UNIT II

Descriptive Statistics: Measures of central tendency: Arithmetic Mean, Median, Mode, Geometric mean, Harmonic mean; Measures of Dispersion:Range, Quartile deviation, Mean deviation, Standard deviation. Definition of Moments; Measures of Skewness: Karl pearson's coefficient of skewness, Bowley's coefficient of skewness; Kurtosis.

2.1 Measures of Central Tendency

Q1. Define Average. What are the functions of an average.

Ans : **(Imp.)**

Introduction

Summarisation of the data is a necessary function of any statistical analysis. As a first step in this direction, the huge mass of unwieldy data are summarised in the form of tables and frequency distributions. In order to bring the characteristics of the data into sharp focus, these tables and frequency distributions need to be summarised further. A measure of central tendency (or) an average is very essential and an important summary measure in any statistical analysis. An average is a single value which can be taken as representative of the whole distribution.

Definitions :

The average of a distribution has been defined in various ways. Some of the important definitions are :

(i) **According to Clark and Sekkade,** "An average is an attempt to find one single figure to describe the whole of figures".

(ii) **According to Murry R. Spiegal,** "Average is a value which is typical or representative of a set of data".

(iii) **According to Croxton and Cowden,** "An average is a single value within the range of the data that is used to represent all the values in the series. Since an average is somewhere within the range of data it is sometimes called a measure of central value".

(iv) **According to Sipson and Kafka,** "A measure of central tendency is a typical value around which other figures congregate".

Functions

1. **To present huge mass of data in a summarised from**

   It is very difficult for human mind to grasp a large body of numerical figures. A measure of average is used to summarise such data into a single figure which makes it easier to understand and remember.

2. **To facilitate comparison**

   Different sets of data can be compared by comparing their averages. For example, the level of wages of workers in two factories can be compared by mean (or average) wages of workers belonging to each of them.

3. **To help in decision-making**

   Most of the decisions to be taken in research, planning, etc., are based on the average value of certain variables. For example, if the average monthly sales of a company are falling, the sales manager may have to take certain decisions to improve it.

Q2. State the objectives of Average.

Ans :

i) **Representative of the group**

   The human mind cannot retain all details of large number of activities and their inter-relations, so averages are a must. An average represents all the features of a group, hence the results about the whole group can be deduced from it.

ii) Brief description

An average gives us simple and brief description of the main features of the whole data.

iii) Helpful in comparison

The measures of central tendency or averages reduce the data to a single value which is highly useful for making comparative studies.

For example, comparing the per capita income of two countries, we can conclude that which country is richer.

iv) Helpful in formulation of policies

Averages help to develop a business in case of a firm or help the economy of a country to develop. For example, in case of an aviation company, the management will be interested to know about the average number of persons boarding plane on the desired certain route. In such a case, a finance minister or finance secretary would apply some economic measures to increase per capita income if he feels that it is lowest as compared to other developed country's per capita income.

v) Basis of other statistical Analysis

Other statistical devices such as mean deviation, co-efficient of variation, co-relation, analysis of time series and index numbers are also based on the averages; and hence the use of averages becomes compulsory.

Q3. Explain the significance of an average.

Ans :

- The measures of central tendency provides a single value which represents the features of the whole group.
- The single/central value is calculated by contracting the mass of data into one single value.
- It helps us to understand briefly about the entire group through one single value.
- The measures of central tendency also provides a comparison which can be carried out at a point of time or over a period of time.
- The comparison may be made between two or more data sets related to each other.
- For example, the percentage profit of different firms producing pens. But, at the time of drawing presumptions from the data, the factors influencing data at different points should also be considered and it should also be compared with the same value of measure of central tendency.
- One measure of central tendency cannot be compared with the different measure of central tendency i.e., one set of values of arithmetic mean cannot be compared with the other set of values of median or mode.
- It implies that a data set should contain only one value of central tendency.
- Every measure of central tendency has its own importance and process for calculation and application.

Q4. What are the characteristics of a good average ?

Ans :

1. It should be rigidly defined

An average should be rigidly defined so that there is no scope for confusion or manipulation. The average value will become very unstable and non representative of the base data if it is not well defined. It is ideal to use an average that is mathematically defined by way of formula.

2. It should be easy to calculate and simple to follow

Calculation of an average should be simple to understand. It should be easy to calculate, preferably without the help of calculators. If an average is too complex to understand and calculate, its use will be very limited. It should also be capable of expression in simple numerical terms without advanced mathematical intricacies.

3. It should be based on all observations in the series

An average will be truly representative of the whole mass of data if it is computed from all the observations.

4. **It should not be affected much by an extreme values**

An average will be representative of the data only if it can set off extreme values against each other. A few very small or very large observations should not unduly affect the value of a good average.

5. **It should be capable of further statistical processing**

An average should be capable of being used in calculation of other statistical measures such as standard deviation, correlation, etc.

6. **It should be capable of further algebraic treatment**

An average should lend itself readily to further algebraic treatment. For example, if averages of two (or) more sets of data are known, it should be possible to obtain the average of combined group.

7. **It should possess sampling stability**

An average should be least affected by sampling. If we take independent random samples of the same size from a given population and compute the average for each of these samples, the values so obtained from different samples should not be very different from one another.

8. **It should be representative of the data**

A good average should represent maximum characteristics of the underlying data.

Q5. What are the limitations of averages?

Ans :

1. **Misleading Conclusions**

Average is a single numerical figure representing the characteristics of a given distribution. This number is vulnerable to errors in interpretation and can lead to misleading conclusions. For example, if the maximum and minimum temperatures of a particular city are 48° F and 2° F respectively, the average may still work out to 24° F. Looking at the average, the weather conditions might look very comfortable, but that is not really the case.

2. **Choice of Average**

There are different types of averages. Different types of averages are suitable for different objectives. The utility of average depends on a proper and judicious choice of the average. A wrong choice of the average might lead to erroneous results.

3. **Incomplete Picture**

An average does not provide the complete picture of a distribution. There may be number of distribution having the same average but differing widely in their structure and constitution. To form a complete idea about the distribution, the measures of central tendency are to be supplemented by some more measures such as dispersion, skewness and kurtosis.

4. **Inadequate Representation of Data:** In certain types of distribution like U- shaped or J-shaped distribution, an average, which is only a single point of concentration, does not adequately represent the centre series.

5. **Absurd Conclusions**

Sometimes, an average might throw up very absurd results that are not realistic. For example, the average size of a family might work out to a fractional number, which is not realistic.

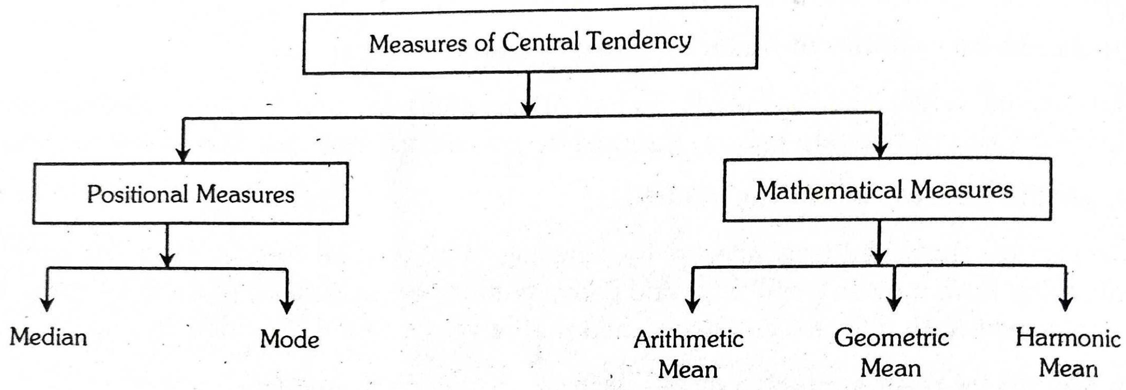
Q6. Explain the different types of averages.

(OR)

What are the various measures of central tendency?

Ans :

Various measures of averages can be classified into the following two categories :



2.2 Arithmetic Mean

Q7. What is arithmetic mean? State the merits and demerits of arithmetic mean.

Ans :

Meaning

Arithmetic Average (or) Mean of a series is the figure obtained by dividing the total "Values of the various items by their number. In other words it is the sum of the values divided by their number. Arithmetic means is the most widely used measure of central tendency.

Merits

1. It is rigidly defined. Hence, different interpre-tations by different persons are not possible.
2. It is easy to understand and easy to calculate. In most of the series it is determinate and its value is definite.
3. It takes all values into consideration. Thus, it is more representative.
4. It can be subjected to further mathematical treatment. The properties of Arithmetic mean are separately explained elsewhere in the chapter.
5. It is used in the computation of various other statistical measures.
6. It is possible to calculate arithmetic average even if some of the details of the data are lacking. For example, it can be known even when only the number of items and their aggregate value are known, and details of various items are not available. Similarly, the aggregate value of items can be calculated if the number of items and the average are known.
7. Of all averages, arithmetic average is least affected by fluctuations of sampling. Thus, it is the most stable measure of central tendency.
8. It provides a good basis for comparison.
9. Arithmetic mean is impacted by every observation. It gives weight to all items in direct proportion to their size.

Demerits

While Arithmetic mean satisfies most of the conditions of an ideal average, it suffers from certain drawbacks. Some of the demerits or limitations of Arithmetic Mean are listed below:

1. It cannot be determined by inspection.
2. It cannot be located graphically.
3. It cannot be used in the study of qualitative phenomena.
4. It can be significant impacted by extreme values and may lead to erroneous conclusions. Abnormal items may considerably affect this average, particularly when the number of items is not large. Thus, it is desirable not to use arithmetic average when the distribution is unevenly spread.

Q8. How to calculate arithmetic mean for individual series ?

Ans :

Individual Series

The process of computing mean in case of individual observations (i.e., where frequencies are not given) is very simple. Add together the various values of the variable and divide the total by the number of items. Symbolically :

$$\bar{X} = \frac{X_1 + X_2 + X_3 + ... + X_n}{N}$$

(OR)

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

Here

X = Arithmetic Mean

ΣX = Sum of all the values of the variable X, i.e., $X_1, X_2, X_3.....X_n$

N = Number ...tions.

Steps

The formula invo... steps in calculating mean :

i) Add together all the values of the variable X and obtain the total, i.e., XX.

ii) Divide this total by the number of the observations, *i.e., N.*

Shortcut method

The arithmetic mean can be calculated by using what is known as an arbitrary origin. When deviations are taken from an arbitrary origin, the formula for calculating arithmetic mean is

$$\bar{X} = A + \frac{\Sigma d}{N} Cx$$

A is the assumed mean and

d is the deviation of items from assumed mean

i.e., d = (X – A*).

Steps

1. Take an assumed mean.
2. Take the deviations of items from the assumed mean and denote these deviations by d.
3. Obtain the sum of these deviations,
4. Apply the formula : $X = A + \frac{\sum d}{N}$.

Q9. How to calculate Arithmetic mean for discrete series ?

Ans :

Calculation of Arithmetic Mean

In discrete series arithmetic mean may be computed by applying

i) Direct method (or)

ii) Short-cut method

i) Direct method

The formula for computing mean is $\bar{X} = \frac{\Sigma fX}{N}$

Where,

f = Frequency;

X = The variable in question;

N* = Total number of observations, i.e., Σf.

Steps

i) Multiply the frequency of each row with the variable and obtain the total If ΣX.

ii) Divide the total obtained by step (i) by the number of observations, i.e., total frequency.

ii) Short-cut Method

According to this method,

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

where A = Assumed mean; d = (X – A); N - Total number of observations, i.e., Σf.

Steps

i) Take an assumed mean.

ii) Take the deviations of the variable X from the assumed mean and denote the deviations by d

iii) Multiply these deviations with the respective frequency and take the total Σfd.

iv) Divide the total obtained in third step by the total frequency.

Q10. How to calculate arithmetic mean for continuous series ?

Ans :

Calculation of Arithmetic Mean

In continuous series, arithmetic mean may be computed by applying any of the following methods:

i) Direct method,

ii) Short-cut method.

i) Direct Method

When direct method is used

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

where

m = Mid-point of various classes

f = The frequency of each class

N = The total frequency.

Steps

i) Obtain the mid-point of each class and denote it by m.

ii) Multiply these mid-points by the respective frequency of each class and obtain the total Σfm.

iii) Divide the total obtained in step (i) by the sum of the frequency, i.e., N.

ii) Short-cut Method

When short-cut method is used, arithmetic; computed by applying the following formula :

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

where A - assumed mean; d = deviations of mid-points from assumed mean, i.e., (m - A): N = total number of observations.

Steps

i) Take an assumed mean.

ii) From the mid-point of each class deduct the assumed mean.

iii) Multiply the respective frequencies of each class by these deviations and obtain the total Σfd.

iv) Apply the formula: $X = A + \frac{\sum fd}{N}$

PROBLEMS

1. **Calculate the arithmetic mean of the monthly incomes of the families in a certain locality in Delhi.**

Family	A	B	C	D	E	F	G	H	I	J
Income (f)	85	70	10	75	500	8	42	250	40	36

Sol :

$$\text{Arithmetic Mean } (\overline{X}) = \frac{\sum X}{N} = \frac{85+70+10+75+500+8+42+250+40+36}{10}$$

$$= 1116/10 = 111.6$$

∴ The arithmetic mean of the monthly incomes in the locality is Rs. 111.6.

2. **Calculate arithmetic mean of the following data by direct, short-cut methods :**

Roll No.	1	2	3	4	5	6	7	8	9	10
Marks (Out of 100)	50	65	37	29	92	43	81	36	52	45

Sol :

Calculation of Arithmetic mean by

i) **Direct Method**

$$\text{Arithmetic mean } (\overline{X}) = \frac{\sum X}{N} = \frac{50+65+37+29+92+43+81+36+52+45}{10} = \frac{530}{10} = 53$$

∴ Average marks of the class = 53 marks.

ii) **Short Cut Method**

X	X – A = d
50	– 42
65	– 27
37	– 55
29	– 63
92 A	0
43	– 49
81	– 11
36	– 56
52	– 40
45	– 47
	– 390

$$\bar{X} = A + \frac{\Sigma d}{N} = 92 + \frac{-390}{10}$$

$$= 92 + (-39) = 53.$$

3. **Calculate Arithmetic mean from the following data.** (Imp.)

Marks	0-10	10-20	20-30	30-40	40-50	50-60	60-70	70-80	80-90	90-100
No. of Students	33	53	108	221	153	322	439	526	495	50

Sol :

Calculation of Arithmetic mean

Marks	No. of students (f)	Midvalue (M)	fM
0-10	33	5	165
10-20	53	15	795
20-30	108	25	2700
30-40	221	35	7735
40-50	153	45	6885
50-60	322	55	17,710
60-70	439	65	28,535
70-80	526	75	39,450
80-90	495	85	42,075
90-100	50	95	4750
	$\Sigma f = 2400$		$\Sigma fM = 1,50,800$

$$\text{Arithmetic mean} = \frac{\Sigma fM}{\Sigma f} = \frac{1,50,800}{2400} = 62.83$$

2.3 Median

Q11. Define median. What are the characteristics of median?

