decreases with respect to another.
- No Correlation – when there is no linear dependence or no relation between the two variables.

1. Calculate Karl Pearson’s co-efficient of Correlation for the following data:

X: 9 8 7 6 5 4 3 2 1
 Y: 15 16 14 13 11 12 10 8 9

X	Y	$x - \bar{x}$	X^2	$y - \bar{y}$	Y^2	XY
9	15	4	16	3	9	12
8	16	3	9	4	16	12
7	14	2	4	2	4	4
6	13	1	1	1	1	1
5	11	0	0	-1	1	0
4	12	-1	1	0	0	0

3	10	-2	4	-2	4	4
2	8	-3	9	-4	16	12
1	9	-4	16	-3	9	12
$\Sigma X = 45$	$\Sigma Y = 108$		$\Sigma X^2 = 60$		$\Sigma Y^2 = 60$	$\Sigma XY = 57$

Formula for correlation of $r = \frac{\Sigma XY}{\sqrt{\Sigma X^2 \Sigma Y^2}}$ where $X = x - \bar{x}$; $Y = y - \bar{y}$; $\bar{x} = \frac{\Sigma X}{N}$; $\bar{y} = \frac{\Sigma Y}{N}$

$$\bar{x} = \frac{\Sigma X}{N} = \frac{45}{9} = 5; \bar{y} = \frac{\Sigma Y}{N} = \frac{108}{9} = 12; \Sigma XY = 57; \Sigma X^2 = 60; \Sigma Y^2 = 60$$

$$r = \frac{\Sigma XY}{\sqrt{\Sigma X^2 \Sigma Y^2}} = \frac{57}{\sqrt{60 \times 60}} = \frac{57}{\sqrt{3600}} = \frac{57}{60} = 0.95 \text{ (perfect or positive correlation)}$$

2. Calculate Correlation for the following data:

X: 10 6 9 10 12 13 11 9

Y: 9 4 6 9 11 13 8 4

X	Y	$x - \bar{x}$	X^2	$y - \bar{y}$	Y^2	XY
10	9	0	0	2	4	0
6	4	-4	16	-2	4	64
9	6	-1	1	1	1	1
10	9	0	0	2	4	0
12	11	2	4	4	16	64
13	13	3	9	5	25	225
11	8	1	1	3	9	9
9	4	-1	9	1	1	1
$\Sigma X = 80$	$\Sigma Y = 64$		$\Sigma X^2 = 36$		$\Sigma Y^2 = 64$	$\Sigma XY = 364$

$$\bar{x} = \frac{\Sigma X}{N} = \frac{80}{8} = 10; \bar{y} = \frac{\Sigma Y}{N} = \frac{64}{8} = 8; \Sigma XY = 364; \Sigma X^2 = 36; \Sigma Y^2 = 64$$

$$r = \frac{\Sigma XY}{\sqrt{\Sigma X^2 \Sigma Y^2}} = \frac{364}{\sqrt{36 \times 64}} = \frac{364}{\sqrt{2304}} = \frac{364}{48} = 7.58 \text{ (perfect or positive correlation)}$$

SPEARMAN'S RANK CORRELATION

The Spearman's rank-order correlation is the nonparametric version of the Pearson product-moment correlation. Spearman's correlation coefficient, (ρ , also signified by r_s) measures the strength and direction of association between two ranked variables.

1. The two judges in a Beauty competition rank the 12 entries as follows:

Judge I : 1 2 3 4 5 6 7 8 9 10 11 12

Judge II: 12 9 6 10 3 5 4 7 8 2 11 1

Judge I	Judge II	d = R ₁ - R ₂	d ²
1	12	-11	121
2	9	-7	49
3	6	-3	9
4	10	-6	36
5	3	2	4
6	5	1	1
7	4	3	9
8	7	1	1
9	8	1	1
10	2	8	64
11	11	0	0
12	1	11	121
			$\Sigma d^2 = 416$

$$R = 1 - \frac{6\Sigma d^2}{N(N^2-1)} ; \Sigma d^2 = 416; N = 12$$

$$R = 1 - \frac{6(416)}{12(12^2-1)} = 1 - \frac{2496}{12(144-1)} = 1 - \frac{2496}{12(143)} = 1 - \frac{2496}{1716} = 1 - 1.4545$$

R = 0.4545 (It is a negative correlation)

Calculate Spearman's Rank correlation from the following data: (Tied or Repeated ranks)

