STATISTICS:
Volume I

Dr. Manoj K. Bhowal,
Dr. Pronob Barua

Asian Books Private Limited

STATISTICS

Volume I

Second Edition

Dr. Manoj K. Bhowal

Senior Lecturer in Statistics

J.B. College, Jorhat

Dr. Pronob Barua

Lecturer in Statistics J.B. College, Jorhat

Asian Books Private Limited

7/28, Mahavir Lane, Vardan House, Ansari Road,
Darya Ganj, New Delhi - 110002

Asian Books Private Limited

Registered and Editorial Office
7/28, Mahavir Lane, Vardan House, Ansari Road, Darya Ganj,
New Delhi-110002
E-Mail : asian@asianbookindia.com
World Wide Web : http://www.asianbooksindia.com
Phones : 23287577, 23282098, 23271887, 23259161
Fax : 91 11 23262021

Sales Offices

Bangalore	103, Swiss Complex No. 33, Race Course Road, Bangalore - 560 001 Ph. : (080) 22200438 Fax : 91 80 2256583 Email : asianblr@blr.vsnl.net.in
Chennai	Palani Murugan Building No. 21, West Cott Road, Royapettah, Chennai - 600014 Ph. : (044) 28486927, 32979058 Email : asianmds@vsnl.net
Delhi	7/28, Mahavir Lane, Vardan House, Ansari Road, Darya Ganj, New Delhi - 110002. Phones : (011) 23287577, 23282098, 23271887, 23259161, Fax : 91 11 23262021 Email : asian@asianbooksindia.com
Guwahati	6, G.N.B. Road, Panbazar Guwahati, Assam - 781 001 Ph. : (0361) 2513020, 2635729 Email : asianghy1@sancharnet.in
Hyderabad	3-5-315, St. No. 7, Vittalwodi, Narayanaguda Hyderabad - 500 029 Phones : (040) 23220112, 23220113 Email : hydasian@eth.net
Kolkata	10 A, Hospital Street, Calcutta - 700 072 Phone : (033) 22153040, 32506706, Fax : 91 33 22159899 Email : calasian@vsnl.net
Mumbai	Shop No. 3 & 4, Ground Floor Shilpin Centre 40, G.D. Ambekar Marg, Sewree Wadala Estate, Wadala, Mumbai - 400031 Phones : (022) 32037931, 24157611/12, 32458947 Email : asianbk@hathway.com
Pune	Shop No. 5-8 Ground Floor, Shaan Brahma Complex Near Ratan Theatre, Budhwar Peth, Pune-02 Phones : (020) 32304554, Fax : 91-20-32911509 Email : asianpune@asianbooksindia.com

© Authors

1st Edition 2006
2nd Edition 2008
ISBN 978-81-8412-085-1

All rights reserved. No part of this publication may be reproduced, 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 publishers.

Published by **Kamal Jagasia** for Asian Books Pvt. Ltd., 7/28, Mahavir Lane, Vardhan House, Darya Ganj, New Delhi-110002.
Typeset at Aoushi Enterprises, Delhi.
Printed at Gopaljee Enterprises, Delhi.

Preface to the second edition

This book has been thoroughly revised and a few examples taken from recent question papers have been incorporated in almost all the chapters. We hope that the book will be found more useful to all those who read it. This book contains materials as per Assam Higher Secondary Education Council. However, this will be of great help to the students of Economics, Commerce, Business Management and other related disciplines.

While revising the book, we have been greatly benefited from our students Miss Manashi Changmai and Miss Priya Mech. Asian Books Private Limited deserve warmest thanks for their interest and sincere efforts in bringing out the revised edition of the book.

Comments and suggestions from teachers and students are cordially invited for further improvement in the materials of the book in subsequent editions.

14th July, 2008
Jorhat

Authors

PREFACE

This book Statistics Vol. I written as a text for Higher Secondary first year students. We are confident that this will meet the needs of the students. Every effort has been made to make the language simple. We would be grateful to the readers for pointing out any mistakes and for sending their valuable suggestions to bring about further improvement in this book.

We would like to thank all the faculty members of the department of Statistics, J.B. College, Jorhat. We will be failing in our duties if we do not pay thanks to our students Khiromohan, Kaushik and Supahi.

We would like to thank Mrs. Ruma Bhowal and Mrs. Nilima Barua for their encouragement towards the completion of the book.

Finally, we are grateful to Asian Books Pvt. Ltd for publishing the book.

July 13, 2005 **Authors**

CONTENTS

STATISTICS

SYLLABUS FOR HIGHER SECONDARY FIRST YEAR COURSE

One Paper | Three hours | Marks 100

Unitwise Distribution of Marks and Periods :

Unit No.		**Title**	**Marks**	**Periods**
Unit-1 :	(a)	Algebra	12	20
	(b)	Calculus	12	20
Unit-2 :		**Descriptive Statistics**	40	60
Unit-3 :		**Applied Statistics :**		
	(a)	Index Number	12	10
	(b)	Vital Statistics	12	10
	(c)	Time Series	12	10
			100	130

Unitwise distribution of course contents

Unit-1 : (a) Algebra : Laws of indices. Logarithms, A, P. and G. P. Series, Permutation and combination, Binomial theorem for positive integral index, Statement and applications of Binomial theorem for any index, Exponential series and logarithmic series, Idea of sets and set operations.

(b) (i) Differential Calculus: Functions, Limit of a function. Derivatives, Rules of Differentiation of sum, Difference, product, quotient of functions and Function of a function. (Trigonometric functions are to be avoided).

(ii) Integral Calculus : Integration as the reverse of differentiation (simple cases only such as

$\int x^n dx, \int \frac{1}{x} dx, \int e^x dx, \int \frac{1}{(x^2-a^2)} dx, \int \frac{1}{(a^2-x^2)} dx,$ method of

substitution (simple examples only), Definite integrals (simple examples only).

Unit-2 : Descriptive Statistics :

Meaning of Statistics — Statistical data and Statistics subject. Origin, development, Scope and limitations of Statistics. Idea of Statistical population and sample. Different types of data-primary and secondary data and methods of their collection.

Time series data. Spatial data. Attribute (qualitative) data and Variable (quantitative) data.

Frequency distribution, Graphical representation of frequency distribution–Histogram, Frequency polygon, Frequency curve, Ogive.

Measures of location—Arithmetic mean, Geometric mean, Harmonic mean, Median, Mode and their properties.

Partition Values — Quartiles, Deciles, Percentiles. Graphical location of Mode, Quartiles, Deciles and Percentiles.

Measures of Dispersion — Range, Inter-quartile Range, Quartile deviation. Mean Deviation, Standard Deviation, Coefficient of variation. Ideal measures of Dispersion;

Idea of Skewness and Kurtosis (without moments).

Bi — variate distribution, Scatter diagram. Correlation and regression, Karl Pearson's Correlation coefficient and its properties. Two regression lines (without derivation), Relation between correlation coefficient and regression coefficients.

Unit-3 : Applied Statistics :

(a) **Index Numbers :** Idea and uses of index numbers, problems in the construction of index numbers, simple and weighted index numbers, Laspeyre's, Paasche's Marshall-Edgeworth and Fisher's 'ideal' index numbers, Tests for a good index number— Time reversal and Factor reversal tests only, Consumer price index number — their construction and use.

(b) **Vital Statistics :** Vital rates and ratios.'Mortality rates — Crude death rate, Age specific death rate and standardized death rates. Fertility rate and total fertility rate. Reproduction rates - Gross reproduction rate and Net reproduction rate.

(c) **Time Series :** Meaning, Components and uses of time series. Determination of trend by the methods of graphic, semi-averages, moving averages and least squares.

UNIT I

STATISTICS

"This page is Intentionally Left Blank"

1

INTRODUCTION

1.1 ORIGIN OF STATISTICS

Statistics was born as the Science of Kings. It had its origin in the needs of the administrators in the ancient days for collecting and maintaining quantitative information about their population and wealth with a view to make policies for citizens. The word 'Statistics' resembles with Latin word 'Status', Italian word 'Statista', German word 'Statistik' carrying the same meaning of a political state. In the early years when this branch of science was not much developed, the data were collected by the government officials for the purpose of administration.

About 3050 B.C. the first collection of data was made in Egypt regarding the population and wealth for raising the pyramids. Such collection of data later held in England, Germany and many other western countries in the middle ages. In India during the period of Chandragupta Maurya a system of collection of vital statistics was evolved and births and deaths were started to be registered. Emperor Akbar made a wider administrative use of statistical surveys in the country. The histories of almost all the countries of the world reveal that statistics was used as a vital instrument for administration of the various activities of the states for which it was regarded as the 'Science of Statecraft'.

1.2 DEVELOPMENT OF STATISTICS

Statistics is an old science. For the last few centuries, it has remained a part of mathematics as the original work was done by mathematicians like Pascal (1623-1662), J. Bernoulli (1654-1705), De Moivre (1667-1754), Laplace (1749-1827), Gauss (1777-1855) and many others. The Royal statistical society of London was founded in 1834 and the American Statistical Association in 1839.

The theory of probability and the normal distribution have been very important in the development of the subject and they are now of primary importance in the theory of Statistics. Blaise Pascal (1623-62), Fermat (1601-65), Huygens (1629-95) developed some new ideas. A first book on probability was published in 1654. Jacob Bernoulli (1654-1705), Abraham de Moivre (1667-1754) and P.S. Laplace (1749-1827) made great contributions to the early theory and application of probability. Abraham de Moivre developed the equation of the normal curve (1733). Much later, Laplace and Gauss (1777-1855) developed the same results independently of each other.

The biologist Charles Darwin (1809-82) and Mendal (1822-84) were not statisticians and they did not spend their time in placing statistics on a firm foundation although they used statistics in their work.

Karl Pearson (1857-1936), initially a mathematical physicist, after becoming interested in the evolution spent nearly half a century in serious statistical research. He helped to found the journal 'Biometrica'. Sir Ronald Fisher (1890-1962) made many important contributions to the subject. Fisher and his students applied statistics in many fields, particularly agriculture, biology and genetics. Some of the basic theory on hypothesis testing was presented by J. Neyman and E.S. Pearson as late as 1936 and 1938.

Since the last 1920's interest in the application of statistical methods to all types of problems has grown rapidly.

1.3 MEANING OF STATISTICS

The word 'Statistics' is used in two senses–Plural and Singular. In its plural form it refers to the statistical data collected in a systematic manner with some definite aim or object in view. In singular form, it means statistical methods or the subject itself. It includes the methods and principles concerned with collection, analysis and interpretation of numerical data.

1.4 DEFINITIONS OF STATISTICS

Different authors have given different definitions of Statistics in two different senses viz. plural and singular. Some of the definitions are given below.

Statistics in plural sense

1. "Statistics are numerical statement of facts in any department of enquiry placed in relation to each other"–A.L. Bowley.
2. "Statistics are the classified facts representing the conditions of the people in a state...specially those facts which can be stated in numbers or in tables of numbers or in any tabular or classified arrangement."–Webster.
3. "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 standards of accuracy, collected in a systematic manner for a predetermined purpose and placed in relation to each other."–Horace Secrist.

In plural sense Horace Secrist's definition seems to be the most exhaustive.

Statistics in singular sense

In singular sense, different authors have defined Statistics differently.

1. "Statistics may be called the science of counting."–A.L. Bowley.
2. "Statistics may rightly be called the science of averages."–A.L. Bowley.

3. "Statistics is the science of measurement of social organism, regarded as a whole in its manifestations"–A.L. Bowley.
4. "Statistics may be defined as the science of collection, presentation, analysis and interpretation of numerical data."–Croxton and Cowden.
5. "Statistics may be regarded as a body of methods for making wise decisions in the face of uncertainity".–Wallis and Roberts.

1.5 FUNCTIONS OF STATISTICS

We find today that there is hardly a phase of human activity which does not find statistical techniques useful. The evergrowing usefulness of Statistics is due to the functions it performs. Some of the most important functions of Statistics are

(i) The complex mass of data are made simple and understandable with the help of statistical methods.
(ii) To interpret the various characteristics of data, classification is done by the application of improved techniques of Statistics.
(iii) To study relationship between two or more phenomena statistical methods are used.
(iv) Statistics provides various informations in analysis of data that enlarges individual experience.
(v) Statistics helps in formulating policies in different fields.
(vi) Statistical methods are highly useful tools in analysing the past data and predicting future trends.
(vii) Statistics helps us to arrive at any correct and dependable conslusion.
(viii) Statistics helps in decision making in the face of uncertainity.
(ix) The correctness of the laws of the different branches of knowledge can only be tested with the help of Statistics.
(x) One important function of Statistics is to provide techniques for making comparisons.

1.6 SCOPE AND IMPORTANCE OF STATISTICS

It is very difficult to describe clearly and concisely the importance of any particular subject. Statistics being a comprehensive science, it has its application to almost all the branches of human knowledge and therefore it becomes more difficult to describe its scope. Even in our day-to-day life we apply statistical methods knowingly or unknowingly. Starting as a science of king now it has covered all branches of science viz. social, physical and natural. The physical sciences do not benefit as much from Statistics as social sciences.

It would be pointless to try to mention all the areas in which statistical methods are used. To mention only a few, Statistics is becoming increasingly important in Agriculture, Business and Commerce, Economics, Astronomy, Biology, Engineering, Medical, Meterology, Physics, Psychology, Sociology, Insurance, Operations Research, Computer Programming, Disaster Management, Demography, Actuarial Science and so on. The use of statistical methods in each of these fields grew in a different way. Statistics is not at the same stage of development in all the different fields of study.

1.7 LIMITATIONS OF STATISTICS

The field of Statistics, though widely used in all areas of human knowledge and widely applied in a variety of disciplines, it has its limitations. It cannot be applied to all situations. Some of the limitations are

(i) Statistics deals only with aggregates and no importance is given to individual items.
(ii) Statistics studies only quantitative characteristics and does not study qualitative characteristics.
(iii) Laws of Statistics are true on an average
(iv) Statistical results are only approximately correct.
(v) Statistics does not reveal the entire story rather it is a helping hand.
(vi) Statistics can be misused and hence should be applied only by experts.

1.8 STATISTICAL POPULATION AND SAMPLE

By Population we generally mean human population but in Statistics population is the aggregate of some individuals or objects, under study. It is also called Universe. A population may be finite or infinite depending upon the number of elements in it. For example, population of students in a class is finite whereas the population of sand particles in a river beach is infinite. Again a population may be real or hypothetical. For example, population of books in bookstall is real whereas the population of outcomes of a die throwing many a number of times is hypothetical.

The number of units in a population is called population size.

A Sample is a small representative part taken from the population to know the characteristic of the population under study. The number of units in a sample is called its sample size. For example to know the average height of 100 students, if we select 5 students and by measuring them infer about the average height of 100 students, then 5 students constitute a sample.

1.9 CENSUS AND SAMPLING

If information is collected from each and every individual of a population, the method of collecting information is known as census. For example after every ten years population census is conducted where every household is examined.

Again, if information is carried out on a properly selected representative sample, the method of collecting information is known as sampling. Thus, sampling is a techniques which enables us to draw inferences about the entire population simply by studying a few of them. Sampling requires less time, labour and money and hence it is more economical than census.

EXERCISE–1

1. Write a short note on the origin and development of the science of Statistics.
2. Give one definition of Statistics which you think to be the best.
3. Write a note on the scope and limitations of Statistics. **(AHSEC 1994)**
4. Discuss the limitations of Statistics. **(AHSEC 1995, 2005, 2008)**
5. "Statistical thinking will one day be as necessary for efficient citizenship as the ability to read and write." Comment. **(AHSEC 1996)**
6. Give the idea of statistical population. **(AHSEC 1993)**
7. Distinguish between
   (i) Population and Sample. **(AHSEC 1995, 2002, 2004)**
8. Define sample and population **(AHSEC 2006)**

2

COLLECTION OF DATA

2.1 INTRODUCTION

Statistical investigation means search for knowledge with the help of statistical methods. Only that knowledge can be obtained through statistical investigation, which can be expressed quantitatively. Thus statistical investigation or enquiry means some sort of investigation where the relevant information can be collected in quantitative terms i.e., in the form of.data. All numerical statements of facts are not statistical data. Statistical data are numerically expressed aggregate of facts, collected in a systematic manner for a predetermined purpose and placed in relation to each other. Statistical data are either enumerated or estimated.

The collection of data is an important part of any statistical investigation. Collection of data means the methods that are to be employed for getting the required information from the units under investigation. It depends on the nature, object and scope of enquiry. Statistical data can be either Primary or Secondary.

2.2 PRIMARY AND SECONDARY DATA

Data which are collected for the first time by the investigator himself for a specific purpose are known as primary data. Such data are collected by means of a census or survey. Secondary data are those which have been previously collected by some agency for their own purpose but are now used in a different connection.

Some examples of primary data are Reserve Bank of India Bulletin (Monthly), annual report of railway board etc. The examples of secondary data are Monthly Abstract of Statistics, International Labour Bulletin (monthly) etc.

2.3 DISTINCTION BETWEEN PRIMARY AND SECONDARY DATA

The difference between primary data and secondary data is largely a matter of degree. Data which are primary in the hands of one may be secondary for others. The differences between the two are as follows:

(i) Primary data are those data which are collected for the first time and thus original in character whereas, secondary data are obtained from some one else's records.

(ii) Primary data are in the shape of raw data to which statistical methods are applied but the secondary data are like finished products since they have been processed statistically.

(iii) After statistical treatment the primary data lose their original shape and becomes secondary data. Primary data once published become secondary data.

2.4 COLLECTION OF PRIMARY DATA

For the collection of primary data, the following methods can be adopted.

A. Direct Personal Investigation

According to this method, the investigator himself personally goes to the persons and collects the necessary information. The investigator establishes personal contact with the informants and conducts on the spot enquiry.

This method gives more reliable results because of personal approach. This method is suitable when the field of enquiry is small and a high degree of accuracy is desired.

But the method is costly and time consuming. Also, the personal bias of the investigator may enter into the data.

B. Indirect Oral Investigation

This method is used incase where informants are reluctant to supply information and is very difficult to contact them directly. Under this method indirect evidence of third parties who are in touch with facts desired, is recorded. Enquiry committees and commissions mostly use this method.

This method is economic and time saving. Here the direct contact with the primary source is not necessary. A wide area can be brought under investigation.

The informations obtained from the third parties may not be reliable at times.

C. Investigation through Local Agencies

Here there is no formal collection of any data but the local correspondents residing in different areas collect the information and report the same to the authority. This method is adopted by newspapers and periodicals.

This method is very cheap and yield result easily and a wide area can be covered.

But the reliability of data may be a matter of doubt.

D. Mailed Questionnaire Method

In this method the schedules of questions known as questionnaires are mailed to the informants with the request of quick response after duly filled in. Generally this method is adopted by research institutes and private bodies.

This method is the least expensive method. The bias of the investigator is completly ruled out. By this method original data are collected.

This method cannot be applied when the informants are illiterate and success of the method depends upon the co-operation of the respondents.

E. Schedules sent through Enumerators

Questionnaire is a list of questions which are answered by the respondent himself in his own handwriting while schedule is the method of getting answers to the questions in a form which are filled by the interviewers in a face to face situation.

In this method the enumerators go to the respondents with the schedule and record their replies. Population census is done by this technique.

This method can be applied even when the informants are illiterate. By this method maximum possible results can be obtained. Here method of substitution can be applied if there is non-response.

This method is quite expensive and time consuming. However, the accuracy largely depend upon the skill and efficiency of the investigator.

2.5 DRAFTING OF QUESTIONNAIRE

The person framing the questionnaire needs a thorough knowledge of the field of enquiry. There is no hard and fast rule for drafting a questionnaire but the following points should be borne in mind.

(i) The size of the questionnaire should be as small as possible.

(ii) The questions should be relevant to the purpose of the investigation.

(iii) Questions should be arranged in a logical sequence.

(iv) Questions should be simple, clear and unambiguous.

(v) Personal and sensitive questions should be avoided.

(vi) Questions should be framed to facilitate the respondent by giving them the questions such as–Multiple choice, Tick-options, True or False etc. In other words, questions should be capable of objective answer.

(vii) Questionnaires should look attractive.

(viii) Questions should be framed in conformity with the educational background and mental capacity of the respondents.

2.6 COLLECTION OF SECONDARY DATA

The published sources of secondary data are

(i) Official publications of the international organisations like the U.N.O., W.H.O, International Labour Organisations (I.L.O), World Bank etc.

(ii) Official publications of the Govt. organisations like Central Statistical Organisation (C.S.O.) National Sample Survey Organisation (N.S.S.O.), Directorate of Economics and Statistics etc.

(iii) Official publications of the semi-govt. organisations and local authorities.

(iv) Publications of Universities and Research institutes.

(v) Reports of various Committees and Commissions.
(vi) Publications of various Commercial and Financial institutes.
(vii) Newspapers, Magazines and Journals etc.

Beside all these there are various unpublished sources also from where secondary data may be obtained.

2.7 PRECAUTIONS IN USING SECONDARY DATA

Secondary data should not be used blindly, and they should be used with extreme care. The following points are to be examined before using secondary data.

