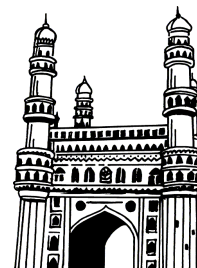


Rahul's ✓
Topper's Voice

AS PER
CBCS SYLLABUS



B.Com.

II Year III Sem

Latest 2022 Edition

BUSINESS STATISTICS-I

- ☞ Study Manual
- ☞ FAQ's & Important Questions
- ☞ Short Question and Answers
- ☞ Choose the Correct Answer
- ☞ Fill in the blanks
- ☞ Solved Previous Question Papers
- ☞ Solved Model Papers

Useful for :

Osmania University
Kakatiya University
Satavahana University
Mahatma Gandhi University
Palamuru University
Telangana University

Price :
₹.179/-

- by -

WELL EXPERIENCED LECTURER



Rahul Publications™

Hyderabad. Ph : 66550071, 9391018098

All disputes are subjects to Hyderabad Jurisdiction only

B.Com.

II Year III Sem

(All Universities in Telangana)

BUSINESS STATISTICS-I

Inspite of many efforts taken to present this book without errors, some errors might have crept in. Therefore we do not take any legal responsibility for such errors and omissions. However, if they are brought to our notice, they will be corrected in the next edition.

© No part of this publications should be reporduced, stored in a retrieval system, or transmitted in any form or by any means, electronic, mechanical, photocopying, recording and/or otherwise without the prior written permission of the publisher

Price ` . 179-00

Sole Distributors :

☎ : 66550071, Cell : 9391018098

VASU BOOK CENTRE

Shop No. 2, Beside Gokul Chat, Koti, Hyderabad.

Maternity Hospital Opp. Lane, Narayan Naik Complex, Koti, Hyderabad.

Near Andhra Bank, Subway, Sultan Bazar, Koti, Hyderabad -195.

BUSINESS STATISTICS-I

CONTENTS

STUDY MANUAL

FAQ's and Important Questions	V - IX
Unit - I	1 - 48
Unit - II	49 - 92
Unit - III	93 - 154
Unit - IV	155 - 204
Unit - V	205 - 236
Logarithms Tables	237 - 240

SOLVED PREVIOUS QUESTION PAPERS

December - 2017	241 - 243
May / June - 2018	244 - 246
November / December - 2018	247 - 248
June / July - 2019	249 - 250
November / December - 2019	251 - 262
July - 2021	263 - 274

SOLVED MODEL PAPERS

Model Paper - I	275 - 277
Model Paper - II	278 - 280

SYLLABUS

UNIT - I

INTRODUCTION:

Origin and Development of Statistics - Definition - Importance and Scope - Limitations of Statistics - Distrust of Statistics.

Statistical Investigation: Planning of statistical investigation - Census and Sampling methods - Collection of primary and secondary data - Statistical errors and approximation - classification and Tabulation of data - Frequency distribution

UNIT - II

DIAGRAMMATIC AND GRAPHIC PRESENTATION:

Diagrammatic presentation: One Dimensional and Two Dimensional Diagrams - Pictograms - Cartograms - Graphic presentation: Technique of Construction of Graphs - Graphs of Frequency Distribution - Graphs of Time Series or Histograms

UNIT - III

MEASURES OF CENTRAL TENDENCY:

Introduction -Significance - Arithmetic Mean - Geometric Mean - Harmonic Mean – Mode - Median - Quartiles and Percentiles - Simple and Weighted Averages - Uses and Limitations of different Averages

UNIT - IV

MEASURES OF DISPERSION, SKEWNESS AND KURTOSIS:

Measures of Dispersion: Significance - Characteristics - Absolute and Relative Measures - Range- Quartile Deviation - Mean Deviation- Standard Deviation - Coefficient of Variation

Measures of Skewness - Karl Pearson's Coefficient of Skewness - Bowley's Coefficient of Skewness - Kelly's Measure of Skewness - Kurtosis: Mesokurtosis, Platy kurtosis and Leptokurtosis

UNIT - V

CORRELATION:

Meaning -Types - Correlation and Causation - Methods: Scatter Diagram - Karl Person's Coefficient of Correlation - Probable Error and Interpretation of Coefficient of Correlation - Rank Correlation - Concurrent Deviation Method

Contents

UNIT - I

Topic	Page No.
1.1 Origin and Development of Statistics	1
1.1.1 Definition	2
1.1.2 Importance	4
1.1.3 Scope	5
1.1.4 Limitations	7
1.2 Distrust of Statistics	9
1.3 Statistical Investigation	10
1.3.1 Planning of Statistical Investigation	10
1.4 Census and Sampling Methods	12
1.5 Collection of Primary and Secondary Data	15
1.6 Statistical Errors and Approximation	27
1.7 Classification and Tabulation of Data	27
1.8 Frequency Distribution	32
➤ Short Question and Answers	40 - 46
➤ Choose the Correct Answer	47 - 47
➤ Fill in the blanks	48 - 48

UNIT - II

2.1 Diagrammatic Presentation	49
2.2 Types of Diagrams	51
2.2.1 One Dimensional Diagrams	52
2.2.2 Two Dimensional Diagrams	63
2.2.3 Pictograms	68
2.2.4 Cartograms	68
2.3 Graphic Presentation	69
2.3.1 Technique of Construction of Graphs	69
2.4 Types of Graphs	72
2.4.1 Graphs of Frequency Distribution	73
2.4.1.1 Histogram	73
2.4.1.2 Frequency Polygon	73

Topic	Page No.
2.4.1.3 Frequency Curve	74
2.4.1.4 Ogive Curve	75
2.4.2 Graphs of Time Series or Historigrams	82
2.4.2.1 Horizontal Line Graph	82
2.4.2.2 Band Graphs	82
2.4.2.3 Net Balance Graphs / Silhouette	83
2.4.2.4 Range Graph	83
2.4.2.5 Z-Curve	83
➤ Short Question and Answers	84 - 89
➤ Choose the Correct Answer	90 - 90
➤ Fill in the blanks	91 - 92
UNIT - III	
3.1 Measures of Central Tendency	93
3.1.1 Introduction	93
3.1.2 Significance	94
3.1.3 Types of Averages	95
3.2 Arithmetic Mean	96
3.2.1 Weighted Arithmetic Mean	103
3.2.2 Combined Mean	104
3.3 Geometric Mean	110
3.3.1 Weighted Geometric Mean	114
3.4 Harmonic Mean	118
3.4.1 Weighted Harmonic Mean	119
3.5 Mode	123
3.6 Median	131
3.6.1 Quartiles and Percentiles	140
3.7 Relationship among Mean, Median and Mode	144
➤ Exercise Problems	146 - 147
➤ Short Question and Answers	148 - 150
➤ Choose the Correct Answer	151 - 153
➤ Fill in the blanks	154 - 154

Topic	Page No.
UNIT - IV	
4.1 Measures of Dispersion	155
4.1.1 Significance	155
4.1.2 Characteristics	156
4.2 Types of Measures of Dispersion	156
4.2.1 Absolute and Relative Measures	156
4.3 Range	156
4.4 Quartile Deviation	160
4.5 Mean Deviation	165
4.6 Standard Deviation	171
4.6.1 Coefficient of Variation	176
4.7 Measures of Skewness	183
4.7.1 Karl Pearson's Coefficient of Skew-ness - Bowley's Coefficient of Skew-ness - Kelly's Measure of Skewness	184
4.8 Kurtosis	195
4.8.1 Mesokurtosis, Platy Kurtosis and Leptokurtosis	195
➤ Exercise Problems	197 - 198
➤ Short Question and Answers	199 - 201
➤ Choose the Correct Answer	202 - 203
➤ Fill in the blanks	204 - 204
UNIT - V	
5.1 Correlation	205
5.1.1 Meaning	205
5.1.2 Types of Correlation	207
5.2 Correlation and Causation	210
5.3 Methods of Correlation	211
5.3.1 Scatter Diagram	211
5.3.2 Karl Person's Coefficient of Correlation	214
5.3.3 Probable Error and Interpretation of Coefficient of Correlation	216
5.4 Rank Correlation	222

Topic	Page No.
5.5 Concurrent Deviation Method	228
➤ Exercise Problems	230 - 230
➤ Short Question and Answers	231 - 234
➤ Choose the Correct Answer	235 - 235
➤ Fill in the blanks	236 - 236

Frequently Asked & Important Questions

UNIT - I

1. Define the term statistics in both Plural and Singular Sense.

Ans : (July-21, June-19, May-18, Dec.-18, Dec.-17, Imp.)

Refer Unit-I, Q.No. 2.

2. What are the characteristics of statistics?

Ans : (July-21, May-18, Imp.)

Refer Unit-I, Q.No. 3.

3. Explain the importance of statistics in economic analysis and planning.

Ans : (Dec.-18, Dec.-17, Imp.)

Refer Unit-I, Q.No. 4.

4. Explain the various functions of statistics.

Ans : (June-19, May-18, Imp.)

Refer Unit-I, Q.No. 6.

5. List out the limitations of statistics.

Ans : (June-19, Dec.-18, Dec.-17, Imp.)

Refer Unit-I, Q.No. 7.

6. What are the various stages involved in Planning of Statistical Investigation?

Ans : (Dec.-19, Dec.-18, Imp.)

Refer Unit-I, Q.No. 10.

7. Explain the various methods of collecting primary data.

Ans : (July-21, May-18, Imp.)

Refer Unit-I, Q.No. 16.

8. What are the differences between Primary and Secondary Data ?

Ans : (Dec.-18)

Refer Unit-I, Q.No. 23.

UNIT - II

1. Define Diagram. State the importance of diagrams.

Ans : (July-21, Imp.)

Refer Unit-II, Q.No. 1.

2. The following table shows the results of B.Com. students of a university for the last three years. Represent the same by a multiple bar diagram.

Year	Exam Result of B.Com Students			
	I Division	II Division	III Division	Failed
2014	50	150	250	150
2015	60	200	300	140
2016	50	250	350	150

Sol : (Dec.-17, Imp.)

Refer Unit-II, Prob. 1.

3. Represent the following data related to the monthly expenditure of a family through a Pie-Diagram.

Item	Food	Clothing	Rent	Education	Others
Monthly Expenditure in Rs.	4,800	4,000	7,200	3,600	2,400

Sol : (June-19)

Refer Unit-II, Prob. 6.

4. State the general rules for constructing graph.

Ans : (Imp.)

Refer Unit-II, Q.No. 17.

5. Distinguish between Diagrams and Graphs.

Ans : (May-18, Imp.)

Refer Unit-II, Q.No. 19.

6. How is a Histogram different from a Bar Diagram?

Ans : (June-19, Imp.)

Refer Unit-II, Q.No. 27.

UNIT - III

1. What are the various measures of central tendency?

Sol : (July-21)

Refer Unit-III, Q.No. 6.

2. The mean of the following series is ` 115.86. Find the missing figures.

House Rent (`)	110	112	113	117	7	125	128	130
Number of Houses	25	27	13	15	14	8	6	2

Sol : (Dec.-17, Imp.)

Refer Unit-III, Q.No. 7.

3. Calculate geometric mean for the following data :

1111, 111, 11, 1.1, 0.11, 0.011

Sol : (May-18, Dec.19)

Refer Unit-III, Prob. 12.

4. 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 : (Dec.-17, Imp.)

Refer Unit-III, Prob. 26.

5. Calculate median from the following data

Marks above	20	30	40	50	60	70	80	90
Frequency	200	195	177	157	142	124	84	16

Sol : (Dec.-17)

Refer Unit-III, Prob. 35.

6. Calculate Median from the following data.

Marks	0 - 9	10 - 19	20 - 29	30 - 39	40 - 49	50 - 59	60 - 69
No. of Students	6	10	20	25	16	12	5

Sol : (June-19, Imp.)

Refer Unit-III, Prob. 37.

UNIT - IV

1. 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 : (May-18, Imp.)

Refer Unit-IV, Prob. 3.

2. 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 : (Dec.-17, Imp.)

Refer Unit-IV, Prob. 10.

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

Ans : (Imp.)

Refer Unit-IV, Q.No. 15.

4. 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

Sol : (Dec.-18, Imp.)

Refer Unit-IV, Prob. 15.

5. 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 : (June-19)

Refer Unit-IV, Prob. 23.

6. 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

Sol : (May-18, Imp.)

Refer Unit-IV, Q.No. 29.

UNIT - V

1. What do you mean by correlation? Explain various types of correlation.

Ans : (Dec.-18, Dec.-17)

Refer Unit-V, Q.No. 1, 5.

2. What is Karl Pearson's Coefficient of Correlation? Explain properties of Coefficient of Correlation.

Ans : (Imp.)

Refer Unit-V, Q.No. 10.

3. Explain briefly about probable error.

Ans : (May-18, Imp.)

Refer Unit-V, Q.No. 13.

4. Find the coefficient of correlation from the following data :

X	46	54	56	56	58	60	62
Y	36	40	44	54	42	58	54

Sol : (May-18)

Refer Unit-V, Prob. 1.

5. What is Rank correlation ? State the properties rank correlation.

Ans : (Imp.)

Refer Unit-V, Q.No. 15.

6. Calculate the coefficient of correlation from the following data by the spearman's Rank Differences method :

Prices of Tea (₹)	75	88	95	70	60	80	81	50
Prices of Coffee(₹)	120	134	150	115	110	140	142	100

Sol : (Dec.-18)

Refer Unit-V, Prob. 10.

UNIT I

INTRODUCTION:

Origin and Development of Statistics - Definition - Importance and Scope - Limitations of Statistics - Distrust of Statistics.

Statistical Investigation: Planning of statistical investigation - Census and Sampling methods - Collection of primary and secondary data - Statistical errors and approximation - classification and Tabulation of data - Frequency distribution

1.1 ORIGIN AND DEVELOPMENT OF STATISTICS

Q1. Explain the Origin and Development of Statistics.

(OR)

State the evolution of statistics.

Ans :

Introduction

The word 'statistic' has been derived from the Latin word 'status' or the Italian word 'statistic' or the German word 'statistic' or the French word 'statistique' all of which mean a political state'. The 'State' collected data pertaining to population and their wealth, in order to plan new taxes and to fund wars. Hence, statistics was also known as 'science of state craft'.

Evolution

Statistics is said to be as old as recorded history. One of the earliest known statistics is the census conducted by the Emperors of Egypt in connection with construction of pyramids. Statistics have been found to be use in India even before 300 B.C. Historical evidences about the prevalence of a good system of collection of data are available in Kautilya's Arthshastra'.

Most of the data collection activity in the medieval times relates to population and output. There have been some studies on the movement of heavenly bodies by Tycho Brahe and Johannes Kepler in the sixteenth century, which led to discovery of three laws relating to movement of planets. This helped Sir Isaac Newton to formulate the theory of gravitation. The use of statistics increased in the seventeenth century, which saw the

birth of 'Vital Statistics' Captain John Graunt of London, known as the Father of Vital Statistics, introduced a systematic study of birth and death statistics. The computation of mortality tables and calculation of life expectancy at different stages led to the establishment of the life Insurance Institution in London in 1698. However, all these studies were carried out under the name of 'Political Arithmetic'.

The first study of the theory of probability was also done in the mid seventeenth century, inspired by an attempt to estimate the chances of winning (or) losing in a gamble. The study of J. Bernoulli, which contained and 'Law of Large numbers', was published in 1713. De-Moivre published his famous "Doctrine of Chance" in 1718 and also discovered the normal probability curve. The term 'statistics' is said to have been used as a 'subject matter' in 1749 by Gottfried Achenwall. The modern theory of statistics is said to have been formulated by L. A.J. Quetelet who introduced the concept of 'average' and deviation from average. He discovered the principles of 'Constancy of Great Numbers' which forms the basis of sampling. There have been some great contribution to the science of statistics in the nineteenth and twentieth centuries. Karl Pearson who conceptualized 'Chi-Square test.' founded the greatest statistical laboratory in England. Sir Ronald Fisher pioneered the study of Estimation Theory' and also applied statistics to a variety of diversified fields such as genetics, biometry, psychology, agriculture and education. The application of scientific methods has considerably widened in the last 2 decades and there is hardly a field that does not make use of statistics. The use of information technology had made it easy and cost effective to use statistics.

Originally, statistics was restricted to recording and classification of data. Today, it has bloomed into a science and an analytical tool with significant applications in almost all disciplines ranging from Mathematics and Economics to areas such as Sociology, Psychology, etc.

1.1.1 Definition

Q2. Define the term statistics in both Plural and Singular Sense.

(OR)

Define statistics in singular sense.

(OR)

Define the statistics in plural sense.

(July-21, June-19, May-18,

Ans : Dec.-18, Dec.-17, Imp.)

Statistics in Singular Sense

In singular sense, statistics acts as a helpful device used for the purpose of collecting, classifying, presenting and interpreting the data. It is otherwise known as 'Analytical statistics'.

Some of the definitions of statistics are presented as follows,

- (a) **According to A.L. Bowley**, "Statistics is the science of counting".
- (b) **According to A.L. Bowley**, "Statistics may rightly be called the science of averages".
- (c) **According to Turtle**, "Statistics is a body of principles and techniques of collecting, classifying, presenting, comparing and interpreting the quantitative data".

Thus, the above definitions specify that statistics in singular sense, is a science which comprises different statistical methods used-for collection, organization, classification, presentation and interpretation of data.

Statistics in Plural Sense

In plural sense, the term "Statistics" implies systematical collection of numerical facts like statistics of population, production, price-level, national income and so on.

Definition of Statistics in Plural Sense-Numerical Data

1. **According to 'Dr. A.L. Bowley**, "statistics are numerical statements of facts in any department of enquiry placed in relation to each other".
2. **According to 'L.R. Comor'**, "Statistics are measurements, enumeration or estimates of natural or social phenomena, systematically arranged so as to exhibit their inter-relations".
3. **According to 'Prof. Yule and Prof. Kendall'**, "By statistics we mean quantitative data affected to a marked extent by multiplicity of causes".
4. **According to 'Horace Secrist'**, "By statistics we mean aggregate of facts affected to a marked extent by multiplicity of causes numerically expressed, enumerated or estimated according to a reasonable standard of accuracy, collected in a systematic manner for a predetermined purpose and placed in relation to each other".
5. **According to Webster**, "Statistics are classified facts respecting the conditions of the people in a state - especially those facts which can be stated in numbers or in tables of numbers or in any other tabular or classified arrangement."
6. **According to Achenwall**, "A collection of noteworthy facts concerning state, both historical and descriptive."
7. **According to H. Secrist**, By statistics we mean aggregate of facts affected to a marked extent by a multiplicity of causes, numerically expressed, enumerated or estimated according to a reasonable standard of accuracy, collected in a systematic manner for a predetermined purpose and placed in relation to each other."

Q3. What are the characteristics of statistics?

(OR)

Explain the characteristics of statistics.

(OR)

Enumerate the characteristics of a statistics.

Ans : (July-21, May-18, Imp.)

1. Statistics are aggregates of facts

Single and unconnected figures are not statistics, for example, if it is stated that Raj, a student, secured 50 marks, it is not statistics. However, marks secured by the students of the class would constitute statistics. A single figure relating to production income, marks height, etc. cannot be regarded as statistics but aggregates of such facts would be regarded as statistics.

2. Statistics are affected to a marked extent by multiplicity of causes

Numerical facts are affected by a multiplicity of factors. For example, the price, of a commodity is affected by number of factors such as supply, demand, imports, exports, money in circulation, competitive products in the market and so on. It is very difficult to study separately the effect of these factors on the price of the commodity. In physical sciences, it is possible to isolate the effect of various factors on a single item, but statistics are commonly used in social sciences and in social sciences it is very difficult to study the effect of any one factor separately. In statistical methods, the effects of various factors affecting a particular phenomenon are generally studied in a combined form, though attempts are also made to study the effects of different factors separately as well.

3. Statistics are numerically expressed

Only numerical data constitutes statistics. Qualitative expressions like good, bad, young, old etc. cannot be regarded as statistics. Statement like the standard of living of people in India has improved 'or production of petroleum products has increased in India do not constitute statistics. But Statements like "Rice production in India in 2003-04 is expected to be 86.4 million tones as against 72.66 million tons in 2002-03" (Source Economic Survey 2003-2004) is statistics.

4. Statistics are enumerated or estimated according to reasonable standards of Accuracy

Numerical information can be either enumerated or estimated. If they are enumerated i.e., actually counted or measured, they are supposed to be exact and accurate. If complete enumeration is not possible because of the large size of high cost, then data is estimated by using sampling technique. Estimated figures cannot be absolutely accurate and precise. The degree of accuracy depends to a large extent on the particular purpose for which information is collected and the nature of the particular problem about which the data is collected. There cannot be uniform standard of accuracy for all types of data collection. For example, while calculating the marks of students in an entrance examination or the number of votes received by a candidate in an election, it is necessary that it should be absolutely accurate, but while calculating the number of persons watching a cricket match on Television, the numbers need not be accurate.

5. Statistics are collected in a systematic manner

The data should be collected in a very systematic manner. For any socioeconomic survey, a proper schedule depending on the object of enquiry should be prepared and trained investigators should be used to collect the data. An attempt should be made to reduce personal bias to the minimum. Data collected in haphazard manner may give inaccurate results.

6. Statistics are collected for a predetermined purpose

The Purpose of collecting the data should be pre-determined. Otherwise, the data collected may not serve any purpose and may become useless. One should not waste time and money in collecting information that is irrelevant for the enquiry. For example, if the purpose of enquiry is to measure the cost of living of low-income group people, we should

collect information about the items that are generally consumed by them. Hence for such an index it is useless to collect information on items such as cars, refrigerators, cell-phones etc.

7. Statistics should be placed in relation to each other

Statistical information is collected mostly for the purpose of comparison. If the data collected cannot be compared, then much of the purpose of collection will be lost. The information collected should be homogeneous and not heterogeneous in character. Statistical data are often compared period wise or region-wise. For example, the data relating to population of a country for different years or population of different countries in some fixed period will be regarded as statistics. But data relating to the size of shoe of an individual and his intelligence quotient do not constitute statistics.

Statistics should contain the above characteristics. In the absence of such characteristics, numerical data cannot be called Statistics. Hence, "all statistics are numerical " "statements of facts, but all numerical statements of facts are not statistics."

1.1.2 Importance

Q4. Explain the importance of statistics.

(OR)

Explain the importance of statistics in economic analysis and planning.

Ans : (Dec.-18, Dec.-17, Imp.)

1. Importance to the State

We know that the subject of statistics originated for helping the ancient rulers in the assessment of their military and economic strength. Gradually its scope was enlarged to tackle other problems relating to political activities of the State. In modern era, the role of State has increased and various governments of the world also take care of the welfare of its people. Therefore, these governments require much greater

information in the form of numerical figures for the fulfillment of welfare objectives in addition to the efficient running of their administration.

In a democratic form of government, various political groups are also guided by the statistical analysis regarding their popularity in the masses. Thus, it can be said that it is impossible to think about the functioning of modern state in the absence of statistics.

2. Importance in economics

Statistics is an indispensable tool for a proper understanding of various economic problems. It also provides important guidelines for the formulation of various economic policies.

Almost every economic problem is capable of being expressed in the form of numerical figures, e.g., the output of agriculture or of industry, volume of exports and imports, prices of commodities, income of the people, distribution of land holding, etc. In each case, the data are affected by a multiplicity of factors. Further, it can be shown that the other conditions prescribed for statistical data are also satisfied. Thus, we can say that the study of various economic problems is essentially the one of a statistical nature."

3. Importance in national income accounting

The system of keeping the accounts of income and expenditure of a country is known as national income accounting. These accounts contain information on various macro-economic variables like national income, expenditure, production, savings, investments, volume of exports and imports, etc. The national income accounts of a country are very useful in having an idea about the broad features of its economy or of a particular region. The preparation of these accounts require data, regarding various variables, at the macro-level. Since such information is very difficult, if not impossible, to obtain, is often estimated by using techniques and principles of statistics.

4. Importance in planning

Planning is indispensable for achieving faster rate of growth through the best use of a nation's resources. It also requires a good deal of statistical data on various aspects of the economy. One of the aims of planning could be to achieve a specified rate of growth of the economy. Using statistical techniques, it is possible to assess the amounts of various resources available in the economy and accordingly determine whether the specified rate of growth is sustainable or not. The statistical analysis of data regarding an economy may reveal certain areas which might require special attention, e.g., a situation of growing unemployment or a situation of rising prices during past few years. Statistical techniques and principles can also guide the Government in adopting suitable policy measures to rectify such situations. In addition to this, these techniques can be used to assess various policies of the Government in the past. Thus, it is rather impossible to think of a situation where planning and evaluation of various policies can be done without the use of statistical techniques. In view of this it is sometimes said that, "Planning without statistics is a ship without rudder and compass Hence statistics is an important tool for the quantification of various planning policies.

5. Importance in business

With the increase in size of business of a firm and with the uncertainties of business because of cut throat competition, the need for statistical information and statistical analysis of various business situations has increased tremendously. Prior to this, when the size of business used to be smaller without much complexities, a single person, usually the owner or manager of a firm, used to take all decisions regarding its business. For example, he used to decide, from where the necessary raw materials and other factors of production were to be obtained, how much of output will be produced, where it will be sold, etc. This type of decision making was usually based on experience and expectations of this single man and as such had no scientific basis.

1.1.3 Scope**Q5. Explain briefly about the scope of statistics.**

Ans : (Dec.-17)

Statistics is viewed not just as a device for collecting numerical data, but as a means of sound techniques for their handling, analysis and for drawing valid inference from them. From this perspective, the scope (or) subject matter to statistics can be broadly studied under 2 hands namely.

- (A) Statistical Methods and
- (B) Applied Statistics

(A) Statistical Methods

Statistical methods are the tools that are in the hands of the statistician. They include all the general principles and techniques that are commonly used in the collection, analysis and interpretation of data. These methods are applicable to all kinds of data. The stage involved in study of any kind of data are:

- (i) Observation and Collection
- (ii) Organization
- (iii) Presentation
- (iv) Analysis and
- (v) Interpretation.

(B) Applied Statistics

Applied Statistics deals with application of Statistical methods of specific problems or concrete forms. To illustrate, if a software services firm is experiencing attrition (loss of personnel), it may be worthwhile to investigate the reasons for the same. Special techniques can be employed to understand the underlying trends. For example, a correlation analysis could probably indicate that increasing opportunities (growth in the industry) and attrition are highly correlated. All such techniques and the results obtained by employing such techniques form part of applied Statistics.

Applied Statistics can be further classified into different categories. These categories are

- (i) **Descriptive Statistics:** Descriptive Statistics deals with data that is known. They describe the main features of such data. They are 'basic' to any statistical analysis and bring out such characteristics of data which could have escaped attention of the statistician. Tabulation, Averages and trends are examples of such descriptive statistics.
- (ii) **Scientific Statistics:** Scientific statistics deals with formulation of statistical laws. These laws are based on quantitative data and generally hold true. For example, when we refer to the properties of standard normal distribution. Scientific statistics are heavily used for the purpose of business forecasting.
- (iii) **Analytical Statistics:** This includes methods such as correlation, regression, etc that help in establishing functional relationship between variables. Two or more sets of data are compared and analyzed to arrive at the relationship between them. Such relationship could be continuous or at a defined period of time.
- (iv) **Inferential Statistics:** Statistical methods that help us arrive at certain conclusion based on study of sample data are part of inferential statistics. It must be noted that no guess work is being made. The size of the sample, criteria for inclusion of a person/activity/thing into the sample, etc are all clearly defined. Based on the results of the study on the sample, inferences about the total population can be drawn.
- (v) **Inductive Statistics:** Statistical methods that help in arriving at general consensus based on a study of random observations are part of inductive statistics. Unlike inferential statistics, the sample chosen under inductive statistics may not 'representative'. Hence, the conclusions arrived at need to be double checked.

Q6. What are the functions of statistics.

(OR)

Explain the various functions of statistics.

Ans : (June-19, May-18, Imp.)

The, following are the main functions of statistics:

1. Presents facts in numerical figures

The first function of statistics is to present a given problem in terms of numerical figures. We know that the numerical presentation helps in having a better understanding of the nature of a problem. Facts expressed in words are not very useful because they are often vague and are likely to be understood differently by different people. For example, the statement that a large proportion of total work force of India is engaged in agriculture, is vague and uncertain. On the other hand, the statement that 70% of the total work force is engaged in agriculture is more specific and easier to grasp. Similarly, the statement that the annual rate of inflation in a country is 10% is more convincing than the statement that prices are rising.

2. Presents complex facts in a simplified form

Generally a problem to be investigated is represented by a large mass of numerical figures which are very difficult to understand and remember. Using various statistical methods, this large mass of data can be presented in a simplified form. This simplification is achieved by the summarization of data so that broad features of the given problem are brought into focus. Various statistical techniques such as presentation of data in the form of diagrams, graphs, frequency distributions and calculation of average, dispersion, correlation, etc., make the given data intelligible and easily understandable.

3. Studies relationship between two (or) more phenomena

Statistics can be used to investigate whether two or more phenomena are related. For example, the relationship between income

and consumption, demand and supply, etc., can be studied by measuring correlation between relevant variables. Furthermore, a given mathematical relation can also be fitted to the given data by using the technique of regression analysis.

4. Provides techniques for the comparison of phenomena

Many a times, the purpose of undertaking a statistical analysis is to compare various phenomena by computing one or more measures like mean, variance, ratios, percentages and various types of coefficients. For example, when we compute the consumer price index for a particular group of workers, then our aim could be to compare this index with that of previous year or to compare it with the consumer price index of a similar group of workers of some other city, etc. Similarly, the inequalities of income in various countries may be computed for the sake of their comparison.

5. Enlarges individual experiences

An important function of statistics is that it enlarges human experience in the solution of various problems. In the words of A.L. Bowley, "the proper function of statistics, indeed is to enlarge individual experience." Statistics is like a master key that is used to solve problems of mankind in every field. It would not be an exaggeration to say that many fields of knowledge would have remained closed to the mankind forever but for the efficient and useful techniques and methodology of the science of statistics.

6. Helps in the formulation of policies

Statistical analysis of data is the starting point in the formulation of policies in various economic, business and government activities. For example, using statistical techniques a firm can know the tastes and preferences of the consumers and decide to make its product accordingly. Similarly, the Government policies regarding taxation, prices, investments, unemployment, imports and exports, etc., are also guided by statistical studies in the relevant areas.

7. Helps in forecasting

The success of planning by the Government or of a business depends to a large extent upon the accuracy of their forecasts. Statistics provides a scientific basis for making such forecasts. Various techniques used for forecasting are time series analysis, regression analysis, etc.

8. Provides techniques for testing of a hypothesis

A hypothesis is a statement about some characteristics of a population (or universe). For example, the statement that average height of students of a college is 66 inches, is a hypothesis. Here students of the college constitute the population. It is possible to test the validity of this statement by the use of statistical techniques.

9. Provides techniques for making decisions under uncertainty

Many a times we face an uncertain situation where any one of the many alternatives may be adopted. For example, a person may face a situation of rain or no rain and he wants to decide whether to take his umbrella or not. Similarly, a businessman might face a situation of uncertain investment opportunities in which he can lose or gain. He may be interested in knowing whether to undertake a particular investment or not. The answer to such problems are provided by the statistical techniques of decision-making under uncertainty.

1.1.4 Limitations

Q7. Discuss the limitations of statistics.

(OR)

State the various limitations of statistics.

(OR)

List out the limitations of statistics.

Ans : (June-19, Dec.-18, Dec.-17, Imp)

1. Statistics deals with numerical facts only

Broadly speaking there are two types of facts:

- (a) quantitative and
- (b) qualitative.

(a) **Quantitative:** Quantitative facts are capable of being represented in the form of numerical figures and therefore, are also known as numerical facts. These facts can be analysed and interpreted with the help of statistical methods.

(b) **Qualitative:** Qualitative facts, on the other hand, represent only the qualitative characteristics like honesty, intelligence, colour of eyes, beauty, etc. and statistical methods cannot be used to study these types of characteristics. Sometimes, however, it is possible to make an indirect study of such characteristics through their conversion into numerical figures. For example, we may assign a number 0 for a male and 1 for a female, etc.

2. **Statistics deals only with groups and not with individuals**

Statistical studies are undertaken to study the characteristics of a group rather than individuals. These studies are done to compare the general behaviour of the group at different points of time or the behaviour of different groups at a particular point of time. For example, the economic performance of a country in a year is measured by its national income in that year and by comparing national income of various years, one can know whether performance of the country is improving or not. Further, by comparing national income of different countries, one can know its relative position vis-a-vis other countries.

3. **Statistical results are true only on the average**

Statistical results give the behaviour of the group on the average and these may not hold for an individual of that very group. Thus, the statement that average wages of workers of a certain factory is ` 1,500 p.m. does not necessarily mean that each worker is getting this wage. In fact, some of the workers may be getting more while others less than or equal to ` 1,500. Further, when value of a variable is estimated by using some explanatory

variable, the estimated value represents the value on the average for a particular value of the explanatory variable. In a similar way, all the laws of statistics are true only on the average.

4. **Statistical results are only approximately true**

Most of the statistical studies are based on a sample taken from the population. Under certain circumstances the estimated data are also used. Therefore, conclusions about a population based on such information are to be true only approximately.

5. **Statistical methods constitute only one set of methods to study a problem**

A given problem can often be studied in many ways. Statistical methods are used to simplify the mass of data and obtain quantitative results by its analysis. However, one should not depend entirely on statistical results. These results must invariably be supplemented by the results of alternative methods of analyzing the problem. It should be kept in mind that statistics is only a means and not an end.

6. **Statistics are liable to be misused**

Statistical data are likely to be misused to draw any type of conclusion. If the attitude of the investigator is biased towards a particular aspect of the problem, he is likely to collect only such data which give more importance to that aspect. The conclusions drawn on the basis of such information are bound to be misleading. Suppose, for example, the attitude of the Government is biased and it wants to compute a price index which should show a smaller rise of prices than the actual one. In such a situation, the Government might use only those price quotations that are obtained from markets having lower prices.

7. **Statistics must be used only by experts**

Statistics, being a technical subject, is very difficult for a common man to understand. Only the experts of statistics can use it correctly and derive right conclusions from the analysis.

1.2 DISTRUST OF STATISTICS

Q8. Define Distrust of statistics.

(OR)

What do you mean by distrust of statistics? Explain the causes and measures to restore trust in statistics.

Ans : **(May-18, Imp.)**

Although statistics plays an important role in many areas, there is still the possibility of statistics to be misused by the biased people. The use of statistics by such people leads to deceptive interpretation of a given problem. Among the common people, this makes a feeling of distrust about statistics. Many authors expressed their feeling about the distrust of statistics, which are as follows,

1. "There are three kinds of lies: lies, demand lies and statistics".
2. "Statistics can prove anything".
3. "Statistics is a rainbow of lies".
4. "Statistics are like clay of which you can make god or devil, as you please".

These statements show a feeling of distrust of statistics.

Major Causes of Distrust

The following are some of the causes of distrust for a common man:

1. Statistics is a Tough Subject

Normally, common man is not familiar with statistics. It is a technical subject by nature, hence it is very tough for a common man to use it. There is a possibility of misleading conclusions when statistics is used by inexperienced persons. As a result the common man does not trust the outcomes acquired by a statistical analysis.

2. Statistics has its Demerits

At the time of executing statistical analysis of a problem, the outcomes are again possible to be misleading if the limitations of statistics are not kept in mind. For a common man, this leads to other cause of distrust of statistics.

3. Different Outcomes are Acquired by Different People

It is possible that outcome could be different if two different persons analyse the same problem by gathering their own information.

4. Statistics is Accountable to be Misused

One more cause of distrust of statistics is that it is accountable to be misused. One can use statistical data to represent a wrong picture of a specified problem.

5. Analysis of Data Without any Reference to a Specific Problem

The outcomes are possible to be incorrect, if the collection and analysis of data are carried out without any reference to a specific problem. Hence, it is very essential to specify the problem to be investigated and the procedure of gathering data and its analysis should always be done with reference to the problem. Statistics is a technique of investigation and hence, its uses will rely on the nature of the specific problem. Therefore, it is said that, "statistics should not be used in the same way as a drunken man uses lamppost for support rather than for illumination".

Because of the above reason, people do not trust in the outcomes acquired through statistical analysis. Nevertheless, it should be remembered that statistics is merely a tool which could be used in right or wrong ways, as it is evident from the statement that, "figures won't lie but liars figure".

Hence, it is concluded that statistics is not a cause of distrust but the persons who are making the wrong use of it are liable for this distrust among the usual people. Therefore, to bring back faith in statistics and its outcomes, one must take care of the causes of statistics.

Measures to Restore Trust in Statistics

The following are the measures recommended to restore trust in statistics,

1. Statistics Should be Used Only by its Experts

W.I. King states that, "the science of statistics is a most useful servant, but only of great value

to those who understand its proper use". Therefore, it is very essential that statistics should be only used by its experts as to achieve the main cause of statistical analysis, i.e., to give exact description of a phenomenon.

2. Self-restraints of the Statistician

The person who is made responsible to collect and analyse the data should exercise self-control of the greatest order. He should be neutral while searching for the truth. Yule and Kendall states that, "Statistics is one of those sciences whose adepts must exercise the self-restraints of an artist".

3. Free Discussion

Marshall states that, "Statistical arguments are often misleading at first sight but free discussion clears away statistical fallacies". The statistician should work in an atmosphere which is free and open to criticism. Based on the free discussions, he must try to make better conclusion of his analysis.

4. Limitations of Statistics Should Always be Considered

The most important factor accountable for the distrust of statistics is the ignorance towards the limitations of statistics. Therefore, it is essential for a statistician to keep in mind the limitations of statistics while drawing the conclusions.

5. Eternal Vigilance

To ensure error free statistical results, the statistician is required to be very careful right from the stage of collection to the analysis and interpretation of data. Apart from this, he must try to check possible misuse of data. Wallis and Robert states that, "Eternal vigilance is the price of freedom from serious statistical blunders".

6. Conclusions must be Drawn Very Carefully

A statistician must study whether all aspects associated with the given problem has been considered or not, before reaching at conclusions. Croxton and Cowden states that, "A statistician is a scientist and must, therefore, consider all relevant data before coming to any conclusion.

1.3 STATISTICAL INVESTIGATION

Q9. What is Statistical Investigation ?

(OR)

Define the term Statistical Investigation.

Ans : (Dec.-18)

The search for knowledge, done by analysing numerical facts, is known as statistical investigation. The investigation of the levels of living of the inhabitants of a particular area, the investigation of relationship between rainfall and yield of a crop etc. are some examples of statistical investigation.

1.3.1 Planning of Statistical Investigation

Q10. What are the various stages involved in Planning of Statistical Investigation?

(OR)

What are the points to be considered planning a Statistical Investigation?

Ans : (Dec.-19, Dec.-18, Imp.)

In a statistical investigation, planning plays a vital role because all feasible preventive steps are extracted during planning. In planning a statistical investigation, the following preliminary steps are involved,

Step-1: Investigation Objective

There must be complete explanation regarding the objectives of the statistical investigation. For all possible investigation object must be clearly defined because it may help to remove unrelated data. However, investigation may leads to false conclusions, waste of time and money if the objectives are not defined clearly.

Step-2: Investigation Scope

The type of data, the subject matter and geographical location are some at the concepts which must be covered in the scope of investigation. The sample methods are used if the scope of the investigation is extensive.

Step-3: Sources of Statistical Data

Primary and secondary are the two sources of statistical data. The data gathered by a researcher, or investigator for the first time is referred as primary data. Whereas, the data which is already gathered by someone for other intention is referred as secondary data. However, the choice of decision, rely on the objective and scope of the investigation.

Step-4: Methods of Data Collection

Census or sample method would be used if primary data is collected. All possible item of population or universe is investigated in case of census method. Whereas, group of representative items of the population or the universe are taken and studied in case of sample method. The decision rely on the convenience of time, the degree of accuracy desired and the nature of the problem studied.

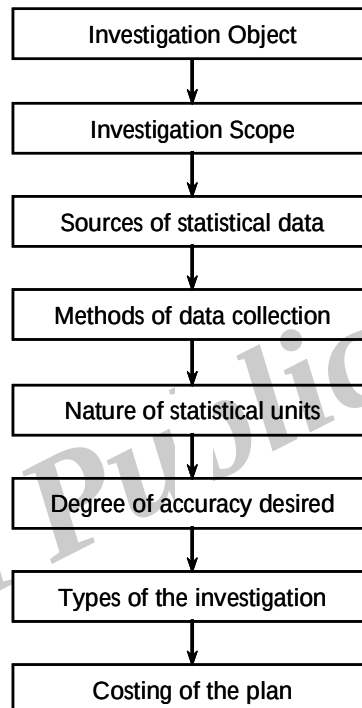


Fig.: Planning of Statistical Investigation

Step-5: Nature of Statistical Units

For the purpose of investigation there must be complete explanation of statistical units. A statistical unit is the evaluation through which the magnitude of a variable is evaluated or expressed. The following are some of the essentials that must satisfy by a statistical unit,

- (i) It must be suitable for the purpose of the investigation.
- (ii) It must be clearly defined.
- (iii) It must be definite and specific by nature.
- (iv) It must be stable and standard.
- (v) It must ensure uniformity and homogeneity.

Step-6: Degree of Accuracy Desired

Perfect accuracy is rarely feasible in any statistical investigation. However, a valid conclusions a suitable standard of accuracy is perfect enough.

Step-7: Type of the Investigation

The different kinds of statistical investigation include the following,

- (i) Official or semi-official or unofficial
- (ii) Primary or secondary
- (iii) Census or sample.
- (iv) Original or repetitive.
- (v) Direct or indirect.
- (vi) Confidential or non-confidential and
- (vii) Regular or Ad hoc.

Based on the availability of time, accuracy desired and the nature of the problem studied the of decisions choice will depends.

Step-8: Costing of the Plan

The last step in the statistical investigation is costing of the plan. The costs may be incurred on preliminary work, personal interview, editing the data, tabulating the data, analyzing the data and the final preparation of the actual report are very important to be evaluated.

Q11. Explain the various steps in Statistical Investigation

Ans : (June-19, Imp.)

- 1. Planning of Investigation :** The first stage of any statistical inquiry is the pre-planned nature, purpose and scope.
- 2. Sources of Information :** To make an inquiry successful it is essential to determine the source of data. Source of data may be internal or external, or primary or secondary.
- 3. Types of inquiry :** Types of inquiry to be conducted includes the object and scope of such an inquiry. It is something experimental, direct or indirect inquiry.
- 4. Editing of data :** After collecting the data, editing makes it comparable and meaningful. Data should be properly classified in the form of schedules and tables.

5. Units of data collection : It means units of data should be uniform throughout the study. Otherwise, it will provide mis leading conclusions.

6. Degree of Accuracy : Degree of accuracy must be kept in mind.

7. Analysis of Data : Data should be analysed, prepared by using various statistical measures like averages and dispersion.

8. Interpretation : After analysing the data properly. It must be interpreted to suggest meaningful and impartial conclusions.

1.4 CENSUS AND SAMPLING METHODS**Q12. Discuss in detail about census method with its merits and demerits.**

Ans :

Meaning

The census method deals with the investigation of the entire population.

In this case of enumeration, information is collected for each and every unit. A unit may be house hold, field, factory, shop or any other similar unit in any specified geographical area to be covered under study. The aggregate of all the units under consideration is called the 'population' or the 'universe'.

These words in statistics have meaning different from the usual ones. For example, if the statistical enquiry is related to the living conditions of textile workers, the population or the universe consists of all the workers in various textile mills in India. The method of obtaining information for each and every unit belonging to the population is defined as complete enumeration or census inquiry.

For example, the income of each individual has to be assessed for purposes of income-tax assessment, a list of all voters has to be prepared for election purposes, applications of all the persons have to be scrutinised for the purpose of recruitment.

Merits

The merits of census method are as follows,

- (i) The data collected through census enumeration are more accurate and reliable since the information is sought from each item comprised in the population.
- (ii) By making an extensive study in census enumeration, information can be obtained from each and every item and more facts relating to the problem may also be brought out. For example, population census gathers information not only in numbers, but also collects information on age, marital status, employment, education, sources of income, etc.

Demerits

The demerits of census method are as follows,

- (i) The complete enumeration of population requires time, manpower and administrative personnel.
- (ii) Apart from organizational difficulties, it is often not practical to adopt this method in many cases. Only big organizations and government having sufficiently large resources at their disposal can afford this method.
- (iii) The census method is useless in case the results are urgently required. It is likely that by the time these are available, the conditions might have changed completely.
- (iv) It may be unnecessary if a sample can yield equally reliable results.

Q13. Define sampling. Explain the advantages and disadvantages of sampling.

Ans : (Dec.-17)

Meaning

A 'sample' is a part of the 'Population'. It is subset of the entire set of units that fall within the scope of study. For example, to study the brand preference of consumers in aerated drinks (Coca Cola vs. Pepsi) we don't have to ask each and every person who consumes aerated drinks about their choice. Only a small number of persons are covered, who form the 'sample'. The number of units (or 'elements') in the sample is called sample size.

The method of obtaining data or conducting a statistical investigation by studying a sample, rather than the entire population, is 'sampling method'. The most important aspect of this method is that the sample should be 'representative' of the population.

Advantages:

1. It is very cost-effective in comparison to census method
2. It has lower turnaround time (i.e. takes less time) and will provide the latest updates, rather than outdated information.
3. Accuracy can be ensured by appropriate selection of sample and by conducting various statistical tests.
4. It offers great flexibility in organizing the statistical investigation if it requires wide coverage.

Disadvantages

1. Choice of sample is of paramount importance. Wrong sample chosen could lead to erroneous results.
2. There is lot of scope for statistical error to creep in
3. There is lot of scope for individual bias and manipulation of data.

Q14. Explain various methods of sampling.

(OR)

Discuss various sampling techniques.

Ans : (Dec.-17)

There are various methods of sampling, which can be broadly divided into two types Random Sampling and Non-Random Sampling methods. Irrespective of the method being followed. It is of paramount importance that the sample is representative of the population.

1. Random Sampling Methods

A random sample is one which is selected in such a way that every item in the population has equal chance of getting included in the sample, 'Random' does not mean 'haphazard'. For example,

if a study is being conducted to find out whether the general public has liked a particular movie, a person stands outside the movie theatre and keeps asking the opinion of the people coming out of the theatre at random. The various random sampling methods are :

i) Simple Random Sampling

Under this method, the entire 'population' is taken as one single composite unit for the purpose of selecting a sample. No attempt is made to study the 'characteristic' of the sample and make sure that the sample is 'representative' of all characters of population. Such sample can be drawn by

- (a) **Lottery:** All the 'elements' or 'constituents' of a population are numbered on identical slips of paper and put in a drum. After rotating the drum for quite some time required number (which is the decided sample size) of slips are picked up by blindfold selection. All the constituents represented by the selected slips form the random sample. This is the simplest and hence most popular method.

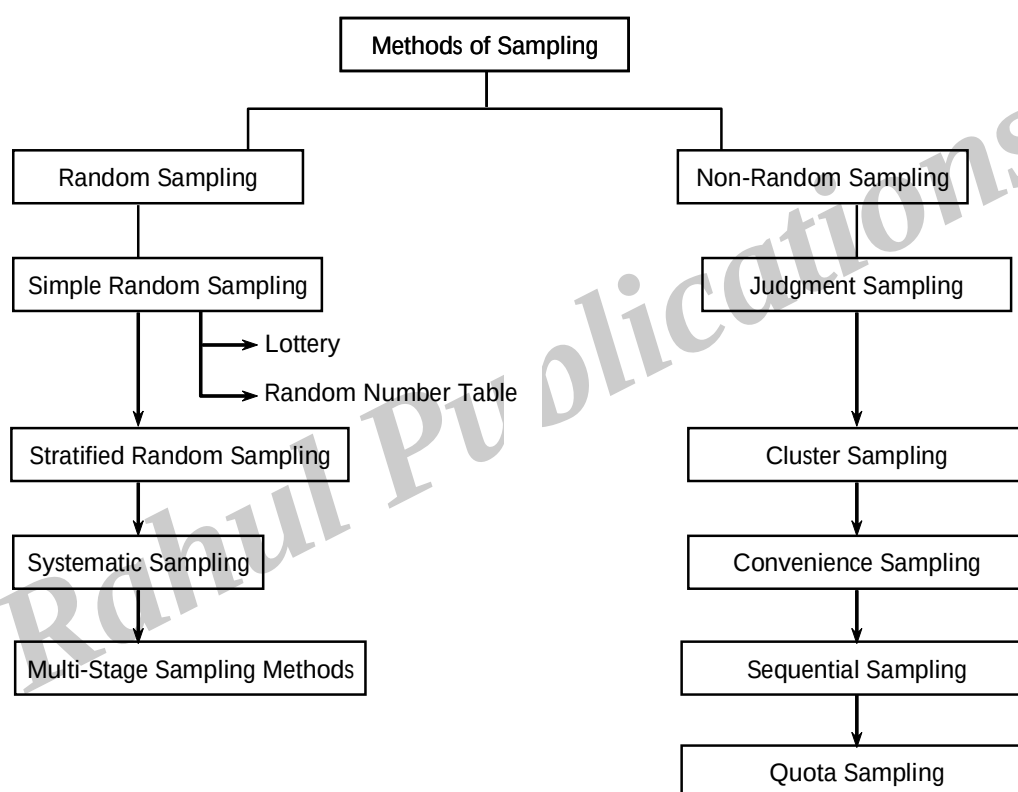


Fig.: Methods of Sampling

- (b) **Random Number Table:** We have ready made 'random number tables', that are so constructed that each of the number 0, 1, 2, 3, 4, 5, 6, 7, 8 and 9 appear with approximately the same frequency and independently of each other. In order to choose a sample, all the elements of the population are allocated numbers (0 to 9 or 00 to 99 or 000 to 999, etc depending on sample size). We then select, at random, any page of the random number table and then any row, column or diagonal of the table and pick up the numbers given in that row, column or diagonal. The elements represented by the numbers form the sample.

This method is scientific, economical and representative. However, complete list of all elements of population may not be available. Also, numbering them may become very cumbersome. However, this method is well suited where the size of the population is not very large.

ii) Stratified Random Sampling

If the 'population' consists of diverse segments that can be clearly distinguished, such population is first classified into groups each representing a diverse segment, and then simple random techniques are used within each group, to come up with a total sample that is more representative of the diverse population. These diverse segments are called a 'strata' (Stratum, if singular) and this method is called 'stratified random sampling'. The classification of population into various 'strata' can be done on the basis of its Geographic, sociological or Economic characteristics. The size of sample from each, strata is normally proportional to the total size of that strata to the total population size. However, it can also be of disproportional size.

(iii) Systematic Sampling

In case of systematic sampling, the first sample unit is -Selected at random. Thereafter, depending on sample size, the rest of the units are selected automatically by using a systematic process.

(iv) Multistage Random Sampling

Under this method, the final sample is - obtained after a drill down exercise, with random selection at each stage.

2. Non-Random Sampling Methods

Non-Random sampling methods rely on 'selection of samples based on 'choice' of the investigator, rather than 'chance'.

The various non random sampling methods are under..

(i) Judgment Sampling

Judgment sampling relies on the capability of the investigator to select the appropriate sample, rather than the use of techniques. An experienced and unbiased investigator can 'focus' on the issue and pick up the sample that meets the objectives of the study rather than live with statistical errors arising out of tools and techniques. However, an inexperienced or biased investigator can arrive at totally misleading conclusions by not following proper random sampling methods.

(ii) Cluster Sampling

In this method, the entire population is divided into some recognizable sub-groups called 'clusters'. A random sample of the clusters is drawn and then all the units belonging to the clusters are considered as part of the sample.

(iii) Convenience Sampling

Under this method, a sample is obtained by selecting "such units that are convenient to locate or contact.

(iv) Sequential Sampling

Sequential sampling is not a separate method of "drawing a sample. It is more of a process. A sample is drawn and the statistical investigation is conducted. If the result are satisfactory, nothing else is done. However, if study of the sample is inconclusive, another sample is drawn. This process may go on for any number of iterations until the investigator can arrive at a conclusion. This process is applied in Statistical Quality Control (SQC).

(v) Quota Sampling

It is a combination of stratified Sampling and Judgment sampling. In this method, the population is divided into 'quotas'. Each quota consists of units with specified characteristics. Each investigator or enumerator is told about the number of units he/she needs to select from the quota. This is normally done using judgment. This should not be confused with cluster sampling. In cluster sampling, only few out of all clusters are picked and all units belonging to selected cluster are selected. In Quota sampling all quotas are selected but sampling is applied within each quota.

**1.5 COLLECTION OF PRIMARY AND
SECONDARY DATA**

Q15. Define data. What are the different sources of collecting data?

Ans : (July-21)

Meaning

Data refers to the information which is collected with the help of research for the purpose of making appropriate decisions.

Types of Data Sources

Usually data is collected through two different sources i.e., primary and secondary and based on these sources data is classified into two types,

- I) Primary data
- II) Secondary data.

I) Primary Data

Primary data refers to the data collected specifically which is for the purpose of research problem. It is the first hand information collected by the research firm or by an external agent with the objective of solving a research problem. There are different methods of collecting primary data. Researchers can conduct experiments to gather the required information. Other methods include questionnaires, mails, interviews of individuals, families, organisations, representatives etc.

II) Secondary Data

Primary data for one party sometimes acts as the secondary data for the other party. Secondary data refers to the existing data that had been collected with an objective other than for research. It could be the data collected by the firm itself for any other purpose, or by any external party for the same or other research problem. For example, inventory records maintained by a firm as part of their routine operating function acts as the internal secondary data of the firm for evaluating the seasonal demand for its product. Alternatively, data gathered on industry demand by a marketing research firm can also be used for the same research problem.

Q16. Explain the various methods of collecting primary data.

Ans : (July-21, May-18, Imp.)

Primary data is one, which is collected by the investigator himself for the purpose of a specific inquiry or study. Such data is original in character and is generated by surveys conducted by individuals or research institutions. Primary data collection is necessary when a researcher cannot find the data needed in secondary sources.

Example

If a researcher is inserted to know the impact of afternoon meal scheme for the school children, he has to undertake a survey and collect data on the opinion of parents and children by asking relevant questions. Such a data collected for the purpose is called primary data.

Conducting primary research is a useful skill to acquire as it can greatly supplement research in secondary sources, such as journals, magazines, or books. Primary research is an excellent skill to learn as it can be useful in a variety of settings including business, personal, and academic.

Methods of Collecting Primary Data

The various methods of collecting primary data are :

- (a) Interviews
 - Personal Interview
 - Telephone Interview
- (b) Surveys
- (c) Observations
- (d) Analysis
- (e) Questionnaires
- (f) Schedules

(a) Interviews

Interviews are non-on-one or small group question and answer sessions. Interviews will provide a lot of information from a small number of people and are useful when you want to get an expert or knowledgeable opinion on a subject. The interviews can be :

- **Personal Interview :** The investigator follows a rigid procedure and seeks answers to a set of pre-conceived questions through personal interviews. This method of collecting data is usually carried out in a structured way where output depends upon the ability of the interviewer to a large extent.

- **Telephone Interviews :** This method of collecting information involves contacting the respondents on telephone itself. This is not a very widely used method but it plays an important role in industrial surveys in developed regions, particularly, when the survey has to be accomplished in a very limited time.

(b) Surveys

Surveys are a form of questioning that is more rigid than interviews and that involve larger groups of people. Surveys will provide a limited amount of information from a large group of people and are useful when you want to learn what a larger population thinks.

(c) Observations

Observations involve taking organized notes about occurrences in the world. Observations provide you insight about specific people, events, or locales and are useful when you want to learn more about an event without the biased viewpoint of an interview. The information obtained relates to what is currently happening and is not complicated by either the past behaviour or future intentions or attitudes of respondents.

This method is no doubt an expensive method and the information provided by this method is also very limited. As such this method is not suitable in inquiries where large samples are concerned.

(d) Analysis

Analysis involves collecting data and organizing it in some fashion based on criteria you develop. They are useful when you want to find some trend or pattern. A type of analysis would be to record commercials on three major television networks and analyze gender roles.

(e) Questionnaire

The questionnaire is an important tool for gathering primary data. Questionnaire are mailed to the respondents with a required to return after completing the same. It is most extensively used method in various economic

and business surveys. Questionnaire to be used must be prepared very carefully so that it may prove to be effective in collecting the relevant information. Poorly constructed questions can result in large errors and invalidate the research data, so significant effort should be put into the questionnaire.

(f) Schedules

Under this method the enumerators are appointed and given training. They are provided with schedules containing relevant questions. These enumerators go to respondents with these schedules. Data are collected by filling up the schedules by enumerators on the basis of replies given by respondents. Much depends upon the capability of enumerators so far as this method is concerned. Some occasional field checks on the work of the enumerators may ensure sincere work.

Q17. What do you mean by Mailed Questionnaire Method ? Explain the points to be kept in mind while drafting a good questionnaire.