Ans :

Meaning

If a group of N observations is arranged in ascending or descending order of magnitude, then the middle value is called median of these observations and is denoted by M.

That is, $M = \frac{N+1}{2}$ th observation.

Definition

According to Croxton and Cowden, "The median is that value which divides a series so that one half or more of the items are equal to or less than it and one half or more of the items are equal to or greater than it."

Characteristics

- Unlike the arithmetic mean, the median can be computed from open-ended distributions. This is because it is located in the median class-interval, which would not be an open-ended class.
- The median can also be determined graphically whereas the arithmetic mean cannot be ascertained in this manner.
- As it is not influenced by the extreme values, it is preferred in case of a distribution having extreme values.
- In case of the qualitative data where the items are not counted or measured but are scored or ranked, it is the most appropriate measure of location.

Q12. What are the advantages and disadvantages of median?

Ans :

Advantages

(i) The median, unlike the mean, is unaffected by the extreme values of the variable.

(ii) It is easy to calculate and simple to understand, particularly in a series of individual observations and a discrete series. .

(iii) It is capable of further algebraic treatment. It is used in calculating mean deviation.

(iv) It can be located by inspection, after arranging the data in order of magnitude.

(v) It can be determined graphically.

(vi) Median can be calculated in case of open-end classes.

(vii) Median is defined rigidly.

Disadvantages

(i) For calculation, it is necessary to arrange the data, other averages do not need any such arrangement.

(ii) It is amenable to algebraic treatment in a limited sense. Median cannot be used to calculate the combined median of two or more groups, like mean.

(iii) Median is affected more by sampling fluctuations than the mean.

(iv) It is not based on all observations. So it is a positional average.

Q13. How median is calculated for individual series ?

Ans :

(i) Arrange the data in ascending or descending order of magnitude. (Both arrangements would give the same answer.)

(ii) In a group composed of an odd number of values such as 7, add 1 to the total number of values and divide by 2. Thus, 7+1 would be 8 which divided by 2 gives 4 the number of the value starting at either end of the numerically arranged groups will be the median value.

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

(iii) In a group composed an even number of values, then median $= \frac{N}{2} + 1^{th}$ item

(or)

$$\frac{\text{Two middle values}}{2}$$

Q14. How median is calculated for discrete series ?

Ans :

(i) Arrange the data in ascending or descending order of magnitude,

(ii) Find out the cumulative frequencies.

(iii) Apply the formula: Median = Size of $\left(\frac{N+1}{2}\right)^{th}$ item .

(iv) Now look at the cumulative frequency column and find that total which is either equal to Size of $\frac{N+1}{2}$ or next higher to that and determine the value of the variable corresponding to it. That gives the value of median.

Q15. How median is calculated for continuous series ?

Ans :

1. Arrange the given data in ascending order
2. Calculate cumulative frequency
3. Find N_1 as $N_1 = N/2$ for Median
4. In C.f find the value equal to or just greater than N_1
5. Find f and L just one step below the C.f

$$\text{Median} = L + \frac{\frac{N}{2} - F}{f} \times C$$

L = Lower limit of the median class,

F = Cumulative frequency of the class preceding the median class (or) sum of the frequencies of all classes lower than the median class.

f = Simple frequency of the median class

C = The class interval of the median class.

PROBLEMS

4. Calculate Median from the following.

61 62 63 61 63 64 64 60 65 63 64 65 66 64

Sol :

Step 1

Arrange the data in ascending order for the given i.e.,

60, 61, 61, 62, 63, 63, 63, 64, 64, 64, 65, 65, 65, 66

Step 2

Apply the formulae for :

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

$$= \frac{14+1}{2} = \text{Size of 7.5th item}$$

when 7.5 is equalled to 7th and 8th items of the data. Hence 7th = 63, 8th item = 64

$$\text{median} = \frac{63+64}{2} = 63.5$$

5. Calculate median from the following.

22, 26, 14, 30, 18, 11, 35, 41, 12, 32

Sol :

Step I : Arrange the data in ascending order i.e.,

11. 12, 14, 18, 22. 26, 30, 32, 35, 41

Step II : Calculation of given ascending to formulae

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

$$= \frac{10+1}{2} = \frac{11}{2} = \text{Size of } 5.5\text{th item}$$

$$\text{Median} = \frac{5\text{th item} + 6\text{th item}}{2} = \frac{22+26}{2} = \frac{48}{2} = 24.$$

6. From the following data, find the value of median.

Income (₹)	200	250	130	270	300	230
No. of Persons	34	36	26	30	16	40

Sol :

Step I : Arrange the data in ascending order.

Income (Rs.) (x)	No. of Persons (F)	Cumulative Frequency (CF)
130	26	26
200	34	60
230	40	100
250	36	136
270	30	166
300	16	182
	N=182	

Step II: Apply the formulae for determination of media.

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

$$\text{Median} = \text{Size of } \frac{182+1}{2} \text{th item}$$

$$\text{Median} = \text{Size of } \frac{183}{2} = 91.5 \text{ item}$$

Median Size of 91.5 is representing in 100 at cumulative frequency which is representing in corresponding (x) column is 230 in income level.

$\therefore$ Median = 230.

7. **From the following data, compute Median**

Class Interval	0-10	10-20	20-30	30-40	40-50
Frequency	15	100	170	120	40

Sol : (Imp.)

Marks	Frequency	Cumulative Frequency (CF)
0 - 10	15	15
10 - 20	100	115 CF
L 20 - 30	170 f	285
30 - 40	120	405
40 - 50	40	445

$$\text{Median} = L + \frac{\frac{N}{2} - F}{f} \times C$$

$$\text{Median class interval} = \text{Size of } \frac{N}{2}\text{th item}$$

$$\text{Median class interval} = \text{Size of } = \frac{445}{2} = 222.50 \text{ item}$$

∴ 222.50th item consisting in CF 285 which is representing in corresponding class interval is 20-30.

$$\text{Median} = 20 + \left[\frac{\frac{N}{2} - F}{f}\right] \times C = 20 + \left(\frac{222.50 - 115}{170}\right) \times 10$$

$$\text{Median} = 20 + 0.632 \times 10$$

$$\text{Median} = 20 + 6.32$$

$$\text{Median} = 26.32$$

2.4 Mode

Q16. Define mode.

Ans :

Meaning

Mode may be defined as the value that occurs most frequently in a statistical distribution or it is defined as that exact value in the ungrouped data if each sample which occurs most frequently.

Definitions

(i) **According to Croxton and Cowden,** "The mode of a distribution is the value at the point around which the items tend to be most heavily concentrated. It may be regarded as the most typical of a series of values."

(ii) **According to A.M. Tuttle,** "Mode is the value which has the greatest frequency density in its immediate neighbourhood."

(iii) **According to Zizek,** "The mode is the value occurring most frequently in a series of items and around which the other items are distributed most densely."

Every distribution cannot have a unique value of Mode. It can have two or even more than two modal values. Such distributions are known as Uni-Modal, Bimodal and Multi Modal.

Its graphical representation is given below.

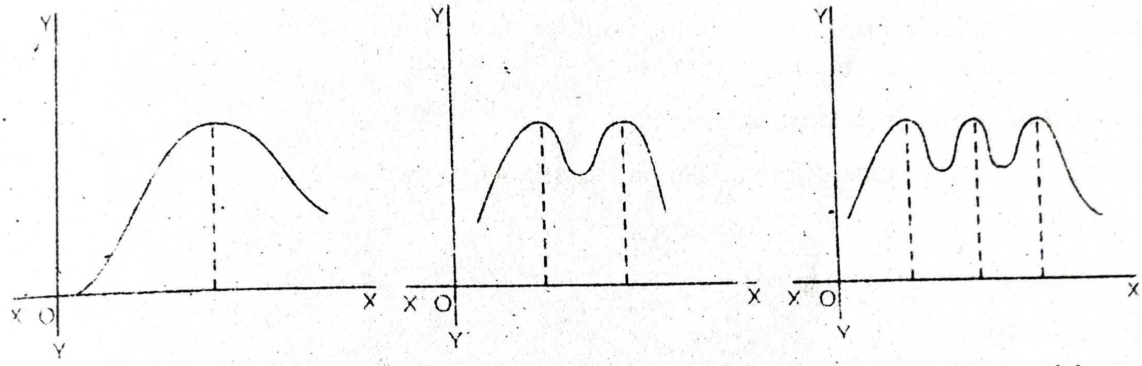


Fig.: Unimodel Fig.: Bimodal Fig.: Multimodal

Q17. Explain the merits and demerits of Mode.

Ans :

Merits

1. Mode is easy to understand and calculate.
2. It is not influenced much by items on the extremes.
3. It can be located even if the class-intervals are of unequal magnitudes, provided the modal class and the preceding and succeeding it are of the same magnitude.
4. It can be computed for distributions which have open end classes.
5. Mode is not an isolated value like the median. It is the term that occurs most in the series.
6. Mode is not a fictional value that is not found in the series.
7. It can be determined by inspection.
8. It can be located graphically
9. It has wide business application

Demerits

1. Calculation of Mode does not consider all the items of the series. Thus, it is not fully representative of the entire data.
2. It is not rigidly defined.
3. It is not capable of further mathematical treatment.

4. Mode is sometimes indeterminate. There may be 2 (Bi-modal) or more (Multimodal) values.
5. Mode is significantly impacted by fluctuations of sampling. Hence, it is less reliable.
6. Mode is considerably influenced by the choice of grouping. A change in the size of the class interval will change the value of the mode". It is a very unstable average and its true value is difficult to determine.

Q18. How mode is calculated for individual and discrete series.

Ans :

Determination of Mode in Individual Series

The steps involved in calculating mode for individual series are as follows,

1. First arrange the data in ascending or descending order.
2. Check which value is repeating maximum number of times. The value repeating maximum number of times is considered as the value of mode.

Determination of Mode in Discrete Series

In discrete series the value of mode can be determined in two ways,

(i) Inspection

(ii) Grouping and analysis table.

(i) Inspection: Steps

(a) Maximum value or highest value is selected from frequency column

(b) The value corresponding to highest frequency value is considered as mode.

(ii) Grouping and Analysis Table

If there exists an error of maximum frequency, frequency has small value preceding or succeeding it and items are highly focused on any one side then under such case grouping and analysis table are prepared.

Grouping Table

It consists of six columns as follows,

- In column 1-highest frequency is selected and highlighted with a circle or a tick (✓) marks.
- In column 2-frequency values are grouped in two's
- In column 3-Ignore first frequency and group the remaining in two's.
- In column 4-frequencies are grouped in three's
- In column 5-Ignore first frequency and group the remaining frequency values in three's.
- In column 6-Ignore first two frequencies and group remaining frequencies in three's

After preparing the grouping table the maximum value in each column should be marked with a circle.

Analysis Table

- Analysis table is prepared by taking the column number on left side and probable values of mode on right side of the table.

- The values of variable corresponding to the highest frequencies are taken at the top of the analysis table and probable values of mode are determined.
- Finally, the maximum marked variable is considered as 'Mode'.

Q19. How mode is calculated for continuous series.

Ans :

The steps involving in calculating mode in continuous series are as follows,

1. In continuous series the class of model class can be determined in two ways,
   (i) By Inspection (or)
   (ii) By preparing grouping and analysis table.
2. Value of mode is ascertained by using the following formula,

$$\boxed{\text{Mode} = L + \frac{\Delta_1}{\Delta_1 + \Delta_2} \times C}$$

Where,

L = Lower limit of model class

f_1 = Frequency of model class

$\Delta_1 = f - f_1$

$\Delta_2 = f - f_2$

C = Class interval.

PROBLEMS

8. **Compute the modal value for the following observations and give your reasons.**
10, 12, 13, 10, 18, 16, 15, 10, 11, 17, 10, 16, 11, 10, 12, 18, 19, 20, 15, 9

Sol :

Arrange the data in ascending order,

Observations	No.of times repeated (or) appeared
9	1
10	5
11	2
12	2
13	1
15	2
16	2
17	1
18	2
19	1
20	1

Among of all the above number 10 is repeated at maximum i.e. 5 times than the other numbers. Hence mode = 10.

9. Find the mode of the following data.

0, 1, 6, 7, 2, 3, 7, 6, 6, 2, 6, 0, 5, 6, 0

Sol :

Observations	No. of times repeated
0	3
1	1
2	2
3	1
5	1
6	5
7	2

Among of all the above number 6 is repeated at maximum times i.e., 5 times.
Hence mode = 6.

10. Data on monthly income of 70 persons are given below. Calculate value of mode.

Income	0 – 10	10 – 20	20 – 30	30 – 40	40 – 50
Persons	8	14	19	17	12

Sol :

Mode can be calculated using the given formula,

$$\text{Mode} = L + \frac{\Delta_1}{\Delta_1 + \Delta_2} \times C$$

Where

$$\Delta_1 = f - f_1$$

$$\Delta_2 = f - f_2$$

From the above distribution, 19 is the highest frequency and hence 20 – 30 is the modal class.

	Income	Persons	
	0 – 10	8	
	10 – 20	14	(f_1)
L	20 – 30	19	(f)
	30 – 40	17	(f_2)
	40 – 50	12	

Here,

$L = 20 \quad f = 19 \quad f_1 = 14 \quad f_2 = 17$

$\Delta_1 = f - f_1$

$19 - 14 = 5$

$\Delta_2 = f - f_2$

$19 - 17 = 2$

$$\text{Mode} = 20 + \frac{5}{5+2} \times 10$$

$$= 20 + \frac{5}{7} \times 10$$

$$= 20 + 7.142 = 27.142$$

∴ The calculated value of mode is ₹ 27.142.

11. **Calculate mode from the following data :**

Marks No. of	**0-10**	**10-20**	**20-30**	**30-40**	**40-50**	**50-60**	**60-70**
Frequency	**5**	**15**	**20**	**20**	**32**	**14**	**14**

Sol :

	Marks	**Frequency**	
	0 - 10	5	
	10 - 20	15	
	20 - 30	20	
	30 - 40	20	f_1
L	40 - 50	32	f
	50 - 60	14	f_2
	60 - 70	14	

$$\text{Mode} = L + \frac{\Delta_1}{\Delta_1 \times \Delta_2} \times C$$

$\Delta_1 = f - f_1$

$\Delta_1 = 32 - 20 = 12$

$\Delta_2 = f - f_2$

$\Delta_2 = 32 - 14 = 18$

$$40 + \frac{12}{12+18} \times 10$$

$$40 + \frac{120}{30}$$

$$40 + 4 = 44$$

2.5 Geometric Mean

Q20. Define Geometric Mean. Explain how Geometric Mean is calculated for individual series.

Ans :

Geometric mean is the n^{th} root of the product of n items of a series. If there are 2 numbers, say a and b, the Geometric mean of the two numbers is the square root of the product of the 2 numbers. Similarly, if there are 3 numbers, Geometric mean of the three numbers would be the cube root of the product of the 3 numbers. Thus, Geometric mean would be $(abc)^{1/3}$. This concept can be applied to as many numbers as possible.