(i) The authority which had collected the data should be reliable and dependable.
(ii) The purpose of the enquiry under investigation should have uniformity with the original purpose for which data were collected.
(iii) The interval between the time periods when data were originally collected to the time when it is going to be used should not be much.
(iv) The definition of units in which the data were collected.
(v) There should not be a significant variation between the degree of accuracy required for the present purpose and the original data.

2.8 TYPES OF DATA

1. Qualitative (Attribute) and Quantitative (Variable) Data

The data which are measurable i.e., direct quantitative measurement is fairly possible are called variable data e.g. height, weight, length, breadth, wage etc. Various important statistical methods like measures of central tendency, dispersion, skewness etc. can be used in the analysis of such data.

Attributes are qualitative characteristics which cannot be measured quantitatively, such as literacy, honesty, beauty, blindness, deafness etc. Thus, attribute data are those for which it is not possible to measure the magnitude. An investigator can only study the presence or absence of a particular quantity in a group of individuals. It is not possible to measure the magnitude of blindness or literacy quantitatively, but one can count the number of persons who have these attributes and those who do not have.

2. Time Series data or Chronological data

The data which are collected chronologically i.e., in accordance with time are called time-series data. Thus, in time series data time is the most important and pre-dominant factor. Sometimes they are called historical data as the data collected relate to either past or present. In such data the level of phenomenon measured are related to the change in time. The following is an example of time-series data.

Table No. 2.1

The Production of Paddy Under Golaghat sub-division of Assam

Year	Production of Paddy (in Metric tons)
1999-2000	94, 214
2000-2001	1,76,433
2001-2002	1,67,639
2002-2003	1,75,087
2003-2004	1,73,050

3. Spatial or Geographical Data

The data which are collected geographically i.e. in accordance with geographical region are called spatial data. Here place is the dominating factor. The place may be countries, states, districts etc.

Geographical data are usually listed either in alphabetical order or size of the frequencies. The following is an example of geographical data:

Table No. 2.2

The Population of Five Districts of Assam According to 1991 Census

Districts	Population ('000)
Dibrugarh	750
Golaghat	450
Jorhat	700
Sibsagar	400
Tinsukia	600

SAMPLE QUESTIONNAIRE

OBJECT

Questionnaire for collecting information regarding the socio- economic conditions of the people of a locality.

Use Tick Mark (√) whereever applicable.

1. Name :
   ....
2. Sex : Male ☐ Female ☐
3. Age : ....
4. Name of Father/Husband : ....
5. Religion : ....
6. Caste : General ☐ OBC ☐ MOBC ☐ SC ☐ ST ☐
7. No. of members in the family : ....
8. No. of members reading in different classes :
   School ....
   College ....
   University ....
9. Monthly income of the family Rs.: ....
10. No. of earning members : ....
11. No. of dependents : ....
12. Monthly expenditure (Rs.)
    Food
    Housing
    Clothing
    Education
    Misc.
13. Average Monthly Savings (Rs.) : ....
14. Living Status :
    (a) Own house : Yes/No.
    (b) Do you have :Car ☐ Scooter ☐ Washing Machine ☐
    T.V. ☐ Computer ☐ Inverter ☐

EXERCISE-2

1. What do you mean by statistical data? **(AHSEC 1993, 2001, 2002)**
2. What do you mean by Primary data and Secondary data? **(AHSEC 1990, 1992, 1993)**
3. Distinguish between Primary data and Secondary data. **(AHSEC 1995, 1998, 2000, 2002, 2004, 2007)**
4. Write a note on different methods of collecting Primary data. **(AHSEC 1991, 1993, 2001, 2008)**
5. What are the different methods of collecting Primary data? **(AHSEC 1994)**
6. Describe the qualities that a good questionnaire should have. **(AHSEC 1994)**
7. What do you mean by qualitative data and quantitative data. **(AHSEC 1990)**
8. Distinguish between qualitative data and quantitative data. **(AHSEC 2000, 2007)**
9. State the difference between questionnaire and schedule. **(AHSEC 1995, 1998, 2005)**
10. Select the correct answer. **(AHSEC 1999)**

    Data taken from a Research Journal will be considered as:

    (i) Primary data

    (ii) Secondary data

    (iii) Both Primary data and Secondary data.

    (iv) None of the above. **[Ans. (ii)]**
11. What is secondary data? Name the sources from which this type of data can be collected. **(AHSEC 2003, 2006)**
12. What precautions should be taken in the use of secondary data?
13. What are spatial data and Time-series data? Give one example of each.
14. Fill up the gaps.

    (i) A questionnaire is filled up by the __________. **(AHSEC 1990)**

    (ii) A schedule is filled up by the __________. **(AHSEC 2001)**

    (iii) __________ is a suitable method of collecting data in case where the informants are literate and spread over a vast area.

    (iv) Data can be obtained through a statistical __________.

    (v) Mailed questionnaire method cannot be used when the informants are __________. **[Ans. (i) Respondent (ii) Enumerator (iii) Mailed questionnaire method (iv) Investigation (v) Illiterate]**
15. It is proposed to conduct a sample survey to obtain information on literacy of the population of your locality. Draft a suitable questionnaire for that purpose. **(AHSEC 1996)**
16. Distinguish between time series data and spatial data. **(AHSEC 2008)**

3

FREQUENCY DISTRIBUTION AND GRAPHICAL REPRESENTATION

3.1 INTRODUCTION

A frequency distribution refers to data classified on the basis of some variable that can be measured such as wage, age, number of children etc. A variable refers to the characteristic that varies in magnitude in a frequency distribution. A *variable* may be either discrete or continuous. A *discrete variable* is that which always takes an integral value. For example, the number of children, the number of employees etc. A *continuous variable* can take every value both integral and fractional within the range of possibilities, such as the height or weight of individuals etc. Generally continuous data are obtained through measurements while discrete data are derived by counting.

Frequency distribution is simply a table in which the data are grouped into classes and the number of cases which fall in each class (frequencies) are recorded. The word 'distribution' refers to the way in which the items are distributed in different classes.

3.2 TYPES OF FREQUENCY DISTRIBUTION

1. Discrete Frequency Distribution

Here we count the number of times each value of the variable occurs in the data. The occurance of the variable is counted with the help of vertical bars, called Tally marks. An example is given below.

Table 3.1

No. of Children	Tally Mark	Frequency
0					3
1						4
2	~~				~~		6
3						4
4					3
Total		**20**

2. Continuous Frequency Distribution

There are certain variables which will not have a distinct integer value and in those cases the frequency distribution will be continuous. Also, when the data set is very large, it becomes necessary to condense the data into a suitable number of classes and then combined frequencies are assigned to the respective classes. An example of continuous frequency distribution is given below:

Table 3.2

Heights (cm)	Tally Mark	Frequency
145-150	|||	3
150-155	~~||||~~ ||	7
155-160	~~||||~~ ~~||||~~	10
160-165	~~||||~~ ~~||||~~ ||||	14
165-170	~~||||~~ ~~||||~~	10
170-175	||||	4
175-180	||	2
Total		**50**

3.3 CONSTRUCTION OF CONTINUOUS FREQUENCY DISTRIBUTION

The following technical terms are associated with the construction of grouped or continuous frequency distributions.

A. Range

The range of a frequency distribution may be defined as the difference between the lower limit of the first class interval and the upper limit of the last class interval. In the example given in Table 3.2., range = 180–145 = 35.

B. Number of Classes

The number of classes should neither be too large nor too small. Normally between 6 and 15 classes are considered to be adequate.

C. Class Intervals

The difference between the upper limit and the lower limit of a class is known as class interval. In the class 145-150, class interval is 5. The size of the class interval is determined by the number of classes and the range.

D. Class Limits

The class limits are the lowest and the highest values that can be included in the class.

E. Class frequency

The number of observations corresponding to a particular class is known as the frequency of that class. In the above example the class frequency of the class 170-175 is 4.

F. Mid-Point

It is the value lying half-way between the lower and the upper class limits of a class-interval.

$$\text{Mid point of a class} = \frac{\text{upper limt of the class + lower limit of the class}}{2}$$

G. Class-Boundaries

The upper and lower class limits of the exclusive type classes are called class boundaries.

$$\text{Upper class boundary} = \text{Upper class limit} + \frac{d}{2}$$

$$\text{Lower class boundary} = \text{Lower class limit} - \frac{d}{2}$$

Where d = gap between the upper limit of any class and lower limit of the succeeding class.

H. Width of a Class

The width of a class interval is the difference between the class boundaries.

Width of a class = Upper class boundary – Lower class boundary.

I. Relative Frequency and Percentage Frequency

The frequency of a class when expressed as a rate of the total frequency of the distribution is called relative frequency and percentage frequency is the ratio of class-frequency to the total frequency expressed as percentage.

$$\text{Relative Frequency} = \frac{\text{Class frequency}}{\text{Total frequency}}$$

$$\text{Percentage Frequency} = \frac{\text{Class frequency}}{\text{Total frequency}} \times 100$$

J. Frequency Density

It is the ratio of the class frequency to the width of that class.

$$\text{Frequency density} = \frac{\text{Class frequency}}{\text{Width of the class}}$$

K. Cumulative Frequency

If for each class the frequencies given are aggregates of the preceeding frequencies, they are known as cumulative frequencies.

L. Methods of Forming Class-intervals

There are two ways of forming class intervals.

Exclusive Method

Under this method a class interval is such that the lower limit is included in the preceeding class where as the upper class limit is included in the succeeding class interval.

Inclusive Method

If a interval is such that the lower as well as the upper class limits are included in the same class interval, it is called inclusive class interval.

Table 3.3

Inclusive Class	Tally Mark	Exclusive Class	Tally mark
10-19	𝍸	10-20	𝍸
20-29	ǀǀǀ	20-30	ǀǀǀ
30-39	ǀ	30-40	ǀ
40-49	ǀ	40-50	ǀ
Total	**10**	**Total**	**10**

In inclusive class interval 30 will be included in the class 30-39 and in exclusive class interval 30 will be included in the class 30-40 instead of 20-30 class.

Example 1: The following are the daily wages of 40 workers.

10	26	24	16	16	23	28	23
25	18	10	11	20	21	19	18
15	13	22	17	15	29	29	12
34	15	14	18	22	24	30	38
17	32	36	20	19	27	33	31

(i) Form a frequency distribution table taking 4 as the class interval.

(ii) Find the percentage of workers getting wage below Rs. 32.

(AHSEC 1994)

Solution: (i) Here, maximum observation = 38

minimum observation = 10

$\therefore$ Difference = 38 − 10 = 28

$$\text{Total number of classes} = \frac{28}{4} = 7$$

Wage	Tally Mark	Number of Workers	Cumulative Frequency
10-14	𝍸	5	5
14-18	𝍸 III	8	13
18-22	𝍸 III	8	21
22-26	𝍸 II	7	28
26-30	𝍸	5	33
30-34	IIII	4	37
34-38	III	3	40
Total		**40**	

(ii) Number of workers below Rs. 32 = 33 + 2 = 35

[From the last column we see that number of workers below Rs. 30 is 33 and observing the data we find 2 persons]

∴ Percentage of workers getting wage below

$$\text{Rs. } 32 = \frac{35}{40} \times 100\%$$

$$= 87.5\%$$

Example 2: Form a frequency distribution from the following data by Inclusive Method taking 4 as the magnitude of class intervals.

Take the data from previous example.

Solution:

Class Interval	Tally Mark	Frequency
10-13	𝍸	5
14-17	𝍸 III	8
18-21	𝍸 III	8
22-25	𝍸 II	7
26-29	𝍸	5
30-33	IIII	4
34-37	II	2
38-41	I	1
Total		**40**

3.4 GRAPHICAL REPRESENTATION OF FREQUENCY DISTRIBUTION

An important function of Statistics is to present complex and unwidely data in such a manner that they would be easily understandable. Numerical figures are not always interesting and with the increase in size their complexity increases. A graphic representation is the geometrical image of a set of data. Graph gives a better picture than when the data are arranged in tabular form. They help us in studying cause and effect relationship between two variables. They are useful for studying both time series and frequency distribution.

Advantage of Graphic Representation

(i) It simplifies the complexity of data and makes it readily understandable.
(ii) It attracts attention of people.
(iii) It saves time and efforts to understand the facts.
(iv) It makes the comparison easy.
(v) A graph studies relationship between two or more variables.

Graphs of Frequency Distribution

The most commonly used graphs of a frequency distribution are

(i) Histogram
(ii) Frequency Polygon
(iii) Frequency Curve
(iv) Cumulative Frequency Curve or Ogive.

Histogram

Histogram is a bar diagram which represents a frequency distribution with continuous classes. The width of all bars is equal to class interval and heights of the bars are proportional to the frequency of the respective classes. Each rectangle is joined with the other so as to give a continuous picture.

In case of unequal class intervals there will be rectangles of unequal width. If it is desired to keep width of all the rectangles equal then, height will be adjusted proportionally so that area of the rectangle remains the same.

Example 3: Draw a histogram from the following data.

Class interval :	**10-20**	**20-30**	**30-40**	**40-50**	**50-60**	**60-70**	**70-80**
Frequency :	**3**	**10**	**14**	**24**	**17**	**14**	**3**

Solution:

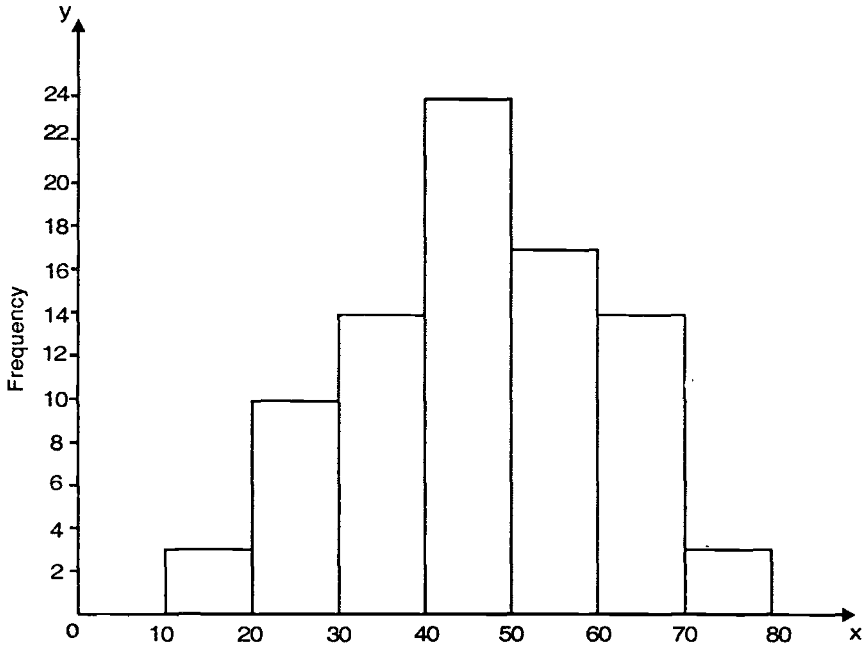
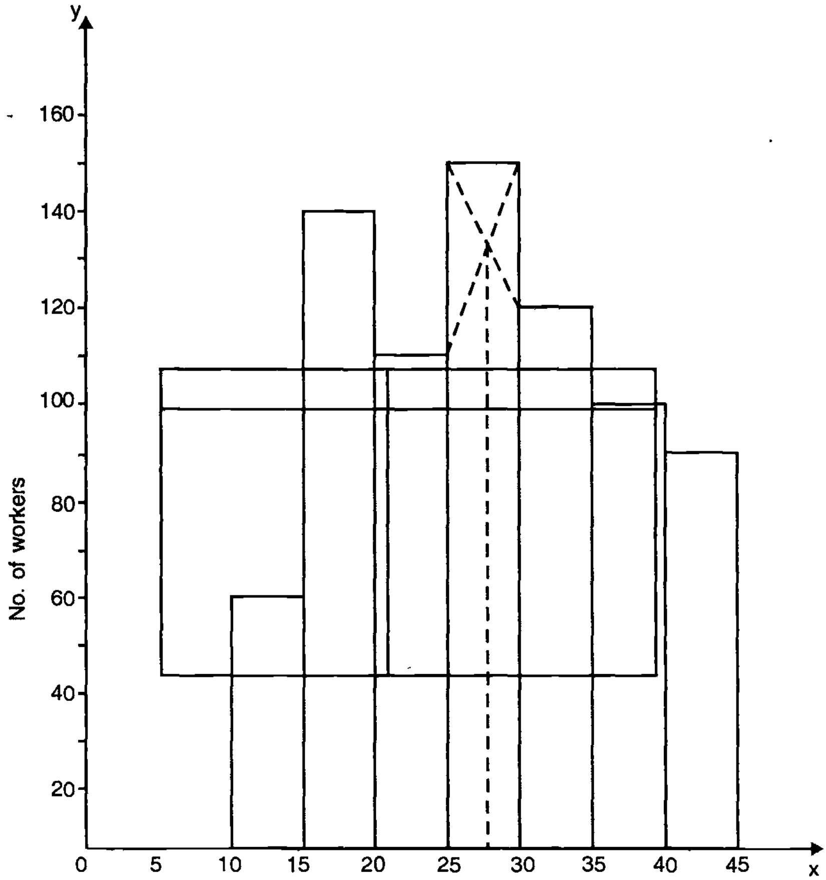


Fig. 3.1

Example 4: Draw a histogram for the following distribution and find the modal wage.

Wage (Rs.) :	**10-15**	**15-20**	**20-25**	**25-30**	**30-35**	**35-40**	**40-45**
No. of Workers :	**60**	**140**	**110**	**150**	**120**	**100**	**90**

Solution: Mode can be located from histogram of frequency distribution. The first step is to draw a histogram and take the modal class rectangle. Join lines from the top of the rectangle of the modal group to the top of the adjacent rectangles, as in the graph 3.2. Vertical line is then drawn downwards from the point of intersection of these two lines. Where this line meets the horizontal axis indicates the position of the mode. From figure 3.2 it is evident that the mode is approximately Rs. 27.9.



Wage (Rs.)
Fig. 3.2

A Few Points to Remember

1. To draw histogram from discrete frequency distribution consider the given values as the midpoint of the continuous classes assuming that the frequencies are uniformly distributed throughout each class.
2. If the class intervals are of inclusive type then change the class limits into class boundaries.
3. Histogram and Historigram : The word histogram should not be confused with the term historigram which stands for time series graphs. Graphs of continuous time series are known as historigrams. If the absolute value of the variable are taken into consideration the graphs obtained are known as absolute historigrams. Historigrams may be constructed on the natural scale or on the ratio scale.

Frequency Polygon

If we join the middle points of the tops of the adjacent rectangles of the histogram with line segments, frequency polygon is obtained. It is not essential to draw histogram in order to obtain frequency polygon. This consists in plotting the frequencies along y-axis against the mid values of the corresponding classes along x-axis and then joining the points by straight lines. The frequency polygon should be extended to the base at both the ends.

Example 5 : Draw a histogram and frequency polygon from the following data.

Wages	No. of Workers
10-15	50
15-20	140
20-25	110
25-30	150
30-35	120
35-40	100
40-45	80

Solution:

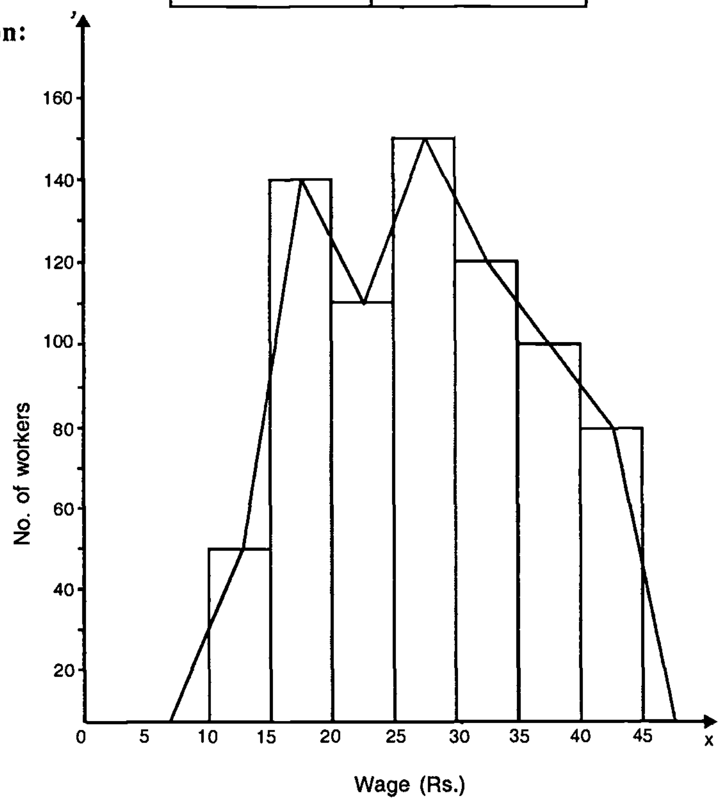


Fig. 3.3

Frequency Curve

If in a grouped distribution, the class intervals become very small so that the number of observations increases, the histogram or the frequency polygon shall tend to a smooth continuous curve, known as frequency curve. Instead of joining the points of the frequency polygon by lines if we join them by a smooth free hand curve, frequency curve is obtained.

Example 6 : Draw a frequency curve from the following data.