Ans :

(June-19)

Under this method a questionnaire is prepared and sent to the informants by post. A questionnaire consists of number of questions relating to the field of enquiry. Space is provided for answers to be filled by the informants (called respondents). This is sent with a covering letter explaining in details the aims and objective of collecting information. The respondents are requested to extend their full co-operation by furnishing correct replies within a reasonable time. Respondents are taken into confidence by assuring them that the information supplied by them will be kept confidential. To ensure quick and better response, the return postage expenses are usually borne by the investigator. This method is generally used by the research scholars, non-government officials, individuals and even by government. This method is less expensive and information can be collected from a wide area. Success of the method depends upon cooperation of informants. Generally, it is found that informants adopt an attitude of indifference towards such enquiries.

There is always some uncertainty about their response. Cooperation on part of informants may be difficult to presume. The information supplied by the informants may not be correct and it is very difficult to verify its accuracy. Moreover, this method can be adopted only when informants are literate people.

Drafting a Questionnaire

Questionnaire is the major medium of communication Between Investigator and respondents. It should be drafted with utmost care and caution so that all relevant and essential information for enquiry is collected without any ambiguity or difficulty. Drafting a good questionnaire is a skilled job requiring care, efficiency and experience. No hard and fast rules can be laid down for framing a questionnaire. However, the following points may be kept in mind while framing a questionnaire.

1. **Size:** The number of questions must be restricted to the minimum possible. More the number of questions, lesser the likelihood of it being answered by the respondents.
2. **Clarity:** Questions should be clear and unambiguous. Simple language must be used. Words that can have multiple meanings or interpretations should be explained.
3. **Sequence:** The questions should be arranged in a logical sequence. This will help the respondent to understand the context and answer appropriately.
4. **Easy to Respond:** It should be easy for the respondent to answer the questions. It is preferable to provide multiple choices as answers, of which the respondent will generally choose the most appropriate answer (unless multiple ticks are allowed). Open ended questions should not require lengthy or elaborate answers.
5. **Confidentiality:** The questionnaire should make all questions that reveal the identity of the respondent (such as Name, Telephone Number, address, etc) optional. The respondents should also be assured that information provided by them would be kept confidential.

6. **Questions to be avoided:** Questions that involve prestige or status, leading questions (Questions that lead the respondent to an answer e.g Is Hero Honda your favorite motor cycle brand), questions of sensitive or personal nature, questions whose answers require detailed calculations, subjective & irrelevant questions need to be avoided.
7. **Look and Feel:** Good quality paper and printer must be used. The layout and presentation should be made interesting.
8. **Post Survey Usage:** Mode of Tabulation and Analysis of data collected should also be kept in mind at the time of design of questionnaire.
9. **Cheeks and Accuracy:** The questionnaire must contain internal checks to ensure accuracy of data being collected.

Q18. What are the advantages and disadvantages of primary data.

Ans :

Advantages

1. **Reliability**
The information collected for primary data is more reliable than those collected from the secondary data because this information is collected directly from the respondents.
2. **Availability of a Wide Range of Techniques**
There are lot of techniques that can be employed, which means that all information necessary can be obtained by using the appropriate techniques, enabling all areas of the research topic to be answered and investigating thoroughly and effectively.
3. **Addresses Specific Research Issues**
The organisation asking for the research has the complete control on the process and the research is streamlined as far as its objectives and scope is concerned. Researching company can be asked to concentrate their efforts to find data regarding specific market rather than concentration on mass market. Primary research is designed to collect the information the marketer wants to know, and report it in ways that benefit the marketer.

4. Greater Control

Not only does primary research enable the marketer to focus on specific issues, it also enables the marketer to have a higher level of control over how the information is collected. In this way the marketer can decide on such issues as size of project (e.g., how many responses), location of research (e.g., geographic area) and time frame for completing the project.

5. Efficient Spending for Information

Unlike secondary research, where the marketer may spend for information that is not needed, primary data collection focuses on issues specific to the researcher and improves the chances that research funds will be spent efficiently.

6. Proprietary Information

Information collected by the marketer using primary research is their own and is generally not shared with others. Thus, information can be kept hidden from competitors, and potentially offer an "information advantage" to the company that undertook the primary research.

Disadvantages**1. Cost**

Compared to secondary research, primary data may be very expensive since there is a great deal of marketer involvement and the expense in preparing and carrying-out research can be high. Skilful persons are required, which is also not an economic process, it is costly.

2. Time Consuming

To be done correctly primary data collection requires the development and execution of a research plan. Going from the start-point of deciding to undertake a research project to the end-point of having results is often much longer than the time it takes to acquire secondary data.

3. Not Always Feasible

Some research projects, while potentially offering information that could prove quite valuable, are not within the reach of a marketer. Many are just too large to be carried-out by all but the largest companies, and some are not feasible at all.

4. Large Volume of Data

Since the data collected by primary methods remains in a very large amount so it becomes very complicated to handle and maintain all the data. Large volumes of data also create difficulty in the data processing.

5. Reluctancy of Respondents

In many cases, the respondents remain reluctant to give the answers of the researchers' questions. Sometimes they give such answers which create business in the research.

Q19. Explain the various ways Editing of primary data.

Ans :

1. Editing for Completeness

The editor should see that each schedule and questionnaire is complete in all respects, i.e., answer to each and every question has been furnished. If some questions have not been answered and those questions are of vital importance the informants should be contacted again either personally or through correspondence. It may happen that in spite of best efforts a few questions remain unanswered. In such questions, the editor should mark "No answer" in the space provided for answers and if the questions are of vital importance then the schedule or questionnaire should be dropped.

2. Editing for Consistency

While editing the data for consistency, the editor should see that the answers to questions are not contradictory in nature. If there are mutually contradictory answers, he should try to obtain the correct answers either by referring back the questionnaire or by

contacting, wherever possible, the informant in person. For example, if amongst others, two questions in questionnaire are: (a) Are you married? (b) State the number of children you have, and the reply to the former question is 'no' and to the latter 'three', then there is contradiction and it should be clarified.

3. Editing for Accuracy

The reliability of conclusions depends basically on the correctness of information. If the information supplied is wrong, conclusions can never be valid. It is, therefore, necessary for the editor to see that the information is accurate in all respects. However, this is one of the most difficult tasks of the editor. If the inaccuracy is due to arithmetical errors, it can be easily detected and corrected. But if the cause of inaccuracy is faulty information supplied, it may be difficult to verify it, e.g., information relating to income, age, etc.

4. Editing for homogeneity

By homogeneity we mean the condition in which all the questions have been understood in the same sense. The editor must check all the questions for uniform interpretation. For example, as to the question of income, if some informants have given monthly income, others annual income and still others weekly income or even daily income, no comparison can be made. Similarly, if some persons have given the basic income whereas others the total income, no comparison is possible. The editor should check up that the information supplied by the various people is homogeneous and uniform.

Q20. Explain the various methods of collecting secondary data.

Ans :

Secondary data are those data which have been already collected and analysed by some earlier agency for its own use; and later the same data are used by a different agency.

Secondary data is data that is neither collected directly by the user nor specifically for the user, often under conditions not known to the user like

Government reports. Secondary information has already been collected for some other purposes. It may be available from internal sources, or may have been collected and published by another organisation. Secondary data is cheaper and more quickly available than primary data, but needs processing before it is useful. Since the data has been collected for another purpose by somebody else, it may not be fully useful, the context could have changed or data could have been doctored.

According to W.A. Neiswanger, "A secondary source is a publication, reporting the data which have been gathered by other authorities and for which others are responsible".

The purpose of collecting secondary data is to make some changes, or to review the needs, to make solutions based on normal database management. To get some new ideas or it was also for the purpose of time saving. If researcher has short time and researcher has to complete an objective, secondary data in this regard is the best way to save time and complete task.

Tools and Techniques of Collecting Secondary Data

In most of the studies the investigator finds it impracticable to collect first-hand information on all related issues and as such he makes use of the data collected by others. There is a vast amount of published information from which statistical studies may be made and fresh statistics are constantly in a state of production. The sources of secondary data can broadly be classified under two heads :

1. Internal Secondary Data

Data that originate within the firm for which the research is being conducted are internal data. They may be adapted for the marketing research purposes. They may be formal data and informal data. Formal data are available on a regularly scheduled basis, such as monthly, quarterly or annually in a form that allows comparisons through time. Informal data reports basic marketing knowledge and are available on a non-recurring basis.

Sales analysis and invoicing are considered important sources of internal secondary data.

i) Sales Analysis

Sales analysis is an important tool of marketing research. It is the first step in the marketing research program and acts as a basis for the development of further marketing research. It reveals the current operating problems in the marketing area where the scope for marketing research can be adequately explored in smaller organisations, sales analysis is an important source of marketing information. It provides a major share of the factors for marketing research.

ii) Invoice Analysis

Company invoices have been a very useful source of information. A copy of an invoice is preserved and information from it may be punched, tabulated, processed and summarized to provide suitable information to the researcher. The invoice data may be classified according to customer, nature of product, region and area. The invoice record may be of immense use provided it has been used with precaution and scientifically.

iii) Accounting Records

The basis for accounting records concerned with sales is the sales invoice. The usual sales invoice has a sizable amount of information on it, which generally includes name of customer, location of customer, items ordered, quantities ordered, quantities shipped, dollar extensions, back orders, discounts allowed and date. In addition, the invoice often contains information on sales territory, sales representative and warehouse of shipment. This information, when supplemented by data on costs and industry and product classification, as well as from sales calls, provides the basis for a comprehensive analysis of sales by product, customer, industry, geographic area, sales territory, and sales representative, as well as the profitability of

each sales category. Unfortunately, most firms' accounting systems are designed primarily for tax reasons rather than for decision support.

2. External Secondary Data

The second form of secondary data is external sources which are generally published and are available in different forms and from different sources. Although external secondary data may be obtained from different sources, some of the sources are given here.

i) Libraries

Researchers first attend libraries to find out relevant data pertaining to research. They provide many sources where suitable data may be obtained. Public libraries, colleges and University libraries contain a large amount of business information, which provides sources of other data. Management books, theses, management journals and other publications can be consulted in these libraries. Management institutes, research institutes, banks, insurance companies, public utility companies and manufacturing units have maintained adequate libraries.

ii) Literature

A great amount of secondary data is available from literature, particularly on marketing subjects. With the development of marketing researches in different countries, new and interesting facts are coming into the picture, which are available in various publications. Consultations of this literature may provide proper guidance pertaining to publication, which can be used from time to time.

iii) Periodicals

Business periodicals published fortnightly, monthly, quarterly, semi annually and annually are often consulted by the marketing executives and researchers to plan and design their marketing research. Periodicals on

economics, finance, trade, transport, industry, labour and management are being prepared by the Government as well as by the non-government agencies. Journals of the marketing association, management association, research agencies, advertising agencies and other related periodicals are becoming very common in India and abroad.

iv) Census and Registration Data

Census and registration data have become very comprehensive sources of marketing research. Previously, these were concentrated only to one population census, but it now extends too many areas. Census and registration data includes Census of Population, Census of Agriculture, Census of Cattle, Census of Trade, Census of Transport, Census of Industry, Census of Banking and Finance etc.

v) Trade Associations

Trade associations may be an excellent source of data pertaining to an industry. The trade association of one industry may exchange data with the trade association of another industry, and within one industry a firm may exchange data with another firm with the help of trade association of the industry.

vi) Government Departments

Different government departments have different data, which are not available in libraries. But these are very useful for understanding various aspects of the economy. The researchers can utilize them for the purposes of their research. Information and data pertaining to agriculture, industry, trade, transport, banking and finance can be obtained from the respective ministries of the Government of India.

vii) Private Sources

Private sources include varied sources available in the form of books, monographs,

bulletins, journals, commercial reports and so on. They are priced and publicly circulated. Some of the sources include extensive original research and some summarise the research findings of other person. Many of them are statements of facts and opinions. The All India Management Association, the Indian Marketing Association, Commerce Pvt. Ltd., Capital, the Economic Times and Financial Express etc. are important private institutions which supply suitable information and data to the public in the form of journals, books and newspapers.

viii) Commercial Data

There are several institutions and companies, which purchase and sell marketing information and data. Some of these companies are solely engaged in marketing research. They collect information and data directly from field surveys. Some such companies collect and process the secondary data and supply them to their subscribers.

ix) Financial Data

The financial data of reputed concerns are available in several magazines, newspapers, journals and in summary of statistics. The Directorate of Income Tax publishes information pertaining to taxes and income ranges. Such information and data are useful to forecast the market potential of a particular product. Private institutions such as the Economic Times of India, Commerce Private Ltd. etc. are publishing assets and investment-wise data of several large companies. The market researchers are indirectly benefited by such data and information.

x) International Organisations

International organisations such as the International Monetary Fund, the World Bank, the United Nations Organisation, the Asian Bank, the African Bank, Foreign embassies etc. publish several useful statistics,

which can be used by researchers. The statistics may relate to the population problem, trade, institutions, culture, agriculture, regional festivals, superstitions, education, consumption, transportation, forestry, manufacturing and so on. There are several other publications such as the World Almanac, Thomas Register etc., which publish much useful information for marketing researchers.

xi) References and Bibliography

In every publication, the researcher can find references and a bibliography which can be very good sources of information of marketing research. The researcher can consult them for further information and data.

xii) Volumes of Statistics

There are several private and public organisations, which prepare a summary of statistics. In India, the Indian Statistical Institute publishes the Statistical Abstract. Commerce Pvt. Ltd., the Times of India Ltd. and the Financial Express compile directories of different subjects. The Government of India publishes the economic survey of India wherein statistics relating to every field of economic activities are compiled in a suitable form.

xiii) Advertising Agencies

Advertising agencies have proved to be very useful sources of marketing research. Recently, a large number of agencies have come into the findings of the advertising researches for their clients. Advertising agencies sometimes, publish reviews, resumes and tests of marketing researches. The consumers' behaviour, consumption pattern and demographic features are generally revealed by these agencies.

xiv) Other Sources

There are several other sources of marketing researches. Individuals conduct their own researches, which may be purchased by other institutions. Marketing associations, management associations and individual business houses have been conducting marketing researches for other researches. There are a large number of researches organisations in foreign countries, which are selling their research findings to organisations requiring knowing the outcome of their researches.

Q21. What are the advantages and disadvantages of Secondary Data?

Ans :

Advantages

1. Economy

Such data is cheaper. The amount of money spent in acquiring secondary data is generally less than that needed to obtain primary data. The various secondary data from libraries can be obtained at no cost.

2. Quickness

Most of these data are 'instant' since they already exist and merely need to be discovered. Thus, the time in collecting secondary data is largely search time and usually requires few hours or days.

3. Quality

An individual investigator cannot match the quality or size of the firms that obtain much of the existing secondary data. The information is gathered by trained personnel specialised in data collection. Also many organisations may not release their data to individual researcher but may give it to firms.

4. No Need of Measuring Instruments

When information is gathered from secondary sources, there is a problem in

designing information gathering instruments as information is already collected by someone else. The only problem is to locate the appropriate source and method of recording desired information.

5. Availability

Secondary data is sometimes available even in those cases where primary investigations are not possible.

6. Bases for Comparison

Secondary data is useful in the case of exploratory researches as they provide increased understanding of the problem.

7. Useful in Exploratory Research

Secondary data act as a basis for comparison after primary data is collected.

8. Generates Feasible Alternatives

Secondary data are very useful in generating viable alternatives to solve problems. The multiplicity of data sources, research approaches and managerial styles usually lead to a number of possibilities which should be examined by the researcher.

Disadvantages

It is however difficult to find secondary data that exactly fits into the needs of some specific research investigation. The problems experienced in respect of secondary sources can be in terms of :

1. Relevance

The data may not fit into the needs of investigation. There may be difference in the units of measurement; there may be surrogated data; discrepancy of classes and data may pertain to some other period of time.

2. Accuracy

It is observed that it is rather difficult to measure the degree of approximations used in the collection of information as well as the competence of the investigator in motivating the persons to supply the desired information.

3. Existence of Obsolete Information

Information may be outdated or obsolete.

4. Non-Disclosure of Research Findings

All the findings of a research study may not be made public.

5. Seldom Catering to the Need

The available data may not suit the current purpose of research, due to incompleteness, generalities and so on.

6. Other Limitations

- i) There may be difficulties in the identification of the source.
- ii) Errors may be there in recording or transferring information from secondary sources.
- iii) The facilities or capabilities of the agency that originally collect the data might be questionable.

Q22. Explain the various ways editing of secondary data.

Ans :

Since secondary data have already been obtained it is highly desirable that a proper scrutiny of such data is made before they are used by the investigator. In fact, the user has to be extra-cautious while using secondary data. In this context, Prof. Bowley rightly points out that secondary data should not be accepted at their face value." The reason is that such data may be erroneous in many respects due to bias, inadequate size of the sample, substitution, errors of definition, arithmetical errors, etc. Even if there is no error such data may not be suitable and adequate for the purpose of the enquiry. Hence Before using such data, the investigator should consider the following aspects:

Whether the data are suitable for the Purpose of Investigation in View

Before using secondary data the investigator must ensure that the data are suitable for the

purpose of the equity. The suitability of data can be judged in the light of the nature and scope of investigation. For example, if the object of enquiry is to study the wage levels including allowances of workers and the data relate to basic wages alone, such data would not be suitable for the immediate purpose. It may be difficult to find data which exactly fit the needs of the present project.

Quite often secondary data do not satisfy immediate needs because they have been compiled for other purpose. Even when directly pertinent to the subject under study, secondary data may be just enough off the point to make them of little or no use. The value of secondary data is frequently impaired by:

- (i) Variation in the units of measurement: consumer income, for example, may be measured by individual, family household, spending units or tax return.
- (ii) Definition of classes may be different: for example, definition of literate, educated, poor may vary from researcher to researcher.
- (iii) Variation in the date/period to which the data related: the data available may relate to a different time period and not serve the purpose of researcher. Data published to promote the interests of a particular group whether it is political, social or commercial are suspect.

Whether the Data are Adequate for the Investigation If it is found that the data are suitable for the purpose of investigation, they should be tested for adequacy. Adequacy of the data is to be judged in the light of the requirements of the survey and the geographical area covered by the available data. For example, in the illustration given above, if our object is to study the wage rates of the workers in sugar industry in India and if the available data cover only the State of U.P., it would not serve the purpose. The question of adequacy may also be considered in the light of the time period for which the data are available. For example, for studying trend of prices we may use data for the last 8-10 years but from the source known to us data may be available for the last 2-3 years only which would not serve the purpose.

Whether the Data are Reliable It is very difficult to find out whether the secondary data are reliable or not. The following tests, if applied, may be helpful to determine how far the given data are reliable:

- (i) Which specific method of data collection was used? If a source fails to give a detailed description of its method of data collection, researchers should be hesitant about using the information provided. When the methodology is described, researchers should subject it to a painstaking examination.

Data published to promote the interests of a particular group whether it is political, commercial or social are suspect.

- (ii) Was the collecting agency unbiased or did it "have an axe to grind"?
- (iii) If the enumeration was based on a sample, was the sample representative?
- (vi) Were the enumerators capable and properly trained? Incompetent or poorly trained enumerators cannot be depended upon to produce useful result.
- (v) Was there a proper check on the accuracy of field work?
- (vi) Was the editing, tabulating and analysis carefully and conscientiously done? Carelessness in either one or more of these functions can render of little value the findings of an otherwise valuable study.
- (vii) What degrees of accuracy was desired by the compiler? How far was it achieved?

Q23. What are the differences between Primary and Secondary Data ?

Ans :

(Dec.-18)

Basis of Difference	Primary Data	Secondary Data
1) Meaning	Primary data is the one, which is collected by the investigator himself for the purpose of a specific inquiry or study.	Secondary data are those data which have been already collected and analysed by some earlier agency for its own use; and later the same data are used by a different agency
2) Cost	The cost of obtaining primary data is typically more.	Secondary data is typically available for free or for the subscription fee to the database magazine or journal.
3) Sources	A primary data source is a publication in which the data are published by the same authority which gathered and analysed them.	A secondary data source is a publication, reporting the data which have been gathered by other authorities and for which others are responsible.
4) Methods	It is normally collected through experiments, surveys, questionnaire focus groups, interviews.	The method or way of collecting secondary data includes books, journals, census data biographies, articles and databases.
5) Reliability	The information collected for primary data is more reliable than those collected from the secondary data.	Whereas secondary data are less reliable because these information are not collected for that particular purpose.
6) Scientific Method	Primary data follows the scientific method. A hypothesis is formed, data is collected from an experiment based on the hypothesis and the hypothesis is proven correct or not.	Secondary data does not start with a hypothesis as the data is already collected. Patterns and insights are found within the secondary data and then the observation on that data is made.
7) Precaution	No extra precautions are required in primary data.	Secondary data need more care and attention.
8) Form of Data	Primary data are in the shape of raw material.	Secondary data are usually in the shape of readymade products.
9) Accuracy	Primary research is tailored specifically for the project and tends to be more accurate. Primary data is customized.	While secondary data can provide plenty of information, it is less accurate because the data collected was not collected specifically for the questions. Secondary data is not customized.
10) Example	One's own questionnaire.	Data from a magazine. Journal, etc.

1.6 STATISTICAL ERRORS AND APPROXIMATION

Q24. Define Statistical Error. What are the various sources of Statistical Errors.

Ans :

Meaning

In statistics, the word “error” is used in a special sense. It is voluntary in nature and can be corrected. It is not considered as mistake. Mistakes in statistical calculations takes place because of defects in the working of human minds and also due to some faults in machines. But in statistics, ‘error’ means the difference between the true value and the estimated value of an item.

Definition

According to Boddington ‘statistical error’ means “the divergence between the actual number and the estimate which is made either by approximation or by any other method”.

Sources

There are several factors responsible for the occurrence of statistical errors. The important sources of statistical errors are as follows,

1. Errors of Origin

These errors occur because of the improper definition of statistical units, untrained staff, the personal bias and faulty questionnaires.

2. Errors of Inadequacy

These errors occur because of insufficient sample data or smallness of sample or in complete information. The errors may continue to occur if the sample has items which are not representative of universe.

3. Errors of Manipulation

These errors occur at the time of counting, weighing, measurement and approximation.

1.7 CLASSIFICATION AND TABULATION OF DATA

Q25. Define classification. What are the objectives of classification?

Ans :

Meaning

The largest amount of data and the greatest amount of detail may not convey the most useful

information for decision making. An important aspect of Statistics is to organize and present data so as to convey critical information quickly.

Classification means grouping of a whole into different groups or classes. However, each of these groups should have a common characteristic.

For example, all living things can be classified into plants and animals. Animals can be classified into Amphibians, Reptiles, and Mammals etc.

Definition

According to Horace Secrets, “Classification is the process of arranging data into sequences and groups according to their common characteristics, or separating them into different but related parts”.

Objectives

1. To condense the mass of data into a form that is easily understandable.
2. To enable the user to get a proper grasp of the significance of the information contained in the data.
3. To sort data in such a manner that irrelevant details are ignored and relevant details only are considered.
4. To present the data in tabular form and enable the user to apply statistical tools and techniques on the same
5. To enable interpretation, analysis and generalization of data
6. To facilitate comparative study of variables.
7. To establish relationships between variable on study of various characteristics of data
8. To bring out similarities and disparities in data collected.
9. To provide an executive summary of the data collected at a glance.

Q26. What are the different types of classifications?

Ans :

The nature of classification depends upon the purpose and objective of investigation. The following are some very common types of classification :

1. Geographical (or spatial) classification

When the data are classified according to geographical location or region, it is called a geographical classification.

2. Chronological classification

When the data are classified on the basis of its time of occurrence, it is called a chronological classification. Various time series such as National Income figures (annual), annual output of wheat, monthly expenditure of a household, daily consumption of milk, etc., are some examples of chronological classification.

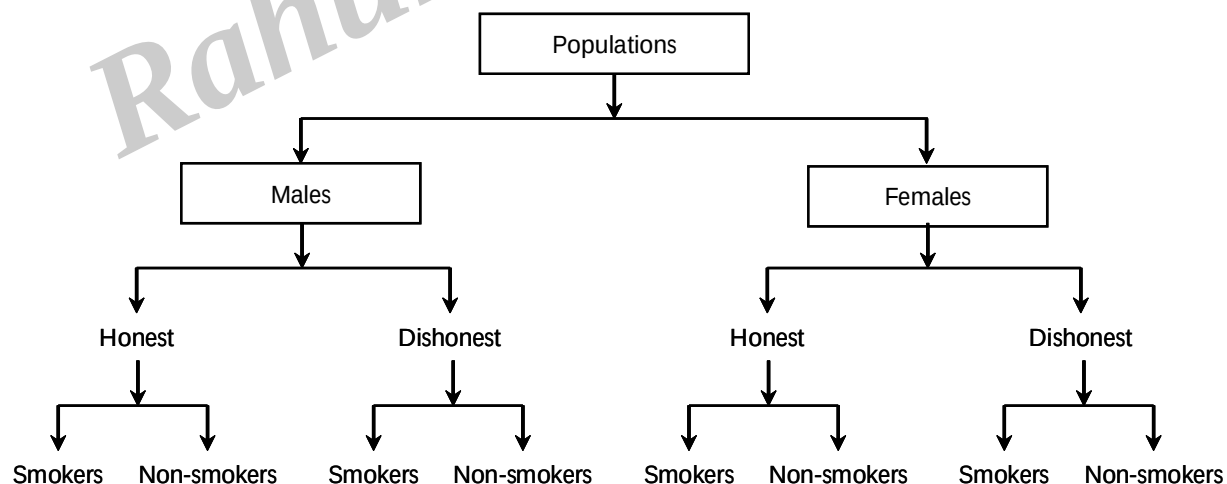
3. Conditional classification

When the data are classified according to certain conditions, other than geographical or chronological, it is called a conditional classification.

4. Qualitative classification or classification according to attributes

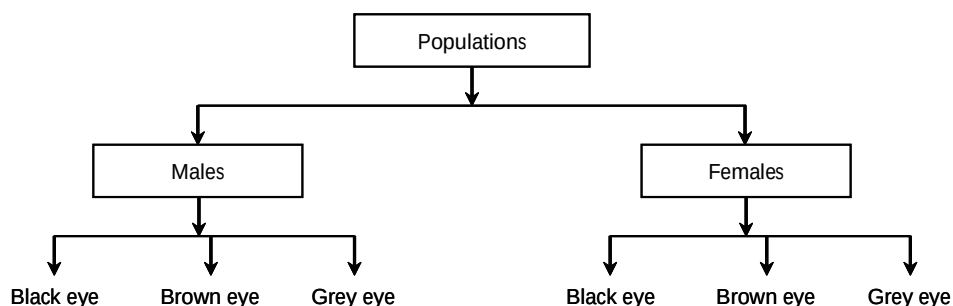
When the characteristics of the data are non-measurable, it is called a qualitative data. The examples of non-measurable characteristics are sex of a person, marital status, colour, honesty, intelligence, etc. These characteristics are also known as attributes. When qualitative data are given, various items can be classified into two or more groups according to a characteristic. If the data are classified only into two categories according to the presence or absence of an attribute, the classification is termed as dichotomous or two-fold classification. On the other hand, if the data are classified into more than two categories according to an attribute, it is called a manifold classification. For example, classification of various students of a college according to the colour of their eyes like black, brown, grey, blue, etc. The conditional classification, given above, is also an example of a manifold classification.

If the classification is done according to a single attribute, it is known as a one-way- classification. On the other hand, the classification done according to two or more attributes is known as a two-way or multi-way classification respectively. The example of a three-way classification, where population is dichotomised according to each attribute, sex, honesty and smoking habit, is shown as following:



We note that there will be eight sub-groups of individuals like (male, honest, smokers), (male, honest, non-smokers), etc.

In the classification, given above, the population is dichotomised with respect to each of the three attributes. There may be situations where classification with respect to one attribute is dichotomous while it is manifold with respect to the other. A two-way classification of this type is shown as:



5. Quantitative classification or classification according to variables

In case of quantitative data, the characteristic is measurable in terms of real numbers and is termed as variable, e.g., weight, height, income, the number of children in a family, the number of crime cases in a city, life of an electric bulb of a company, etc. A variable can take a different value corresponding to a different item of the population or universe.

Variables can be of two types:

- (a) **Discrete Variable:** A discrete variable can assume only some specific values in a given interval. For example, the number of children in a family, the number of rooms on each floor of a multistoried building, etc.
- (b) **Continuous Variable:** A continuous variable can assume any value in a given interval. For example, monthly income of a worker can take any value, say, between ` 1,000 to 2,500. The income of a worker can be ` 1,500.25, etc. Similarly, the life of an electric bulb is a continuous variable that can take any value from 0 to ∞ .

Q27. Explain the different methods of classifications?

Ans :

The three major methods of classification of data are as follows,

1. Exclusive method
2. Inclusive method
3. Open-end class interval method

1. Exclusive Method

In exclusive method, the upper limit of one class would be the lower limit of the successive class when the class intervals in a frequency distribution are fixed. This is explained below,

Marks	Number of Students
0 – 5	2
5 – 10	4
10 – 15	5
15 – 20	12
20 – 25	16
25 – 30	22
30 – 35	25
35 – 40	27

2. Inclusive Method

In inclusive method, the upper limit of one class is also included in the similar class itself when the class intervals in a frequency distribution are fixed. This is explained below,

Weekly Wages (in `)	No. of Workers
0 – 50	7
51 – 100	11
101 – 150	14
151 – 200	16
201 – 250	18
251 – 300	22
301 – 350	26
351 – 400	29

3. Open-end Class Interval Method

An open-end class interval method does not provide any upper limit and lower limit. This is explained below.

Monthly Income (in `)	Number of Persons
Below 500	8
501 – 1000	14
1001 – 1500	22
1501 – 2000	27
2001 – 2500	34
2501 – 3000	38
Above 3001	43

Q28. What is tabulation? What are the objectives of tabulation?

Ans :

Meaning

A Table is a systematic arrangement of statistical data in columns and rows. Rows are horizontal arrangements, whereas columns are vertical. Tabulation is systematic presentation of data in a form suitable for analysis and interpretation.

Objectives

- Simplify Data:** The main objective of tabulation is to simplify a mass of information into its simplest form, making it easy to understand.
- Provide Clarity:** Tabulation provides clarity on data collected, Many of the questions that are related to the objective of the study can be easily answered through tabulation

3. **Facilitate Comparison:** Tabulation facilitates easy comparison. The two competing options that need to be evaluated can be placed side by side and compared.
4. **Trend Analysis:** Tabulation helps to catch the trend by more observation. Tabulated data is well placed to perform statistical calculation of trend and other features of data.
5. **Economy of Space and Time:** Presentation of data in tabular form saves space without any compromise on quality or completeness of data. Not only presentation but usage of data also results in savings of time for the users of such data.
6. **Enables Easy Reference:** Tabulated information with appropriate title and number are easy to remember and convenient to refer at future date. Most Research papers will have a separate Table of contents listing down the various table in which useful statistical data is presented and their page numbers.
7. **Detections of Errors and Omissions:** Particularly in case of numeric data, tabulation serves as a control. If any of the totals of rows or columns are not as they should be, it gets immediately known.
8. **Facilitate Statistical Processing:** It is only after tabulation that data becomes fit for statistical processing we can apply various statistical methods and techniques such as correlation, calculation of various measures of central tendency and dispersion etc, only after data is correctly tabulated.
9. **Clarity on Characteristics of Data:** A concise tabular form clearly reveals the characteristics of data and highlights its significant characteristics.

Q29. Explain the Essential parts of a table.

Ans :

1. Table Number

Indicates the serial number of the table. Table Numbers help in easy identification and reference of the table in future. Hence, they should be stated in clear and visible terms. It is preferred to write the Table Number at the top of the table, usually in center.

2. Table Title

Indicates the subject matter of the table. The title should be as unambiguous as possible. It should be clearly worded and indicate the nature of data contained in the table. The title should ideally indicate then 'what, where and How of the classified data and hint at the period of which it relates. Titles should not be too long, However, if the title is becoming very lengthy, it is advised to have a 'catch-line' with the rest of title stated in the next line. The table title should be very prominent in bold, and at the center, just below the Table Number.

3. Captions

It refers to the headings of the columns. Captions usually have a heading and also sub-heads, depending on the type of data being tabulated. A caption should be brief, concise and self-explanatory. Care should be taken that each of the column names does not spill into other column areas. It is preferable to number the captions and the sub-heads for case of reference, particularly when calculations are involved.

4. Sub

It refers to the headings of the rows. As with captions, rows can also have further subdivisions. However, stubs do not need to be very brief. Often, the choice of arrangement is done in such a way that the more descriptive items are taken as stubs. As in case of captions, it is desirable to number the stubs as well.

5. Body

It refers to the numerical information entered into the table. The entire statistical data that is to be presented constitutes the 'body'. Tabulation is an effort to present this 'body' in a more meaningful manner.

6. Head Note

It is a brief explanatory statement with respect to some or all parts of the table to facilitate better understanding of the nature of information provided in the table.

7. Foot Note

Footnotes provide further explanation with regard to the contents of the table. They may include exceptions to data, classifications etc.

8. Source Data

In case the table is not generated from primary data collected by the statistical investigator, the source from which data is collected is given at the bottom of the table. The source is stated below the footnotes. Stating the source adds credibility to the data. Usually, we state the 'original source' of data along with other indicators such as page number, table number, etc so that it is easy for anybody to refer to it at any time.

Q30. What are the differences between classification and tabulation?

Ans :

Basis for Comparison	Classification	Tabulation
Meaning	Classification is the process of grouping data into different categories, on the basis of nature, behavior, or common characteristics.	Tabulation is a process of summarizing data and presenting it in a compact form, by putting data into statistical table.
Order	After data collection	After classification
Arrangement	Attributes and variables	Columns and rows
Purpose	To analyse data	To present data
Bifurcates data into	Categories and sub-categories	Headings and sub-headings

1.8 FREQUENCY DISTRIBUTION**Q31. What is frequency distribution? Explain briefly about discrete and continuous frequency distribution.**

Ans :

Arrangement of raw data in the form of row or column is called data array.

Frequency distribution refers to the tabular arrangement of data, when arranged into groups or categories according to conveniently established divisions of the range of the observations. In a frequency distribution, raw data is expressed in distinct groups called 'classes'. The number of observations that fall into each of these classes is known as 'frequency'. Thus, a frequency distribution has two parts, the left part represents the classes and the right part represents the frequencies.

When data is described by a continuous variable, it is called continuous data and when described using discrete variable it is called discrete data.

The following are the two examples of discrete and continuous frequency distributions.

No. of Children	No. of Families
0	10
1	40
2	80
3	100
4	250
5	150
6	50
Total	680

**Table (1): Discrete Frequency
Distribution**

Weight (lb.)	No. of Persons
100 – 110	10
110 – 120	15
120 – 130	40
130 – 140	45
140 – 150	20
150 – 160	4
Total	134

**Table (2): Continuous Frequency
Distribution**

i) Construction of a Discrete Frequency Distribution

The process of preparing frequency distribution is very simple. For discrete data, place all the possible variables values in ascending order in one column and then prepare another column for 'Tally marks' (|) to count the number of marks. A block of five marks is prepared and some space is left in between the blocks. The frequency column refers to the number of "Tally marks", a particular class will contain. The construction of a discrete frequency distribution is illustrated as follows.

In a survey of 35 families in a village, the number of children per family was recorded and the following data obtained:

1	0	2	3	4	5	6
7	2	3	4	0	2	5
8	4	5	12	6	3	2
7	6	5	3	3	7	8
9	7	9	4	5	4	3

'Tally marks' are used to condense this data into a discrete frequency distribution as shown below.

Frequency Distribution of the Number of Children

No. of Children	Tallies	Frequency
0		2
1		1
2		4
3	<del> </del>	6
4	<del> </del>	5
5	<del> </del>	5
6		3
7		4
8		2
9		2
10	-	0
11	-	0
12		1
		Total 35

ii) Formation of a Continuous Frequency Distribution

Before constructing the frequency distribution for continuous data, it is necessary to identify some of the important terms that are used frequently.

➤ **Class Limits**

The class limits are the lowest and the highest values that can be included in the class. For example, take the class 20-40. The lowest value of the class is 20 and the highest 40. The two boundaries of class are known as the lower limit and the upper limit of the class. The lower limit of a class is the value below which there can be no item in the class. The upper limit of a class is the value above which no item can belong to that class. Of the class 70-89, 70 is the lower limit and 89 upper limit, i.e., in this class there can be no value which is less than 70 or more than 89. Similarly, if we take the class 90-109, there can be no value in that class which is less than 90 or more than 109. The way in which class limits are stated depends upon the nature of the data.

➤ **Class Intervals**

The difference between the Upper and Lower Limit of a class is known as class interval of that class. For example, in the class 100-200, the class interval is 100 [Le., 200 minus 100). An important decision while constructing a frequency distribution is about the width of the class interval, Le., whether it should be 10, 20, 50, 100, 500, etc. The decision would depend upon a number of factors such as the range in the data, i.e., the difference between the smallest and largest item, the details required and number of classes to be formed etc.

➤ Class Midpoint

It is the value lying halfway between the lower and upper class limits of a class-interval. Mid-point of a class is ascertained as follows :

$$\text{Mid-point of a class} = \frac{\text{Upper limit of the class} + \text{Lower limit of the class}}{2}$$

For the purpose of further calculations in statistical work the mid-point of each class is taken to represent that class.

➤ Class Frequency

The number of observations corresponding to a particular class is known as the frequency of that class or the class frequency. In the following illustration, the frequency of the class 1000 –1100 is 50 which implies that there are 50 persons having income between Rs. 1000 and Rs.1100. If we add together the frequencies of all individual classes, we obtain the total frequency. Thus, in the same problem, the total frequency of the six classes is 550 which means that in all there are 550 persons whose income has been studied.

Types of Class Interval

There are different ways in which limits of a class interval can be expressed such as,

1. Exclusive Method

The class intervals are arranged in such a way that the upper limit of one class is the lower limit of the next class. For example, a table below represents an exclusive method for class intervals.

Income (₹)	No. of Persons
1000 – 1100	50
1100 – 1200	100
1200 – 1300	200
1300 – 1400	150
1400 – 1500	40
1500 – 1600	10
	Total 550

2. Inclusive Method

Under the 'inclusive' method of classification. The upper limit of one class is included in that class itself. The following example illustrates the method:

Income (₹)	No. of Persons
1000 – 1099	50
1100 – 1199	100
1200 – 1299	200
1300 – 1399	150
1400 – 1499	40
1500 – 1599	10
	Total 550

In the class 1000 – 1099 we include persons whose income is between Rs. 1000 and Rs. 1099. If the income of a person is exactly Rs. 1100 he is included in the next class. The above example makes it clear that there is no confusion here of the type we find under the exclusive method. We may have classes like 1000 – 1099.5 or 1000 – 1099.99, and so on.

Q32. State the merits and demerits of frequency distribution.

Ans :

Merits

The following are the merits of frequency distribution,

1. To obtain knowledge about the important features of the data, the data is expressed in a compact form.
2. In frequency distribution it is possible to note quickly the pattern of distribution of observations in different classes.
3. To analyze and reveal the data characteristics, statistical techniques can be used.

Demerits

The following are the demerits of frequency distribution,

1. The identity of individual observations may be lost in the process of grouping. It becomes difficult to notice how the observations are put in each class and are distributed.
2. In different classes, there will be high clustering of observations, if the grouped observations are less than five. As a result, some important data may not be revealed.

Q33. What are the different types of frequency distribution? Explain with examples.

Ans :

The following are the three types of frequency distributions,

1. Cumulative frequency distribution
2. Relative frequency distribution
3. Percentage frequency distribution.

1. Cumulative Frequency Distribution

The normal frequency distribution simply tells us about the number of occurrence of a particular observation (i.e., value of the variable). However, it cannot provide the information such as “total number of observations whose value is less than or greater than a particular variable value”. In such a situation,

cumulative frequency distribution can be implemented. This frequency distribution can be obtained by adding all the successive frequencies of the variable values according to some principles. These are two types of cumulative frequencies, they are,

- (i) 'Less than' cumulative frequency
- (ii) 'More than' cumulative frequency.

The less than cumulative frequency can be obtained simply by taking the first frequency as it is and then adding all the succeeding frequencies to it.

On the other hand, the more than cumulative frequency can be obtained by taking the total number of frequencies as the first frequency and subtracting all the succeeding frequencies.

Example

The marks of 100 students in a test were recorded and shown by the following frequency distribution.

Marks	No. of Student
10-19	4
20-29	10
30-39	16
40-49	20
50-59	20
60-69	10
70-79	10
80-89	10

Table: Construct the Cumulative Frequency Table

Solution :

The cumulative frequency table is given below. If we keep on adding the frequency of each class starting from the frequency of the very first class, we get cumulative frequency.

Class Interval	Frequency	Cumulative (Less than)	Cumulative (More than)
10-19	4	4	100
20-29	10	$4 + 10 = 14$	$100 - 4 = 96$
30-39	16	$14 + 16 = 30$	$96 - 10 = 86$
40-49	20	$30 + 20 = 50$	$86 - 16 = 70$
50-59	20	$50 + 20 = 70$	$70 - 20 = 50$
60-69	10	$70 + 10 = 80$	$50 - 20 = 30$
70-79	10	$80 + 10 = 90$	$30 - 10 = 20$
80-89	10	$90 + 10 = 100$	$20 - 10 = 10$
	100		

Table: Cumulative Frequency

Upper Limits	Cumulative Frequency (Less than)
Less than 19	4
Less than 29	14
Less than 39	30
Less than 49	50
Less than 59	70
Less than 69	80
Less than 79	90
Less than 89	100

Table: (1)

Lower Limits	Cumulative Frequency (More than)
More than 10	100
More than 20	96
More than 30	86
More than 40	70
More than 50	50
More than 60	30
More than 70	20
More than 80	10

Table: (2)

From the above tables (1) and (2), it can be observed that cumulative frequencies are corresponding to upper limit and lower limits of class intervals. The 'more than' cumulative frequencies are corresponding to the lower limits of class intervals and the 'less than' cumulative frequencies are corresponding to upper limit of class intervals.

2. Relative Frequency Distribution

At times it may be desirable to convert class frequencies to relative class frequencies to show the percentage of the total number of observations in each class. For example, we may be interested in knowing the percentage of employees earning less than Rs. 2,000 per month.

In order to convert a frequency distribution to a relative frequency distribution, each of the class frequencies is divided by the total number of frequencies so that the relative frequencies would always total 1.

3. Percentage Frequency Distribution

Percentage frequency distribution is calculated after calculating the relative frequency distribution. The formulae for calculating percentage frequency distribution is given below,

$$\text{Percentage Frequency Distribution} = \text{Relative Frequency} \times 100$$

(or)

$$\frac{\text{Frequency Observation}}{\text{Total Number of Observation}} \times 100$$

Rahul Publications

Short Question and Answers

1. Explain the various functions of statistics.

Ans :

The, following are the main functions of statistics:

(i) Presents facts in numerical figures

The first function of statistics is to present a given problem in terms of numerical figures. We know that the numerical presentation helps in having a better understanding of the nature of a problem. Facts expressed in words are not very useful because they are often vague and are likely to be understood differently by different people.

(ii) Presents complex facts in a simplified form

Generally a problem to be investigated is represented by a large mass of numerical figures which are very difficult to understand and remember. Using various statistical methods, this large mass of data can be presented in a simplified form. This simplification is achieved by the summarization of data so that broad features of the given problem are brought into focus.

(iii) Studies relationship between two (or) more phenomena

Statistics can be used to investigate whether two or more phenomena are related. For example, the relationship between income and consumption, demand and supply, etc., can be studied by measuring correlation between relevant variables. Furthermore, a given mathematical relation can also be fitted to the given data by using the technique of regression analysis.

2. Define distrust of statistics.

Ans :

Although statistics plays an important role in many areas, there is still the possibility of statistics to be misused by the biased people. The use of statistics by such people leads to deceptive interpretation of a given problem. Among the common people, this makes a feeling of distrust about statistics. Many authors expressed there feeling about the distrust of statistics, which are as follows,

- (i) "There are three kinds of lies: lies, demand lies and statistics".
- (ii) "Statistics can prove anything".
- (iii) "Statistics is a rainbow of lies".
- (iv) "Statistics are like clay of which you can make god or devil, as you please".

These statements shows a feeling of distrust of statistics.

3. Define the statistics in plural sense.

Ans :

Definition of Statistics in Plural Sense-Numerical Data

- (i) According to 'Dr. A.L. Bowley', "statistics are numerical statements of facts in any department of enquiry placed in relation to each other".
- (ii) According to 'L.R. Comor', "Statistics are measurements, enumeration or estimates of natural or social phenomena, systematically arranged so as to exhibit their inter-relations".
- (iii) According to 'Prof. Yule and Prof. Kendall', "By statistics we mean quantitative data affected to a marked extent by multiplicity of causes".

- (iv) **According to 'Horace Secrist'**, "By statistics we mean aggregate of facts affected to a marked extent by multiplicity of causes numerically expressed, enumerated or estimated according to a reasonable standard of accuracy, collected in a systematic manner for a predetermined purpose and placed in relation to each other".
- (v) **According to Webster**, "Statistics are classified facts respecting the conditions of the people in a state - especially those facts which can be stated in numbers or in tables of numbers or in any other tabular or classified arrangement."
-

4. State the various limitations of statistics.

Ans :

(i) Statistics deals with numerical facts only

Broadly speaking there are two types of facts:

- (a) quantitative and
- (b) qualitative.

(a) Quantitative

Quantitative facts are capable of being represented in the form of numerical figures and therefore, are also known as numerical facts. These facts can be analysed and interpreted with the help of statistical methods.

(b) Qualitative

Qualitative facts, on the other hand, represent only the qualitative characteristics like honesty, intelligence, colour of eyes, beauty, etc. and statistical methods cannot be used to study these types of characteristics. Sometimes, however, it is possible to make an indirect study of such characteristics through their conversion into numerical figures. For example, we may assign a number 0 for a male and 1 for a female, etc.

(ii) Statistics deals only with groups and not with individuals

Statistical studies are undertaken to study the characteristics of a group rather than individuals. These studies are done to compare the general behaviour of the group at different points of time or the behaviour of different groups at a particular point of time. For example, the economic performance of a country in a year is measured by its national income in that year and by comparing national income of various years, one can know whether performance of the country is improving or not. Further, by comparing national income of different countries, one can know its relative position vis-a-vis other countries.

(iii) Statistical results are true only on the average

Statistical results give the behaviour of the group on the average and these may not hold for an individual of that very group. Thus, the statement that average wages of workers of a certain factory is ` 1,500 p.m. does not necessarily mean that each worker is getting this wage. In fact, some of the workers may be getting more while others less than or equal to ` 1,500. Further, when value of a variable is estimated by using some explanatory variable, the estimated value represents the value on the average for a particular value of the explanatory variable. In a similar way, all the laws of statistics are true only on the average.

5. What are the differences between Primary and Secondary Data ?

Ans :

(Dec.-18)

Basis of Difference	Primary Data	Secondary Data
1) Meaning	Primary data is the one, which is collected by the investigator himself for the purpose of a specific inquiry or study.	Secondary data are those data which have been already collected and analysed by some earlier agency for its own use; and later the same data are used by a different agency
2) Cost	The cost of obtaining primary data is typically more.	Secondary data is typically available for free or for the subscription fee to the database magazine or journal.
3) Sources	A primary data source is a publication in which the data are published by the same authority which gathered and analysed them.	A secondary data source is a publication, reporting the data which have been gathered by other authorities and for which others are responsible.
4) Methods	It is normally collected through experiments, surveys, questionnaire focus groups, interviews.	The method or way of collecting secondary data includes books, journals, census data biographies, articles and databases.
5) Reliability	The information collected for primary data is more reliable than those collected from the secondary data.	Whereas secondary data are less reliable because these information are not collected for that particular purpose.

6. Explain the various steps in Statistical Investigation

Ans :

- (i) **Planning of Investigation** : The first stage of any statistical inquiry is the pre-planned nature, purpose and scope.
- (ii) **Sources of Information** : To make an inquiry successful it is essential to determine the source of data. Source of data may be internal or external, or primary or secondary.
- (iii) **Types of inquiry** : Types of inquiry to be conducted includes the object and scope of such an inquiry. It is something experimental, direct or indirect inquiry.
- (iv) **Editing of data** : After collecting the data, editing makes it comparable and meaningful. Data should be properly classified in the form of schedules and tables.
- (v) **Units of data collection** : It means units of data should be uniform throughout the study. Otherwise, it will provide misleading conclusions.
- (vi) **Degree of Accuracy** : Degree of accuracy must be kept in mind.
- (vii) **Analysis of Data** : Data should be analysed, prepared by using various statistical measures like averages and dispersion.
- (viii) **Interpretation** : After analysing the data properly. It must be interpreted to suggest meaningful and impartial conclusions.

7. Characteristics of a statistics.

Ans :

(i) Statistics are aggregates of facts

Single and unconnected figures are not statistics, for example, if it is stated that Raj, a student, secured 50 marks, it is not statistics. However, marks secured by the students of the class would constitute statistics. A single figure relating to production income, marks height, etc. cannot be regarded as statistics but aggregates of such facts would be regarded as statistics.

(ii) Statistics are affected to a marked extent by multiplicity of causes

Numerical facts are affected by a multiplicity of factors. For example, the price, of a commodity is affected by number of factory such as supply, demand, imports, exports, money in circulation, competitive products in the market and so on. It is very difficult to study separately the effect of these factors on the price of the commodity. In physical sciences, it is possible to isolate the effect of various factors on a single item, but statistics are commonly used in social sciences and in social sciences it is very difficult to study the effect of any one factor separately. In statistical methods, the effects of various factors affecting a particular phenomenon are generally studied in a combined form, though attempts are also made to study the effects of different factors separately as well.

(iii) Statistics are numerically expressed

Only numerical data constitutes statistics. Qualitative expressions like good, bad, young, old etc. cannot be regarded as statistics. Statement like the standard of living of people in India has improved 'or production of petroleum products has increased in India do not constitute statistics. But Statements like "Rice production in India in 2003-04 is expected to be 86.4 million tones as against 72.66 million tons in 2002-03".

8. Define sampling.

Ans :

A 'sample' is a part of the 'Population'. It is subset of the entire set of units that fall within the scope of study. For example, to study the brand preference of consumers in aerated drinks (Coca Cola vs. Pepsi) we don't have to ask each and every person who consumes aerated drinks about their choice. Only a small number of persons are covered, who form the 'sample'. The number of units (or 'elements') in the sample is called sample size.

The method of obtaining data or conducting a statistical investigation by studying a sample, rather than the entire population, is 'sampling method'. The most important aspect of this method is that the sample should be 'representative' of the population.

9. Stratified Random Sampling

Ans :

If the 'population' consists of diverse segments that can be clearly distinguished, such population is first classified into groups each representing a divers segment, and then simple random techniques are used within each group, to come up with a total sample that is more representative of the diverse population. These diverse segments are called a 'strata' (Stratum, if singular) and this method is called 'stratified random sampling. The classification of population into various 'strata' can be done on the basis of its Geographic, sociological or Economic characteristics. The size of sample from each, strata is normally proportional to the total size of that .strata to the total population size. However, it can also be of disproportional size.

10. Non-Random Sampling Methods

Ans :

(i) Judgment Sampling

Judgment sampling relies on the capability of the investigator to select the appropriate sample, rather than the use of techniques. An experience and unbiased investigator can 'focus' on the issue and pick up the sample that meets the objectives of the study rather

than live with statistical errors arising out of tools and techniques. However, an inexperienced or biased investigator can arrive at totally misleading conclusions by not following proper random sampling methods.

(ii) Cluster Sampling

In this method, the entire population is divided into some recognizable sub-groups called 'clusters'. A random sample of the clusters is drawn and then all the units belonging to the clusters are considered as part of the sample.

(iii) Convenience Sampling

Under this method, a sample is obtained by selecting "such units that are convenient to locate or contact.

(iv) Sequential Sampling

Sequential sampling is not a separate method of "drawing a sample. It is more of a process. A sample is drawn and the statistical investigation is conducted. If the result are satisfactory, nothing else is done. However, if study of the sample is inconclusive, another sample is drawn. This process may go on for any number of iterations until the investigator can arrive at a conclusion. This process is applied in Statistical Quality Control (SQC).

(v) Quota Sampling

It is a combination of stratified Sampling and Judgment sampling. In this method, the population is divided into 'quotas'. Each quota consists of units with specified characteristics. Each investigator or enumerator is told about the number of units he/she needs to select from the quota. This is normally done using judgment. This should not be confused with cluster sampling. In cluster sampling, only few out of all clusters are picked and all units belonging to selected cluster are selected. In Quota sampling all quotas are selected but sampling is applied within each quota.

11. Types of Data Sources

Ans :

Usually data is collected through two different sources i.e., primary and secondary and based on these sources data is classified into two types,

- I) Primary data
- II) Secondary data.

I) Primary Data

Primary data refers to the data collected specifically which is for the purpose of research problem. It is the first hand information collected by the research firm or by an external agent with the objective of solving a research problem. There are different methods of collecting primary data. Researchers can conduct experiments to gather the required information. Other methods include questionnaires, mails, interviews of individuals, families, organisations, representatives etc.

II) Secondary Data

Primary data for one party sometimes acts as the secondary data for the other party. Secondary data refers to the existing data that had been collected with an objective other than for research. It could be the data collected by the firm itself for any other purpose, or by any external party for the same or other research problem. For example, inventory records maintained by a firm as part of their routine operating function acts as the internal secondary data of the firm for evaluating the seasonal demand for its product. Alternatively, data gathered on industry demand by a marketing research firm can also be used for the same research problem.

12. Questionnaire

Ans :

The questionnaire is an important tool for gathering primary data. Questionnaire are mailed to the respondents with a required to return after completing the same. It is most extensively used method in various economic and business surveys.

Questionnaire to be used must be prepared very carefully so that it may prove to be effective in collecting the relevant information. Poorly constructed questions can result in large errors and invalidate the research data, so significant effort should be put into the questionnaire.

13. Define Statistical Error.

Ans :

Meaning

In statistics, the word “error” is used in a special sense. It is voluntary in nature and can be corrected. It is not considered as mistake. Mistakes in statistical calculations takes place because of defects in the working of human minds and also due to some faults in machines. But in statistics, ‘error’ means the difference between the true value and the estimated value of an item.

Definition

According to Boddington ‘statistical error’ means “the divergence between the actual number and the estimate which is made either by approximation or by any other method”.

14. Define classification.

Ans :

Meaning

The largest amount of data and the greatest amount of detail may not convey the most useful information for decision making. An important aspect of Statistics is to organize and present data so as to convey critical information quickly.

Classification means grouping of a whole into different groups or classes. However, each of these groups should have a common characteristic.

For example, all living things can be classified into plants and animals. Animals can be classified into Amphibians, Reptiles, and Mammals etc.

Definition

According to Horace Secrets, “Classification is the process of arranging data into sequences and groups according to their common characteristics, or separating them into different but related parts”.

15. What is tabulation?

Ans :

Meaning

A Table is a systematic arrangement of statistical data in columns and rows. Rows are horizontal arrangements, whereas columns are vertical. Tabulation is systematic presentation of data in a form suitable for analysis and interpretation.

Objectives

1. **Simplify Data:** The main objective of tabulation is to simplify a mass of information into its simplest form, making it easy to understand.
2. **Provide Clarity:** Tabulation provides clarity on data collected, Many of the questions that are related to the objective of the study can be easily answered through tabulation

3. **Facilitate Comparison:** Tabulation facilitates easy comparison. The two competing options that need to be evaluated can be placed side by side and compared.
 4. **Trend Analysis:** Tabulation helps to catch the trend by more observation. Tabulated data is well placed to perform statistical calculation of trend and other features of data.
 5. **Economy of Space and Time:** Presentation of data in tabular form saves space without any compromise on quality or completeness of data. Not only presentation but usage of data also results in savings of time for the users of such data.
-

16. What is frequency distribution? Explain briefly about discrete and continuous frequency distribution.

Ans :

Arrangement of raw data in the form of row or column is called data array.

Frequency distribution refers to the tabular arrangement of data, when arranged into groups or categories according to conveniently established divisions of the range of the observations. In a frequency distribution, raw data is expressed in distinct groups called 'classes'. The number of observations that fall into each of these classes is known as 'frequency'. Thus, a frequency distribution has two parts, the left part represents the classes and the right part represents the frequencies.

When data is described by a continuous variable, it is called continuous data and when described using discrete variable it is called discrete data.