Individual Observations

In case of individual observation series, the procedure to calculate geometric mean is the, same as that of arithmetic mean. The only difference is that in geometric mean, we take the sum total of log values of all the items and divide it by the number of items i.e.

$$\log\ g = \frac{\log(X_1) + \log(X_2) + \ldots + \log(X_n)}{N},$$

$$g = \text{Antilog}\left(\frac{\Sigma \log X}{N}\right)$$

Steps

(i) Take log of all the given items of the data

(ii) Add the log values i.e. find Σ log X

(iii) Divide Σ log X by the number of items in a given problem.

(iv) Head the antilog value from the antilog tables. This will give us the value of g which is the required geometric mean.

Example

Calculate geometric mean from the following data :

x : 10 110 135 120 50 59 60 7

Sol. :

Calculation of Geometric mean

Size of items (X)	Logarithmic values (log X)	
10	log 10	1.0000
110	log 110	2.0414
135	log 135	2.1303
120	log 120	2.0792
50	log 50	1.6990
59	log 59	1.7709
60	log 60	1.7782
7	log 7	0.8451
N = 8		Σ log X = 13.3441

Geometric mean (g)

$$g = \text{Antilog}\left(\frac{\sum \log X}{N}\right) = \text{Antilog}\left(\frac{13.3441}{8}\right)$$

$$= \text{Antilog } 1.6680$$

$$= 46.56$$

Q21. How to calculate the Geometric Mean for discrete series ?

Ans :

Formula for calculation of geometric mean through the following :

$$g = \text{Antilog}\left(\frac{\sum f \log X}{N}\right)$$

Steps

i) Take log values of all the items of a given series.

ii) Multiply each log value to its respective frequencies.

iii) Add the values and divide by the total number of frequencies.

iv) Take the value of antilog from the antilog table and the result would be the geometric mean.

Example

Find the geometric mean from the following data.

X	2	3	5	6	4
Frequencies (f) :	10	15	18	12	7

Sol :

Calculation of Geometric mean

X	f	log X	f log X
2	10	0.3019	3.0100
3	15	0.4771	7.1565
5	18	0.6990	12.5820
6	12	0.7782	9.3384
4	7	0.6021	4.2147
	N = 62		Σ f log X = 36.3016

The formula is for geometric mean is

$$g = \text{Antilog}\left\{\frac{\sum f \log X}{N}\right\}$$

$$g = \text{Antilog}\left\{\frac{36.3016}{62}\right\}$$

$$g = \text{Antilog } [0.5855]$$

Thus the G.M. is $g = 3.850$

Q22. How to calculate the Geometric Mean for continuous series ?

Ans :

The process of computing Geometric mean in case of continuous series involves the following steps

(i) Find the mid value of each class – m.

(ii) Find the logarithm of the mid - value – log m.

(iii) Multiply the logs of m by their respective frequency f log m.

(iv) Add up all the products – $\Sigma f \log m$

(v) Divided f log m by $\Sigma f - \dfrac{\Sigma f \log m}{N}$

(vi) Find out the antilog of the result of step 5 and this will give the answer.

The formula is : G.M. = Anti log = $\dfrac{\Sigma f \log m}{N}$

Example : Find out the geometric mean

Yield of wheat	No. of farms
7.5 – 10.5	5
10.5 – 13.5	9
13.5 – 16.5	19
16.5 – 19.5	23
19.5 – 22.5	7
22.5 – 25.5	4
25.5 – 28.5	1

Sol :

Calculation of Geometric Mean

C.I	Mid Value (m)	log m	f	f log m
7.5 – 10.5	9	0.9542	5	4.7710
10.5 – 13.5	12	1.0792	9	9.7128
13.5 – 16.5	15	1.1761	19	22.3459
16.5 – 19.5	18	12.2553	23	28.8719
19.5 – 22.5	21	1.3222	7	9.2554
22.5 – 25.5	24	1.3802	4	5.5208
25.5 – 28.5	27	1.4314	1	f 1.4314
			N = 68	Σ logm = 81.9092

$$\text{G.M.} = \text{Antilog}\ \frac{\Sigma f \log m}{N}$$

$$= \text{Antilog}\ \frac{81.9092}{68}$$

$$= \text{Anti log } 1.2045 = 16.02.$$

Q23. State the merits and demerits of geometric mean.

Ans :

Merits

1. It is rigidly defined. Hence, different interpretations by different persons are not possible.
2. It takes all values into consideration. Thus, it is more representative.
3. It can be subjected to further mathematical treatment. The properties of Geometric mean have been separatelu explained.
4. It has a bias towards lower values.
5. It is not affected much by presence of extremely small or extremely large observations.
6. It is not much affected by the fluctuations of sampling.

Demerits

1. It is neither simple to understand nor easy to calculate.
2. It cannot be determined by inspection.
3. It cannot be located graphically.
4. It cannot be used in the study of qualitative phenomena.
5. It may be a fictitious value i.e. One that does not exist in the series.
6. It cannot be computed if any value in a series is zero or negative.
7. It brings out the property of the ratio of change and not of absolute difference as in the case of arithmetic mean.
8. It cannot be calculated even if a single observation is missing or lost.

Q24. Explain briefly about Weighted Geometric Mean.

Ans :

In weighted geometric mean weights are given along with variable 'x'.

Steps :

The following steps are followed to calculate weighted geometric mean,

1. Take the logarithms of variable x.
2. Multiply values of log x with their respective weights.
3. Obtain the total of Wlogx and represent this total as Σ(Wlogx).
4. Divide Σ(Wlogx) by the total of weights, [i.e., ΣW]
5. Find out the Antilog for the value obtained in step 4.

$$\boxed{G.M_w = \text{Anti}\log\left[\frac{\Sigma(W \text{ Log } x}{\Sigma W}\right]}$$

PROBLEMS

12. Calculate geometric mean of the following :

50 72 54 82 93

Sol :

Calculation of Geometric Mean

X	log of X
50	1.6990
72	1.8573
54	1.7324
82	1.9138
93	1.9685
	Σ Log X = 9.1710

$$\text{G.M.} = \text{Antilog } \frac{\Sigma \text{ Log X}}{\text{N}}$$

$$= \text{Antilog } \frac{9.1710}{5}$$

$$= \text{Antilog } (1.8342) = 68.26.$$

13. Calculate geometric mean for the following data :

1111, 111, 11, 1.1, 0.11, 0.011

Sol :

Calculation of Geometric Mean

X	Log x
11111	3.0457
111	2.0458
11	1.0414
1.1	0.0414
0.11	–0.9586
0.011	–1.9586
	Σ logx = 3.2157

We have M = 6, Σ logx = 3.2157

$$\therefore \quad GM = \text{Antilog}\left[\frac{\Sigma \log x}{N}\right]$$

$$= \text{Antilog}\left[\frac{3.2157}{6}\right]$$

$$= \text{Antilog}\ [0.5360]$$

$$GM = 3.4356$$

14. The following table gives the weight of 31 persons in a sample survey. Calculate geometric mean

Weight (lbs)	130	135	140	145	146	148	149	150	157
No. of persons	3	4	6	6	3	5	2	1	1

Sol :

Size of item	Frequency f	log X	f log X
130	3	2.1139	6.3417
135	4	2.1303	8.5212
140	6	2.1461	12.8766
145	6	2.1614	12.9684
146	3	2.1644	6.4932
148	5	2.1703	10.8515
149	2	2.1732	4.3464
150	1	2.1761	2.1761
157	1	2.1959	2.1959
	N = 31		$\Sigma f \log X = 66.7710$

$$\text{G.M.} = \text{Antilog}\ \frac{\Sigma f \log X}{N}$$

$$= \text{Antilog}\ \frac{66.7710}{31} = \text{Antilog of } (2.1539)$$

$$\text{G.M.} = 142.5$$

15. Find the geometric mean from the following data :

X	10	20	30	40	50	60
f	12	15	25	10	6	2

Sol :

Calculation of Geometric Mean

X	f	log x	f log x
10	12	1.0000	12.0000
20	15	1.3010	19.5150
30	25	1.4771	36.9295
40	10	1.6021	16.0210
50	6	1.6990	10.1940
60	2	1.7782	3.5564
	N = 70		**98.2139**

$$\therefore \quad \text{Geometric Mean (GM)} = \text{Antilog}\left[\frac{\Sigma f \log x}{N}\right]$$

$$= \text{Antilog}\left[\frac{98.2139}{70}\right]$$

$$= \text{Antilog}\,(1.4031)$$

$$\text{G.M} = 25.2988$$

16. Calculate geometric mean for the following data,

C.I.	0-8	8-16	16-24	24-32	32-40	40-48	48-56	56-64
Frequency	5	8	12	15	20	18	13	9

Sol :

Calculation of Geometric Mean

C.I	Mid-points (x)	Frequency (f)	Log x	f Log x
0 - 8	4	5	0.6021	3.0105
8 - 16	12	8	1.0792	8.6336
16 - 24	20	12	1.3010	15.6120
24 - 32	28	15	1.4472	21.708
32 - 40	36	20	1.5563	31.126
40 - 48	44	18	1.6435	29.583
48 - 56	52	13	1.7160	22.308
56 - 64	60	9	1.7782	16.0038
		N = 100		**Σf. logx = 147.9849**

$$\text{Geometric Mean (GM)} = \text{Antilog}\left[\frac{\Sigma f \log x}{N}\right]$$

$$G.M = \text{Antilog}\left[\frac{147.9849}{100}\right]$$

$$G.M = \text{Antilog}\ (1.479849)$$

$$\boxed{\because \text{Antilog} = 10^x \text{ i.e., } 10^{1.479849}}$$

$$\therefore \quad G.M = 30.1890$$

2.6 Harmonic Mean

Q25. What is Harmonic Mean? Explain how Harmonic Mean is calculated for individual series.

Ans :

Meaning

"The reciprocal of the arithmetic mean of the reciprocal of individual observations" is known as Harmonic mean. It is represented as follows,

$$H.M = \frac{N}{\left(\frac{1}{x_1} + \frac{1}{x_2} + \frac{1}{x_3} + \ldots + \frac{1}{x_n}\right)}$$

When the number of observations are large then Harmonic mean is stated as follows,

$$H.M = \frac{N}{\Sigma(1/x)}$$

The steps involved in calculating harmonic mean for individual observation are,

1. Take the reciprocal of each value of variable x $\left[\text{i.e., } \frac{1}{x}\right]$

2. Find out the sum of $\frac{1}{x}$ $\left(\text{i.e., } \Sigma\left[\frac{1}{x}\right]\right)$

3. Divide the number of observations [N] by $\Sigma\frac{1}{x}$

$$H.M = \frac{N}{\Sigma\left[\frac{1}{x}\right]}$$

Q26. Explain how Harmonic Mean is calculated for discrete series.

Ans :

The steps involved in calculating harmonic mean for discrete series are as follows,

1. Find the reciprocal of each value of variable x.

2. Multiply reciprocal value of x with their corresponding frequencies and obtain the sum as $\Sigma\left[f \times \frac{1}{x}\right]$.

3. Divide the sum of frequency [N] by $\Sigma\left[f \times \frac{1}{x}\right]$.

The formula to calculate Harmonic mean for discrete series is as follows,

$$H.M = \frac{N}{\Sigma\left[f \times \frac{1}{x}\right]} = \frac{N}{\left(\frac{f}{x}\right)}$$

Q27. Explain how Harmonic Mean is calculated for continuous series.

Ans :

The steps involved in calculating harmonic mean for continuous series are as follows,

1. Take the mid-points of various class intervals and denote them as m.
2. Take the reciprocal of each value of the variable "x'.
3. Multiply reciprocal of x $\left(\text{i.e., } \frac{1}{m}\right)$ with their corresponding frequencies and obtain the sum $\Sigma\left[f \times \frac{1}{m}\right]$.
4. Divide the sum of frequency [N] by $\Sigma\left[f \times \frac{1}{m}\right]$.

The formula to calculate Harmonic mean for continuous series is as follows,

$$H.M = \frac{N}{\Sigma\left[f \times \frac{1}{x}\right]} = \frac{N}{\Sigma\left[\frac{f}{x}\right]}$$

Q28. What is Weighted Harmonic Mean ?

Ans :

When information about speed as well as distance is given in the problem then weighted Harmonic mean is calculated.

Weighted Harmonic mean is calculated by using the following formula,

$$H.M_w = \frac{\Sigma W}{\left[\frac{1}{a} \times W_1\right] + \left[\frac{1}{b} \times W_2\right] + \left[\frac{1}{c} \times W_3\right]}$$

(or)

$$H.M_w = \frac{\Sigma W}{\Sigma\left(\frac{W}{X}\right)}$$

PROBLEMS

17. **Compute Harmonic mean of the following :**

20, 40, 50, 80, 100

Sol :

X		1/x
20	1/20	0.05
40	1/40	0.025
50	1/50	0.02
80	1/80	0.0125
100	1/100	0.0100
		Σ 1/x 0.1175

$$\text{Harmonic Mean (H.M)} = \frac{N}{\Sigma\left(\frac{1}{x}\right)}$$

$$\text{Harmonic Mean (H.M)} = \frac{5}{0.1175}$$

Harmonic Mean (H.M) = 42.55.

18. **Calculate Harmonic Mean from the following data :**

70, 90, 178, 152 and 174

Sol :

Calculation of Harmonic Mean

X	1/X
70	0.014
90	0.011
178	0.005
152	0.006
174	0.005
	Σ(1/X) = 0.041

$$\text{Harmonic Mean (H.M)} = \frac{N}{\Sigma\left(\frac{1}{X}\right)}$$

$$= \frac{5}{0.041}$$

$= 121.9 = 122$

19. An aeroplane flies along the four sides of a square at varying speeds of 200, 400, 600 and 800 miles per hour respectively. What is the average speed of the plane ?

Sol :

Thus, apply H.M.

$$H.M = \frac{N}{\Sigma\left(\frac{1}{X}\right)} \Rightarrow \frac{4}{\frac{1}{200}+\frac{1}{400}+\frac{1}{600}+\frac{1}{800}}$$

$$= \frac{4}{\frac{12+6+4+3}{2400}} \Rightarrow \frac{4\times 2400}{25} \Rightarrow \frac{9600}{25}$$

$$= 384 \text{ miles per hour.}$$

20. From the following table compute the Harmonic Mean,

Marks	0-10	10-20	20-30	30-40	40-50
No. of students	7	12	2	14	15

Sol :

Marks	Mid - Points (m)	No. of Students (f)	f/m
0 – 10	5	7	1.4
10 – 20	15	12	0.8
20 – 30	25	2	0.08
30 – 40	35	14	0.40
40 – 50	45	15	0.3333
		N=50	**3.0133**

$$\therefore H.M = \frac{N}{\Sigma\left(\frac{f}{m}\right)} = \frac{50}{3.0133}$$

$$\therefore H.M = 16.5931.$$

21. Calculate the Harmonic mean from the following data

x	10	12	14	16	18	20
f	5	18	20	10	6	1

Sol :

Calculation of Harmonic Mean

x	f	f/x
10	5	0.500
12	18	1.500
14	20	1.428
16	10	0.625
18	6	0.333
20	1	0.050
	N = 60	**4.436**

$$H.M = \frac{N}{\Sigma\left[\frac{f}{x}\right]}$$

$$= \frac{60}{4.436}$$

$$= 13.526$$

22. Calculate the Harmonic Mean from the following data.

X	10	20	40	60	20
f	1	3	6	5	4

Sol :

x	f	f/x
10	1	0.1000
20	3	0.15
40	6	0.15
60	5	0.083
20	4	0.2
	N = 19	**0.683**

$$H.M = \frac{N}{\Sigma\left[\frac{f}{x}\right]}$$

$$= \frac{19}{0.683} = 27.8$$

2.7 Measures of Dispersion

Q29. Define Dispersion. What are the objectives of Measuring Dispersion.