Class interval :	**1-2**	**2-3**	**3-4**	**4-5**	**5-6**	**6-7**	**7-8**
Frequency :	**3**	**10**	**20**	**24**	**20**	**14**	**3**

Solution:

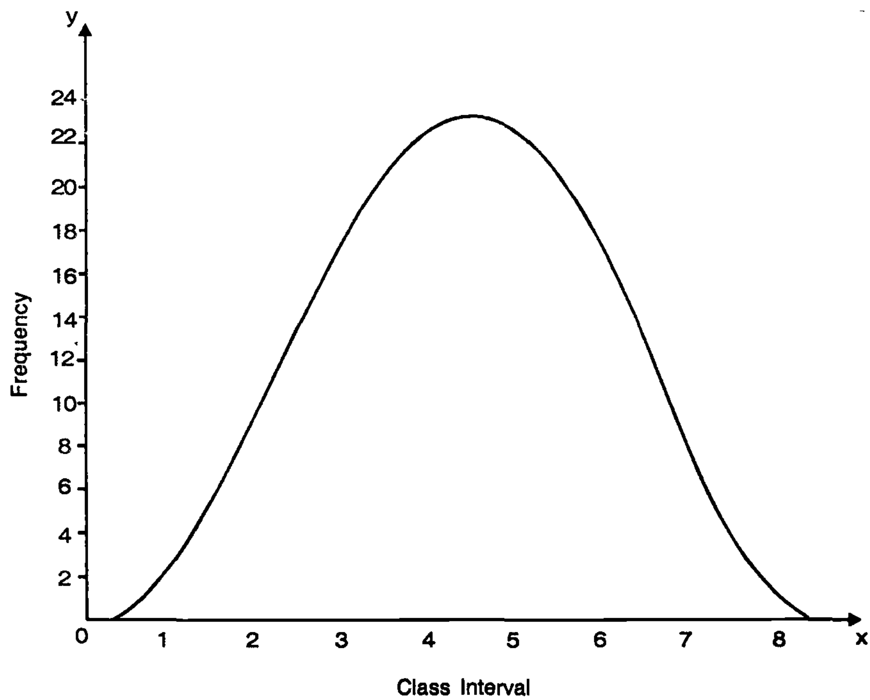


Fig. 3.4

Ogive

An ogive is a graph of cumulative frequencies of a frequency distribution of continuous series. In drawing an ogive the class boundaries are plotted on the x-axis and the cumulative frequencies on the y-axis and the resulting curves are known as ogives.

Ogives are of two types

(i) *Less Than Ogive* : Here 'less than' cumulative frequencies are plotted against upper class boundaries of the respective class intervals.

(ii) *More Than Ogive* : Here 'greater than' cumulative frequencies are plotted against the lower class boundaries.

Less than ogive is an increasing curve and more than ogive is a decreasing curve. The foot of the perpendicular from the intersection of two ogives gives the value of median.

Uses of Ogives

Ogives are useful for locating partition values viz. Median, Quartiles, Deciles. Percentiles etc. They are also useful to determine the number of proporiion of items above or below a particular value. With the help of ogivc we can compare more than one distributions.

3.5 LOCATION OF PARTITION VALUES FROM OGIVES

Various partition values can be located graphically with the help of ogives. First of all, we draw an ogive by plotting the given data on a graph. For finding positional values, we mark the position of frequency on y-axis and draw a line from this mark till it intersects with the ogive. From this point of intersection we draw a vertical line meeting x-axis. The point where the line touches the x-axis is the required positional value.

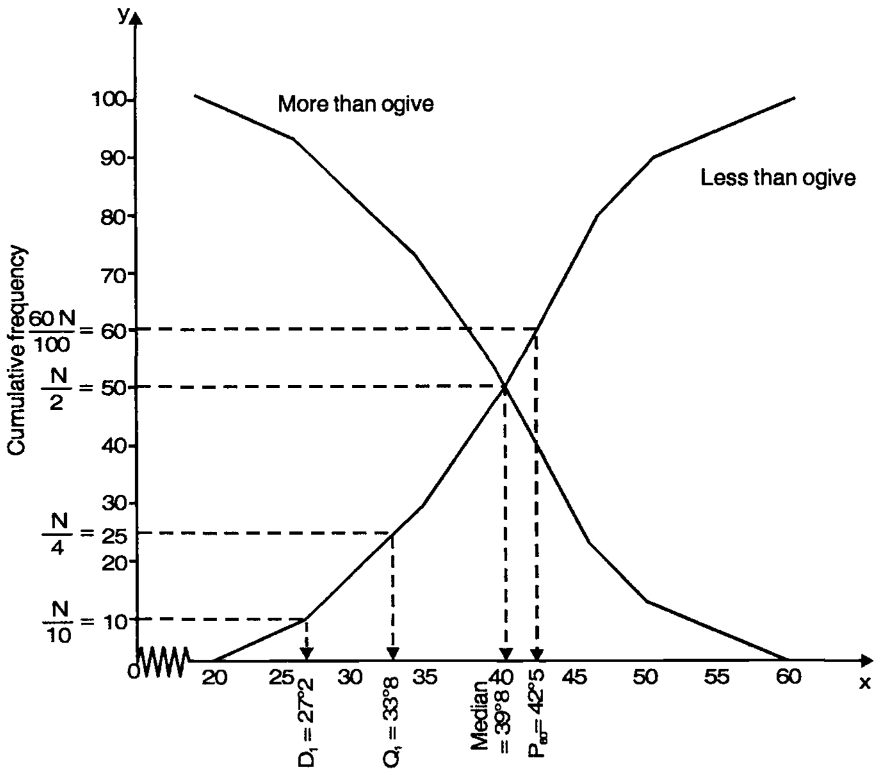
Example 7 : Draw two gives from the following data.

Class	Frequency
20-25	6
25-30	9
30-35	13
35-40	23
40-45	19
45-50	15
50-55	9
55-60	6

From your ogive find median, first quartile, first decile and 60^{th} percentile.

Solution: First we shall calculate cumulative frequency

Class	Frequency	Cumulative Frequency	
		Less than	More than
20-25	6	6	100
25-30	9	15	94
30-35	13	28	85
35-40	23	51	72
40-45	19	70	49
45-50	15	85	30
50-55	9	94	15
55-60	6	100	6



Class interval
Fig. 3.5

Note : Partition values will be discussed in section 4.6 in Chapter 4.

EXERCISE-3

1. The following data relate to loan advance to 40 farmers by the co-operative Bank (the figures are in hundred rupees)

12	8	10	9	5	14	12	10	14	20
18	19	10	7	6	9	10	12	13	15
18	12	8	12	17	20	15	12	14	17
11	15	18	17	19	12	15	5	5	9

   (a) Construct a frequency table taking 5 as the class interval.

   (b) Draw increasing and decreasing ogives : what is the significance of their point of intersection. **(AHSEC 1999)**

2. Distinguish between Histogram and Historigam. **(AHSEC 2000)**
3. Explain following terms with rough sketch
   (i) Histogram
   (ii) Frequency Polygon
   (iii) Cumulative Frequency Curve **(AHSEC 2001)**
4. What is cumulative frequency curve? Write a note on its uses. **(AHSEC 1993)**
5. What do you mean by ogive? What are the different types of ogives? Explain their uses. **(AHSEC 1996)**
6. The following table gives the frequency distribution of the number of heads in hundred tosses with 4 coins. Draw a histogram of the frequency distribution.

No of heads :	0	1	2	3	4	Total
Frequency :	6	28	36	25	5	100

(AHSEC 1996)

7. Define the following terms
   (i) Class interval
   (ii) Class limit
   (iii) Class boundary **(AHSEC 1996)**
8. Write a short note on Histogram. **(AHSEC 1995)**
9. Write down the names of the frequency graphs. How is mode determined from graph?
10. Fill up the gaps.
   (i) With the help of ogive one can determine __________. **(AHSEC 1999)**
   (ii) In drawing histograms the class intervals should be __________. **(AHSEC 1997)**
   (iii) Using ogive we can determine a particular measure of central tendency, namely __________. **(AHSEC 1997)**

(iv) In a histogram the height of the rectangles are always _________ to the respective class frequencies. **(AHSEC 1998)**

Ans. (i) Partition values(ii)Continuous (iii) Median (iv) Proportional.

11. From the following data, determine the median graphically (use graph paper).

Height (cm) :	150-155	155-160	160-165	165-170	170-175	175-180	180-185
No. of persons :	5	8	20	25	18	10	4

(AHSEC 1991)

12. Fill up the gap.

(i) There are _________ % observations on the LHS of the third quartile (Q3) of a frequency curve. **(AHSEC 1997)**

[Ans. 75]

13. State the difference between discrete variable and continuous variable. **(AHSEC 1998)**

14. (a) Which of the following statements is true?
    (i) Histogram and historigram are similar in look.
    (ii) Simple bar diagrams and histograms are similar in look.
    (iii) None of the above.

    (b) Select the correct answer:
    With the help of ogive one can determine _________.
    (i) Median
    (ii) Deciles
    (iii) Quartiles
    (iv) All the above. **(AHSEC 1999)**

[Ans. (a)-(ii), (b)-(iv)]

15. How do you determine median graphically? **(AHSEC 2002, 2008)**

16. Explain the advantages of graphic representation of statistical data.

17. Draw a histogram from the given data and locate the position of mode from the graph.

Weight (kg) :	35-40	40-45	45-50	50-55	55-60	60-65
No. of person :	12	30	22	32	18	10

18. Below is the frequency distribution of marks of 100 students.

Marks	No. of Students	Marks	No. of Students
20-29	7	60-69	9
30-39	11	70-79	14
40-49	24	80-89	2
50-59	32	90-99	1

Draw less than ogive and use it to determine median.

19. Draw a histogram and frequency curve from the following data

Marks :	0-10	10-20	20-30	30-40	40-50	50-60	60-70
No. of students:	4	8	11	15	12	7	3

(AHSEC 2008)

20. Draw a histogram for the following distribution and find the modal wage:

Wages (in Rs.)	10-15	15-20	20-25	25-30
No. of workers:	40	120	190	150

(AHSEC 2007)

21. Draw a histogram and frequency polygon from the following data.

Wages (in Rs.)	10-15	15-20	20-25	25-30	30-35	35-40
No. of workers:	40	120	90	140	120	80

4

MEASURES OF CENTRAL TENDENCY

4.1 INTRODUCTION

So far, we have learned the techniques of data collection and summarising and condensing data into various tables and graphs. Now, we shall deal with some arithmetic procedures that can be used for analysing and interpreting data.

Measures of central tendency also known as averages describe the tendency of individual items to cluster or concentrate around the centre in a frequency distribution. The most important objective of measures of central tendency is to determine a single representative value for the entire series. They perform the function of giving a concise picture of the entire group and hence they are useful as a basis for comparison with other groups of data.

The following measures of central tendency are commonly used in practice.

(i) Arithmetic Mean (A.M.) or simply mean

(ii) Median

(iii) Mode

(iv) Geometric Mean (G.M.)

(v) Harmonic Mean (H.M.)

4.2 CHARACTERISTIC OF AN IDEAL AVERAGE

An average to be considered as an ideal one should possess the following characteristics.

(i) It should be easy to understand and easy to calculate.

(ii) It should be rigidly defined i.e., it should have one and only one interpretation.

(iii) It should be based on all the observations of the series.

(iv) It should be capable of further algebraic treatment i.e., its use in further statistical theory is enhanced.

(v) It should be affected least by fluctuations of sampling.

(vi) It should be affected least by extreme values i.e., too large or too small values.

Σ–Notation

Greek alphabet Σ (capital sigma) is used to denote sum. If a variable x takes the values $x_1, x_2, \ldots, x_n$ then the sum

$$x_1 + x_2 + \ldots + x_n = \sum_{i=1}^{n} x_i \text{ or } \sum x$$

It implies that minimum value i can take is 1 and maximum value is n and i takes only integral values.

Rules of Σ

$$\sum_{i=1}^{n}(x_i \pm y_i) = \sum_{i=1}^{n} x_i \pm \sum_{i=1}^{n} y_i$$

$$\sum_{i=1}^{n}(c x_i) = c \sum_{i=1}^{n} x_i$$

$$\sum_{i=1}^{n} c = nc$$

c is a constant

4.3 ARITHMETIC MEAN (A.M.)

The AM of a series is obtained by adding the values of the series and dividing by the number of items.

Ungrouped Data

If $x_1, x_2, \ldots, x_n$ are the values of the variable x then A.M. denoted by $\bar{x}$ is given by.

$$\bar{x} = \frac{x_1 + x_2 + \ldots + x_n}{n} = \frac{\sum_{i=1}^{n} x_i}{n} = \frac{\sum x}{n}$$

Grouped data : - If $x_1, x_2 \ldots x_n$ are the values of the variable x with frequencies $f_1, f_2 \ldots f_n$ then

$$\bar{x} = \frac{x_1 f_1 + x_2 f_2 + \ldots + x_n f_n}{f_1 + f_2 + \ldots + f_n} = \frac{\sum_{i=1}^{n} x_i f_i}{\sum_{i=1}^{n} f_i} = \frac{\sum xf}{N}$$

Where $N = \Sigma f$ = Total frequency.

In case of continuous frequency distribution x is taken as the mid point of the class intervals.

Short-cut Method

AM can also be calculated by this method

Ungrouped data

$$\bar{x} = A + \frac{\sum d}{n}; \qquad d = x - A$$

A = Assumed mean

Grouped data

$$\bar{x} = A + \frac{\sum fd}{N}; \qquad d = x - A$$

In case of continuous data with equal class intervals

$$\bar{x} = A + \frac{\sum fd'}{N} \times h; \qquad d' = \frac{x - A}{h}$$

h = class width

This method is also known as step-deviation method which is based on property 4 of A.M.

The assumed mean A is arbitrary, but it should be chosen in such a way that the value of x is reduced considerably. Generally, it is taken some where in the centre of x values.

Example 1 : Find the AM of the numbers 5, 8, 12, 15, 20.

Solution: $$\bar{x} = \frac{\sum x}{n}$$

$$= \frac{5+8+12+15+20}{5} = \frac{60}{5} = 12$$

using short-cut method.

x	$d = x - A$
5	-7
8	-4
12	0
15	3
20	8
	$\Sigma d = 0$

Assumed mean, $A = 12$

$$\bar{x} = A + \frac{\sum d}{n}$$

$$= 12 + \frac{0}{5}$$

$$= 12$$

Example 2 : For the discrete frequency distribution given below find AM.

Wages (in Rs.) :	**40**	**45**	**50**	**60**	**65**
No. of persons :	**2**	**6**	**25**	**10**	**3**

Solution:

Wages (x)	No. of Persons (f)	xf
40	2	80
45	6	270
50	25	1250
60	10	600
65	3	195
	$N = \Sigma f = 46$	$\Sigma xf = 2395$

$$\therefore \quad \bar{x} = \frac{\sum xf}{N} = \frac{2395}{46} = 52.07$$

using short-cut method

x	f	$d = x - 50$	fd
40	2	– 10	– 20
45	6	– 5	– 30
50	25	0	0
60	10	10	100
65	3	15	45
	$N = 46$		$\Sigma fd = 95$

$$\therefore \quad \bar{x} = A + \frac{\sum fd}{N}; \quad A = 50$$

$$= 50 + \frac{95}{46}$$

$$= 52.07$$

Example 3 : Find AM of the distribution of marks given below:

Marks :	**0-20**	**20-40**	**40-60**	**60-80**	**80-100**
No. of Students :	**5**	**15**	**30**	**12**	**8**

Solution:

Marks	No. of students (f)	Mid Point (x)	xf
0-20	5	10	50
20-40	15	30	450
40-60	30	50	1500
60-80	12	70	840
80-100	8	90	720
	$N = 70$		$\Sigma xf = 3560$

$$\therefore \quad \bar{x} = \frac{\Sigma xf}{N} = \frac{3560}{70} = 50.9$$

using step-deviation method

Marks	f	x	$d' = \frac{x - 50}{20}$	fd'
0-20	5	10	-2	-10
20-40	15	30	-1	-15
40-60	30	50	0	0
60-80	12	70	1	12
80-100	8	90	2	16
	$N = 70$			$\Sigma fd' = 3$

Now, $$\bar{x} = A + \frac{\Sigma fd'}{N} \times h$$

$$= 50 + \frac{3}{70} \times 20$$

$$= 50.9$$

4.4 PROPERTIES OF AM

1. The algebraic sum of deviations of a set of values from their AM is zero.

Symbolically, (a) $\Sigma(x_i - \bar{x}) = 0$ for ungrouped data

(b) $\Sigma f_i (x_i - \bar{x}) = 0$ for grouped data

***Proof*:**

(a) $$\text{LHS} = \Sigma (x_i - \bar{x})$$

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

$$= n \frac{1}{n} \Sigma x_i - n\bar{x} \qquad [\sum_{i=1}^{n} c = nc]$$

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

$$= 0$$

(b)

$$\text{LHS} = \Sigma f_i (x_i - \bar{x})$$

$$= \Sigma f_i x_i - \Sigma f_i \bar{x}$$

$$= N. \frac{1}{N} \Sigma f_i x_i - \bar{x}\, \Sigma f_i \qquad [\therefore\ \bar{x} \text{ a constant}]$$

$$= N\bar{x} - \bar{x}.N \qquad [\Sigma f_i = N]$$

$$= 0$$

2. The sum of the squares of deviations about mean is the least. Mathematically, $\Sigma(x_i - A)^2$ is minimum when $A = \bar{x}$, where A is any arbitrary value.

Proof: Let S denote the sum of squares of the deviations of the values from any arbitrary value.

$$S = \Sigma (x_i - A)^2$$

$$= \Sigma (x_i - \bar{x} + \bar{x} - A)^2$$

$$= \Sigma [(x_i - \bar{x})^2 + (\bar{x} - A)^2 + 2 (x_i - \bar{x})(\bar{x} - A)]$$

$\because (\bar{x} - A)^2$ is a constant

$$= \Sigma(x_i - \bar{x})^2 + n(\bar{x} - A)^2 + 0$$

[By property 1, $\Sigma (x_i - \bar{x}) = 0$]

The value of S is the sum of two non negative values. The first one is a fixed number since $\bar{x}$ is a fixed number. Thus the value of S will be the least when $n(\bar{x} - A)^2$ becomes the least i.e., zero.

$$\therefore \quad n(\bar{x} - A)^2 = 0$$

$$\Rightarrow \quad (\bar{x} - A) = 0$$

$$\Rightarrow \quad \bar{x} = A$$

3. If n_1 and n_2 are the number of observations and $\bar{x}_1$ and $\bar{x}_2$ be their respective means of the two series then mean $\bar{x}$ of the combined series is

$$\bar{x} = \frac{n_1\bar{x}_1 + n_2\bar{x}_2}{n_1 + n_2}$$

(this result can be generalised to more than two events).

Proof: Suppose $x_{11}, x_{12}, ..., x_{1n_1}$ be the n_1 observations of the first series and $x_{21}, x_{2j}, ..., x_{2n_2}$ be the n_2 observations of the second series.

$$\therefore \qquad \bar{x}_1 = \frac{1}{n_1}\sum_{i=1}^{n_1} x_{1i} \qquad\qquad \bar{x}_2 = \frac{1}{n_2}\sum_{j=1}^{n_2} x_{2j}$$

$$\Rightarrow \qquad n_1\bar{x}_1 = \sum_{i=1}^{n_1} x_{1i} \qquad \Rightarrow \qquad n_2\bar{x}_2 = \sum_{j=1}^{n_2} x_{2j}$$

The mean $\bar{x}$ of the combined series.

$$\bar{x} = \frac{(x_{11}+x_{12}+...+x_{1n_1})+(x_{21}+x_{22}+...+x_{2n_2})}{n_1+n_2}$$

$$= \frac{\sum_{i=1}^{n_1} x_{1i} + \sum_{j=1}^{n_2} x_{2j}}{n_1+n_2}$$

$$= \frac{n_1\bar{x}_1 + n_2\bar{x}_2}{n_1+n_2}$$

4. If the relation between two variables x and y is $y = a + bx$, where a and b are constants then $\bar{y} = a + b\bar{x}$. **(AHSEC 1996)**

Proof: $y = a + bx$

$\Rightarrow \quad \Sigma y = \Sigma a + \Sigma bx$ (Taking Σ over both sides)

$\Rightarrow \quad \dfrac{\Sigma y}{n} = \dfrac{\Sigma a}{n} + \dfrac{b\Sigma x}{n}$ (Dividing both sides by n)

$\Rightarrow \quad \bar{y} = \dfrac{na}{n} + b\bar{x}$

$\Rightarrow \quad \bar{y} = a + b\bar{x}$

Note: This property indicates that AM is dependent of both change in origin and scale.

Example 4 : A test in algebra given to 400 high school children of whom 150 were boys and 250 were girls. The mean score of boys was 72 and that of girls was 73. Find the mean score of the combined group.

Solution: Given n_1 = number of boys = 150 $\qquad \bar{x}_1$ = mean score of boys = 72.

n_2 = number of girls = 250 $\bar{x}_2$ = mean score of girls = 73.

$$\bar{x} = \frac{n_1\bar{x}_1 + n_2\bar{x}_2}{n_1 + n_2}$$

$$= \frac{150 \times 72 + 250 \times 73}{150 + 250}$$

$$= \frac{10800 + 18250}{400}$$

$$= 72.6$$

Hence, the mean of the combined group is 72.6.