Choose the Correct Answers

1. Which of the following is not a stage in statistical investigation [a]
(a) Plainning the statistical investigation (b) Collection of data
(c) Editing of data (d) Electronic media reports to the respondents
2. Questionnaire is used to collect [b]
(a) Primary Data (b) Secondary data
(c) Secondary records (d) Unpublished data
3. The sources of collection of data are [c]
(a) Primary Records (b) Internal records
(c) Secondary records (d) External records
4. Which one of the following is note the feature of primary data ? [d]
(a) These are original data
(b) Investigator himself collects it
(c) These are not concerned with an objective
(d) None of these
5. The scope of the investigation depends on [d]
(a) The objective (b) Availability of time
(c) Resoureces (d) Only 1 and 2 and 3
6. Classification is the process of arranging data in [d]
(a) Different rows and columns (b) Different rows
(c) Different columns (d) None of these
7. M.A. Sturges rule for determining the number of classes is [a]
(a) $K = 1 + 1.322 \log_{10} N$ (b) $K = 1 - 3.3 \log N$
(c) $K = 3 + 1.322 \log N$ (d) $K = 1 + 2.3 \log N$
8. While constructing a frequency distribution table the number of classes should be [b]
(a) Between 4 and 9 (b) Between 5 and 15
(c) Between 4 and 25 (d) More than 25
9. Geographical classification means arrangement of data according to [a]
(a) Location (b) Attributes
(c) Class intervals (d) Time
10. Complex table does not mean [d]
(a) Manifold table (b) One way table
(c) Three way table (d) None of these

Fill in the blanks

1. Statistics was also known as _____.
2. _____ methods are the tools that are in the hands of the statistician.
3. _____ deals with application of Statistical methods of specific problems or concrete forms.
4. _____ statistics deals with formulation of statistical laws.
5. _____ facts are capable of being represented in the form of numerical figures.
6. The search for knowledge, done by analysing numerical facts, is known as _____.
7. A 'sample' is a part of the _____.
8. A _____ sample is one which is selected in such a way that every item in the population has equal chance of getting included in the sample.
9. _____ sampling relies on the capability of the investigator to select the appropriate sample, rather than the use of techniques.
10. SQC stands for _____.
11. _____ data refers to the data collected specifically which is for the purpose of research problem.
12. _____ are a form of questioning that is more rigid than interviews and that involve larger groups of people.
13. The _____ is an important tool for gathering primary data.
14. _____ means grouping of a whole into different groups or classes.
15. When the data are classified according to geographical location or region, it is called a _____ classification.
16. When the data are classified on the basis of its time of occurrence, it is called a _____ classification.

ANSWER

1. Science of State Craft
2. Statistical
3. Applied Statistics
4. Scientific
5. Quantitative
6. Statistical investigation
7. Population
8. Random
9. Judgment
10. Statistical Quality Control
11. Primary
12. Surveys
13. Questionnaire
14. Classification
15. Geographical
16. Chronological

UNIT II

DIAGRAMMATIC AND GRAPHIC PRESENTATION:

Diagrammatic presentation: One Dimensional and Two Dimensional Diagrams- Pictograms - Cartograms - Graphic presentation: Technique of Construction of Graphs - Graphs of Frequency Distribution - Graphs of Time Series or Histograms.

2.1 DIAGRAMMATIC PRESENTATION

Q1. Define Diagram. State the importance of diagrams.

Ans : (July-21, Imp.)

Meaning

Diagram is a visual presentation of statistical information. The pictorial presentation helps in proper understanding the data. Diagrams are of differ types like pie diagram, rectangles, lines, pictures and maps.

Importance

1. Diagrams reduce huge amount of data into simple figures. Thus, it is the best option for making short duration presentations to top management, Government, customer, etc who. cannot go into details.
2. Diagrams bring out the essence of the underlying data with great clarity. A person going through the data may get lost in detail, but will get the message immediately if the same data is presented as a diagram
3. Diagrams help in highlighting the trends in underlying data that cannot be easily spotted by going through the data.
4. Diagrams create a positive and lasting impression. The audience can remember the details for a longer time.
5. Diagrams have the ability to reach a greater audience people who cannot read can also understand diagrams.
6. Diagrams facilitate comparison of data. Interesting conclusions can be drawn by merely juxtaposing two figures.

7. Diagrams have universal acceptance. There is not restriction of language, time or place.
8. Diagrams bring forth the characteristics of data, highlighting hidden facts and relationship.
9. Diagrams are helpful in garnering the attention of the user and create interest in the subject.
10. Diagrams save time, effort and costs. It is very economical to prepare and present diagrams.

Q2. Explain the general rules for constructing diagrams.

(OR)

What are the components of diagrams.

Ans :

1. Title

Each diagrams must be given a suitable title to convey the main idea it is intended to portray.

2. Scale

Scale should be selected consistent with the size of observations to be displayed. The diagram should be neither too small nor too large. The scale should specify the size of the unit and what it represents such as marks scored, in numbers price in rupees etc.

3. Proportion between width and height

A diagram should be displayed in the center of the page. It should ideally maintain a proportion of 1:1.414 between the smaller side and the larger side respectively.

4. Choice of Diagram

The choice of a particular type of diagram has to be made depending on nature of data magnitude of observations and type of people Who use the diagram. For example, data relating to frequency curves and time series is best represented by means of graphs.

5. Index

An index illustrating different types of shades, colors, lines, designs etc used for presenting the data should be provided for easy understanding of the diagram by the user.

6. Footnotes

Classifications regarding certain aspects of the diagram may be provided as a footnote.

7. Source Note

The source from where the data has been collected may be provided.

8. Simplicity

Diagrams are meant to make complex data simple. Hence, too much information should not be cluttered into a single diagram.

9. Neatness

Lastly, diagrams are visual aids and hence need to be neatly presented. A variety of devices such as dots, broken lines etc are used to make the diagram attractive.

10. Communicative

Diagrams must be intelligible. They must convey a message.

11. Self-Explanatory

Any person looking at the diagram should be able to understand the underlying message. It must be self-explanatory.

12. Reliable

Diagrams should be reliable. They should not mislead the user, in an attempt to make the presentation more attractive data accuracy must be maintained.

Q3. What are the functions of Diagrams?

Ans : (Dec.-17, Imp.)

Following points highlights the functions of diagrams.

1. Diagrams give a clear picture of the data, which can be understood by a common man easily and quickly.
2. By using diagrammatic presentation data can be short end.
3. Diagrams assist in locating median and mode values.
4. Diagrammatic presentation detect computation errors.
5. They can be used for carrying out investigation.

Q4. State the advantages and disadvantages of diagrams.

Ans : (July-21, Imp.)

Advantages**1. Easy to understand**

Diagrammatic data presentation makes it easier for a common man to understand the data. Diagrams are usually attractive and impressive and many newspapers and magazines use them frequently to explain certain facts or phenomena. Modern advertising campaigns also use diagrams.

2. Reveals hidden facts

When you classify and tabulate data, some facts are not revealed. Diagrammatic data presentation helps in bringing out these facts and also relations.

3. Quick to grasp

Usually, when the data is represented using diagrams, people can grasp it quickly.

4. Easy to compare

Diagrams make it easier to compare data.

5. Universally accepted

Almost all fields of study like Business, economics, social institutions, administration, etc. use diagrams. Therefore, they have universal acceptability.

Disadvantages**1. Provides vague ideas**

While diagrams offer a vague idea about the problem, it is useful only to a common man. An expert, who seeks an exact idea of the problem cannot benefit from them.

2. Limited information

Classified and tabulated data provides more information than diagrams.

3. Low precision

Diagram offer a low level of precision of values.

4. Restricts further data analysis

Diagrams do not allow the user to analyze the data further.

5. Portrays limited characteristics

Diagrams tend to portray only a limited number of characteristics. Therefore, it is difficult to understand a large number of characteristics using diagrams.

6. A possibility of misuse

Sometimes diagrams are misused to present an illusory picture of the problem.

7. Fail to present a meaningful look in certain situations

If the data has various measurements and wide variation, then diagrams do not present a meaningful look.

8. Careful usage

If diagrams are drawn on a false baseline, then the user must analyze them carefully.

2.2 TYPES OF DIAGRAMS**Q5. Explain the different types of diagrams.**

Ans :

1. One Dimensional Diagrams

One-dimensional diagram means the construction of a diagram on the basis of one dimension. This could be height in case of vertical diagrams (or) length in case of horizontal diagrams.

One dimensional diagrams can be line diagrams or bar diagrams.

2. Two-Dimensional Diagrams

In two-dimensional diagrams, both length and width of the observations are represented by bars. The area of a two-dimensional diagram is equal to the product of its length and width. So, two-dimensional diagrams are also called 'area diagrams'.

3. Three-Dimensional Diagrams

Three dimensional diagrams are diagrams in which three dimensions, namely, length breadth and height are taken into account. They are also termed as volume diagrams. The objective of presenting data in 3 dimensions is to highlight the magnitude of the data through volumes of the corresponding diagrams. Diagrams in the shape of cubes, spheres, cylinders, blocks etc are commonly used. Three dimensional diagrams are specially useful if there are very wide variations between the smallest and the largest magnitudes.

4. Pictograms

Pictures are more attractive and appealing to the eye and have a Tasting impression on the mind. Pictograms is the technique of presenting statistical data through appropriate pictures. It is a very popular device for presenting statistical facts to a layman who does not have any mathematical background. The magnitude of the particular phenomenon under study is of the pictures being proportional to the values of the different magnitudes to be presented.

5. Cartograms

Cartograms are used to present data pertaining to geographical regions with the help of maps. For example, the distribution of rainfall in different regions of a country can be shown with the help of maps or cartograms. The different geographical regions are depicted on a map and the

quantum of rainfall in the regions may be shown by dots, different shades or colours etc. Cartograms are simple and elementary forms of visual presentation and are easy to understand. They are generally used when the regional or geographic comparisons are to be highlighted.

2.2.1 One Dimensional Diagrams

Q6. Define One Dimensional Diagrams. Explain different types of One Dimensional Diagrams.

Ans :

One Dimensional Diagrams: One-dimensional diagram means the construction of a diagram on the basis of one dimension. This could be height in case of vertical diagrams or length in case of horizontal diagrams.

One dimensional diagrams can be line diagrams or bar diagrams.

Types

(i) Line Diagram

A line diagram involves drawing multiple vertical lines. Each vertical line represents a particular frequency. The variety (x) values are presented on X axis and the corresponding frequencies for each value of X are presented on Y axis. Both the axis are drawn as per a suitable scale. Let us understand the preparation with the help of an illustration.

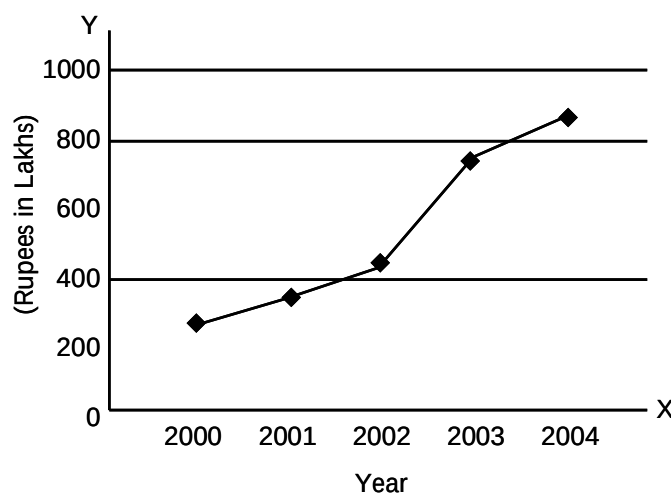
Example

Present the following frequency distribution in the form of a line diagram

Year	2000	2001	2002	2003	2004
Sales (Rupees in lakhs)	287	360	439	743	889

Sol :

X axis will represent Year and Y axis will represent sales in lakhs of rupees. One line needs to be drawn for each year in the ratio of the magnitude of the given data. Identical gap should be provided in between the different lines. The scale of one centimeter on Y-axis is equal to Rupees Two Hundred lakhs.



(ii) Bar Diagram

Bar diagrams are one of the most commonly used diagrams for presenting data. This type of diagram is generally constructed when the data represent different values over a period or different situations. They are largely used by business and Government to present economy related data such as industrial / agricultural production inflation exports. Forex reserves etc.

Bar diagrams are used specifically for categorical data or series. They consist of a group of equidistant rectangles. Bars may be drawn vertically (or) horizontally. The concept of Bar diagrams is similar to that of line diagram.

As bar diagrams are one dimensional the height/ length of the bars alone is important. The height/ length of the bars should be in proportion to magnitude of observations. The bars can be of any width or thickness, but such thickness (or) width should make the diagram attractive.

Q7. What are the advantages and disadvantages of Bar diagrams?

Ans :

Advantages

The chief advantages of a bar diagram can be outlined as under:

1. It is very simple to draw and read as well.
2. It is the only form of diagram which can represent a large number of data on a piece of paper.
3. It can be drawn both vertically and horizontally.
4. It gives a better look and facilitates comparison.

Disadvantages

1. It cannot exhibit a large number of aspects of the data.
2. The width of the bars are fixed arbitrarily by a drawer.

Q8. Explain the different types of Bar diagrams.

Ans :

Bar diagrams are of the following types :

- (a) Simple bar diagrams
- (b) Sub-divided bar diagrams
- (c) Multiple bar diagrams
- (d) Percentage bar diagrams
- (e) Deviation bars.

(a) Simple Bar Diagrams

A simple bar diagram is used to represent only one variable. For example, the figures of sales, production, population, etc., for various years may be shown by means of a simple bar diagram. Since these are of the same width and only the length varies, it becomes very easy for the reader to study the relationship. Simple bar diagrams are very popular in practice. This can be either vertical or horizontal. In practice, vertical bars are more popular. However, an important limitation of such diagrams is that they can present only one classification or one category of data. For example, while presenting the population for the last five decades, one can only depict the total population in the simple bar diagrams, and not its sex-wise distribution.

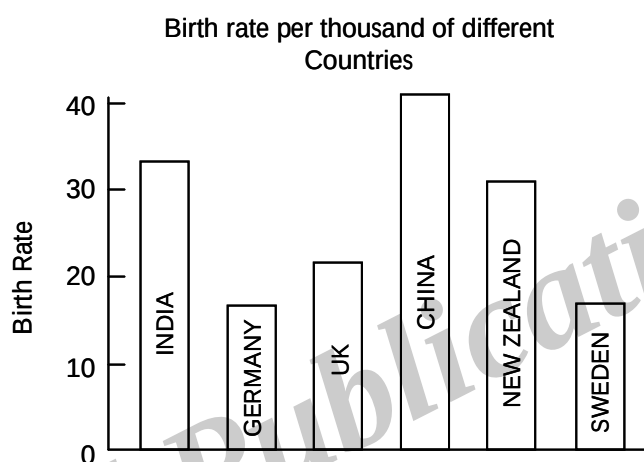
Example

Following table gives the birth rate per thousand of different countries over a certain period :

Country	Birth Rate	Country	Birth Rate
India	33	China	40
Germany	16	New Zaland	30
U.K.	20	Sweden	15

Represent the above data by a suitable diagram.

Sol :

**(b) Sub-divided Bar Diagrams**

In a sub-divided bar diagram each bar representing the magnitude of a given phenomenon is further sub-divided in its various components. Each component occupies a part of the bar proportional to its share in the total.

For example, the number of students in various courses for B.Com., M.Com., BA. and M.A. in various colleges may be represented by a sub-divided bar diagram. While constructing such a diagram, the various components in each bar should be kept in the same order. A common and helpful arrangement is that of presenting each bar in the order of magnitude from the largest component at the base of the bar to the smallest at the end. To distinguish between the different components, it is useful to use different shades or colours, index or key should be given explaining these differences.

Sub-divided bar diagrams should not be used where the number of components is more than 10 or 12, for, in that case, the diagram will be overloaded with information which cannot be easily compared and understood.

The component bar diagram can be used to represent either the absolute data or distribution ratios such as percentage distributions. It is, in fact, an excellent method for presenting a set of distribution ratios diagrammatically. The sub-divided bar diagrams can be constructed both on horizontal and vertical bases.

Example :

The growth of production for the period 1991-92 to 1997-98 is given below:

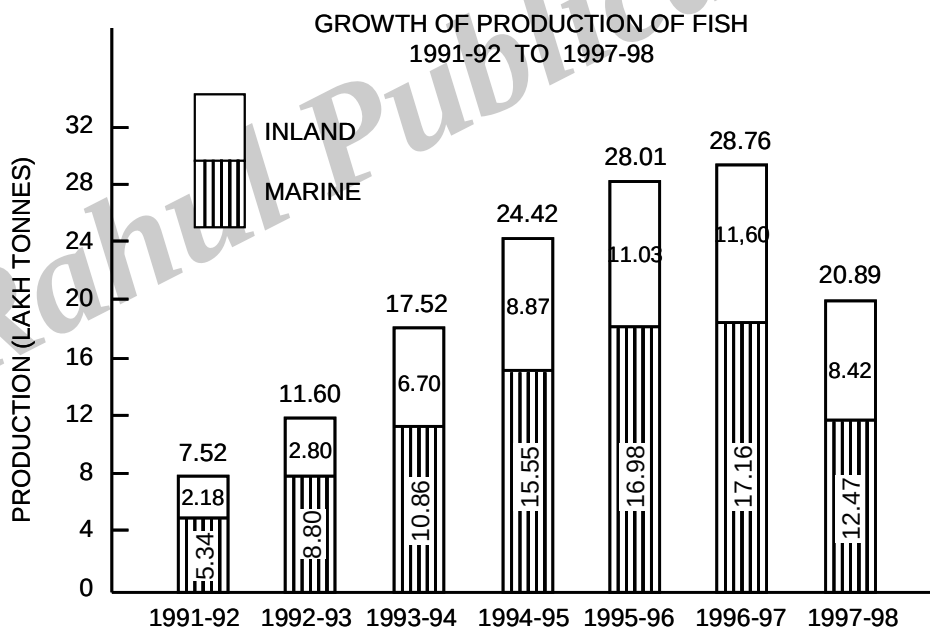
(lakh tonnes)

Year	Marine	Inland	Total
1991 - 92	5.34	2.18	7.52
1992 - 93	8.80	2.80	11.60
1993 - 94	10.86	6.70	17.56
1994 - 95	15.55	8.87	24.42
1995 - 96	16.98	11.03	28.01
1996 - 97	17.16	11.60	28.76
1997 - 98	12.47	8.42	20.89

Represent the data by a suitable diagram.

Sol :

The above data can be represented by a sub-divided bar diagram.

**(c) Multiple Bars Diagrams**

In a multiple bar diagram two or more sets of interrelated data are represented. The technique of drawing such a diagram is the same as that of simple bar diagram. The only difference is that since more than one phenomenon is represented, different shades, colours, dots or crosses are used to distinguish between the bars. Wherever a comparison between two or more related variables is to be made, multiple bar diagram should be preferred.

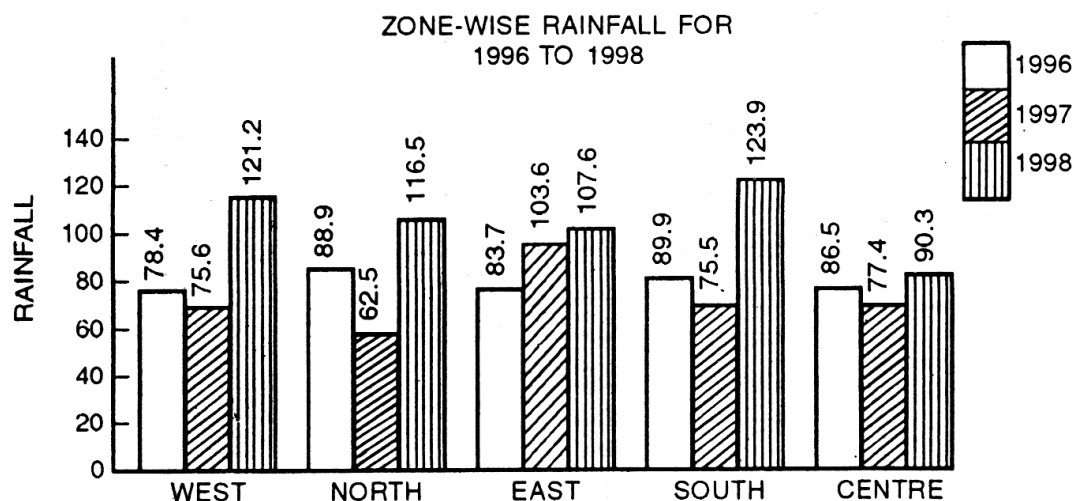
Example

The regional rainfall indices during the year 1997 to 1998 are given below :

Year	Zones				
	West	North	East	South	Centre
1996	78.4	88.9	83.7	89.9	86.5
1997	75.6	62.5	103.6	75.5	77.4
1998	121.2	116.5	107.6	123.9	90.3

Represent the data by a multiple bar diagram.

Sol :

**(d) Percentage Bar Diagrams**

Percentage bars are particularly useful in statistical work which requires the portrayal of relative changes in data. When such diagrams are prepared, the length of the bars is kept equal to 100 and segments are cut in these bars to represent the components (percentages) of an aggregate.

Example

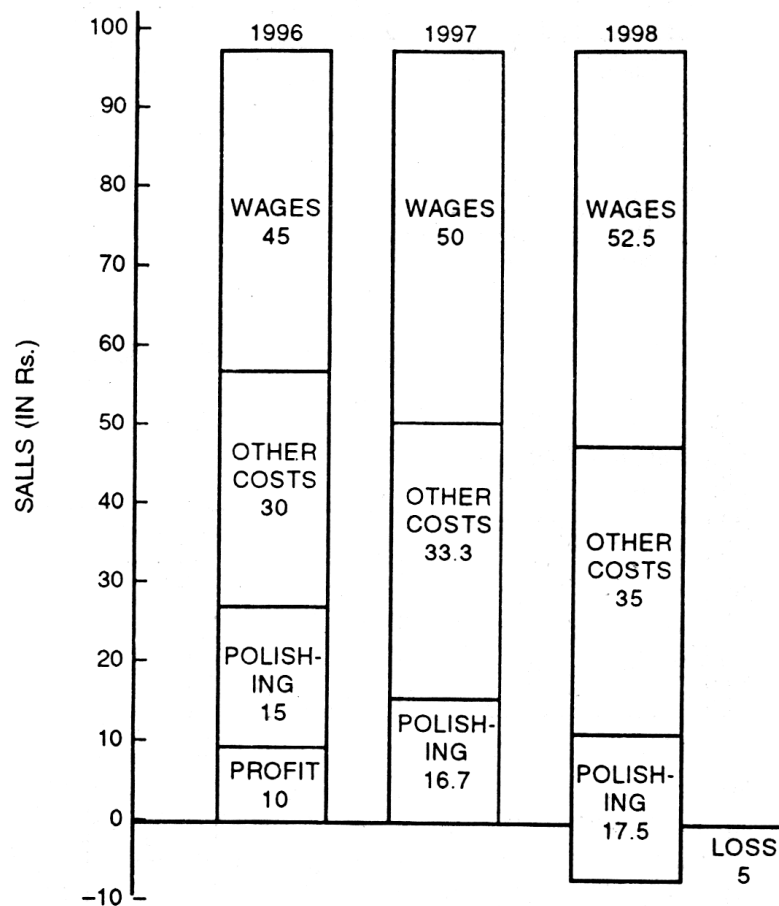
Represent the following by sub-divided bar diagram drawn on the percentage basis :

Particulars		1996	1997	1998
1.	Cost per chair			
	(a) Wages	9	15	21
	(b) Other costs	6	10	14
	(c) Polishing	3	5	7
	Total	18	30	42
2.	Proceeds per chair	20	30	40
3.	Profit (+) / Loss (-)	+2	-	-2

Sol :

Take the sale price per chair as Rs. 100 and express the other figures in , percentages. The percentages so obtained are given below:

Particulars	1996	1997	1998
Wages	45.0	50.0	52.5
Other costs	30.0	33.3	35.0
Polishing	15.0	16.7	17.5
Total costs	90.0	100.0	105.0
Sale price	100.0	100.0	100.0
Profit or loss	+10.0	—	-5.0



(e) Deviation Bars

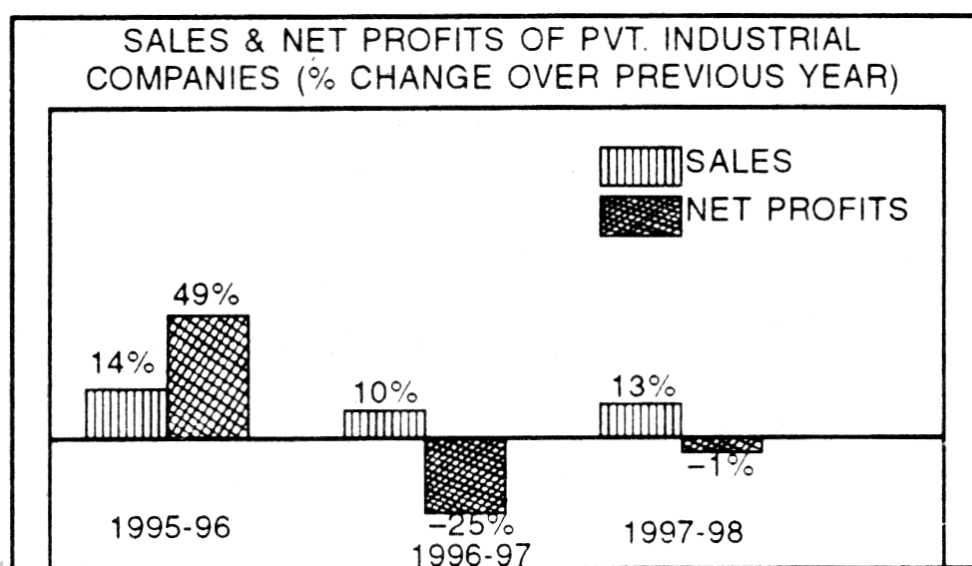
Deviation bars are popularly used for representing net quantities - excess or deficit, ie., net profit, net loss, net exports or unports, etc. Such bars can have both positive and negative values. Positive values are shown above the base line and negative values below it.

Example :

Present the following data by a suitable diagram showing the Sales and Net Profits of private industrial companies:

Year	Sales	Net Profits
1995 - 96	14%	49%
1996 - 97	10%	- 25%
1997 - 98	13%	-1 %

Sol :



Q9. How is percentage bar diagram different from multiple bar diagrams.

Ans :

(Dec.-18, Imp.)

Percentage Bar Diagrams

It comprises one or more segmented bars. The magnitude of each bar is 100%. The only difference between the sub-divided bar chart and percentage sub-divided bar chart is that in the former type, absolute quantities are presented whereas, in the later type, values are expressed in terms of percentage.

Multiple Bar Diagram

It is used when the number of commodities need to be compared with respect to two or more variables. In multiple bar charts, the values of the variables must be arranged in some order, (either ascending or descending). This is done in order to identify and differentiate the variable. The rules for drawing multiple bar chart is similar to that of simple bar chart with only one distinctive feature i.e., for representing the various variables different bars are drawn in multiple bar diagram.

PROBLEMS ON ONE DIMENSIONAL DIAGRAMS

1. The following table shows the results of B.Com. students of a university for the last three years. Represent the same by a multiple bar diagram.

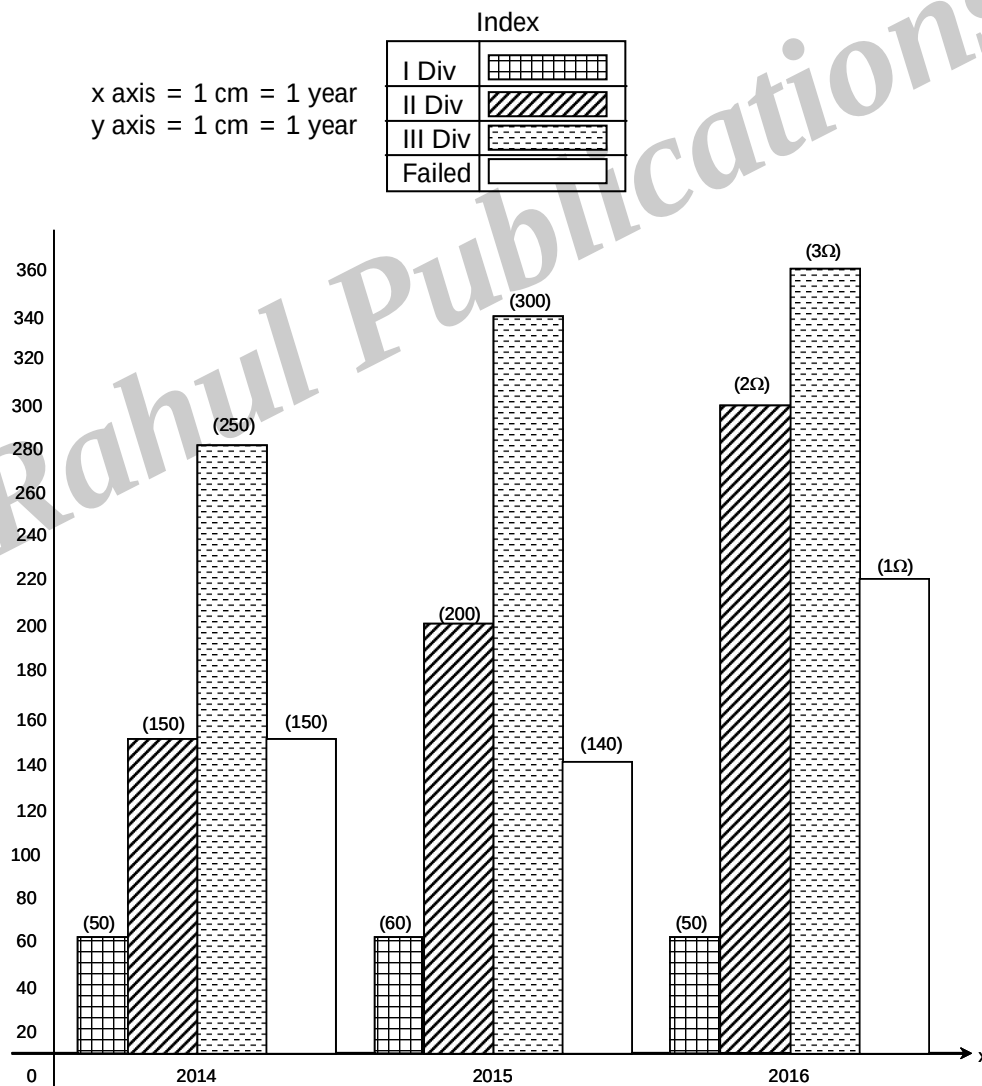
Year	Exam Result of B.Com Students			
	I Division	II Division	III Division	Failed
2014	50	150	250	150
2015	60	200	300	140
2016	50	250	350	150

Sol :

(Dec.-17, Imp.)

Diagram showing the results of can students of a university for the last three years.

Scale :



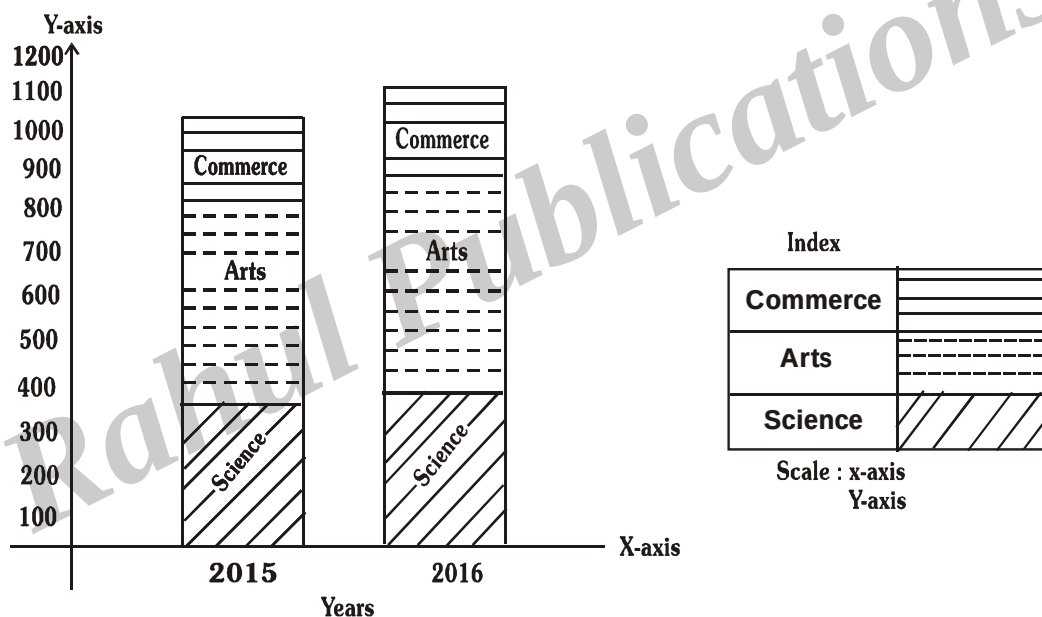
2. Present the following data in a sub-divided bar diagram.

Year	Faculty		
	Science	Arts	Commerce
2015	240	560	220
2016	280	610	280

Sol :

(May-18)

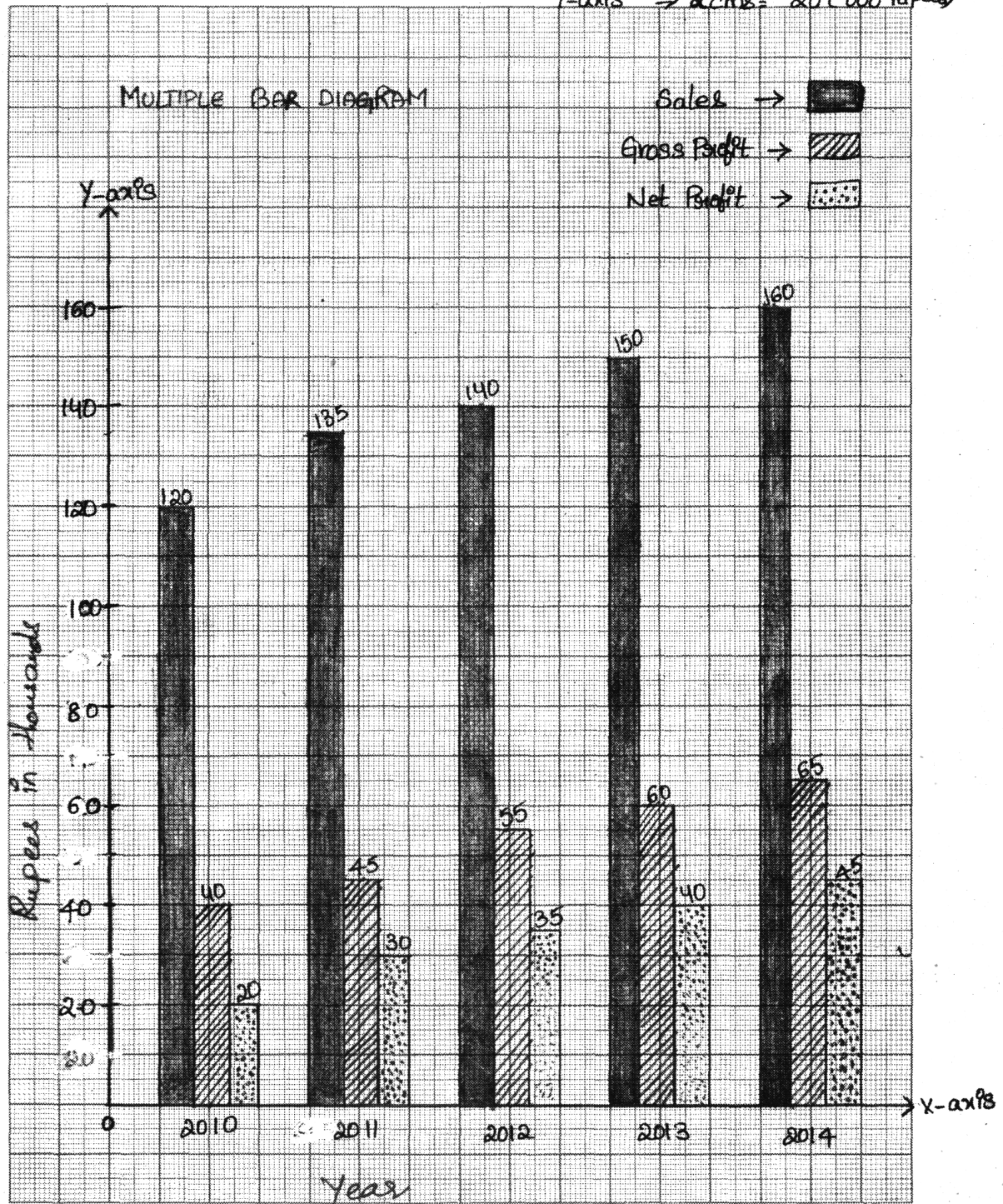
Year	Science	Arts	Commerce	Total
2015	240	560	220	1020
2016	280	610	280	1170



3. Draw a multiple bar diagram from the following data.

Year	Sales (000 `)	Gross Profit (000 `)	Net Profit (000 `)
2010	120	40	20
2011	135	45	30
2012	140	55	35
2013	150	60	40
2014	160	65	45

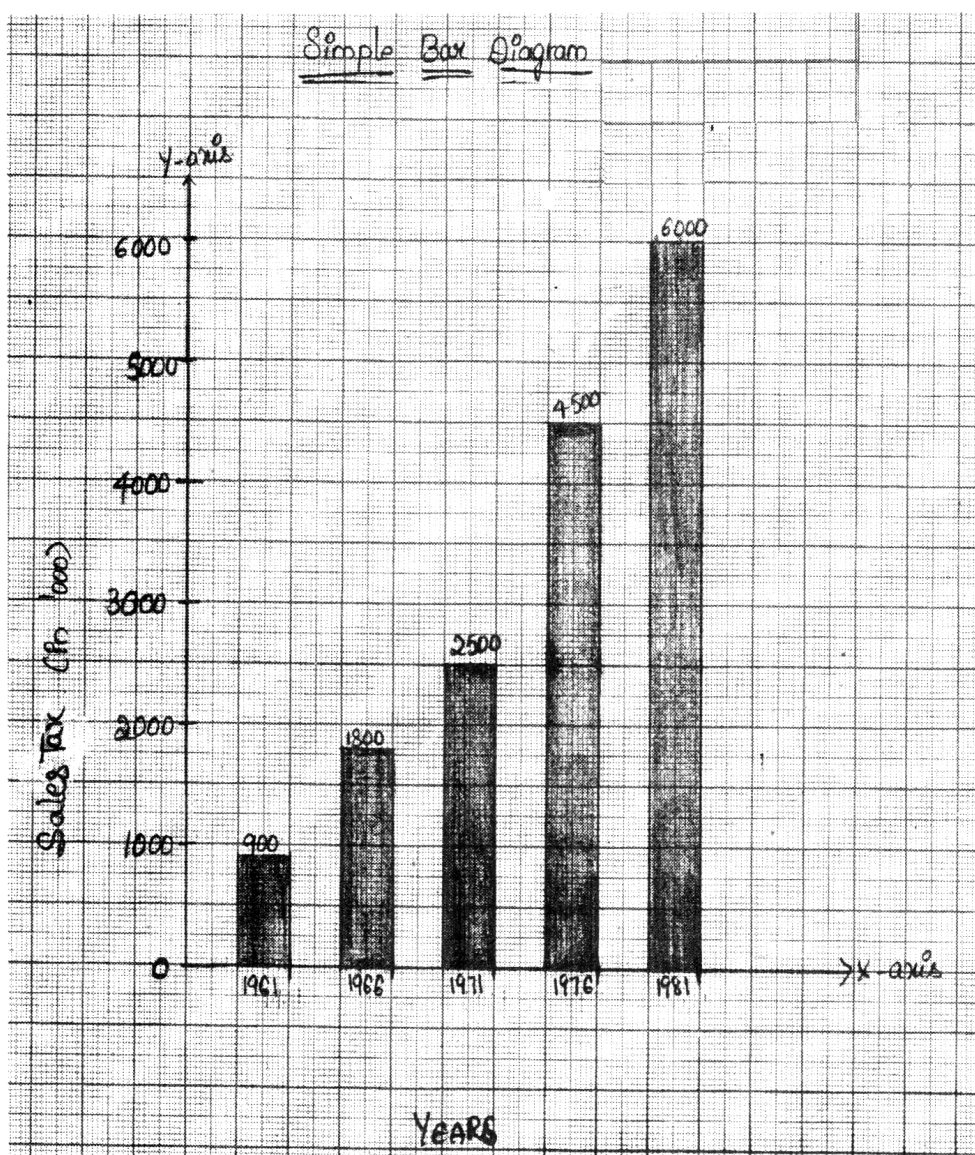
Sol :

Scale :X-axis $\Rightarrow 2 \text{ cms} = 1 \text{ year}$ Y-axis $\Rightarrow 2 \text{ cms} = 20 (\text{'000 rupees})$ 

4. Represent the below given data by a simple bar diagram :

Year	Sales Tax (in '000)
1961	900
1966	1800
1971	2500
1976	4500
1981	6000

Sol :



2.2.2 Two Dimensional Diagrams

Q10. Explain briefly two dimensional diagrams.

(OR)

What are two dimensional diagrams?

Ans :

(Dec.-18)

Meaning

Two dimensional (2D) diagrams are also referred as 'surface diagrams' (or) 'area diagrams'. These diagrams are constructed with respect to length and width of the given data,

Types

Some of the important 2-D diagrams are listed below.

1. Squares
2. Rectangles
3. Circles.

1. Squares

When variation between given numerical figures is high then choice of squares diagram is more suitable instead of rectangles diagram. Like rectangles diagram here given numerical figures are represented by areas of squares.

We know that area of a square = (side) $\times$ (side) = (side)². So, we take (side)² = given numerical figure $\Rightarrow$ side of square = $\sqrt{\text{given numerical figure}}$.

Remember that the base line would be same for all squares.

In other words, we follow the following steps for the construction of the square diagram:

Step 1: Take the given numerical observations/figures as areas of the corresponding squares.

Step 2: Take square roots of the given numerical observations/figures as sides of the corresponding squares.

Step 3: Construct the corresponding squares like rectangle diagrams.

2. Rectangles

In rectangles diagram given numerical figures are represented by areas of the rectangles.

Area of a rectangle = (length) $\times$ (breadth).

So, rectangles diagram is drawn by taking one of the two variables as lengths and another variable as breadths of the rectangles along two axes.

3. Circles

Another form of preparing the two dimensional diagram is circle diagram. As in square diagram we took given numerical figures/observations as the areas of the corresponding squares. Similarly, here we take given numerical figures/observations as areas of the corresponding circles.

Area (A) of a circle = πr^2 ,

where r is radius of the circle

Therefore, we follow the following steps for the construction of the circle diagram:

Step 1: Take the given numerical observations/figures as areas of the corresponding circles.

Step 2: Take squares of the radii (r^2) of the corresponding circles proportional to the given numerical figures/observations as sides of the corresponding squares.

Step 3: Take radii (r) of the corresponding circles proportional to the square roots of the given numerical figures/observations.

Step 4: Construct the corresponding circles like rectangles/squares diagrams.

Q11. Define pie diagram. Explain various steps to construct a pie diagram.

Ans :

(Imp.)

Pie diagram/chart is used when the requirement of the situation is to know the relationship between whole of a thing and its parts, i.e. pie chart provides us the information that how the entire thing is divided up into different parts.

Steps

Steps used for constructing a pie chart.

Step 1: Find the total of different parts.

Step 2: Find the sector angles (in degrees) of each part keeping in mind that total angle around the centre of a circle is of 360° .

Step 3: Find the percentage of each part taking the total obtained in step 1 as 100 percent.

Step 4: Draw a circle and divide it into sectors, where each sector (or area of the sector) of the circle with corresponding angles obtained in step 2 will represent the size of corresponding parts. Diagram thus obtained is nothing but pie chart fitted to the given data.

Q12. State the limitations of pie diagram.

Ans :

The following are the limitations of the pie diagram/chart:

- (i) For accurate reading and interpretation, particularly when data are divided into a large number of components or the difference among the values of components is very small, the pie diagram is less effective than the bar diagrams.
- (ii) Attractiveness of a pie chart suffers if the number of parts of the whole is more than 7 or 8. That is, pie chart should be avoided if number of parts of the whole is more than 7 or 8.
- (iii) In terms of comparison, the pie diagram appears inferior to simple bar diagram or divided bar diagram.
- (iv) Pie chart is used only when total of the parts make a meaningful whole.
- (v) Pie chart should not be used if observations of the different parts are not mutually exclusive.
- (vi) Pie chart should not be used if observations of the different parts are observed at different time.

Q13. Distinguish between One dimensional diagram and Two dimensional diagram.*Ans :***(June-19)****One Dimensional Diagram**

In this case only the length dimensional is given the importance : These diagrams are either bar or line diagrams.

Merits :

1. These are easy to construct
2. These are easy to understand
3. Comparison can be made easily by this device

Two Dimensional Diagram

As in single bars it was mentioned that the width of each bar should be equal for a certain variable or items. But in this case not only the length but the width also is taken proportionately in case of rectangles.

But where the items are represented in square terms we use the technique of squares or circles. These are also known as area diagrams.

PROBLEMS ON TWO DIMENSIONAL DIAGRAMS

5. Draw a suitable diagram to represent the following data relating to the production cost of manufacture.

Cost of material = ₹ 38,400

Cost of labour = ₹ 30,720

Direct expenses = ₹ 11,520

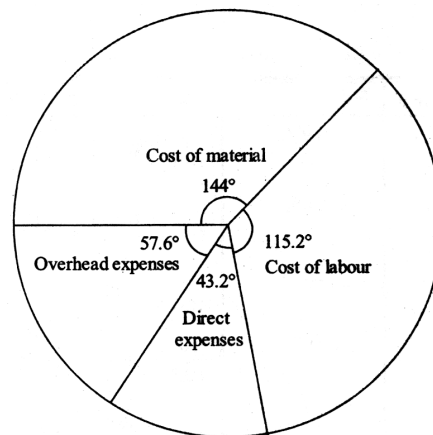
Overhead expenses = ₹ 15,360

Sol :

$$\begin{aligned} \text{Total Production Cost} &= \text{Cost of Material} + \text{Cost of Labour} + \text{Direct Expenses} + \text{Overhead Expenses} \\ &= 38,400 + 30,720 + 11,520 + 15,360 = 96,000. \end{aligned}$$

Computations for construction of pie diagram for the given data are as follows,

Item of Expenditure	Expenditure in (₹)	Expenditure in Percentage (%)	Expenditure in Degree (°)
Cost of material	38,400	$\frac{38400 \times 100}{96000} = 40\%$	$\frac{40 \times 360}{100} = 144^\circ$
Cost of labour	30,720	$\frac{30720 \times 100}{96000} = 32\%$	$\frac{32 \times 360}{100} = 115.2^\circ$
Direct expenses	11,520	$\frac{11520 \times 100}{96000} = 12\%$	$\frac{12 \times 360}{100} = 43.2^\circ$
Overhead expenses	15,360	$\frac{15360 \times 100}{96000} = 16\%$	$\frac{16 \times 360}{100} = 57.6^\circ$
Total	96,000	100%	360°



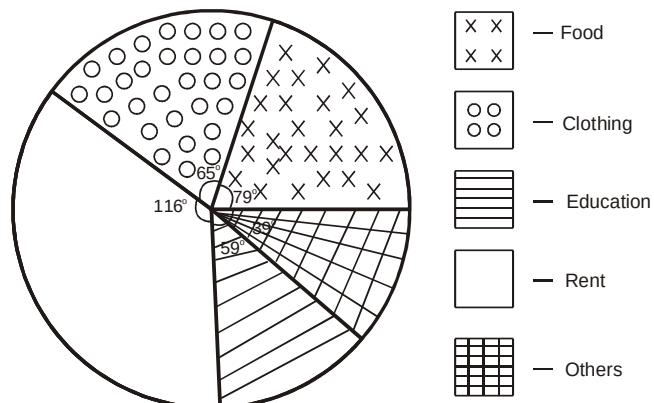
6. Represent the following data related to the monthly expenditure of a family through a Pie-Diagram.

Item	Food	Clothing	Rent	Education	Others
Monthly Expenditure in Rs.	4,800	4,000	7,200	3,600	2,400

Sol :

(June-19)

Items	Expenditure	Ratio or Angle
Food	4,800	$\frac{4,800}{22,000} \times 360 = 79^\circ$
Clothing	4,000	$\frac{4,000}{22,000} \times 360 = 65^\circ$
Rent	7,200	$\frac{7,200}{22,000} \times 360 = 118^\circ$
Education	3,600	$\frac{3,600}{22,000} \times 360 = 59^\circ$
Others	2,400	$\frac{2,400}{22,000} \times 360 = 39^\circ$
	22,000	360°



7. Draw a Pie diagram by showing the following of values :

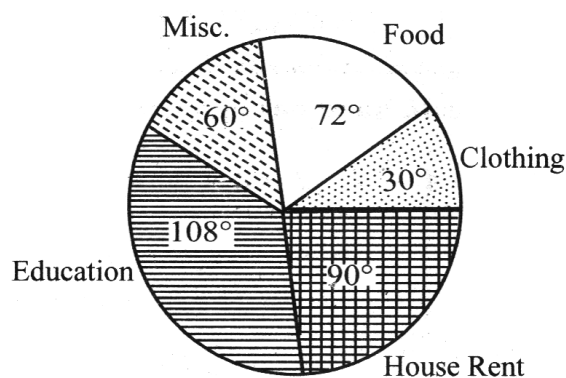
Item of Expenditure	₹
Food	12,000
Clothing	5,000
House Rent	15,000
Education	18,000
Misc.	10,000

Sol :

Calculation for Pie Chart

Items of Expenditure (a)	₹ (b)	Angle at the Centre (c) = $\frac{(b)}{60,000} \times 360^\circ$
Food	12,000	$\frac{12,000}{60,000} \times 360^\circ = 72^\circ$
Clothing	5,000	$\frac{5,000}{60,000} \times 360^\circ = 30^\circ$
House Rent	15,000	$\frac{15,000}{60,000} \times 360^\circ = 90^\circ$
Education	18,000	$\frac{18,000}{60,000} \times 360^\circ = 108^\circ$
Miscellaneous	10,000	$\frac{10,000}{60,000} \times 360^\circ = 60^\circ$
	60,000	360°

Pie diagram representing items of expenditure.



2.2.3 Pictograms

Q14. What do you mean by Pictograms? Explain the merits and demerits of Pictograms.

Ans :

Pictures are more attractive and appealing to the eye and have a lasting impression on the mind. Pictograms is the technique of presenting statistical data through appropriate pictures. It is a very popular device for presenting statistical facts to a layman who does not have any mathematical background. The magnitude of the particular phenomenon under study is of the pictures being proportional to the values of the different magnitudes to be presented.

Merits

Some of the advantages of pictographs are,

1. The main advantage of pictograph is its attractiveness.
2. Pictographs are more valuable and appealing to a large number of population.
3. They are helpful in creating images which cannot be forgotten quickly.
4. They are easy to understand and grasp.
5. They are self-explanatory in nature.
6. They are helpful in attracting the attention of people at exhibitions, malls, fairs etc, where huge crowd is present.

Demerits

Some of the demerits of pictographs are,

1. Constructing pictograph is a difficult task.
2. One of the major drawback of pictograph is, it does not give detailed information.

2.2.4 Cartograms

Q15. Explain briefly about Cartograms.

Ans :

- Representation of the numerical facts with the help of a map is known as cartogram.
- By representing the facts by maps, the impact of the results on different geographical areas may be shown and compared also.
- Maps are helpful in comparative study of various districts of a state or different states of a country.

For example, the production of wheat in different geographical areas can also be represented by cartogram. The quantities on the map can be shown in many ways, such as through shades or colours or by dots or by placing pictograms in each geographical area or by the appropriate numerical figure in each geographical area.

2.3 GRAPHIC PRESENTATION

2.3.1 Technique of Construction of Graphs

Q16. Define graph. Explain the Technique of Construction of Graphs.

Ans :

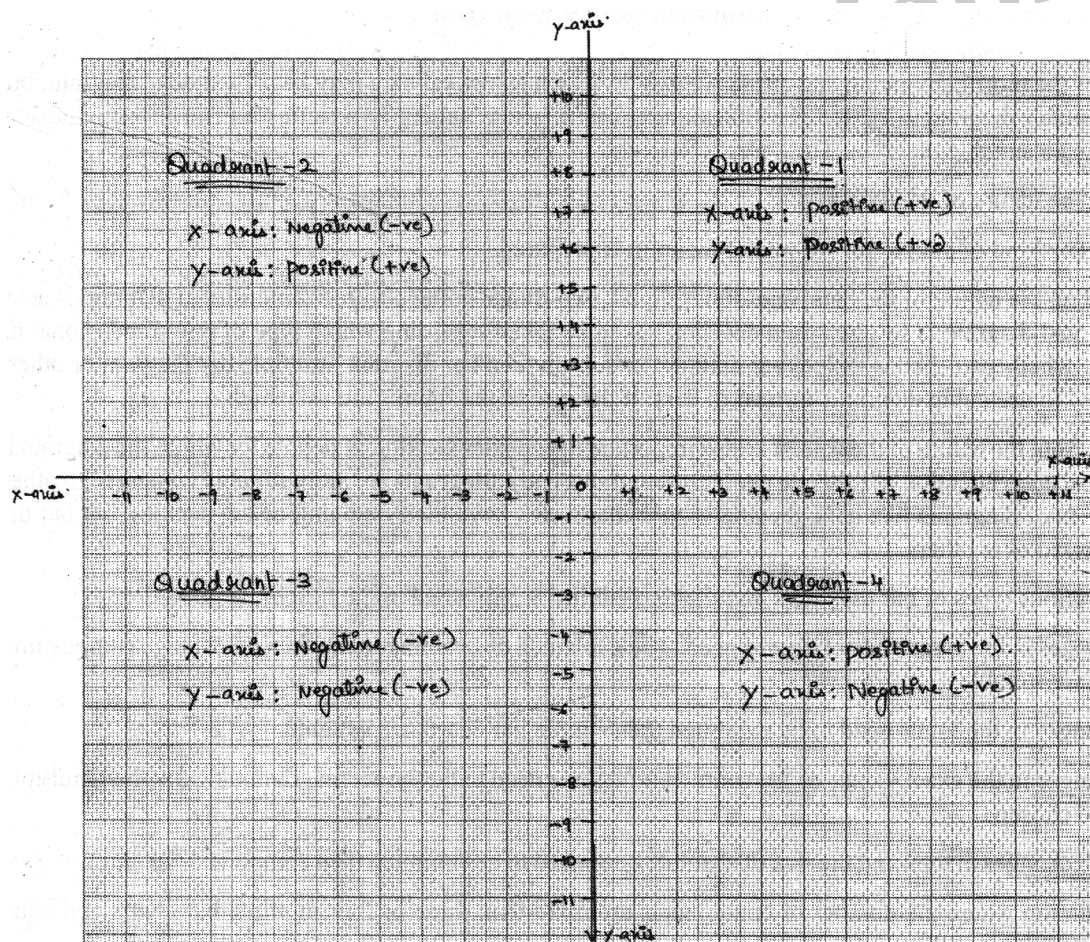
(July-21, Imp.)

Meaning

Graph or chart is defined as the pictorial representation of statistical data in graphical form. The statistical graphs are used to represent a set of data to make it easier to understand and interpret statistical information.

Graph paper is used for the construction of graphs. The graph sheet is an arrangement of horizontal and vertical intersecting lines. In any graph, firstly two simple lines are drawn intersecting each other. The intersecting point is called as origin or the 'zero' point and the two lines are called as 'coordinate axes', where the horizontal line is known as x-axis or 'abscissa' and the vertical line is known as Y-axis or 'ordinate'.

The following figure depicts a graph paper with X-axis and Y-axis and four quadrants.



In the above figure, the whole graph sheet is divided into four sections known as 'quadrants'. The point O is the origin point. In quadrant-1, both values of X-axis and Y-axis are positive. In Quadrant-2,

values of X-axis is negative but values of Y-axis are positive. In quadrant-3, values of both X-axis and Y-axis are negative. In quadrant -4, X-axis values are positive and Y-axis values are negative. Usually the independent variable is shown on X-axis and dependent variable is shown on Y-axis. The scale on X-axis and Y-axis may or may not be same.

In a graph paper, the scale on X-axis and Y-axis starts with zero as origin. The curve drawn on the graph is interpreted based on the distance between the curve and the base line. In case if series of changes presented from a norm of 100% then 100% is taken as a base line. There is no fixed standard regarding the scale to be taken on X-axis and Y-axis, it completely depends upon the paper size and the given data.

The last step in the construction of graph is plotting the given data by taking into consideration the values of X and Y. The various points plotted on the graph need to be joined with a straight line.

Q17. State the general rules for constructing graph.

Ans : (Imp.)

1. Neatness

One of the basic requirements of a graph is to present data in a manner that is visually pleasing. Neatness is of paramount importance in ensuring that the graph catches the attention of the intended audience.

2. Title

Every graph must have a clear and comprehensive title so that the user of the graph understands what is being presented by means of the graph.

3. Footnote

Any explanation that is necessary to explain and is not obvious needs to be explained with the help of a footnote. For example, if a graphs is representing the revenues of a corporate for the previous 5 years and the current year it may not catch the attention of the user that

revenues on the last 5 years are audited. But the current year revenues may be unedited. This needs to be explained with the help of a footnote. It is common to use special symbols such as *, #, \$ etc to indicate the place where a footnote is required.

4. Structural Framework

The position of the axes should be so chosen that the graph looks proportionate and attractive. In drawing the graph it is customary to plot the independent variable along the X-axis and the dependent variable along the Y-axis. For instance, if census details are being plotted on a graph paper, the time factor (Year) is taken along X-axis and the population of the country in different years is taken along Y-axis.