Ans :

Meaning

The concept of dispersion is related to the extent of scatter or variability in observations. The variability, in an observation, is often measured as its deviation from a central value. A suitable average of all such deviations is called the measure of dispersion. Since most of the measures of dispersion are based on the average of deviations of observations from an average, they are also known as the averages of second order.

Definitions

As opposed to this, the measures of central tendency are known as the averages of first order. Some important definitions of dispersion are given below:

(i) **According to A.L. Bowley,** "Dispersion is the measure of variation of the items."

(ii) **According to Connor,** "Dispersion is the measure of extent to which individual items vary."

(iii) **According to Simpson and Kafka,** "The measure of the scatteredness of the mass of figures in a series about an average is called the measure of variation or dispersion."

(iv) **According to Spiegel,** "The degree to which numerical data tend to spread about an average value is called variation or dispersion of the data."

Objectives

The main objectives of measuring dispersion of a distribution are:

1. **To test reliability of an average:** A measure of dispersion can be used to test the reliability of an average. A low value of dispersion implies that there is greater degree of homogeneity among various items and, consequently, their average can be taken as more reliable or representative of the distribution.

2. **To compare the extent of variability in two or more distributions:** The extent of variability in two or more distributions can be compared by computing their respective dispersions. A distribution having lower value of dispersion is said to be more uniform or consistent.

3. **To facilitate the computations of other statistical measures:** Measures of dispersions are used in computations of various important statistical measures like correlation, regression, test statistics, confidence intervals, control limits, etc.

4. **To serve as the basis for control of variations:** The main objective of computing a measure of dispersion is to know whether the given observations are uniform or not.

Q30. Explain the types of Measures of Dispersion.

Ans :

There are different measures of dispersion. These measures can be classified into:

A) Absolute measures and

B) Relative measures.

i) An 'Absolute' measure is one that is expressed in terms of the same unit in which the variable (or given data) is measured.

ii) A 'Relative" measure of dispersion is expressed as a pure number (without any units) which enables comparison of the levels of dispersion from a central tendency across different series (stated in different units). These measures are also called as "Co-efficient(s) of Dispersion". The important methods of studying variation are listed as under:

A) Absolute measures

(i) The Range

(ii) Inter Quartile Range and Quartile Deviation.

(iii) The Mean Deviation (or) Average Deviation.

(iv) The Standard Deviation and Variance

(v) The Lorenz Curve

B) **Relative measures**

(i) Coefficient of Range

(ii) Coefficient of Quartile Deviation.

(iii) Coefficient of Mean Deviation.

(iv) Coefficient of Variation

2.8 Range

Q31. What is Range ? Explain.

Ans :

Range is the simplest method of studying dispersion. It is defined as the difference between the value of the smallest item and the value of the largest item included in the distribution. Symbolically,

$$\boxed{\text{Range} = L - S}$$

where

L = Largest item, and

S = Smallest item.

The relative measure corresponding to range, called the coefficient of range, is obtained by applying the following formula :

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

If the averages of the two distributions are about the same, a comparison of the range indicates that the distribution with the smaller range has less dispersion, and the average of that distribution is more typical of the group.

Q32. What are the uses of range ?

Ans :

(i) **Quality Control**

The object of quality control is to keep a check on the quality of the product without 100% inspection. When statistical methods of quality control are used, control charts are prepared and in preparing these charts range plays a very important role. The idea basically is that if the range - the difference between the largest and smallest mass produced items-increases beyond a certain point, the production machinery should be examined to find out why the items produced have not followed their usual more consistent pattern

(ii) **Fluctuations in the Share Prices**

Range is useful in studying the variations in the prices of stocks and shares and other commodities that are sensitive to price changes from one period to another.

(iii) **Weather Forecasts**

The meteorological department does make use of the range in determining, say, the difference between the minimum temperature and the maximum temperature. This information is of great concern to the general public because they know as to within what limits the temperature is likely to vary on a particular day.

(iv) Everyday Life

The range is a most commonly used measure of dispersion in everyday life. Questions of the form "what is the minimum and maximum temperature on a particular day"? "What is the difference between the wages earned by workers of a particular factory" ? How much one spends on petrol in his car/scooter in a month"- are all usually answered in the form of range. Answers to questions such as these are usually given in the form of 'Between such and such]. Regardless of the crudity of expression the answer is still a range.

PROBLEMS

23. The following are the wages of 10 workers of a factory. Find the range of variation and also compute the coefficient of range.

275, 200, 370, 240, 100, 290, 400, 500, 180, 350

Sol :

Step I

Arrange the data in ascending order i.e.

100, 180, 200, 240, 275, 290, 350, 370, 400, 500

Range = L – S

where as = L = 500, S = 100

Range (R) = 500 – 100 = 400

$$\text{Coefficient of range} = \frac{L-S}{L+S} = \frac{500-100}{500+100}$$

$$\text{Coefficient of range} = \frac{400}{600} = 0.667.$$

24. Find the range and its coefficient from the following data.

Marks	0-10	10-20	20-30	30-40	40-50	50-60	60-70	70-80
No. of students	4	12	20	18	15	8	2	1

Sol :

Calculation of Range and Co-efficient of Range

Marks	No. of Students
0 – 10	4
10 – 20	12
20 – 30	20
30 – 40	18
40 – 50	15
50 – 60	8
60 – 70	2
70 – 80	1
	N = 80

Range-when Actual Values are Taken

R = L − S

Where,

R = Range

L = Largest Value = 80

S = Smallest Value = 10

$\therefore$ R = 80 − 10 = 70

Co-efficient of Range

$$CR = \frac{L-S}{L+S}$$

Where,

CR = Co-efficient of range

L = Largest value = 80

S = Smallest value = 10

$$\therefore CR = \frac{80-10}{80+10}$$

$$= \frac{70}{90} = 0.778$$

2.9 Quartile Deviation

Q33. What is Quartile Deviation ?

Ans :

It is based on two extreme items and it fails to take account of the scatter within the range. From this there is reason to believe that if the dispersion of the extreme items is discarded, the limited range thus established might be more instructive.

For this purpose there has been developed a measure called the interquartile range, the range which includes the middle 50 per cent of the distribution. That is, one quarter of the observations at the lower end, another quarter of the observations at the upper end of the distribution are excluded in computing the interquartile range. In other words, interquartile range represents the difference between the third quartile and the first quartile.

$$\boxed{\text{Inter quartile range} = Q_3 - Q_1}$$

Very often the interquartile range is reduced to the form of the Semi-interquartile range or quartile deviation by dividing it by 2.

$$\boxed{\text{Quartile Deviation or Q.D.} = \frac{Q_3 - Q_1}{2}}$$

Coefficients of Quartile Deviation

Quartile deviation is an absolute measure of dispersion. Its relative measure is the coefficient of Quartile deviation.

Coefficients of Quartile Deviation

$$= \frac{\frac{Q_3 - Q_1}{2}}{\frac{Q_3 + Q_1}{2}}$$

$$= \frac{Q_3 - Q_1}{Q_3 - Q_1}$$

It is used to compare the degree of variation in different series.

Q34. Explain merits and demerits of quartile deviations.

Ans :

Merits of Quartile Deviation

1. It is the simple to calculate and very easy to understand.
2. It is not impacted by extreme values
3. It can be computed for open ended distributions and for data containing unequal classes

Demerits of Quartile Deviation

1. It is impacted by sample size and composition of the sample.
2. It is not amenable to algebraic or statistical treatment
3. It does not tell us anything about the spread of the various data items across the measure of central tendency.

Q35. How to calculate quartile deviations?

Ans :

Individual Series

Low Quartile (Q_1),

$$= \text{Size of } \frac{N+1}{4} \text{th item}$$

N= No. of observations

Upper Quartile (Q_3),

$$= \text{Size of } \frac{3(N+1)}{4} \text{th item.}$$

Discrete Series

i) $Q_1 = \text{Size of } \frac{N+1}{4} \text{th item}$

ii) $Q_3 = \text{Size of } \frac{3(N+1)}{4} \text{th item}$

iii) Quartile Deviation (Q.D) $= \frac{Q_3 - Q_1}{2}$

Continuous Series

$$Q_1 = L + \frac{\frac{N}{4} - F}{f} \times C$$

$$Q_3 = L + \frac{\frac{3N}{4} - F}{f} \times C$$

PROBLEMS

25. Find out the value of quartile deviation and its coefficient for the following data :

Roll No	1	2	3	4	5	6	7
Marks	25	33	45	17	35	20	55

Sol :

Calculation of Quartile Deviation

Marks arranged in ascending order 17 20 25 33 35 45 55

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

N = No. of observations.

$$= \text{Size of } \left[\frac{7+1}{4}\right]^{th} \text{Item} = 2^{nd} \text{ Item} = 20$$

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

$$= \text{Size of } 3\left[\frac{7+1}{4}\right]^{th} \text{ Item} = 6^{th} \text{ Item} = 45$$

$$\text{Quartile Deviation (Q.D)} = \frac{Q_3 - Q_1}{2} = \frac{45-20}{2} = \frac{25}{2} = 12.5$$

$$\text{Coefficient of Q.D} = \frac{Q_3 - Q_1}{Q_3 + Q_1} = \frac{45-20}{45+20} = \frac{25}{65} = 0.38.$$

26. **Find out the value of quartile deviation from the following data**

Roll No.	1	2	3	4	5	6	7
Marks :	30	42	60	18	45	24	75

Sol :

Rearrange into ascending order

x = 18, 24, 30, 42, 45, 60, 75

$$\text{Quartile deviation } Q_1 = \text{size of } \left(\frac{N+1}{4}\right)^{th} \text{ item}$$

$$\text{Size of } \left(\frac{7+1}{4}\right)^{th} \text{ item} = 2^{nd} \text{ item}$$

The size of 2^{nd} item is 24 i.e. $Q_1 = 24$

$$Q_3 = \text{size of } 3\left(\frac{7+1}{4}\right)^{th} \text{ item}$$

$$= 6^{th} \text{ item} = 60, \text{ then } Q_3 = 60$$

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

$$= \frac{60-24}{2} = \frac{36}{2} = 18$$

Q.D = 18

27. **Calculate the Quartile Deviation and Coefficient of Quartile Deviation of the following distribution.**

Marks (X)	0-10	10-20	20-30	30-40	40-50	50-60	60-70	70-80	80-90	90-100
No.of Students (f)	11	18	25	28	30	33	22	15	12	10

Sol :

Calculation of Quartile Deviation

X	f	C.f
0-10	11	11
10-20	18	29 F
L 20-30	25 f	54 (Q_1)
30-40	28	82
40-50	30	112
50-60	33	145 F
L 60-70	22 f	167 (Q_3)
70-80	15	182
80-90	12	194
90-100	10	204

Calculation of Q_1

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

$$= \left(\frac{204}{4}\right)^{th} \text{Item}$$

$$= 51^{th} \text{ Item}$$

Size of 51^{th} Item lies in CF 54. So Q_1 class interval = 20–30

$$Q_1 = L + \frac{\frac{N}{4} - F}{f} \times C$$

$$= 20 + \frac{51 - 29}{25} \times 10$$

$$= 20 + \frac{22 \times 10}{25}$$

$$= 20 + \frac{220}{25}$$

$$= 20 + 8.8 = 28.8.$$

Calculation of Q_3

$$= Q_3 = 3\left(\frac{N}{4}\right)^{th}$$

$$= 3\left(\frac{204}{4}\right)^{th} \text{ Item} = 3(51)^{th} \text{ Item} = 153^{th} \text{ Item}$$

Size of 153^{th} Item lies in CF 167. So Q_3 class interval = 60 – 70

$$Q_3 = L_3 + \frac{\frac{3N}{4} - F}{f} \times C$$

$$= 60 + \frac{153 - 145}{22} \times 10$$

$$= 60 + \frac{80}{22}$$

$$= 60 + 3.64$$

$$= 63.64$$

Calculation of Quartile Deviation

$$\text{Quartile Deviation (Q.D)} = \frac{Q_3 - Q_1}{2} = \frac{63.64 - 28.8}{2} = 17.4$$

$$\text{Coefficient of Q.D} = \frac{Q_3 - Q_1}{Q_3 + Q_1} = \frac{63.64 - 28.8}{63.64 + 28.8} = \frac{34.84}{92.44} = 0.37$$

2.10 Mean Deviation

Q36. Define mean deviation. State the merits and demerits of mean deviation.

Ans :

Meaning

"Mean Deviation of a series is the arithmetic average of the deviations of various items from a measure of central tendency (either mean, median or mode)". Theoretically, deviations can be taken from any of the three averages mentioned above, but in actual practice it is calculated either from mean or from Median. While Calculating deviations algebraic signs are not taken into account.

Merits

(i) It is rigidly defined

(ii) It is not least impacted by sampling fluctuations

(iii) It takes into account every single value in the series

(iv) Mean deviation from Median is least impacted due to extreme values

(v) It is extensively used in multiple fields such as Economics, Commerce, etc as it is the best measure for comparison of two or more series.

Demerits

(i) It is relatively difficult to compute

(ii) It is not amenable to further algebraic or statistical treatment.

(iii) It is difficult for a layman to understand as to why or when a particular average should be considered for calculation of Mean Deviation. The Mean Deviations obtained by taking the Mean, Median and Mode as average differ widely

(iv) It is not effective for open ended series, particularly when the average is Arithmetic Mean.'

Q37. How to calculate mean deviation for individual series ?

Ans :

Individual Series

If X_1, X_2, X_3, X_N are N given observations then the deviation about an average. A is given by

$$\frac{\Sigma|D|}{N}$$

Steps

i) Compute the median of the series.

ii) The deviations of items from median ignoring ± signs and denote these deviations by $|D|$.

iii) Obtain the total of these deviations, i.e., $\Sigma|D|$.

iv) Divide the total obtained in step

v) By the number of observations.