Example 5 : Calculate the missing frequency from the following data you are given that mean is 50.9.

Marks	No. of Students
0-20	5
20-40	?
40-60	30
60-80	12
80-100	8

Solution: Let the missing frequency be f_2.

Marks	**No. of Students (f)**	**Mid Point (x)**	xf
0-20	5	10	50
20-40	f_2	30	$30f_2$
40-60	30	50	1500
60-80	12	70	840
80-100	8	90	720
	$N = 55 + f_2$		$\Sigma xf = 3110 + 30f_2$

$$\bar{x} = \frac{\Sigma xf}{N}$$

$$\Rightarrow \quad 50.9 = \frac{3110 + 30f_2}{55 + f_2}$$

$$\Rightarrow \quad 2799.5 + 50.9\,f_2 = 3110 + 30\,f_2$$

$$\Rightarrow \quad 20.9\,f_2 = 310.5$$

$$\Rightarrow \quad f_2 = 14.9$$

$$\therefore \quad f_2 = 15 \qquad \text{[since frequency can not be in fraction]}$$

Example 6 : If the relation between the variables x and y is $2x - 3y + 4 = 0$ and $\bar{y} = 6$, then what will be the value of $\bar{x}$?

Solution: If $2x - 3y + 4 = 0$

$\Rightarrow \quad 3y = 2x + 4$

$\Rightarrow \quad 3\bar{y} = 2\bar{x} + 4$

Now $\quad \bar{y} = 6$

$\therefore \quad 18 = 2\bar{x} + 4$

$\Rightarrow \quad 2\bar{x} = 18 - 4$

$\Rightarrow \quad \bar{x} = 7$

Example 7 : The mean salary paid to 1000 employees of an establishment was found to be Rs. 180.40. Later on, after disbursement of salary, it was discovered that the salary of two employees was wrongly entered as Rs. 297 and 165. Their correct salaries were Rs. 197 and 185. Find the correct AM.

Solution: $\quad \bar{x} = \dfrac{\Sigma x}{n}$

$\Rightarrow \quad \Sigma x = n\bar{x}$

Given $\quad n = 1000$

$\bar{x} = 180.40$

$\therefore \quad \Sigma x = 1000 \times 180.40$

$= 180400$

Corrected $\quad \Sigma x = \Sigma x -$ wrong items + correct items

$= 180400 - (297 + 165) + (197 + 185)$

$= 180320$

$\therefore \quad$ Corrected $\bar{x} = \dfrac{180320}{1000} = 180.32$

Thus corrected mean salary is Rs. 180.32

Merits and Demerits of AM.

Merits:

(i) It is easy to understand and easy to calculate.

(ii) It is based on all the observations of the series.

(iii) It is rigidly defined.

(iv) It does not necessitate the arranging of data as median or mode does.

(v) It is capable of further algebraic treatment.

(vi) It is not much affected by fluctuations of sampling.

Demerits:

(i) It is greatly affected by extreme values.

(ii) AM can not be located by inspection nor it can be located graphically.

(iii) It is not suitable when some items are missing.

(iv) AM can not be used in case of qualitative data.

(v) Unless some assumptions are made, AM is difficult to compute in case of open-end classes.

4.5 MEDIAN

Median is that value of a variable which divides the data into two equal halves so that same proportion of values lie above and below the median value.

Ungrouped data

Here after arranging the data in ascending or descending order the middle most item is termed as median. In case of odd number of items there is one middle term but incase of even number of items there are two middle terms and median is the AM of two values.

(i) When number of items n is odd–

$$\text{Median} = \text{value of } \left(\frac{n+1}{2}\right)^{th} \text{ item}$$

(ii) When n is even–

$$\text{Median} = \frac{\text{value of} \left(\frac{n}{2}\right)^{th} \text{ item} + \text{value of} \left(\frac{n}{2}+1\right)^{th} \textit{ item}}{2}$$

Example 8 : Find the median of

(i) 10, 6, 15, 2, 3.

(ii) 10, 6, 2, 3, 8, 15.

Solution: (i) Here $n = 5$. After arranging in ascending order we get 2, 3, 6, 10, 15.

$\therefore$ Median = 6

(ii) Here $n = 6$, so arranging in ascending order we get the values as 2, 3, 6, 8, 10, 15.

$$\therefore \quad \text{Median} = \frac{6+8}{2} = 7$$

Median for Discrete data

For finding median in case of discrete data cumulative frequencies are calculated and the value of the variable corresponding to cumulative frequency $\frac{N+1}{2}$ gives the median.

Example 9 : Find the median of the following frequency distribution.

Marks :	**5**	**10**	**15**	**20**	**25**	**30**	**35**	**40**	**45**	**50**
No. of students :	**2**	**5**	**12**	**14**	**15**	**11**	**13**	**9**	**7**	**2**

Solution:

Marks (x)	No. of Students (f)	Cumulative Frequency
5	2	2
10	5	7
15	12	19
20	14	33
25	15	48
30	11	59
35	13	72
40	9	81
45	7	88
50	2	90
	N = 90	

Now, Median = value of $\frac{N+1}{2}$ th item

$= \text{value of } \frac{90+1}{2}$ th item

$=$ value of 45.5 th item

Here we see that 45.5 is greater than the cumulative frequency 33 but less than 48 corresponding to $x = 25$

$\therefore$ Median = 25

Median for Continuous Frequency Distribution

For computing median in case of continuous frequency distribution we shall have to determine the median class. Median class can be defined as that class for which the cumulative frequency is just greater than $\frac{N}{2}$. The value of median now can be obtained by the formula.

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

where

L = Lower limit of the median class

N = Σf = Total frequency

F = Cumulative frequency of the class preceeding the median class.

f = frequency of the median class.

h = Width of the median class.

Example 10 : Find median from the following frequency distribution.

Marks :	**0-20**	**20-40**	**40-60**	**60-80**	**80-100**
No. of students :	**5**	**15**	**30**	**12**	**8**

Solution: First we shall calculate the cumulative frequency distribution.

Marks	f	Cumulative frequency
0-20	5	5
20-40	15	20
40-60	30	50
60-80	12	62
80-100	8	70
	$N = 70$	

Median class is 40-60 since cumulative frequency just greater than $\frac{N}{2} = \frac{70}{2}$ = 35 is 50.

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

$$= 40 + \frac{\frac{70}{2} - 20}{30} \times 20$$

$$= 40 + \frac{15}{30} \times 20$$

$$= 50$$

Example 11 : Find median from the following data.

Value :	**10-14**	**15-19**	**20-24**	**25-29**	**30-34**	**35-39**
Frequency :	**5**	**10**	**13**	**20**	**7**	**5**

Solution: Here the class-intervals are not continuous. So first of all these are made continuous by making class-boundaries i.e., by subtracting 0.5 from the lower limit and adding 0.5 to the upper limit and then the calculation of median is done.

Class boundaries	f	Cumulative Frequency
9.5-14.5	5	5
14.5-19.5	10	15
19.5-24.5	13	28
24.5-29.5	20	48
29.5-34.5	7	55
34.5-39.5	5	60
	$N = 60$	

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

$$= 24.5 + \frac{\frac{60}{2} - 28}{20} \times 5$$

$$= 24.5 + 0.5$$

$$= 25$$

Example 12 : Calculate median from the following cumulative frequency distribution.

Marks	No. of Students
Below 10	3
Below 20	8
Below 30	17
Below 40	20
Below 50	22

Solution:

Marks	f	Cumulative Frequency
0-10	3	3
10-20	$8 - 3 = 5$	8
20-30	$17 - 8 = 9$	17
30-40	$20 - 17 = 3$	20
40-50	$22 - 20 = 2$	22
	$N = 22$	

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

$$= 20 + \frac{\frac{22}{2} - 8}{9} \times 10$$

$$= 20 + 3.33$$
$$= 23.33$$

Example 13 : Find the missing frequency when median is given as Rs. 25.

Expenditure (Rs.) :	**0-10**	**10-20**	**20-30**	**30-40**	**40-50**
No. of families :	**14**	**25**	**27**	—	**15**

Solution: Let the missing frequency be f_4

Expenditure	f	Cumulative Frequency
0-10	14	14
10-20	25	39
20-30	27	66
30-40	f_4	$66 + f_4$
40-50	15	$81 + f_4$
	$N = 81 + f_4$	

Since median is 25, median class is 20-30

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

$$\Rightarrow \qquad 25 = 20 + \frac{\frac{81 + f_4}{2} - 39}{27} \times 10$$

$$\Rightarrow \qquad 5 = \frac{f_4 + 3}{54} \times 10$$

$$\Rightarrow \qquad f_4 = 24$$

Hence the missing frequency is 24.

Merits and Demerits of Median

Merits:

(i) It is easy to understand and calculate.

(ii) It can be calculated for frequency distribution having open-end classes.

(iii) In many cases it can be obtained by inspection.

(iv) It is not affected at all by extreme values.
(v) It is the only average to deal with qualitative data.

Demerits:

(i) For calculating median arrangement of data is necessary.
(ii) Combined median of two or more series cannot be obtained.
(iii) Compared to AM median is affected more by fluctuations of sampling.
(iv) It cannot be precisely determined in a series comprising of an even number of items.

4.6 OTHER PARTITION VALUES

A measure of central tendency is necessarily a measure of location but the reverse may not be true. Five different measures of central tendency are such values of a distribution around which most of the observations lie. So, measure of central tendency are also the measure of location. Beside these there are other measures of location also which are not measure of central tendency such as Quartiles, Deciles and Percentiles. All these partition values can be located from ogives as we have already discussed in chapter 3.

Quartiles

The values of the variable dividing a series into four equal parts are called the quartiles. Hence there are three quartiles. Below Q_1 lie 25% of the total number of observations and below Q_3 lie 75% of the total number of observations. Here Q_2 = median.

Ungrouped data

$$Q_i = \text{size of } \frac{\text{i}\,(n+1)}{4} \text{th item.} \qquad \text{i} = 1, 2, 3$$

Grouped data

$$Q_i = L + \frac{\frac{\text{i}\,N}{4} - F}{f} \times h; \qquad i = 1, 2, 3$$

The computational procedure is basically the same as that of computing the median.

Deciles

The values of the variable which divide an arranging series into ten equal parts are called Deciles. Hence, there are nine deciles, each part containing 10% of the total. In case of deciles D_5 = median.

Ungrouped data

$$D_j = \text{Size of } \frac{j\,(n+1)}{10} \text{ th item}; \qquad j = 1, 2, \ldots, 9$$

Grouped data

$$D_j = L + \frac{\frac{jN}{10} - F}{f} \times h; \qquad j = 1, 2, \ldots, 9$$

Percentiles

The values of the variable which divide an arranging series into 100 equal parts, are called percentiles. Hence, there are 99 percentiles. In case of percentiles P_{50} = median.

Ungrouped data

$$P_k = \text{size of } \frac{k\,(n+1)}{100} \text{ th item}; \qquad k = 1, 2, \ldots, 99$$

Grouped data

$$P_k = L + \frac{\frac{kN}{100} - F}{f} \times h; \qquad k = 1, 2, \ldots, 99$$

Example 14 : Calculate median, Q_1, Q_3, D_6 and P_{58} from the marks of 10 students.

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

Solution: We first arrange the data as follows

Serial No.	Marks
1	11
2	12
3	14
4	18
5	22
6	26
7	30
8	32
9	35
10	41

$$\text{Median} = \text{size of } \frac{n+1}{2} \text{ th item}$$

$$= \text{size of } \frac{10+1}{2} \text{ th item}$$

$= \text{size of } 5.5 \text{ th item}$

$$= \text{size of } \frac{\text{5th item} + \text{6th item}}{2}$$

$$= \frac{22+26}{2}$$

$= 24$

$$Q_1 = \text{size of } \frac{n+1}{4} \text{ th item}$$

$$= \text{size of } \frac{10+1}{4} \text{ th item}$$

$= \text{size of } 2.75 \text{ th item}$
$= \text{2nd item} + .75 \text{ (size of 3 rd item-size of 2 nd item)}$
$= 12 + 0.75\ (14–12)$
$= 13.5$

$$Q_3 = \text{size of } \frac{3(n+1)}{4} \text{ th item}$$

$= \text{size of } 8.25 \text{ th item}$
$= \text{8th item} + 0.25 \text{ (size of 9th item-size of 8 th item)}$
$= 32 + 0.25\ (35 - 32)$
$= 32.75$

$$D_6 = \text{size of } \frac{6(n+1)}{10} \text{ th item}$$

$= \text{size of } 6.6 \text{ th item}$
$= \text{6 th item} + 0.6 \text{ (size of 7 th item-size of 6 th item)}$
$= 26 + 0.6\ (30 - 26)$
$= 28.4$

$$P_{58} = \text{size of } \frac{58(n+1)}{100} \text{ th item}$$

$= \text{size of } 6.38 \text{ th item}$
$= \text{6 th item} + 0.38 \text{ (size of 7 th item-size of 6th item)}$
$= 26 + 0.38\ (30 - 26)$
$= 27.52$

Example 15 : Find Q_1, D_4 and P_{50} from the following data.

Class :	**5-10**	**10-15**	**15-20**	**20-25**	**25-30**	**30-35**	**35-40**
Frequency :	**10**	**15**	**25**	**40**	**35**	**20**	**5**

Solution:

Class	f	Cumulative Frequency
5-10	10	10
10-15	15	25
15-20	25	50
20-25	40	90
25-30	35	125
30-35	20	145
35-40	5	150
	N = 150	

Q_1 class is the class which has cumulative frequency just > $\frac{N}{4} = \frac{150}{4} =$ 37.5. Thus, 15-20 is the Q_1 class.

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

$$= 15 + \frac{\frac{150}{4} - 25}{25} \times 5$$

$$= 15 + 2.5$$

$$= 17.5$$

The cumulative frequency just > $\frac{4N}{10}$ = 60 is 90 which is in the class 20-25. Thus 20-25 class is the D_4 class.

$$D_4 = L + \frac{\frac{4N}{10} - F}{f} \times h$$

$$= 20 + \frac{60 - 50}{40} \times 5$$

$$= 20 + 1.25$$

$$= 21.25$$

Again the cumulative frequency just > $\frac{50N}{100}$ = 75 is 90. Hence 20-25 is the P_{50} class

$$P_{50} = L + \frac{\frac{50N}{100} - F}{f} \times h$$

$$= 20 + \frac{75-50}{40} \times 5$$

$$= 20 + 3.13$$

$$= 23.13$$

4.7 MODE

Mode is the value of the variable which occurs most frequently or repeat itself the maximum number of times. Mode may be regarded as the most typical of a series of values. If there is only one mode, the distribution is called unimodal and if there are two modes, the distribution is called bimodal.

Ungrouped data

Mode cannot be easily determind from individual series unless it is converted to a discrete (or continuous) frequency distribution.

Example 16 : Find the modal size from the given sizes of an item.

7, 5, 6, 7, 8, 7, 4, 7, 6 ,5

Solution:

Size :	4	5	6	7	8
Frequency :	1	2	2	4	1

From the frequency distribution we see that the size 7 occurs with maximum frequency of 4. Hence mode is 7.

Mode for Continuous Frequency Distribution

For computing mode in case of continuous frequency distribution we shall have to determine the modal class. The class corresponding to the maximum frequency is called the modal class and the value of mode now can be obtained by the formula.

$$\text{Mode} = L + \frac{f_1 - f_0}{2f_1 - f_0 - f_2} \times h$$

where

L = Lower limit of the modal class.

f_1 = frequency of the modal class

f_0 = frequency of the class preceeding the modal class.

f_2 = frequency of the class succeeding the modal class.
h = width of the modal class.

Example 17 : Calculate mode from the following frequency distribution.

Marks :	**0-20**	**20-40**	**40-60**	**60-80**	**80-100**
No. of Students :	**5**	**15**	**30**	**12**	**8**

Solution: Modal class is 40-60 since the frequency in the class is maximum.

$$\text{Mode} = L + \frac{f_1 - f_0}{2f_1 - f_0 - f_2} \times h$$

$$= 40 + \frac{30-15}{2\times 30 - 15 - 12} \times 20$$

$$= 40 + \frac{15\times 20}{33}$$

$$= 40 + 9.09$$

$$= 49.09$$

Merits and Demerits of Mode

Merits:

(i) It is easy to understand and calculate.
(ii) It is not at all affected by extreme values.
(iii) It can be located graphically.
(iv) It gives the most representative value of a series.

Demerits:

(i) It is not rigidly defined.
(ii) It is not based on all the observations.
(iii) It is not capable of further algebraic treatment.

4.8 RELATIONSHIP AMONG MEAN, MEDIAN AND MODE

For a symmetrical distribution the mean, median and mode are identical. But if the distribution is moderately asymmetrical, there is an emperical relationship between them. The relationship is

$$\text{Mean} - \text{Mode} = 3\,(\text{Mean} - \text{Median})$$

$$\Rightarrow \quad \text{Mode} = 3\,\text{Median} - 2\,\text{Mean}$$

Example 18 : In a moderately asymmetrical distribution the mode and mean are 32.1 and 35.4 respectively. Calculate the median.

Solution: $\text{Mode} = 3\,\text{Median} - 2\,\text{Mean}$

$$\Rightarrow \quad 32.1 = 3\,\text{Median} - 2 \times 35.4$$

$\Rightarrow$ 3 Median = 32.1 + 70.8

$\Rightarrow$ Median = 34.3

4.9 GEOMETRIC MEAN

GM of a set of n numbers is the nth root of their product. If one value of a series is zero, the GM is zero and if one or more values are negative, the GM is meaningless.

Ungrouped data

If $x_1, x_2, ..., x_n$ are the values of a variable x then

$$G = \sqrt[n]{x_1, x_2 ...... x_n}$$
$$= (x_1, x_2, ... x_n)^{1/n}$$

For computational easiness we can take logarithm

$$\log G = \frac{1}{n}[\log x_1 + \log x_2 + ... + \log x_n]$$

$$= \frac{1}{n} \Sigma \log x_i$$

$$\therefore \quad G = \text{Antilog}\left(\frac{1}{n} \Sigma \log x_i\right)$$

Grouped data

If $x_1, x_2, ..., x_n$ are the values of a variable x with frequencies $f_1, f_2, ..., f_n$ then

$$G = \left(x_1^{f_1}.x_2^{f_2} .... x\right)^{\frac{1}{N}}$$

$$\therefore \quad G = \text{Antilog}\left(\frac{1}{N} \Sigma f_i \log x_i\right)$$

Example 19 : Find the GM of 4, 6 and 9.

Solution:

$$G = \sqrt[3]{4 \cdot 6 \cdot 9}$$
$$= (2^2 \cdot 2 \cdot 3 \cdot 3^2)^{1/3}$$
$$= (2^3.3^3)^{1/3}$$
$$= 2.3$$
$$= 6$$

Example 20 : Calculate GM of 5.3, 4.7, 4.2, 3.2, 5.8.

Solution:

x	$\log x$
5.3	.7243
4.7	.6721
4.2	.6232
3.2	.5051
5.8	.7634
	$\Sigma \log x = 3.2881$

Now,
$$G = \text{Antilog}\left(\frac{1}{n}\Sigma \log x\right)$$

$$= \text{Antilog}\left(\frac{1}{5} \times 3.2881\right)$$

$$= \text{Antilog}(.6576)$$
$$= 4.545$$

Example 21 : Calculate GM of the following distribution.

Class :	**0-10**	**10-20**	**20-30**	**30-40**	**40-50**
Frequency :	**10**	**17**	**25**	**20**	**8**

Solution:

class	f	x	$\log x$	$f.\log x$
0-10	10	5	0.6990	6.9900
10-20	17	15	1.1761	19.9937
20-30	25	25	1.3979	34.9475
30-40	20	35	1.5441	30.8820
40-50	8	45	1.6532	13.2256
	N = 80			$\Sigma f.\log x = 106.0388$

$$G = \text{Antilog}\left(\frac{1}{N}\Sigma f \log x\right)$$

$$= \text{Antilog}\left(\frac{106.0388}{80}\right)$$

$$= \text{Antilog}(1.3255)$$
$$= 21.16$$

Example 22 : If n_1 and n_2 are the sizes, G_1 and G_2 the geometric means of two series respectively, the geometric mean G, of the combined series is given by

$$\log G = \frac{n_1 \log G_1 + n_2 \log G_2}{n_1 + n_2}$$

Solution: Let $x_{11}, x_{12}, \ldots\ldots, x_{1n_1}$ and $x_{21}, x_{22}, \ldots\ldots, x_{2n_2}$ be n_1 and n_2 items of two series.

$$G_1 = (x_{11}.x_{12} \ldots\ldots x_{1n_1})^{\frac{1}{n_1}} \Rightarrow \log G_1 = \frac{1}{n_1} \sum_{i=1}^{n_1} \log x_{1i}$$

$$G_2 = \left(x_{21}.x_{22} \ldots x_{2n_2}\right)^{\frac{1}{n_2}} \Rightarrow \log G_2 = \frac{1}{n_2} \sum_{j=1}^{n_2} \log x_{2j}$$

the GM, G of the combined series

$$G = (x_{11}.x_{12} \ldots\ldots x_{1n_1} \cdot x_{21}.x_{22} \ldots\ldots x_{2n_2})^{\frac{1}{n_1+n_2}}$$

$$\Rightarrow \qquad \log G = \frac{1}{n_1 + n_2} \left[\sum_{i=1}^{n_1} \log x_{1i} + \sum_{j=1}^{n_2} \log x_{2j} \right]$$

$$= \frac{1}{n_1 + n_2} [n_1 \log G_1 + n_2 \log G_2]$$

Merits and Demerits of GM

Merits:

(i) It is based on all observations.