5. Scale

The scale along both the axes (X-axis and Y-axis) should be so chosen that the entire data can be accommodated in the available space without crowding. The ratio between the horizontal and vertical scales that need to be considered. The figure must be sufficiently small for the whole of it to be visible at once. If the range of data is very high. Minute accuracy can be sacrificed to some extent. All important variations must be clearly visible.

6. False Base Line

In a graph, the vertical scale normally starts with 0. However if the minimum value of the dependent variable is very far (greater than) from '0' and the fluctuations of the variable are relatively small, the fluctuations can be more effectively portrayed by using a false base line. In such a situation, the vertical scale is broken and the space between the origin '0' and the minimum value (or some convenient value near that) of the dependent variable is omitted drawing two zigzag horizontal lines above the base line. The scale along Y-axis is framed accordingly.

False base line technique is quite extensively used for magnifying the minor fluctuations in a time series data. It also saves

and enhances the look and feel of the graph. However, proper care should be taken to interpret graphs in which false base line is used.

7. Line Designs

If more than one variable is to be depicted on the same graph, the different graphs so obtained should be distinguished from each other by the use of different lines viz., dotted lines, broken lines, dash-dot lines, thin or thick lines etc., and index to identify them should be given.

8. Source Note

The source note states the source from where the underlying data presented in the graph has been collected. The source note adds credibility to the data presented.

9. Index

If multiple shades, designs or colors have been used in the graph, they need to be explained to the user with the help of a legend.

10. Simplicity

The graph presented should be simple such that a layman can also understand the underlying message without having any mathematical / statistical background.

11. Ratio or Logarithmic Scale

These scales are used to display proportional changes in data. In order to display proportional or relative changes in the magnitude, the ratio or logarithmic scale should, be used instead of natural or arithmetic scale is used to display absolute changes.

Q18. State the Advantages and Disadvantages of graphs.

Ans : (July-21)
Advantages

Following are the merits

1. A graph is more attractive and impressive than a table of figures

2. the help of graphs, comparison between two or more phenomena can be made very easy.
3. Since the data become visible at a glance in a graph, one can understand it easily and can study the tendency and fluctuations of data.
4. The impressions created by the figures presented in a tabular than those created by the figures presented in a tabular form.
5. Graphs are also helpful in interpretation, extra population and forecasting.
6. Correlation between two series can be studied easily with the help of graphs.
7. Certain statistical measures like, median, mode, quartiles etc can be determined by drawing the graphs of frequency distribution.
8. It needs no special knowledge of mathematics to understand a graph.
9. Apart from simplicity, it saves the time and energy of the statistician as well as observer.

Disadvantages

It suffers from the following limitations

1. Graphs may be misused by taking false scale
2. Since a curve shows tendency and fluctuations, actual values are not known
3. Accuracy is rather not possible in a graph
4. Graphs cannot be quoted in support of a statement
5. Only one or two characteristics can be depicted in a graph. If more features are shown, the graph becomes difficult to follow.

Q19. What are the differences between Diagrams and Graphs.

(OR)

Distinguish between Diagrams and Graphs.

Ans : (May-18, Imp.)

The diagrams and graphs are statistical techniques of representing the data. However, we can make the following distinctions between the two

Diagrams	Graphs
- Both plain paper and graph sheet can be used. - A diagram does not represent any mathematical relationship between two variable. - Diagrams are constructed for categorically data, including time series and spatial series. - Diagrams are more attractive to the eye and as such are better suited for publicity & propaganda. - Diagrams are not helpful in statistical analysis. - Median and mode cannot be estimated. - It is used for comparison only. - Data's are represented by bars. Rectangles.	- Graph sheets are used. - A graph represents mathematical relationship between two variables. - Graphs are more appropriate for representing time series and frequency distribution. - Graphs are less attractive to the eye. - Graphs are very much used in statistical analysis. - The value of median and mode can be estimated. - It represents a mathematical relationship between two variables. - Data are presented by points or dots line.

2.4 TYPES OF GRAPHS

Q20. What are the different types of graphs?

Ans :

Graphs can be broadly classified under two types. The two types of graphs are :

- Time series graphs (or) Historigrams
- Frequency Distribution graphs

1. Time series graphs or Historigrams

Time series graphs have time as an independent variable on the X-axis (horizontal axis) and dependent variables on the Y-axis (vertical axis), Two-or more variables dependent on time can be conveniently presented on the same graph provided the magnitudes do not vary a great deal. A suitable scale has to be devised for presenting all the variables.

2. Frequency Distribution graphs

Frequency graphs are designed to reveal clearly the characteristics features of a frequency data. In a frequency graph the size or the value of the item is presented on the horizontal axis and the frequency or the number of items on the vertical axis. The scale of measurement has to be indicated separately on both the axis. The following types of frequency graph are used:

- Histogram
- Frequency Polygon
- Smooth Frequency Curve
- Ogive

2.4.1 Graphs of Frequency Distribution

Q21. Define Graphs of Frequency Distribution. State the different types of Graphs of Frequency Distribution.

Ans :

Meaning

Frequency graphs are designed to reveal clearly the characteristics features of a frequency data. In a frequency graph the size or the value of the item is presented on the horizontal axis and the frequency or the number of items on the vertical axis. The scale of measurement has to be indicated separately on both the axis.

The following types of frequency graph are used:

- Histogram
- Frequency Polygon
- Smooth Frequency Curve
- Ogive

2.4.1.1 Histogram

Q22. Define Histogram. State the rules for construction of Histogram.

Ans :

(Imp.)

Meaning

A histogram is a graphical method of presenting data. Where the observations are located on a horizontal axis and the frequency of those observations is depicted along the vertical axis. It is one of the most popular and commonly used devices for charting continuous frequency distributions.

Construction of Histogram

Construction of a histogram involves erecting a series of adjoining vertical rectangles on the sections of the horizontal axis (X-axis) with bases (sections) equal to the width of the corresponding class intervals and heights proportional to the corresponding frequencies of each class. The length and breadth of the rectangles should reflect the class intervals and corresponding frequencies. Construction of Histogram will vary depending on whether the class Intervals are equal or unequal.

(i) When class intervals are equal

Each class interval is drawn on the x-axis by a base proportional to the magnitude of the class interval. On each class interval base, erect a rectangle with height proportional to the corresponding frequency of the class. The series of adjacent rectangles so formed gives the histogram of the frequency distribution and its area represents the total frequency of the distribution as distributed across the different classes.

(ii) When Class Intervals are Unequal

In case of unequal class intervals, if the "rectangles are erected on different class intervals as explained above the areas would not be proportional to the frequency and will give undue weight age to those having larger class intervals. In order to correct this situation, if class intervals are unequal, they should be converted into equal class intervals.

2.4.1.2 Frequency Polygon

Q23. Define Frequency Polygon. What are the different ways of Frequency Polygon.

Ans :

A frequency polygon is a graph representing frequency distribution. It has more than four sides and is very helpful in comparing several frequency distributions plotted on the same graph. This feature of frequency polygon is an advantage over histogram.

Different Ways to Draw/Construct a Frequency Polygon

The frequency polygon can be constructed in two ways. They are,

- (a) Using histograms
- (b) Using midpoints.

(a) Using Histograms

A histogram is drawn by using the given data. Identify the midpoints of the upper horizontal side of the rectangles. Join the midpoints by a line. The figure thus obtained is known as

'frequency polygon'. The area of the frequency polygon is approximated by joining the ends of the polygon with the base line and by assuming that both the frequencies of the class before the first class and the class after the last class are zero.

(b) Using Midpoints

Calculate the midpoints of the various class intervals. Plot the midpoints and their corresponding frequencies on the graph. Join these points by a straight line. The figure thus formed is called 'frequency polygon'.

Q24. What are the advantages of frequency polygon over histogram ?

Ans :

The advantages of frequency polygon over histogram are as follows :

1. Frequency polygon is more simple when compared to histogram.
 2. Compared to histogram, frequency polygon gives a clear outline of the data.
 3. With an increase in the number of classes and number of observations, the frequency polygon becomes more and more smoother and curved.
 4. In case of frequency polygon, several frequency distributions can be located on the same axis which makes some comparisons easy. But this is not possible in case of histograms, we need to draw separate graphs for every distribution in order to compare histograms. In order to compare frequency distributions graphically, frequency polygons are chosen over histograms.
-

2.4.1.3 Frequency Curve

Q25. Define Frequency Curve. What are the different ways of Frequency Curve.

Ans :

Frequency curve is a smooth free hand curve drawn through the vertices of frequency polygon. It is a smooth free hand curve drawn to fit into the histogram in a manner that it passes through the tops of rectangular blocks. The area included under the curve is approximately the same as that of the polygon.

The basic drawing a frequency curve is to present graphically the area covered by a histogram in a more symmetrical fashion. If the frequency curve is correctly drawn, it can be used for interpolation purposes.

Types of Frequency Curves

There are different types of frequency curves:

1. Symmetrical bell-shaped

The curve is bell shaped and can be bisected at the center to result in two equal sides mirroring each other.

2. Asymmetrical bell shaped

The curve is bell shaped but cannot be bisected at the center. The peak is either towards the left or right of the center. If the peak is towards the left, it is termed as positively skewed. If the peak is towards the right, it is termed as Negatively skewed

3. Upward Sloping or J-shaped

This curve is not bell shaped. It keeps rising as we move from left to right. It loosely resembles the shape of the alphabet.

4. Downward Sloping or Reverse J-shaped

This curve is also not bell shaped. It peaks in the beginning and then slowly keeps slopping down as we move from left to right. It loosely resembles the mirror image of the alphabet.

5. U Shaped

As the name suggests, it loosely resembles the shape of the alphabet U. It is a downward sloping curve to begin with but gets converted into an upward sloping curve after some time.

6. Bi-modal Curve

This kind of Curve is obtained when there are two peak points (may not be equal) and frequencies are concentrated around them.

7. Multi-modal (or Mixed)

This kind of curve is obtained when there are multiple peak points (not necessarily equal) and frequencies are concentrated around these multiple points.

2.4.1.4 Ogive Curve**Q26. Explain briefly about Ogive Curves.**

Ans :

(Imp.)

Ogive is a graphic representation of a cumulative frequency of a distribution. It is also known as Cumulative Frequency Polygon. It is pronounced as 'ogive'. The ogive is drawn by joining the cumulative frequencies plotted against the corresponding class boundaries.

There are two types of ogives. It can be of 'less than' or 'more than' type based on whether the cumulative frequencies used are of 'less than' or more than 'type.

1. 'Less than' Ogive

The "less than ogive" like an elongated S which starts from the lowest class boundary on the horizontal axis and gradually rises upward, ending at the highest class boundary corresponding to the total frequency of the distribution. To construct this curve, we plot the 'less than' cumulative frequencies against the upper class boundaries of the respective classes. The points plotted are joined by a smooth free hand curve. It is upward sloping from left to right.

2. 'More than' Ogive

The "more than ogive" has the appearance of an elongated S turned upside down. It starts from the lowest class boundary on the horizontal axis and gradually slopes downward, ending at the highest class boundary corresponding to the total frequency of the distribution. To construct this curve, we plot the 'more than' cumulative frequencies against the lower class boundaries of the respective classes. The points plotted are joined by a smooth free hand curve. It is downward sloping from left to right.

Ogive is very useful for graphic computation of partition values, viz, Median, Quartiles, Deciles, Percentiles, etc. of which we will learn in the forthcoming chapters. It is alright to draw both the 'less than' ogive and 'more than' ogive on the same graph. If done so, they intersect at a point. A perpendicular from the point of intersection on the X-axis gives the value of median.

Ogive is used for estimating cumulative frequencies of values falling within the range of values represented on a graph. For example one can determine the total number of assesses paying not less than a certain sum of money, or the number of workers who are receiving wages that are not more than a certain amount, etc.

Ogive can be used with advantage over frequency curves. For each of the distribution, different ogive can be constructed on the same graph and they are generally less over-lapping than the corresponding frequency curves.

Q27. How is a Histogram different from a Bar Diagram?

Ans :

(June-19, Imp.)

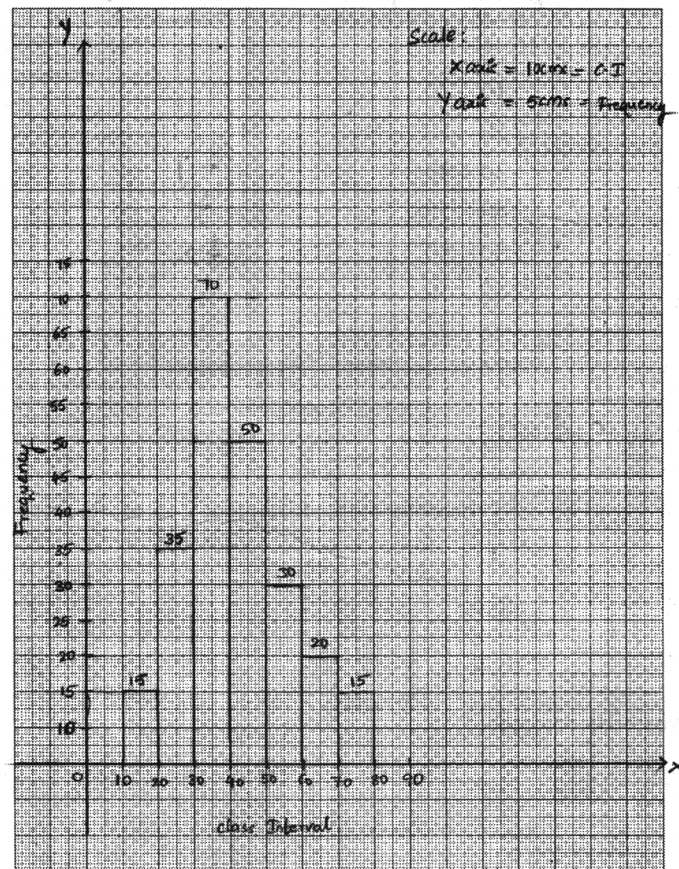
S.No.	Basis	Histogram	Bar graph
1.	Meaning	Histogram refers to a graphical representation that displays data by way of bars to show the frequency of numerical data.	Bar graph is a pictorial representation of data that uses bars to compare different categories of data.
2.	Indicates	Distribution of non-discrete variables.	Comparison of discrete variables.
3.	Presents	Quantitative data	Categorical data
4.	Spaces	Bars touch each other hence there are no spaces between bars.	Bars do not touch each other, hence, there are spaces between bars.
5.	Elements	Elements are grouped together, so that they are considered as ranges	Elements are taken as individual entities.
6.	Can bars be recorded	No	Yes
7.	Width of bars	Need not to be same	Same

PROBLEMS ON FREQUENCY DISTRIBUTION GRAPHS

8. Construct Histogram for the following distribution.

Class Interval	Frequency
10 – 20	15
20 – 30	35
30 – 40	70
40 – 50	50
50 – 60	30
60 – 70	20
70 – 80	15

Sol :

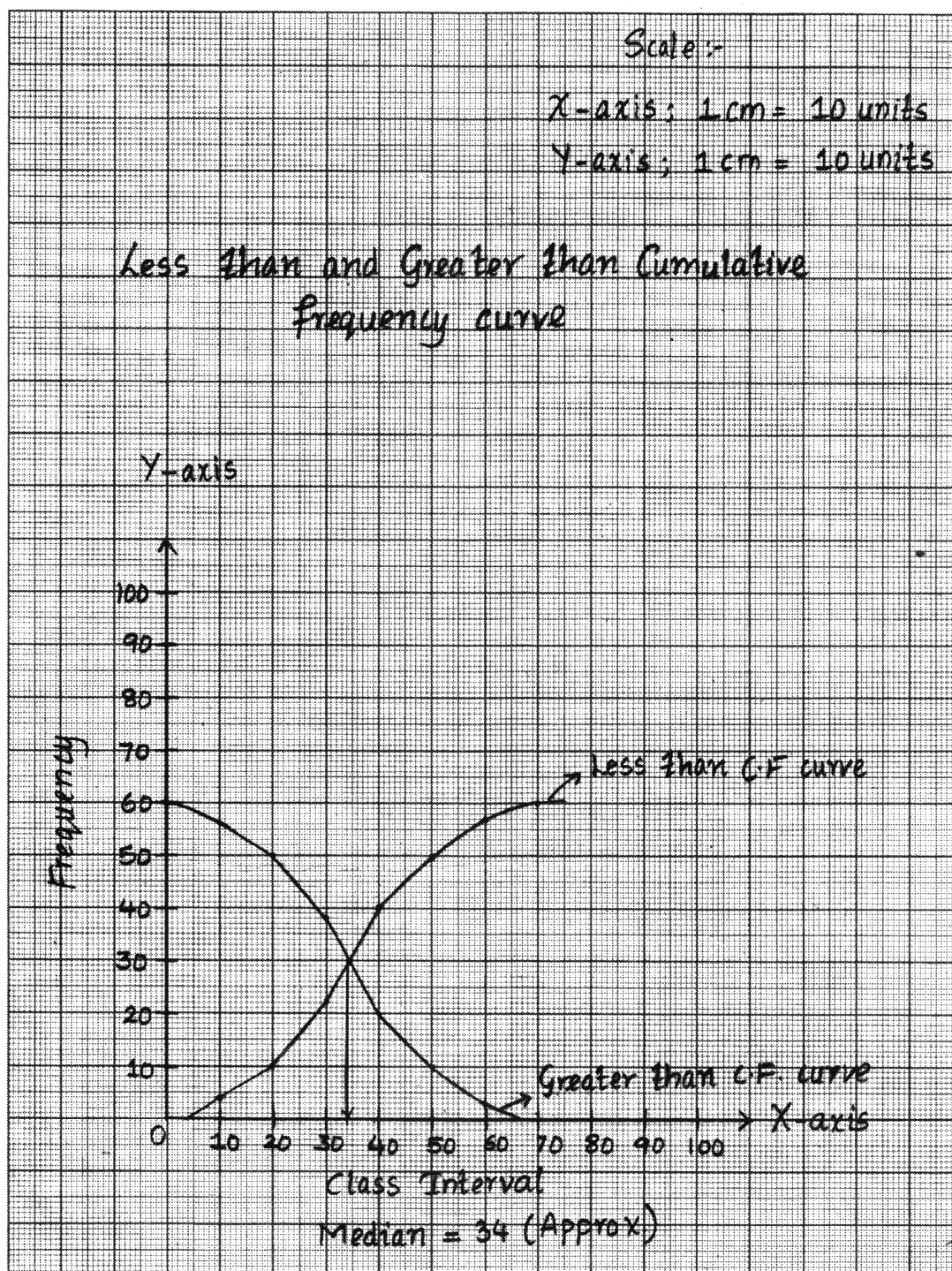


9. Draw less than cumulative frequency curve and greater than cumulative frequency curve for the following data and also find median with the help of the diagram.

C.I.	0 - 10	10 - 20	20 - 30	30 - 40	40 - 50	50 - 60	60 - 70
Frequency	4	6	12	18	10	7	3

Sol :

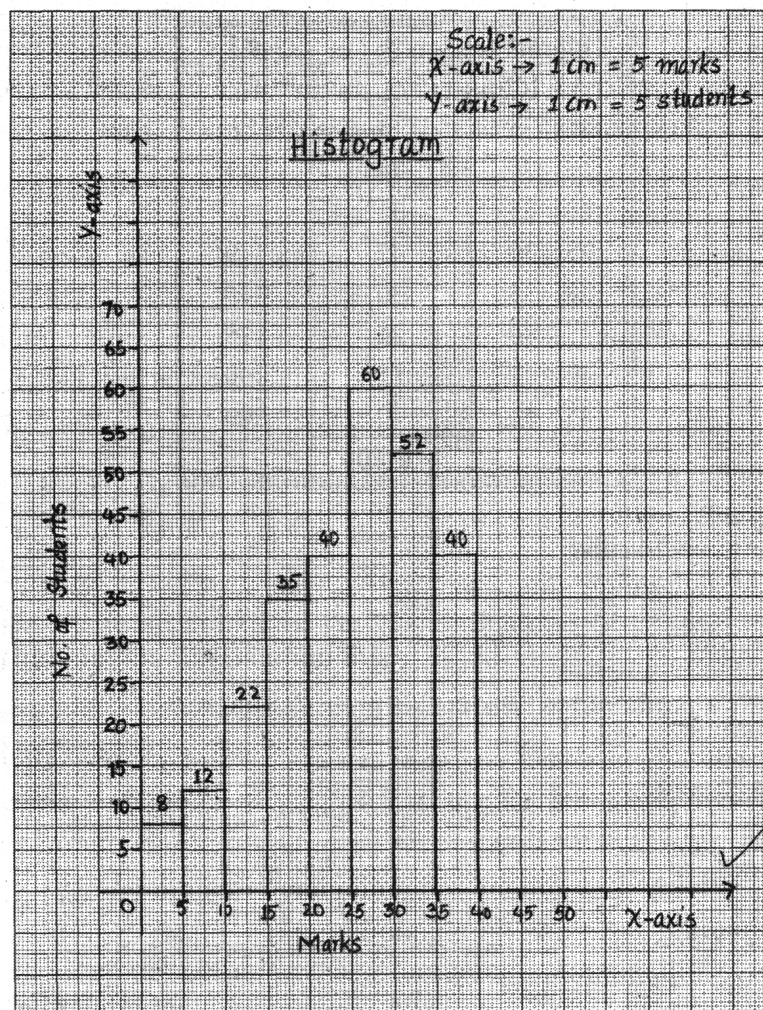
C.I	Frequency (F)	Less than C.F	Greater than C.F
0 - 10	4	4	$54 + 4 = 60$
10 - 20	6	$4 + 6 = 10$	$50 + 6 = 56$
20 - 30	12	$10 + 12 = 22$	$38 + 12 = 50$
30 - 40	18	$22 + 18 = 40$	$20 + 18 = 38$
40 - 50	10	$40 + 10 = 50$	$10 + 10 = 20$
50 - 60	7	$50 + 7 = 57$	$3 + 7 = 10$
60 - 70	3	$57 + 3 = 60$	3



10. Present the following data by Histogram.

Marks	0 - 5	5 - 10	10 - 15	15 - 20	20 - 25	25 - 30	30 - 35	35 - 40
No. of Students	8	12	22	35	40	60	52	40

Sol :



11. From the following data draw 'less than' and 'more than' ogives.

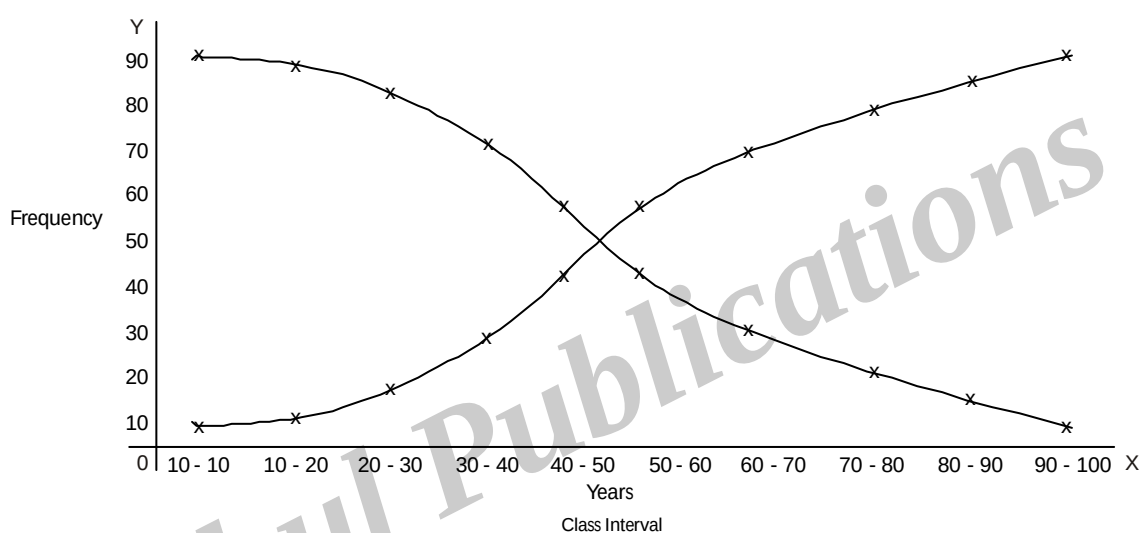
Marks	0-10	10-20	20-30	30-40	40-50	50-60	60-70	70-80	80-90	90-100
No. of Students	5	4	8	12	16	15	10	8	5	2

Sol :

(May-18)

Class interval	Frequency	Less than cumulative	Greater than cumulative frequency
0 - 10	5	5	85
10 - 20	4	9	80
20 - 30	8	17	76

30 – 40	12	29	68
40 – 50	16	45	56
50 – 60	15	60	40
60 – 70	10	70	25
70 – 80	8	78	15
80 – 90	5	83	7
90 – 100	2	85	2



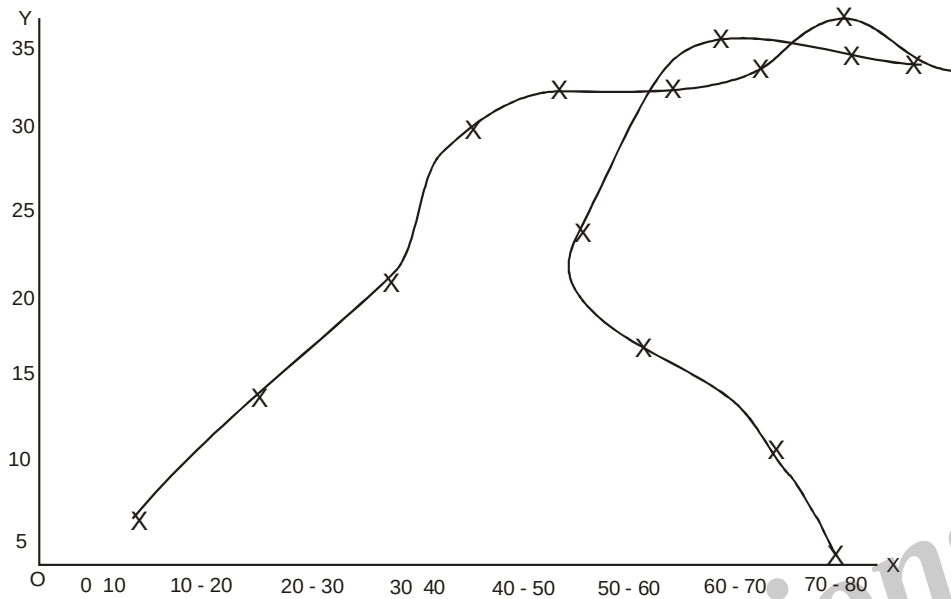
12. From the following data construct Ogive Curves and locate Median.

Profit in Lakhs Rs.	0 – 10	10 – 20	20 – 30	30 – 40	40 – 50	50 – 60	60 – 70	70 – 80
No. of Companies	3	8	12	16	11	10	7	2

Sol :

(Dec.-18)

C.I	F	Less than C.F	More than C.F
0 – 10	3	3	69
10 – 20	8	11	66
20 – 30	12	23	58
30 – 40	16	39	46
40 – 50	11	50	30
50 – 60	10	60	19
60 – 70	7	67	9
70 – 80	2	69	2



Curve Intersect in the class 50 – 60

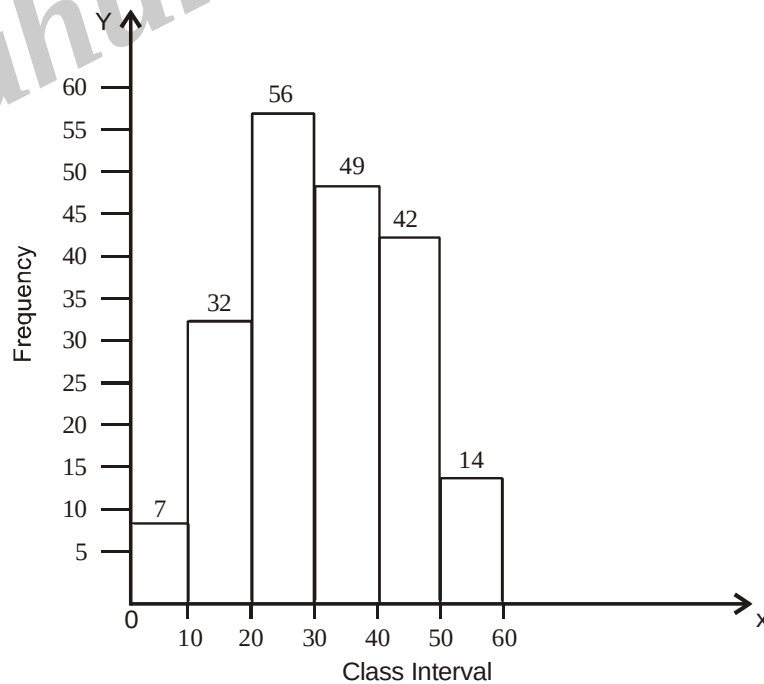
Median lies in the class 50 - 60

13. Prepare a Histogram for the following data.

Sales	0 – 10	10 – 20	20 – 30	30 – 40	40 – 50	50 – 60
No. of Co.s	7	32	56	49	42	14

Sol :

(June-19)



2.4.2 Graphs of Time Series or Histograms

Q28. What do you mean by time series? Explain different types of time series graphs.

Ans :

Time series graphs have time as an independent variable on the X-axis (horizontal axis) and dependent variables on the Y-axis (vertical axis). Two or more variables dependent on time can be conveniently presented on the same graph provided the magnitudes do not vary a great deal. A suitable scale has to be devised for presenting all the variables.

Depending on the type of data and the number of variables. Time Series graphs can be further classified into

- Horizontal Line Graphs
- Component or Band Graphs
- Silhouette or Net balance Graphs
- Range or variation Graphs and
- Z-curve in this case

2.4.2.1 Horizontal Line Graph

Q29. Explain briefly about Horizontal Line Graph.

(OR)

Define false base line.

Ans : (June-19)

Horizontal line graphs can be represented by one variable and two or more variables.

- **Historiogram/Horizontal Line Graph with One Variable**

These graphs represent only one variable graphically. Here the time factor is taken along the x-axis and other variable i.e., size of the study is taken along the y-axis on an appropriate scale. The points on the graph are connected with the straight line.

False Base Line

The vertical scale must start with zero for representing the given variable effectively, when the fluctuations are maximum and the starting values are very far from zero. Hence, the false base line is used to magnify the vertical scale and to plot the values.

In a false base line, the vertical scale is broken into two parts and the values of the dependent variable from zero to the lowest value are omitted by drawing two zig-zag horizontal lines above the base line i.e., X-axis.

Objectives

The following are the main objectives behind drawing false base line,

1. To identify the fluctuations in the given data easily.
2. To avoid wastage of space in the graph paper.
3. To provide clear visual communication through graph.

➤ Historiogram/Horizontal Line Graphs with Two or More Variables

The time series data of two or more related variables such as variables which can be determined in the same units and belongs to the same period of time can be easily plotted together in the same graph by using the same scale for all the variables along the y-axis and same scale for time period along the x-axis for each variable. This method is same as historiogram of one variable. Hence, we get curves for each variable. By using various types of lines such as thin, thick, dotted, dash, dash-dotted lines etc, one curve can be differentiated from another. Thus, a list of this effect must be provided for identification of curves properly.

2.4.2.2 Band Graphs

Q30. Define Band graphs.

Ans :

A Band graph is prepared when a variable has several components. Hence Band graph is also known as Component graph. The given data is

converted into cumulative data and then the constituent parts of the variable are plotted one over the other. They are distinguished by different shades. Band graph is also known as Belt curve.

2.4.2.3 Net Balance Graphs / Silhouette

Q31. Explain in detail Net Balance Graphs.

(OR)

Define Silhouette.

Ans :

Meaning

This graph is used when two different but related variables need to be presented with clear emphasis on the difference between the two.

Method-1

The difference between two sets of the variables for different time periods is obtained. In some cases, differences may have negative value.

Apart from the two histograms, another histogram is to be drawn on the same graph (i.e., Net balance). A part of the graph will be below the x-axis which is corresponding to negative values of the net balance.

Method-2

In this method, two histograms are drawn from one variable each on the same graph. The difference between the two variables is indicated by filling or shading the gap properly which expresses the positive and negative balances.

2.4.2.4 Range Graph

Q32. Explain briefly about Range Graph.

Ans :

Meaning

Range is the deviation that exists between maximum and minimum values of the variables. It indicates the difference between the two extreme values of variables. Range graph is also known as zone graph or variation graph, which is used to display and focus on the range of variation of a variable for each period.

Example

The range of variation is pointed out in the following cases,

1. Temperature on different days.
2. Readings of blood pressure on different days of an individual.
3. Price of a commodity on different time periods etc.

The range or zone graph is most suitable for time series and it helps the users to have an idea regarding the changes in the size of the variables under study. The range graph or zone graph or variation graph can be constructed in any one of the following methods:

Method-1

In this method, the minimum and maximum values of the variable of each period are plotted and joined with straight lines in order to get the range lines. Average value of each variable (midpoint) is plotted and joined with straight lines to attain the range graph. Hence, the range graph indicates the minimum, maximum and average value of the variable for each time period.

Method-2

This method involves plotting of the two histograms on the graph i.e.,

1. With respect to the maximum values of the variable for different periods and
2. With respect to the minimum values.

2.4.2.5 Z-Curve

Q33. Define Z-Curve.

Ans :

Meaning

A Z-chart is a special kind of graph where figures are plotted at regular intervals over a period of one year. It is slightly complicated by the fact that it incorporates three graphs on the same chart and these curves or lines taken together tend to look like the letter Z. The easiest approach is to build up the chart line by line.

Short Question and Answers

1. Distinguish between Diagrams and Graphs.

Ans :

Diagrams	Graphs
- Both plain paper and graph sheet can be used. - A diagram does not represent any mathematical relationship between two variable. - Diagrams are constructed for categorically data, including time series and spatial series. - Diagrams are more attractive to the eye and as such are better suited for publicity & propaganda. - Diagrams are not helpful in statistical analysis.	- Graph sheets are used. - A graph represents mathematical relationship between two variables. - Graphs are more appropriate for representing time series and frequency distribution. - Graphs are less attractive to the eye. - Graphs are very much used in statistical analysis.

2. What are the functions of Diagrams?

Ans :

Following points highlights the functions of diagrams.

- (i) Diagrams give a clear picture of the data, which can be understood by a common man easily and quickly.
- (ii) By using diagrammatic presentation data can be short end.
- (iii) Diagrams assist in locating median and mode values.
- (iv) Diagrammatic presentation detect computation errors.
- (v) They can be used for carrying out investigation.

3. What are two dimensional diagrams?

Ans :

Two dimensional (2D) diagrams are also referred as 'surface diagrams' (or) 'area diagrams'. These diagrams are constructed with respect to length and width of the given data,

Types

Some of the important 2-D diagrams are listed below.

- (i) Squares
- (ii) Rectangles
- (iii) Circles.

(i) Squares

When variation between given numerical figures is high then choice of squares diagram is more suitable instead of rectangles diagram. Like rectangles diagram here given numerical figures are represented by areas of squares.

We know that area of a square = (side) $\times$ (side) = (side)². So, we take (side)² = given numerical figure $\Rightarrow$ side of square = $\sqrt{\text{given numerical figure}}$.

Remember that the base line would be same for all squares.

In other words, we follow the following steps for the construction of the square diagram:

Step 1: Take the given numerical observations/figures as areas of the corresponding squares.

Step 2: Take square roots of the given numerical observations/figures as sides of the corresponding squares.

Step 3: Construct the corresponding squares like rectangle diagrams.

(ii) Rectangles

In rectangles diagram given numerical figures are represented by areas of the rectangles.

Area of a rectangle = (length) $\times$ (breadth).

So, rectangles diagram is drawn by taking one of the two variables as lengths and another variable as breadths of the rectangles along two axes.

(iii) Circles

Another form of preparing the two dimensional diagram is circle diagram. As in square diagram we took given numerical figures/observations as the areas of the corresponding squares. Similarly, here we take given numerical figures/observations as areas of the corresponding circles.

Area (A) of a circle = πr^2 ,
where r is radius of the circle

4. Define false base line.

Ans :

The vertical scale must start with zero for representing the given variable effectively, when the fluctuations are maximum and the starting values are very far from zero. Hence, the false base line is used to magnify the vertical scale and to plot the values.

In a false base line, the vertical scale is broken into two parts and the values of the dependent variable from zero to the lowest value are omitted by drawing two zig-zag horizontal lines above the base line i.e., X-axis.

5. Importance of diagrams.

Ans :

- (i) Diagrams reduce huge amount of data into simple figures. Thus, it is the best option for making short duration presentations to top management, Government, customer, etc who cannot go into details.
- (ii) Diagrams bring out the essence of the underlying data with great clarity. A person going through the data may get lost in detail, but will get the message immediately if the same data is presented as a diagram
- (iii) Diagrams help in highlighting the trends in underlying data that cannot be easily spotted by going through the data.
- (iv) Diagrams create a positive and lasting impression. The audience can remember the details for a longer time.
- (v) Diagrams have the ability to reach a greater audience people who cannot read can also understand diagrams.
- (vi) Diagrams facilitate comparison of data. Interesting conclusions can be drawn by merely juxtaposing two figures.

6. Explain the different types of diagrams.

Ans :

(i) One Dimensional Diagrams

One-dimensional diagram means the construction of a diagram on the basis of

one dimension. This could be height in case of vertical diagrams (or) length in case of horizontal diagrams.

One dimensional diagrams can be line diagrams or bar diagrams.

(ii) Two-Dimensional Diagrams

In two-dimensional diagrams, both length and width of the observations are represented by bars. The area of a two-dimensional diagram is equal to the product of its length and width. So, two-dimensional diagrams are also called 'area diagrams'.

(iii) Three-Dimensional Diagrams

Three dimensional diagrams are diagrams in which three dimensions, namely, length breadth and height are taken into account. They are also termed as volume diagrams. The objective of presenting data in 3 dimensions is to highlight the magnitude of the data through volumes of the corresponding diagrams. Diagrams in the shape of cubes, spheres, cylinders, blocks etc are commonly used. Three dimensional diagrams are specially useful if there are very wide variations between the smallest and the largest magnitudes.

(iv) Pictograms

Pictures are more attractive and appealing to the eye and have a Tasting impression on the mind. Pictograms is the technique of presenting statistical data through appropriate pictures. It is a very popular device for presenting statistical facts to a layman who does not have any mathematical background. The magnitude of the particular phenomenon under study is of the pictures being proportional to the values of the different magnitudes to be presented.

(v) Cartograms

Cartograms are used to present data pertaining to geographical regions with the help of maps. For example, the distribution of rainfall in different regions of a country can be shown with the help of maps or

cartograms. The different geographical regions are depicted on a map and the quantum of rainfall in the regions may be shown by dots, different shades or colours etc. Cartograms are simple and elementary forms of visual presentation and are easy to understand. They are generally used when the regional or geographic comparisons are to be highlighted.

7. Limitations of pie diagram.

Ans :

The following are the limitations of the pie diagram/chart:

- (i) For accurate reading and interpretation, particularly when data are divided into a large number of components or the difference among the values of components is very small, the pie diagram is less effective than the bar diagrams.
- (ii) Attractiveness of a pie chart suffers if the number of parts of the whole is more than 7 or 8. That is, pie chart should be avoided if number of parts of the whole is more than 7 or 8.
- (iii) In terms of comparison, the pie diagram appears inferior to simple bar diagram or divided bar diagram.
- (iv) Pie chart is used only when total of the parts make a meaningful whole.

8. What do you meant by Pictograms?

Ans :

Pictures are more attractive and appealing to the eye and have a Tasting impression on the mind. Pictograms is the technique of presenting statistical data through appropriate pictures. It is a very popular device for presenting statistical facts to a layman who does not have any mathematical background. The magnitude of the particular phenomenon under study is of the pictures being proportional to the values of the different magnitudes to be presented.

9. Explain briefly about Cartograms.

Ans :

- Representation of the numerical facts with the help of a map is known as cartogram.
- By representing the facts by maps, the impact of the results on different geographical area may be shown and to be compared also.
- Maps are helpful in comparative study of various districts of a state or different states of a country.

For example, the production of wheat in different geographical areas can also be represented by cartogram. The quantities on the map can be shown in many ways, such as through shades or colours or by dots or by placing pictograms in each geographical area or by the appropriate numerical figure in each geographical area.

10. State the Advantages and Disadvantages of graphs.

Ans :

Advantages

Following are the merits

- (i) A graph is more attractive and impressive than a table of figures.
- (ii) the help of graphs, comparison between two or more phenomena can be made very easy.
- (iii) Since the data become visible at a glance in a graph, one can understand it easily and can study the tendency and fluctuations of data.
- (iv) The impressions created by the figures presented in a tabular than those created by the figures presented in a tabular form.
- (v) Graphs are also helpful in interpretation, extra population and forecasting.
- (vi) Correlation between two series can be studied easily with the help of graphs.

- (vii) Certain statistical measures like, median, mode, quartiles etc can be determined by drawing the graphs of frequency distribution.
- (viii) It needs no special knowledge of mathematics to understand a graph.
- (ix) Apart from simplicity, it saves the time and energy of the statistician as well as observer.

Disadvantages

It suffers from the following limitations

- (i) Graphs may be misused by taking false scale.
- (ii) Since a curve shows tendency and fluctuations, actual values are not known.
- (iii) Accuracy is rather not possible in a graph.
- (iv) Graphs cannot be quoted in support of a statement.

11. Define Histogram.

Ans :

A histogram is a graphical method of presenting data. Where the observations are located on a horizontal axis and the frequency of those observations is depicted along the vertical axis. It is one of the most popular and commonly used devices for charting continuous frequency distributions.

Construction of Histogram

Construction of a histogram involves erecting a series of adjoining vertical rectangles on the sections of the horizontal axis (X-axis) with bases (sections) equal to the width of the corresponding class intervals and heights proportional to the corresponding frequencies of each class. The length and breadth of the rectangles should reflect the class intervals and corresponding frequencies. Construction of Histogram will vary depending on whether the class intervals are equal or unequal.

(i) When class intervals are equal

Each class interval is drawn on the x-axis by a base proportional to the magnitude of the class interval. On each class interval base, erect a rectangle with height proportional

to the corresponding frequency of the class. The series of adjacent rectangles so formed gives the histogram of the frequency distribution and its area represents the total frequency of the distribution as distributed across the different classes.

(ii) When Class Intervals are Unequal

In case of unequal class intervals, if the "rectangles are erected on different class intervals as explained above the areas would not be proportional to the frequency and will give undue weight age to those having larger class intervals. In order to correct this situation, if class intervals are unequal, they should be converted into equal class intervals.

12. Define Frequency Polygon.

Ans :

A frequency polygon is a graph representing frequency distribution. It has more than four sides and is very helpful in comparing several frequency distributions plotted on the same graph. This feature of frequency polygon is an advantage over histogram.

Different Ways to Draw/Construct a Frequency Polygon

The frequency polygon can be constructed in two ways. They are,

- (a) Using histograms
- (b) Using midpoints.

(a) Using Histograms

A histogram is drawn by using the given data. Identify the midpoints of the upper horizontal side of the rectangles. Join the midpoints by a line. The figure thus obtained is known as 'frequency polygon'. The area of the frequency polygon is approximated by joining the ends of the polygon with the base line and by assuming that both the frequencies of the class before the first class and the class after the last class are zero.

(b) Using Midpoints

Calculate the midpoints of the various class intervals. Plot the midpoints and their corresponding frequencies on the graph. Join these points by a straight line. The figure thus formed is called 'frequency polygon'.

13. Ogive Curves.

Ans :

(Imp.)

Ogive is a graphic representation of a cumulative frequency of a distribution. It is also known as Cumulative Frequency Polygon. It is pronounced as 'ogive'. The ogive is drawn by joining the cumulative frequencies plotted against the corresponding class boundaries.

There are two types of ogives. It can be of 'less than' or 'more than' type based on whether the cumulative frequencies used are of 'less than' or more than 'type.

(i) 'Less than' Ogive

The "less than ogive" like an elongated S which starts from the lowest class boundary on the horizontal axis and gradually rises upward, ending at the highest class boundary corresponding to the total frequency of the distribution. To construct this curve, we plot the 'less than' cumulative frequencies against the upper class boundaries of the respective classes. The points plotted are joined by a smooth free hand curve. It is upward sloping from left to right.

(ii) 'More than' Ogive

The "more than ogive" has the appearance of an elongated S turned upside down. It starts from the lowest class boundary on the horizontal axis and gradually slopes downward, ending at the highest class boundary corresponding to the total frequency of the distribution. To construct this curve, we plot the 'more than' cumulative frequencies against the lower class boundaries of the respective classes. The points plotted are joined by a smooth free hand curve. It is down ward sloping from left to right.

Ogive is very useful for graphic computation of partition values, viz, Median, Quartiles, Deciles, Percentiles, etc. of which we will learn in the forthcoming chapters. It is alright to draw both the 'less than' ogive and 'more than' ogive on the same graph. If done so, they intersect at a point. A perpendicular from the point of intersection on the X-axis gives the value of median.

14. Define Silhouette.

Ans :

This graph is used when two different but related variable need to be presented with clear emphasis on the difference between the two.

Method-1

The difference between two sets of the variables for different time periods is obtained. In some cases, differences may have negative value.

Apart from the two historigrams, another historigram is to be draw on the same graph (i.e., Net balance). A part of graph will be below the x-axis which is corresponding to negative values of the net balance.

Method-2

In this method, two historigrams are drawn from one variable each on the same graph. The difference between the two variables is indicated by filling or shading the gap properly which express the positive and negative balances.

15. Range Graph.

Ans :

Range is the deviation that exist between maximum and minimum values of the variables. It indicates the difference between the two extreme values of variables. Range graph is also known as zone graph or variation graph, which is used to display and focus on the range of variation of a variable for each period.

Example

The range of variation is pointed out in the following cases,

- (i) Temperature on different days.
- (ii) Readings of blood pressure on different days of an individual.
- (iii) Price of a commodity on different time periods etc.

The range or zone graph is most suitable for time series and it helps the users to have an idea regarding, the changes in the size of the variables under study. The range graph or zone graph or variation graph can be constructed in any one of the following methods:

Method-1

In this method, the minimum and maximum values of the variable of each period are plotted and joined with straight line in order to get the range lines. Average value of the each variable (midpoint) is plotted and joined with straight line to attain the range graph. Hence, the range graph indicates the minimum, maximum and average value of the variable for each time period.

Method-2

This method involves plotting of the two historigrams on graph i.e.,

- (i) With respect to the maximum values of the variable for different periods and
- (ii) With respect to the minimum values.

Choose the Correct Answer

1. Diagrams are Tools of [b]
(a) Collection of data (b) Presentation of data
(c) Classification of data (d) Analysis of date
2. One of the following should be avoided in the presentation of data : [d]
(a) Pictograms (b) Rectangles
(c) Bars (d) Pie-diagrams.
3. Spheres and cylinders In a two dimensional diagram [a]
(a) Length and breadth are considered.
(b) Only width is considered.
(c) Only height is considered
(d) Length, breadth and thickness are considered.
4. For presentation of geographical distribution of data most suitable diagram is [b]
(a) Pictogram (b) Multiple Bar Diagram
(c) Dimensional diagram (d) Cartograph
5. A frequency polygon is a [a]
(a) Frequency curve (b) Bar diagram
(c) Modified ogive (d) None of these.
6. Ratio scale is suitable for [b]
(a) Low range of magnitude
(b) Wide range of magnitude or relative changes
(c) Zero and negative values
(d) Absolute changes.
7. Graphs are tools of [b]
(a) Collection of data (b) Analysis of data
(c) Presentation of data (d) Interpretation of data.
8. What type of frequency distribution is used for ogive curve : [a]
(a) Cumulative frequency (b) Percentage frequencies
(c) Simple frequencies (d) Open-end series.
9. What is meant by False Base line ? [a]
(a) Assumed base line (b) Not an original base line.
(c) Moves towards Y-axis (d) Moves towards X-axis.
10. Diagram are the most common type of diagrams used in practice. [c]
(a) Simple (b) Sub-divided
(c) Bar (d) All the above

Fill in the blanks

1. _____ create a positive and lasting impression.
2. Classifications regarding certain aspects of the diagram may be provided as a _____.
3. In two-dimensional diagrams, both _____ and _____ of the observations are represented by bars.
4. Three dimensional diagrams are diagrams in which three dimensions, namely, _____, _____ and _____.
5. A _____ diagram is used to represent only one variable.
6. In a _____ bar diagram two or more sets of interrelated data are represented.
7. _____ bars are popularly used for representing net quantities - excess or deficit.
8. _____ diagram is used when the requirement of the situation is to know the relationship between whole of a thing and its parts.
9. In terms of comparison, the _____ diagram appears inferior to simple bar diagram or divided bar diagram.
10. _____ paper is used for the construction of graphs.
11. _____ line technique is quite extensively used for magnifying the minor fluctuations in a time series data.
12. _____ graphs have time as an independent variable on the X-axis and dependent variables on the Y-axis,
13. Frequency graphs are designed to reveal clearly the characteristics features of a _____.
14. _____ curve is a smooth free hand curve drawn through the vertices of frequency polygon.
15. Ogive is also known as _____.
16. Band graph is also known as _____.
17. _____ is the deviation that exist between maximum and minimum values of the variables.
18. A _____ is a special kind or graph where figures are plotted at regular intervals over a period of one year.

ANSWERS

1. Diagrams
2. Footnote
3. Length, Width
4. Length, Breadth, Height.
5. Simple bar

6. Multiple
7. Deviation
8. Pie
9. Pie
10. Graph
11. False base
12. Time series
13. Frequency data.
14. Frequency
15. Cumulative Frequency Polygon.
16. Component graph.
17. Range
18. Z-chart

Rahul Publications

UNIT III

MEASURES OF CENTRAL TENDENCY:

Introduction -Significance - Arithmetic Mean - Geometric Mean - Harmonic Mean – Mode - Median - Quartiles and Percentiles - Simple and Weighted Averages - Uses and Limitations of different Averages

3.1 MEASURES OF CENTRAL TENDENCY

3.1.1 Introduction

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. Spiegel**, “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 form**

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.

3.1.2 Significance

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.

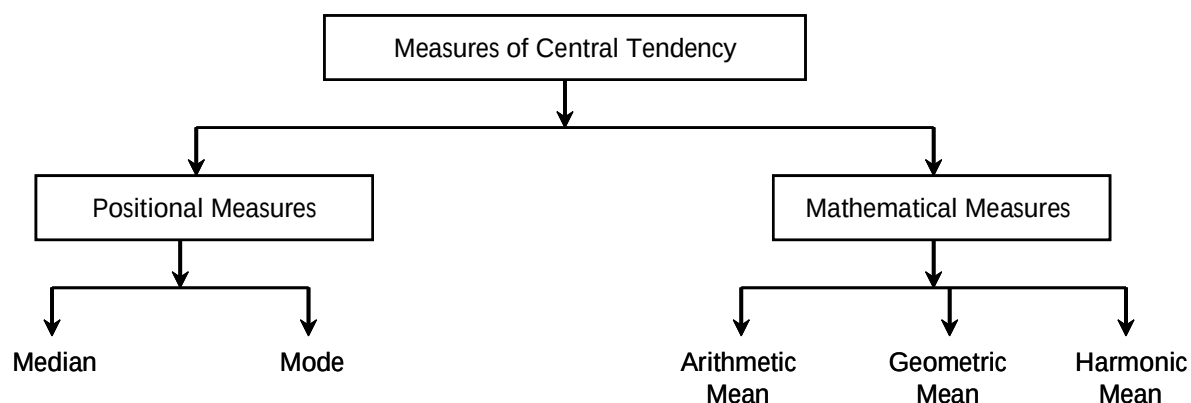
3.1.3 Types of Averages

Q6. Explain the different types of averages. (OR)

What are the various measures of central tendency?

Ans : **(July-21)**

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



3.2 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 interpretations 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 significantly 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 :

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

(OR)

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

Here

$\bar{X}$ = Arithmetic Mean

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

N = Number of observations.

Steps

The formula involves two steps in calculating mean :

- i) Add together all the values of the variable X and obtain the total, i.e., $\sum X$.
- ii) Divide this total by the number of the observations, i.e., N .

Example

The following table gives the monthly income of 10 employees in an office :

Income (Rs.) 4,780 5,760 6,690 7,750 4,840 4,920 6,100 7,810 7,050 6,950 Calculate the arithmetic mean of incomes.

Sol. :

Let income be denoted by the symbol X .

Calculation of Arithmetic Mean

Employee	Monthly Income (Rs.) X
1	4,780
2	5,760
3	6,690
4	7,750
5	4,840
6	4,920
7	6,100
8	7,810
9	7,050
10	6,950
<u>N = 10</u>	<u>ΣX = 62,650</u>

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

$$\sum X = 62,650, N = 10$$

$$\bar{X} = \frac{62,650}{10} = 6,265.$$

Hence the average income is Rs. 6,265.

Short-cut 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

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

where

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 : $\bar{X} = A + \frac{\sum d}{N}$.

Example

From above example calculate arithmetic mean by taking 7,000 as the assumed mean.

Sol. :

Calculation of Arithmetic Mean

Employee	Income (Rs.) X	(X-7000)
1	4,780	-2220
3	6,690	-310
4	7,750	+750
5	4,840	-2160
6	4,920	-2080
7	6,100	-900
8	7,810	+810
9	7,050	+50
10	6,950	-50
<u>N = 10</u>		<u>Σd = - 7350</u>

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

$$A = 7000, \sum d = -7350, N = 10$$

$$\bar{X} = 7000 - \frac{7,350}{10} = 7,000 - 735$$

$$= 6,265.$$

Hence the average income is Rs. 6,265.

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

Where,

f = Frequency;

X = The variable in question;

N* = Total number of observations, i.e., $\sum f$.

Steps

- i) Multiply the frequency of each row with the variable and obtain the total ΣfX .
- ii) Divide the total obtained by step (i) by the number of observations, i.e., total frequency.

Example

From the following data of the marks obtained by 60 students of a class, calculate the arithmetic mean :

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

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

Calculation of Arithmetic Mean

Marks X	No. of students f	fX
20	8	160
30	12	360
40	20	800
50	10	500
60	6	360
70	4	280
	<u>$N = 60$</u>	<u>$\Sigma fX = 2,460$</u>

$$\bar{X} = \frac{\Sigma fX}{N} = \frac{2,460}{60} = 41$$

Hence the average marks = 41.

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.

Example

Calculate arithmetic mean by the short-cut method using frequency distribution of above example.

Sol. :

Calculation of Arithmetic Mean

Marks X	No.of students f	(X - 40) d	fd
20	8	-20	-160
30	12	-10	-120
40-A	20	0	0
50	10	+10	+100
60	6	+20	+120
70	4	+30	+120
	<u>60</u>		<u>60</u>

$$\bar{X} = A + \frac{\Sigma fd}{N} = 40 + \frac{60}{60} = 40 + 1 = 41$$

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

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

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.

Example

From the following data compute arithmetic mean by direct method :

Marks:	0 – 10	10 – 20	20 – 30	30 – 40	40 – 50	50 – 60
No. of students :	5	10	25	30	20	10

Sol. :

Calculation of Arithmetic Mean by Direct Method

Marks	Mid-point (m)	No. of Students (f)	fm
0 – 10	5	5	25
10 – 20	15	10	150
20 – 30	25	25	625
30 – 40	35	30	1,050
40 – 50	45	20	900
50 – 60	55	10	550
		<u>N = 100</u>	<u>$\Sigma fm = 3,300$</u>

$$\bar{X} = \frac{\Sigma fm}{N} = \frac{3,300}{100} = 33.$$

ii) Short-cut Method

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

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

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

Calculate arithmetic mean by the short-cut method from the data of above example.

Calculation of Arithmetic Mean

Marks	Mid-points (m)	No. of Students (f)	(m – 35) d	fd
0 – 10	5	5	–30	–150
10 – 20	15	10	–20	–200
20 – 30	25	25	–10	–250
30 – 40	35–A	30	0	0
40 – 50	45	20	+10	+200
50 – 60	55	10	+20	+200
		<u>N = 100</u>		<u>Σ fd = – 200</u>

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

$$\bar{X} = 35 + \frac{-200}{100}$$

$$= 35 + (-2) = 33$$

3.2.1 Weighted Arithmetic Mean

Q11. Explain in detail about Weighted Arithmetic Mean.

Ans :

The calculation of arithmetic mean is based on an assumption that all items are equally important. But in practice, situations might come wherein the relative importance of various items will not be same. Under such situations weighted arithmetic mean is calculated as follows,

$$\bar{X}_w = \frac{\sum WX}{\sum W}$$

Where,

$\bar{X}_w$ = Weighted arithmetic mean

X = Variable values

W = Weights

Steps

The following steps are followed to calculate weighted arithmetic mean.