The relative measure corresponding to the mean deviation, called the coefficient of mean deviation, is obtained by dividing mean deviation by the particular average used in computing mean deviation. Thus, if mean deviation has been computed from median, the coefficient of mean deviation shall be obtained by dividing mean deviation by median.

$$\text{Coefficient of M.D.} = \frac{\text{M.D.}}{\text{Median}}$$

If mean has been used while calculating the value of mean deviation, in such a case coefficient of mean deviation shall be obtained by dividing mean deviation by the mean.

Q38. How to calculate mean deviation for discrete series ?

Ans :

In discrete series the formula for calculating mean deviation is

$$\text{M.D.} = \frac{\Sigma f|D|}{N}$$

$|D|$ denotes deviation from median ignoring signs.

N = Sum of frequency

Steps

i) Calculate the median of the series.

ii) Take the deviations of the items from median ignoring signs and denote them by $|D|$.

iii) Multiply these deviations by the respective frequencies and obtain the total $\Sigma f|D|$

iv) Divide the total obtained in Step (ii) by the number of observations. This gives us the value of mean deviation.

Q39. How to calculate mean deviation for continuous series ?

Ans :

For calculating mean deviation in continuous series the procedure remains the same as discussed above. The only difference is that here we have to obtain the mid-point of the various classes and take deviations of these points from median. The formula is same, i.e..

$$\text{M.D.} = \frac{\Sigma f|D|}{N}$$

Steps

i) Calculate the median of the series.

ii) Take the deviations of the items from median ignoring signs and denote them by $|D|$.

iii) Multiply these deviations by the respective frequencies and obtain the total $Sf|D|$.

iv) Divide the total obtained in step (ii) by the number of observations. This gives us the value of mean deviations.

PROBLEMS

28. **Calculate mean deviation of the follo-wing from mean and median**

2, 6, 11, 14, 16, 19, 23

Sol. :

X	$X-\overline{X}=\lvert D\rvert$	$X-\text{Median}=\lvert D\rvert$
2	11	12
6	7	8
11	2	3
14	1	0
16	3	2
19	6	5
23	10	9
91	**40**	**39**

(a) Mean $(\overline{X}) = \dfrac{\Sigma x}{N} = \dfrac{91}{7} = 13$

MD through mean $= \dfrac{\Sigma\lvert D\rvert}{N} = \dfrac{40}{7} = 5.71$

(b) Median $= \text{size of} \left(\dfrac{N+1}{N}\right)^{th}$ item

$= \text{size of} \left(\dfrac{7+1}{2}\right)$

$= 8/2 = 4^{th}$ observation $= 14$

(c) Median = size of 4th Item = 14

Mean Deviation (MD) through

Median $= \dfrac{\Sigma\lvert D\rvert}{M}$

Mean Deviation (MD) $= \dfrac{39}{7}$

Mean Deviation (MD) = 5.57

29. **From the following data calculate mean deviation from the median.**

C.I.	16-20	21-25	26-30	31-35	36-40	41-45	46-50	51-55	56-60
Frequency	8	15	13	20	11	7	3	2	1

Sol : (Imp.)

Calculation of Mean Deviation from Median

C.I	Frequency (F)	C.F	Mid-Value (m)	$\|M-A\| = \|D\|$ $\|M - 31.5\|$	$F\|M-A\| = f\|D\|$
16 - 20	8	8	18	13.5	108.0
21 - 25	15	23	23	8.5	127.5
26 - 30	13	36 F	28	3.5	45.5
L 31 - 35	20 f	56	33	1.5	30.0
36 - 40	11	67	38	6.5	71.5
41 - 45	7	74	43	11.5	80.5
46 - 50	3	77	48	16.5	49.5
51 - 55	2	79	53	21.5	43.0
56 - 60	1	80	58	26.5	26.5
	N = 80				$\Sigma f\|D\| = 582$

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

$$= \frac{80}{2} = 40^{th} \text{ item}$$

C.F is just greater than 40 is 56

Median lies in the class 31-35. However, the real limit of this class under exclusive method is 30.5 – 35.5.

$$\text{Median} = L + \frac{\frac{N}{2} - F}{f} \times C$$

$$L = 30.5,\ F = 36,\ f = 20 \ \ C = 5$$

$$= 30.5 + \frac{40-36}{20} \times 5$$

$$= 30.5 + \frac{4}{20} \times 5$$

$$= 30.5 + [0.2] \times 5$$

$$= 30.5 + 1 = 31.5$$

$$\text{M.D from Median} = \frac{\Sigma f|D|}{N} = \frac{582}{80} = 7.275$$

2.11 Standard Deviation

Q40. What is standard deviation ? Explain its merits and demerits.

Ans : **(Imp.)**

The standard deviation concept was introduced by Karl Pearson in 1823. It is by far the most important and widely used measure of studying dispersion. Its significance lies in the fact that it is free from those defects from which the earlier methods suffer and satisfies most of the properties of a good measure of dispersion. Standard deviation is also known as root mean square deviation for the reason that it is the square root of the mean of the squared deviation from the arithmetic mean. Standard deviation is denoted by the small Greek letter σ (read as sigma).

The standard deviation measures the absolute dispersion (or variability of distribution; the greater the amount of dispersion or variability), the greater the standard deviation, the greater will be the magnitude of the deviations of the values from their mean. A small standard deviation means a high degree of uniformity of the observation as well as homogeneity of a series; a large standard deviation means just the opposite.

Merits

1. It is rigidly defined
2. It takes into account every single value in the series
3. It is amenable to further algebraic or statistical treatment.
4. It is extensively used in various other statistical calculations such as correlation, regression, sampling etc

Demerits

1. It is relatively difficult to compute
2. It is calculated with only Arithmetic Mean as the average. Standard deviation from other averages such as Median is not an effective measure of dispersion.

Q41. How to compute the standard deviations for individual series ?

Ans :

Individual Series

In case of individual observations standard deviations may be calculate by applying any of the following two methods, i.e.

i) By taking deviations of the items from the actual mean.

ii) By taking deviations of the items from an assumed mean.

Deviation taken from actual mean : when deviations taken from actual mean the following formula is applied.

$$\sigma = \sqrt{\frac{\Sigma x^2}{N}}$$

where as

$X = (X - \bar{X})$

N = Total number of observations.

Steps

i) Calculate the actual for the series i.e. $\bar{X}$

ii) Find X deviation from $\bar{X}$ i.e. $x = (X - \bar{X})$

iii) Square these deviations and obtain the total Σx^2

iv) Divide ΣX^2 by the total number of observations and extract the square root.

Deviations from Assumed Mean

$$\sigma = \sqrt{\frac{\Sigma d^2}{N} - \left(\frac{\Sigma d}{N}\right)^2}$$

Steps

➢ Take the deviations of the items from an assumed mean, i.e., obtain (X – A). Denote these deviations by d. Take the total of these deviations. i.e., obtain Σd.

➢ Square these deviations and obtain the total Σd^2.

➢ Substitute the value of Σd^2, Σd and N in the above formula.

Q42. How to compute the standard deviations for discrete series ?

Ans :

For calculating standard deviation in discrete series, any of the following methods may be applied:

(a) Actual mean method.

(b) Assumed mean method.

(c) Step deviation method.

(a) Actual Mean Method. When this method is applied, deviations are taken from the actual mean, i.e., we find $(X - \bar{X})$ and denote these deviations by x. These deviations are then squared and multiplied by the respective frequencies. The following formula is applied :

$$\sigma = \sqrt{\frac{\Sigma fx^2}{N}}, \text{ where } x = (X - \bar{X})$$

N = Sum of observation/Frequency.

However, in practice this method is rarely used because if the actual mean is in fractions the calculations take a lot of time.

(b) Assumed mean method

When this method is used, the following formula is applied :

$$\sigma = \sqrt{\frac{\Sigma fd^2}{N} - \left(\frac{\Sigma fd}{N}\right)^2}$$

where $d = (X - A)$

N= Sum of frequency

Steps

i) Take the deviations of the items from an assumed mean and denote these deviations by d.

ii) Multiply these deviations by the respective frequencies and obtain the total, Sfd.

iii) Obtain the squares of the deviations, i.e., calculate d^2.

iv) Multiply the squared deviations by the respective frequencies, and obtain the total, Sfd^2.

v) Substitute the values in the above formula.

(c) Step Deviation Method: When this method is used we take deviations of midpoints from an assumed mean and divide these deviations by the width of class interval, i.e., 'C'. In case class intervals are unequal, we divide the deviations of midpoints by the lowest common factor and use 'C' in the formula for calculating standard deviation.

The formula for calculating standard deviation is :

$$\sigma = \sqrt{\frac{\Sigma fd^2}{N} - \left(\frac{\Sigma fd}{N}\right)^2} \times C$$

Where,

$$d = \frac{(X - A)}{C} \text{ and } C = \text{class interval.}$$

The use of the above formula simplifies calculations.

Q43. How to compute the standard deviations for continuous series ?

Ans :

In continuous series any of the methods discussed above for discrete frequency distribution can be used. However, in practice it is the step deviation method that is most used. The formula is

$$= \sqrt{\frac{\Sigma fd^2}{N} - \left(\frac{\Sigma fd}{N}\right)^2} \times C$$

$$\text{where } d = \frac{(m - A)}{C}, \quad C = \text{class interval}$$

Steps

i) Find the mid-points of various classes.

ii) Take the deviations of these mid-points from an assumed mean and denote these deviations by d.

iii) Wherever possible take a common factor and denote this column by d.

iv) Multiply the frequencies of each class with these deviations and obtain Σfd.

v) Square the deviations and multiply them with the respective frequencies of each class and obtain Σfd^2.

30. Calculate standard deviation for the following data:

Sl. No.	Weekend Income (₹)
1	270
2	350
3	258
4	282
5	218
6	202
7	364
8	184

Sol :

Calculation of Standard Deviation

Sl.No.	X	$x = [X - \bar{X}]$	x^2
1	270	4	16
2	350	84	7056
3	258	– 8	64
4	282	16	256
5	218	– 48	2304
6	202	– 64	4096
7	364	98	9604
8	184	– 82	6724
	$\Sigma X = 2128$		$\Sigma X^2 = 30120$

(i) Arithmetic Mean $(\bar{X}) = \frac{\Sigma X}{N} = \frac{2128}{8}$

$= 266$

(ii) Standard Deviation $(\sigma) = \sqrt{\frac{\Sigma x^2}{N}}$

$= \sqrt{\frac{30120}{8}}$

$= \sqrt{3765}$

$= 61.35$

31. From the following information calculate standard deviation.

X	4.5	14.5	24.5	34.A	44.5	54.5	64.5
F	1	5	12	22	17	9	4

Sol :

Calculation of Standard Deviation

X	f	X – A = d	d^2	fd	fd^2
4.5	1	–30	900	–30	900
14.5	5	–20	400	–100	2000
24.5	12	–10	100	–120	1200
34.5A	22	0	0	0	0
44.5	17	10	100	170	1700
54.5	9	20	400	180	3600
64.5	4	30	900	120	3600
	70			220	13,000

$$\sigma = \sqrt{\frac{\Sigma fd^2}{N} - \left(\frac{\Sigma fd}{N}\right)^2}$$

$$\sigma = \sqrt{\frac{13000}{70} - \left(\frac{220}{70}\right)^2}$$

$$\sigma = \sqrt{185.71 - (3.14)^2}$$

$$\sigma = \sqrt{185.71 - 9.8596}$$

$$\sigma = \sqrt{175.85}$$

$$\sigma = 13.26$$

32. **Find out the mean and standard deviation of the following data :**

Age under	10	20	30	40	50	60	70	80
No.of persons dying	15	30	53	75	100	110	115	125

Sol :

Calculation of Standard Deviation

Age	Cf	f	Midvalue (m)	m–A=d	d	d^2	fd^2	fd
0 - 10	15	15	5	–30	–3	9	135	–45
10 - 20	30	15	15	–20	–2	4	60	–30
20 - 30	53	23	25	–10	–1	1	23	–23
30 - 40	75	22	35 A	0	0	0	0	0
40 - 50	100	25	45	10	1	1	25	25
50 - 60	110	10	55	20	2	4	40	20
60 - 70	115	5	65	30	3	9	45	15
70 - 80	125	10	75	40	4	16	160	40
		125					488	-2

i) Mean $\bar{X} = A + \left(\frac{\sum fd'}{N}\right) \times C$

Mean $\bar{X} = 35 + \left(\frac{25}{125}\right) \times 10$

Mean $\bar{X} = 35 + (0.016) \times 10$

Mean $\bar{X} = 35 + 0.16 = 35.16$

$35 + \left(\frac{-2}{125}\right) \times 10$

$35 - 0.16 = 34.84$

ii) Standard deviation $(\sigma) = \sqrt{\frac{\sum fd^2}{N} - \left(\frac{\sum fd}{N}\right)^2} \times C$

$$= \sqrt{\frac{488}{125} - \left(\frac{2}{125}\right)^2} \times 10$$

$$= \sqrt{3.904 - (0.016)^2} \times 10$$

$$= \sqrt{3.904 - 0.0003} \times 10$$

$$= \sqrt{3.9037} \times 10 = 1.9757 \times 10 = 19.76.$$

33. Calculate Standard Deviation and Coefficient of Variation from the following data:

Marks	0-10	10-20	20-30	30-40	40-50	50-60	60-70	70-80
No. of Students	5	7	14	28	12	9	6	2

(Imp.)

Sol :

Calculation of Standard Deviation

Marks	Frequency	Midvalue	d	d^2	fd^2	fd
0 - 10	5	5	–3	9	45	–15
10 - 20	7	15	–2	4	28	–14
20 - 30	14	25	–1	1	14	–14
30 - 40	28	35 A	0	0	0	0
40 - 50	12	45	1	1	12	12
50 - 60	9	55	2	4	36	18
60 - 70	6	65	3	9	5	18
70 - 80	2	75	4	16	32	8
	83				221	13

$$\text{Mean} = A + \frac{\Sigma fd}{N} \times C$$

$$= 35 + \frac{13}{83} \times 10$$

$$= 35 + 15 = 36.5$$

$$\text{S.D} = \sqrt{\frac{\Sigma fd^2}{N} - \left(\frac{\Sigma fd}{N}\right)^2} \times C$$

$$= \sqrt{\frac{221}{83} - \left(\frac{13}{83}\right)^2} \times 10$$

$$= \sqrt{2.66 - 0.024} \times 10$$

$$= \sqrt{2.636} \times 10$$

$$= 1.623 \times 10$$

$$= 16.23$$

$$\text{Coefficient of variation} = \frac{\text{S.D}}{\text{Mean}} \times 100$$

$$= \frac{16.23}{36.5} \times 100$$

$$= 0.44 \times 100 = 44.46$$

2.12 Definition of Moments

Q44. What do you understand by moment?

Ans :

The word 'moment' in statistic has been introduced to describe the characteristics of a distribution either empirically or theoretically. It gives the maximum information about the original data with respect to few quantities. It is used to provide convenient and unifying method to summarize various descriptive statistical measures namely measures of central tendency, measures of dispersion and measures of skewness and kurtosis.