(ii) It is not affected by extreme values compared to AM.

(iii) It is rigidly defined.

(iv) It is suitable for further mathematical treatment.

Demerits:

(i) It is not easy to understand and calculate.

(ii) It cannot be computed if any one of the observations is zero and negative.

4.10 HARMONIC MEAN

HM of set of values is the reciprocal of the mean of reciprocals of the values.

Ungrouped data

It $x_1, x_2, \ldots, x_n$ are the values of a variable x then

$$H = \frac{1}{\dfrac{\frac{1}{x_1} + \frac{1}{x_2} + \ldots\ldots + \frac{1}{x_n}}{n}} = \frac{n}{\Sigma\left(\frac{1}{x_i}\right)}$$

Grouped data

If x_1, x_2, ..., x_n are the values of x with respective frequencies f_1, f_2, ..., f_n then

$$H = \frac{N}{\frac{f_1}{x_1} + \frac{f_2}{x_2} + \ldots\ldots + \frac{f_n}{x_n}} = \frac{N}{\Sigma\left(\frac{f_i}{x_i}\right)}$$

Example 23 : Calculate the harmonic mean of 4, 8 and 12.

Solution:

$$H = \frac{3}{\frac{1}{4} + \frac{1}{8} + \frac{1}{12}}$$

$$= \frac{3}{\frac{6+3+2}{24}}$$

$$= \frac{72}{11} = 6.55$$

Example 24 : Calculate HM.

Class :	**10-20**	**20-30**	**30-40**	**40-50**	**50-60**
Frequency :	**4**	**5**	**10**	**8**	**3**

Solution:

Class	f	x	
10-20	4	15	.27
20-30	5	25	.20
30-40	10	35	.29
40-50	8	45	.18
50-60	3	55	.05
	N = 30		$\Sigma\left(\frac{f}{x}\right)$ **= 0.99**

$$\therefore \qquad H = \frac{N}{\Sigma\left(\frac{f}{x}\right)} = \frac{30}{.99} = 30.3$$

Note: In averaging rates and ratios involving speed, time and distance, sometimes it becomes difficult to decide whether AM or HM is the appropriate one. If the given ratios are expressed as x per unit of y then for finding the average ratio we use.

(i) HM if x's are given
(ii) AM if y's are given

Example 25 : Suppose a person travels 10 miles at 4 mph and again 12 miles at 5 mph. What is his average speed?

Solution: Here the rates are miles (x) per hour (y). Further miles (x) are given hence HM will be the appropriate average.

$$H = \frac{f_1 + f_2}{\frac{f_1}{x_1} + \frac{f_2}{x_2}}$$

$$= \frac{10+12}{\frac{10}{4} + \frac{12}{5}} = \frac{22}{4.9} = 4.49$$

$\therefore$ Average speed is 4.49 mph.

Example 26 : A man travels 10 hours at 4 mph and again 12 hours at 5 mph. What is his average speed?

Solution: Here the rates are miles (x) per hour (y) and hours (y) are given. Hence AM will be the appropriate average.

$$\text{AM} = \frac{f_1 x_1 + f_2 x_2}{f_1 + f_2}$$

$$= \frac{10 \times 4 + 12 \times 5}{10 + 12} = \frac{100}{22} = 4.55$$

$\therefore$ Average speed is 4.55 mph.

Merits and Demerits of HM

Merits:

(i) It is based on all observations.
(ii) It gives largest weight to smaller items.
(iii) It is rigidly defined.
(iv) It is useful in calculating rates.

Demerits:

(i) It is not easy to understand and calculate.
(ii) It cannot be calculated if value of any item is zero.

Example 27 : Prove that

(i) AM $\geq$ GM $\geq$ HM

(ii) AM × HM = (GM)²

Solution: We shall prove the result for two observations only. The result can be generalized for any number of observations.

Let x_1 and x_2 be two real numbers which are non-zero and non-negative. Then

$$AM = \frac{x_1 + x_2}{2}$$

$$GM = \sqrt{x_1 x_2}$$

$$HM = \frac{2}{\frac{1}{x_1} + \frac{1}{x_2}}$$

Consider, $(\sqrt{x_1} - \sqrt{x_2})^2 \geq 0$

$\Rightarrow \quad x_1 + x_2 - 2\sqrt{x_1 x_2} \geq 0$

$\Rightarrow \quad \frac{x_1 + x_2}{2} \geq \sqrt{x_1 x_2}$

$\Rightarrow \quad AM \geq GM$ ...(i)

Again, $\left(\frac{1}{\sqrt{x_1}} - \frac{1}{\sqrt{x_2}}\right)^2 \geq 0$

$\Rightarrow \frac{1}{x_1} + \frac{1}{x_2} - 2\frac{1}{\sqrt{x_1}} \cdot \frac{1}{\sqrt{x_2}} \geq 0$

$\Rightarrow \quad \sqrt{x_1 x_2} \geq \frac{2}{\frac{1}{x_1} + \frac{1}{x_2}}$

$\Rightarrow \quad GM \geq HM$ ...(ii)

Thus from (i) and (ii) we get, AM ≥ GM ≥ HM

(ii) LHS = AM × HM

$$= \frac{x_1 + x_2}{2} \quad \frac{2}{\frac{1}{x_1} + \frac{1}{x_2}}$$

$$= x_1 + x_2 \frac{x_1 x_2}{x_2 + x_1}$$

$$= x_1 x_2$$

$$= (GM)^2$$

For more than two observations this relation holds but the observations must be in Geometric Progression (GP).

Example 28 : If AM and GM of two numbers are 30 and 18 respectively, find the numbers.

Solution: Let the two numbers be x_1 and x_2

A/q, $\frac{x_1 + x_2}{2} = 30 \Rightarrow x_1 + x_2 = 60$ ...(i)

$\sqrt{x_1 x_2} = 18 \Rightarrow x_1 x_2 = 324$ ...(ii)

Putting $x_1 = 60 - x_2$ in (ii) we get,

$$(60 - x_2)\, x_2 = 324$$

$\Rightarrow \quad x_2^2 - 60\, x_2 + 324 = 0$

$\Rightarrow x_2^2 - 54\, x_2 - 6x_2 + 324 = 0$

$\Rightarrow \quad (x_2 - 54)(x_2 - 6) = 0$

$\Rightarrow \quad x_2 = 54$ or 6

$\therefore$ From (i) $x_1 = 6$ or 54

Hence the two numbers are 6 and 54.

4.11 USES OF DIFFERENT MEASURES OF CENTRAL TENDENCY

AM : It is the most important, widely used and best measure of central tendency. In the determination of average income, average price, average cost of production, average sales etc. i.e., those phenomenon which are capable of direct quantitative measurements, AM is the most appropriate measure.

Median

Median is most useful average in case of open end classes frequency distributions. To find the average of qualitative data median is the most suitable one.

Mode

Mode is mostly used in business and commerce. Meterological forecasts are based on mode.

GM

GM is the most useful when smaller items are to be given importance. They are used in the construction of Index Numbers. GM is the best in cases where ratios, percentages, etc. are considered.

HM

HM is specially useful in averaging rates and ratios where time is variable and distance is constant.

EXERCISE–4

1. What do you mean by measures of location? What are the different measures of location?
2. Which is the best measure of location? Give reasons. **(AHSEC 1999, 2007)**
3. Name the commonly used measures of location. **(AHSEC 1994, 2000)**
4. Write down the characteristics of an ideal measure of location. **(AHSEC 1997, 2000, 2001, 2006)**
5. Name a measure of location which is not a measure of central tendency. **(AHSEC 1995, 2005)**

   [Ans. Deciles]
6. Show that for a grouped distribution

   $\sum_{i=1}^{n} f_i(x_i - \bar{x}) = 0$ **(AHSEC 1995, 1999, 2007)**
7. Two sets with 28 and 36 observations have mean 2.9 and 5.6 respectively, find the mean of the combined series. **(AHSEC 2000)**

   [Ans. 4.42]
8. Show that sum of the deviations of the variate values about their mean is zero. **(AHSEC 2000)**
9. What do you mean by central tendency? Define different measures of central tendency. **(AHSEC 1991)**
10. "Arithmetic mean is a measure of location as well as a measure of central tendency"- Discuss. **(AHSEC 1993)**
11. Prove that for the n observations $x_1, x_2, ..., x_n$ with A.M $\bar{x}$.

    $\sum_{i=1}^{n} (x_i - \bar{x}) = 0$ **(AHSEC 2005, 2006)**
12. Which measure of location will be suitable to compare the following.

    (i)Average intelligence of students.

    (ii)Average size of shoes sold in a shop. **(AHSEC 1996)**

    [Ans. (i) Median (ii) Mode]
13. Which mean is commonly used in the following cases.

    (i)In computing rate of population growth.

    (ii)In computing the average speed. **(AHSEC 1996)**

    [Ans. (i) GM (ii) HM]
14. Which measure of location will be suitable to compare the sale of shirts of different sizes of a cloth shop? **(AHSEC 1995)**

    [Ans. Mode]

15. Find the AM of the square of the first n natural numbers. **(AHSEC 1995)**

[Ans. $\frac{(n+1)(2n+1)}{6}$]

16. What are the measures of central tendency? Why are they called measures of central tendency?
17. Find the AM and GM of the first n natural numbers.
18. Find the GM of 1, 2, 8, 16 **(AHSEC 1999, 2007)**

[Ans. 4]

19. What do you mean by the median of a frequency distribution? How would you determine it graphically? **(AHSEC 2002)**
20. Two sets with 30 and 40 observations have means 12.8 and 17.5 respectively, find the mean of the combined series. **(AHSEC 2002)**

[Ans. 15.49]

21. Prove that A.M $\geq$ GM $\geq$ HM **(AHSEC 1998, 2005, 2007)**
22. Calculate mean from the following distribution.

Class Interval	Frequency
0-10	2
10-20	7
20-30	10
30-40	1

(AHSEC 1992)

[Ans. 20]

23. An Aircraft travels 500 km at 200 km per hour and returns over the same route at 250 km per hour. Calculate the average speed of the whole journey. **(AHSEC 1992)**

[Ans. 222.2 kmph]

24. A variable x takes the values 2, 3, 7, 6, 5, 9. Find the GM. **(AHSEC 1993)**

[Ans. 4.74]

25. A variable x takes the values 2, 4, 5, 7, 9. Find AM and GM. **(AHSEC 1991)**

[Ans. AM = 5.4, GM = 4.79]

26. The mean of a set of 10 observations is 30.7 and the mean of another set of 12 observations is 18.5; find the mean of the combined set. **(AHSEC 1990)**

[Ans. 24.05]

27. Find the two numbers whose AM is 9 and GM is 7.2. **[Ans. 14.4, 3.6]**
28. The mean marks of 100 students were found to be 40. Later on it was discovered that a score of 5. was misread as 83. Find the correct mean.

[Ans. 39.22]

29. Find out the mode from the following series.

Size :	0-5	5-10	10-15	15-20	20-25	25-30	30-35
Frequency :	1	2	5	14	10	9	2

[Ans. 18.46]

30. Calculate Q_1, Q_2 and Q_3 from the given data.

Income :	Below 30	31-40	41-50	51-60	61-70	71-80	81 and above
No. of Person :	69	167	207	65	58	27	10

[Ans. 35.4, 43.7, 51.9]

31. From the following series trace out the missing frequencies if its median is 27.5 and number of items is 50.

Marks :	0-10	10-20	20-30	30-40	40-50	50-60
Frequency :	4	?	20	?	7	3

[Ans. 6, 10]

32. Determine the value of median from the follwing series.

Marks :	0-10	10-15	15-20	20-25	25-30
No. of Students :	7	5	8	38	42

[Ans. 23.95]

33. What is the relationship among mean, median and mode of a symmetrical distribution? **[Ans. Me = Md = Mo]**

34. What is the relationship among median, quartiles, deciles and percentiles? **[Ans. Md = $Q_2 = D_5 = P_{50}$]**

35. For a group of 50 students, the AM of x and y are 50 and 75 respectively. If $z = x + 2y$, what must be the AM of z? **[Ans. 200]**

36. Fill in the blanks:

(i) Mode = __________ – 2 Mean.

(ii) Mean – Mode = __________ (Mean – Median) **(AHSEC 2001)**

(iii) Mode is the value that has the greatest __________.

(iv) Median is an average of __________.

(v) For a qualitative phenomenon __________ is a suitable average.

(vi) HM is the reciprocal of the __________ mean of reciprocal of the values.

(vii) GM is the __________ root of the product of n observations.

(viii) Quartiles are measures of __________. **(AHSEC 1993)**

[Ans. (i) 3 Median (ii) 3 (iii) Frequency (iv) Position (v) Median (vi) Arithmetic (vi) nth (viii) location]

(ix) If u = $\frac{x-a}{h}$, then $\bar{x}$ = __________. **(AHSEC 2008)**

(x) The sum of deviations of the values from their mean is always is __________. **(AHSEC 2008)**

[Ans. (ix) a+$h\bar{u}$ (x) zero.]

37. Locate Median, Q_3, D_4 and P_{70} graphically from the following data

Age group :	0-10	10-20	20-30	30-40	40-50	50-60	60-70
No. of persons :	20	30	15	10	5	15	5

38. How would you determine Quartiles, Deciles and Percentiles graphically?

39. Write the value of $\sqrt{AM \times HM}$, where AM is the Arithmetic Mean and HM is the Harmonic Mean of two values x_1 and x_2. **(AHSEC 2007)**

[Ans. $\sqrt{x_1 x_2}$ **]**

40. Explain: Quartiles, Ogive, Deciles, Percentiles. **[AHSEC 2007]**

5

MEASURES OF DISPERSION

5.1 INTRODUCTION

In the preceeding chapter we have considered certain measures which are essential to study the central tendency of a frequency distribution. The other aspects of frequency distribution which we shall consider here are dispersion or variability of the observations. A measure of central tendency does not give the complete information about some distribution and we should use measure of dispersion to supplement the information. Measures of dispersion give an idea about the scattering of values around the central value of the data. The measures of dispersion are called averages of second order. These measures should attain large values if the observations are widely scattered from the centre and should have small values for the data having observation close to the mean.

There are two types of measures of dispersion.

(i) Absolute measure

(ii) Relative measure.

The absolute measures of dispersion are used to measure the variability of a given data expressed in same unit, while the relative measures are used to compare the variability of two or more sets of observations expressed in different units.

There are four absolute measures of dispersion and they are

(i) Range

(ii) Quartile Deviation (Q.D.) or semi-interquartile range

(iii) Mean deviation (MD)

(iv) Standard deviation (SD)

The relative measures are their co-efficients.

5.2 RANGE

The difference between the largest and smallest values of a set of data is known as Range.

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

$$\text{Co-efficient of Range} = \frac{L - S}{L + S}$$

Example 1: Calculate the range and co-efficient of range from the following data:

70, 75, 60, 22, 58, 80, 36.

Solution:

$$\begin{aligned} \text{Largest value } (L) &= 80 \\ \text{Smallest value } (S) &= 22 \\ \therefore \quad \text{Range} &= L - S = 80 - 22 = 58 \end{aligned}$$

$$\text{co-efficient of range} = \frac{L-S}{L+S}$$

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

$$= .57$$

Example 2 : From the following data calculate range and its co-efficient

Class :	**10-20**	**20-30**	**30-40**	**40-50**	**50-60**
Frequency :	**3**	**7**	**10**	**8**	**2**

Solution:

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

$$\text{co-efficient of Range} = \frac{L-S}{L+S}$$

$$= \frac{60-10}{60+10}$$

$$= \frac{50}{70} = .71$$

Merits and Demerits of Range

Merits

(i) It is the simplest measure of dispersion.

(ii) It is easy to understand and calculate.

(iii) It is rigidly defined.

Demerits

(i) It is not based on all observations. It is based on two extreme observations only.

(ii) It is affected much by fluctuations of sampling.

(iii) It is not capable of further algebraic treatment.

5.3 QUARTILE DEVIATION

Difference between third and first quartile is known as interquartile range and half of this difference is known as semi-interquartile range or $Q.D.$

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

$$\text{co-efficient of } Q.D. = \frac{Q_3 - Q_1}{Q_3 + Q_1}$$

Example 3 : Calculate the *Q.D.* and its co-efficient from the following data:

Class **:**	**5 – 10**	**10 – 15**	**15 – 20**	**20 – 25**	**25 – 30**	**30 – 35**	**35 – 40**
Frequency **:**	**10**	**15**	**25**	**40**	**35**	**20**	**5**

Solution:

Class	f	cumulative frequencies
5-10	10	10
10-15	15	25
15-20	25	50
20-25	40	90
25-30	35	125
30-35	20	145
35-40	5	150
	$N = 150$	

$$Q_1 = L + \frac{\frac{N}{4} - F}{f} \times h \text{ ; } 15 - 20 \text{ is the } Q_1 \text{ class.}$$

$$= 15 + \frac{\frac{150}{4} - 25}{25} \times 5 = 15 + 2.5$$

$$= 17.5$$

$$Q_3 = L + \frac{\frac{3N}{4} - F}{f} \times h \text{ ; } 25 - 30 \text{ is the } Q_3 \text{ class.}$$

$$= 25 + \frac{3 \times \frac{150}{4} - 90}{35} \times 5 = 25 + 3.2 = 28.21$$

$$\therefore \quad Q.D. = \frac{Q_3 - Q_1}{2} = \frac{28.21 - 17.5}{2}$$

$$= 5.4$$

$$\text{co-efficient of } Q.D. = \frac{Q_3 - Q_1}{Q_3 + Q_1} = \frac{28.21 - 17.5}{28.21 + 17.5}$$

$$= \frac{10.7}{45.7} = .23$$

Merits and Demerits of Q.D.

Merits

(i) It is easy to understand and calculate.

(ii) It is not affected by extreme values.

(iii) It is rigidly defined.

Demerits

(i) It is based on 50% observations only.

(ii) It is not suitable for further algebraic treatment.

(iii) It is affected much by fluctuations of sampling.

5.4 MEAN DEVIATION

Mean Deviation of a set of values is the AM of absolute deviations when deviations are taken from an average A. (usually mean, median or mode)

Ungrouped data: If $x_1, x_2, \ldots, x_n$ are the values of a variable then

$$MD = \frac{1}{n}\Sigma|x_i - A|$$

Grouped data: If $x_1, x_2, \ldots, x_n$ are the values of a variable with frequencies $f_1, f_2, \ldots, f_n$ then

$$MD = \frac{1}{N}\Sigma f_i|x_i - A|$$

where A may be mean or median or mode.

$$\text{co-efficient of } MD = \frac{MD}{A}$$

Example 4: The following are the marks of 7 students in Mathematics. Find

(i) *MD* from mean and its co-efficient

(ii) *MD* from median and its co-efficient

18, 26, 15, 20, 17, 12, 25

Solution:

(i)

x	$(x-\bar{x})$	$\lvert x-\bar{x}\rvert$
18	–1	1
26	7	7
15	–4	4
20	1	1
17	–2	2
12	–7	7
25	6	6
$\Sigma x = 133$		**$\Sigma \lvert x - \bar{x}\rvert = 28$**

$$\bar{x} = \frac{\sum x}{n} = \frac{133}{7} = 19$$

$$\therefore MD \text{ from mean} = \frac{\sum |x-\bar{x}|}{n} = \frac{28}{7} = 4$$

$$\text{co-efficient of } MD \text{ from mean} = \frac{MD}{Mean} = \frac{4}{19} = .21$$

(ii) Arranging the marks in ascending order we get,

12, 15, 17, 18, 20, 25, 26

$\therefore$ Median = 18

x	(x–Median)	$\lvert x$ – Median$\rvert$
18	0	0
26	8	8
15	–3	3
20	2	2
17	–1	1
12	–6	6
25	7	7
		$\Sigma \lvert x - \text{Median}\rvert = 27$

$$\therefore MD \text{ from median} = \frac{\sum |x - Median|}{n}$$

$$= \frac{27}{7} = 3.9$$

co-efficient of MD from median $= \dfrac{MD}{Median} = \dfrac{3.9}{18} = .22$

Example 5 : Find *MD* from mean from the following discrete distribution.