1. First multiply the value of each variable (X) with their respective weights (W) and calculate the sum $[\sum WX]$.
2. Divide the total [i.e., $\sum WX$] by the total of weights [i.e., $\sum W$]

$$\bar{X}_w = \frac{\sum W(X)}{\sum W}$$

Frequency Distribution

In frequency distribution, each value of frequency [f] is multiplied with their respective variable values [x] and then values of fx are multiplied with their respective weighted [W]. Calculate the total of W(f(x)) column and represent it as $\Sigma W(fx)$. Finally, divide the $\Sigma W(f(x))$ by the total of weights (i.e., ΣW) in order to get the value of weighted arithmetic mean. The formula is as follows,

$$\bar{X}_w = \frac{\Sigma W(fx)}{\Sigma W}$$

Q12. State the uses of Weighted Arithmetic Mean.

Ans :

- (i) Importance of all items in a series is not equal.
- (ii) Classes of the same group contain widely varying frequencies.
- (iii) It is desired to calculate the average of the series from the average of its component parts.
- (iv) Ratios, percentages or rates are being averaged or
- (v) There is a change either in the proportion of values of items or in the proportion of their frequencies.

3.2.2 Combined Mean

Q13. Explain briefly about Combined Mean.

Ans :

If we have the arithmetic mean and number of items of two or more than two related groups, we can compute combined average of these groups by applying the following formula :

$$\bar{X}_{12} = \frac{N_1\bar{X}_1 + N_2\bar{X}_2}{N_1 + N_2}$$

$\bar{X}_{12}$ = combined mean of the two groups

$\bar{X}_1$ = arithmetic mean of first group

$\bar{X}_2$ = arithmetic mean of second group

N_1 = number of items in the first group

N_2 = number of items in the second group.

If we have to find out the combined mean of three sub-groups the above formula can be extended as follows :

$$\bar{X}_{123} = \frac{N_1\bar{X}_1 + N_2\bar{X}_2 + N_3\bar{X}_3}{N_1 + N_2 + N_3}$$

Example

The mean height of 25 male workers in a factory is 61 inches and the mean height of 35 female workers in the same factory is 58 inches. Find the combined mean height of 60 workers in the factory.

Sol. :

$$\bar{X}_{12} = \frac{N_1\bar{X}_1 + N_2\bar{X}_2}{N_1 + N_2}$$

whereas $N_1 = 25$, $\bar{X}_1 = 61$, $N_2 = 35$, $\bar{X}_2 = 58$

$$\bar{X}_{12} = \frac{(25 \times 61) + (35 \times 58)}{25 + 35} = \frac{1525 + 2030}{60} = \frac{3555}{60} = 59.25$$

Thus the combined mean height of 60 workers is 59.25 inches.

PROBLEMS ON ARITHMETIC MEAN

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 :

$$\begin{aligned} \text{Arithmetic Mean } (\bar{X}) &= \frac{\sum X}{N} = \frac{85 + 70 + 10 + 75 + 500 + 8 + 42 + 250 + 40 + 36}{10} \\ &= 1116/10 = 111.6 \end{aligned}$$

∴ 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**

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

∴ Average marks of the class = 53 marks.

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

$$\begin{aligned}\bar{X} &= A + \frac{\sum d}{N} = 92 + \frac{-390}{10} \\ &= 92 + (-39) \\ &= 53.\end{aligned}$$

3. The following data related to the distance travelled by 520 villagers to buy their weekly requirements :

Miles Travelled	2	4	6	8	10	12	14	16	18	20
No. of Villagers	38	104	140	78	48	42	28	24	16	2

Sol :

$$\text{Arithmetic Mean} = \frac{\sum fx}{N}$$

Miles Travelled (x)	No. of villagers (f)	fx
2	38	76
4	104	416
6	140	840
8	78	624
10	48	480
12	42	504
14	28	392
16	24	384
18	16	288
20	2	40
	<u>N = 520</u>	<u>Σfx = 4044</u>

$$\begin{aligned}(\bar{X}) &= \frac{\sum fx}{N} = \frac{4044}{520} \\ &= 7.7769.\end{aligned}$$

4. Find mean from the following frequency data :

Class-interval	Frequency
15 - 25	4
25 - 35	11
35 - 45	19
45 - 55	14
55 - 65	0
65 - 75	2

Sol :

Calculation of Arithmetic Mean

Class-interval	Frequency (f)	Mid value (m)	fm
15 - 25	4	20	80
25 - 35	11	30	330
35 - 45	19	40	760
45 - 55	14	50	700
55 - 65	0	60	0
65 - 75	2	70	140
	<u>50</u>		<u>2010</u>

$$\text{Arithmetic mean } (\bar{X}) = \frac{\sum fm}{N} = \frac{2010}{50}$$

$$(\bar{X}) = 40.2$$

5. Calculate Arithmetic mean from the following data.

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 :

(May-18)

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
	<u>$\Sigma f = 2400$</u>		<u>$\Sigma fM = 1,50,800$</u>

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

6. Find the missing frequency from the following data :

Marks	0 - 10	10 - 20	20 - 30	30 - 40	40 - 50	50 - 60
No. of Students	5	15	20	-	20	10

The arithmetic mean is 34 marks.

Sol :

Let the missing frequency be " x_1 "

Calculation of Missing Frequency

Class (x)	Frequency (f)	Mid Values (m)	fm
0 - 10	5	5	25
10 - 20	15	15	225
20 - 30	20	25	500
30 - 40	x_1	35	$35x_1$
40 - 50	20	45	900
50 - 60	10	55	550
	$N = 70 + x_1$		$\Sigma fm = 2200 + 35x_1$

Given, Mean ($\bar{X}$) = 34

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

$$34 = \frac{2200 + 35x_1}{70 + x_1}$$

$$34(70 + x_1) = 2200 + 35x_1$$

$$2380 + 34x_1 = 2200 + 35x_1$$

$$2380 - 2200 = 35x_1 - 34x_1$$

$$\therefore x_1 = 180$$

$\therefore$ Missing frequency x_1 = would be 180.

7. The mean of the following series is ` 115.86. Find the missing figures.

House Rent (`)	110	112	113	117	7	125	128	130
Number of Houses	25	27	13	15	14	8	6	2

Sol :

(Dec.-17, Imp.)

Let missing figure be ' x_1 '

Calculation of Missing Figure

House Rent (₹) (x)	No. of Houses	fx
110	25	2,750
112	17	1,904
113	13	1,469
117	15	1,755
x_1	14	$14x_1$
125	8	1,000
128	6	768
130	2	260
	<u>N = 100</u>	<u>$\Sigma fx = 9,906 + 14x_1$</u>

Given,

$$\text{Mean } (\bar{x}) = 115.86$$

$$\bar{x} = \frac{\Sigma fx}{N}$$

$$115.86 = \frac{9,906 + 14x_1}{100}$$

Cross multiply the above information,

$$115.86 \times 100 = 9,906 + 14x_1$$

$$11,586 = 9,906 + 14x_1$$

$$-14x_1 = 9,906 - 11,586$$

$$-14x_1 = -1,680$$

$$x_1 = \frac{-1,680}{-14}$$

$$x_1 = 120$$

$\therefore$ Missing figure x_1 is 120.

8. Find the combined mean from the following data.

Group	A	B	C
Number	100	125	150
Mean	25	10	15

Sol :

(May-18, Imp.)

Given that,

$$N_1 = 100 \quad N_2 = 125 \quad N_3 = 150$$

$$\bar{X}_1 = 25 \quad \bar{X}_2 = 10 \quad \bar{X}_3 = 15$$

The formula for combined mean is,

$$\bar{X}_{13} = \frac{N_1 \bar{X}_1 + N_2 \bar{X}_2 + N_3 \bar{X}_3}{N_1 + N_2 + N_3}$$

$$\bar{X}_{13} = \frac{100(25) + 125(10) + 150(15)}{100 + 125 + 150}$$

$$\bar{X}_{13} = \frac{2,500 + 1,250 + 2,250}{375}$$

$$\bar{X}_{13} = \frac{6,000}{375}$$

$$\bar{X}_{13} = 16.$$

9. The mean weight of 120 students of B.Com Class is 60 Kgs. The mean weight of boys is 55 Kgs and that of girls is 45 Kgs. Find the number of boys and girls.

Sol :

(June-19)

$$\bar{X}_{12} = \frac{\bar{X}_1 N_1 + \bar{X}_2 N_2}{N_1 + N_2}$$

$$\bar{X}_{12} = 60, N_1 + N_2 = 120 \quad \bar{X}_1 = 55,$$

$$\bar{X}_2 = 45 \quad N_1 = 120 - N_2$$

$$60 = \frac{55(120 - N_2) + 45(N_2)}{120}$$

$$60 \times 120 = 6600 - 55N_2 + 45N_2$$

$$7200 - 6600 = -10 N_2$$

$$600 = -10N_2$$

$$N_2 = \frac{600}{-10}, N_2 = -60$$

$$\therefore N_1 = 180, N_2 = -60$$

10. Mean of 100 observations is found to be 40. If at the time of computation, two items are wrongly taken as 30 and 27 instead of 3 and 72. Find correct mean.

Sol :

Given that,

$$N = 100 \text{ and } \bar{X} = 40$$

$$\Sigma X = N\bar{X}$$

$$\Sigma X = 100 \times 40$$

$$= 4,000$$

Correct mean is calculated as follows,

$$\Sigma X = \Sigma X - \text{Wrong items} + \text{Correct items}$$

$$= 4,000 - (30 + 27) + (3 + 72)$$

$$= 4,000 - 57 + 75$$

$$= 4,018$$

$$\therefore \text{Correct Mean } (\bar{X}) = \frac{\text{Correct } \Sigma X}{N}$$

$$= \frac{4,018}{100}$$

$$\text{Correct Mean} = 40.18$$

3.3 GEOMETRIC MEAN

Q14. 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.

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

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

Steps

- (i) Take log of all the given items of the data
- (ii) Add the log values i.e. find $\sum \log X$
- (iii) Divide $\sum \log X$ 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	x_1	log 10	1.0000
110	x_2	log 110	2.0414
135	x_3	log 135	2.1303
120	x_4	log 120	2.0792
50	x_5	log 50	1.6990
59	x_6	log 59	1.7709
60	x_7	log 60	1.7782
7	x_8	log 7	0.8451
N = 8		$\sum \log X = 13.3441$	

Geometric mean (g)

$$\begin{aligned}
 g &= \text{Antilog} \left(\frac{\sum \log X}{N} \right) = \text{Antilog} \left(\frac{13.3441}{8} \right) \\
 &= \text{Antilog } 1.6680 \\
 &= 46.56
 \end{aligned}$$

Q15. 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
	<u>N = 62</u>		<u>Σ f log X = 36.3016</u>

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 $g = 3.850$

Q16. 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 – Σf log m
- (v) Divided f log m by Σf – $\frac{\sum f \log m}{N}$

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

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

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	1.2553	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	1.4314
			<u>N = 68</u>	<u>$\Sigma \log m = 81.9092$</u>

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

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

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

Q17. 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 separately 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.

3.3.1 Weighted Geometric Mean**Q18. 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 $W \log x$ and represent this total as $\Sigma(W \log x)$.
4. Divide $\Sigma(W \log x)$ by the total of weights, [i.e., ΣW]
5. Find out the Antilog for the value obtained in step 4.

$$G.M_w = \text{Antilog} \left[\frac{\Sigma(W \log x)}{\Sigma W} \right]$$

PROBLEMS

11. 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
	$\Sigma \text{Log X} = 9.1710$

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

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

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

12. Calculate geometric mean for the following data :

1111, 111, 11, 1.1, 0.11, 0.011

Sol :

(May-18, Dec.19)

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
	$\Sigma \log x = 3.2157$

We have M = 6, $\Sigma \log x = 3.2157$

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

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

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

$$\text{GM} = 3.4356$$

13. 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
	<u>N = 31</u>		<u>$\sum f \log X = 66.7710$</u>

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

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

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

14. Find the geometric mean from the following data :

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

Sol :

(May-18)

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

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

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

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

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

15. 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		$\sum f \cdot \log x = 147.9849$

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

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

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

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

$$\therefore \text{G.M} = 30.1890$$

3.4 HARMONIC MEAN

Q19. 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,

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

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

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

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 $\frac{1}{x}$ (i.e., $\sum \left[\frac{1}{x} \right]$)

3. Divide the number of observations [N] by

$$\sum \frac{1}{x}$$

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

Q20. 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

$$\text{sum as } \sum \left[f \times \frac{1}{x} \right].$$

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

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

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

Q21. 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 (i.e., $\frac{1}{m}$) with their corresponding frequencies and obtain the

$$\text{sum } \sum \left[f \times \frac{1}{m} \right].$$

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

$$\left[f \times \frac{1}{m} \right].$$

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

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

3.4.1 Weighted Harmonic Mean

Q22. 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,

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

(or)

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

PROBLEMS

16. Compute Harmonic mean of the following :

20, 40, 50, 80, 100

Sol :

X	1/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
		<u><u>Σ 1/x 0.1175</u></u>

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

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

$$\text{Harmonic Mean (H.M)} = 42.55.$$

17. 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
	$\Sigma(1/X) = 0.041$

$$\begin{aligned}\text{Harmonic Mean (H.M)} &= \frac{N}{\Sigma\left(\frac{1}{X}\right)} \\ &= \frac{5}{0.041} \\ &= 121.9 = 122\end{aligned}$$

18. 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 : (June-19)

Thus, apply H.M.

$$\begin{aligned}\text{H.M} &= \frac{N}{\Sigma\left(\frac{1}{X}\right)} \Rightarrow \frac{4}{\frac{1}{200} + \frac{1}{400} + \frac{1}{600} + \frac{1}{800}} \\ &= \frac{4}{\frac{12+6+4+3}{2400}} \Rightarrow \frac{4 \times 2400}{25} \Rightarrow \frac{9600}{25} \\ &= 384 \text{ miles per hour.}\end{aligned}$$

19. 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

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

$$\therefore \text{H.M} = 16.5931.$$

20. Calculate the Harmonic mean from the following data

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

Sol :

(Dec.-17, Imp.)

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

$$\begin{aligned} \text{H.M} &= \frac{N}{\sum \left[\frac{f}{x} \right]} \\ &= \frac{60}{4.436} \\ &= 13.526 \end{aligned}$$

21. 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

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

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

22. Calculate Harmonic mean for the following distribution.

Class	Frequency
5 - 10	1
10 - 15	9
15 - 20	29
20 - 25	54
25 - 30	11
30 - 35	5

Sol :

Calculation of Harmonic Mean

Class	Mid - Points (m)	Frequency (f)	f/m
5 - 10	7.5	1	0.13
10 - 15	12.5	9	0.72
15 - 20	17.5	29	1.65
20 - 25	22.5	54	2.4
25 - 30	27.5	11	0.4
30 - 35	32.5	5	0.15
		N = 109	$\sum \left(\frac{f}{m} \right) = 5.45$

$$\begin{aligned}\therefore \text{H.M} &= \frac{N}{\sum \left(\frac{f}{m} \right)} \\ &= \frac{109}{5.45} \\ \text{H.M} &= 20\end{aligned}$$

3.5 MODE

Q23. 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.

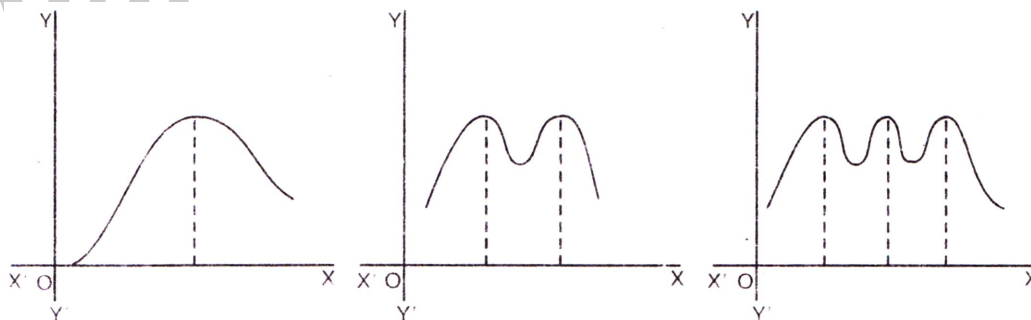


Fig.: Unimodal

Fig.: Bimodal

Fig.: Multimodal

Q24. 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.

Q25. 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.

Example:

Find the mode from the following observation,

Marks [x]: 5, 29, 25, 10, 29, 29, 20, 18, 15, 30.

Sol :

Calculation of Mode

Marks (Arranged in ascending order)	5	10	15	18	20	25	29	30
No. of Times occurs	1	1	1	1	1	1	3	1

Since, 29 is repeated maximum number of times i.e., 3 times.

∴ Mode = 29

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.

Example

Size of Garment	27	28	29	30	31	32
No. of persons	5	15	20	70	60	15

Sol :

- By inspection the highest value is selected from frequency column i.e., (70)
- The value of the variable corresponding to 70 is 30
- ∴ Mode = 30

(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'.

Q26. 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,

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

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 ON MODE**23. 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.

24. 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.

25. 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
20 – 30	19
30 – 40	17
40 – 50	12

L

(f_1)

(f)

(f_2)

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.

26. 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 :

(Dec.-17, Imp.)

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

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

$$\Delta_1 = f - f_1$$

$$\Delta_2 = f - f_2$$

$$\Delta_1 = 32 - 20 = 12$$

$$\Delta_2 = 32 - 14$$

$$= 18$$

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

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

$$40 + 4 = 44$$

27. Calculate Modal Weight from the following data.

Wt. in Lbs	100-110	110-120	120-130	130-140	140-150	150-160	160-170	170-180
No. of Persons	4	6	20	32	33	17	8	2

Sol :

(June-19)

Wt. in Lbs	No. of Persons
100 – 110	4
110 – 120	6
120 – 130	20
130 – 140	32 f_1
L 140 – 150	33 f
150 – 160	17 f_2
160 – 170	8
170 – 180	2

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

$$L = 140, \Delta_1 = f - f_1$$

$$33 - 32 = 1$$

$$\Delta_2 = f - f_2$$

$$33 - 17 = 16$$

$$140 + \frac{1}{1+16} \times 10$$

$$140 + \frac{10}{17}$$

$$140 + 0.58 = 140.58$$

28. Calculate mode for the following data :

Marks	No. of Students
0 – 10	2
10 – 20	5
20 – 30	11
30 – 40	14
40 – 50	17
50 – 60	13
60 – 70	6
70 – 80	2

Sol :

Calculation of Mode

Marks	No. of Students
0 – 10	2
10 – 20	5
20 – 30	11
30 – 40	14 (f_1)
L 40 – 50	17 (f)
50 – 60	13 (f_2)
60 – 70	6
70 – 80	2

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

Mode lies in the class interval 40 – 50 since the highest frequency 17 exists there.

$\therefore$ L = Lower limit = 40

$$\Delta_1 = f - f_1$$

$$\Delta_1 = 17 - 14 = 3$$

$$\Delta_2 = f - f_2$$

$$\Delta_2 = 17 - 13 = 4$$

C = Class interval = 10

$$\therefore \text{Mode} = 40 + \frac{3}{3+4} \times 10$$

$$= 40 + \frac{3}{7} \times 10$$

$$= 40 + 0.428 \times 10$$

$$= 40 + 4.29$$

$$\text{Mode} = 44.29$$

29. Find out the Mode for the data given below

Size	0 - 5	5 - 10	10 - 15	15 - 20	20 - 25	25 - 30	30 - 35
Frequency	2	4	20	8	20	18	64

Sol :

Calculation of Mode

Marks	No. of Students
0 – 5	2
5 – 10	4
10 – 15	20
15 – 20	8
20 – 25	20
25 – 30	18 (f_1)
L 30 – 35	64 (f)

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

Mode lies in the class interval 30 – 35 since the highest frequency 64 exists there.

$\therefore$ L = Lower limit = 30

$$\Delta_1 = f - f_1$$

$$\Delta_1 = 64 - 18 = 46$$

$$\Delta_2 = f - f_2$$

$$\Delta_2 = 64 - 0 = 64$$

$$C = \text{Class interval} = 5$$

$$\therefore \text{Mode} = 30 + \frac{46}{46 + 64} \times 5$$

$$= 30 + \frac{46}{110} \times 5$$

$$= 30 + 0.418 \times 5$$

$$= 30 + 2.09$$

$$\text{Mode} = 32.09$$

3.6 MEDIAN

Q27. 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.
-

Q28. 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.
-

Q29. How median is calculated for individual series ?

Ans :

Steps

- (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.

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

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

(or)

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

Example

From the following data of the wages of 7 workers, compute the median wage :

Wages (in Rs.) 4100 4150 6080 7120 5200 6160 7400

Sol. :

Calculation of Median

Sl.No.	Wages arranged in ascending order
1	4100
2	4150
3	5200
4	6080
5	6160
6	7120
7	7400

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

$$= \frac{7+1}{2} = 4^{\text{th}} \text{ item}$$

$$= \text{Rs. } 6080.$$

Q30. How median is calculated for discrete series ?

Ans :

Steps

- (i) Arrange the data in ascending or descending order of magnitude,
 (ii) Find out the cumulative frequencies.

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

Example

From the following data, find the value of median :

Income (Rs.)	5,000	5,500	6,800	8,000	8,500	7,800
No. of persons	24	26	20	20	2	30

Calculation of Median

Income arranged in ascending order	No. of person (f)	c.f.
5,000	24	24
5,500	26	50
6,800	20	70
7,800	30	100
8,000	20	120
8,500	2	122

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

$$= \text{Size of 61.5th item.}$$

Size of 61.5th item = Rs. 6,800. Hence the median income is Rs. 6,800.

Q31. 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 ON MEDIAN**30. Calculate Median from the following.**

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

*Sol :***Procedure****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

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

31. 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

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

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

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

32. 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 :

Procedure :

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
	<u>N = 182</u>	

Step II: Apply the formulae for determination of media.

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

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

$$\begin{aligned} \text{Median} &= \text{Size of } \frac{183}{2} \\ &= 91.5 \text{ item} \end{aligned}$$

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.

33. 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 :

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

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

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

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

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

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

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

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

$$\text{Median} = 26.32$$

34. From the data given below compute median of the average deposit balance of savings for the branch during march 1988.

Average deposit balance (₹)	No. of deposits
Less than 100	26
100-200	68
200-300	145
300-400	242
400-500	188
500-600	65
600-700	16

Sol :

Average deposit balance	No. of deposit (f)	Cf
0 - 100	26	26
100 - 200	68	94
200 - 300	145	239 F
L 300 - 400	242 f	481
400 - 500	188	669
500 - 600	65	734
600 - 700	16	750

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

Median class interval = Size of $\frac{750}{2} = 375^{\text{th}}$ item

∴ Median class interval is 300 – 400

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

$$\text{Median} = 300 + \left[\frac{375 - 239}{242} \right] \times 100$$

$$\text{Median} = 300 + \left[\frac{136}{242} \right] \times 100$$

$$\text{Median} = 300 + 0.5619 \times 100$$

$$\text{Median} = 300 + 56.198$$

$$\text{Median} = 356.198 \text{ (or) } 356.20.$$

35. Calculate median from the following data

Marks above	20	30	40	50	60	70	80	90
Frequency	200	195	177	157	142	124	84	16

Sol :

(Dec.-17)

C.I	Frequency	Cumulative frequency
20 - 30	5	5
30 - 40	18	23
40 - 50	20	43
50 - 60	15	58
60 - 70	18	76 F
L 70 - 80	40 f	116
80 - 90	68	184
90 - 100	16	200 → N

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

$$\frac{N}{2} = \frac{200}{2} = 100$$

$$70 + \frac{100 - 76}{40} \times 10$$

$$70 + \frac{24 \times 10}{40}$$

$$70 + \frac{240}{40}$$

$$= 70 + 6$$

$$= 76$$

36. Compute Median from the following data.

Mid Values	115	125	135	145	155	165	175	185	195
Frequency	6	25	48	72	116	60	38	22	3

Sol :

(Dec.-18)

Mid Values	Frequency	Class interval	CF
115	6	110 - 120	6
125	25	120 - 130	31
135	48	130 - 140	79
145	72	140 - 150	151 F
155	116 f	L 150 - 160	267
165	60	160 - 170	327
175	38	170 - 180	365
185	22	180 - 190	387
195	03	190 - 200	390

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

$$\frac{N}{2} = \frac{390}{2} = 195$$

$$= 150 + \frac{195 - 151}{116} \times 10$$

$$= 150 + \frac{440}{116}$$

$$= 150 + 3.7$$

$$\text{Median} = 153.7$$

37. Calculate Median from the following data.

Marks	0 - 9	10 - 19	20 - 29	30 - 39	40 - 49	50 - 59	60 - 69
No. of Students	6	10	20	25	16	12	5

Sol :

(June-19, Imp.)

C.I	f	Cumulative Frequency
0 - 9.5	6	6
9.5 - 19.5	10	16
19.5 - 29.5	20	36 F
L 29.5 - 39.5	25 f	61
39.5 - 49.5	16	77
49.5 - 59.5	12	89
59.5 - 69.5	5	94 = N

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

$$= \frac{94}{2} = 47$$

$$29.5 + \frac{47 - 36}{25} \times 10$$

$$29.5 + \frac{11 \times 10}{25}$$

$$29.5 + \frac{10}{25}$$

$$29.5 + 4.5$$

$$= 33.9$$

3.6.1 Quartiles and Percentiles

Q32. Describe briefly about Quartiles and Percentiles of individual, discrete and continuous series.

Ans :

(i) Quartiles

Quartiles refers to the values dividing the given series into four equal parts i.e., Q_1 , Q_2 , Q_3 and Q_4 .

Where,

Q_1 = First quartile / lower quartile.

Q_2 = Second quartile. It is nothing but median.

Q_3 = Third quartile / upper quartile

Q_4 = Fourth quartile Q_4 can be ignored.

∴ The procedure for calculating quartile is same as that of procedure for calculating the median.

Formula

Lower Quartile

In Individual Series and Discrete Series

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

In Continuous Series

$$Q_1 = \text{Size of } \frac{N}{4} \text{ item}$$

Upper Quartile

In individual and discrete series

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

In Continuous Series

$$Q_3 = \text{Size of } \frac{3N}{4} \text{ item}$$

(ii) Percentiles

Percentiles refers to the values which divide the given series into 100 equal parts. It means that there will be 99 percentiles i.e., $P_1, P_2, P_3, \dots, P_{99}$

Formula

In Individual and Discrete Series

$$P_{50} = \text{Size of } \frac{50(N+1)}{100} \text{ item}$$

In Continuous Series

$$P_{50} = \text{Size of } \frac{50N}{100} \text{ item}$$

PROBLEMS ON QUARTILES AND PERCENTILES

38. Calculate Quartiles, and 8th Decile from the following data the marks obtained by 10 students in an Examination.

27 31 19 35 23 16 40 46 17 37

Sol :

The data has to be arranged in ascending order:

16 17 19 23 27 31 35 37 40 46

$$Q_1 = \text{Size of } \left(\frac{N+1}{4} \right)^{\text{th}} \text{ item} = \text{Size of } \left(\frac{10+1}{4} \right)^{\text{th}} = 2.75^{\text{th}} \text{ item}$$

$$= \text{Size of } 2^{\text{nd}} \text{ item} + 0.75 (\text{Size of } 3^{\text{rd}} \text{ item} - \text{Size of } 2^{\text{nd}} \text{ item})$$

$$= 17 + 0.75 (19-17)$$

$$= 17 + 1.5$$

$$= 18.5 \text{ marks}$$

$$Q_3 = \text{Size of } 3 \left(\frac{N+1}{4} \right)^{\text{th}} \text{ item} = \text{Size of } 3 \left(\frac{10+1}{4} \right)^{\text{th}} \text{ item}$$

$$= \text{Size of } 8.25^{\text{th}} \text{ item.}$$

$$= \text{Size of } 8^{\text{th}} \text{ item} + 0.25 (\text{Size of } 9^{\text{th}} \text{ item} - \text{Size of } 8^{\text{th}} \text{ item})$$

$$= 37 + 0.25 (40-37)$$

$$= 37 + .75$$

$$= 37.75$$

$$D_8 = \text{Size of } 8 \left(\frac{N+1}{4} \right)^{\text{th}} \text{ item} = \text{Size of } 8 \left(\frac{10+1}{4} \right)^{\text{th}} \text{ item}$$

$$= \text{Size of } 8.8^{\text{th}} \text{ item} = \text{Size of } 8^{\text{th}} \text{ item} + 0.8 (9^{\text{th}} \text{ item} - 8^{\text{th}} \text{ item})$$

$$= 37 + 0.8 (40 - 37)$$

$$= 37 + 2.4$$

$$= 39.4$$

39. Calculate Lower Quartile and Upper quartile for 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 :

i) Lower Quartile, i.e. Q_1

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

$$Q_1 = \frac{14+1}{4} = \text{Size of 3.75th item}$$

when 3.75th item is exactly does not appearing from the data.

Hence, Q_1 formulae is

Value of 3rd item + 0.75 (value of 4th item – Value of 3rd item)

$$Q_1 = 61 + 0.75 (64 - 61)$$

$$Q_1 = 61 + 0.75 (3)$$

$$Q_1 = 61 + 2.25$$

$$Q_1 = 63.25$$

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

$$Q_3 = \text{Size of } \frac{3(14+1)}{4} \text{th item}$$

$$Q_3 = \text{Size of 3 (3.75)th item}$$

$$Q_3 = \text{Size of 11.25th item}$$

where 11.25th item is exactly does not appearing from the given data.

Hence Q_3 formulae is

$$Q_3 = \text{Value of 11th item} + 0.25 (\text{value of 12th} - \text{Value of 11th item})$$

$$Q_3 = 64 + 0.25 (65 - 64)$$

$$Q_3 = 64 + 0.25 (1)$$

$$Q_3 = 64.25$$

40. Calculate 8th Decile and Sixty first percentile from the following data of marks obtained by 10 students in an examination.

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

i) Deciles 8 = Size of $\frac{8(N+1)}{10}$ th item

$$= \frac{8(10+1)}{10} = \frac{8(11)}{10} = \frac{88}{10} = \text{Size of 8.8th item}$$

$$\text{Deciles 8} = 8^{\text{th}} \text{ item} + 0.8 (9^{\text{th}} \text{ item} - 8^{\text{th}} \text{ item})$$

$$\text{Deciles 8} = 32 + 0.8 (35 - 32)$$

$$\text{Deciles 8} = 32 + (0.8 \times 3)$$

$$\text{Deciles 8} = 32 + 2.4$$

$$\text{Deciles 8} = 34.4$$

$$\text{ii) } 6^{\text{th}} \text{ percentile} = \text{Size of } \frac{61(N+1)}{100}^{\text{th}} \text{ item}$$

$$6^{\text{th}} \text{ percentile} = \frac{61(10+1)}{100}$$

$$6^{\text{th}} \text{ percentile} = \frac{61(11)}{100}$$

$$6^{\text{th}} \text{ percentile} = \frac{671}{100} = \text{Size of } 6.71 \text{ item}$$

$$6^{\text{th}} \text{ percentile} = 6^{\text{th}} \text{ item} + 0.71 \\ (7^{\text{th}} \text{ item} - 6^{\text{th}} \text{ item})$$

$$6^{\text{th}} \text{ percentile} = 26 + 0.71 (30 - 26)$$

$$6^{\text{th}} \text{ percentile} = 26 + 0.71 (4)$$

$$6^{\text{th}} \text{ percentile} = 26 + 2.84$$

$$6^{\text{th}} \text{ percentile} = 28.4.$$

3.7 RELATIONSHIP AMONG MEAN, MEDIAN AND MODE

Q33. Describe the relationship between Mean, Median and Mode.

Ans :

Mode touches the peak of the curve indicating maximum frequency. Median divides the area of the curve in two equal halves and Mean is the centre of gravity. The three values are inter-related. The following points list the relationship between the various averages.

- In a distribution, the relative positions of mean, median, and the mode depend upon the Skewness of the distribution. If the distribution is symmetrical, then Mean = Median = Mode.
- If the distribution is positively skewed (the longer tail of the distribution is towards the

right), the Mode will be lesser, than the Median, which in turn will be lower than the Arithmetic Mean. In case of a negatively skewed distribution, the Mode will be greater than the Median, which in turn will be greater than the Arithmetic Mean.

- In a given distribution, if all the observations are positive, Arithmetic Mean is greater than Geometric Mean, which in turn is greater than Harmonic Mean. With a distribution of moderate Skewness, median tends to be approximately $1/3^{\text{rd}}$ as far away from the mean as from the mode.
- Mode = 3 Median - 2 Mean
- Mean - Mode = 3 (Mean - Median). Thus, the difference between Mean and Mode is three times, the difference between mean and median.

PROBLEMS

41. If the Mode and Mean of a moderately skewed series are 30.2 and 20.4 respectively, what would be its Median?

Sol : (Dec.-18, Imp.)

We know that $3\text{Median} - 2\text{Mean} = \text{Mode}$

Given mode = 30.2 & Mean = 20.4

Median = ?

$$3\text{Median} = \text{Mode} + 2 \text{ Mean}$$

$$3 \text{ Median} = 30.2 + 2(20.4)$$

$$3 \text{ Median} = 30.2 + 40.8$$

$$3 \text{ Median} = 71.0$$

$$\text{Median} = \frac{71}{3} = 23.6$$

42. If mode = 16, mean = 15.6 find the median.

Sol :

Given that,

$$\text{Mean} = 15.6 \text{ and Mode} = 16$$

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

$$16 = 3 \text{ Median} - (2 \times 15.6)$$

$$16 + 31.2 = 3 \text{ Median}$$

$$47.2 = 3 \text{ Median}$$

$$\text{Median} = \frac{47.2}{3}$$

$$\text{Median} = 15.733$$

43. Given Median = 20.6, Mode = 26, find Mean.

Sol :

Given that,

$$\text{Median} = 20.6$$

$$\text{Mode} = 26$$

$$\text{Mean} = ?$$

Substituting Median and Mode values in the formula,

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

$$26 = 3(20.6) - 2 \text{ Mean}$$

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

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

$$2 \text{ Mean} = 35.8$$

$$\text{Mean} = \frac{35.8}{2}$$

$$\text{Mean} = 17.9$$

44. Mode = 32.1 and Mean = 35.4, Median = ?

Sol :

Given that,

$$\text{Mean} = 35.4$$

$$\text{Mode} = 32.1$$

$$\text{Median} = ?$$

$$32.1 = 3 \text{ Median} - (2 \times 35.4)$$

$$32.1 + 70.8 = 3 \text{ Median}$$

$$102.9 = 3 \text{ Median}$$

$$\text{Median} = \frac{102.9}{3}$$

$$\text{Median} = 34.3$$

Exercise Problems

1. Calculate the arithmetic mean from the following data :

Age (Years)	18 – 21	22 – 25	26 – 35	36 – 45	46 – 55
No. of Employees	8	32	54	36	20

[Ans: 28.91]

2. Calculate the arithmetic average from the following data :

Value	Frequency
Less than 10	4
Less than 20	16
Less than 30	40
Less than 40	76
Less than 50	96
Less than 60	112
Less than 70	120
Less than 80	125

[Ans: 37.58]

3. The mean weight of 150 students of B.Com. Class is 60 Kgs. The mean weight of boys is 70 Kgs. And that of girls is 55 Kgs. Find the number of boys and girls in the class.

[Ans: Boys 50, Girls 100]

4. The average Salary of Male Employees in a firm was Rs. 5,200 and that of Females was Rs. 4,200. The Average Salary of all employees was Rs. 5,000. Find out percentage of Male and Female Employees.

[Ans: Males-80%, Females-20%]

5. For the following Data, Compute Median.

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

[Ans: 26.32]

6. Compute median from the following data,

Mid – Values	115	125	135	145	155	165	175	185	195
Frequency	6	25	48	72	116	60	38	22	3

[Ans: 153.8]

7. From the following data of weight of 122 persons, determine the modal weight.

Weight (in lbs)	No. of Persons
100 – 110	4
110 – 120	6
120 – 130	20
130 – 140	32
140 – 150	33
150 – 160	17
160 – 170	8
170 – 180	2

[Ans: Modal Weight = 140.59 lbs]

8. The following table given 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 - 33.33]

9. From the following data compute the value of harmonic mean.

Marks	10	20	30	40	50
No. of Students	20	30	50	15	5

[Ans: H.M = 21.257]

10. If the mode and mean of a moderately asymmetrical series are respectively 16 inches and 15.6 inches, what would be its most probable median ?

[Ans: Median = 15.73]

Short Question and Answers

1. Define Average.

Ans :

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. Spiegel**, "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".

2. State the uses of Weighted Arithmetic Mean.

Ans :

- (i) Importance of all items in a series is not equal.

- (ii) Classes of the same group contain widely varying frequencies.
- (iii) It is desired to calculate the average of the series from the average of its component parts.
- (iv) Ratios, percentages or rates are being averaged or
- (v) There is a change either in the proportion of values of items or in the proportion of their frequencies.

3. Combined Mean

Ans :

If we have the arithmetic mean and number of items of two or more than two related groups, we can compute combined average of these groups by applying the following formula :

$$\bar{X}_{12} = \frac{N_1\bar{X}_1 + N_2\bar{X}_2}{N_1 + N_2}$$

$\bar{X}_{12}$ = combined mean of the two groups

$\bar{X}_1$ = arithmetic mean of first group

$\bar{X}_2$ = arithmetic mean of second group

N_1 = number of items in the first group

N_2 = number of items in the second group.

If we have to find out the combined mean of three sub-groups the above formula can be extended as follows :

$$\bar{X}_{123} = \frac{N_1\bar{X}_1 + N_2\bar{X}_2 + N_3\bar{X}_3}{N_1 + N_2 + N_3}$$

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. State the merits and demerits of geometric mean.

Ans :

Merits

- (i) It is rigidly defined. Hence, different interpretations by different persons are not possible.
- (ii) It takes all values into consideration. Thus, it is more representative.
- (iii) It can be subjected to further mathematical treatment. The properties of Geometric mean have been separately explained.
- (iv) It has a bias towards lower values.
- (v) It is not affected much by presence of extremely small or extremely large observations.
- (vi) It is not much affected by the fluctuations of sampling.

Demerits

- (i) It is neither simple to understand nor easy to calculate.
- (ii) It cannot be determined by inspection.
- (iii) It cannot be located graphically.
- (iv) It cannot be used in the study of qualitative phenomena.
- (v) It may be a fictitious value i.e. One that does not exist in the series.
- (vi) It cannot be computed if any value in a series is zero or negative.
- (vii) It brings out the property of the ratio of change and not of absolute difference as in the case of arithmetic mean.
- (viii) It cannot be calculated even if a single observation is missing or lost.

6. What is Harmonic Mean?

Ans :

Meaning

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

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

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

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

7. 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, Bi-modal and Multi Modal.

8. merits and demerits of Mode.*Ans :***Merits**

- (i) Mode is easy to understand and calculate.
- (ii) It is not influenced much by items on the extremes.
- (iii) 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.
- (iv) It can be computed for distributions which have open end classes.
- (v) Mode is not an isolated value like the median. It is the term that occurs most in the series.
- (vi) Mode is not a fictional value that is not found in the series.
- (vii) It can be determined by inspection.
- (viii) It can be located graphically
- (ix) It has wide business application

Demerits

- (i) Calculation of Mode does not consider all the items of the series. Thus, it is not fully representative of the entire data.
- (ii) It is not rigidly defined.
- (iii) It is not capable of further mathematical treatment.
- (iv) Mode is sometimes indeterminate. There may be 2 (Bi-modal) or more (Multimodal) values.
- (v) Mode is significantly impacted by fluctuations of sampling. Hence, it is less reliable.
- (vi) 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.

9. 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.

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

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."

10. 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.

Choose the Correct Answer

1. Any measure indicating the centre of a set of data, arranged in an increasing or decreasing order of magnitude, is called a measure of: [c]
(a) Skewness (b) Symmetry
(c) Central tendency (d) Dispersion
2. The measure of central tendency listed below is: [b]
(a) The raw score (b) The mean
(c) The range (d) Standard deviation
3. The total of all the observations divided by the number of observations is called: [a]
(a) Arithmetic mean (b) Geometric mean
(c) Median (d) Harmonic mean
4. While computing the arithmetic mean of a frequency distribution, the each value of a class is considered equal to: [b]
(a) Class mark (b) Lower limit
(c) Upper limit (d) Lower class boundary
5. Change of origin and scale is used for calculation of the: [a]
(a) Arithmetic mean (b) Geometric mean
(c) Weighted mean (d) Lower and upper quartiles
6. The sum of the squares of the deviations of the values of a variable is least when the deviations are measured from: [d]
(a) Harmonic mean (b) Geometric mean
(c) Median (d) Arithmetic mean
7. Step deviation method or coding method is used for computation of the: [a]
(a) Arithmetic mean (b) Geometric mean
(c) Weighted mean (d) Harmonic mean
8. If the arithmetic mean of 20 values is 10, then sum of these 20 values is: [c]
(a) 10 (b) 20
(c) 200 (d) $20 + 10$
9. The sample mean of first $\bar{X}$ natural numbers is: [d]
(a) $n(n+1)/2$ (b) $(n+1)/2$
(c) $n/2$ (d) $(n+1)/2$

10. When the values in a series are not of equal importance, we calculate the: [c]
(a) Arithmetic mean (b) Geometric mean
(c) Weighted mean (d) Mode
11. The midpoint of the values after they have been ordered from the smallest to the largest or the largest to the smallest is called: [b]
(a) Mean (b) Median
(c) Lower quartile (d) Upper quartile
12. The suitable average for qualitative data is: [b]
(a) Mean (b) Median
(c) Mode (d) Geometric mean
13. The values of the variate that divide a set of data into four equal parts after arranging the observations in ascending order of magnitude are called: [a]
(a) Quartiles (b) Deciles
(c) Percentiles (d) Difficult to tell
14. The middle value of an ordered series is called: [d]
(a) Median (b) 5th decile
(c) 50th percentile (d) All the above
15. Suitable average for averaging the shoe sizes for children is: [b]
(a) Mean (b) Mode
(c) Median (d) Geometric mean
16. A measurement that corresponds to largest frequency in a set of data is called: [c]
(a) Mean (b) Median
(c) Mode (d) Percentile
17. Which of the following average cannot be calculated for the observations 2, 2, 4, 4, 6, 6, 8, 8, 10, 10 ? [c]
(a) Mean (b) Median
(c) Mode (d) All of the above
18. The model letter of the word "STATISTICS" is: [d]
(a) S (b) T
(c) Both S and I (d) Both S and T
19. In a moderately symmetrical series, the arithmetic mean, median and mode are related as: [a]
(a) $\text{Mean} - \text{Mode} = 3(\text{Mean} - \text{Median})$ (b) $\text{Mean} - \text{Median} = 2(\text{Median} - \text{Mode})$
(c) $\text{Median} - \text{Mode} = (\text{Mean} - \text{Median})/2$ (d) $\text{Mode} - \text{Median} = 2\text{Mean} - 2\text{Median}$

20. The sum of the deviations taken from mean is: [d]
(a) Always equal to zero (b) Some times equal to zero
(c) Never equal to zero (d) Less than zero
21. Taking the relevant root of the product of all non-zero and positive values are called: [b]
(a) Arithmetic mean (b) Geometric mean
(c) Harmonic mean (d) Combined mean
22. The suitable average for computing average percentage increase in population is: [a]
(a) Geometric mean (b) Harmonic mean
(c) Combined mean (d) Population mean
23. If any value in a series is negative, then we cannot calculate the: [c]
(a) Mean (d) Median
(c) Geometric mean (d) Harmonic mean
24. The ratio among the number of items and the sum of reciprocals of items is called: [c]
(a) Arithmetic mean (b) Geometric mean
(c) Harmonic mean (d) Mode
25. Harmonic mean gives less weightage to: [b]
(a) Small values (b) Large values
(c) Positive values (d) Negative values
26. If all the items in a variable are non zero and non negative then: [a]
(a) $A.M > G.M > H.M$ (b) $G.M > A.M > H.M$
(c) $H.M > G.M > A.M$ (d) $A.M < G.M < H.M$
27. The averages are affected by change of: [c]
(a) Origin (b) Scale
(c) Both (a) and (b) (d) None of the above

Fill in the blanks

1. An _____ is an attempt to find one single figure to describe the whole of figures.
2. A _____ is a typical value around which other figures congregate
3. _____ of a series is the figure obtained by dividing the total Values of the various items by their number.
4. _____ is the n^{th} root of the product of n items of a series.
5. The reciprocal of the arithmetic mean of the reciprocal of individual observations is known as _____.
6. _____ is the value which has the greatest frequency density in its immediate neighbourhood.
7. If a group of N observations is arranged in ascending or descending order of magnitude, then the middle value is called _____.
8. Q_3 denotes _____.
9. _____ refers to the values which divide the given series into 100 equal parts.
10. Mode = 3 _____.

ANSWERS

1. Average
2. Measure of central tendency
3. Mean
4. Geometric mean
5. Harmonic mean
6. Mode
7. Median
8. Third Quartile
9. Percentiles
10. Median-2 Mean

UNIT IV

MEASURES OF DISPERSION, SKEWNESS AND KURTOSIS:

Measures of Dispersion: Significance - Characteristics - Absolute and Relative Measures - Range - Quartile Deviation - Mean Deviation- Standard Deviation- Coefficient of Variation

Measures of Skewness - Karl Pearson's Coefficient of Skewness - Bowley's Coefficient of Skewness - Kelly's Measure of Skewness - Kurtosis: Mesokurtosis, Platy kurtosis and Leptokurtosis

4.1 MEASURES OF DISPERSION

Q1. 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.

4.1.1 Significance

Q2. Explain the Significance of Measuring Dispersion.

(OR)

Explain the purpose of Dispersion.

Ans :

Measures of variation are needed for four basic purposes :

1. To determine the reliability of an average.
2. To serve as a basis for the control of the variability.
3. To compare two or more series with regard to their variability.
4. To facilitate the use of other statistical measures.

4.1.2 Characteristics

Q3. What are the Characteristics of a good Measures of Dispersion.

Ans :

- (i) It should be easy to calculate.
- (ii) It should be easy to understand.
- (iii) It should be rigidly defined.
- (iv) It should be based on all the observations.
- (v) It should be capable of further mathematical treatment.
- (vi) It should not be unduly affected by extreme observations.
- (vii) It should not be much affected by the fluctuations of sampling.

4.2 TYPES OF MEASURES OF DISPERSION

4.2.1 Absolute and Relative Measures

Q4. Explain the types of Measures of Dispersion.

Ans :

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

- i) Absolute measures and
 - ii) 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 "Coefficient(s) of Dispersion". The important methods of studying variation are listed as under:

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

Relative measures

- (i) Coefficient of Range
- (ii) Coefficient of Quartile Deviation.
- (iii) Coefficient of Mean Deviation.
- (iv) Coefficient of Variation

4.3 RANGE

Q5. 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,

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

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 :

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

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.

Q6. What are the merits and limitations of Range ?

Ans :

The merits and limitations of Range can be enumerated here.

Merits (or) Advantages

- (i) Amongst all the methods of studying dispersion range is the simplest to understand and the easiest to compute.
 - It is one of those measures which are rigidity defined.
- (ii) It takes minimum time to calculate the value of range. Hence, if one is interested in getting a quick rather than a very accurate picture of variability one may compute range.
 - It gives us the total picture of the problem even with a single glance.
 - It is used to check the quality of a product for quality control. Range plays an important role in preparing R-charts, thus quality is maintained.

Demerits (or) Limitations (or) Drawbacks

1. Range is not based on all the terms. Only extreme items reflect its size. Hence range cannot be completely representative of the data as all other middle values are ignored.
2. Due to above reason range is not a reliable measure of dispersion.
3. Range does not change even the least even if all the terms and variables are changed.
4. Range is too much affected by fluctuation of sampling. Range changes from sample to sample. As the size of sample increases range increases and vice versa.
5. It does not tell us anything about the variability of other data.
6. For open-end intervals, range is indeterminate because lower and upper limits of first and last interval are not given.

Q7. 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 ON RANGE

1. 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 $L = 500$, $S = 100$

Range (R) = $500 - 100 = 400$

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

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

2. The following are the marks of 100 students of a class. Find the range of variation of marks and also compute coefficient of range.

Marks	10-20	20-30	30-40	40-50	50-60	60-70	70-80
No. of students	8	16	26	20	12	10	8

Sol :

From the given data

Highest (L) marks is 80 and

Lowest (S) marks is 10. Hence

Range (R) = $L - S$

(R) = $80 - 10 = 70$

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

Coefficient of Range = $\frac{80 - 10}{80 + 10}$

Coefficient of Range = $\frac{70}{90}$

Coefficient of Range = 0.77.

3. 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 :

(May-18, Imp.)

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

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

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

4.4 QUARTILE DEVIATION

Q8. What is Quartile Deviations ?

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.

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

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

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

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

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

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

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

Q9. 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.

Q10. How to calculate quartile deviations?

Ans :

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

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

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

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

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 ON QUARTILE DEVIATION

4. 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

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

N = No. of observations.

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

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

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

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

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

5. 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 :

(Dec.-17)

Rearrange into ascending order

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

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

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

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

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

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

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

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

$$Q.D = 18$$

6. Compute Quartile Deviation and its coefficient for the following data.

Height (in inches)	58	59	60	61	62	63	64	65	66
No. of Students	15	20	32	35	33	22	20	10	8

Sol :

Calculation of Quartile Deviation

Height	f	C.f
58	15	15
59	20	35
60	32	67 (Q1)
61	35	102
62	33	135
63	22	157 (Q3)
64	20	177
65	10	187
66	8	195
	N = 195	

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

N = No. of observations.

$$= \text{Size of } \left[\frac{195+1}{4} \right]^{\text{th}} \text{ Item}$$

$$= 49^{\text{th}} \text{ Item} = 60$$

$$\therefore Q_1 = 60$$

$$Q_3 = \text{Size of } 3 \left[\frac{195+1}{4} \right]^{th} \text{ Item}$$

$$= \text{Size of } 3 \left[\frac{196}{4} \right]^{th} \text{ Item} = 147^{th} \text{ Item} = 63$$

$$\therefore Q_3 = 63$$

$$\text{Quartile Deviation (Q.D)} = \frac{Q_3 - Q_1}{2} = \frac{63 - 60}{2} = 1.5$$

$$\text{Coefficient of Q.D} = \frac{Q_3 - Q_1}{Q_3 + Q_1} = \frac{63 - 60}{63 + 60} = \frac{3}{123} = 0.024$$

7. 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 51th Item lies in CF 54. So Q_1 class interval = 20–30

$$\begin{aligned}
 Q_1 &= L + \frac{\frac{N}{4} - F}{f} \times C \\
 &= 20 + \frac{51 - 29}{25} \times 10 \\
 &= 20 + \frac{22 \times 10}{25} \\
 &= 20 + \frac{220}{25} \\
 &= 20 + 8.8 = 28.8.
 \end{aligned}$$

Calculation of Q_3

$$\begin{aligned}
 &= Q_3 = 3\left(\frac{N}{4}\right)^{th} \\
 &= 3\left(\frac{204}{4}\right)^{th} \text{ Item} = 3(51)^{th} \text{ Item} = 153^{th} \text{ Item}
 \end{aligned}$$

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

$$\begin{aligned}
 Q_3 &= L_3 + \frac{\frac{3N}{4} - F}{f} \times C \\
 &= 60 + \frac{153 - 145}{22} \times 10 \\
 &= 60 + \frac{80}{22} \\
 &= 60 + 3.64 \\
 &= 63.64
 \end{aligned}$$

Calculation of Quartile Deviation

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

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

4.5 MEAN DEVIATION

Q11. 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.

Q12. 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

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

Steps

- i) Compute the median of the series.
- ii) The deviations of items from median ignoring $\pm$ signs and denote these deviations by $|D|$.
- iii) Obtain the total of these deviations, i.e., $\sum |D|$.
- 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.

Example

Calculate mean deviation and its coefficient from (i) Arithmetic mean (ii) Median in respect of marks obtained by 7 students given below

Marks (Out of 25) 7, 5, 12, 10, 8, 11, 6

Sol. :

Steps

- (i) Rearrange the data in Ascending order i.e., 5, 6, 7, 8, 9, 10, 11, 12
- (ii) Determine the required i.e., (a) Arithmetic Mean (b) Median

$$\begin{aligned}
 \text{(a) Arithmetic mean } (\bar{X}) &= \frac{\Sigma X}{N} \\
 &= \frac{3+6+7+8+10+11+12}{7} \\
 &= \frac{59}{7} = 8.43
 \end{aligned}$$

$$\text{(b) Median} = \text{Size of } \left(\frac{N+1}{2} \right)^{\text{th}} \text{ term}$$

$$\text{Median} = \text{Size of } \left(\frac{7+1}{2} \right) = \frac{8}{2}$$

= 4 Size of 4th item

Median = Size of 4th item is 8.

Marks	Deviation (D) from mean = 8.43	Deviation from Median = 8
5	3.43	3
6	2.43	2
7	1.43	1
8	0.43	0
10	1.57	2
11	2.57	3
12	3.57	4
	15.43	15

$$\text{(i) Mean Deviation from Mean} = \frac{\Sigma |D|}{N}$$

$$\text{Mean Deviation from Mean} = \frac{15.43}{7}$$

$$\text{Mean Deviation from Mean} = 2.204$$

$$\text{(ii) Mean Deviation from Mean} = \frac{\Sigma |D|}{N}$$

$$\text{Mean Deviation from Mean} = \frac{15}{7}$$

$$\text{Mean Deviation from Mean} = 2.143$$

(iii) Coefficient of Mean Deviation

$$\begin{aligned}
 \text{(a) Through Mean} &= \frac{MD}{\bar{X}} \\
 &= \frac{2.204}{8.43} \\
 &= 0.262
 \end{aligned}$$

$$\begin{aligned}
 \text{(b) Through Median} &= \frac{MD}{\text{Median}} \\
 &= \frac{2.143}{8} \\
 &= 0.268
 \end{aligned}$$

Q13. How to calculate mean deviation for discrete series ?

Ans :

In discrete series the formula for calculating mean deviation is

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

|D| denotes deviation from median ignoring signs.

N = Sum of frequency

Steps

- Calculate the median of the series.
- Take the deviations of the items from median ignoring signs and denote them by |D|.
- Multiply these deviations by the respective frequencies and obtain the total $\Sigma f |D|$
- Divide the total obtained in Step (ii) by the number of observations. This gives us the value of mean deviation.

Example

Calculate and Compare the coefficient of M.D. through median.

Age (years)	16	17	18	19	20	21	22	23	24
No. of Boys	4	5	7	12	20	13	5	0	4
No. of Girls	2	0	4	8	15	10	6	3	2

Sol :

i) Calculation M.D for boys

X	f	X - M = D (M = 20)	f D	Cf
16	4	+4	16	4
17	5	+3	15	9
18	7	+2	14	16
19	12	+1	12	28
20	20	0	0	48
21	13	1	13	61
22	5	2	10	66
23	0	3	0	66
24	4	4	16	70
	70		96	

$$\begin{aligned}\text{Median} &= \frac{N+1}{2} \text{th item} \\ &= \frac{70+1}{2} = 35.5 \text{th item}\end{aligned}$$

35.5th item is representing in Cf at 20.

Hence median = 20

$$\text{M.D} = \frac{\sum f|D|}{N} = \frac{96}{70} = 1.371$$

$$\begin{aligned}\text{Co-efficient of M.D} &= \frac{\text{MD}}{\text{median}} = \frac{1.371}{20} \\ &= 0.0685\end{aligned}$$

ii) Calculation of MD for Girls

X	f	X - M = D (M = 20)	f D	Cf
16	2	4	8	2
17	0	3	0	2
18	4	2	8	6
19	8	1	8	14
20	15	0	0	29
21	10	1	10	39
22	6	2	12	45
23	3	3	9	48
24	2	4	8	50
	50		63	

$$\text{a) Median} = \frac{N+1}{2} \text{th item}$$

$$= \frac{50+1}{2} = \frac{51}{2}$$

$$= 25.5 \text{th item}$$

25.5th item is representing in Cf at 20.

Hence, Median = 20.

$$\text{b) M.D} = \frac{\sum f|D|}{N}$$

$$= \frac{63}{50}$$

$$= 1.26$$

$$\text{c) Co-efficient of M.D.} = \frac{\text{MD}}{\text{median}}$$

$$= \frac{1.26}{20}$$

$$= 0.063$$

Q14. 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.,

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

Steps

- Calculate the median of the series.
- Take the deviations of the items from median ignoring signs and denote them by |D|.
- Multiply these deviations by the respective frequencies and obtain the total $\sum f|D|$.
- Divide the total obtained in step (ii) by the number of observations. This gives us the value of mean deviations.

PROBLEMS

8. Calculate mean deviation of the following from mean and median

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

Sol. : (Dec.-18)

X	$X - \bar{X} = D $	$X - \text{Median} = D $
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 ($\bar{X}$) = $\frac{\Sigma x}{N} = \frac{91}{7} = 13$

MD through mean = $\frac{\Sigma |D|}{N} = \frac{40}{7} = 5.71$

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

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

= $\frac{8}{2} = 4^{\text{th}}$ observation = 14

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

Mean Deviation (MD) through

Median = $\frac{\Sigma |D|}{M}$

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

Mean Deviation (MD) = 5.57

9. Calculate the mean deviation and its coefficient from the following data.

Income of Members (₹)
1,000
1,500
2,500
3,000
3,500
4,000

Sol :

Calculation of Mean Deviation

Income of Members (x)	Deviation from Median $ D = x - A (2500)$
1,000	1,500
1,500	1,000
2,000	500
2,500	0
3,000	500
3,500	1,000
4,000	1,500
N = 7	$\Sigma D = 6,000$

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

= $\frac{7+1}{2} = \frac{8}{2} = 4^{\text{th}}$ item,

Median deviation = $\frac{\Sigma |D|}{N}$

Where,

$|D|$ = Deviation from median ignoring signs.