Basically, the measures of skewness and kurtosis uses the moment concept to determine the shape of a distribution. In addition to that, it also determines and tests the symmetry as well as the normality of a distribution.

Thus, it can be inferred that the moment plays an important role in studying the pattern of distribution in order to have a clear idea about the nature of the distribution.

Q45. Explain different types of moments. **(Imp.)**

Ans :

These are difference of the values from central (or) Non central value.

There are two types of moments

1. Non-central moments
2. Central moments

1. **Non-central Moments**

The r^{th} non central moment of the variable x about any arbitrary point A is denoted by $\mu'r$ and is defined as

$$\mu'r = \frac{1}{N}\sum_{i=1}^{n} f_i(x_i - A)^r; N = \sum_{i=1}^{n} f_i$$

r = 0

$$\mu'_0 = \frac{1}{N}\sum_{i=1}^{n} f_i(x_i - A)^0 = \frac{1}{N}\sum_{i=1}^{n} f_i = 1$$

r = 1

$$\mu'_1 = \frac{1}{N}\sum_{i=1}^{n} f_i(x_i - A)'$$

$$= \frac{1}{N}\sum_{i=1}^{n} f_i x_i - \frac{1}{N}\sum_{i=1}^{n} f_i A$$

$$\mu'_1 = \bar{x} - A$$

r = 2

$$\mu'_2 = \frac{1}{N}\sum_{i=1}^{n} f_i(\bar{x} - A)^2$$

If A = 0, these are known as moments about origin.

2. Central Moments

The r[th] central moment of the variable x about the mean $\bar{x}$ is denoted by μ_r and is defined as,

$$\mu_r = \frac{1}{N}\sum_{i=1}^{n} f_i(x_i - \bar{x})^r;\ N = \sum_{i=1}^{n} f_i$$

r = 0

$$\mu_0 = \frac{1}{N}\sum_{i=1}^{n} f_i(x_i - \bar{x})^0$$

$$\mu_0 = 1$$

r = 1

$$\mu_1 = \frac{1}{N}\sum_{i=1}^{n} f_i(x_i - \bar{x})$$

$$= \frac{1}{N}\sum_{i=1}^{n} f_i x_i - \frac{1}{N}\sum_{i=1}^{n} f_i \bar{x}$$

$$= \bar{x} - \bar{x} = 0$$

$$\mu_1 = 0$$

r = 2

$$\mu_2 = \frac{1}{N}\sum_{i=1}^{n} f_i(x_i - \bar{x})^2$$

$$= \sigma^2 = \text{Variance}$$

Q46. Explain Relation between the moments about the mean interms of moments about any arbitrary point.

Ans :

Relation between the central moments in terms of non central moment :

$$\mu_r = \frac{1}{N}\sum_{i=1}^{n} f_i(x_i - \bar{x})^r$$

$$= \frac{1}{N}\sum_{i=1}^{n} f_i [x_i - A + A - \bar{x}]^r; \text{ where 'A' is a constant.}$$

$$= \frac{1}{N}\sum_{i=1}^{n} f_i [x_i - A - \mu_1']^r \text{ where } \mu_1' = \bar{x} - A$$

The binomial expansion

$$(x-y)^n = x^n - nc_1 x^{n-1} y + nc_2 x^{n-2} y^2 - nc_3 x^{n-3} y^3 + \ldots$$

$$\therefore \mu_r = \frac{1}{N}\sum_{i=1}^{n} f_i \Big[(x_i - A)^r - rc_1 (x_i - A)^{r-1} \mu_1' + rc_2 (x_i - A)^{r-2} \mu_1'^2$$

$$- rc_3 (x_i - A)^{r-3} \mu_1'^3 + rc_4 (x_i - A)^{r-4} \mu_1'^4 - \ldots\ldots \Big]$$

$$= \frac{1}{N}\sum_{i=1}^{n} f_i (x_i - A)^r - rc_1 \mu_1' \frac{1}{N}\sum_{i=1}^{n} f_i (x_i - A)^{r-1} + rc_2 \mu_1'^2 \frac{1}{N}\sum_{i=1}^{n} f_i (x_i - A)^{r-2}$$

$$- rc_3 \mu_1'^3 \frac{1}{N}\sum_{i=1}^{n} f_i (x_i - A)^{r-3} + rc_4 \mu_1'^4 \frac{1}{N}\sum_{i=1}^{n} f_i (x_i - A)^{r-4} - \ldots\ldots$$

$$= \mu_r' - rc_1 \mu_1' \mu_{r-1}' + rc_2 \mu_1'^2 \mu_{r-2}' - rc_3 \mu_1'^3 \mu_{r-3}' + rc_4 \mu_1'^4 \mu_{r-4}' - \ldots (1)$$

On substituting $r = 1,2,3,4$ in (1), we get

$$\mu_1 = \mu_1' - 1c_1 \mu_1' \mu_0'$$

$$= \mu_1' - \mu_1' = 0 \quad (\because \mu_0' = 1)$$

$$\therefore \boxed{\mu_1 = 0}$$

$$\mu_2 = \mu_2' - 2c_1 \mu_1' \mu_1' + 2c_2 \mu_1'^2 \mu_0'$$

$$= \mu_2' - 2\mu_1'^2 + \mu_1'^2 \quad (\because \mu_0' = 1)$$

$$\boxed{\mu_2 = \mu_2' - \mu_1'^2}$$

$$\mu_3 = \mu_3' - 3c_1 \mu_1' \mu_2' + 3c_2 \mu_1'^2 \mu_1' - 3c_3 \mu_1'^3 \mu_0'$$

$$= \mu_3' - 3\mu_2' \mu_1' + 3\mu_1'^3 - \mu_1'^3 \quad (\because \mu_0' = 1)$$

$$\boxed{\mu_3 = \mu_3' - 3\mu_2' \mu_1' + 2\mu_1'^3}$$

$$\mu_4 = \mu_4' - 4c_1 \mu_1' \mu_3' + 4c_2 \mu_1'^2 \mu_2' - 4c_3 \mu_1'^3 \mu_1' + 4c_4 \mu_1'^4 \mu_0'$$

$$= \mu_4' - 4\mu_3' \mu_1' + 6\mu_2' \mu_1'^2 - 4\mu_1'^4 + \mu_1'^4 \quad (\because \mu_0' = 1)$$

$$\boxed{\mu_4 = \mu_4' - 4\mu_3' \mu_1' + 6\mu_2' \mu_1'^2 - 3\mu_1'^4}$$

Q47. Establish the relationship between the moments about any arbitrary points in terms of the moments about the mean.

Ans :

From the definition of non - central moments we know that,

$$\mu_r' = \frac{1}{N}\Sigma f_i (x_i - A)^r$$

$$\mu_r' = \frac{1}{N}\Sigma f_i (x_i - \bar{x} + \bar{x} - A)^r$$

$$\mu_r' = \frac{1}{N} f_i (z_i + \mu_1)^r$$

Since $\mu_1' = \bar{x} - A$

$$z_i = x_i - \bar{x}$$

Using Binomial expansion, we get,

$$\mu_r' = \frac{1}{N}\Sigma f_i \left[z_i^r + rC_1 z_i^{r-1}\mu_1' + rC_2 z_i^{r-2}\mu_1'^2 + rC_3 z_i^{r-3}\mu_1'^3 + ... + \mu_1'^r \right]$$

$$\mu_r' = \frac{1}{N}\Sigma f_i z_i^r + rC_1 \frac{1}{N}\Sigma f_i z_i^{r-1}.\mu_1' + rC_2 \frac{1}{N}\Sigma f_i z_i^{r-2}\mu_1'^2 + rC_3 \frac{1}{N}\Sigma f_i z_i^{r-3}\mu_1'^3 + ... + \frac{1}{N}\Sigma f_i \mu_1'^r$$

$$\mu_r' = \frac{1}{N}\Sigma f_i (x_i - \bar{x})^r + rC_1 \frac{1}{N}\Sigma f_i (x_i - \bar{x})^{r-1}\mu_1' + rC_2 \frac{1}{N}\Sigma f_i (x_i - \bar{x})^{r-2}\mu_1'^2 + ... + \mu_1'^r$$

$$\mu_r' = \mu_r + rC_1\mu_{r-1}\mu_1' + rC_2\mu_{r-2}\mu_1'^2 + rC_3\mu_{r-3}\mu_1'^3 + ... + \mu_1'^r$$

If r =2,

$$\mu_2' = \mu_2 + 2C_1\mu_1\mu_1' + 2C_2\mu_0\mu_1'^2$$

$$\mu_2' = \mu_2 + 2\mu_1\mu_1' + \mu_1'^2$$

$$\boxed{\therefore \mu_2' = \mu_2 + \mu_1'^2} \qquad (\because \mu_1 = 0)$$

If r = 3,

$$\mu_3' = \mu_3 + 3C_1\mu_2\mu_1' + 3C_2\mu_1\mu_1'^2 + 3C_3\mu_0\mu_1'^3$$

$$\mu_3' = \mu_3 + 3\mu_2\mu_1' + 0 + \mu_1'^3$$

$$\boxed{\therefore \mu_3' = \mu_3 + 3\mu_2\mu_1' + \mu_1'^3}$$

If r = 4,

$$\mu_4' = \mu_4 + 4C_1\mu_3\mu_1' + 4C_2\mu_2\mu_1'^2 + 4C_3\mu_1\mu_1'^3 + 4C_4\mu_0\mu_1'^4$$

$$\boxed{\mu_4' = \mu_4 + 4\mu_3\mu_1' + 6\mu_2\mu_1'^2 + \mu_1'^4}$$

Note

The mean and variance from μ_1' and μ_2' about any arbitrary origin A can be defined as, $\bar{x} = \mu_1' + A$ and $\sigma^2 = \mu_2 - \mu_1$.

2.13 Measures of Skewness

Q48. Explain briefly about Skewness.

Ans : **(Imp.)**

Introduction

The word skewness refers to lack of symmetry. Non-normal or asymmetrical distribution is called skew distribution. Two frequency distributions may have the same mean and standard deviation and yet may differ with respect to another characteristics- the skewness or, asymmetry of the distribution. Any measure of skewness indicates the difference between the manner in which items are distributed in a particular distribution compared with a normal distribution. Lack of symmetry or skewness in frequency distributions is due to the existence of a longer tail on one side (either to the left or to the right), which has no counterpart on the other side. If the larger tail is on the right, we say that the distribution is positively skewed; whereas if the longer tail is on the left side.

Definitions

Some important definitions of skewness are as follows :

- **According to Corxton & Cowden** "When a series is not symmetrical it is said to be asymmetrical or skewed."
- **According to Morris Hamburg** "Skewness refers to the asymmetry or lack of symmetry in the shape of a frequency distribution."
- **According to Simpson & Kafka** "Meas-ures of skewness tell us the direction and the extent of skewness. In symmetrical distribution the mean, median and mode are identical. The more the mean moves away from the mode, the larger the asymmetry or skewness."
- **According to Garrett** "A distribution is said to be 'skewed' when the mean and the median fall at different points in the distribution, and the balance (or centre of gravity) is shifted to one side or the other – to left or right."

1. **Positively Skewed Distribution**

If the longer tail of the distribution is towards the higher values (or) right hand side, the skewness is positive. Positive skewness occurs when mean is increased by some unusually high values thus satisfying the following properties,

(i) Mean > Median > Mode

(ii) Right tail is longer than its left tail

(iii) $(Q_3 - \text{Median}) > (\text{Median} - Q_1)$.

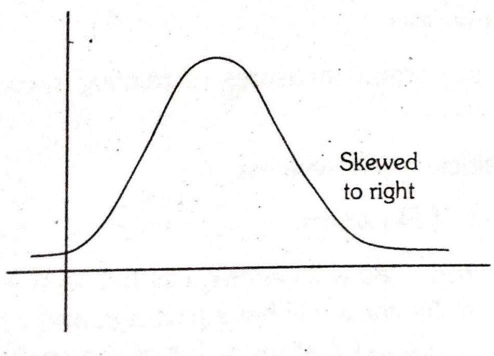


Fig.: Positively Skewed Distribution

2. **Negatively Skewed Distribution**

If the longer tail is towards the lower value or left hand side, the skewness is negative.

Negative skewness arises when mean is decreased by some extremely low values, thus satisfying the following properties,

(i) Mean < Median < Mode

(ii) Left tail is longer than its right tail

(iii) (Q_3 – Median) < (Median - Q_1)

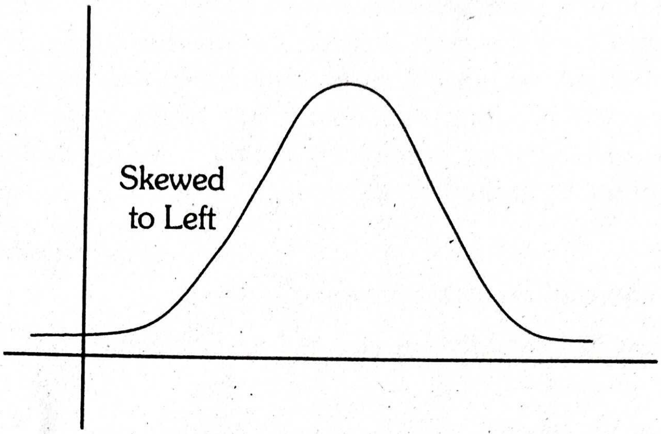


Fig.: Negatively Skewed Distribution

2.13.1 Karl Pearson's Coefficient of Skew-ness - Bowley's Coefficient of Skewness

Q49. Explain the different measures of skewness.

Ans :

Skewness can be measured absolutely and relatively. Absolute measures are also known as measures of skewness whereas relative measures are termed as the coefficients of skewness.

1. **Absolute Measures of Skewness**

In a skewed distribution the three measures of central tendency differ. Accordingly skewness may be worked out in absolute amount with the help of the following formulae :

Absolute Skewness $S_K = \overline{X} - \text{Mode}$

Absolute Skewness $S_K = \overline{X} - \text{Median}$

Absolute Skewness $S_K = \text{Median} - \text{Mode}$

2. **Relative Measures of Skewness**

The following are the four important measures of relative skewness, termed as coefficients of skewness :

(i) The Karl Pearson's Coefficient of Skewness.

(ii) The Bowley's Coefficient of Skewness.

The results obtained by these formulae will generally lie between +1 and -1. When the distribution is positively skewed, the coefficient of skewness will have plus sign and when it is negatively skewed it will have negative sign. It should be remembered that the value of the coefficient will never exceed 1.

(i) **Karl Pearson's Coefficient of Skewness**

This method of measuring skewness, also known as Pearsonian Coefficient of Skewness, was suggested by Karl Pearson, a great British Biometrician and Statistician. It is based upon the difference between mean and mode. This difference is divided by standard deviation to give a relative measure. The formula thus becomes :

$$Sk_p = \frac{\text{Mean} - \text{Mode}}{\text{Standard Deviation}}$$

Sk_p = Karl Pearson's coefficient of skewness

When Mode is ill-defined

Coefficient of Skewness

$$(S_{K_P}) = \frac{3\,(\text{Mean} - \text{Median})}{\sigma}$$

There is no limit to this measure in theory and this is a slight drawback. But in practice the value given by this formula is rarely very high and usually lies between ± 1.