***Items*:**	**5**	**15**	**25**	**35**	**45**
***Frequency* :**	**5**	**8**	**15**	**16**	**6**

Solution:

x	f	xf	$(x-\bar{x})$	$\lvert x-\bar{x}\rvert$	$f\lvert x-\bar{x}\rvert$
5	5	25	–22	22	110
15	8	120	–12	12	96
25	15	375	–2	2	30
35	16	560	8	8	128
45	6	270	18	18	108
	N = 50	**Σxf = 1350**			**$\Sigma f\lvert x-\bar{x}\rvert = 472$**

$$\bar{x} = \frac{\sum xf}{N} = \frac{1350}{50} = 27$$

$$\therefore MD \text{ from mean} = \frac{\sum f\left|x-\bar{x}\right|}{N} = \frac{472}{50} = 9.44$$

Example 6 : Calculate *MD* from mean from the following frequency distribution

***Items*:**	**10-20**	**20-30**	**30-40**	**40-50**	**50-60**
***Frequency*:**	**7**	**10**	**20**	**10**	**3**

Solution:

Class	f	x	xf	$\lvert x-\bar{x}\rvert$	$f\lvert x-\bar{x}\rvert$
10–20	7	15	105	18.4	128.8
20–30	10	25	250	8.4	84
30–40	20	35	700	1.6	32
40–50	10	45	450	11.6	116
50–60	3	55	165	21.6	64.8
	N = 50		**Σxf = 1670**		**$\Sigma f\lvert x-\bar{x}\rvert$ = 425.6**

$$\bar{x} = \frac{\sum xf}{N} = \frac{1670}{50} = 33.4$$

$$\therefore MD \text{ from mean} = \frac{\sum f|x-\bar{x}|}{N} = \frac{425.6}{50} = 8.5$$

Merits and Demerits

Merits

(i) It is easy to understand and calculate.

(ii) It is based on all observations.

(iii) It is not very much affected by extreme values.

(iv) It is rigidly defined.

Demerits

(i) Mean deviation ignores the algebraic signs of the deviations and hence it is not capable of further algebraic treatment.

(ii) It is not an accurate measure, particularly when it is calculated from mode.

Note:

1. Mean deviation is least when it is measured from mean.
2. Mean deviation about mean is independent of the change in origin but not of scale.

5.5 STANDARD DEVIATION (SD)

SD of a set of values is the positive square root of the A.M. of squared deviation when deviations are taken from mean. It is denoted by σ (Greek small letter sigma). The square of *SD* in called variance. Thus $SD = \sqrt{\text{variance}}$.

Ungrouped data: If $x_1, x_2, \ldots, x_n$ are the values of x then

$$\sigma = \sqrt{\frac{1}{n}\Sigma(x_i - \bar{x})^2}$$

$$= \sqrt{\frac{1}{n}\Sigma x_i^2 - \bar{x}^2}$$

Grouped data: If $x_1, x_2, \ldots, x_n$ are the values of x with respective frequencies $f_1, f_2, \ldots, f_n$, then

$$\sigma = \sqrt{\frac{1}{N}\Sigma f_i(x_i - \bar{x})^2}$$

$$= \sqrt{\frac{1}{N}\sum f_i x_i^2 - \bar{x}^2}$$

Short cut method for calculating SD

SD can also be calculated by this method.

Ungrouped data

$$\sigma = \sqrt{\frac{\sum d^2}{n} - \left(\frac{\sum d}{n}\right)^2} \text{ where } d = x - A,$$

A = Assumed mean.

Grouped data:

$$\sigma = \sqrt{\frac{\sum fd^2}{N} - \left(\frac{\sum fd}{N}\right)^2} \text{ where } d = x - A$$

In case of continuous data with equal class intervals.

$$\sigma = \sqrt{\frac{\sum fd'^2}{N} - \left(\frac{\sum fd'}{N}\right)^2} \times h \,;\, d' = \frac{x - A}{h.}$$

h = class width

Generally A is taken somewhere in the center of x values.

Example 7: Find the *SD* of the following numbers:

27, 60, 40, 30, 32.

Solution:

x	x^2
27	729
60	3600
40	1600
30	900
32	1024
$\Sigma x = 189$	$\Sigma x^2 = 7853$

$$\bar{x} = \frac{\sum x}{n} = \frac{189}{5} = 37.8$$

$$\therefore \quad \sigma = \sqrt{\frac{1}{n}\sum x^2 - \bar{x}^2} = \sqrt{\frac{7853}{5} - (37.8)^2}$$

$$= \sqrt{1570.6 - 1428.84} = 11.9$$

Using short-cut Method

x	$d = x - 40$	d^2
27	−13	169
60	20	400
40	0	0
30	−10	100
32	−8	64
	$\Sigma d = -11$	$\Sigma d^2 = 733$

$$\therefore \qquad \sigma = \sqrt{\frac{\sum d^2}{n} - \left(\frac{\sum d}{n}\right)^2} = \sqrt{\frac{733}{5} - \left(\frac{-11}{5}\right)^2}$$

$$= \sqrt{146.6 - 4.84} = 11.9$$

Example 8: Find *SD* from the following.

Items:	**5**	**15**	**25**	**35**	**40**
Frequency:	**5**	**8**	**15**	**16**	**6**

Solution:

x	f	$d = x - A$	fd	fd^2
5	5	−20	−100	2000
15	8	−10	−80	800
25	15	0	0	0
35	16	10	160	1600
40	6	15	90	1350
	$\Sigma f = 50$		$\Sigma fd = 70$	$\Sigma fd^2 = 5750$

$$\sigma = \sqrt{\frac{1}{N}\sum fd^2 - \left(\frac{\sum fd}{N}\right)^2} = \sqrt{\frac{5750}{50} - \left(\frac{70}{50}\right)^2}$$

$$= \sqrt{115 - 1.96} = 10.63$$

Example 9: Calculate standard deviation from the following data

class interval :	**10–19**	**20–29**	**30–39**	**40–49**	**50–59**	**60–69**	**70–79**
Frequency :	**3**	**61**	**223**	**137**	**53**	**19**	**4**

Solution :

Class interval	f	x	d'	fd'	fd'^2
10–19	3	14.5	–3	–9	27
20–29	61	24.5	–2	–122	244
30–39	223	34.5	–1	–223	223
40–49	137	44.5	0	0	0
50–59	53	54.5	1	53	53
60–69	19	64.5	2	38	76
70–79	4	74.5	3	12	36
	$N = 500$			**$\Sigma fd' = -251$**	**$\Sigma fd'^2 = 659$**

$$\sigma = \sqrt{\frac{\sum fd'^2}{N} - \left(\frac{\sum fd'}{N}\right)^2} \times h$$

$$= \sqrt{\frac{659}{500} - \left(\frac{-251}{500}\right)^2} \times 10$$

$$= \sqrt{1.318 - .252} \times 10 = 1.032 \times 10 = 10.32$$

5.6 PROPERTIES OF SD

1. *SD* is the minimum value of root mean square deviation.

Proof: For a frequency distribution *SD* is

$$\sigma = \sqrt{\frac{1}{N}\sum f_i\left(x_i - \bar{x}\right)^2}$$

Root mean square deviation is

$$s = \sqrt{\frac{1}{N}\sum f_i\left(x_i - A\right)^2}\text{ , A is any arbitrary number.}$$

Now

$$s^2 = \frac{1}{N}\sum f_i\left(x_i - A\right)^2$$

$$= \frac{1}{N}\sum f_i\left(x_i - \bar{x} + \bar{x} - A\right)^2$$

$$= \frac{1}{N}\sum f_i\left[\left(x_i - \bar{x}\right)^2 + \left(\bar{x} - A\right)^2 + 2\left(x_i - \bar{x}\right)\left(\bar{x} - A\right)\right]$$

$$= \frac{1}{N}\sum f_i\left(x_i - \bar{x}\right)^2 + \left(\bar{x} - A\right)^2 \frac{1}{N}\sum f_i + 2\left(\bar{x} - A\right).$$

$$\frac{1}{N}\sum f_i\left(x_i - \bar{x}\right)$$

But $\sum f_i\left(x_i - \bar{x}\right) = 0$, by the properties of AM

$$\therefore \qquad s^2 = \frac{1}{N}\sum f_i\left(x_i - \bar{x}\right)^2 + \left(\bar{x} - A\right)^2$$

$$\Rightarrow \qquad s^2 = \sigma^2 + d^2 \;;\; d = \bar{x} - A$$

s^2 will be least when $d = 0$

$$\Rightarrow \qquad \bar{x} - A = 0$$

$$\Rightarrow \qquad \bar{x} = A$$

Hence mean square deviation is the minimum when $A = \bar{x}$. Consequently, SD is the minimum value of root mean square deviation.

2. SD is independent of change in origin but not of scale.

Proof: Let x be the original variable taking values $x_1, x_2, \ldots, x_n$ then

$$\sigma_x = \sqrt{\frac{1}{n}\Sigma\left(x_i - \bar{x}\right)^2}$$

After changing origin and scale the new variable becomes

$$u_i = \frac{x_i - A}{h}\;; A \text{ and } h \text{ are constants}$$

$$\Rightarrow \qquad x_i = A + h\, u_i$$

$$\Rightarrow \qquad \bar{x} = A + h\, \bar{u}$$

$$\therefore \qquad x_i - \bar{x} = h\left(u_i - \bar{u}\right)$$

$$\text{Now,} \qquad \sigma_x^2 = \frac{1}{n}\sum\left(x_i - \bar{x}\right)^2 = \frac{1}{n}\sum\left\{h\left(u_i - \bar{u}\right)\right\}^2$$

$$= h^2\,\frac{1}{n}\sum\left(u_i - \bar{u}\right)^2 = h^2\,\sigma_u^2$$

$$\therefore \qquad \sigma_x = h\,\sigma_u\,,$$

which involves h and not A. Thus we see that SD is independent of change in origin but not of scale.

3. For any discrete distribution SD is not less than MD from mean.

Proof: Let the discrete frequency distribution be

$$x \;:\; x_1\, x_2 \,.....x_n$$
$$f \;:\; f_1\, f_2 \,.....f_n$$

we want to show

$SD \not< MD$ from mean

$\Rightarrow \quad (SD)^2 \geq (MD \text{ from mean})^2$

$$\Rightarrow \quad \frac{1}{N}\sum fi\left(x_i - \bar{x}\right)^2 \geq \left[\frac{1}{N}\sum fi\left|x_i - \bar{x}\right|\right]^2$$

Putting $\left|x_i - \bar{x}\right| = z_i$ we write.

$$\frac{1}{N}\sum f_i z_i^2 \geq \left[\frac{1}{N}\sum f_i z_i\right]^2$$

$$\Rightarrow \quad \frac{1}{N}\sum f_i z_i^2 - \left[\frac{1}{N}\sum f_i z_i\right]^2 \geq 0$$

$$\Rightarrow \quad \frac{1}{N}\sum f_i\left(z_i - \bar{z}\right)^2 \geq 0$$

$\Rightarrow \quad \sigma_z^2 \geq 0$, which is always true. Hence the result.

4. Combined SD: Let there be two sets of values of x with n_1 and n_2 observations and let $\bar{x}_1$ and $\bar{x}_2$ be their means and let σ_1 and σ_2 be their standard deviations. Then the SD of the combined series

$$\sigma^2 = \frac{1}{n_1 + n_2}\left[n_1\left(\sigma_1^2 + d_1^2\right) + n_2\left(\sigma_2^2 + d_2^2\right)\right]$$

where $\quad d_1 = \bar{x}_1 - \bar{x}$

$\qquad\qquad d_2 = \bar{x}_2 - \bar{x}$

Solution:

Let $x_{11}, x_{12}, \ldots, x_{1n_1}$ and $x_{21}, x_{22}, \ldots, x_{2n_2}$ be two series, then

$$\bar{x}_1 = \frac{1}{n_1}\sum_{1=1}^{n_1} x_{1i} \qquad \Rightarrow n_1\bar{x}_1 = \sum_{1=1}^{n_1} x_{1i}$$

$$\bar{x}_2 = \frac{1}{n_2}\sum_{j=1}^{n_2} x_{2j} \qquad \Rightarrow n_2\bar{x}_2 = \frac{1}{n_2}\sum_{j=1}^{n_2} x_{2j}$$

$$\sigma_1^2 = \frac{1}{n_1}\sum_{j=1}^{n_1}\left(x_{1i} - \bar{x}_1\right)^2 \qquad \Rightarrow n_1\sigma_1^2 = \sum_{j=1}^{n_1}\left(x_{1i} - \bar{x}_1\right)^2$$

$$\sigma_2^2 = \frac{1}{n_2}\sum_{j=1}^{n_2}\left(x_{2j} - \bar{x}_2\right)^2 \quad \Rightarrow n_2\sigma_2^2 = \sum_{j=1}^{n_2}\left(x_{2j} - \bar{x}_2\right)^2$$

The mean $\bar{x}$ of the combined series

$$\bar{x} = \frac{1}{n_1 + n_2}\left[\sum_{i=1}^{n_1} x_{1i} + \sum_{j=1}^{n_2} x_{2j}\right]$$

$$= \frac{n_1\bar{x}_1 + n_2\bar{x}_2}{n_1 + n_2}$$

The variance σ^2 of the combined series

$$\sigma^2 = \frac{1}{n_1 + n_2}\left[\sum_{i=1}^{n_1}\left(x_{1i} - \bar{x}\right)^2 + \sum_{j=1}^{n_2}\left(x_{2j} - \bar{x}\right)^2\right] \qquad (1)$$

Now, $\sum_{i=1}^{n_1}\left(x_{1i} - \bar{x}\right)^2$

$$= \sum_{i=1}^{n_1}\left(x_{1i} - \bar{x}_1 + \bar{x}_1 - \bar{x}\right)^2$$

$$= \sum_{i=1}^{n_1}\left[\left(x_{1i} - \bar{x}_1\right)^2 + \left(\bar{x}_1 - \bar{x}\right)^2 + 2\left(x_{1i} - \bar{x}_1\right)\left(\bar{x}_1 - \bar{x}\right)\right]$$

$$= \sum_{i=1}^{n_1}\left(x_{1i} - \bar{x}_1\right)^2 + n_1\left(\bar{x}_1 - \bar{x}\right)^2 + 2\left(\bar{x}_1 - \bar{x}\right)\sum_{i=1}^{n_1}\left(x_{1i} - \bar{x}_1\right)$$

$$= n_1\sigma_1^2 + n_1\left(\bar{x}_1 - \bar{x}\right)^2 + 0 \text{ [By the property of AM]}$$

$$= n_1\sigma_1^2 + n_1 d_1^2 \text{ where } d_1 = \bar{x}_1 - \bar{x}$$

Similarly we get,

$$\sum_{j=1}^{n_2}\left(x_{2j} - \bar{x}\right)^2 = n_2\sigma_2^2 + n_2 d_2^2 \text{ where } d_2 = \bar{x}_2 - \bar{x}$$

$\therefore$ from (i) we get,

$$\sigma^2 = \frac{1}{n_1 + n_2}\left[n_1\sigma_1^2 + n_1 d_1^2 + n_2\sigma_2^2 + n_2 d_2^2\right]$$

$$= \frac{1}{n_1 + n_2}\left[n_1\left(\sigma_1^2 + d_1^2\right) + n_2\left(\sigma_2^2 + d_2^2\right)\right]$$

Example 10: For a group containing 100 observations, the arithmetic mean and standard deviation are 8 and $\sqrt{10.5}$ respectively. For 50 observations selected from these 100 observations, the mean and standard deviation are 10 and 2 respectively. Calculate mean and standard deviation for the other half. (AHSEC 2001)

Solution:

Given $n = 100,\ \bar{x} = 8$

$\sigma = \sqrt{10.5}$

$\Rightarrow \quad \sigma^2 = 10.5,\ n_1 = 50$

$\bar{x}_1 = 10,\ \sigma_1 = 2$

$\bar{x}_2 = ?,\ \sigma_2 = ?$

Now,

$$\bar{x} = \frac{n_1\bar{x}_1 + n_2\bar{x}_2}{n_1 + n_2}$$

$$\Rightarrow \quad 8 = \frac{50 \times 10 + 50\bar{x}_2}{100} \qquad [\because n_2 = 100 - 50 = 50]$$

$\Rightarrow \quad 50\,\bar{x}_2 = 800 - 500$

$\Rightarrow \quad \bar{x}_2 = 6$

$\therefore \quad d_1 = \bar{x}_1 - \bar{x} = 10 - 8 = -2 \Rightarrow d_1^2 = 4$

$d_2 = \bar{x}_2 - \bar{x} = 6 - 8 = -2 \Rightarrow d_2^2 = 4$

$$\therefore \quad \sigma^2 = \frac{1}{n_1 + n_2}\left[n_1\left(\sigma_1^2 + d_1^2\right) + n_2\left(\sigma_2^2 + d_2^2\right)\right]$$

$\Rightarrow \quad 100 \times 10.5 = 50(4 + 4) + 50(\sigma_2^2 + 4)$

$\Rightarrow \quad 50\sigma_2^2 = 1050 - 600.$

$\Rightarrow \quad \sigma_2^2 = 9$

$\Rightarrow \quad \sigma_2 = 3 \qquad (\because \sigma \geq 0)$

Example 11: Mean and *S.D.* of 18 observations are found to be 7 and 4 respectively. But on comparing the original data it was found that an observation 12 was miscopied as 21 in the calculations. Calculate correct mean and *S.D.*

Solution:

Given $n = 18,\ \bar{x} = 7,\ \sigma = 4$

$$\bar{x} = \frac{\sum x}{n} \Rightarrow \sum x = n\,\bar{x} = 18 \times 7 = 126$$

Also, $\sigma^2 = \frac{1}{n}\sum x^2 - \bar{x}^2$

$$\Rightarrow \quad \sum x^2 = n\left(\sigma^2 + \bar{x}^2\right) = 18\,(16 + 49) = 1170$$

$$\text{corrected } \sum x = 126 - 21 + 12 = 117$$

$$\text{corrected} \sum x^2 = 1170 - (21)^2 + (12)^2$$
$$= 1170 - 441 + 144 = 873$$

$$\therefore \text{corrected mean} = \frac{117}{18} = 6.5$$

$$\text{corrected } \sigma^2 = \frac{873}{18} - (6.5)^2 = 48.5 - 42.25$$
$$= 6.25$$
$$\Rightarrow \text{corrected } \sigma = 2.5$$

Example 12: Find the mean and standard deviation of first *n* natural numbers.

(AHSEC 1998)

Solution: The values of x are 1, 2, 3, ..., n

$$\text{Mean} = \frac{\sum x}{n}$$
$$= \frac{1+2+...+n}{n}$$
$$= \frac{n(n+1)}{2}.\frac{1}{n} = \frac{n+1}{2}$$

$$SD = \sqrt{\frac{\sum x^2}{n} - \bar{x}^2} = \sqrt{n\frac{(n+1)(2n+1)}{6}\frac{1}{n} - \left(\frac{n+1}{2}\right)^2}$$
$$= \sqrt{\frac{n+1}{2}\left\{\frac{2n+1}{3} - \frac{n+1}{2}\right\}}$$
$$= \sqrt{\frac{n+1}{2}.\frac{4n+2-3n-3}{6}}$$
$$= \sqrt{\frac{n+1}{2}.\frac{n-1}{6}} = \sqrt{\frac{n^2-1}{12}}$$

Example 13: If $\sum x^2 = 256$, $\bar{x} = 0$ and n = 16 find σ.

Solution:

$$\sigma = \sqrt{\frac{1}{n}\sum x^2 - \bar{x}^2}$$
$$= \sqrt{\frac{256}{16} - 0} = \sqrt{16}$$
$$= 4$$

Example 14: If SD of x is 5, find SD of (i) $2x-3$ and (ii) $\frac{x}{5}+1$

Solution: We know that when $x = a + hu$, $\sigma_x = h\sigma_u$

(i) $S.D.\ (2x-3) = 2\ SD\ (x) = 2 \times 5 = 10$

(ii) $SD\left(\frac{x}{5}+1\right) = \frac{1}{5}\ SD(x) = \frac{1}{5} \times 5 = 1$

Example 15: Find the SD of two values 6 and 10

Solution :

$$\text{SD of two values} = \frac{1}{2} \times \text{ Difference of two values}$$

$$= \frac{1}{2}\ (10-6) = 2$$

Example 16: Find the SD of the numbers 6, 6, 8, 8.

Solution: SD of 6, 6, 8, 8

$$= \text{SD of 6 and 8}$$

$$= \frac{1}{2} \times \text{ Difference of two values}$$

$$= \frac{1}{2}.(8-6)$$

$$= 1$$

Example 17: SD of two quantities is 5. If one of them is 12 find the other.

Solution: Let the other be x

Now, $\frac{1}{2}(12-x) = 5$

$\Rightarrow \quad 12-x = 10$

$\Rightarrow \quad x = 2$ when $x < 12$

Again $\frac{1}{2}(x-12) = 5$

$\Rightarrow \quad x - 12 = 10$

$\Rightarrow \quad x = 22$ when $x > 12$

Example 18: The mean of 5 observations is 4.4 and the variance is 8.24. If three of the five observations are 1, 2 and 6, find the other two.

Solution: Let the two observations be x_1 and x_2

Since $\bar{x} = \frac{\sum x}{n}$

$\Rightarrow \quad \sum x = n\bar{x}$

$\Rightarrow 1 + 2 + 6 + x_1 + x_2 = 5 \times 4.4$

$\Rightarrow \quad x_1 + x_2 = 13$ (i)

Again, $\sigma^2 = \frac{1}{n}\sum x^2 - \bar{x}^2$

$\Rightarrow \quad \sum x^2 = n\left(\sigma^2 + \bar{x}^2\right)$

$\Rightarrow 1^2 + 2^2 + 6^2 + x_1^2 + x_2^2 = 5\ (8.24 + 19.36)$

$\Rightarrow \quad x_1^2 + x_2^2 = 97$ (ii)

Since $(x_1 - x_2)^2 = (x_1 + x_2)^2 - 2x_1x_2$

$= 13^2 - 2\left[(x_1 + x_2)^2 - (x_1^2 + x_2^2)\right]$

$= 169 - 2[169 - 97]$

$= 169 - 144$

$= 25$

$\therefore \quad x_1 + x_2 = 5$ or -5 (iii)

from (i) and (iii) we get,

$x_1 = 9$ or 4

$\Rightarrow \quad x_2 = 4$ or 9

$\therefore$ The two observations are 4 and 9.