M.D = $\frac{6,000}{7} = 857.14$

$\therefore$ Coefficient of M.D = $\frac{857.14}{2500} = 0.343$

10. 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 :

(Dec.-17, 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				$\Sigma f D = 582$

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

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

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.

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

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

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

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

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

$$= 30.5 + 1 = 31.5$$

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

11. Calculate the mean deviation and the coefficient from the following data.

Class	0 – 10	10 – 20	20 – 30	30 – 40	40 – 50	50 – 60	60 – 70	70 – 80
Frequency	5	8	12	15	20	14	12	6

Sol :

Calculation of Mean Deviation from Median

Class	Frequency (f)	Cumulative Frequency	Mid-value (m)	D = m – md D = m – 43	f x – md f D
0 – 10	5	5	5	38	190
10 – 20	8	13	15	28	224
20 – 30	12	25	25	18	216
30 – 40	15	40 F	35	8	120
L 40 – 50	20 f	60	45	2	40
50 – 60	14	74	55	12	168
60 – 70	12	86	65	22	264
70 – 80	6	92	75	32	192
	N = 92				Σf D = 1414

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

$$= \frac{92}{2} = 46^{\text{th}} \text{ item}$$

∴ Median lies between 40 – 50

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

$$= 40 + \frac{46 - 40}{20} \times 10$$

$$= 40 + \frac{6}{20} \times 10$$

$$= 40 + \frac{6}{20} = 40 + 3 = 43$$

$$(i) \quad \text{Mean deviation (M.D)} = \frac{\Sigma F |D|}{N} = \frac{1414}{92} = 15.37$$

$$(ii) \quad \text{Coefficient of M.D} = \frac{\text{M.D}}{\text{Median}} = \frac{15.37}{43} = 0.36$$

4.6 STANDARD DEVIATION

Q15. 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.

Q16. 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.

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

where as

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

N = Total number of observations.

Steps

- Calculate the actual for the series i.e. $\bar{X}$
- Find X deviation from $\bar{X}$ i.e. $x = (X - \bar{X})$
- Square these deviations and obtain the total $\sum x^2$
- Divide $\sum x^2$ by the total number of observations and extract the square root.

Example

Find standard deviation of the following distribution.

10, 12, 14, 16, 18, 20, 22, 24, 26, 28

Sol :

X	$x = (X - \bar{X})$	x^2
10	- 9	81
12	- 7	49
14	- 5	25
16	- 3	9
18	- 1	1
20	1	1
22	3	9
24	5	25
26	7	49
28	9	81
190		330

$$i) \quad \bar{X} = \frac{\sum X}{N} = \frac{190}{10} = 19$$

$$ii) \quad \sigma = \sqrt{\frac{\sum x^2}{N}} = \sqrt{\frac{330}{10}} = \sqrt{33}$$

$$\sigma = 5.745.$$

Deviations from Assumed Mean

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

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 $\sum d$.
- Square these deviations and obtain the total $\sum d^2$.
- Substitute the value of $\sum d^2$, $\sum d$ and N in the above formula.

Q17. 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 :

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

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.

Example

Calculate Standard Deviation for the following distribution.

Size of item	4	5	6	7	8	9	10
Frequency	6	8	10	15	12	10	9

Sol :

Size of Item(X)	f	fx	$x = (X - \bar{X})$	x^2	fx^2
4	6	24	-3.21	10.3041	61.8246
5	8	40	-2.21	4.8841	39.0128
6	10	60	-1.21	1.4641	14.6410
7	15	105	-0.21	0.0441	0.6615
8	12	96	0.79	0.6241	7.4892
9	10	90	1.79	3.2041	32.041
10	9	90	2.79	7.7841	70.0569
	70	505			225.787

i) Mean $(\bar{X}) = \frac{\sum fx}{N} = \frac{505}{70} = 7.21$

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

$$\sigma = \sqrt{\frac{225.787}{70}} = \sqrt{3.2255}$$

$$\sigma = 1.79$$

(b) Assumed Mean Method. When this method is used, the following formula is applied :

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

where $d = (X - A)$

N = Sum of frequency

Steps

- Take the deviations of the items from an assumed mean and denote these deviations by d .
- Multiply these deviations by the respective frequencies and obtain the total, $\sum fd$.
- Obtain the squares of the deviations, i.e., calculate d^2 .
- Multiply the squared deviations by the respective frequencies, and obtain the total, $\sum fd^2$.
- Substitute the values in the above formula.

Example

Calculate Standard Deviation for the following distribution.

Size of item	4	5	6	7	8	9	10
Frequency	6	8	10	15	12	10	9

Sol :

Deviations from Assumed Mean

X	f	X-A=d	d ²	fd	fd ²
4	6	-3	9	-18	54
5	8	-2	4	-16	32
6	10	-1	1	-10	10
7A	15	0	0	0	0
8	12	1	1	12	12
9	10	2	4	20	40
10	9	3	9	27	81
	70			15	229

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

$$\sigma = \sqrt{\frac{229}{70} - \left(\frac{15}{10}\right)^2}$$

$$\sigma = \sqrt{3.27 - (0.21)^2}$$

$$\sigma = \sqrt{3.27 - 0.0441}$$

$$\sigma = \sqrt{3.2259}$$

$$\sigma = 1.79$$

- (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 :

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

where,

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

The use of the above formula simplifies calculations.

Example

The annual salaries of a group of employees are given in the following table :

Salaries (in Rs. 000)	45	50	55	60	65	70	75	80
Number of persons	3	5	8	7	9	7	4	7

Calculate the standard deviation of the salaries.

Sol. :

Calculation of Standard Deviation

Salaries X	No. of persons f	(X - 60)/5	fd	fd ²
45	3	-3	-9	27
50	5	-2	-10	20
55	8	-1	-8	8
60	7	0	0	0
65	9	+1	+9	9
70	7	+2	+14	28
75	4	+3	+12	36
80	7	+4	+28	112
N = 50			Σfd = 36	Σfd ² = 240

$$\sigma = \sqrt{\frac{\sum fd^2}{N} - \left(\frac{\sum fd}{N}\right)^2} \times C \quad \text{Here, } \sum fd^2 = 240, N = 50, \sum fd = 36, C = 5$$

$$\sqrt{\frac{240}{50} - \left(\frac{36}{50}\right)^2} \times 5$$

$$\sigma = \sqrt{\frac{\sum fd^2}{N} - \left(\frac{\sum fd}{N}\right)^2} \times 5 = \sqrt{4.8 - 5184} \times 5 = 10.35$$

Q18. 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

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

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

Steps

- Find the mid-points of various classes.
- Take the deviations of these mid-points from an assumed mean and denote these deviations by d.
- Wherever possible take a common factor and denote this column by d.
- Multiply the frequencies of each class with these deviations and obtain Σfd.
- Square the deviations and multiply them with the respective frequencies of each class and obtain Σfd².

Example

Find out standard deviation.

Age in years	0-10	10-20	20-30	30-40	40-50
No. of persons	18	30	22	10	6

Sol :

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

Age in years	f	Mid value	m - A = d	d	d ²	fd ²	fd
0 - 10	18	5	-20	-2	4	72	-36
10 - 20	30	15	-10	-1	1	30	-30
20 - 30	22	(25) A	0	0	0	0	0
30 - 40	10	35	10	1	1	10	10
40 - 50	6	45	20	2	4	24	12
	86					136	-44

$$\sigma = \sqrt{\frac{136}{86} - \left[\frac{-44}{86}\right]^2} \times 10$$

$$\sigma = \sqrt{1.5814 - (0.51)^2} \times 10$$

$$\sigma = \sqrt{1.5814 - (-0.2618)} \times 10$$

$$\sigma = \sqrt{1.3196} \times 10$$

$$\sigma = 1.149 \times 10$$

$$\sigma = 11.49.$$

4.6.1 Coefficient of Variation**Q19. Define Coefficient of Variation.**

Ans :

Coefficient of Variation (CV) was proposed by Karl Pearson. It is used to compare the variability of two (or) more distributions. A distribution with greater C.V. is considered as more variable or less consistent, less unit form, less stable or less homogeneous distribution and the distribution with less C.V. is considered as less variable or more consistent, more uniform, more stable or more homogeneous distribution.

$$\text{Coefficient of variation} = \frac{\text{Standard deviation}}{\text{Arithmetic mean}} \times 100$$

PROBLEMS ON STANDARD DEVIATION

12. 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
	$\Sigma X = 2128$		$\Sigma x^2 = 30120$

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

$$= 266$$

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

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

$$= \sqrt{3765}$$

$$= 61.35$$

13. 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

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

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

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

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

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

$$\sigma = 13.26$$

14. 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

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

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

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

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

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

$$35 - 0.16 = 34.84$$

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

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

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

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

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

15. 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

Sol :

(Dec.-18, Imp.)

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	54	18
70 - 80	2	75	4	16	32	8
	83				221	13

$$\begin{aligned}\text{Mean} &= A + \frac{\Sigma fd}{N} \times C \\ &= 35 + \frac{13}{83} \times 10 \\ &= 35 + 15 = 36.5\end{aligned}$$

$$\begin{aligned}\text{S.D} &= \sqrt{\frac{\Sigma fd^2}{N} - \left(\frac{\Sigma fd}{N}\right)^2} \times C \\ &= \sqrt{\frac{221}{83} - \left(\frac{13}{83}\right)^2} \times 10 \\ &= \sqrt{2.66 - 0.024} \times 10 \\ &= \sqrt{2.636} \times 10 \\ &= 1.623 \times 10 \\ &= 16.23\end{aligned}$$

$$\begin{aligned}\text{Coefficient of variation} &= \frac{\text{S.D}}{\text{Mean}} \times 100 \\ &= \frac{16.23}{36.5} \times 100 \\ &= 0.44 \times 100 = 44.46\end{aligned}$$

16. From the marks of two students given below. Find out who is more consistent.

A	25	50	45	30	70	42	36	48	34	60
B	10	70	50	20	95	55	42	60	48	80

Sol :

(June-19)

Calculation of Standard Deviation

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

$$\bar{X}_A = \frac{440}{10} = 44$$

$$\bar{X}_B = \frac{530}{10} = 53$$

Calculation of Standard Deviation

$$b_A = \sqrt{\frac{\sum xA^2}{N}} = \sqrt{\frac{1710}{10}} = \sqrt{171} = 13.076$$

$$b_B = \sqrt{\frac{\sum x_B^2}{N}} = \sqrt{\frac{5928}{10}} = \sqrt{592.8} = 24.347$$

$$\text{Coefficient of variation A} = \frac{\text{S.D of A}}{\bar{X}_A} \times 100 = \frac{13.076}{44} \times 100 = 29.72$$

$$\text{Coefficient of variation B} = \frac{\text{S.D of B}}{\bar{X}_B} \times 100 = \frac{24.347}{53} \times 100 = 45.94$$

The coefficient of variation 'B' is more consistent.

17. Find the standard deviation and coefficient of variation from the following data

Marks	Upto 10	Upto 20	Upto 30	Upto 40	Upto 50	Upto 60	Upto 70	Upto 80
No. of Students	12	30	65	107	157	202	222	230

Sol:

Converting the given cumulative frequencies into simple frequencies and taking class intervals as 0 – 10, 10 – 20, 20 – 30 etc.

Calculation of Standard Deviation and Coefficient of Variation

Marks	Mid point	No. of Student (f)	$d = \frac{X - 35}{10}$ (A = 35)	d^2	fd	fd ²
0 – 10	5	12	-3	9	-36	108
10 – 20	15	18	-2	4	-36	72
20 – 30	25	35	-1	1	-35	35
30 – 40	(35) A	42	0	0	0	0
40 – 50	45	50	1	1	50	50
50 – 60	55	45	2	4	90	180
60 – 70	65	20	3	9	60	180
70 – 80	75	8	4	16	32	128
		N = 230			Σfd = 125	Σfd ² = 753

$$\text{Standard Deviation } (\sigma) = \sqrt{\left[\frac{\sum fd^2}{N} \right] - \left(\frac{\sum fd}{N} \right)^2} \times C$$

$$\sum fd^2 = 753; \sum fd = 125; N = 230 \text{ and } C = 10$$

$$\sigma = \sqrt{\left[\frac{753}{230} \right] - \left(\frac{125}{230} \right)^2} \times 10$$

$$\sigma = \sqrt{3.274 - (0.543)^2} \times 10$$

$$\sigma = \sqrt{3.274 - 0.295} \times 10 = \sqrt{2.979} \times 10 = 1.726 \times 10 = 17.26$$

$$\therefore \sigma = 17.26$$

$$\text{Coefficient of Variation (C.V)} = \frac{\sigma}{\bar{X}} \times 100$$

$$\sigma = 17.26, \bar{X} = ?$$

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

$$A = 35; \sum fd = 125; N = 230 \text{ and } C = 10$$

$$\bar{X} = 35 + \frac{125}{230} \times 10 = 35 + 0.543 \times 10 = 35 + 5.43 = 40.43$$

$$\therefore \bar{X} = 40.43$$

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

$$\text{C.V} = \frac{17.26}{40.43} \times 100 = 0.4269 \times 100$$

$$\therefore \text{C.V} = 42.69.$$

18. From the following data, which of the two makes shows more consistency

Life (No. of Years)	0-5	5-10	10-15	15-20	20-25
Make x No. of cars	8	12	17	10	3
Make y No. of cars	6	10	20	12	2

Sol :

Life	x	y	$x = x - \bar{x}$	$y = y - \bar{y}$	x^2	y^2
0-5	8	6	-2	-4	4	16
5-10	12	10	2	0	4	0
10-15	17	20	7	+10	49	100
15-20	10	12	0	+2	0	4
20-25	3	2	-7	-8	49	64
			0	0	106	184

By using coefficient of variation we conclude that which of the two makes shows more consistency

$$\text{Coefficient of variation} = \frac{\sigma}{\bar{x}} \times 100$$

$$\text{Mean of x series } (\bar{x}) = \frac{\Sigma x}{n} = \frac{50}{5} = 10$$

$$(\bar{y}) = \frac{\Sigma y}{n} = \frac{50}{5} = 10$$

$$\text{S.D of x series } \sigma x = \sqrt{\left(\frac{\Sigma(x - \bar{x})^2}{n}\right)} = \sqrt{\frac{106}{5}} = 4.604$$

$$\text{S.D of y series } \sigma y = \sqrt{\left(\frac{\Sigma(y - \bar{y})^2}{n}\right)} = \sqrt{\frac{184}{5}} = 6.066$$

$$\text{Coefficient of variation x} = \frac{\sigma}{\bar{x}} \times 100$$

$$\frac{4.604}{10} \times 100 = 46.04$$

$$\text{Coefficient of variation y} = \frac{\sigma}{\bar{y}} \times 100$$

$$\frac{6.066}{10} \times 100 = 60.06$$

Comparing two makes x is less than y so y is more consistency.

4.7 MEASURES OF SKEWNESS

Q20. Explain briefly about Skewness.

Ans :

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** "Measures 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) $\text{Mean} > \text{Median} > \text{Mode}$
- (ii) Right tail is longer than its left tail
- (iii) $(Q_3 - \text{Median}) > (\text{Median} - Q_1)$.

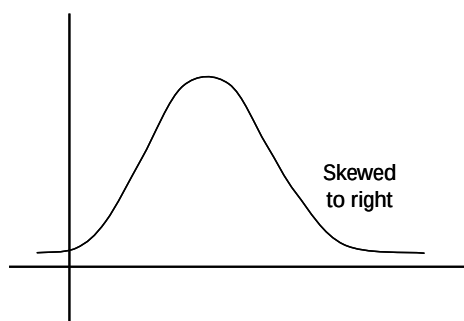


Fig.: Positively Skewed Distribution

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) $\text{Mean} < \text{Median} < \text{Mode}$
- (ii) Left tail is longer than its right tail
- (iii) $(Q_3 - \text{Median}) < (\text{Median} - Q_1)$

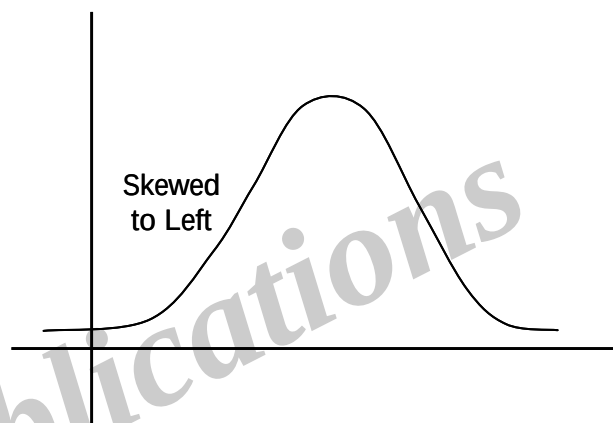


Fig.: Negatively Skewed Distribution

4.7.1 Karl Pearson's Coefficient of Skewness - Bowley's Coefficient of Skewness - Kelly's Measure of Skewness

Q21. 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 :

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

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

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

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.
- (ii) The Kelly'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.

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

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

where, M = Median

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

(iii) Kelly's Coefficient of Skewness

Bowley's measure discussed above neglects the two extreme quarters of the data. It would be better for a measure to cover the entire data especially because in measuring skewness, we are often interested in the more extreme items. Bowley's measure can be extended by taking any two deciles equidistant from the median or any two percentiles equidistant from the median. Kelly has suggested the following formula for measuring skewness upon the 10th and the 90th percentiles (or the first and ninth deciles) :

$$Sk_K = \frac{P_{10} + P_{90} - 2 \text{ Med.}}{P_{90} - P_{10}} \text{ also}$$

$$Sk = \frac{D_1 + D_9 - 2 \text{ Med.}}{D_9 - D_1}$$

Sk_K = Kelly's coefficient of skewness.

This measure of skewness has one theoretical attraction if skewness is to be based on percentiles. However, this method is not popular in practice and generally Karl Pearson's method is used.

PROBLEMS ON KARL PEARSON COEFFICIENT OF SKEWNESS

19. 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

$$\begin{aligned} &= \frac{\text{Mean} - \text{Mode}}{\text{Standard Deviation}} = \frac{30.08 - 33.5}{13.405} \\ &= - 0.255 \end{aligned}$$

Co-efficient of Variation

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

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

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

Sol :

Co-efficient of Skewness

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

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

21. 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 : (Dec.-17)

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

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

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

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

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

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

$$\sigma = 22.5$$

22. 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 : (May-18)

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.

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

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

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

$\therefore$ Standard Deviation (s) = 16

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

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

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

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

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

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

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

Median can be calculated as follows,

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

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

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

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

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

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

23. 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 :

(June-19)

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 - 70	10	65	3	9	30	90
	<u>200</u>				-34	454

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

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

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

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

$$= 35 - 1.7 = 33.3$$

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

$$\Delta_1 = f - f_1$$

$$= 60 - 40 = 20$$

$$\Delta_2 = f - f_2$$

$$= 60 - 32 = 28$$

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

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

$$= 30 + 4.1 = 34.1$$

$$\begin{aligned}
 S.D &= \sqrt{\left(\frac{\Sigma fd^2}{N}\right) - \left(\frac{\Sigma fd}{N}\right)^2} \times C \\
 &= \sqrt{\frac{454}{200} - \left(\frac{-34}{200}\right)^2} \times 10 = \sqrt{2.27 + 0.0289} \times 10 \\
 &= \sqrt{2.2989} \times 10 = 1.516 \times 10 = 15.16 \\
 S_{KP} &= \frac{\text{Mean} - \text{Mode}}{S.D} = \frac{33.3 - 34.1}{15.16} = \frac{-0.18}{15.16} \\
 &= -0.0531
 \end{aligned}$$

24. Calculate Karl Pearson Coefficient of Skewness from the following data,

Wages	No. of Students
25 – 75	25
75 – 125	30
125 – 175	40
175 – 225	35
225 – 275	30
275 – 325	25
325 – 375	30
375 – 425	45

Sol :

Calculation of Karl Pearson's Coefficient of Skewness

Wages	Mid points (X)	f	$d = \frac{X - A}{10}$ A = 200	d ²	fd	fd ²
25 – 75	50	25	– 15	225	– 375	1,40,625
75 – 125	100	30	– 10	100	– 300	90,000
125 – 175	150	40	– 5	25	– 200	40,000
175 – 225	(200) A	35	0	0	0	0
225 – 275	250	30	5	25	150	22,500
275 – 325	300	25	10	100	250	62,500
325 – 375	350	30	15	225	450	2,02,500
375 – 425	400	45	20	400	900	8,10,000
		N = 260			Σfd = 875	Σfd² = 13,68,125

Karl Pearson's coefficient of Skewness,

$$S_{kp} = \frac{\text{Mean} - \text{Mode}}{\sigma}$$

Calculation of Mean ($\bar{X}$)

$$\text{Mean } (\bar{X}) = A + \frac{\Sigma fd}{N} \times C$$

$A = 200, \Sigma fd = 875,$
 $N = 260 \text{ and } C = 50$

$$\text{Mean } (\bar{X}) = 200 + \frac{875}{260} \times 50$$

$$\text{Mean } (\bar{X}) = 368.26$$

Calculation of Mode

Mode lies in the class 175 – 225

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

$$L = 175$$

$$\Delta_1 = f - f_1$$

$$\Delta_1 = 35 - 40 = 5$$

$$\Delta_2 = f - f_2$$

$$\Delta_2 = 35 - 30 = 5$$

$$C = 50$$

$$\begin{aligned} \text{Mode} &= 175 + \frac{5}{5+5} \times 50 \\ &= 175 + 25 \\ &= 200 \end{aligned}$$

Calculation of Standard Deviation (σ)

$$\begin{aligned} \sigma &= \sqrt{\left(\frac{\Sigma fd^2}{N} \right) - \left(\frac{\Sigma fd}{N} \right)^2} \times C \\ &= \sqrt{\frac{1,368,125}{260} - \left(\frac{875}{260} \right)^2} \times 50 \\ &= \sqrt{5262.73} \times 50 \\ &= 3623.09 \end{aligned}$$

Karl Pearson coefficient of Skewness,

$$\begin{aligned} S_{kp} &= \frac{\text{Mean} - \text{Mode}}{\sigma} \\ &= \frac{368.26 - 200}{3623.09} \end{aligned}$$

$$\therefore S_{kp} = 0.046.$$

25.

Measure	Place - A	Place - B
Mean	150	140
Median	142	155
Standard deviation	30	55
Third quartile	195	260
First quartile	62	80

From the above information calculate Karl Pearson's coefficients of skewness and also Bowley's coefficient skewness for both places separately.

Sol :

Karl Pearson's coefficient of skewness

$$S_{kp} = \frac{3(\text{Mean} - \text{Median})}{\text{Standard Deviation}(\sigma)}$$

Bowley's coefficient of skewness,

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

Place A

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

$$= \frac{3(150 - 142)}{30}$$

$$= \frac{3(8)}{30} = \frac{24}{30}$$

$$S_{kp} = 0.8$$

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

$$= \frac{195 + 62 - 2(142)}{195 - 62}$$

$$= \frac{257 - 284}{133}$$

$$= \frac{-27}{133}$$

$$S_{kb} = -0.203$$

Place B

$$S_{kp} = \frac{3(\text{Mean} - \text{Median})}{\sigma} = \frac{3(140 - 155)}{55}$$

$$= \frac{3(-15)}{55} = \frac{-45}{55}$$

$$S_{kp} = -0.82$$

$$S_{kb} = \frac{Q_3 + Q_1 - 2 \text{ Median}}{Q_3 - Q_1} = \frac{260 + 80 - 2(155)}{260 - 80}$$

$$= \frac{340 - 310}{180} = \frac{30}{180}$$

$$S_{kb} = 0.167$$

PROBLEMS ON BOWLEY'S COEFFICIENT OF SKEWNESS

- 26. 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

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

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

- 27. From the data given below, calculate and Bowley's Coefficients of Skewness.**

Mean = 160, Median = 152, Q_1 = 72, Q_3 = 205

and Standard Deviation = 40.

Sol :

Bowley's Co-efficient of Skewness

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

$$= \frac{205 + 72 - 2(152)}{205 - 72}$$

$$= \frac{205 + 72 - 304}{133} = \frac{-27}{133} = -0.203$$

28. From the following marks obtained by 7 students in a tutorial group, calculate Bowley's Coefficient of Skewness.

Roll No.	1	2	3	4	5	6	7
Marks	1	5	12	22	17	9	4

Sol :

Calculation of Bowley's coefficient of skewness. Arranging the data in an ascending order.

Marks (X) : 1 4 5 9 12 17 22

$$i) \quad Q_1 = \text{Size of } \left(\frac{N+1}{4} \right) \text{th item} = \left(\frac{7+1}{4} \right) \text{th item} = 2\text{nd item} = 4$$

$$ii) \quad Q_3 = \text{Size of } \frac{3(N+1)}{4} \text{th item} = \frac{3(7+1)}{4} \text{th item} = 6\text{th item} = 17$$

$$\text{Median} = \text{Size of the } \left(\frac{N+1}{2} \right) \text{th item} = \left(\frac{7+1}{2} \right) \text{th item} = 4\text{th item} = 9$$

$$\begin{aligned} \text{Bowley's coefficient of skewness} &= \frac{Q_3 + Q_1 - 2 \text{ Median}}{Q_3 - Q_1} \\ &= \frac{17 + 4 - 2 \times 9}{17 - 4} = \frac{21 - 18}{13} = \frac{3}{13} = + 0.231 \end{aligned}$$

29. 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

Sol :

(May-18, Imp.)

	Variable	Frequency	Cumulative frequency
	10-20	5	5
	20-30	9	14F
L	30-40	14 f	28 (Q ₁)
	40-50	20	48F
L	50-60	25 f	73F (Median)
L	60-70	15 f	88 (Q ₃)
	70-80	8	96
	80-90	4	100=N

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

Calculation of Median

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

Median lies in the class of 50 – 60

$$\begin{aligned} &= 50 + \frac{50 - 48}{25} \times 10 \\ &= 50 + \frac{2 \times 10}{25} \\ &= 50 + \frac{20}{25} = 50 + 0.8 = 50.8 \end{aligned}$$

Calculation of Q_3 and Q_1

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

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

Q_3 lies in the class of 60 – 70

$$\begin{aligned} &= 60 + \frac{75 - 73}{15} \times 10 = 60 + \frac{2 \times 10}{15} \\ &= 60 + \frac{20}{15} = 60 + 1.33 = 61.33 \end{aligned}$$

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

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

Q_1 lies in the class of 30 – 40

$$\begin{aligned} &= 30 + \frac{25 - 14}{14} \times 10 \\ &= 30 + \frac{11}{14} \times 10 \\ &= 30 + \frac{110}{14} \\ &= 30 + 7.86 = 37.86 \end{aligned}$$

Bowley's coefficient of skewness

$$\begin{aligned} &= \frac{Q_3 + Q_1 - 2 \text{ median}}{Q_3 - Q_1} = \frac{61.33 + 37.86 - 2(50.8)}{61.33 - 37.86} \\ &= \frac{99.19 - 101.6}{23.47} = \frac{-2.41}{23.47} \end{aligned}$$

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

30. 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.

(Dec.-18, Imp.)

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

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

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

Median lies in the class of 40 – 50

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

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

$$= 40 + 4.33 = 44.33$$

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 $Q_1 = Q_1$

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

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

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 skewness

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

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

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

$$= 0.0428.$$

31. 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 :

(June-19)

Quartile Deviation (Q.D) = 40,

Median = 32

$SK_B = 0.5$

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

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

$$= \theta_3 - \theta_1$$

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

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

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

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

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

$$40 - 16 = 2\theta_1$$

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

$$= 12$$

$$\therefore Q_3 = 80 + Q_1$$

$$= 80 + 12$$

$$= 92$$

PROBLEM ON KELLY'S COEFFICIENT OF SKEWNESS

32. Calculate percentile co-efficient of skewness from the following positional measures given below:

$$P_{90} = 110,$$

$$P_{10} = 50.67,$$

$$\text{Median} = 75.03$$

Sol :

Kelly's Coefficient of Skewness (Using Percentiles)

$$\begin{aligned} S_k &= \frac{P_{10} + P_{90} - 2 \text{ Median}}{P_{90} - P_{10}} \\ &= \frac{50.67 + 110 - 2(75.03)}{100 - 50.67} \\ &= \frac{160.67 - 150.06}{59.33} \\ &= \frac{10.61}{59.33} \\ &= 0.179 \end{aligned}$$

Kelly's Coefficient of Skewness = 0.179.

4.8 KURTOSIS
4.8.1 Mesokurtosis, Platy Kurtosis and Leptokurtosis

Q22. Define kurtosis, Explain briefly about measuring 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

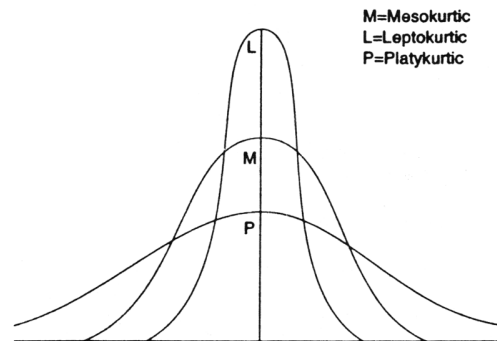
Measures of kurtosis tell us the extent to which a distribution is more peaked or flat-topped than the normal curve. If a curve is more peaked than the normal curve, it is called 'leptokurtic'. In such a case items are more closely bunched around the mode.

(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'.

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.

Rahul Publications

Exercise Problems

1. Calculate median and mean Deviation for the following frequency distribution.

Age (years)	No. of Persons	Age (years)	No. of persons
1 - 5	7	26 - 30	18
6 - 10	10	31 - 35	10
11 - 15	16	36 - 40	5
16 - 20	32	36 - 40	1
21 - 25	24		

(Ans : Median = 19.95, M.D. = 7, 103).

2. Calculate the standard deviation from the following data:

Marks in Cost Accounting	No. of Students
0 - 10	5
10 - 20	7
20 - 30	14
40 - 50	9
50 - 60	6
60 - 70	2

(Ans : σ = 15.57)

3. Calculate coefficient of variation and Karl Pearson's Coefficient of Skewness.

Income (Rs.) Below	500	600	700	800	900	1000
No. of Employees	8	20	40	70	87	100

(Ans : C.V. = 19.64, Co. eff. of SK = .0.13)

4. Calculate coefficient of skewness based on quartiles from the following data.

Size	Frequency
10 - 20	6
20 - 30	10
30 - 40	18
40 - 50	30
50 - 60	12
60 - 70	10
70 - 80	6
80 - 90	2

(Ans : 0.04)

5. Calculate Bowley's coefficient of skewness of the following data:

Weight (lbs)	No. of Persons
Under 100	1
100 - 109	14
110 - 119	66
120 - 129	122
130 - 139	145
140 - 149	121
150 - 159	65
160 - 169	31
170 - 179	12
180 - 189	5
190 - 199	2
200 and over	2

(Ans: + 0.023)

Rahul Publications

Short Question and Answers

1. 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."

2. Purpose of Dispersion.

Ans :

Measures of variation are needed for four basic purposes :

- (i) To determine the reliability of an average.
- (ii) To serve as a basis for the control of the variability.

- (iii) To compare two or more series with regard to their variability.
- (iv) To facilitate the use of other statistical measures.

3. Characteristics of a good Measures of Dispersion.

Ans :

- (i) It should be easy to calculate.
- (ii) It should be easy to understand.
- (iii) It should be rigidly defined.
- (iv) It should be based on all the observations.
- (v) It should be capable of further mathematical treatment.
- (vi) It should not be unduly affected by extreme observations.
- (vii) It should not be much affected by the fluctuations of sampling.

4. 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,

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

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 :

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

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.

5. 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.

6. What is Quartile Deviations ?

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.

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

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

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

7. State the merits and demerits of mean deviation.

Ans :

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.

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. Define Coefficient of Variation.**Ans :**

Coefficient of Variation (CV) was proposed by Karl Pearson. It is used to compare the variability of two (or) more distributions. A distribution with greater C.V. is considered as more variable or less consistent, less unit form, less stable or less homogeneous distribution and the distribution with less C.V. is considered as less variable or more consistent, more uniform, more stable or more homogeneous distribution.

Coefficient of variation

$$= \frac{\text{Standard deviation}}{\text{Arithmetic mean}} \times 100.$$

11. Skewness.**Ans :****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** "Measures 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."

Choose the Correct Answers

1. The scatter in a series of values about the average is called: [b]
(a) Central tendency (b) Dispersion
(c) Skewness (d) Symmetry
2. The measurements of spread or scatter of the individual values around the central point is called: [a]
(a) Measures of dispersion (b) Measures of central tendency
(c) Measures of skewness (d) Measures of kurtosis
3. The measures used to calculate the variation present among the observations in the unit of the variable is called: [c]
(a) Relative measures of dispersion (b) Coefficient of skewness
(c) Absolute measures of dispersion (d) Coefficient of variation
4. The measures used to calculate the variation present among the observations relative to their average is called: [d]
(a) Coefficient of kurtosis (b) Absolute measures of dispersion
(c) Quartile deviation (d) Relative measures of dispersion
5. The degree to which numerical data tend to spread about an average value called: [c]
(a) Constant (b) Flatness
(c) Variation (d) Skewness
6. The measures of dispersion can never be: [c]
(a) Positive (b) Zero
(c) Negative (d) Equal to 2
7. If all the scores on examination cluster around the mean, the dispersion is said to be: [a]
(a) Small (b) Large
(c) Normal (d) Symmetrical
8. If there are many extreme scores on all examination, the dispersion is: [a]
(a) Large (b) Small
(c) Normal (d) Symmetric
9. Given below the four sets of observations. Which set has the minimum variation? [d]
(a) 46, 48, 50, 52, 54 (b) 30, 40, 50, 60, 70
(c) 40, 50, 60, 70, 80 (d) 48, 49, 50, 51, 52
10. Which of the following is an absolute measure of dispersion? [c]
(a) Coefficient of variation (b) Coefficient of dispersion
(c) Standard deviation (d) Coefficient of skewness
11. The measure of dispersion which uses only two observations is called: [c]
(a) Mean (b) Median
(c) Range (d) Coefficient of variation

12. The measure of dispersion which uses only two observations is called: [a]
(a) Range (b) Quartile deviation
(c) Mean deviation (d) Standard deviation
13. The range of the values -5, -8, -10, 0, 6, 10 is: [d]
(a) 0 (b) 10
(c) -10 (d) 20
14. Half of the difference between upper and lower quartiles is called: [b]
(a) Interquartile range (b) Quartile deviation
(c) Mean deviation (d) Standard deviation
15. Which measure of dispersion can be computed in case of open-end classes? [c]
(a) Standard deviation (b) Range
(c) Quartile deviation (d) Coefficient of variation
16. The average of squared deviations from mean is called: [b]
(a) Mean deviation (b) Variance
(c) Standard deviation (d) Coefficient of variation
17. Which of the following measures of dispersion is expressed in the same units as the units of observation? [b]
(a) Variance (b) Standard deviation
(c) Coefficient of variation (d) Coefficient of standard deviation
18. Which of the following is a unit free quantity: [c]
(a) Range (b) Standard deviation
(c) Coefficient of variation (d) Arithmetic mean
19. The value of standard deviation changes by a change of: [b]
(a) Origin (b) Scale
(c) Algebraic signs (d) None
20. The standard deviation one distribution divided by the mean of the distribution and expressing in percentage is called: [d]
(a) Coefficient of Standard deviation (b) Coefficient of skewness
(c) Coefficient of quartile deviation (d) Coefficient of variation
21. Departure from symmetry is called: [c]
(a) Second moment (b) Kurtosis
(c) Skewness (d) Variation
22. In a symmetrical distribution, the coefficient of skewness will be: [a]
(a) 0 (b) Q1
(c) Q3 (d) 1
23. The lack of uniformity or symmetry is called: [a]
(a) Skewness (b) Dispersion
(c) Kurtosis (d) Standard deviation
24. The lower and upper quartiles of a distribution are 80 and 120 respectively, while median is 100. The shape of the distribution is: [c]
(a) Positively skewed (b) Negatively skewed
(c) Symmetrical (d) Normal

Fill in the blanks

1. The concept of _____ is related to the extent of scatter or variability in observations.
2. The degree to which numerical data tend to spread about an average value is called _____.
3. _____ of a series is the arithmetic average of the deviations of various items from a measure of central tendency.
4. The standard deviation concept was introduced by _____ in 1823.
5. Standard deviation is denoted by the small Greek letter _____.
6. _____ was proposed by Karl Pearson.
7. Coefficient of variation _____.
8. The word skewness refers to _____.
9. _____ refers to the asymmetry or lack of symmetry in the shape of a frequency distribution.
10. _____ can be used to test the reliability of an average.

ANSWERS

1. Dispersion
2. Variation
3. Mean Deviation
4. Karl Pearson
5. σ
6. Coefficient of Variation (CV)
7. $\frac{\text{Standard deviation}}{\text{Arithmetic mean}} \times 100$
8. Lack of Symmetry
9. Skewness
10. A measure of dispersion

UNIT V

CORRELATION:

Meaning - Types - Correlation and Causation - Methods: Scatter Diagram - Karl Person's Coefficient of Correlation - Probable Error and Interpretation of Coefficient of Correlation - Rank Correlation - Concurrent Deviation Method

5.1 CORRELATION

5.1.1 Meaning

Q1. Define the term correlation.

(OR)

What do you mean by correlation?

Ans :

(Dec.-18)

Meaning

Correlation is the study of the linear relationship between two variables. When there is a relationship of 'quantitative measure between two set of variables, the appropriate statistical tool for measuring the relationship and expressing each in a precise way is known as correlation.

For example, there is a relationship between the heights and weights of persons, demand and prices of commodities etc.

Correlation analysis is the statistical tool we can use to describe the degree to which one variable is linearly related to another.

Definitions

- (i) **According to L.R. Connor** "If two or more quantities vary in sympathy so that movements in one tend to be accompanied by corresponding movements in other(s) then they are said to be correlated."
- (ii) **According to A.M. Tuttle** "Correlation is an analysis of covariation between two or more variables".
- (iii) **According to Croxton and Cowden** "When the relationship is of a quantitative nature, the appropriate statistical tool for discovering and measuring the relationship and expressing it in a brief formula is known as correlation."
- (iv) **According to Ya Lun Chou** "Correlation analysis attempts to determine the 'degree of relationship' between variable".

Q2. Explain the Significance of Measuring Correlation.

(OR)

How do you say that the correlation between the two variables is significant (or) not.

Ans :

(Dec.-18)

1. Correlation is very useful to economists to study the relationship between variables, like price and quantity demanded. To businessmen, it helps to estimate costs, sales, price and other related variables.
2. In economic theory we come across several types of variables which show some kind of relationship. For example, there exists a relationship between price, supply and quantity demanded; convenience, amenities, and service standards are related to customer retention; yield a crop related to quantity of fertilizer applied, type of soil, quality of seeds, rainfall and so on. Correlation analysis helps in measuring the degree of association and direction of such relationship.
3. The relation between variables can be verified and tested for significance, with the help of the correlation analysis. The effect of correlation is to reduce the range of uncertainty of our prediction.
4. The coefficient of correlation is a relative measure and we can compare the relationship between variables, which are expressed in different units.
5. Correlations are useful in the areas of health care such as determining the validity and reliability of clinical measures or in expressing how health problems are related to certain biological or environmental factors. For example, correlation coefficient can be used to determine the degree of inter-observer reliability for two doctors who are assessing a patient's disease.
6. Sampling error can also be calculated.
7. Correlation is the basis for the concept of regression and ratio of variation.
8. The decision making is heavily facilitated by reducing the range of uncertainty and hence empowering the predictions.

Q3. Explain the scope of correlation analysis.

Ans :

- (i) One of the variable may be affecting the other:** A correlation coefficient calculated from the data on quantity demanded and corresponding price of tea would only reveal that the degree of association between them is very high. It will not give us any idea about whether price is affecting demand of tea or vice-versa. In order to know this, we need to have some additional information apart from the study of correlation.

For example if, on the basis of some additional information, we say that the price of tea affects its demand, then price will be the cause and quantity will be the effect. The causal variable is also termed as independent variable while the other variable is termed as dependent variable.

- (ii) The two variables may act upon each other:** Cause and effect relation exists in this case also but it may be very difficult to find out which of the two variables is independent.

For example, if we have data on price of wheat and its cost of production, the correlation between them may be very high because higher price of wheat may attract farmers to produce more wheat and more production of wheat may mean higher cost of production, assuming that it is an increasing cost industry. Further, the higher cost of production may in turn raise the price of wheat. For the purpose of determining a relationship between the two variables in such situations, we can take any one of them as independent variable.

- (iii) The two variables may be acted upon by the outside influences:** In this case we might get a high value of correlation between the two variables, however, apparently no cause and effect type relation seems to exist between them.

For example, the demands of the two commodities, say X and Y, may be positively correlated because the incomes of the consumers are rising. Coefficient of correlation obtained in such a situation is called a spurious or nonsense correlation.

- (iv) **A high value of the correlation coefficient may be obtained due to sheer coincidence (or pure chance):** This is another situation of spurious correlation. Given the data on any two variables, one may obtain a high value of correlation coefficient when in fact they do not have any relationship. For example, a high value of correlation coefficient may be obtained between the size of shoe and the income of persons of a locality.

Q4. State the applications of correlation.

Ans :

- (a) Relationship between advertising expenditure and sales revenue of a company for a period of time.
 - (b) Relationship between time or the day and the temperatures recorded for a particular city.
 - (c) Relationship between ages of the husbands and wives for a group of couples under study.
 - (d) Relationship between heights and weights of a group of students.
 - (e) Relationship between age and blood pressure of men.
 - (f) Relationship between imports of raw material and exports of finished goods of a firm in a particular year.
 - (g) Relationship between amount invested (capital) and profits earned by different firms.
 - (h) Relationship between price and demand for a particular commodity.
-

5.1.2 Types of Correlation

Q5. Explain various Types of Correlation.

(or)

What are the different Types of correlations.

Ans :

(Dec.-17)

Broadly speaking, there are four types of correlation, namely,

- A) Positive correlation,
- B) Negative correlation,
- C) Linear correlation and
- D) Non-Linear Correlation.

A) Positive correlation

If the values of two variables deviate in the same direction i.e., if increase in the values of one variable results, on an average, in a corresponding increase in the values of the other variable or if a decrease in the values of one variable results, on an average, in a corresponding decrease in the values of the other variable, the corresponding correlation is said to be positive or direct.

Examples

- i) Sales revenue of a product and expenditure on Advertising.
- ii) Amount of rain fall and yield of a crop (up to a point)
- iii) Price of a commodity and quantity of supply of a commodity
- iv) Height of the Parent and the height of the Child.
- v) Number of patients admitted into a Hospital and Revenue of the Hospital.
- vi) Number of workers and output of a factory.

- i) **Perfect Positive Correlation** : If the variables X and Y are perfectly positively related to each other then, we get a graph as shown in fig. below.

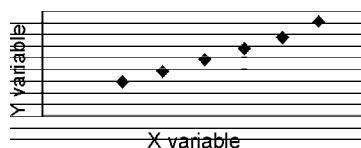


Fig.: Perfect Positive Correlation ($r = +1$)

- ii) **Very High Positive Correlation** : If the variables X and Y are related to each other with a very high degree of positive relationship then we can notice a graph as in figure below.

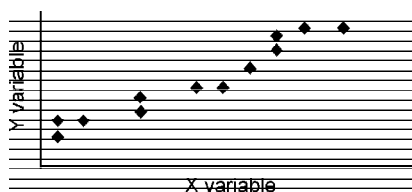


Fig.: Very High Positive Correlation ($r = \text{nearly } +1$)

- iii) **Very Low Positive Correlation** : If the variables X and Y are related to each other with a very low degree of positive relationship then we can notice a graph as in fig. below.

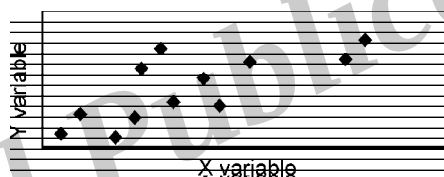


Fig.: Very Low Positive Correlation ($r = \text{near to } +0$)

B) Negative Correlation

Correlation is said to be negative or inverse if the variables deviate in the opposite direction i.e., if the increase (decrease) in the values of one variable results, on the average, in a corresponding decrease (increase) in the values of the other variable.

Examples

1. Price and demand of a commodity.
2. Sales of Woolen garments and the day temperature.

- i) **Perfect Negative Correlation** : If the variables X and Y are perfectly negatively related to each other then, we get a graph as shown in fig. below.

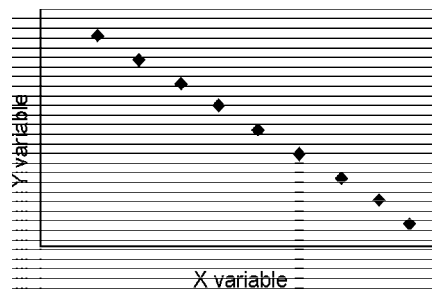


Fig.: Perfect Negative Correlation ($r = -1$)

- ii) **Very High Negative Correlation** : If the variables X and Y are related to each other with a very high degree of negative relationship then we can notice a graph as in fig. below.

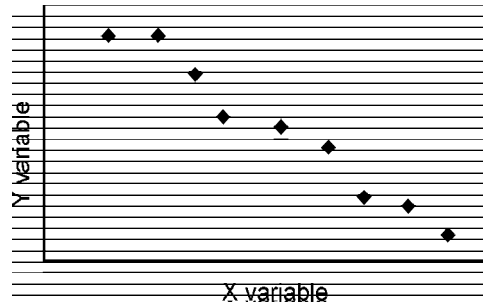


Fig.: Very High Negative Correlation ($r = \text{near to } -1$)

- iii) **Very low Negative Correlation** : If the variables X and Y are related to each other with a very low degree of negative relationship then we can notice a graph as in fig. below.

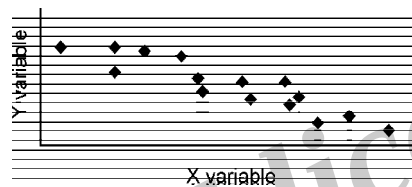


Fig.: Very Low Negative Correlation
($r = \text{near to } 0 \text{ but negative}$)

- iv) **No Correlation** : If the scatter diagram show the points which are highly spread over and show no trend or patterns we can say that there is no correlation between the variables.

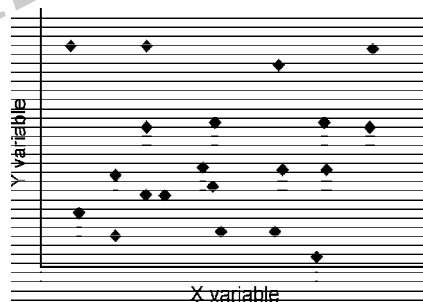


Fig.: No Correlation ($r = 0$)

C) Linear Correlation

Two variables are said to be linearly related if corresponding to a unit change in one variable there is a constant change in the other variable over the entire range of the values.

If two variables are related linearly, then we can express the relationship as

$$Y = a + bX$$

where ' a ' is called as the "intercept" (If $X = 0$, then $Y = a$) and ' b ' is called as the "rate of change" or slope.

If we plot the values of X and the corresponding values of Y on a graph, then the graph would be a straight line as shown in fig. below.

Example

X	1	2	3	4	5
Y	8	11	14	17	20

For a unit change in the value of x, a constant 3 units changes in the value of y can be noticed. The above can be expressed as : $Y = 5 + 3x$.

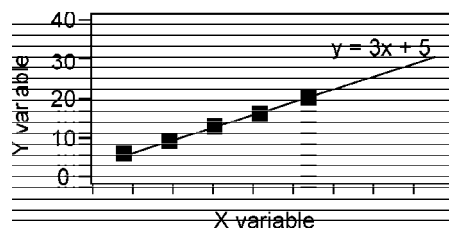


Fig.: Linear Correlation

D) Non Linear (Curvilinear) Correlation

If corresponding to a unit change in one variable, the other variable does not change in a constant rate, but change at varying rates, then the relationship between two variables is said to be nonlinear or curvilinear as shown in fig. below. In this case, if the data are plotted on the graph, we do not get a straight line curve.

Mathematically, the correlation is non-linear if the slope of the plotted curve is not constant. Data relating to Economics, Social Science and Business Management do exhibit often non-linear relationship. We confine ourselves to linear correlation only.

Example

X	-3	-2	-1	0	1	2	3
Y	9	4	1	0	1	4	9

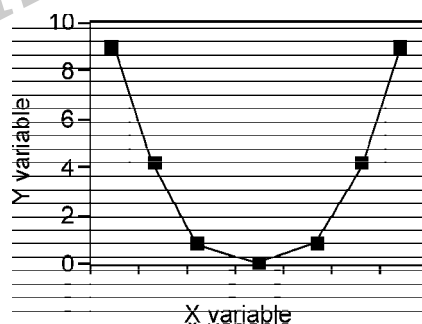


Fig.: Non Linear Correlation

5.2 CORRELATION AND CAUSATION

Q6. Explain briefly about Correlation and Causation relationship. State the various reasons behind such relationships.

Ans :

Correlation analysis describes the degree and direction of the relationship between the two variables under study. But, it does not reflect upon the cause and effect relationship between the variables.

If the variables have cause and effect relationship, then they have a high degree of correlation. But the converse is not true i.e., high degree of correlation between two variables need not imply a cause and effect relationship between them.

Reasons

The reasons for such high degree of correlation (where cause and effect relationship does not exist) are as follows,

(a) Pure Chance

The correlation may be due to pure chance especially in a small sample. Such a correlation may arise either because of pure random sampling variation or because of the bias of the investigator in selecting the sample. For, if correlation exists between income and height, weight and IQ level, education level and dress size etc., then such correlations could be by pure chance only.

(b) Mutual Dependence

Both the correlation variables may be mutually influencing each other, so that neither can be designated as the cause and the other the effect. This is particularly found in economic variables like demand and price.

(c) Same Factors Influencing

Both the correlated variables may be influenced by one or more other variables i.e., a high degree of correlation between the variables may be due to some causes affecting each variable with the same effect. Example, salaries of nurses and advertising expenses both can be influenced by economic condition of the country.

5.3 METHODS OF CORRELATION

Q7. Explain various methods of Correlation.

Ans :

The various method of studying linear correlation can be shown diagrammatically as follows:

- (i) Scatter Diagram
- (ii) Two way frequency table
- (iii) Spearman's Rank Correlation Method
- (iv) Method of Concurrent Deviations
- (v) Method of Least Squares.

5.3.1 Scatter Diagram

Q8. What is Scatter Diagram? Explain the procedure of Scatter Diagram.

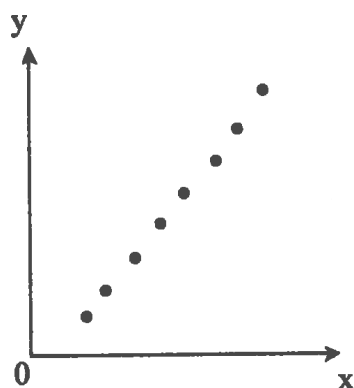
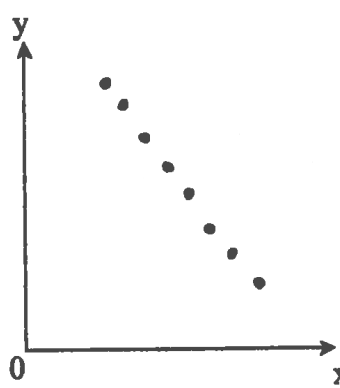
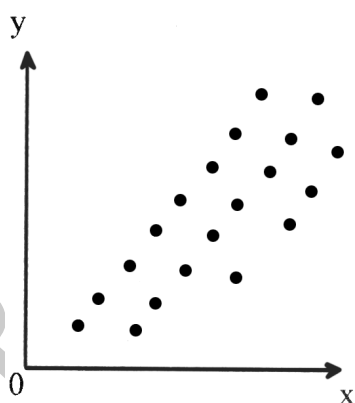
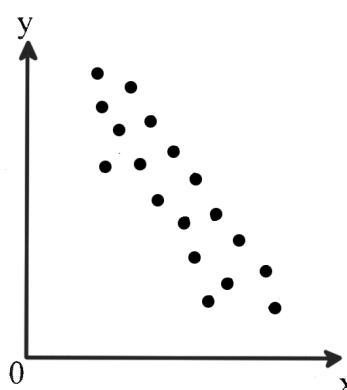
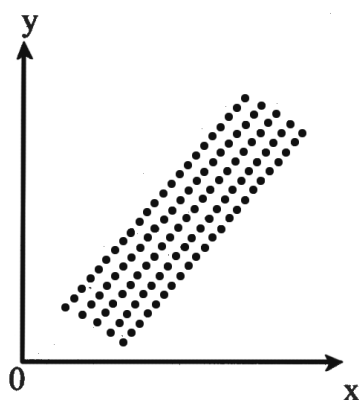
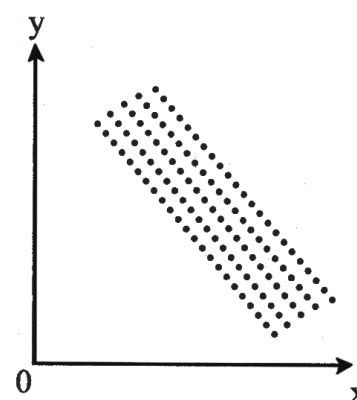
Ans :

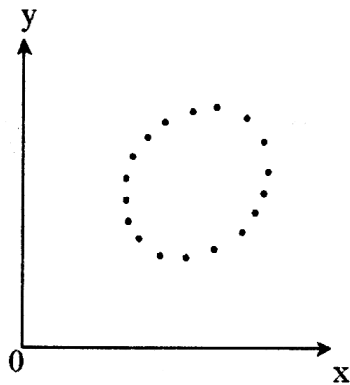
Scatter diagram method is the simplest way of diagrammatic representation of a bivariate distribution and helps in ascertaining the correlation between the two variables under study i.e., it portrays the relationship between these two variables graphically.

Procedure of Scatter Diagram

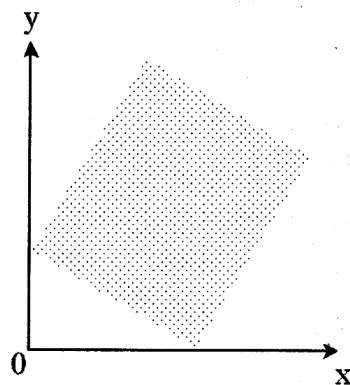
Given pair of values (x_1, y_1) (x_2, y_2) ... (x_n, y_n) of two variables X and Y. Take the independent variables on the X axis and the dependent variable on the Y-axis. The V points denoted by the pair of values are plotted on the graph. The diagram of dots thus obtained is the scatter diagram. Regarding the correlation between the two variables, the scatter diagram can be interpreted as follows,

- (i) If the points reveal any upward or downward trend, the variables are said to be correlated, otherwise uncorrelated.
- (ii) If the points are very close to each other, a good amount of correlation exists, else poor correlation exists.
- (iii) Upward trend indicates positive correlation and downward trend indicates negative correlation.

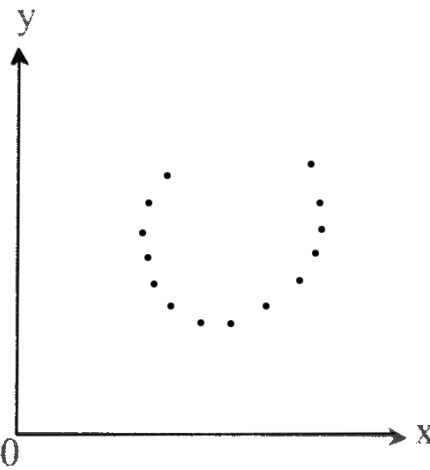
Different Forms of Scatter Diagram**Perfect Positive Correlation****Perfect Negative Correlation****Low Degree Positive Correlation****Low Degree Negative Correlation****High Degree Positive Correlation****High Degree negative Correlation**



No Correlation



No Correlation



No Correlation

Q9. State the Merits and Demerits of Scatter Diagram.

Ans :

Merits

- Scatter diagram is a simple and attractive method of finding out the nature of correlation between two variables.
- It is a non-mathematical method of studying correlation. It is easy to understand.
- We can get a rough idea at a glance whether it is a positive or negative correlation.
- It is not influenced by extreme items.
- It is a first step in finding out the relationship between two variables.

Demerits

- The major limitation of the method is that it only gives a visual picture of the relationship of two variables. It only tells us whether there is correlation between the variables, and if so, then in which direction, positive (or) negative.
- It does not give an idea about the precise degree of relationship as it is not amenable to mathematical treatment.