(ii) **Bowley's Coefficient of Skewness**

It is based on quartiles (Q_3 and Q_1).

Bowley's Coeff. of Sk.

$$S_{K_B} = \frac{(Q_3 - M) - (M - Q_1)}{(Q_3 - M) + (M - Q_1)}$$

$$\boxed{S_{K_B} = \frac{Q_3 + Q_1 - 2M}{Q_3 - Q_1}}$$

where, M = Median

This measure is called the quartile measure of skewness and values of the coefficient, thus obtained vary between ± 1.

PROBLEMS

34. **Calculate Karl Pearson's Coefficient of Skewness and coefficient of variation from the following data.**

Mode = 33.5, Mean = 30.08,

Standard deviation = 13.405

Sol :

Co-efficient of Skewness

$$= \frac{\text{Mean - Mode}}{\text{Standard Deviation}} = \frac{30.08 - 33.5}{13.405}$$

$= -0.255$

Co-efficient of Variation

$$= \frac{\sigma}{\bar{X}} \times 100$$

$$= \frac{13,405}{30.08} \times 100 = 44.56.$$

35. Calculate Karl Pearson's Coefficient of Skewness when $\bar{X}$ = 20, Mode = 20, = 13.62.

Sol :

Co-efficient of Skewness

$$= \frac{\text{Mean - Mode}}{\text{Standrd Deviation}}$$

$$= \frac{20-20}{13.62} = \frac{0}{13.62}\ 0.$$

36. In a certain distribution the following results were obtained.

Mean = 45 Median = 48 Karl Pearson co-efficient of skewness = – 0.4, Calculate the value of standard deviation.

Sol :

Given that S_{kp} = – 0.40, $\bar{x} = 45, M = 48$

$$S_{kp} = 3\left(\frac{\bar{x}-M}{\sigma}\right)$$

$$-0.40 = 3\left(\frac{45-48}{\sigma}\right)$$

$$-0.40 = 3\left(\frac{-3}{\sigma}\right)$$

$$-0.40 = \frac{-9}{\sigma}$$

$$\sigma = \frac{+9}{+0.40}$$

$$\sigma = 22.5$$

37. For a distribution, if mean is 200, the co-efficient variation is 8 and Karl Pearson's co-efficient of skewness is 0.3. Find the mode and median.

Sol :

Given that,

Arithmetic Mean ($\bar{X}$) = 200

Co-efficient of Variation (C, V) = 8

Skewness = 0.3

With the help of coefficient of variation, standard deviation can be determined.

$$C.V = \frac{\sigma}{x} \times 100 = \frac{\sigma}{200} \times 100$$

$$\sigma = \frac{8 \times 200}{100}$$

$$\sigma = \frac{8 \times 2}{1}$$

$\therefore$ Standard Deviation (s) = 16

Karl person's coefficient of skewness is used to determine the mode.

$$\boxed{\therefore \text{Coefficient of skewness } (SK_p) = \frac{\text{Mean} - \text{Mode}}{\sigma}}$$

$$0.3 = \frac{200 - \text{Mode}}{16}$$

$$0.3 \times 16 = 200 - \text{Mode}$$

$$4.8 = 200 - \text{Mode}$$

$$\text{Mode} = 200 - 4.8$$

$$\boxed{\therefore \text{Mode} = 195.2}$$

Median can be calculated as follows,

$$\boxed{\text{Mode} = 3\ \text{Median} - 2\ \text{Mean}}$$

$$195.2 = 3\ \text{Median} - 2 \times 200$$

$$195.2 + 400 = 3\ \text{Median}$$

$$595.2 = 3\ \text{Median}$$

$$\text{Median} = \frac{595.2}{3}$$

$$\boxed{\therefore \text{Median} = 198.4}$$

38. From the following data calculate Karl Pearson's Coefficient of Skewness.

Marks	0 - 10	10 - 20	20 - 30	30 - 40	40 - 50	50 - 60	60 - 70
No. of Students	12	28	40	60	32	18	10

Sol : (Imp.)

Marks	No. of Students (f)	Mid (value) (X)	$d = \frac{X-A}{10}$	d^2	fd	fd^2
0 - 10	12	5	– 3	9	– 36	108
10 - 20	28	15	– 2	4	– 56	112
20 - 30	f_1 40	25	– 1	1	– 40	40
L 30 – 40	f 60 A	35 A	0	0	0	0
40 – 50	f_2 32	45	1	1	32	32
50 - 60	18	55	2	4	36	72
60 - 17	10	65	3	9	30	90
	200				– 34	454

Karl Pearsons Coefficient of Skewness $= \frac{\text{Mean} - \text{Mode}}{\text{S.D}}$

$$\text{Mean} = A + \frac{\Sigma fd}{N} \times C$$

$$= 35 + \frac{(-34)}{200} \times 10$$

$$= 35 - \frac{340}{200}$$

$$= 35 - 1.7 = 33.3$$

$$\text{Mode} = L + \frac{\Delta_1}{\Delta_1 + \Delta_2} \times C$$

$$\Delta_1 = f - f1$$

$$= 60 - 40 = 20$$

$$\Delta_2 = f - f2$$

$$= 60 - 32 = 28$$

$$= 30 + \frac{20}{20+28} \times 10$$

$$= 30 + \frac{200}{48}$$

$$= 30 + 4.1 = 34.1$$

$$\text{S.D} = \sqrt{\left(\frac{\Sigma fd^2}{N}\right) - \left(\frac{\Sigma fd}{N}\right)^2} \times C$$

$$= \sqrt{\frac{454}{200} - \left(\frac{-34}{200}\right)^2} \times 10 = \sqrt{2.27 + 0.0289} \times 10$$

$$= \sqrt{2.2989} \times 10 = 1.516 \times 10 = 15.16$$

$$S_{KP} = \frac{\text{Mean} - \text{Mode}}{\text{S.D}} = \frac{33.3 - 34.1}{15.16} = \frac{-0.18}{15.16}$$

$$= -0.0531$$

39. Calculate Coefficient of Skewness from the following information.

First Quartile (Q_1) = 14 cm

Third Quartile (Q_3) = 25 cm

Median = 18 cm

Sol :

Bowley's Coefficient of Skewness

$$= \frac{Q_3 + Q_1 - 2M}{Q_3 - Q_1}$$

$$= \frac{25 + 14 - 2(18)}{25 - 14} = \frac{3}{11} = 0.273.$$

40. Calculate the coefficient of skewness based on quartiles

Variable	10-20	20-30	30-40	40-50	50-60	60-70	70-80	80-90
Frequency	5	9	14	20	25	15	8	4

(Imp.)

Sol :

	Variable	Frequency	Cumulative frequency
	10-20	5	5
	20-30	9	14F
L	30-40	14 f	28 (Q1)
	40-50	20	48F
L	50-60	25 f	73F (Median)
L	60-70	15 f	88 (Q3)
	70-80	8	96
	80-90	4	100=N

$$\text{Coefficient of skewness} = \frac{Q_3 + Q_1 - 2\text{median}}{Q_3 - Q_1}$$

culation of Median

$$\text{Median} = L + \frac{\frac{N}{2} - F}{f} \times C = \frac{N}{2} = \frac{100}{2} = 50^{th} \text{ observation}$$

Median lies in the class of 50 – 60

$$= 50 + \frac{50-48}{25} \times 10$$

$$= 50 + \frac{2 \times 10}{25}$$

$$= 50 + \frac{20}{25} = 50 + 0.8 = 50.8$$

Calculation of Q_3 and Q_1

$$Q_3 = L + \frac{\frac{3N}{4} - F}{f} \times C$$

$$\frac{3N}{4} = \frac{3 \times 100}{4} = \frac{300}{4} = 75^{th} \text{ observation}$$

Q_3 lies in the class of 60 – 70

$$= 60 + \frac{75-73}{15} \times 10 = 60 + \frac{2 \times 10}{15}$$

$$= 60 + \frac{20}{15} = 60 + 1.33 = 61.33$$

$$Q_1 = L + \frac{\frac{N}{4} - F}{f} \times C$$

$$\frac{N}{4} = \frac{100}{4} = 25^{th} \text{ observation}$$

Q_1 lies in the class of 30 – 40

$$= 30 + \frac{25-14}{14} \times 10$$

$$= 30 + \frac{11}{14} \times 10$$

$$= 30 + \frac{110}{14}$$

$$= 30 + 7.86 = 37.86$$

Bowley's coefficient of skewness

$$= \frac{Q_3 + Q_1 - 2\text{ median}}{Q_3 - Q_1} = \frac{61.33 + 37.86 - 2(50.8)}{61.33 - 37.86}$$

$$= \frac{99.19 - 101.6}{23.47} = \frac{-2.41}{23.47}$$

$$\therefore S_{K_B} = -0.1027$$

41. **Calculate Coefficient of Skewness based on Quarties from the following data:**

Class Interval	10 - 20	20 - 30	30 - 40	40 - 50	50 - 60	60 - 70	70 - 80	80 - 90
f	6	10	18	30	12	10	6	2

Sol.

C.I	F	CF
10 - 20	6	6
20 - 30	10	16 F
L 30 - 40	18 f	34 F (Q1)
L 40 - 50	30 f	64 F (Median)
L 50 - 60	12 f	76 (Q3)
60 - 70	10	86
70 - 80	6	92
80 - 90	2	94
	94	

$$\text{Bowley's Coefficient of Skewness} = \frac{Q_3 + Q_1 - 2\text{ Median}}{Q_3 - Q_1}$$

Calculation of Median

$$= L + \frac{\frac{N}{2} - F}{f} \times C$$

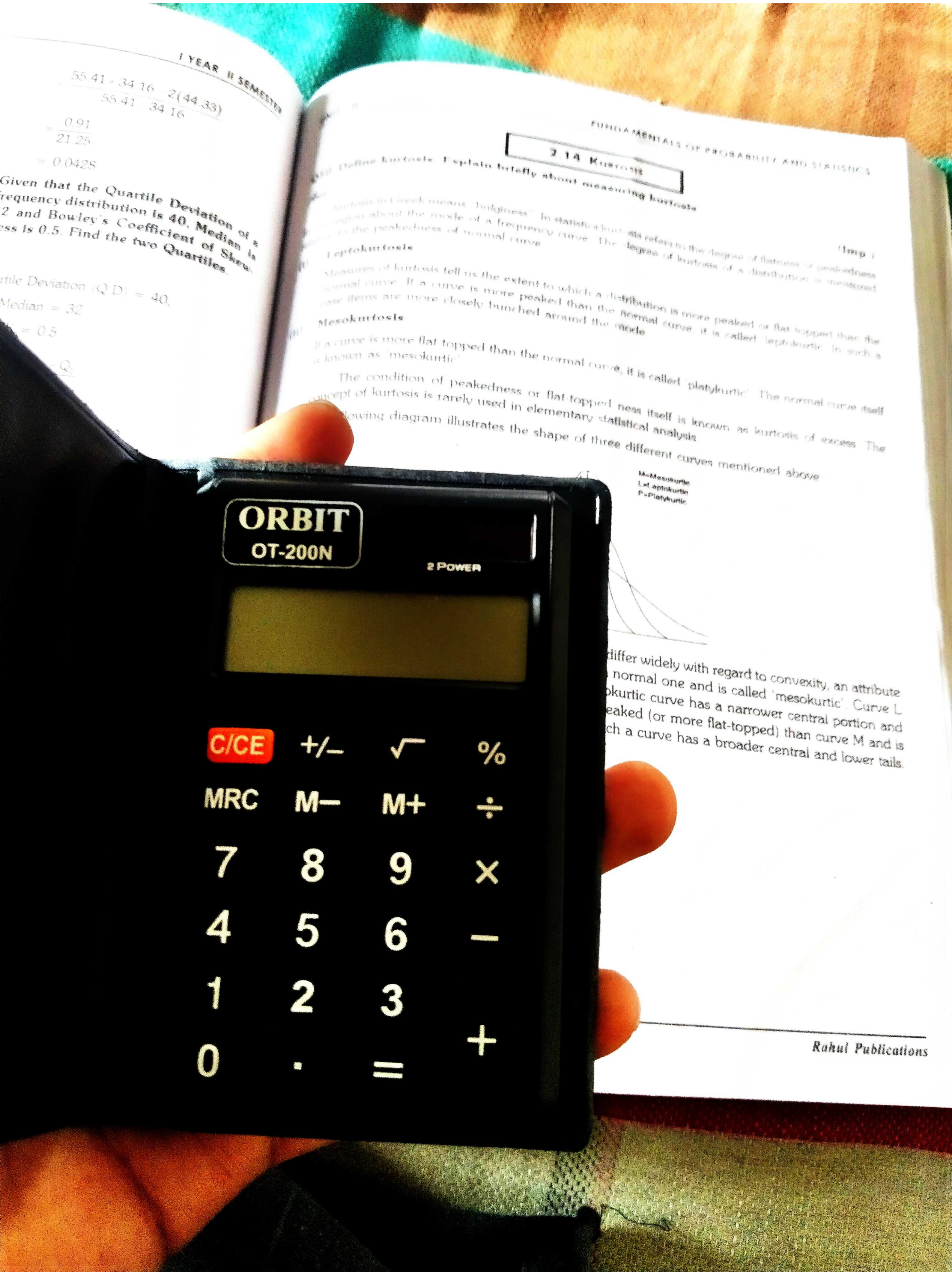
$$\frac{N}{2} = \frac{94}{2} = 47^{th} \text{ item}$$

Median lies in the class of 40 – 50

$$= 40 + \frac{47 - 34}{30} \times 10$$

$$= 40 + \frac{130}{30}$$

$$= 40 + 4.33 = 44.33$$


I YEAR II SEMESTER
= 0.0428
Given that the Quartile Deviation of a frequency distribution is 40, Median is 32 and Bowley's Coefficient of Skewness is 0.5. Find the two Quartiles.
Median = 32
FUNDAMENTALS OF PROBABILITY AND STATISTICS
2.14 Kurtosis
Define kurtosis. Explain briefly about measuring kurtosis
(Imp.)
Leptokurtosis
Mesokurtosis
If a curve is more flat topped than the normal curve, it is called 'platykurtic'. The normal curve itself is known as 'mesokurtic'.
The condition of peakedness or flat topped ness itself is known as kurtosis of excess. The concept of kurtosis is rarely used in elementary statistical analysis.
following diagram illustrates the shape of three different curves mentioned above
M=Mesokurtic
L=Leptokurtic
P=Platykurtic
differ widely with regard to convexity, an attribute
normal one and is called 'mesokurtic'. Curve L
kurtic curve has a narrower central portion and
eaked (or more flat-topped) than curve M and is
ch a curve has a broader central and lower tails.
Rahul Publications
ORBIT
OT-200N
2 POWER
C/CE
+/–
√
%
MRC
M–
M+
÷
7
8
9
×
4
5
6
–
1
2
3
+
0
.
=

Calculation of Q_3

$$Q_3 = L + \frac{\frac{3N}{4} - F}{f} \times C$$

$$\frac{3N}{4} = \frac{3 \times 94}{4} = 70.5$$

Q_3 lies in the class of 50 – 60

$$= 50 + \frac{70.5 - 64}{12} \times 10$$

$$= 50 + \frac{6.5}{12} \times 10$$

$$= 50 + \frac{6.5}{12}$$

$$= 50 + 5.41$$

$$= 55.41$$

Calculation of Q1 = Q_1

$$= L + \frac{\frac{N}{4} - F}{f} \times C$$

$$\frac{N}{4} = \frac{94}{4} = 23.5$$

Q_1 lies in the class of 30 – 40

$$= 30 + \frac{23.5 - 16}{18} \times 10$$

$$= 30 + \frac{7.5}{18} \times 10$$

$$= 30 + \frac{7.5}{18}$$

$$= 30 + 4.16$$

$$= 34.16$$

Coefficient of skerimes

$$= \frac{Q_3 + Q_1 - 2\text{ Median}}{Q_3 - Q_1}$$

$$= \frac{55.41 + 34.16 - 2(44.33)}{55.41 - 34.16}$$

$$= \frac{0.91}{21.25}$$

$$= 0.0428.$$

42. Given that the Quartile Deviation of a frequency distribution is 40, Median is 32 and Bowley's Coefficient of Skewness is 0.5. Find the two Quartiles.