5.7 RELATIONSHIP BETWEEN DIFFERENT MEASURES OF DISPERSION

For moderately asymmetrical distribution or symmetrical distribution, the following relations hold

1. $Q.D. = \frac{2}{3} SD$

2. MD from mean $= \frac{4}{5} SD$

3. MD from mean $= \frac{6}{5} QD.$

Example 19: For a group of items the value of *QD* is 30. Find out the most likely value of variance.

Solution:

$$Q.D. = \frac{2}{3}\text{SD} \Rightarrow \text{SD} = \frac{2}{3}Q.D.$$

$$= \frac{3}{2} \times 30 = 45$$

$\therefore$ Variance = $(\text{SD})^2 = 2025$

Example 20: What is the probable value of *MD* where $Q_3 = 40$ and $Q_1 = 15$.

Solution:

$$QD = \frac{Q_3 - Q_1}{2} = \frac{40-15}{2} = 12.5$$

Now, $$MD = \frac{6}{5}QD$$

$\Rightarrow$ $$MD = \frac{6 \times 12.5}{5} = 15$$

Merits and Demerits of *SD*.

Merits:

SD is the best measure of dispersion because of the following merits.

(i) It is based on all observations.

(ii) It is suitable for further mathematical treatment.

(iii) It is less affected by fluctuations of sampling than other measures of dispersion.

(iv) It is rigidly defined.

(v) It has wide applications in statistical theory.

Demerits:

(i) It is difficult to calculate

(ii) It is affected by extreme values.

5.8 USES OF DIFFERENT MEASURES OF DISPERSION

Range: Range is widely used in statistical Quality control (*S.Q.C*), weather forecasting, stock market fluctuations etc.

Quartile Deviation: QD is used in elementary descriptive statistics where the distribution is open end. In case of attributes, QD is more appropriate.

Mean Deviation: MD is frequently used by the economic and business statisticians due to its simplicity. It is used in computing the distribution of wealth in a community.

Standard Deviation: SD is used in statistical theory such as skewness, kurtosis, correlation and regression analysis, sampling theory and tests of significance etc. It is the most widely used measure of dispersion.

5.9 CO-EFFICIENT OF VARIATION

To compare the variability of two series, the relative measure of dispersion based on standard deviation called the co-efficient of standard deviation is used.

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

The co-efficient of SD multiplied by 100 gives the co-efficient of variation. This was introduced by Karl Pearson.

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

It indicates the relationship between the SD and the AM expressed in percentage. This is a pure number independent of units.

For comparing the variability in two sets of data we compute the *CV* for each. The series having higher *CV* has higher degree of variability.

Example 21: If $n = 5$, $\bar{x} = 4$, $\sum x^2 = 90$, find the co-efficient of variation

Solution:

$$\sigma^2 = \frac{1}{n}\sum x^2 - \bar{x}^2 = \frac{1}{5} \times 90 - 16$$

$$= 18 - 16 = 2$$

$$\therefore \qquad \sigma = 1.41, \text{CV} = \frac{\sigma}{\bar{x}} \times 100$$

$$= \frac{1.41 \times 100}{4} = 35.25\%$$

Example 22: For a distribution, the co-efficient of variation is 35.3% and the value of arithmetic mean is 4. Find the value of SD.

Solution:

$$\text{CV} = \frac{\sigma}{\bar{x}} \times 100$$

$$\Rightarrow \qquad \sigma = \frac{CV \times \bar{x}}{100}$$

$$= \frac{35.3 \times 4}{100} = 1.412$$

Example 23: Co-efficient of variation of two series are 60% and 80%. Their standard deviations are 20 and 16. What are their arithmetic means?

Solution:

For the first series CV = $\frac{\sigma}{\bar{x}} \times 100$

$$\Rightarrow \quad \bar{x} = \frac{\sigma \times 100}{CV}$$

$$= \frac{100 \times 20}{60} = 33.3$$

For the second series

$$\bar{x} = \frac{100 \times 16}{80} = 20$$

Example 24: Find the co-efficient of variation from the following data:

Wages:	12-13	13-14	14-15	15-16	16-17	17-18	18-19
No. of persons :	15	30	44	60	30	14	7

Solution:

Wages	f	x	$d = x$-15.5	fd	fd^2
12–13	15	12.5	–3	–45	135
13–14	30	13.5	–2	–60	120
14–15	44	14.5	–1	–44	44
15–16	60	15.5	0	0	0
16–17	30	16.5	1	30	30
17–18	14	17.5	2	28	56
18–19	7	18.5	3	21	63
	$N = 200$			$\Sigma fd = -70$	$\Sigma fd^2 = 448$

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

$$= \sqrt{\frac{448}{200} - \left(\frac{-70}{200}\right)^2}$$

$$= \sqrt{2.24 - .1225} = 1.46$$

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

$$= 15.5 + \frac{(-70)}{200}$$

$$= 15.5 - .35$$

$$= 15.15$$

$$CV = \frac{\sigma}{\bar{x}} \times 100$$

$$= \frac{1.46}{15.15} \times 100 = 9.64\%$$

Example 25: Calculate co-efficient of variation of the two section and find which section is more consistent:

Marks	Section *A*	Section *B*
	f	*f*
10–20	2	3
20–30	5	7
30–40	10	8
40–50	5	5
50–60	3	2

Solution:

Section A

Marks	f	x	$d' = \frac{x-35}{10}$	fd'	fd'^2
10–20	2	15	–2	–4	8
20–30	5	25	–1	–5	5
30–40	10	35	0	0	0
40–50	5	45	1	5	5
50–60	3	55	2	6	12
	$N = 25$			$\Sigma fd' = 2$	$\Sigma fd'^2 = 30$

$$\bar{x} = A + \frac{\sum fd'}{N} \times h$$

$$= 35 + \frac{2}{25} \times 10$$

$$= 35 + .8 = 35.8$$

$$\sigma = \sqrt{\frac{\sum fd'^2}{N} - \left(\frac{\sum fd'}{N}\right)^2} \times h$$

$$= \sqrt{\frac{30}{25} - \left(\frac{2}{25}\right)^2} \times 10$$

$$= \sqrt{1.2 - .0064} \times 10 = 10.93$$

$$\therefore \qquad CV = \frac{\sigma}{\bar{x}} \times 100$$

$$= \frac{10.93}{35.8} \times 100 = 30.53\%$$

Section –B

Marks	f	x	$d' = \frac{x-35}{10}$	fd'	fd'^2
10–20	3	15	–2	–6	12
20–30	7	25	–1	–7	7
30–40	8	35	0	0	0
40–50	5	45	1	5	5
50–60	2	55	2	4	8
	$N = 25$			**$\Sigma fd' = -4$**	**$\Sigma fd'^2 = 32$**

$$\bar{x} = A + \frac{\sum fd'}{N} \times h$$

$$= 35 + \frac{(-4)}{25} \times 10$$

$$= 35 - 1.6 = 33.4$$

$$\sigma = \sqrt{\frac{\sum fd'^2}{N} - \left(\frac{\sum fd'}{N}\right)^2} \times h$$

$$= \sqrt{\frac{32}{25} - \left(\frac{-4}{25}\right)^2} \times 10$$

$$= \sqrt{1.28 - .0256} \times 10$$

$$= 1.12 \times 10 = 11.2$$

$$CV = \frac{\sigma}{\bar{x}} \times 100$$

$$= \frac{11.2}{33.4} \times 100 = 33.53\%$$

Since $CV_A < CV_B$, Section A is more consistent than Section B.

5.10 LORENZ'S CURVE

A graphic method of studying dispersion was adopted by Dr. Max. O. Lorenz who studied the distribution of wealth and income and the curve is known as Lorenz curve. It is commonly used to show the equality of income or wealth in a country and sometimes to make comparisons between countries.

In drawing this curve the size of item and frequency both are cumulated and converted to percentages. These percentages are plotted on a graph paper. If the points lie in a straight line then there is no dispersion, the straight line is called line of equal distribution. The curve would be away from the line of equal distribution if there exists dispersion. The further away is the curve from the line of equal distribution, the greater would be the dispersion.

The drawback of this method is that no numerical measurement of dispersion is possible to get.

Example 26: Draw a Lorenz curve from the following data:

Income ('000 Rs.)	**No. of Persons**	**('000)**
	Group *A*	**Group *B***
10	**8**	**15**
20	**7**	**6**
40	**5**	**2**
50	**3**	**1**
80	**2**	**1**

Solution:

Here income and frequencies are made cumulative and then percentages are calculated by taking the respective totals as 100.

Percentages relating to the number of persons would be shown on the x-axis from left to right from 100 to 0. The income percentages would be shown on the y axis from 0 and go upto 100 at the top. Then this point is joined by a line to the point of origin. This line in called the line of equal distribution. We see that in Group A the distribution is uneven and in Group B the distribution is still more uneven. The variation in Group B is greater than the variation in Group A.

Income			Group A			Group B		
Rupees ('000)	Cumulated Income	Cumulated Percentage	No. of Persons ('000)	Cumulated Number	Cumulated Percentage	No. of Persons	Cumulated Number ('000)	Cumulated Percentages
10	10	5	8	8	32	15	15	60
20	30	15	7	15	60	6	21	84
40	70	35	5	20	80	2	23	92
50	120	60	3	23	92	1	24	96
80	200	100	2	25	100	1	25	100

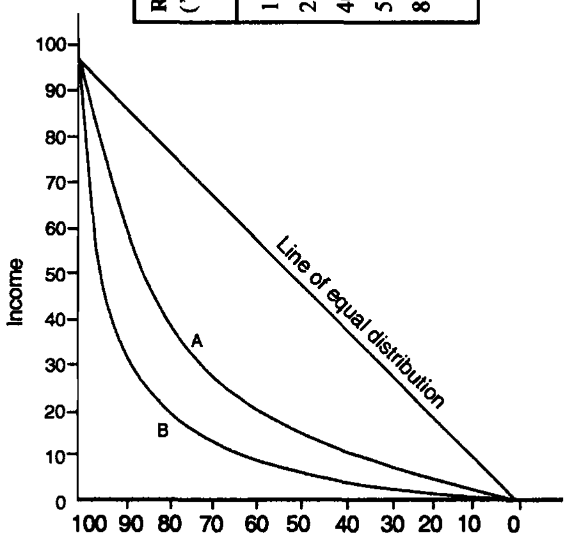


Fig. 5.1

EXERCISE–5

1. What do you mean by measures of dispersion ? **(AHSEC 1990, 1992, 2002)**
2. Which measure of dispersion is the best and why? **(AHSEC 1990, 1992)**
3. Which is greater mean deviation or standard deviation? Justify your comment by calculating the following distribution. **(AHSEC 2002)**

Class Interval	Frequency
0–5	2
5–10	7
10–15	13
15–20	6
20–25	2

[Ans. *SD* > *MD*]

4. Show that standard deviation is the least root mean square deviation. **(AHSEC 1991)**
5. Show that mean deviation about mean is not greater than standard deviation. **(AHSEC 1993)**
6. Define standard deviation. Prove that *SD* of a variable whose values are all equal is zero. **(AHSEC1996)**
7. Define *SD*. For a distribution of height mean = 150 cm and S.D. = 5 cm. For a distribution of weight mean = 55 kg and *SD* = 2kg. Compare the dispersion of the distributions. **(AHSEC 1999)**
8. Calculate *SD* of the following distribution.

Wages per week (in Rs.) :	32–34	35–37	38–40	41–43	44–46
No. of workers :	14	62	99	18	7

(AHSEC 2000)

[Ans. 2.57]

9. Find co-efficient of variation from the following distribution.

x:	10	11	12	13	14
f:	3	12	18	12	5

(AHSEC 2000)

[Ans. 8.77%]

10. Co- efficient of variation of two series are75% and 90%. Their standard deviations are respectively 15 and 20. What are their arithmetic means? **(AHSEC 2001)**

[Ans. 20, 22.22]

11. Define standard deviation. Why is standard deviation called ideal measure of dispersion? **(SHSEC 1993, 2001, 2003)**

12. Calculate co-efficient of variation from the following data:

Marks	No. of Students
Below 20	8
" 40	20
" 60	50
" 80	70
" 100	80

(AHSEC 2003)
[Ans. 22.27%]

13. Find the mean and variance of the following : **(AHSEC 2004)**

Measurement	No. of Articles
More than 80	5
" 70	14
" 60	34
" 50	65
" 40	110
" 30	150
" 20	170
" 10	176
" 0	180

[Ans. 45.2, 282.95]

14. Write a short note on co-efficient of variation. **(AHSEC 2004, 2008)**
15. Examine the correctness of the statement : "Standard deviation is not affected by the change of origin" **(AHSEC 2005)**
16. Find the range of variation of the following values

Weight (kg.) :40, 51, 47, 39, 60, 48, 64, 51, 57

(AHSEC 2005)
[Ans. 25]

17. What do you mean by absolute and relative measures of dispersion?
18. Write short note on Lorenz Curve.
19. Calculate mean and *SD* of the following distribution. **(AHSEC 1990)**

Class :	1–4	5–8	9–12	13–16	17–20
Frequency :	5	14	21	13	7

[Ans 10.7, 4.47]

20. Given $u = \frac{x - 10}{5}$ and $\bar{u} = 15$, $\sigma_u = 3$, find $\bar{x}$ and σ_u. **(AHSEC 1998)**

[Ans. $\bar{x} = 85$, $\sigma_x = 15$]

21. Define Co-efficient of variation. State its applications. **(AHSEC 1991)**

22. The mean and variance of x are 50 and 9 respectively, if $u = \frac{x-5}{3}$, then find the mean and variance of u. **(AHSEC 1992)**

[Ans. 15, 1]

23. For a set of n_1 Observations mean $= m_1$, $SD = \sigma_1$. For another set of n_2 observations mean $= m_2$ and $SD = \sigma_2$. Find the mean and variance of both the sets combined. **(AHSEC 1992)**

24. For a distribution, the CV is 22.5% and the value of average is 7.5. Find out the value of SD.

(AHSEC 1996)

[Ans. 1.6875]

25. State whether the following statements are true (T) or False (F).

(AHSEC 1996)

(i) Variance is always non-negative. **(AHSEC 2008)**

(ii) Mean Deviation is minimum when calculated from the median.

(iii) Mean, SD and CV has the same unit

(vi) If each value in a set of a 5 observations is 10, then its mean is 10 and variance is 1. **[Ans. (i) T, (ii) T, (iii) F, (iv) F]**

26. Fill in the blanks:

(i) Co-efficient of variation is .... of unit of measurement.

(AHSEC 1998)

(ii) Mean Deviation about .... is always least.

(AHSEC 1998)

(iii) The concept of Lorenz Curve was put forward by ....

(iv) Variance is equal to the square of ....

(v) MD = ... SD.

(vi) SD is denoted by ....

Ans.(i) independent, (ii) median, (ii) Max O. Lorenz, (iv) SD, (v) $\frac{4}{5}$, (vi) σ.

(vii) Standard deviation is a measure of _____________. **[AHSEC 2007]**

[Ans. Dispersion]

(viii)The standard deviation is affected by the change of ___________.

[AHSEC 2006]

[Ans. scale]

27. Distinguish between measures of central tendency and measures of dispersion. **[AHSEC 2008]**

28. If $U = \dfrac{x-a}{h}$, then prove that $\sigma_x = h\sigma_u$ **[AHSEC 2008]**

29. The AM and SD of 100 observations are 50 and 10 respectively, Find the new AM and SD if 2 is added to each observation. **[AHSEC 2008]**

[Ans. 52, 10]

30. What are the different measures of dispersion. **[AHSEC 2006]**

31. 'Standard deviation is the best measure of dispersion' Explain. **[AHSEC 2006]**

6

SKEWNESS AND KURTOSIS

The central tendency and dispersion are two important measures to characterise a distribution. But we can not get a complete idea of a distribution unless we know skewness and kurtosis in addition to central tendency and dispersion. Skewness enables us to study the symmetry or asymmetry of the distribution while kurtosis refers to the flatness or peakedness of the curve of the distribution.

6.1 SKEWNESS

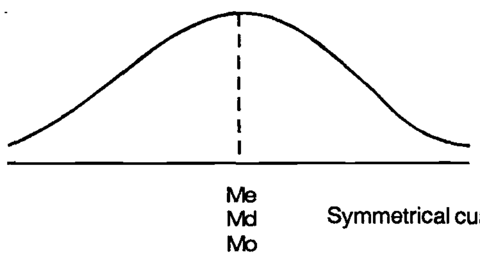
The absence of symmetry in a distribution is known as skewness. When a distribution is not symmetrical, it is called skewed or asymmetrical. The symmetry of a distribution shows that at equal distances on either side of the centre, the frequencies are evenly distributed. We study skewness to have an idea about the shape of the curve which we can draw with the help of given frequency distribution.

In a symmetrical distribution

(i) Mean = Median = Mode

(ii) Q_3 – Median = Median – Q_1

(iii) The curve of the given data is symmetrical.



Symmetrical curve.

Figure 6.1

A distribution is said to be positively skewed if the longer tail lies towards the right side. Here mean > median > mode.

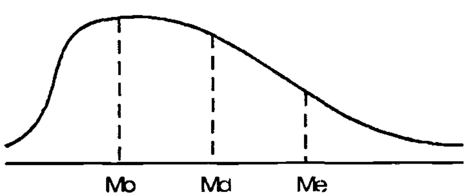


Figure 6.2 Positive Skewness

Again a distribution is said to be negatively skewed if the longer tail lies towards the left side. Here mean < median < mode.

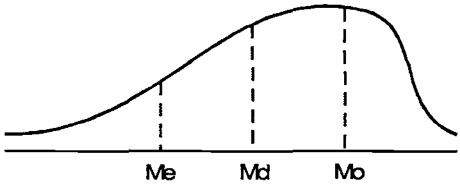


Figure 6.3 Negative Skewness

6.2 MEASURES OF SKEWNESS

Absolute measure

The larger the distance between the mean and the mode, the greater is the skewness.

Skewness (Sk) = Mean – Mode

This measure is not suitable for comparison

Relative Measure

1. Karl Pearson's co-efficient of skewness

It is given by

$$Sk = \frac{\text{Mean-mode}}{SD}$$

If mode is not well defined we use the formula

$$Sk = \frac{3(\text{Mean} - \text{Median})}{S.D.}$$

Here $-3 \leq Sk \leq 3$. In practice it is rarely obtained.

2. Bowley's co-efficient of skewness

This is based on quartiles and is given by

$$Sk = \frac{Q_3 + Q_1 - 2Q_2}{Q_3 - Q_1}$$

Here $-1 \leq Sk \leq 1$

The co-efficient of skewness is a pure number and is zero for a symmetrical distribution.

3. Kelly's co-efficient of Skewness

$$Sk = \frac{P_{90} + P_{10} - 2P_{50}}{P_{90} - P_{10}}$$

$$= \frac{D_9 + D_1 - 2D_5}{D_9 - D_1}$$

Where P_i and D_i denote i^{th} percentile and decile respectively.

6.3 KURTOSIS

Kurtosis means peakedness of a distribution. Measuring kurtosis means to examine the extent of peakedness of a distribution as compared to the normal or symmertical distribution. Normal curve is neither flat nor peak and is called Mesokurtic. The curves with greater peakedness than the normal curve are called Leptokurtic. The curves which are flatter than the normal curve are called Platykurtic. We can say that Kurtosis tells us about the middle portion of a frequency curve.

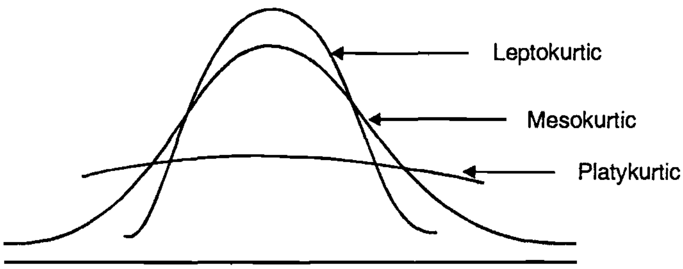


Figure 6.4

Kelly has given a measure of kurtosis based on percentiles. His formula is

$$Sk = \frac{P_{75} - P_{25}}{P_{90} - P_{10}}$$

If $Sk > .26315$, the distribution is Platykurtic

$Sk < .26315$, the distribution is Leptokurtic.

Example 1: For a distribution Karl Pearson's co-efficient of skewness is .32, SD is 6.5 and mean is 29.6. Find mode and median.