5.3.2 Karl Person's Coefficient of Correlation

Q10. What is Karl Pearson's Coefficient of Correlation? Explain properties of Coefficient of Correlation.

Ans :

(Imp.)

Karl Parson's Coefficient of Correlation is arrived at with the help of a statistical formula that takes into account the mean and standard deviation of the two variables, the number of such observations and the covariance between them. Since Karl Pearson's coefficient of correlation is a number, it can describe the strength of the correlation in greater detail and more objectively. A value of -1 signifies "absolute" negative correlation, a value between -1 and -0.5 signifies strong negative correlation, a value between -0.5 and -0.25 signifies moderate negative correlation and a value between -0.25 and 0 signifies weak negative correlation. Similarly, a value of $+1$ signifies "absolute" positive correlation, a value between $+1$ and $+0.5$ signifies strong positive correlation, a value between $+0.5$ and $+0.25$ signifies moderate positive correlation and a value between $+0.25$ and 0 signifies weak positive correlation

Properties of Karl Pearson's Coefficient of Correlation

1. It is based on Arithmetic Mean and Standard Deviation.
2. It lies between -1
3. It measures both direction as well as degree of change. If r is less than 0 , there is negative correlation, which means the direction of change of the two variables will be opposite. If r is more than 0 , there is positive correlation, which means the direction of change of the two, variables will be same. Higher the value of r , greater is the degree of correlation. Hence, Karl Pearson's coefficient of correlation is said to be the ideal measure of correlation.
4. It is independent of change in scale. In other words, if a constant amount is added/ subtracted from all values of a variable, the value of r does not change.
5. It is independent of change in origin. Thus, if a constant amount is multiplied with or divides all values of a variable, r does not change.
6. It is independent of direction. In other words, Correlation of X and Y is same as Correlation of Y and X .
7. It is a pure number without any units. In other words, it is independent of the unit of measurement of the 2 variables.
8. It takes into account all items of the variable(s).
9. It does not prove causation but is simply a measure of co-variation.
10. Correlation coefficient of two variables X and Y is the Geometric Mean of two regression coefficients, regression coefficient of X on Y and regression coefficient of Y on X . Symbolically,
$$r = \text{Square root of } (b_{xy} * b_{yx})$$
11. Correlation coefficient can be calculated between two unrelated variables and such a number can be misleading. Such correlation is called accidental correlation, spurious correlation or non sense correlation.

Q11. Explain Merits and Demerits of Coefficient of Correlation.*Ans :***Merits**

1. It takes into account all items of the variable(s).
2. It is a numerical measure and hence more objective.
3. It measures both direction as well as degree of change.
4. It facilitates comparisons between two series.
5. It is capable of further Algebraic treatment
6. It is more practical and hence popular and is more commonly used.

Demerits

1. It is not easy to calculate as complex formulae are involved.
2. It is more time consuming compared to methods such as rank correlation
3. It assumes a linear relationship between the two variables which may not be correct
4. It is impacted by extreme values as it is based on mean and standard deviation.
5. It is not easy to interpret.

Q12. Explain the various methods of Coefficient of Correlation.*Ans :***i) Direct Method when deviations are taken from actual mean**

$$r = \frac{\sum xy}{\sqrt{\sum x^2 \times \sum y^2}}$$

Where

$$x = X - \bar{X} \text{ and } y = Y - \bar{Y}$$

Steps :

1. Find the means of the two series ($\bar{X}$, $\bar{Y}$)
2. Take the deviations of X series from the mean of X and denote these deviations as x.
3. Square these deviations and obtain the total. Denote it as $\sum x^2$.
4. Take the deviations of Y series from the Mean of Y and denote these deviations as y.
5. Square these deviations, obtain the total and denote it as $\sum y^2$.
6. Multiply the deviations of X and Y series, obtain the total and denote it $\sum xy$.
7. Substitute the above values in the formula.

ii) Short-Cut Method**When deviations are taken from assumed mean.**

$$r = \frac{N\sum xy - \sum x \sum y}{\sqrt{N\sum x^2 - (\sum x)^2} \sqrt{N\sum y^2 - (\sum y)^2}}$$

Where, $x = X - A_1$

$$y = Y - A_2$$

A_1 = Assumed mean for X series

A_2 = Assumed mean for Y series

The values of coefficient of correlation as obtained by above formulae will always lie between ± 1 . When there is perfect positive correlation its value is +1 and when there is perfect negative correlation, its value is -1. When $r = 0$ means that there is no relationship between the two variables. We normally get values which lie between +1 and -1.

5.3.3 Probable Error and Interpretation of Coefficient of Correlation

Q13. Explain briefly about probable error.

Ans :

(May-18, Imp.)

The probable error of the coefficient of correlation helps in interpretation. The probable error of the coefficient of correlation is obtained as follows:

$$\text{P.E. of } r = 0.6745 \frac{1-r^2}{\sqrt{N}}$$

Where

r = Coefficient of correlation

N = Number of pairs of observations.

If the probable error is added to and subtracted from the coefficient of correlation, it would give two such limits within which we can reasonably expect the value of coefficient of correlation to vary.

Symbolically $P(\rho) = r \pm \text{P.E.}$

Where 'P' denotes the correlation in the population. Suppose, the Coefficient of correlation for a pair of 10 observations is 0.8 and its P.E. is 0.05. the limits of the correlation in the population would be $r \pm \text{P.E.}$ i.e. 0.8 ± 0.05 or 0.75 – 0.85.

If the value of r is less than the probable error then r is not at all significant, i.e. there is no evidence of correlation. If the value of r is more than six times the probable error, it is significant. Hence it can be said that r is significant, when

$$r > 6 \text{ P.E. or } \frac{r}{\text{P.E.}} > 6.$$

Q14. How do you interpret the coefficient of correlation?

Ans :

The investigator must take utmost care while interpreting the value of 'r' and interpretation depends very much on experience.

Rules of Interpretation

1. When $r = 0$, it implies that the variables are uncorrelated.
2. When $r = +1$, it implies that there is a perfect positive relationship between the variables.
3. When $r = -1$, it implies that there is a perfect negative relationship between the variables.
4. The closer the value of 'r' to +1 or -1, the closer the relationship between the variables.
5. The closeness of the relationship is not proportional to r .

**PROBLEMS ON KARL PERSON'S COEFFICIENT OF CORRELATION AND
PROBABLE ERROR**

1. Find the coefficient of correlation from the following data :

X	46	54	56	56	58	60	62
Y	36	40	44	54	42	58	54

Sol :

(May-18)

Calculation of Correlation

X	(X - $\bar{X}$) (X)	Y	(Y - $\bar{Y}$) (Y)	x ²	y ²	xy
46	- 10	36	-11	100	121	110
54	- 2	40	-7	4	49	14
56	0	44	- 3	0	9	0
56	0	54	7	0	49	0
58	+ 2	42	- 5	4	25	-10
60	+ 4	58	11	16	121	44
62	+ 6	54	7	36	49	42
				160	423	200

$$r = \frac{\sum xy}{\sqrt{\sum x^2 \times \sum y^2}}$$

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

$$\bar{X} = \frac{392}{7} = 56,$$

Where,

$$x^2 = 160, y^2 = 1423, xy = 200$$

$$= \frac{200}{\sqrt{160 \times 423}}$$

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

$$\bar{Y} = \frac{328}{7} = 46.85 = 47$$

$$= \frac{200}{260.15} = 0.768$$

2. Calculate Karl Pearson's co-efficient of correlation from the following data

x	3	4	6	7	10
f	9	11	14	15	16

Sol :

(Dec.-17)

Calculation of Karl Pearson's co-efficient of correlation without taking deviation from actual mean or assumed mean.

X	X ²	Y	Y ²	XY
3	9	9	81	27
4	16	11	121	44
6	36	14	196	84
7	49	15	225	105
10	100	16	256	160
$\Sigma X = 30$	$\Sigma X^2 = 210$	$\Sigma y = 65$	$\Sigma Y^2 = 879$	$\Sigma XY = 420$

$$r = \frac{N\Sigma XY - (\Sigma X)(\Sigma Y)}{\sqrt{N\Sigma X^2 - (\Sigma X)^2} \sqrt{N\Sigma Y^2 - (\Sigma Y)^2}}$$

Where, N = 5

$$= \frac{(5 \times 420) - (30 \times 65)}{\sqrt{5 \times 210 - (30)^2} \sqrt{5 \times 879 - (65)^2}}$$

$$= \frac{2100 - 1950}{\sqrt{1050 - 900} \sqrt{4395 - 4225}}$$

$$= \frac{150}{\sqrt{150} \sqrt{170}}$$

$$= \frac{150}{12.25 \times 13.04}$$

$$= \frac{150}{159.74}$$

$$r = 0.94$$

3. Find Karl Pearson's Coefficient of Correlation for the following data:

X	65	66	67	67	68	69	70	72
Y	67	68	65	68	72	72	69	71

Sol :

(Dec.-18)

Calculations of Karl Pearson's coefficient of correlation

X	$X - \bar{X} = x$	x^2	Y	$Y - \bar{Y} = y$	y^2	xy
65	-3	9	67	-2	4	6
66	-2	4	68	-1	1	2
67	-1	1	65	-4	16	4
67	-1	1	68	-1	1	1
68	0	0	72	3	9	0
69	1	1	72	3	9	3
70	2	4	69	0	0	0
72	4	16	71	2	4	8
		$\Sigma x^2 = 36$			$\Sigma y^2 = 44$	$\Sigma xy = 24$

$$\bar{x} = \frac{\Sigma x}{N} = \frac{544}{8} = 68$$

$$\bar{y} = \frac{\Sigma y}{N} = \frac{552}{8} = 69$$

$$r = \frac{\Sigma xy}{\sqrt{\Sigma x^2 \times \Sigma y^2}} = \frac{24}{\sqrt{36 \times 44}} = \frac{24}{\sqrt{1584}} = \frac{24}{39.799}$$

$$r = 0.6031.$$

4. Calculate Karl Pearson's Coefficient of Correlation for the following data.

X	7	6	5	4	3	2	1
Y	18	16	14	12	10	6	8

Sol :

(June-19)

Calculations of Karl Pearson's coefficient of correlation

X	Y	$X - \bar{X}$	$Y - \bar{Y} (y)$	xy	x^2	y^2
7	18	3	6	18	9	36
6	16	2	4	8	4	16
5	14	1	2	2	1	4
4	12	0	0	0	0	0
3	10	-1	-2	+ 2	1	4
2	6	-2	-6	12	4	36
1	8	-3	-4	12	9	16
				54	28	112

$$\bar{X} = \frac{28}{7} = 4$$

$$\bar{Y} = \frac{84}{7} = 12$$

$$r = \frac{\Sigma xy}{\sqrt{\Sigma x^2 \times \Sigma y^2}}$$

$$r = \frac{54}{\sqrt{28 \times 112}}$$

$$= \frac{54}{\sqrt{3136}} = \frac{54}{56} = 0.96$$

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

X	6	8	12	15	18	20	24	28	31
Y	10	12	15	15	18	25	22	26	28

Sol :

Calculation of co-efficient of correlation, assumed mean.

X	X ²	Y	Y ²	XY
6	36	10	100	60
8	64	12	144	96
12	144	15	225	180
15	225	15	225	225
18	324	18	324	324
20	400	25	625	500
24	576	22	484	528
28	784	26	676	728
31	961	28	784	868
$\Sigma X = 162$	$\Sigma X^2 = 3514$	$\Sigma y = 171$	$\Sigma Y^2 = 3587$	$\Sigma XY = 3509$

Co-efficient of correlation,

$$r = \frac{N\Sigma XY - (\Sigma X)(\Sigma Y)}{\sqrt{N\Sigma X^2 - (\Sigma X)^2} \sqrt{N\Sigma Y^2 - (\Sigma Y)^2}}$$

Where, N = 9

$$r = \frac{(9 \times 3509) - (162)(171)}{\sqrt{9 \times 3514 - (162)^2} \sqrt{9 \times 3587 - (171)^2}}$$

$$\begin{aligned}
 &= \frac{31581 - 27702}{\sqrt{31626 - 26244} \sqrt{32283 - 29241}} \\
 &= \frac{3879}{\sqrt{5382} \sqrt{3042}} \\
 &= \frac{3879}{(73.36)(55.15)} \\
 &= \frac{3879}{4046} \\
 r &= 0.958.
 \end{aligned}$$

6. If the Coefficient of Correlation between two variables X and Y is 0.86, the Covariance is 36. And the Standard Deviation of X is 4, find the Standard Deviation of Y.

Sol : (Dec.-18)

Given

$$\sigma_x = 4$$

$$\begin{aligned}
 \text{Covariance} &= \frac{\sum xy}{N} \\
 &= 36
 \end{aligned}$$

$$\begin{aligned}
 \text{Here } r &= \frac{\sum xy}{N \sigma_x \sigma_y} \\
 &= \frac{\sum xy}{N} \times \frac{1}{\sigma_x \cdot \sigma_y}
 \end{aligned}$$

$$r = 36 \times \frac{1}{4 \sigma_y}$$

$$\text{Here } r = 0.86$$

$$0.86 = \frac{36}{4 \cdot \sigma_y}$$

$$\sigma_y = \frac{36}{4 \times 0.86}$$

$$\sigma_y = \frac{9}{0.86}$$

$$= 10.46.$$

7. If $r = 0.917$, $PE = 0.034$, calculate the value of N .

Sol :

Given that,

$$PE_{(r)} = 0.034$$

$$r = 0.917$$

$$N = ?$$

$$PE_{(r)} = 0.6745 \frac{1-r^2}{\sqrt{n}}$$

$$0.034 = 0.6745 \times \frac{1-(0.917)^2}{\sqrt{n}}$$

$$0.034 = 0.6745 \times \frac{1-(0.84)}{\sqrt{n}}$$

$$0.034 = 0.6745 \times \frac{0.16}{\sqrt{n}}$$

$$\frac{0.034}{0.6745} = \frac{0.16}{\sqrt{n}}$$

$$0.0504 = \frac{0.16}{\sqrt{n}}$$

$$\sqrt{n} = \frac{0.16}{0.0504} = 3.174$$

$$n = \sqrt{3.174} = 1.78$$

8. If $b_{xy} = 0.30$ and $b_{yx} = 3.25$, find the co-efficient of correlation.

Sol :

Co-efficient of correlation,

$$r = \sqrt{b_{xy} \times b_{yx}}$$

$$\therefore r = \sqrt{0.30 \times 3.25} \\ = 0.987.$$

9. If $r = 0.6$ and $N = 64$ find out the Probable Error of the Co-efficient of Correlation.

Sol :

$$\text{Probable Error (P, } E_r) = 0.6745 \times \frac{1-r^2}{\sqrt{N}}$$

$$= 0.6745 \times \frac{1-(0.6)^2}{\sqrt{64}}$$

$$= 0.6745 \times \frac{1-0.36}{8}$$

$$= 0.6745 \times \frac{0.64}{8}$$

$$= 0.6745 \times 0.08$$

$$P.E_r = 0.05396$$

5.4 RANK CORRELATION

Q15. What is Rank correlation ? State the properties rank correlation.

Ans :

The Karl Pearson's method is based on the assumption that the population being studied is normally distributed. When it is known that the population is not normal or when the shape of the distribution is not known, there is need for a measure of correlation that involves no assumption about the parameter of the population.

It is possible to avoid making any assumptions about the populations being studied by ranking the observations according to size and basing the calculations on the ranks rather than upon the original observations. It does not matter which way the items are ranked, item number one may be the largest or it may be the smallest. Using ranks rather than actual observations gives the coefficient of rank correlation.

This method of finding out covariability or the lack of it between two variables was developed by the British Psychologist Charles Edward Spearman in 1904.

Properties

1. It is based on subjective ranking of variables.
2. It lies between -1
3. It measures both direction as well as degree of change. If r is less than 0, there is negative correlation, which means the direction of

change of the two variables will be opposite. If r is more than 0, there is positive correlation, which means the direction of change of the two variables will be same. Higher the value of r , greater is the degree of correlation.

4. It is independent of direction. In other words, Correlation of X and Y is same as Correlation of Y and X.
5. It is a pure number without any units. In other words, it is independent of the unit of measurement of the 2 variables.
6. It is not impacted by extreme values as only ranking matters.

Q16. State the Merits and Demerits of rank correlation.

Ans :

Merits

1. It is easy to understand and calculate
2. It is not impacted by extreme values.
3. It is a numerical measure and provides objectivity to subjective ranking.
4. It is the only method of finding correlation with respect to qualitative factors such as honesty, beauty, etc.
5. It measures both direction as well as degree of change.
6. It facilitates comparisons between two series.
7. It can be applied to irregular data also.
8. It is ideal when the number of observations is very small.

Demerits

1. It cannot be applied to grouped data
2. It lacks the precision of Karl Pearson's Coefficient of Correlation.
3. All the information concerning the variable is not used.
4. The Computation becomes complicated as the number of observations increase.

Q17. Explain how correlation is calculated through spearman's rank correlation method.

Ans :

Computation

Spearman's correlation is computed by using the following formula:

$$r_s = 1 - \frac{6\sum D^2}{N^3 - N}$$

Where

r_s = rank coefficient of correlation.

D = Difference of rank between paired items in two series.

N = No. of Pairs of observation.

Spearman's rank correlation coefficient lies between +1 and - 1.

Under this method two types of problems can be given. (i) Where ranks are given (ii) Where ranks are not given.

i) When ranks are given

Steps:

- Take the difference of the two ranks and denote the difference as D .
- Square these differences and obtain the total. Denote it as D^2 and apply the formula.

ii) When ranks are not given

When the actual data is given and ranks are not given, we have to assign the ranks. Ranks can be assigned by taking either the higher value as 1 or the lowest value as 1. The same should be followed for both the variables. Once the ranks are assigned, the rest of the working remains the same.

➤ **Equal Ranks**

Where there is more than one item with the same value a common rank is given to such items. If two individual are ranked equal at fourth place,

they are each given the rank $\frac{4+5}{2} = 4.5$. In the same manner if three items are ranked at 6th place,

they are given $\frac{6+7+8}{3} = 7$ rank.

When equal ranks are given to some items, an adjustment is made in the formula by adding $\frac{1}{12}$ (m^3-m) to the value of ΣD^2 where m stands for the number of items whose ranks are equal. If there are more than one such group of items with common ranks, this value is added as many times the number of such groups. It may be written as.

$$r_s = 1 - \frac{6 \left[\Sigma D^2 + \frac{1}{12}(m^3 - m) + \frac{1}{12}(m^3 - m) \dots \right]}{N^3 - N}$$

PROBLEMS ON RANK CORRELATION

10. Calculate the coefficient of correlation from the following data by the spearman's Rank Differences method :

Prices of Tea (₹)	75	88	95	70	60	80	81	50
Prices of Coffee (₹)	120	134	150	115	110	140	142	100

Sol :

(Dec.-18)

Price of Tea (X)	R ₁	Price of Coffee (Y)	R ₂	R ₁ -R ₂ =d	d ²
75	4	120	4	0	0
88	7	134	5	2	4
95	8	150	8	0	0
70	3	115	3	0	0
60	2	110	2	0	0
80	5	140	6	-1	1
81	6	142	7	-1	1
50	1	100	1	0	0
					6

$$r_s = 1 - \frac{6 \Sigma (d)^2}{(N)^3 - N} = 1 - \frac{6 \times 6}{(8)^3 - 8}$$

$$r_s = 1 - \frac{384}{512 - 8} = 1 - \frac{36}{504}$$

$$r_s = 1 - 0.071$$

$$= 0.238$$

11. Compute Rank coefficient of correlation for the following data :

X	15	20	28	12	40	60	20	80
Y	40	30	50	30	20	10	30	60

Sol :

X	R ₁	Y	R ₂	R ₁ - R ₂ = d	d ²
15	2	40	6	-4	16
20	3.5	30	4	-0.5	0.25
28	5	50	7	-2	4
12	1	30	4	-3	9
40	6	20	2	4	16
60	7	10	1	6	36
20	3.5	30	4	-0.5	0.25
80	8	60	8	0	0
					81.5

$$r_s = 1 - \frac{6 \left[\sum d^2 + \frac{1}{12}(m^3 - m) + \frac{1}{12}(m^3 - m) \right]}{(N)^3 - N}$$

$$r_s = 1 - \frac{6 \left[81.5 + \frac{1}{12}(2^3 - 2) + \frac{1}{12}(3^3 - 3) \right]}{(8)^3 - 8}$$

$$r_s = 1 - \frac{6[81.5 + 0.5 + 2]}{512 - 8}$$

$$r_s = 1 - \frac{6(84)}{504}$$

$$r_s = 1 - \frac{504}{504}$$

$$r_s = 1 - 1$$

$$= 0$$

12. Calculate Rank correlation co-efficient from the following data :

	A	B	C	D	E
x :	1	2	3	5	4
f :	2	1	4	3	5

Sol :

(May-18)

Rank correlation co-efficient

X	R ₁	Y	R ₂	(R ₁ - R ₂) 'D'	D ²
1	5	2	4	1	1
2	4	1	5	-1	1
3	3	4	2	1	1
5	1	3	3	-2	4
4	2	5	1	1	1
					8

$$1 - \frac{6\sum D^2}{N^3 - N}$$

$$1 - \frac{6 \times 8}{5^3 - 5}$$

$$1 - \frac{48}{125 - 5}$$

$$1 - \frac{48}{120}$$

$$1 - 0.4 = 0.6$$

13. From the following data calculate rank correlation

x	20	25	60	45	80	25	55	65	25	75
f	45	50	55	50	60	70	72	78	80	63

Sol :

(Dec.-17)

X	R ₁	Y	R ₂	$\frac{R_1 - R_2}{(D)}$	D ²
20	10	45	10	0	0
25	8	50	8.5	-0.5	0.25
60	4	55	7	-3	9
45	6	50	8.5	-2.5	6.25
80	1	60	6	-5	25
25	8	70	4	4	16
55	5	72	3	2	4
65	3	78	2	1	1
25	8	80	1	7	49
75	2	63	5	-3	9
					$\sum D^2$ 11.95

$$1 - \frac{6 \left(\sum D^2 + \frac{1}{12}(m^3 - m) + \frac{1}{12}(m^3 - m) \right)}{N^3 - N}$$

$$1 - \frac{6 \left\{ 119.5 + \frac{1}{12}(3^3 - 3) + \frac{1}{12}(2^3 - 2) \right\}}{10^3 - 10}$$

$$\frac{1 - 6\{119.5 + 2 + 0.5\}}{990}$$

$$1 - \frac{6(122)}{990}$$

$$1 - \frac{732}{990}$$

$$1 - 0.73 = 0.27$$

14. From the ranks of 10 student's in Accountancy and Statistics given below,

Calculate Rank Correlation Coefficient

Ranks in Accountancy	1	2	3	4	5	6	7	8	9	10
Rank in Statistics	1	3	5	6	7	4	8	10	9	2

Sol :

(June-19)

R_1	R_2	$R_1 - R_2 (D)$	D^2
1	1	0	0
2	3	-1	1
3	5	-2	4
4	6	-2	4
5	7	-2	4
6	4	2	4
7	8	-1	1
8	10	-2	4
9	9	0	0
10	2	8	16
			38

$$r_k = 1 - \frac{6\sum d^2}{N^3 - N} = 1 - \frac{6(38)}{10^3 - 10} = 1 - \frac{228}{1000 - 10} = 1 - \frac{228}{990}$$

$$= 1 - 0.230 = 0.7696$$

5.5 CONCURRENT DEVIATION METHOD

Q18. Explain briefly about Concurrent Deviation Method.

Ans :

Under this method correlation is calculated between the direction of deviation in the two series. The magnitude of change is ignored. The formula is

$$rc = \pm \sqrt{\pm \left(\frac{2c - n}{n} \right)} \text{ where}$$

r_c = Coefficient of concurrent deviation,

c = numbers of concurrent deviations or positive signs,

n = number of pairs of observations compared. It is always one less than the actual number of pairs of observation.

Steps

1. Find the direction of change of X variable. Find whether the second figure in the series is increasing, decreasing or is constant. If it is increasing, put a + sign. If decreasing, put a sign. If constant, put zero. Similarly find the direction of change of other variable compared to the previous number. Denote this column as D_x .
2. Find the direction of change of Y variable in the same manner and denote the column as D_y .
3. Multiply D_x with D_y and determine the value of C (Concurrent deviation) i.e. positive signs and apply the formula.

PROBLEMS ON CONCURRENT DEVIATION METHOD

15. Calculate coefficient of concurrent deviations from the data given below.

X	200	250	290	275	300	250
Y	45	50	60	55	62	50

Sol :

Calculation of Coefficient of Concurrent deviation

Direction of change Direction change

X	D_x	Y	D_y	$D_x D_y$
200		45		+
250	+	50	+	+
290	+	60	+	+
275	-	55	-	+
300	+	62	+	+
250	-	50	-	+
				c = 5

$$n = 5 (N - 1), C = 5$$

$$rc = \pm \sqrt{\pm \left(\frac{2c - n}{n} \right)} = \pm \sqrt{\left(\frac{10 - 5}{5} \right)} = \pm \sqrt{\pm \frac{5}{5}} = 1$$

- 16. Find Coefficient of Concurrent Deviation given that $n = 5$ and number of concurrent deviations (i.e. C) are 5%.**

Sol :

(June-19)

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

$$C = 5, n = 5$$

$$= \pm \sqrt{\pm \left(\frac{(2 \times 5) - 5}{5} \right)}$$

$$= \pm \sqrt{\pm \frac{10 - 5}{5}}$$

$$= \pm \sqrt{\pm \left(\frac{5}{5} \right)}$$

$$= \pm \sqrt{\pm 1} = 1$$

Rahul Publications

Exercises Problems

1. Find the coefficient of correlation between X and Y according to Karl Pearson's formula:

X	1	2	3	4	5	6	7	8	9	10	11
Y	4	7	10	13	16	19	22	25	28	31	34

(Ans: +1)

2. The following data relate to the age of 10 employees and the number of days on which they reported sick in a month.

Age (X)	28	32	38	42	46	52	54	57	58	63
Sick Days (Y)	0	1	3	4	2	5	4	6	7	8

(Ans: $r = +0.93$)

3. Find out the coefficient of correlation of Karl Pearson from the following data:

X	65	66	67	67	68	69	70	72
Y	67	68	65	68	72	72	69	71

(Ans: $r = 0.603$)

4. Calculate the coefficient of correlation between the marks in Statistics (X) and Accountancy (Y) of 12 students from the following data.

X	52	74	93	55	41	23	92	64	40	71	33	71
Y	45	80	63	60	35	40	70	58	43	64	51	75

(Ans: $r = 0.7885$; P.E. = 0.0737)

5. Two judges in a beauty competition rank the 12 entries as follows:

X	1	2	3	4	5	6	7	8	9	10	11	12
Y	12	9	6	10	3	5	4	7	8	2	11	1

What degree of agreement is there between the two judges.

(Ans: 0.454)

6. The ranks of the same 10 students in tests in Accountancy and Statistics are given below:

Ranks in Accountancy	1	2	3	4	5	6	7	8	9	10
Ranks in Statistics	1	3	5	6	7	4	8	10	9	2

Calculate the rank correlation coefficient and comment on the value.

(Ans: 0.479)

Short Question and Answers

1. Explain briefly about probable error.

Ans :

The probable error of the coefficient of correlation helps in interpretation. The probable error of the coefficient of correlation is obtained as follows:

$$\text{P.E. of } r = 0.6745 \frac{1-r^2}{\sqrt{N}}$$

Where

r = Coefficient of correlation

N = Number of pairs of observations.

If the probable error is added to and subtracted from the coefficient of correlation, it would give two such limits within which we can reasonably expect the value of coefficient of correlation to vary.

Symbolically $P(\rho) = r \pm \text{P.E.}$

Where 'P' denotes the correlation in the population. Suppose, the Coefficient of correlation for a pair of 10 observations is 0.8 and its P.E. is 0.05. the limits of the correlation in the population would be $r \pm \text{P.E.}$ i.e. 0.8 ± 0.05 or $0.75 - 0.85$.

If the value of r is less than the probable error then r is not at all significant, i.e. there is no evidence of correlation. If the value of r is more than six times the probable error, it is significant. Hence it can be said that r is significant, when

$$r > 6 \text{ P.E. or } \frac{r}{\text{P.E.}} > 6.$$

2. Types of Correlation.

Ans :

Broadly speaking, there are four types of correlation, namely,

- A) Positive correlation,
- B) Negative correlation,
- C) Linear correlation and
- D) Non-Linear Correlation.

A) Positive correlation

If the values of two variables deviate in the same direction i.e., if increase in the values of one variable results, on an average, in a corresponding increase in the values of the other variable or if a decrease in the values of one variable results, on an average, in a corresponding decrease in the values of the other variable, the corresponding correlation is said to be positive or direct.

Examples

- i) Sales revenue of a product and expenditure on Advertising.
- ii) Amount of rain fall and yield of a crop (up to a point)
- iii) Price of a commodity and quantity of supply of a commodity
- iv) Height of the Parent and the height of the Child.
- v) Number of patients admitted into a Hospital and Revenue of the Hospital.
- vi) Number of workers and output of a factory.

B) Negative Correlation

Correlation is said to be negative or inverse if the variables deviate in the opposite direction i.e., if the increase (decrease) in the values of one variable results, on the average, in a corresponding decrease (increase) in the values of the other variable.

C) Linear Correlation

Two variables are said to be linearly related if corresponding to a unit change in one variable there is a constant change in the other variable over the entire range of the values.

If two variables are related linearly, then we can express the relationship as

$$Y = a + bX$$

where ' a ' is called as the "intercept" (If $X = 0$, then $Y = a$) and ' b ' is called as the "rate of change" or slope.

D) Non Linear (Curvilinear) Correlation

If corresponding to a unit change in one variable, the other variable does not change in a constant rate, but change at varying rates, then the relationship between two variables is said to be nonlinear or curvilinear as shown in fig. below. In this case, if the data are plotted on the graph, we do not get a straight line curve.

Mathematically, the correlation is non-linear if the slope of the plotted curve is not constant. Data relating to Economics, Social Science and Business Management do exhibit often non-linear relationship. We confine ourselves to linear correlation only.

3. What do you mean by correlation?

Ans :

Meaning

Correlation is the study of the linear relationship between two variables. When there is a relationship of 'quantitative measure between two set of variables, the appropriate statistical tool for measuring the relationship and expressing each in a precise way is known as correlation.

For example, there is a relationship between the heights and weights of persons, demand and prices of commodities etc.

Correlation analysis is the statistical tool we can use to describe the degree to which one variable is linearly related to another.

Definitions

- (i) **According to L.R. Connor** "If two or more quantities vary in sympathy so that movements in one tend to be accompanied by corresponding movements in other(s) then they are said to be correlated."
- (ii) **According to A.M. Tuttle** "Correlation is an analysis of covariation between two or more variables".
- (iii) **According to Croxton and Cowden** "When the relationship is of a quantitative nature, the appropriate statistical tool for

discovering and measuring the relationship and expressing it in a brief formula is known as correlation."

- (iv) **According to Ya Lun Chou** "Correlation analysis attempts to determine the 'degree of relationship' between variable".

4. Applications of correlation.

Ans :

- (a) Relationship between advertising expenditure and sales revenue of a company for a period of time.
- (b) Relationship between time or the day and the temperatures recorded for a particular city.
- (c) Relationship between ages of the husbands and wives for a group of couples under study.
- (d) Relationship between heights and weights of a group of students.
- (e) Relationship between age and blood pressure of men.
- (f) Relationship between imports of raw material and exports of finished goods of a firm in a particular year.
- (g) Relationship between amount invested (capital) and profits earned by different firms.
- (h) Relationship between price and demand for a particular commodity.

5. What is Scatter Diagram?

Ans :

Scatter diagram method is the simplest way of diagrammatic representation of a bivariate distribution and helps in ascertaining the correlation between the two variables under study i.e., it portrays the relationship between these two variables graphically.

Procedure of Scatter Diagram

Given pair of values (x_1, y_1) (x_2, y_2) ... (x_n, y_n) of two variables X and Y. Take the independent variables on the X axis and the dependent variable on the Y-axis. The V points denoted by the pair of values are plotted on the graph. The diagram of dots thus obtained is the scatter diagram. Regarding the correlation between the two variables, the scatter diagram can be interpreted as follows,

- (i) If the points reveal any upward or downward trend, the variables are said to be correlated, otherwise uncorrelated.
- (ii) If the points are very close to each other, a good amount of correlation exists, else poor correlation exists.
- (iii) Upward trend indicates positive correlation and downward trend indicates negative correlation.

6. Merits and Demerits of Scatter Diagram.

Ans :

Merits

- Scatter diagram is a simple and attractive method of finding out the nature of correlation between two variables.
- It is a non-mathematical method of studying correlation. It is easy to understand.
- We can get a rough idea at a glance whether it is a positive or negative correlation.
- It is not influenced by extreme items.
- It is a first step in finding out the relationship between two variables.

Demerits

- The major limitation of the method is that it only gives a visual picture of the relationship of two variables. It only tells us whether there is correlation between the variables, and if so, then in which direction, positive (or) negative.
- It does not give an idea about the precise degree of relationship as it is not amenable to mathematical treatment.

7. Merits and Demerits of Coefficient of Correlation.

Ans :

Merits

- It takes into account all items of the variable(s).
- It is a numerical measure and hence more objective.
- It measures both direction as well as degree of change.

- It facilitates comparisons between two series.
- It is capable of further Algebraic treatment
- It is more practical and hence popular and is more commonly used.

Demerits

- It is not easy to calculate as complex formulae are involved.
- It is more time consuming compared to methods such as rank correlation
- It assumes a linear relationship between the two variables which may not be correct
- It is impacted by extreme values as it is based on mean and standard deviation.
- It is not easy to interpret.

8. How do you interpret the coefficient of correlation?

Ans :

The investigator must take utmost care while interpreting the value of 'r' and interpretation depends very much on experience.

Rules of Interpretation

1. When $r = 0$, it implies that three variables are uncorrelated.
2. When $r = +1$, it implies that there is a perfect positive relationship between the variables.
3. When $r = -1$, it implies that there is a perfect negative relationship between the variables.
4. The closer the value of 'r' to $+1$ or -1 , the closer the relationship between the variables.
5. The closeness of the relationship is not proportional to r.

9. What is Rank correlation ?

Ans :

The Karl Pearson's method is based on the assumption that the population being studied is normally distributed. When it is known that the population is not normal or when the shape of the

distribution is not known, there is need for a measure of correlation that involves no assumption about the parameter of the population.

It is possible to avoid making any assumptions about the populations being studied by ranking the observations according to size and basing the calculations on the ranks rather than upon the original observations. It does not matter which way the items are ranked, item number one may be the largest or it may be the smallest. Using ranks rather than actual observations gives the coefficient of rank correlation.

This method of finding out covariability or the lack of it between two variables was developed by the British Psychologist Charles Edward Spearman in 1904.

10. State the Merits and Demerits of rank correlation.

Ans :

Merits

1. It is easy to understand and calculate
2. It is not impacted by extreme values.
3. It is a numerical measure and provides objectivity to subjective ranking.
4. It is the only method of finding correlation with respect to qualitative factors such as honesty, beauty, etc.
5. It measures both direction as well as degree of change.
6. It facilitates comparisons between two series.
7. It can be applied to irregular data also.
8. It is ideal when the number of observations is very small.

Demerits

1. It cannot be applied to grouped data
2. It lacks the precision of Karl Pearson's Coefficient of Correlation.
3. All the information concerning the variable is not used.
4. The Computation becomes complicated as the number of observations increase.

11. Explain briefly about Concurrent Deviation Method.

Ans :

Under this method correlation is calculated between the direction of deviation in the two series. The magnitude of change is ignored. The formula is

$$r_c = \pm \sqrt{\pm \left(\frac{2c - n}{n} \right)} \text{ where}$$

r_c = Coefficient of concurrent deviation,

c = numbers of concurrent deviations or positive signs,

n = number of pairs of observations compared. It is always one less than the actual number of pairs of observation.

Steps

1. Find the direction of change of X variable. Find whether the second figure in the series is increasing, decreasing or is constant. If it is increasing, put a + sign. If decreasing, put a sign. If constant, put zero. Similarly find the direction of change of other variable compared to the previous number. Denote this column as D_x .
2. Find the direction of change of Y variable in the same manner and denote the column as D_y .
3. Multiply D_x with D_y and determine the value of C (Concurrent deviation) i.e. positive signs and apply the formula.

Choose the Correct Answers

1. Who is the chief exponent of the principle of correlation
(a) Bowley (b) Spearman [c]
(c) Karl Pearson (d) Connor
2. What is meant by correlation [a]
(a) Measuring relationship between two variables
(b) Measuring changes in prices
(c) Measuring average correlation between the variables
(d) Prediction
3. Study of relationship between three or more variables is known as [a]
(a) Multiple correlation (b) Simple correlation
(c) Both (d) None
4. Which is the non-mathematical method of studying the correlation between the variables [c]
(a) Scatter diagram method (b) Graphic method
(c) Both (a) and (b) (d) None of these
5. If 'r' is more than 6 times the probable error, correlation is - [c]
(a) Negative (b) Positive
(c) Significant (d) None of the above
6. A process by which we estimate the value of dependent variable on the basis of one or more independent variables is called: [b]
(a) Correlation (b) Regression
(c) Residual (d) Slope
7. The value we would predict for the dependent variable when the independent variables are all equal to zero is called: [c]
(a) Slope (b) Sum of residual
(c) Intercept (d) Difficult to tell
8. The predicted rate of response of the dependent variable to changes in the independent variable is called: [b]
(a) Slope (b) Intercept
(c) Error (d) Regression equation
9. In simple regression equation, the numbers of variables involved are: [c]
(a) 0 (b) 1
(c) 2 (d) 3
10. The dependent variable is also called: [d]
(a) Regress and variable (b) Predict and variable
(c) Explained variable (d) All of these

Fill in the blanks

1. Karl Pearson's coefficient of correlation measures _____.
2. Maximum positive value of coefficient of correlation is _____.
3. Correlation between price and demand is _____.
4. Correlation between price and supply is _____.
5. If the correlation is perfect positive, its value is _____.
6. The independent variable is also called _____.
7. If the value of any regression coefficient is zero, then two variables are _____.
8. The range of regression coefficient is _____.
9. The term regression was used by _____.
10. The correlation coefficient between X and -X is _____.

ANSWERS

1. Linear correlation
2. +1
3. negative
4. Positive
5. +1
6. Regressor
7. Independent
8. -" to +"
9. Galton
10. -1

LOGARITHMS VALUES

											Mean Differences								
	0	1	2	3	4	5	6	7	8	9	1	2	3	4	5	6	7	8	9
10	0000	0043	0086	0128	0170	0212	0253	0294	0334	0374	4	8	12	17	21	25	29	33	37
11	0414	0453	0492	0531	0569	0607	0645	0682	0719	0755	4	8	11	15	19	23	26	30	34
12	0792	0828	0864	0899	0934	0969	1004	1038	1072	1106	3	7	10	14	17	21	24	28	31
13	1139	1173	1206	1239	1271	1303	1335	1367	1399	1430	3	6	10	13	16	19	23	26	29
14	1461	1492	1523	1553	1584	1614	1644	1673	1703	1732	3	6	9	12	15	18	21	24	27
15	1761	1790	1818	1847	1875	1903	1931	1959	1987	2014	3	6	8	11	14	17	20	22	25
16	2041	2068	2095	2122	2148	2175	2201	2227	2253	2279	3	5	8	11	13	16	18	21	24
17	2304	2330	2355	2380	2405	2430	2455	2480	2504	2529	2	5	7	10	12	15	17	20	22
18	2553	2577	2601	2625	2648	2672	2695	2718	2742	2765	2	5	7	9	12	14	16	19	21
19	2788	2810	2833	2856	2878	2900	2923	2945	2967	2989	2	4	7	9	11	13	16	18	20
20	3010	3032	3054	3075	3096	3118	3139	3160	3181	3201	2	4	6	8	11	13	15	17	19
21	3222	3243	3263	3284	3304	3324	3345	3365	3385	3404	2	4	6	8	10	12	14	16	18
22	3424	3444	3464	3483	3502	3522	3541	3560	3579	3598	2	4	6	8	10	12	14	15	17
23	3617	3636	3655	3674	3692	3711	3729	3747	3766	3784	2	4	6	7	9	11	13	15	17
24	3802	3820	3838	3856	3874	3892	3909	3927	3945	3962	2	4	5	7	9	11	12	14	16
25	3979	3997	4014	4031	4048	4065	4082	4099	4116	4133	2	3	5	7	9	10	12	14	15
26	4150	4168	4183	4200	4216	4232	4249	4265	4281	4298	2	3	5	7	8	10	11	13	15
27	4314	4330	4346	4362	4378	4393	4409	4425	4440	4456	2	3	5	6	8	9	11	13	14
28	4472	4487	4502	4518	4533	4548	4564	4579	4594	4609	2	3	5	6	8	9	11	12	14
29	4624	4639	4654	4669	4683	4698	4713	4728	4742	4757	1	3	4	6	7	9	10	12	13
30	4771	4786	4800	4814	4829	4843	4857	4871	4886	4900	1	3	4	6	7	9	10	11	13
31	4914	4928	4942	4955	4969	4983	4997	5011	5024	5038	1	3	4	6	7	8	10	11	12
32	5051	5065	5079	5092	5105	5119	5132	5145	5159	5172	1	3	4	5	7	8	9	11	12
33	5185	5198	5211	5224	5237	5250	5263	5276	5289	5302	1	3	4	5	6	8	9	10	12
34	5315	5328	5340	5353	5366	5378	5391	5403	5416	5428	1	3	4	5	6	8	9	10	11
35	5441	5453	5465	5478	5490	5502	5514	5527	5539	5551	1	2	4	5	6	7	9	10	11
36	5563	5575	5587	5599	5611	5623	5635	5647	5658	5670	1	2	4	5	6	7	8	10	11
37	5682	5694	5705	5717	5729	5740	5752	5763	5775	5786	1	2	3	5	6	7	8	9	10
38	5798	5809	5821	5832	5843	5855	5866	5877	5888	5899	1	2	3	5	6	7	8	9	10
39	5911	5922	5933	5944	5955	5966	5977	5988	5999	6010	1	2	3	4	5	7	8	9	10
40	6021	6031	6042	6053	6064	6075	6085	6096	6107	6117	1	2	3	4	5	6	8	9	10
41	6128	6138	6149	6160	6170	6180	6191	6201	6212	6222	1	2	3	4	5	6	7	8	9
42	6232	6243	6253	6263	6274	6284	6294	6304	6314	6325	1	2	3	4	5	6	7	8	9
43	6335	6345	6355	6365	6375	6385	6395	6405	6415	6425	1	2	3	4	5	6	7	8	9
44	6435	6444	6454	6464	6474	6484	6493	6503	6513	6522	1	2	3	4	5	6	7	8	9
45	6532	6542	6551	6561	6571	6580	6590	6599	6609	6618	1	2	3	4	5	6	7	8	9
46	6628	6637	6646	6656	6665	6675	6684	6693	6702	6712	1	2	3	4	5	6	7	7	8
47	6721	6730	6739	6749	6758	6767	6776	6785	6794	6803	1	2	3	4	5	5	6	7	8
48	6812	6821	6830	6839	6848	6857	6866	6875	6884	6893	1	2	3	4	4	5	6	7	8
49	6902	6911	6920	6928	6937	6946	6955	6964	6972	6981	1	2	3	4	4	5	6	7	8
50	6990	6998	7007	7016	7024	7033	7042	7050	7059	7067	1	2	3	3	4	5	6	7	8
51	7076	7084	7093	7101	7110	7118	7126	7135	7143	7152	1	2	3	3	4	5	6	7	8
52	7160	7168	7177	7185	7193	7202	7210	7218	7226	7235	1	2	2	3	4	5	6	7	7
53	7243	7251	7259	7267	7275	7284	7292	7300	7308	7316	1	2	2	3	4	5	6	6	7
54	7324	7332	7340	7348	7356	7364	7372	7380	7388	7396	1	2	2	3	4	5	6	6	7

											Mean Differences								
	0	1	2	3	4	5	6	7	8	9	1	2	3	4	5	6	7	8	9
55	7404	7412	7419	7427	7435	7443	7451	7459	7466	7474	1	2	2	3	4	5	5	6	7
56	7482	7490	7497	7505	7513	7520	7528	7536	7543	7551	1	2	2	3	4	5	5	6	7
57	7559	7566	7574	7582	7589	7597	7604	7612	7619	7627	1	2	2	3	4	5	5	6	7
58	7634	7642	7649	7657	7664	7672	7679	7686	7694	7701	1	1	2	3	4	4	5	6	7
59	7709	7716	7723	7731	7738	7745	7752	7760	7767	7774	1	1	2	3	4	4	5	6	7
60	7782	7789	7796	7803	7810	7818	7825	7832	7839	7846	1	1	2	3	4	4	5	6	6
61	7853	7860	7868	7875	7882	7889	7896	7903	7910	7917	1	1	2	3	4	4	5	6	6
62	7924	7931	7938	7945	7952	7959	7966	7973	7980	7987	1	1	2	3	3	4	5	6	6
63	7993	8000	8007	8014	8021	8028	8035	8041	8048	8055	1	1	2	3	3	4	5	5	6
64	8062	8069	8075	8082	8089	8096	8102	8109	8116	8122	1	1	2	3	3	4	5	5	6
65	8129	8136	8142	8149	8156	8162	8169	8176	8182	8189	1	1	2	3	3	4	5	5	6
66	8195	8202	8209	8215	8222	8228	8235	8241	8248	8254	1	1	2	3	3	4	5	5	6
67	8261	8267	8274	8280	8287	8293	8299	8306	8312	8319	1	1	2	3	3	4	5	5	6
68	8325	8331	8338	8344	8351	8357	8363	8370	8376	8382	1	1	2	3	3	4	4	5	6
69	8388	8395	8401	8407	8414	8420	8426	8432	8439	8445	1	1	2	2	3	4	4	5	6
70	8451	8457	8463	8470	8476	8482	8488	8494	8500	8506	1	1	2	2	3	4	4	5	6
71	8513	8519	8525	8531	8537	8543	8549	8555	8561	8567	1	1	2	2	3	4	4	5	5
72	8573	8579	8585	8591	8597	8603	8609	8615	8621	8627	1	1	2	2	3	4	4	5	5
73	8633	8639	8645	8651	8657	8663	8669	8675	8681	8686	1	1	2	2	3	4	4	5	5
74	8692	8698	8704	8710	8716	8722	8727	8733	8739	8745	1	1	2	2	3	4	4	5	5
75	8751	8756	8762	8768	8774	8779	8785	8791	8797	8802	1	1	2	2	3	3	4	5	5
76	8808	8814	8820	8825	8831	8837	8842	8848	8854	8859	1	1	2	2	3	3	4	5	5
77	8865	8871	8876	8882	8887	8893	8899	8904	8910	8915	1	1	2	2	3	3	4	4	5
78	8921	8927	8932	8938	8943	8949	8954	8960	8965	8971	1	1	2	2	3	3	4	4	5
79	8976	8982	8987	8993	8998	9004	9009	9015	9020	9025	1	1	2	2	3	3	4	4	5
80	9031	9036	9042	9047	9053	9058	9063	9069	9074	9079	1	1	2	2	3	3	4	4	5
81	9085	9090	9096	9101	9106	9112	9117	9122	9128	9133	1	1	2	2	3	3	4	4	5
82	9138	9143	9149	9154	9159	9165	9170	9175	9180	9186	1	1	2	2	3	3	4	4	5
83	9191	9196	9201	9206	9212	9217	9222	9227	9232	9238	1	1	2	2	3	3	4	4	5
84	9243	9248	9253	9258	9263	9269	9274	9279	9284	9289	1	1	2	2	3	3	4	4	5
85	9294	9299	9304	9309	9315	9320	9325	9330	9335	9340	1	1	2	2	3	3	4	4	5
86	9345	9350	9355	9360	9365	9370	9375	9380	9385	9390	1	1	2	2	3	3	4	4	5
87	9395	9400	9405	9410	9415	9420	9425	9430	9435	9440	0	1	1	2	2	3	3	4	4
88	9445	9450	9455	9460	9465	9469	9474	9479	9484	9489	0	1	1	2	2	3	3	4	4
89	9494	9499	9504	9509	9513	9518	9523	9528	9533	9538	0	1	1	2	2	3	3	4	4
90	9542	9547	9552	9557	9562	9566	9571	9576	9581	9586	0	1	1	2	2	3	3	4	4
91	9590	9595	9600	9605	9609	9614	9619	9624	9628	9633	0	1	1	2	2	3	3	4	4
92	9638	9643	9647	9653	9657	9661	9666	9671	9675	9680	0	1	1	2	2	3	3	4	4
93	9685	9689	9694	9699	9703	9708	9713	9717	9722	9727	0	1	1	2	2	3	3	4	4
94	9731	9736	9741	9745	9750	9754	9759	9763	9768	9773	0	1	1	2	2	3	3	4	4
95	9777	9782	9786	9791	9795	9800	9805	9809	9814	9818	0	1	1	2	2	3	3	4	4
96	9823	9827	9832	9836	9841	9845	9850	9854	9859	9863	0	1	1	2	2	3	3	4	4
97	9868	9872	9877	9881	9886	9890	9894	9899	9903	9908	0	1	1	2	2	3	3	4	4
98	9912	9917	9921	9926	9930	9934	9939	9943	9948	9952	0	1	1	2	2	3	3	4	4
99	9956	9961	9965	9969	9974	9978	9983	9987	9991	9996	0	1	1	2	2	3	3	3	4

ANTI-LOGARITHMS VALUES

											Mean Differences								
	0	1	2	3	4	5	6	7	8	9	1	2	3	4	5	6	7	8	9
00	1000	1002	1005	1007	1009	1012	1014	1016	1019	1021	0	0	1	1	1	1	2	2	2
01	1023	1026	1028	1030	1033	1035	1038	1040	1042	1045	0	0	1	1	1	1	2	2	2
02	1047	1050	1052	1054	1057	1059	1062	1064	1067	1069	0	0	1	1	1	1	2	2	2
03	1072	1074	1076	1079	1081	1084	1086	1089	1091	1094	0	0	1	1	1	1	2	2	2
04	1096	1099	1102	1104	1107	1109	1112	1114	1117	1119	0	1	1	1	1	2	2	2	2
05	1122	1125	1127	1130	1132	1135	1138	1140	1143	1146	0	1	1	1	1	2	2	2	2
06	1148	1151	1153	1156	1159	1161	1164	1167	1169	1172	0	1	1	1	1	2	2	2	2
07	1175	1178	1180	1183	1186	1189	1191	1194	1197	1199	0	1	1	1	1	2	2	2	2
08	1202	1205	1208	1211	1213	1216	1219	1222	1225	1227	0	1	1	1	1	2	2	2	3
09	1230	1233	1236	1239	1242	1245	1247	1250	1253	1256	0	1	1	1	1	2	2	2	3
10	1259	1262	1265	1268	1271	1274	1276	1279	1282	1285	0	1	1	1	1	2	2	2	3
11	1288	1291	1294	1297	1300	1303	1306	1309	1312	1315	0	1	1	1	2	2	2	2	3
12	1318	1321	1324	1327	1330	1334	1337	1340	1343	1346	0	1	1	1	2	2	2	2	3
13	1349	1352	1355	1358	1361	1365	1368	1371	1374	1377	0	1	1	1	2	2	2	3	3
14	1380	1384	1387	1390	1393	1396	1400	1403	1406	1409	0	1	1	1	2	2	2	3	3
15	1413	1416	1419	1422	1426	1429	1432	1435	1439	1442	0	1	1	1	2	2	2	3	3
16	1445	1449	1452	1455	1459	1462	1466	1469	1472	1476	0	1	1	1	2	2	2	3	3
17	1479	1483	1486	1489	1493	1496	1500	1503	1507	1510	0	1	1	1	2	2	2	3	3
18	1514	1517	1521	1524	1528	1531	1535	1538	1542	1545	0	1	1	1	2	2	2	3	3
19	1549	1552	1556	1560	1563	1567	1570	1574	1578	1581	0	1	1	1	2	2	3	3	3
20	1585	1589	1592	1596	1600	1603	1607	1611	1614	1618	0	1	1	1	2	2	3	3	3
21	1622	1626	1629	1633	1637	1641	1644	1648	1652	1656	0	1	1	2	2	2	3	3	3
22	1660	1663	1667	1671	1675	1679	1683	1687	1690	1694	0	1	1	2	2	2	3	3	3
23	1698	1702	1706	1710	1714	1718	1722	1726	1730	1734	0	1	1	2	2	2	3	3	4
24	1738	1742	1746	1750	1754	1758	1762	1766	1770	1774	0	1	1	2	2	2	3	3	4
25	1778	1782	1786	1791	1795	1799	1803	1807	1811	1816	0	1	1	2	2	2	3	3	4
26	1820	1824	1828	1832	1837	1841	1845	1849	1854	1858	0	1	1	2	2	3	3	3	4
27	1862	1866	1871	1875	1879	1884	1888	1892	1897	1901	0	1	1	2	2	3	3	3	4
28	1905	1910	1914	1919	1923	1928	1932	1936	1941	1945	0	1	1	2	2	3	3	4	4
29	1950	1954	1959	1963	1968	1972	1977	1982	1986	1991	0	1	1	2	2	3	3	4	4
30	1995	2000	2004	2009	2014	2018	2023	2028	2032	2037	0	1	1	2	2	3	3	4	4
31	2042	2046	2051	2056	2061	2065	2070	2075	2080	2084	0	1	1	2	2	3	3	4	4
32	2089	2094	2099	2104	2109	2113	2118	2123	2128	2133	0	1	1	2	2	3	3	4	4
33	2138	2143	2148	2153	2158	2163	2168	2173	2178	2183	0	1	1	2	2	3	3	4	4
34	2188	2193	2198	2203	2208	2213	2218	2223	2228	2234	1	1	2	2	3	3	4	4	5
35	2239	2244	2249	2254	2259	2265	2270	2275	2280	2286	1	1	2	2	3	3	4	4	5
36	2291	2296	2301	2307	2312	2317	2323	2328	2333	2339	1	1	2	2	3	3	4	4	5
37	2344	2350	2355	2360	2366	2371	2377	2382	2388	2393	1	1	2	2	3	3	4	4	5
38	2399	2404	2410	2415	2421	2427	2432	2438	2443	2449	1	1	2	2	3	3	4	4	5
39	2455	2460	2466	2472	2477	2483	2489	2495	2500	2506	1	1	2	2	3	3	4	5	5
40	2512	2518	2523	2529	2535	2541	2547	2553	2559	2564	1	1	2	2	3	4	4	5	5
41	2570	2576	2582	2588	2594	2600	2606	2612	2618	2624	1	1	2	2	3	4	4	5	5
42	2630	2636	2642	2649	2655	2661	2667	2673	2679	2685	1	1	2	2	3	4	4	5	6
43	2692	2698	2704	2710	2716	2723	2729	2735	2742	2748	1	1	2	3	3	4	4	5	6
44	2754	2761	2767	2773	2780	2786	2793	2799	2805	2812	1	1	2	3	3	4	4	5	6
45	2818	2825	2831	2838	2844	2851	2858	2864	2871	2877	1	1	2	3	3	4	5	5	6
46	2884	2891	2897	2904	2911	2917	2924	2931	2938	2944	1	1	2	3	3	4	5	5	6
47	2951	2958	2965	2972	2979	2985	2992	2999	3006	3013	1	1	2	3	3	4	5	5	6
48	3020	3027	3034	3041	3048	3055	3062	3069	3076	3083	1	1	2	3	4	4	5	6	6
49	3090	3097	3105	3112	3119	3126	3133	3141	3148	3155	1	1	2	3	4	4	5	6	6

											Mean Differences								
	0	1	2	3	4	5	6	7	8	9	1	2	3	4	5	6	7	8	9
-50	3162	3170	3177	3184	3192	3199	3206	3214	3221	3228	1	1	2	3	4	4	5	6	7
-51	3236	3243	3251	3258	3266	3273	3281	3289	3296	3304	1	2	2	3	4	5	5	6	7
-52	3311	3319	3327	3334	3342	3350	3357	3365	3373	3381	1	2	2	3	4	5	5	6	7
-53	3388	3396	3404	3412	3420	3428	3436	3443	3451	3459	1	2	2	3	4	5	6	6	7
-54	3467	3475	3483	3491	3499	3508	3516	3524	3532	3540	1	2	2	3	4	5	6	6	7
-55	3548	3556	3565	3573	3581	3589	3597	3606	3614	3622	1	2	2	3	4	5	6	7	7
-56	3631	3639	3648	3656	3664	3673	3681	3690	3698	3707	1	2	3	3	4	5	6	7	8
-57	3715	3724	3733	3741	3750	3758	3767	3776	3784	3793	1	2	3	3	4	5	6	7	8
-58	3802	3811	3819	3828	3837	3846	3855	3864	3873	3882	1	2	3	4	4	5	6	7	8
-59	3890	3899	3908	3917	3926	3936	3945	3954	3963	3972	1	2	3	4	5	5	6	7	8
-60	3981	3990	3999	4009	4018	4027	4036	4046	4055	4064	1	2	3	4	5	6	6	7	8
-61	4074	4083	4093	4102	4111	4121	4130	4140	4150	4159	1	2	3	4	5	6	7	8	9
-62	4169	4178	4188	4198	4207	4217	4227	4236	4246	4256	1	2	3	4	5	6	7	8	9
-63	4266	4276	4285	4295	4305	4315	4325	4335	4345	4355	1	2	3	4	5	6	7	8	9
-64	4365	4375	4385	4395	4406	4416	4426	4436	4446	4457	1	2	3	4	5	6	7	8	9
-65	4467	4477	4487	4498	4508	4519	4529	4539	4550	4560	1	2	3	4	5	6	7	8	9
-66	4571	4581	4592	4603	4613	4624	4634	4645	4656	4667	1	2	3	4	5	6	7	9	10
-67	4677	4688	4699	4710	4721	4732	4742	4753	4764	4775	1	2	3	4	5	7	8	9	10
-68	4786	4797	4808	4819	4831	4842	4853	4864	4875	4887	1	2	3	4	6	7	8	9	10
-69	4898	4909	4920	4932	4943	4955	4966	4977	4989	5000	1	2	3	5	6	7	8	9	10
-70	5012	5023	5035	5047	5058	5070	5082	5093	5105	5117	1	2	4	5	6	7	8	9	11
-71	5129	5140	5152	5164	5176	5188	5200	5212	5224	5236	1	2	4	5	6	7	8	10	11
-72	5248	5260	5272	5284	5297	5309	5321	5333	5346	5358	1	2	4	5	6	7	9	10	11
-73	5370	5383	5395	5408	5420	5433	5445	5458	5470	5483	1	3	4	5	6	8	9	10	11
-74	5495	5508	5521	5534	5546	5559	5572	5585	5598	5610	1	3	4	5	6	8	9	10	12
-75	5623	5636	5649	5662	5675	5689	5702	5715	5728	5741	1	3	4	5	7	8	9	10	12
-76	5754	5768	5781	5794	5808	5821	5834	5848	5861	5875	1	3	4	5	7	8	9	11	12
-77	5888	5902	5916	5929	5943	5957	5970	5984	5998	6012	1	3	4	5	7	8	10	11	12
-78	6026	6039	6053	6067	6081	6095	6109	6124	6138	6152	1	3	4	6	7	8	10	11	13
-79	6166	6180	6194	6209	6223	6237	6252	6266	6281	6295	1	3	4	6	7	9	10	11	13
-80	6310	6324	6339	6353	6368	6383	6397	6412	6427	6442	1	3	4	6	7	9	10	12	13
-81	6457	6471	6486	6501	6516	6531	6546	6561	6577	6592	2	3	5	6	8	9	11	12	14
-82	6607	6622	6637	6653	6668	6683	6699	6714	6730	6745	2	3	5	6	8	9	11	12	14
-83	6761	6776	6792	6808	6823	6839	6855	6871	6887	6902	2	3	5	6	8	9	11	13	14
-84	6918	6934	6950	6966	6982	6998	7015	7031	7047	7063	2	3	5	6	8	10	11	13	15
-85	7079	7096	7112	7129	7145	7161	7178	7194	7211	7228	2	3	5	7	8	10	12	13	15
-86	7244	7261	7278	7295	7311	7328	7345	7362	7379	7396	2	3	5	7	8	10	12	13	15
-87	7413	7430	7447	7464	7482	7499	7516	7534	7551	7568	2	3	5	7	9	10	12	14	16
-88	7586	7603	7621	7638	7656	7674	7691	7709	7727	7745	2	4	5	7	9	11	12	14	16
-89	7762	7780	7798	7816	7834	7852	7870	7889	7907	7925	2	4	5	7	9	11	13	14	16
-90	7943	7962	7980	7998	8017	8035	8054	8072	8091	8110	2	4	6	7	9	11	13	15	17
-91	8128	8147	8166	8185	8204	8222	8241	8260	8279	8299	2	4	6	8	9	11	13	15	17
-92	8318	8337	8356	8375	8395	8414	8433	8453	8472	8492	2	4	6	8	10	12	14	15	17
-93	8511	8531	8551	8570	8590	8610	8630	8650	8670	8690	2	4	6	8	10	12	14	16	18
-94	8710	8730	8750	8770	8790	8810	8831	8851	8872	8892	2	4	6	8	10	12	14	16	18
-95	8913	8933	8954	8974	8995	9016	9036	9057	9078	9099	2	4	6	8	10	12	15	17	19
-96	9120	9141	9162	9183	9204	9226	9247	9268	9290	9311	2	4	6	8	11	13	15	17	19
-97	9333	9354	9376	9397	9419	9441	9462	9484	9506	9528	2	4	7	9	11	13	15	17	20
-98	9550	9572	9594	9616	9638	9661	9683	9705	9727	9750	2	4	7	9	11	13	16	18	20
-99	9772	9795	9817	9840	9863	9886	9908	9931	9954	9977	2	5	7	9	11	14	16	18	20

FACULTY OF COMMERCE
B.Com (CBCS) III- Semester Examination, December-2017
(Common for General / Computers / Computer Applications / Advertising /
Foreign Trade/Tax Procedure Courses)
BUSINESS STATISTICS -I

Time: 3 Hours

Max. Marks: 80

PART - A (5 × 4 = 20 Marks)

Note : Answer any FIVE of the following question

1. Define the term 'Statistics' in plural sense. (Unit-I, SQA-3)
2. Limitations of statistics. (Unit-I, SQA-4)
3. Functions of Diagrams. (Unit-II, SQA-2)
4. The mean of the following series is Rs.115.86. Find the missing figure (Unit-III, Prob. 7)

House Rent (RS)	110	112	113	117	?	125	128	130
No. of Houses	25	17	13	15	14	8	6	2

5. Calculate the Harmonic mean from the following data

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

(Unit-III, Prob. 20)

6. Find out the value of quartile deviation from the following data (Unit-IV, Prob. 5)

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

7. 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.