Sol :

Quartile Deviation (Q.D) = 40,

Median = 32

$SK_B = 0.5$

$$QD = \frac{Q_3 - Q_1}{2}$$

$$40 = \frac{Q_3 - Q_1}{2} \Rightarrow 80$$

$$= \theta_3 - \theta_1$$

$$\Rightarrow \theta_3 = 80 + \theta_1$$

$$SK_B = \frac{Q_3 + Q_1 - 2M}{Q_3 - Q_1}$$

$$0.5 = \frac{(80 + Q_1) + Q_1 - 2(32)}{(80 + Q_1) - Q_1}$$

$$0.5 = \frac{80 + 2Q_1 - 64}{80 + Q_1 - Q_1}$$

$$0.5 \times 80 = 16 + 2\theta_1$$

$$40 - 16 = 2\theta_1$$

$$Q_1 = \frac{24}{2}$$

$$= 12$$

$$\therefore \quad Q_3 = 80 + Q_1$$

$$= 80 + 12$$

$$= 92$$

2.14 Kurtosis

Q50. Define kurtosis. Explain briefly about measuring kurtosis.

Ans : **(Imp.)**

Kurtosis in Greek means "bulginess". In statistics kurtosis refers to the degree of flatness or peakedness in the region about the mode of a frequency curve. The degree of kurtosis of a distribution is measured relative to the peakedness of normal curve.

(i) **Leptokurtosis**

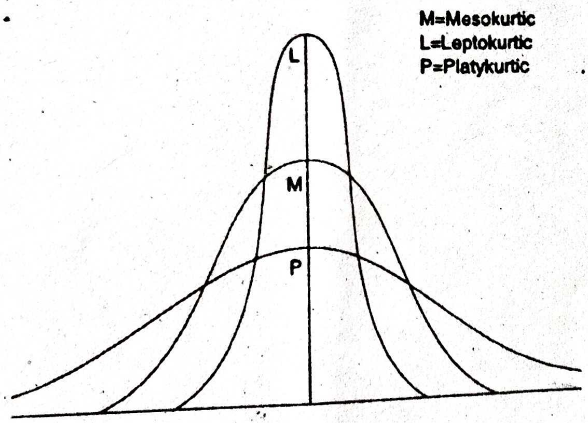
Measures of kurtosis tell us the extent to which a distribution is more peaked or flat-topped than the normal curve. If a curve is more peaked than the normal curve, it is called 'leptokurtic. In such a case items are more closely bunched around the mode.

(ii) **Mesokurtosis**

If a curve is more flat-topped than the normal curve, it is called 'platykurtic'. The normal curve itself is known as 'mesokurtic'.

The condition of peakedness or flat-topped ness itself is known as kurtosis of excess. The concept of kurtosis is rarely used in elementary statistical analysis.

The following diagram illustrates the shape of three different curves mentioned above :



The above diagram clearly shows that these curves differ widely with regard to convexity, an attribute which Karl Pearson referred to as 'kurtosis'. Curve M is a normal one and is called 'mesokurtic'. Curve L is more peaked than M and is called 'leptokurtic'. A leptokurtic curve has a narrower central portion and higher tails than does the normal curve. Curve P is less peaked (or more flat-topped) than curve M and is called 'platykurtic'. As may be seen from the diagram, such a curve has a broader central and lower tails.

Short Question and Answers

1. What is arithmetic mean?

Ans :

Meaning

Arithmetic Average (or) Mean of a series is the figure obtained by dividing the total "Values of the various items by their number. In other words it is the sum of the values divided by their number. Arithmetic means is the most widely used measure of central tendency.

2. Define median.

Ans :

Meaning

If a group of N observations is arranged in ascending or descending order of magnitude, then the middle value is called median of these observations and is denoted by M.

That is, $M = \frac{N+1}{2}$ th observation.

Definition

According to Croxton and Cowden, "The median is that value which divides a series so that one half or more of the items are equal to or less than it and one half or more of the items are equal to or greater than it."

3. Define mode.

Ans :

Meaning

Mode may be defined as the value that occurs most frequently in a statistical distribution or it is defined as that exact value in the ungrouped data if each sample which occurs most frequently.

Definitions

(i) **According to Croxton and Cowden,** "The mode of a distribution is the value at the point around which the items tend to be most heavily concentrated. It may be regarded as the most typical of a series of values."

(ii) **According to A.M. Tuttle,** "Mode is the value which has the greatest frequency density in its immediate neighbourhood."

(iii) **According to Zizek,** "The mode is the value occurring most frequently in a series of items and around which the other items are distributed most densely."

4. Define Geometric Mean.

Ans :

Geometric mean is the n^{th} root of the product of n items of a series. If there are 2 numbers, say a and b, the Geometric mean of the two numbers is the square root of the product of the 2 numbers. Similarly, if there are 3 numbers, Geometric mean of the three numbers would be the cube root of the product of the 3 numbers. Thus, Geometric mean would be $(abc)^{1/3}$. This concept can be applied to as many numbers as possible.

5. Define Dispersion.

Ans :

Meaning

The concept of dispersion is related to the extent of scatter or variability in observations. The variability, in an observation, is often measured as its deviation from a central value. A suitable average of all such deviations is called the measure of dispersion. Since most of the measures of dispersion are based on the average of deviations of observations from an average, they are also known as the averages of second order.

Definitions

As opposed to this, the measures of central tendency are known as the averages of first order. Some important definitions of dispersion are given below:

(i) **According to A.L. Bowley,** "Dispersion is the measure of variation of the items."

(ii) **According to Connor,** "Dispersion is the measure of extent to which individual items vary."

(iii) **According to Simpson and Kafka,** "The measure of the scatteredness of the mass of figures in a series about an average is called the measure of variation or dispersion."

(iv) **According to Spiegel,** "The degree to which numerical data tend to spread about an average value is called variation or dispersion of the data."

6. **What is Range ?**

Ans :

Range is the simplest method of studying dispersion. It is defined as the difference between the value of the smallest item and the value of the largest item included in the distribution. Symbolically,

$$\boxed{\text{Range} = L - S}$$

where

L = Largest item, and

S = Smallest item.

7. **What is Quartile Deviation ?**

Ans :

It is based on two extreme items and it fails to take account of the scatter within the range. From this there is reason to believe that if the dispersion of the extreme items is discarded, the limited range thus established might be more instructive.

For this purpose there has been developed a measure called the interquartile range, the range which includes the middle 50 per cent of the distribution. That is, one quarter of the observations at the lower end, another quarter of the observations at the upper end of the distribution are excluded in computing the interquartile range. In other words, interquartile range represents the difference between the third quartile and the first quartile.

8. **Define mean deviation.**

Ans :

Meaning

"Mean Deviation of a series is the arithmetic average of the deviations of various items from a measure of central tendency (either mean, median or mode)". Theoretically, deviations can be taken from any of the three averages mentioned above, but in actual practice it is calculated either from mean or from Median. While Calculating deviations algebraic signs are not taken into account.

9. **What is standard deviation ?**

Ans :

The standard deviation concept was introduced by Karl Pearson in 1823. It is by far the most important and widely used measure of studying dispersion. Its significance lies in the fact that it is free from those defects from which the earlier methods suffer and satisfies most of the properties of a good measure of dispersion. Standard deviation is also known as root mean square deviation for the reason that it is the square root of the mean of the squared deviation from the arithmetic mean. Standard deviation is denoted by the small Greek letter σ (read as sigma).

The standard deviation measures the absolute dispersion (or variability of distribution; the greater the amount of dispersion or variability), the greater the standard deviation, the greater will be the magnitude of the deviations of the values from their mean. A small standard deviation means a high degree of uniformity of the observation as well as homogeneity of a series; a large standard deviation means just the opposite.

10. **Moment.**

Ans :

The word 'moment' in statistic has been introduced to describe the characteristics of a distribution either empirically or theoretically. It gives the maximum information about the original data with respect to few quantities. It is used to provide convenient and unifying method to summarize various descriptive statistical measures namely measures of central tendency, measures of dispersion and measures of skewness and kurtosis.

Basically, the measures of skewness and kurtosis uses the moment concept to determine the shape of a distribution. In addition to that, it also determines and tests the symmetry as well as the normality of a distribution.

Thus, it can be inferred that the moment plays an important role in studying the pattern of distribution in order to have a clear idea about the nature of the distribution.

11. Skewness.

Ans :

Some important definitions of skewness are as follows :

- **According to Corxton & Cowden** "When a series is not symmetrical it is said to be asymmetrical or skewed."
- **According to Morris Hamburg** "Skewness refers to the asymmetry or lack of symmetry in the shape of a frequency distribution."
- **According to Simpson & Kafka** "Meas-ures of skewness tell us the direction and the extent of skewness. In symmetrical distribution the mean, median and mode are identical. The more the mean moves away from the mode, the larger the asymmetry or skewness."
- **According to Garrett** "A distribution is said to be 'skewed' when the mean and the median fall at different points in the distribution, and the balance (or centre of gravity) is shifted to one side or the other – to left or right."

12. Define kurtosis.

Ans :

Kurtosis in Greek means "bulginess". In statistics kurtosis refers to the degree of flatness or peakedness in the region about the mode of a frequency curve. The degree of kurtosis of a distribution is measured relative to the peakedness of normal curve.

(i) Leptokurtosis

Measures of kurtosis tell us the extent to which a distribution is more peaked or flat-topped than the normal curve. If a curve is more peaked than the normal curve, it is called 'leptokurtic. In such a case items are more closely bunched around the mode.

(ii) Mesokurtosis

If a curve is more flat-topped than the normal curve, it is called 'platykurtic'. The normal curve itself is known as 'mesokurtic'.

Exercise Problems

1. The following table givens the marks secured by 60 students of a class.

Marks	No.of Students (f)
10-20	8
20-30	12
30-40	20
40-50	10
50-60	7
60-70	3

Calculate the arithmetic mean and geometric mean.

[Ans: A.M = 32.50, G.H=33.13]

2. Calculate (a) the median (b) the mode and (c) two quartiles from the following data.

Age	No.of persons
20-25	100
25-30	140
30-35	200
35-40	360
40-45	300
45-50	240
50-55	140
55-60	120

[Ans: Median = 40, Mode = 38.64, Q_1 = 34, Q_3 =47.08]

3. Calculate co-efficient of M.D. through mean and median.

Marks	5	10	15	20	25
Students	6	7	8	11	8

[Ans : Co-efficient of M.D. through mean = 0.363, through median = 0.383]

4. Calculate the value of coefficeint of mean deviation (from median) of the following data:

Marks	No. of students
10 - 20	2
20 - 30	6
30 - 40	12
40 - 50	18
50 - 60	25
60 - 70	20
70 - 80	10
80 - 90	7

[Ans : Co-efficient of M.D. = 0.2363]

5. Calculate Karl Pearson's Coefficient of Skewness from the following data

Class	**Frequency**
70 - 80	18
60 - 70	22
50 - 60	30
40 - 50	35
30 - 40	21
20 - 30	11
10 - 20	6
0 - 10	5

[Ans: 0.0456]

Choose the Correct Answers

Which measure of central tendency includes the magnitude of scores? [a]

(a) Mean (b) Mode

(c) Median (d) Range

Which of the following is not a disadvantage of using mean? [c]

(a) It is affected by extreme values

(b) It cannot be computed in grouped data with open-ended class intervals

(c) It does not possess the desired algebraic property

(d) None of the above

The two methods of finding mode in a discrete series are ______. [c]

(a) Grouping method and ascending method

(b) Table method and midpoint method

(c) Grouping method and inspecting method

(d) None of the above

When the values in a series do not have equal importance, we calculate the ______. [b]

(a) Mode (b) Weighted mean

(c) Arithmetic mean (d) None of the above

To calculate the median, all the items of a series have to be arranged in a/an ______. [c]

(a) Descending order (b) Ascending order

(c) Ascending or descending order (d) None of the above

Which of the following are methods under measures of dispersion? [d]

(a) Standard deviation (b) Mean deviation

(c) Range (d) All of the above

Which of the following are characteristics of a good measure of dispersion? [d]

(a) It should be easy to calculate

(b) It should be based on all the observations within a series

(c) It should not be affected by the fluctuations within the sampling

(d) All of the above

If all the observations within a series are multiplied by five, then ______. [d]

(a) The new standard deviation would be decreased by five

(b) The new standard deviation would be increased by five

(c) The new standard deviation would be half of the previous standard deviation

(d) The new standard deviation would be multiplied by five

The coefficient of variation is a percentage expression for ______. [a]

(a) Standard deviation (b) Quartile deviation

(c) Mean deviation (d) None of the above

While calculating the standard deviation, the deviations are only taken from ______. [d]

(a) The mode value of a series (b) The median value of a series

(c) The quartile value of a series (d) The mean value of a series

Fill in the Blanks

1. Measures of central tendency are also known as ______.
2. The value of mode can also be ascertained graphically with the help of ______
3. The ______ provides a single value which represents the features of the whole group.
4. Karl Pearson introduced the concept of standard deviation in the year ______.
5. The term variance was used to describe the square of the standard creviation by ______ in 19
6. Coefficient of Variation (CV) was proposed by ______.
7. Relative measures of dispersion is also known as ______.
8. ______ refers to lack of symmetry in a frequency distribution.
9. The ______ system includes measures like median, quartiles, deciles, percentiles, and so on
10. Skewness ranges from negative to positive ______

Answers

1. Averages
2. Histogram
3. Measures of central tendency
4. 1823
5. R. A. Fisher
6. Karl Pearson
7. Coefficient of Dispersion
8. Skewess
9. Percentile
10. Infinity.