X : 32 35 49 60 43 37 43 49 10 20

Y : 40 30 70 20 30 50 72 60 45 25

X	Y	R ₁	R ₂	R ₁ - R ₂	d ²
32	40	8	6	2	4
35	30	7	7.5	-0.5	0.25
49	70	2.5	2	0.5	0.25
60	20	1	10	-9	81
43	30	4.5	7.5	-3	9
37	50	6	4	2	4
43	72	4.5	1	3.5	12.25
49	60	2.5	3	-0.5	0.25
10	45	10	5	5	25
20	25	9	9	0	0
					Σd ² = 136

$$P = 1 - \frac{6\left\{\Sigma d^2 + \frac{1}{12}m(m^2-1) + \frac{1}{12}m(m^2-1) + \frac{1}{12}m(m^2-1) + \frac{1}{12}m(m^2-1) \dots\right\}}{N(N^2-1)}$$

$$P = 1 - \frac{6\left\{136 + \frac{1}{12}x2(2^2-1) + \frac{1}{12}x2(2^2-1) + \frac{1}{12}x2(2^2-1)\right\}}{10(10^2-1)}$$

$$P = 1 - \frac{6\{136 + 0.16(4-1) + 0.16(4-1) + 0.16(4-1)\}}{10(100-1)} = 1 - \frac{6\{136 + 0.16(3) + 0.16(3) + 0.16(3)\}}{10(99)}$$

$$P = 1 - \frac{6\{136 + 0.48 + 0.48 + 0.48\}}{990} = 1 - \frac{6\{137.44\}}{990} = 1 - \frac{824.64}{990} = 1 - 0.83 = 0.17 \text{ (positive correlation)}$$

CONCURRENT DEVIATION

Correlation co-efficient by the method of concurrent deviations is given by the formula

$r_c = \pm \sqrt{\pm \left(\frac{2c-n}{n} \right)}$ where C = No. of concurrent deviations (the no. of positive signs obtained after multiplying Dx & Dy)

Dx & Dy = directions of change of x & y

N = N - 1 (N = no. of pairs of observations)

Note : the significance of \pm signs both inside the square root and outside the squareroot is that we cannot the underroot of (-) sign. Therefore if, $\left(\frac{2c-n}{n} \right)$ is negative , this negative value multiplied with the (-) sign inside, would make it positive . And we take the ssquareroot . But the ultimate result would be negative. If $\left(\frac{2c-n}{n} \right)$ is positive then of course, we get a positive value of the co-efficient of correlation.

The correlation co-efficient calculated using this method is also sometimes known as co-efficient of Concurrent Deviations.

1.Calculate Concurrent Deviations for the following data;

English:	72	70	65	50	49	49	40	39	38	35
Statitics:	95	80	81	45	49	50	60	61	40	38

X	Y	Dx	Dy	DxDy
72	95			
70	80	-	-	+
65	81	-	+	-
50	45	-	-	+
49	49	-	+	-
49	50	0	+	0
40	60	-	+	-

39	61	-	+	-
38	40	-	-	+
35	38	-	-	+
				C = 4

$$C = 4 ; N = n-1 = 10 - 1 = 9$$

$$r_{c=\pm} = \sqrt{\pm \left(\frac{2c-n}{n} \right)} = r_{c=\pm} = \sqrt{\pm \left(\frac{2(4)-9}{9} \right)} = \pm \sqrt{\pm \left(\frac{8-9}{9} \right)} = \pm \sqrt{\pm \left(\frac{-1}{9} \right)}$$

$$r_{c=\pm} = \sqrt{\pm 0.11} = \sqrt{0.11} = \pm 0.33$$

2. Calculate Concurrent deviation from the data:

Year: 1983 1984 1985 1986 1987

Supply: 140 154 160 140 170

Price: 180 160 190 200 210

X	Y	Dx	Dy	DxDy
140	180			
154	160	+	-	-
160	190	+	+	+
140	200	-	+	-
170	210	+	+	+
				C = 2

$$C = 2 ; N = N-1 = 5-1 = 4$$

$$r_{c=\pm} = \sqrt{\pm \left(\frac{2c-n}{n} \right)} = r_{c=\pm} = \sqrt{\pm \left(\frac{2(2)-4}{4} \right)} = \pm \sqrt{\pm \left(\frac{4-4}{4} \right)} = \pm \sqrt{\pm \left(\frac{0}{4} \right)}$$

$$r_{c=\pm} = \sqrt{\pm 0} = 0 \text{ (It is positive)}$$