(Unit-IV, Prob. 21)

8. Types of correlation (Unit-V, SQA-2)

PART - B (5 × 12 = 60 Marks)

(Essay Answer Type)

Note : Answer the following questions not exceeding Four Pages each.

9. (a) Define statistics and explain its scope and importance. (Unit-I, Q.No. 2, 4, 5)

OR

- (b) Define sampling. Discuss various sampling techniques. (Unit-I, Q.No. 13, 14)

10. (a) The following table shows the results of B.Com. students of a university for the last three years. Represent the same by a multiple bar diagram.

Exam Result of B.Com Students				
Year	I Division	II Division	III Division	Failed
2014	50	150	250	150
2015	60	200	300	140
2016	50	250	350	150

(Unit-II, Prob. 1)

OR

- (b) The following table shows the marks obtained by students in a class

Marks No. of	10 – 20	20 – 30	30 – 40	40 – 50	50 – 60	60 – 70	70 – 80
Students	3	10	14	24	17	14	3

Ans : Incomplete Information

11. (a) Calculate median from the following data

(Unit-III, Prob. 35)

Marks above	20	30	40	50	60	70	80	90
Frequency	200	195	177	157	142	124	84	16

OR

- (b) 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

(Unit-III, Prob. 26)

12. (a) 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

(Unit-IV, Prob. 10)

OR

- (b) A firm which was carrying on business from 1st January 2015 gets itself incorporated as a company on 1st May 2015, the first accounts are drawn up

C.I.	0-10	10-20	20-30	30-40	40-50	50-60	60-70	70-80	80-90	90-100
Frequency :	3	5	7	10	12	15	12	6	2	8

Ans : Incomplete Information

13. (a) Calculate Karl Pearson's co-efficient of correlation from the following data (Unit-V, Prob. 2)

x	3	4	6	7	10
f	9	11	14	15	16

OR

- (b) From the following data calculate rank correlation (Unit-V, Prob. 13)

x	20	25	60	45	80	25	55	65	25	75
f	45	50	55	50	60	70	72	78	80	63

FACULTY OF COMMERCE

B.Com (CBCS) III- Semester Examination, May / June - 2018
(Common for General / Computers / Computer Applications / Advertising /
Foreign Trade/Tax Procedure Courses)

BUSINESS STATISTICS -I

Time: 3 Hours

Max. Marks: 80

PART – A (5 x 4 = 20 Marks)**(Short Answer Type)****Note:** Answer any FIVE of the following questions

1. Explain the functions of statistics. (Unit-I, SQA-1)
2. Distrust of Statistics (Unit-I, SQA-2)
3. Distinguish between diagrams and graphs. (Unit-II, SQA-1)
4. Find the combined mean from the following data.

Group	A	B	C
Number	100	125	150
Mean	25	10	15

(Unit-III, Prob. 8)

5. Find the geometric mean from the following data :

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

(Unit-III, Prob. 14)

6. Find the range and its co-efficient 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

(Unit-IV, Prob. 3)

7. 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. (Unit-IV, Prob. 22)
8. Explain briefly about probable error. (Unit-V, SQA-1)

PART – B (5 x 12 = 60 Marks)**Note:** Answer the following questions not exceeding FOUR pages each.

9. (a) Define statistics in both plural and singular sense. (Unit-I, Q.No. 2, 3)
Explain its characteristics.

OR

(b) Explain the various methods of collecting primary data.

(Unit-I, Q.No. 16)

10. a) Present the following data in a sub-divided bar diagram.

Year	Faculty		
	Science	Arts	Commerce
2015	240	560	220
2016	280	610	280

(Unit-II, Prob. 2)

OR

b) From the following data draw 'less than' and 'more than' ogives.

Marks	0-10	10-20	20-30	30-40	40-50	50-60	60-70	70-80	80-90	90-100
No. of Students	5	4	8	12	16	15	10	8	5	2

(Unit-II, Prob. 11)

11. a) Calculate Arithmetic mean from the following data.

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

(Unit-III, Prob. 5)

OR

b) Calculate geometric mean for the following data :

1111, 111, 11, 1.1, 0.11, 0.011

(Unit-III, Prob. 12)

12. a) From the following data, which of the two makes shows more consistency

Life (No. of Years)	0-5	5-10	10-15	15-20	20-25
Make x No. of cars	8	12	17	10	3
Make y No. of cars	6	10	20	12	2

(Unit-IV, Prob. 18)

OR

b) Calculate the co-efficient 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

(Unit-IV, Prob. 29)

13. a) Find the co-efficient of correlation from the following data :

x	46	54	56	56	58	60	62
f	36	40	44	54	42	58	54

(Unit-V, Prob. 1)

OR

- b) Calculate Rank correlation co-efficient from the following data :

	A	B	C	D	E
x :	1	2	3	5	4
f :	2	1	4	3	5

(Unit-V, Prob. 12)

FACULTY OF COMMERCE

B.Com (CBCS) III- Semester Examination, November / December - 2018
(Common for General / Computers / Computer Applications / Advertising /
Foreign Trade/Tax Procedure Courses)
BUSINESS STATISTICS -I

Time : 3 Hours]

[Max. Marks : 80

PART – A (5 × 4 = 20 Marks)**Note :** Answer any FIVE of the following questionANSWERS

1. Limitations of Statistics (Unit - I, SQA-4)
2. Primary Data vs Secondary Data (Unit - I, SQA-5)
3. What are Two Dimensional Diagram? (Unit - II, SQA-3)
4. If the Mode and Mean of a moderately skewed series are
30.2 and 20.4 respectively, what would be its Median? (Unit - III, Prob. 41)
5. Compute Coefficient of Mean Deviation for the following data by
taking deviations from Mean.
2, 6, 11, 14, 18, 19, 23 (Unit - IV, Prob. 8)
6. If the Coefficient of Correlation between two variables X and Y is 0.86, the
Covariance is 36 and the Standard Deviation of X is 4, find the Standard
Deviation of Y. (Unit - V, Prob. 6)
7. Calculate the Harmonic Mean from the following data.

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

(Unit - III, Prob. 21)
8. What do you mean by Correlation? (Unit - V, SQA-3)

PART – B (5 × 12 = 60 Marks)**(Essay Answer Type)****Note :** Answer the following questions not exceeding Four Pages each.

9. (a) Define Statistics in both Plural and Singular Sense and explain
its importance. (Unit - I, Q.No. 2, 4)
- OR**
- (b) What is Statistical Investigation? What are the points to be
considered while planning a Statistical Investigation? (Unit - I, Q.No. 9, 10)
 10. (a) How is Percentage Bar Diagram different from Multiple Bar Diagram? (Unit - II, Q.No. 9)
- OR**

- (b) From the following data construct O'give Curves and locate Median.

Profit in Lakhs Rs.	0 – 10	10 – 20	20 – 30	30 – 40	40 – 50	50 – 60	60 – 70	70 – 80
No. of Companies	3	8	12	16	11	10	7	2

(Unit - II, Prob. 12)

11. (a) Compute Median from the following data.

Mid Classes	115	125	135	145	155	165	175	185	195
Frequency	6	25	48	72	116	60	38	22	3

(Unit - III, Prob. 36)

OR

- (b) Find out the Mode for the data given below

Size	0 – 5	5 – 10	10 – 15	15 – 20	20 – 25	25 – 30	30 – 35
Frequency	2	4	20	8	20	18	64

(Unit - III, Prob. 29)

12. (a) 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

(Unit - IV, Prob. 15)

OR

- (b) Calculate Coefficient of Skewness based on Quartiles 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

(Unit - IV, Prob. 30)

13. (a) Find Kariperson's Coefficient of Correlation for the following data:

X	65	66	67	67	68	69	70	72
Y	67	68	65	68	72	72	69	71

(Unit - V, Prob. 3)

OR

- (b) Compute Spearman's Rank Correlation Coefficient for the following data and comment on the value

Price of Tea (Rs.)	75	88	95	70	60	80	81	50
Price of coffee (Rs.)	120	134	150	55	110	140	142	100

(Unit - V, Prob. 10)

FACULTIES OF COMMERCE

B.Com. (CBCS) III - Semester Examination, June / July - 2019

(Common Paper for General / Computers / Computer Applications / Advertising / Foreign Trade / and Tax Procedure Courses)

BUSINESS STATISTICS -I

Time : 3 Hours]

[Max. Marks : 80

PART - A (5 × 4 = 20 Marks)**Note :** Answer any FIVE of the following question**[Short Answer Type]**

1. Distrust of Statistics (Unit - I, SQA-2)
2. Steps in Statistical Investigation (Unit - I, SQA-6)
3. False Base Line (Unit - II, SQA-4)
4. The mean weight of 120 students of B.Com Class is 60 Kgs. The mean weight of boys is 55 Kgs and that of girls is 45 Kgs. Find the number of boys and girls. (Unit - III, Prob. 9)
5. 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. (Unit - IV, Prob. 31)
6. Explain the Positive and Negative Correlation with suitable examples. (Unit - V, SQA-2)
7. 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 ? (Unit - III, Prob. 18)
8. Find Coefficient of Concurrent Deviation given that $n = 5$ and number of concurrent deviations (i.e. C) are 5%. (Unit - V, Prob. 16)

PART - B (5 × 12 = 60 Marks)**(Essay Answer Type)****Note :** Answer the following questions not exceeding Four Pages each.

9. (a) Define Statistics in Singular Sense. Explain its functions and list out its limitations. (Unit - I, Q.No. 2, 6, 7)
- OR**
- (b) What do you mean by Mailed Questionnaire Method ? Explain the points to be kept in mind while drafting a good questionnaire. (Unit - I, Q.No. 17)
 10. (a) Distinguish between One dimensional and Two dimensional diagrams and represent the following data related to the monthly expenditure of a family through a Pie-Diagram.

Item	Food	Clothing	Rent	Education	Others
Monthly Expenditure in Rs.	4,800	4,000	7,200	3,600	2,400

(Unit - II, Q.No. 13, Prob. 6)

OR

- (b) How is a Histogram different from a Bar Diagram? Prepare a Histogram for the following data.

Sales	0 – 10	10 – 20	20 – 30	30 – 40	40 – 50	50 – 60
No. of Co.s	7	32	56	49	42	14

(Unit - II, Q.No. 27, Prob. 13)

11. (a) Calculate Median from the following data.

Marks	0 – 9	10 – 19	20 – 29	30 – 39	40 – 49	50 – 59	60 – 69
No. of Students	6	10	20	25	16	12	5

(Unit - III, Prob. 37)

OR

- (b) Calculate Mode from the following data.

Wt. in Lbs	100 – 110	110 – 120	120 – 130	130 – 140	140 – 150	150 – 160	160 – 170	170 – 180
No. of Persons	4	6	20	32	33	17	8	2

(Unit - III, Prob. 27)

12. (a) From the marks of two students given below. Find out who is more consistent.

A	25	50	45	30	70	42	36	48	34	60
B	10	70	50	20	95	55	42	60	48	80

(Unit - IV, Prob. 16)

OR

- (b) From the following data calculate Kari 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

(Unit - IV, Prob. 23)

13. (a) Calculate Karl Pearson's Coefficient of Correlation for the following data.

X	7	6	5	4	3	2	1
Y	18	16	14	12	10	6	8

(Unit - V, Prob. 4)

OR

- (b) From the ranks of 10 student's in Accountancy and Statistics given below, calculate Rank Correlation Coefficient and Comment on the value.

Ranks in Accountancy	1	2	3	4	5	6	7	8	9	10
Rank in Statistics	1	3	5	6	7	4	8	10	9	2

(Unit - V, Prob. 14)

FACULTY OF COMMERCE
B.Com II Year III-Semester(CBCS) Examination
(Common for General / Computer Applications / Advertising & Sales Management/
Business Analytics / Honours / Foreign Trade / Tax Procedure Courses)
Novem / December - 2019
BUSINESS STATISTICS - I

Time : 2 Hours]

[Max. Marks : 80

PART - A - (4 × 5 = 20 Marks)

ANSWERS

Note: Answer any **Five** questions in not exceeding 20 lines each

1. Distrust of Statistics (Unit-I, SQA-2)
2. Parts of a Table (Unit-I, SQA-29)
3. False Base Line (Unit-II, SQA-4)
4. Represent the following data with the help of Percentage Bar Diagram

Item of the Expenditure	Expenditure in (₹)
Food	5,000
Rent	10,000
Children's Education	4,000
Clothing	8,000
Medicines	2,000
Others	6,000

Sol :

Items of the Expenditure	Amount	Expenditure (%)	Cumulative (%)
Food	5,000	14.29%	14.29%
Rent	10,000	28.57%	42.86%
Children Education	4,000	11.43%	54.29%
Clothing	8,000	22.86%	77.15%
Medicines	2,000	5.71%	82.86%
Others	6,000	17.14%	100%
	35,000		

5. State whether the data is right or wrong along with the reasons if any

AM = 4.5, GM = 4.192 and HM = 4.02

Sol :

AM = 4.5, GM = 4.192 and HM = 4.02

Since AM > GM > HM

4.5 > 4.192 > 4.02

The data is right

6. Calculate the Geometric Mean for the data given below (Unit-III, Prob.12)

1111, 111, 11, 1.1, 0.11, and 0.011

7. Two workers on the same job show the following results.

Detail	Worker X	Worker Y
Mean time of completing the job (Minutes)	30	25
Standard Deviation (Minutes)	6	4

- (i) Which worker is faster in completing the job?
 (ii) Which worker is more consistent in time he requires in completing the job.

Sol :

Coefficient of variation

$$= \frac{\text{Standard deviation}}{\text{Mean}} \times 100$$

$$X = \frac{6}{30} \times 100 = 20\%$$

$$Y = \frac{4}{25} \times 100 = 16\%$$

- (i) Coefficient of worker X is greater than worker Y

So 'X' appears to be faster in completing the job

- (ii) Worker 'Y' must be consistent

8. A research student calculated the value of R as + 0.7 with 5 observations and concludes that the correlation is highly significant. Is he right? Give your explanation with proper workings.

Sol :

$$\begin{aligned} \text{Probable Error} &= 0.6745 \times \frac{1-r^2}{\sqrt{N}} \\ &= 0.6745 \times \frac{1-(0.7)^2}{\sqrt{5}} \end{aligned}$$

$$= 0.6745 \times \frac{1 - 0.49}{2.236}$$

$$= 0.6745 \times \frac{0.51}{2.236}$$

$$= 0.154$$

$$6 \times \text{Probable Error}$$

$$6 \times 0.154 = 0.924$$

PART - B - (5 × 12 = 60 Marks)

Note: Answer following questions in not exceeding **Four** pages each using internal choice

9. (a) Present the following information in a suitable tabular form, supplying the figures not directly given.

In 2015 out of total 3000 workers in a factory 2325 were members of a Trade Union. The number of women workers employed was 375 out of which 300 did not belong to any Trade union.

in 2018 the number of union workers was 2585 of which 2400 were men. The number of non-union workers were 570 among which 230 were women.

Sol :

Genderwise Distribution of Membership of Trade Union

Trade Union	2015			2018		
	Males	Females	Total	Males	Females	Total
Members	2400	185	2585	2250	75	2325
Non-Members	340	230	570	375	300	675
Total	2740	415	3155	2625	375	3000

OR

- (b) Explain the steps involved in a Statistical Investigation. (Unit-I, Q.No.10)
10. (a) The following data shows the expenditure of two families on various heads during 2018.

Expenditure Head	Expenditure (in ₹)	
	Big Family	Small family
Rent	9,600	5,400
Electricity, telephone and Gas Bills	6,400	3,600
Education	16,000	9,000
Conveyance	6,400	1,800
Food and Provisions	22,400	14,700
Functions and Miscellaneous	3,200	1,500
Total	64,000	36,000

Draw the Pie diagrams for the data given above.

Sol :

S.No.	Expenditure head	Expenditure	Degrees
1.	Rent	9,600	$\frac{9600}{64000} \times 360 = 54^\circ$
2.	Electricity	6,400	36°
3.	Education	16,000	90°
4.	Conveyance	6,400	36°
5.	Food and Provisions	22,400	126°
6.	Funds and miscellaneous	3,200	18°
		64,000	360°

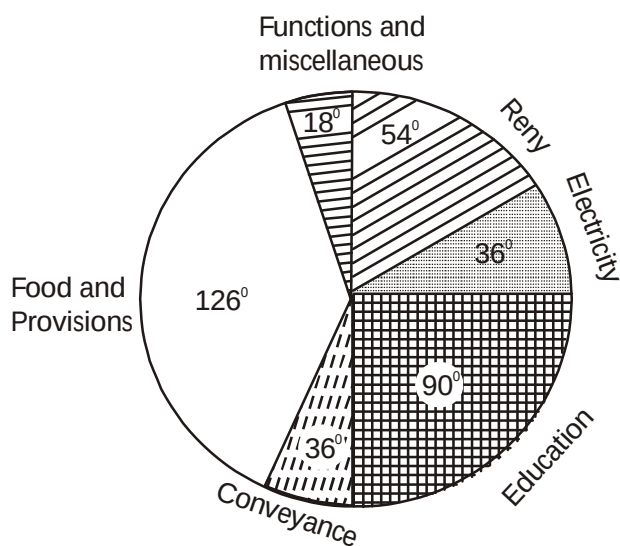


Fig.: Big Family

S.No.	Expenditure head	Expenditure	Degrees
1.	Rent	5,400	$\frac{5400}{36000} \times 360 = 54^\circ$
2.	Electricity	3,600	36°
3.	Education	9,000	90°
4.	Conveyance	1,800	18°
5.	Food and Provisions	14,700	147°
6.	Functions and miscellaneous	1,500	15°
		36,000	360°

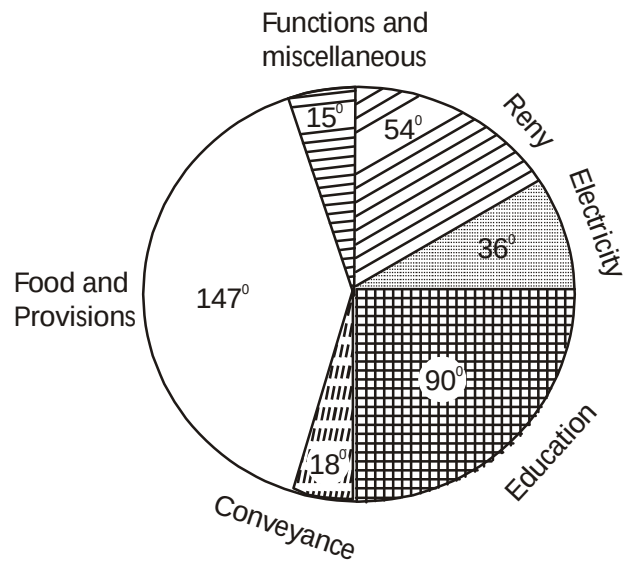


Fig.: Small Family

OR

(b) Draw ogive curves from the following data.

Marks	0 – 10	10 – 20	20 – 30	30 – 40	40 – 50	50 – 60
No. Student	10	20	30	50	40	30

(Refer, July-21, Q.No.12)

11. (a) From the following information Calculate median and Quartiles.

Marks	Below 10	Below 20	Below 30	Below 40	Below 50	Below 60	Below 70	Below 80
No. of Workers	4	10	24	50	70	85	95	100

Sol :

Marks	Frequency	Cumulative frequency
0 – 10	4	4
10 – 20	6	10
20 – 30	14	24 F
30 – 40	26 f	50 (Median) (Q_1)
40 – 50	20	70 F
50 – 60	15 f	85 (Q_3)
60 – 70	10	95
70 – 80	5	100
	100	

$$\text{Median} = \frac{N}{2} = \frac{100}{2} = 50^{\text{th}} \text{ observation}$$

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

$$30 + \frac{50 - 24}{26} \times 10$$

$$30 + \frac{26}{26} \times 10$$

$$30 + 10 \\ = 40$$

Calculation of lower Quartile (Q_1)

$$Q_1 = \frac{N}{4}$$

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

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

$$30 + \frac{25 - 24}{26} \times 10$$

$$30 + \frac{1}{26} \times 10$$

$$30 + \frac{10}{26} = 30 + 0.385 \\ = 30.385$$

Calculation of upper Quartile (Q_3)

$$Q_3 + \frac{3N}{4} = \frac{3 \times 100}{4} = 75^{\text{th}} \text{ observation}$$

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

$$50 + \frac{75 - 70}{15} \times 10$$

$$50 + \frac{5}{15} \times 10$$

$$50 + \frac{50}{15}$$

$$50 + 3.33 = 53.33$$

OR

(b) Computer Mean using Direct Method and Step Deviation Method and Verify the answers.

Wages	0 – 100	100 – 200	200 – 300	300 – 400	400 – 500	500 – 600
Frequency	14	20	42	24	20	10

Sol :

Calculation of Arithmetic Mean using Direct Method

Wages	Frequency(f)	Mid Point(M)	Frequency(M)
0 – 100	14	50	700
100 – 200	20	150	3,000
200 – 300	42	250	10,500
300 – 400	24	350	8,400
400 – 500	20	450	9,000
500 – 600	10	550	5,500
	130		37,100

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

$$= \frac{37,100}{130}$$

$$= 285.3$$

Calculation of Arithmetic mean by step deviation method

Wages	Frequency	Midpoint	d	fd
0 – 100	14	50	– 2	– 28
100 – 200	20	150	– 1	– 20
200 – 300	42	250 A	0	0
300 – 400	24	350	1	24
400 – 500	20	450	2	40
500 – 600	10	550	3	30
	130			46

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

$$250 + \frac{46}{130} \times 100$$

$$250 + 35.3$$

$$= 285.385$$

12. (a) Find Bowley's coefficient of Skewness

Wages more than in(₹)	5	15	25	35	45	55	65
No. of workers	100	80	75	60	55	20	0

Sol :

Marks	Frequency	Cumulative frequency
5 – 15	20(100 – 80)	50 F
15 – 25 (L)	5 (80 – 75) f	25 (Q ₁)
25 – 35	15 (75 – 60)	40
35 – 45	5 (60 – 55)	45 f
45 – 55 (L)	35 (55 – 20) f	80 (Median), Q ₃
55 – 65	20 (20 – 0)	100

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

Calculation of Median

$$= \frac{N}{2} = \frac{100}{2} = 50^{\text{th}} \text{ observation}$$

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

$$45 + \frac{50 - 45}{35} \times 10$$

$$45 + \frac{5 \times 10}{35}$$

$$45 + \frac{50}{35}$$

$$45 + 1.429 = 46.429$$

Calculation of Q₁

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

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

$$15 + \frac{25 - 20}{5} \times 10$$

$$15 + \frac{5}{5} \times 10$$

$$15 + \frac{50}{5} = 15 + 10 = 25$$

Calculation of Q_3

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

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

$$45 + \frac{75 - 45}{35} \times 10$$

$$45 + \frac{30}{35} \times 10$$

$$45 + \frac{300}{35}$$

$$45 + 8.57 = 53.57$$

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

$$= \frac{53.57 + 25 - 2(46.42)}{53.57 - 25}$$

$$= \frac{78.57 - 92.858}{28.57}$$

$$= -0.5001$$

OR

- (b) From the following data, calculate the Raw Moments and convert them into Central moments,

X	0 – 10	10 – 20	20 – 30	30 – 40
F	1	3	4	2

Sol :

Calculation of Raw moments

X	F	Mid Point	$d = \frac{M - 25}{10}$	d^2	fd	fd^2	fd^3	fd^4
0 – 10	1	5	-2	4	-2	4	-8	16
10 – 20	3	15	-1	1	-3	3	-3	3
20 – 30	4	25	0	0	0	0	0	0
30 – 40	2	35	1	0	2	2	2	2
					-3	9	-9	21

$$u_7^1 = \frac{\sum fd}{N} \times C = \frac{-3}{10} \times 10 = -3$$

$$u_2^1 = \frac{\sum fd^2}{N} \times C^2 = \frac{9}{10} \times 100 = 90$$

$$u_3^1 = \frac{\sum fd^3}{N} \times C^3 = \frac{-9}{10} \times 1000 = -900$$

$$u_4^1 = \frac{\sum fd^4}{N} \times C^4 = \frac{21}{10} \times 10,000 = 21,000$$

13. (a) The marks obtained by the students are tabulated below,

Marks in Science	19	25	14	9	25	14	22	15	25
Marks mathematics	30	30	11	22	18	22	14	15	11

Find the rank Correlation Coefficient using repeated ranks formula.

Sol :

Calculation of Rank Correlation

Marks in Science	R_1	Marks in Mathematics	R_2	$R_1 - R_2 = d$	d^2
19	5	30	8.5	-3.5	12.25
25	8	30	8.5	-0.5	0.25
14	2.5	11	1.5	1	1
9	1	22	6.5	-5.5	30.25
25	8	18	5	3	9
14	2.5	22	6.5	-4.5	20.25
22	6	14	3	3	9
15	4	15	4	0	0
25	8	11	1.5	6.5	42.25
					124.25

$$\begin{aligned}
 r_s &= \frac{1 - 6(\sum D^2 + \frac{1}{12}(m^3 - m) + \frac{1}{12}(m^3 - m) + \frac{1}{12}(m^3 - m) + \frac{1}{12}(m^3 - m) + \frac{1}{12}(m^3 - m))}{N^3 - N} \\
 &= \frac{1 - 6(124.25 + \frac{1}{12}(3^3 - 3) + \frac{1}{12}(2^3 - 2) + \frac{1}{12}(2^3 - 2) + \frac{1}{12}(2^3 - 2) + \frac{1}{12}(2^3 - 2))}{9^3 - 9} \\
 &= \frac{1 - 6(124.25 + 2 + 0.5 + 0.5 + 0.5 + 0.5)}{729 - 9} \\
 &= \frac{1 - 6(124.25 + 4)}{720} \\
 &= \frac{1 - 6(128.25)}{720} \\
 &= \frac{1 - 769.5}{720} \\
 &= 1 - 1.06875 = -0.068
 \end{aligned}$$

OR

- (b) Calculate Kari Pearson's Coefficient of Correlation between Advertisement Expenditure and Sales of a firm.

Advertisement Expenses('000 `)	30	56	53	81	73	66	26	89	27	69
Sales (Lakhs `)	38	44	49	77	53	59	51	82	42	75

Sol :

Calculation of Coefficient of Correlation

Advertisement expenses (X)	(X - $\bar{X}$) (x)	(Y)	(Y - $\bar{Y}$) y	x ²	y ²	xy
30	-27	38	-19	729	361	+513
56	-1	44	-13	1	169	+13
53	-4	49	-8	16	64	+32
81	24	77	20	576	400	+480
73	16	53	-4	256	16	-64
66	9	59	+2	81	4	18
26	-31	51	-6	961	36	186
89	32	82	+25	1024	625	800
27	-30	42	-15	900	225	450
69	12	75	+18	144	324	216
570		570		4688	2224	2644

$$\bar{X} = \frac{\Sigma X}{N} = \frac{570}{10} = 57$$

$$\bar{Y} = \frac{\Sigma Y}{N} = \frac{570}{10} = 57$$

$$r = \frac{\Sigma xy}{\sqrt{\Sigma x^2 \times \Sigma y^2}} = \frac{2644}{\sqrt{4688 \times 2224}} = \frac{2644}{\sqrt{10426112}}$$

$$\Rightarrow \frac{2644}{3288.9} = 0.80$$

FACULTY OF COMMERCE
B.Com II Year III-Semester(CBCS) Examination
(Common for General / Computer Applications / Advertising & Sales Management/
Business Analytics / Honours / Foreign Trade / Tax Procedure Courses)
July - 2021
BUSINESS STATISTICS - I

Time : 2 Hours]

[Max. Marks : 80

PART - A - (4 × 5 = 20 Marks)

ANSWERS

Note : Answer the following any **Four** Question.

- | | |
|---|--------------------|
| 1. Limitations of statistics | (Unit-I, Q.No.7) |
| 2. Diagrams Vs Graphs | (Unit-II, SQA-1) |
| 3. Types of Averages | (Unit-III, Q.No.6) |
| 4. Calculate 8 th Decline (D ₁) of the following data. | |
| 22, 26, 14, 30 18, 11, 35, 41, 12, 32 | |

Sol :

Arrange the observation in ascending order

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

Calculation of D₈

$$D_4 = \text{Size of } \left[\frac{8(N+1)}{10} \right]^{\text{th}} \text{ term}$$

$$= 8 \left(\frac{10+1}{10} \right)^{\text{th}} \text{ term}$$

$$= 8 \left(\frac{11}{10} \right)^{\text{th}} \text{ term}$$

$$= 8(1.1)^{\text{th}} \text{ term}$$

$$\text{Size of } (8.8)^{\text{th}} \text{ term} = 8^{\text{th}} \text{ term} + 0.8(9^{\text{th}} \text{ term} - 8^{\text{th}} \text{ term})$$

$$= 32 + 0.8(35 - 32)$$

$$= 32 + 0.8 \times 3$$

$$= 32 + 2.4$$

$$= 34.4$$

$$D_8 = 24.4$$

5. Calculate Mean Deviation and its coefficient. From Arithmetic mean.

7, 5, 12, 10, 8, 11, 16

Sol :

Calculation of Mean Deviation

x	d = x - $\bar{X}$ $\bar{X} = 9.86$
7	2.86
5	4.86
12	2.14
10	0.14
8	1.86
11	1.14
16	6.14
$\Sigma X = 69$	$\Sigma D = 19.22$

Mean

$$\bar{X} = \frac{\Sigma X}{N} = \frac{69}{7} = 9.857 \approx 9.86$$

Mean Deviation form Mean

$$\text{Mean Deviation} = \frac{\Sigma |D|}{N} = \frac{19.44}{7}$$

$$\therefore \text{M.D} = 2.734$$

Coefficient of Mean Deviation

$$\text{Coefficient of Mean Deviation} = \frac{\text{Mean Deviation}}{\text{Mean}}$$

$$= \frac{2.734}{9.86} = 0.277$$

$$\therefore \text{Coefficient of Mean Deviation} = 0.277$$

6. Calculate the Value of standard Deviation

Mean = 45 Medium = 48 Karl person's co-efficient of Skewness = - 0.4

Sol :

Mean = 45

Medium = 48

Karl person co-efficient of Skewness = - 0.4

$$\text{Karl person co-efficient of Skewness } [S_{kp}] = \frac{3(\text{Mean} - \text{Median})}{\sigma}$$

$$\text{Substituting the given values in } S_{kp} \text{ formula, } -1.4 = \frac{3(45 - 48)}{\sigma}, -0.4 = \frac{3(-3)}{\sigma}, -0.4 = \frac{-9}{\sigma}$$

$$\sigma = \frac{-9}{-0.4}$$

$$\sigma = 22.5$$

$$\therefore \text{Standard Deviation} = 22.5$$

7. For a moderately skewed distribution Mean = 160 Mode = 157 and standard Deviation = 40 Find (i) Co-efficient of variation (ii) person's Co-efficient of Skewness.

Sol :

(i) Co-efficient of Variation

$$\begin{aligned} \text{Co-efficient of Variation, C.V} &= \frac{\sigma}{X} \times 100 \\ &= \frac{40}{160} \times 100 \\ \therefore &= 25 \end{aligned}$$

(ii) Person's Co-efficient of Skewness

$$\begin{aligned} \text{Person's Co-efficient of Skewness (SK}_p\text{)} &= \frac{\text{Mean} - \text{Mode}}{\text{Standard deviation}} \\ &= \frac{160 - 157}{40} \\ \therefore \text{SK}_p &= 0.075 \end{aligned}$$

8. Calculate Co-efficient of concurrent Deviation from the following data.

X	59	69	39	49	29
Y	79	69	59	49	39

Sol :

For calculating co-efficient of correlation by concurrent deviation method, the following formula is used,

$$r_c = \pm \sqrt{\frac{\pm 2c - n}{n}}$$

Where,

r_c = Coefficient of correlation

C = Number of positive signs after multiplying d_x with d_y

$n = N - 1$

Calculation of Coefficient of Correlation by Concurrent Deviation Method

X	Y	d_x	d_y	$d_x d_y$
59	79			
69	69	+	-	-
39	59	-	-	+
49	49	+	-	-
29	39	-	-	+
				C = 2

$$r_c = \pm \sqrt{\frac{\pm 2c - n}{n}} = \pm \sqrt{\frac{(2 \times 2)(5 - 1)}{(5 - 1)}}$$

$$= \pm \sqrt{\frac{4 - 4}{4}} = \pm \sqrt{\frac{0}{4}}$$

$$\therefore r_c = 0$$

PART - B - (4 × 15 = 60 Marks)

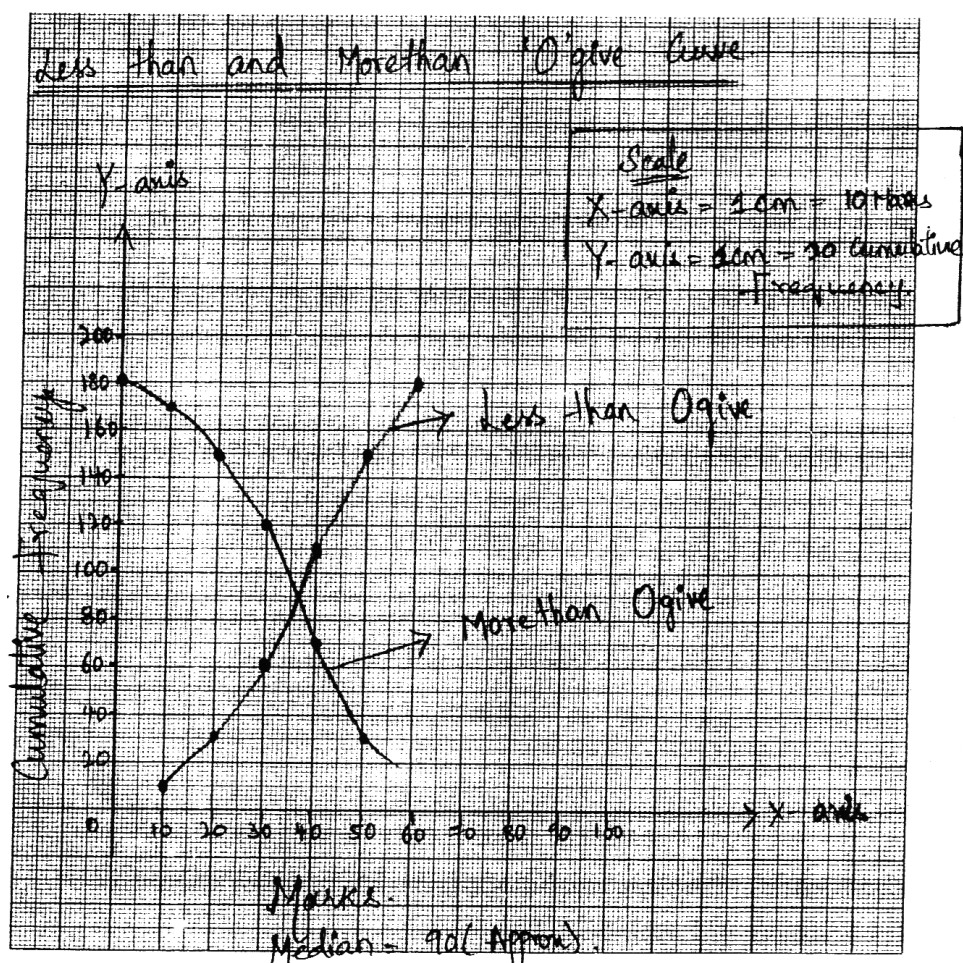
Note : Answer the following any **Four** Question.

- What do you mean by Primary Data ? Explain the various method of Collecting Primary Data. (Unit-I, Q.No.15,16)
- Define the term statistics. Explain the various characteristics of statistics. (Unit-I, Q.No.2,3)
- What is diagrammatic presentation of data? Explain the importance and limitations of graphical presentation of data. (Unit-II, Q.No.1,4,16,18)
- Draw ogive curves from the following data.

Marks	0 – 10	10 – 20	20 – 30	30 – 40	40 – 50	50 – 60
No. Student	10	20	30	50	40	30

Sol :

Marks	No. of Students	Less than Cumulative Frequency	More than Cumulative Frequency
0 – 10	10	10	180
10 – 20	20	(10 + 20) 30	(180 – 10) 170
20 – 30	30	(30 + 30) 60	(170 – 20) 150
30 – 40	50	(60 + 50) 110	(150 – 30) 120
40 – 50	40	(110 + 40) 150	(120 – 50) 70
50 – 60	30	(150 + 30) 180	(70 – 40) 30



13. Calculate Medium from the following data.

Midvalue	115	125	135	145	155	165	175	185	195
Frequency	6	25	48	72	116	60	38	22	3

Sol :

	Class Intervals	Mid Values	Frequency (f)	Cumulative Frequency(c.f)
	110 – 120	115	6	6
	120 – 130	125	25	31
	130 – 140	135	48	79
	140 – 150	145	72	151 F
L	150 – 160	155	116 f	267
	160 – 170	165	60	327
	170 – 180	175	38	365
	180 – 190	185	22	387
	190 – 200	195	3	390
			N = 390	

Median = Size of $\frac{N^{\text{th}}}{2}$ observation

$$= \frac{390}{2} = 195^{\text{th}} \text{ observation}$$

Median lies in the class 150 – 160

$$\begin{aligned} \therefore \text{Median} &= L + \frac{\frac{N}{2} - F}{f} \times C \\ &= 150 + \frac{195 - 151}{116} \times 10 \\ &= 150 + 3.793 \quad \therefore \text{Median} = 153.793 \end{aligned}$$

14. Find out the Mode from the following data.

Size	0 – 5	5 – 10	10 – 15	15 – 20	20 – 25	25 – 30	30 – 35
Frequency	1	2	10	4	10	9	2

Sol :

	C.I	Frequency
	0.5	1
	5 – 10	2
	10 – 15	10
	15 – 20	4 f_1
L	20 – 25	10 f
	25 – 30	9 f_2
	30 – 35	2

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

$$\begin{aligned}\Delta_1 &= f - f_1 \\ &= 10 - 4 = 6\end{aligned}$$

$$\begin{aligned}\Delta_2 &= f - f_2 \\ &= 10 - 9 = 1\end{aligned}$$

$$20 + \frac{6 \times 5}{6 + 1}$$

$$20 + \frac{30}{7}$$

$$\begin{aligned}20 + 4.28 \\ = 24.28\end{aligned}$$

15. Calculation co-efficient of skewness based on quartiles from the following data.

Size	10 – 20	20 – 30	30 – 40	40 – 50	50 – 60	60 – 70	70 – 80	80 – 90
Frequency	6	10	18	30	12	10	6	2

Sol :

Bowley's Co-efficient of Skewness is based on quartiles

Class Intervals	Frequency (f)	Cumulative Frequency (C.F)
10 – 20	6	6
20 – 30	10	16
30 – 40	18	34
40 – 50	30	64
50 – 60	12	76
60 – 70	10	86
70 – 80	6	92
80 – 90	2	94
	N = 94	

Bowley's Coefficient of skewness based on quartites,

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

Calculation of Upper Quartile (Q_3) and Lower Quartile(Q_1)

$$Q_1 = \text{Size of } \frac{N}{4} \text{th item}$$

$$N = 94$$

$$Q_1 = \frac{94}{4} = 23.5^{\text{th}} \text{ observation}$$

$\therefore Q_1$ lies in class 30 – 40

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

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

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

$$= 30 + (0.417 \times 10)$$

$$= 30 + 4.17$$

$$\therefore Q_1 = 34.17$$

$$Q_3 = \text{size of } \frac{3N}{4} \text{th item}$$

$$N = 94$$

$$Q_3 = \frac{3 \times 94}{4}$$

$$= \frac{282}{4} = 70.5^{\text{th}} \text{ observation}$$

$\therefore Q_3$ lies in class 50 – 60

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

$$Q_3 = 50 + \frac{70.5 - 64}{12} \times 10$$

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

$$= 50 + (0.542 \times 10)$$

$$= 50 + 5.42$$

$$\therefore Q_3 = 55.42$$

Calculation of Median

$$\text{Median} = \text{Size of observation } \frac{94}{2} = 47$$

Median lies in class of 40 – 50

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

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

$$= 40 + \frac{13}{30} \times 10$$

$$= 40 + (0.433 \times 10)$$

$$= 40 + 4.33$$

$$\therefore \text{Median} = 44.33$$

Substituting the Q_3 , Q_1 and median values in the SK_B formula,

we get,

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

$$SK_B = \frac{55.42 + 34.17 - 2(44.33)}{55.42 - 34.14}$$

$$= \frac{89.59 - 88.66}{21.25}$$

$$= \frac{0.93}{21.25}$$

$$\therefore SK_B = 0.0438$$

16. Find the Co-efficient of variation of a frequency distribution given that its
Mean = 100 Mode = 102, Karl person's Co-efficient of skewness is -0.4

Sol :

Given that,

$$\text{Mean} = 100$$

$$\text{Mode} = 102$$

$$Sk_p = -0.4$$

$$\text{Coefficient of Variation (C.V)} = \frac{\sigma}{\bar{X}} \times 100 \text{ (OR) } \frac{\text{Standard Deviation(SD)}}{\text{Mean}} \times 100$$

Calculating the value of standard Deviation (SD : σ) by substituting the given values in Karl Pearson's Co-efficient of Skewness formula.

$$\text{Karl Pearson's Co-efficient of skewness}(Sk_p) = \frac{\text{Mean} - \text{Mode}}{\text{S.D}}$$

$$= 0.4 = \frac{100 - 102}{\text{S.D}}$$

$$-0.4 = \frac{-2}{\text{S.D}}$$

$$-0.4 \times \text{S.D} = -2$$

$$\text{S.D} = \frac{-2}{-0.4}$$

$$\text{S.D} = 5$$

$$\therefore \text{Co-efficient of Variation (Cv)} = \frac{5}{100} \times 100 = 5$$

$$\text{Average of Y} = \bar{Y} = \frac{\Sigma Y}{N} = \frac{552}{8} = 69$$

$$\text{Karl Pearson's Coefficient of correlation (r)} = \frac{N\Sigma xdy - (\Sigma dx)(\Sigma dy)}{\sqrt{N\Sigma dx^2 - (\Sigma dx)^2} \sqrt{N\Sigma dy^2 - (\Sigma dy)^2}}$$

$$\Sigma dx = 0, \Sigma dx^2 = 36, \Sigma dy = 0, \Sigma dy^2 = 44, \Sigma dx dy = 24, N = 8$$

$$r = \frac{(8 \times 24) - (0 \times 0)}{\sqrt{8 \times 36 - (0)^2} \sqrt{8 \times 44 - (0)^2}} = \frac{192}{\sqrt{288} \sqrt{352}} = \frac{192}{16.97 \times 18.76} = \frac{192}{318.35}$$

$$\therefore r = 0.60$$

17. Calculate the Karl Pearson's Co-efficient of correlation from the following data.

X	3	4	6	7	10
Y	9	11	14	15	16

Sol :

Calculation of Karl Pearson's co-efficient of correlation by direct method without taking deviations from actual mean or assumed mean.

Calculation of Karl Pearson's Coefficient of Correlation

X	X ²	Y	Y ²	XY
3	9	9	81	27
4	16	11	121	44
6	36	14	196	84
7	49	15	225	105
10	100	16	256	160
$\Sigma X = 30$	$\Sigma X^2 = 270$	$\Sigma Y = 65$	$\Sigma Y^2 = 879$	$\Sigma XY = 420$

$$r = \frac{N\Sigma XY - (\Sigma X)(\Sigma Y)}{\sqrt{N\Sigma X^2 - (\Sigma X)^2} \sqrt{N\Sigma Y^2 - (\Sigma Y)^2}}$$

Where, N = 5

$$r = \frac{(5 \times 420) - (30 \times 65)}{\sqrt{5 \times 270 - (30)^2} \sqrt{5 \times 879 - (65)^2}} = \frac{2100 - 1950}{\sqrt{1050 - 900} \sqrt{4395 - 4225}} = \frac{150}{\sqrt{150} \sqrt{170}} = \frac{150}{12.25 \times 13.04}$$

$$= \frac{150}{159.74}$$

$$\therefore r = 0.94$$

18. Calculate Rank correlation Co-efficient from the following data.

X	99	47	68	57	76	61	30	44	82	58
Y	75	63	72	42	74	41	20	64	68	67

Sol :

Calculation of Co-efficient of Rank Coefficient

X	Y	R ₁	R ₂	d = (R ₁ - R ₂)	d ²
99	75	10	10	0	0
47	63	3	4	-1	1
68	72	7	8	-1	1
57	42	4	3	1	1
76	74	8	9	-1	1
61	41	6	2	4	16
30	20	1	1	0	0
44	64	2	5	-3	9
82	68	9	7	2	4
58	67	5	6	-1	1
					Σd ² = 34

$$R = 1 - \frac{6\Sigma d^2}{n(n^2 - 1)}$$

$$R = 1 - \frac{6 \times 34}{10(10^2 - 1)}$$

$$= 1 - \frac{204}{10(100 - 1)}$$

$$= 1 - \frac{204}{10(99)}$$

$$= 1 - \frac{204}{990}$$

$$\Rightarrow 1 - 0.206$$

$$\therefore R = 0.794$$

FACULTY OF COMMERCE
B.Com (CBCS) III- Semester Examination
Model Paper - I
(Common for General / Computer Applications / Advertising & Sales Management /
Business Analytics / Honours / Foreign Trade / Tax Procedure Courses)
BUSINESS STATISTICS-I

Time: 3 Hours

Max. Marks: 80

PART – A (5 × 4 = 20 Marks)

Note : Answer any FIVE of the following question

ANSWERS

1. What are the differences between Primary and Secondary Data ? (Unit-I, SQA-5)
2. Define distrust of statistics. (Unit-I, SQA-2)
3. What are two dimensional diagrams? (Unit-II, SQA-3)
4. Mean of 100 observations is found to be 40. If at the time of computation, two items are wrongly taken as 30 and 27 instead of 3 and 72. Find correct mean. (Unit-III, Prob. 10)
5. Calculate mean deviation of the follo-wing from mean and median
2, 6, 11, 14, 16, 19, 23 (Unit-IV, Prob. 8)
6. 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. (Unit-IV, Prob. 21)
7. What is Scatter diagra ? (Unit-V, SQA-5)
8. If the Coefficient of Correlation between two variables X and Y is 0.86, the
Covariance is 36. And the Standard Deviation of X is 4, find the Standard
Deviation of Y. (Unit-V, Prob. 6)

PART – B (5 × 12 = 60 Marks)

(Essay Answer Type)

Note : Answer the following questions not exceeding Four Pages each.

9. (a) Define the term statistics in both Plural and Singular Sense. Explain the
characteristics of statistics. (Unit-I, Q.No. 2, 3)
- OR
- (b) Define secondary data. What are the advantages and disadvan-tages of
Secondary Data? (Unit-I, Q.No. 15, 21)

10. (a) Represent the below given data by a simple bar diagram :

Year	Sales Tax (in '000)
1961	900
1966	1800
1971	2500
1976	4500
1981	6000

(Unit-II, Prob. 4)

OR

- (b) Draw a Pie diagram by showing the following of values :

Item of Expenditure	
Food	12,000
Clothing	5,000
House Rent	15,000
Education	18,000
Misc.	10,000

(Unit-II, Prob. 7)

11. (a) 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

(Unit-III, Prob. 19)

OR

- (b) From the following data, compute Median

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

(Unit-III, Prob. 33)

12. (a) Calculate Coefficient of Skewness based on Quartiles 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

(Unit-IV, Prob. 30)

OR

- (b) 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

(Unit-IV, Prob. 14)

13. (a) Calculate Karl Pearson's co-efficient of correlation from the following data

x	3	4	6	7	10
f	9	11	14	15	16

(Unit-V, Prob. 2)

OR

- (b) Compute Rank coefficient of correlation for the following data :

X	15	20	28	12	40	60	20	80
Y	40	30	50	30	20	10	30	60

(Unit-V, Prob. 11)

FACULTY OF COMMERCE
B.Com (CBCS) III- Semester Examination
Model Paper - II

(Common for General / Computer Applications / Advertising & Sales Management /
Business Analytics / Honours / Foreign Trade / Tax Procedure Courses)

BUSINESS STATISTICS-I

Time: 3 Hours

Max. Marks: 80

PART - A (5 × 4 = 20 Marks)

Note : Answer any FIVE of the following question

ANSWERS

1. Explain the various steps in Statistical Investigation. (Unit-I, SQA-1)
2. Distinguish between Diagrams and Graphs. (Unit-II, SQA-1)
3. Define Histogram. (Unit-II, SQA-11)
4. Find the geometric mean from the following data :

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

(Unit-III, Prob. 14)

5. If mode = 16, mean = 15.6 find the median. (Unit-III, Prob. 42)
6. From the data given below, calculate and Bowley's Coefficients of Skewness.
Mean = 160, Median = 152, Q1 = 72, Q3 = 205
and Standard Deviation = 40. (Unit-IV, Prob. 27)
7. 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 (Unit-IV, Prob. 1)
8. Types of Correlation. (Unit-V, SQA-2)

PART - B (5 × 12 = 60 Marks)

(Essay Answer Type)

Note : Answer the following questions not exceeding Four Pages each.

9. (a) Explain the importance of statistics in economic analysis and planning. (Unit-I, Q.No. 4)
- OR
- (b) What is frequency distribution? Explain briefly about discrete and continuous frequency distribution. (Unit-I, Q.No. 31)

10. (a) Draw a multiple bar diagram from the following data.

Year	Sales (000 `)	Gross Profit (000 `)	Net Profit (000 `)
2010	120	40	20
2011	135	45	30
2012	140	55	35
2013	150	60	40
2014	160	65	45

(Unit-II, Prob. 3)

OR

- (b) Construct Histogram for the following distribution.

Class Interval	Frequency
10 – 20	15
20 – 30	35
30 – 40	70
40 – 50	50
50 – 60	30
60 – 70	20
70 – 80	15

(Unit-II, Prob. 8)

11. (a) The mean of the following series is ` 115.86. Find the missing figures.

House Rent (`)	110	112	113	117	7	125	128	130
Number of Houses	25	27	13	15	14	8	6	2

(Unit-III, Prob. 7)

OR

- (b) Calculate mode for the following data :

Marks	No. of Students
0 – 10	2
10 – 20	5
20 – 30	11
30 – 40	14
40 – 50	17
50 – 60	13
60 – 70	6
70 – 80	2

(Unit-III, Prob. 28)

12. (a) Calculate Karl Pearson Coefficient of Skewness from the following data,

Wages	No. of Students
25 – 75	25
75 – 125	30
125 – 175	40
175 – 225	35
225 – 275	30
275 – 325	25
325 – 375	30
375 – 425	45

(Unit-IV, Prob. 24)

OR

- (b) Calculate the mean deviation and the coefficient from the following data.

Class	0 – 10	10 – 20	20 – 30	30 – 40	40 – 50	50 – 60	60 – 70	70 – 80
Frequency	5	8	12	15	20	14	12	6

(Unit-IV, Prob. 11)

13. (a) Find Karl Pearson's Coefficient of Correlation for the following data:

X	65	66	67	67	68	69	70	72
Y	67	68	65	68	72	72	69	71

(Unit-V, Prob. 3)

OR

- (b) From the following data calculate rank correlation

x	20	25	60	45	80	25	55	65	25	75
f	45	50	55	50	60	70	72	78	80	63

(Unit-V, Q.No. 13)