Solution:

Given $S_k = .32$

S.D. = 6.5

Mean = 29.6

$$Sk = \frac{\text{Mean-Mode}}{SD}$$

$$\Rightarrow \quad .32 = \frac{29.6 - \text{Mode}}{6.5}$$

$\Rightarrow$ Mode = 27.52

Since Mode = 3 Median − 2 Mean

$$\Rightarrow \quad \text{Median} = \frac{\text{Mode+2 Mean}}{3}$$

$$= \frac{27.52 + 2 \times 29.6}{3}$$

$$= 28.91$$

Example 2: Karl Pearson's co-efficient of skewness of a distribution is .64, its mean is 82 and mode 50. Find the *SD*.

Solution:

$$Sk = \frac{\text{Mean - Mode}}{SD}$$

$$\Rightarrow \quad .64 = \frac{82 - 50}{SD}$$

$$\Rightarrow \quad \text{SD} = \frac{32}{.64}$$

$$\Rightarrow \quad \text{SD} = 50$$

Example 3: For a frequency distribution the co-efficient of skewness based on quartiles is .6. If the sum of the upper and lower quartiles is 100 and median is 38, find the value of the upper quartile.

Solution:

Given $Sk = .6, Q_3 + Q_1 = 100, Q_2 = 38$

$Q_3 = ?$

Since $$Sk = \frac{Q_3 + Q_1 - 2Q_2}{Q_3 - Q_1}$$

$$\Rightarrow \quad .6 = \frac{100 - 2 \times 38}{Q_3 - Q_1}$$

$$\Rightarrow \quad Q_3 - Q_1 = \frac{24}{.6}$$

$$\Rightarrow \quad Q_3 - Q_1 = 40$$

$$\Rightarrow \quad Q_3 - (100 - Q_3) = 40$$

$$\Rightarrow \quad 2Q_3 = 140$$

$$\Rightarrow \qquad Q_3 = 70$$

Hence the value of the upper quartile is 70.

Example 4: The following are the marks of 100 students in an examination. Calculate Karl Pearson's coefficient of skewness.

Marks	No. of Students
More than 0	150
" 10	140
" 20	100
" 30	80
" 40	80
" 50	70
" 60	30
" 70	14
" 80	0

Solution:

Let us construct the following table

Marks	f	x	cumulative frequency	$d' = \frac{x-35}{10}$	fd'	fd'^2
0–10	10	5	10	–3	–30	90
10–20	40	15	50	–2	–80	160
20–30	20	25	70	–1	–20	20
30–40	0	35	70	0	0	0
40–50	10	45	80	1	10	10
50–60	40	55	120	2	80	160
60–70	16	65	136	3	48	144
70–80	14	75	150	4	56	224
	$N = 150$				$\sum fd' = 64$	$\sum fd'^2 = 808$

Hence mode is not well defined since there are two modal classes corresponding to highest frequency 40. Hence we should use median.

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

$$= 40 + \frac{75-70}{10} \times 10$$

$$= 45$$

$$\text{Mean } (\bar{x}) = A + \frac{\sum fd'}{N} \times h$$

$$= 35 + \frac{64}{150} \times 10 = 39.27$$

$$\text{SD } (\sigma) = \sqrt{\frac{\sum fd'^2}{N} - \left(\frac{\sum fd'}{N}\right)^2} \times h$$

$$= \sqrt{\frac{808}{150} - \left(\frac{64}{150}\right)^2} \times 10$$

$$= \sqrt{5.21} \times 10 = 22.83$$

$\therefore$ Co-efficient of skewness

$$Sk = \frac{3(\text{Mean - Median})}{SD.}$$

$$= \frac{3(39.27 - 45)}{22.83}$$

$$= -.75$$

EXERCISE–6

1. Write a short note on skewness of a frquency distribution.
(AHSEC 1990, 1992, 1993, 1995, 2001, 2002, 2008)

2. Write short note on kurtosis.
(AHSEC 1990, 1993, 1995, 1999, 2002, 2008)

3. A frequency distribution gives the following result:
CV = 5
Karl Pearson's co-efficient of skewness = .5, standard deviation = 2
Find mean and mode of the distritbution.

(AHSEC 1990, 2000) $\left[\textbf{Ans.}\ \begin{array}{r}\textbf{Mean}=\textbf{40}\\ \textbf{Mode}=\textbf{39}\end{array}\right]$

4. For a moderately skew distribution, what relation between mean, median and mode exists?
(AHSEC2005) [Ans. Mode = 3 Median–2 Mean]

5. Write a note of skewness and kurtosis. Draw diagrams where necessary.
(AHSEC 1998)

6. State the order in which mean, median and mode of a distribution lie for a positively skew frequency distribution. **(AHSEC 1991)**
[Ans. Mean>Median>Mode]

7. The approximate relation between mean, median and mode of a moderately skew distribution is __________. **(AHSEC 1990)**
[Ans. Mode = 3 Median –2Mean]

8. The mean, mode and S.D. of a distribution are 40, 52.5 and 10 respectively. Find the co-efficient of skewness. Draw a rough sketch of the frequency curve of the distribution and show the relative positions of the mean, median and mode on the graph.
(AHSEC 1994) [Ans. *Sk* = –1.25]

9. Fill in the blanks.

(i) In a frequency curve when mean>median>mode the curve is said to be _____ curve.
(AHSEC 1998)

(ii) Kurtosis is greater then _________.

(iii) Co-efficient of skewness is ____ than unity.

(iv) When $Sk = 0$, the distribution is ____ shaped .

(v) In a symmetrical distribution, co-efficient Sk = ____________.

[Ans. (i) positively skewed (ii) unity (iii) less (iv) bell (v) O]

10. Karl Pearson's Co-efficient of Skewness of a distribution is .32. Its standard deviation is 6.5 and mean is 29.6. Find mode and median of he distribution.

$$\left[\textbf{Ans.}\ \begin{array}{r}\textbf{Mode} = \textbf{27.5}\\ \textbf{Median} = \textbf{28.9}\end{array}\right]$$

11. In a moderately skew distribution mean is 24.6 and median is 25.1. Find the value of mode.

(AHSEC 1998) [Ans. 26.1]

12. If concentration of variate values about the mode of a symmetrical distribution increases will the kurtosis of the frequency curve increase?

(AHSEC 1991)

[Ans. Yes]

13. Calculate Karl Pearson's co-efficient of skewness from the following data

Class Interval :	0–10	10–20	20–30	30–40	40–50
Frequency :	15	20	30	25	10

(Ans. –0.12)

14. In a distribution, the difference of the two quartiles is 15 and their sum is 35 and the median is 20. Find the co-efficient of skewness.

[Ans. Sk = –•33]

15. In case of skewed distribution, what happens to mean, median and mode?

[AHSEC 2008]

7

CORRELATION AND REGRESSION

7.1 BIVARIATE DISTRIBUTION

In case of univariate data, the data are collected by studying only one characteristic e.g. the mean height of students in a college, each observation represents the height of each student. But many situations arise in which we may have to study two or more variables simultaneously. Such type of data lead to Bi-variate or Multi-variate distributions depending upon the number of characteristics under study e.g. if we study the students of a college with respect to height and weight the data obtained is known as bi-variate data. The concept can be generalised for multivariate data. A few examples of bi-variate distributions may be

(i) the income and expenditure of a group of families
(ii) the price and demand of a set of commodities
(iii) age of husbands and age of wives
(iv) the amount of fertilizers used and yield of a certain crop etc.

For bivariate data we take (x_i, y_i); $i = 1, 2, \ldots n$ and the frequency distribution is called bivariate frequency distribution which is expressed in the form

Variables	Frequency
$(x_1\, y_1)$	f_1
$(x_2\, y_2)$	f_2
....	....
(x_n, y_n)	f_n

If x denotes height and y weight, then f_1 individuals having height x_1 and weight y_1; f_2 individuals height x_2 and weight y_2 and so on. In case of a grouped frequency distribution the data can be presented in a two way table.

7.2 CORRELATION

In a bivariate distribution, we may be intrested to know if there is any relationship between the variables under consideration. Two variables are said to be correlated if with a change in the value of one variable there corresponds a change in the value of the other variable. Correlation is a statistical technique which measures and analyses the degree to which two variables vary with reference to each other. On the other hand if a change in the value of one variable does not bring any change in the value of the other variable, the two variables are said to be uncorrelated i.e., having no relation with each other e.g., price of commodities and height of individuals etc.

When correlation is studied between two variables, it is called simple correlation, when there are three or more variables and we study the combined effect of two or more variables upon a variable not included in that group, our study is of multiple correlation. If, however, we wish to examine the effect of one variable upon a second, after eliminating the effects of other variables, our problem is that of partial correlation.

7.3 TYPES OF CORRELATION

On the basis of nature of relationship between two variables, correlation may be

1. Positive Correlation

If increase (or decrease) in one variable results a corresponding increase (or decrease) in the other variable ie., two variables move in the same direction, variables are said to be positively correlated e.g., heights and weights of children, price and supply of a set of commodities etc., income and expenditure of a group of families etc.

2. Negative Correlation

If increase (or decrease) in one variable results a corresponding decrease (or increase) in the other variable ie., two variables deviate in the opposite direction, variables are said to be negatively correlated e.g. price and demand of a set of commodities, speed and time etc.

3. Perfect Positive Correlation

When changes in two related variables are exactly proportional there is perfect correlation between them.

If increase (or decrease) in one variable results a corresponding and proportional increase (or decrease) in the other variable, variables are said to be perfect positively correlated e.g. Correlation between radius and circumference of a circle, temperature and volume etc.

4. Perfect Negative Correlation

If increase (or decrease) in one variable results a corresponding and proportional decrease (or increase) in the other variable, variables are said to be perfect negatively correlated: e.g., pressure and volume etc.

5. Zero Correlation

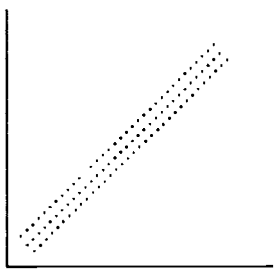
If the change in one variable does not result any change in the other variable, variables are said to be uncorrelated or independent or having zero correlation e.g. price qf commodities and weight of individuals etc.

7.4 SCATTER DIAGRAM

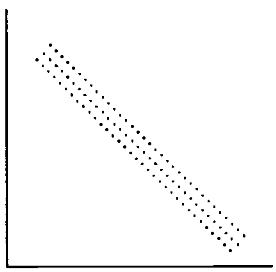
This is the simplest way of representing bivariate data. When data relating to the simultaneous measurement on two variables are available, each pair of observations can be graphically represented by a dot on a graph. Suppose (x_i, y_i); i = 1, 2, ..., n be n pairs of values of the variables X and Y. Taking X on the horizontal axis and Y along vertical axis we get a diagram of n dots. This is called scatter diagram. If these points show some trend, the two variables are correlated and if the points do not reflect any trend the two variables have no correlation. If the trend of the points is upward rising, correlation is positive and if the trend is downward from left top to right bottom, correlations is negative. If the dots lie exactly on an upward rising straight line, then two variables are said to be perfectly positively correlated. Similarly, when all the dots lie exactly on a downward sloping straight line, then two variables are said to be perfect negatively correlated.

When the variation in the values of two variables are in a constant ratio, correlation is said to be linear. On the other hard, if the two variables are such that the amount of change in one variable does not result a constant ratio to the amount of change in the other variable, then correlation would be non-linear or curvilinear. In the first case a straight line $y = a + bx$ would be obtained if all the points are plotted and in the second case we would not get a straight line.

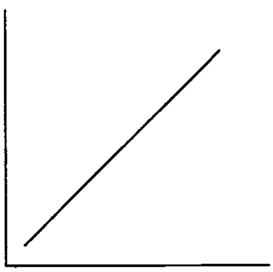
Following are some scattered diagrams representing varied degrees of correlation



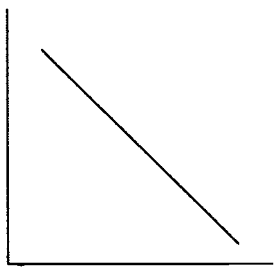
Positive Correlation
Figure 7.1



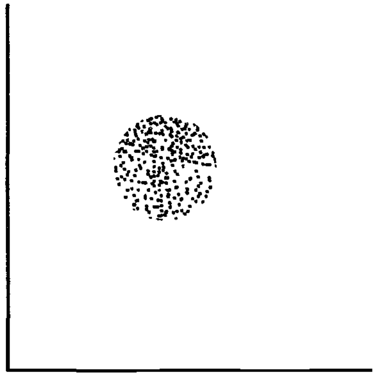
Negative Correlation
Figure 7.2



Perfect positive correlation
Figure 7.3



Perfect negative correlation
Figure 7.4



Zero Correlation
Figure 7.5

7.5 KARL PEARSON'S CO-EFFICIENT OF CORRELATION

Prof. Karl Pearson, the noted statistician and a pioneer in the field of correlation and regression, presented a co-efficient to measure the degree of linear relationship between two variables which is known as Karl Pearson's co-efficient of correlation (or product-moment correlation co-efficient). It is denoted by the symbol 'r'. The correlation co-efficient between two variables X and Y.

$$r_{xy} = \frac{\text{cov}(x, y)}{\sigma_x . \sigma_y}$$

$$\text{where cov}(x, y) = \frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})$$

$$= \frac{1}{n}\sum_{i=1}^{n}(x_i y_i - x_i \bar{y} - \bar{x} y_i - \bar{x}.\bar{y})$$

$$= \frac{1}{n}\sum_{i=1}^{n} x_i y_i - \bar{y}\frac{1}{n}\sum_{i=1}^{n} x_i - \bar{x}\frac{1}{n}\sum_{x=1}^{n} y_i + \frac{1}{n} n\bar{x}.\bar{y}$$

$$= \frac{1}{n}\sum_{i=1}^{n} x_i y_i - \bar{x}.\bar{y} - \bar{x}.\bar{y} + \bar{x}.\bar{y}$$

$$= \frac{1}{n}\sum_{i=1}^{n} x_i y_i - \bar{x}.\bar{y}$$

$$\sigma_x^2 = \frac{1}{n}\sum_{i=1}^{n}\left(x_i - \bar{x}\right)^2$$

$$= \frac{1}{n}\sum_{i=1}^{n}\left(x_i^2 - 2x_i\bar{x} + \bar{x}^2\right)$$

$$= \frac{1}{n}\sum_{i=1}^{n} x_i^2 - 2\bar{x}\ \frac{1}{n}\sum_{i=1}^{n} x_i + \frac{1}{n}n\bar{x}^2$$

$$= \sum_{i=1}^{n} x_i^2 - 2.\bar{x}.\bar{x} + \bar{x}^2$$

$$= \frac{1}{n}\sum_{i=1}^{n} x_i^2 - \bar{x}^2$$

Similarly $\sigma_y^2 = \frac{1}{n}\sum_{i=1}^{n}\left(y_i - \bar{y}\right)^2 = \frac{1}{n}\sum_{i=1}^{n} y_i^2 - \bar{y}^2$

Thus r_{xy} can also be written in the following forms

$$r_{xy} = \frac{\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^{n}\left(x_i - \bar{x}\right)^2}\sqrt{\sum_{i=1}^{n}\left(y_i - \bar{y}\right)^2}}$$

$$= \frac{\sum\left(x - \bar{x}\right)\left(y - \bar{y}\right)}{\sqrt{\sum\left(x - \bar{x}\right)^2 \sum\left(y - \bar{y}\right)^2}}$$

$$r_{xy} = \frac{n\sum_{i=1}^{n} x_i y_i - \left(\sum_{i=1}^{n} x_i\right)\left(\sum_{i=1}^{n} y_i\right)}{\sqrt{n\sum_{1}^{n} x_i^2 - \left(\sum_{i=1}^{n} x_i\right)^2}\sqrt{n\sum_{i=1}^{n} y_i^2 - \left(\sum_{i=1}^{n} y_i\right)^2}}$$

$$= \frac{n\sum xy - \left(\sum x\right)\left(\sum y\right)}{\sqrt{n\sum x^2 - \left(\sum x\right)^2}\sqrt{n\sum y^2 - \left(\sum y\right)^2}}$$

when actual mean is in decimal, the calculations become very tedius and in such cases we may take help of the following formula.

$$r = \frac{n\sum uv - (\Sigma u)(\Sigma v)}{\sqrt{n\sum u^2 - (\Sigma u)^2}\sqrt{n\sum v^2 - (\Sigma v)^2}}; \qquad \begin{array}{l} u = x - A \\ v = y - B \end{array}$$

where A and B are assumed means.

7.6 PROPERTIES OF KARL PEARSON'S CO-EFFICIENT OF CORRELATION (r)

1. $-1 \le r \le 1$ i.e., r lies between -1 and $+1$

Proof: Let X and Y be two variables which take values (x_i, y_i); $i = 1, 2, \ldots n$ with means $\bar{x}$, $\bar{y}$ and standard deviations σ_x, σ_y respectively.

Consider, $$\sum\left[\frac{x-\bar{x}}{\sigma_x} \pm \frac{y-\bar{y}}{\sigma_y}\right]^2 \ge 0$$

$$\Rightarrow \quad \sum\left[\left(\frac{x-\bar{x}}{\sigma_x}\right)^2 + \left(\frac{y-\bar{y}}{\sigma_y}\right)^2 \pm 2\frac{(x-\bar{x})(y-\bar{y})}{\sigma_x\sigma_y}\right] \ge 0$$

$$\Rightarrow \quad \frac{1}{\sigma_x^2}\sum(x-\bar{x})^2 + \frac{1}{\sigma_y^2}\sum(y-\bar{y})^2 \pm \frac{2}{\sigma_x\sigma_y}\sum(x-\bar{x})(y-\bar{y}) \ge 0$$

Dividing both sides by n

$$\frac{1}{\sigma_x^2}.\frac{\sum(x-\bar{x})^2}{n} + \frac{1}{\sigma_y^2}\frac{\sum(y-\bar{y})^2}{n} \pm \frac{2}{\sigma_x\sigma_y}\frac{\sum(x-\bar{x})(y-\bar{y})}{n} \ge 0$$

$$\Rightarrow \quad \frac{1}{\sigma_x^2}.\sigma_x^2 + \frac{1}{\sigma_y^2}.\sigma_y^2 \pm \frac{2}{\sigma_x\sigma_y}\text{cov}(x, y) \ge 0$$

$\Rightarrow 1 + 1 \pm 2r \ge 0$

$\Rightarrow 2 \pm 2r \ge 0$

$\Rightarrow 2(1 \pm r) \ge 0$

$\Rightarrow 1 \pm r \ge 0$

Either $1 + r \ge 0$ or $1 - r \ge 0$

$\Rightarrow r \ge -1$ or $1 \ge r \therefore -1 \le r \le 1$

Interpretation

The least value of r is -1 and the most is $+1$. If $r = +1$, there is perfect positive correlation between two variables. If $r = -1$, there is perfect negative correlation.

If $r = 0$, we say that there is no linear relation between the variable. However, there may be non-linear relationship between the variables.

If r is positive but close to zero, we have weak positive correlation and if r is close to $+1$, we have strong positive correlation,

2. Correlation co-efficient r is independent of change in origin and scale.

Proof: Let X and Y be the original variables and after changing origin and scale new variables

$$U = \frac{X-a}{h} \quad \text{and} \quad V = \frac{Y-b}{k}$$; a, b, h, k are all constants.

$$\Rightarrow \quad X - a = hU \qquad \Rightarrow \quad Y - b = kV$$
$$\Rightarrow \quad X = a + hU \qquad \Rightarrow \quad Y = b + kV$$
$$\Rightarrow \quad \overline{X} = a + h\overline{U} \qquad \Rightarrow \quad \overline{Y} = b + k\overline{V}$$
$$\therefore \quad X - \overline{X} = h\left(U - \overline{U}\right)$$
$$Y - \overline{Y} = k\left(V - \overline{V}\right)$$

Now,

$$r_{xy} = \frac{\sum\left(x-\bar{x}\right)\left(y-\bar{y}\right)}{\sqrt{\sum\left(x-\bar{x}\right)^2\sum\left(y-\bar{y}\right)^2}}$$

$$= \frac{\sum h\left(u-\bar{u}\right).k\left(v-\bar{v}\right)}{\sqrt{\sum h^2\left(u-\bar{u}\right)^2}\sqrt{\sum k^2\left(v-\bar{v}\right)^2}}$$

$$= \frac{hk\sum\left(u-\bar{u}\right)\left(v-\bar{v}\right)}{hk\sqrt{\sum\left(u-\bar{u}\right)^2.\sum\left(v-\bar{v}\right)^2}}$$

$$= r_{uv}$$

$\therefore \quad r_{xy} = r_{uv}$. Hence Proved.

3. Two independent variables are uncorrelated but the converse is not true.

Proof: If two variables are independent then their covariance is zero i.e. cov $(X, Y) = 0$

$$\therefore \quad r_{xy} = \frac{Cov(X,Y)}{\sigma_x\sigma_y} = \frac{0}{\sigma_x.\sigma_y} = 0$$

Thus, if two variables are independent their co-efficient of correlation is zero i.e. independent variables are uncorrelated.

But the converse is not true. If $r_{xy} = 0$, we say that there does not exist any linear correlation between the variables as Karl Peanson's co-efficient of correlation r_{xy} is a measure of only linear relationship. However, there may be strong non-linear or curvilinear relationship even through $r_{xy} = 0$.

As an illustration consider the bivariate distribution

x:	–3	–2	–1	0	1	2	3
y:	9	4	1	0	1	4	9

Applying the formula of r_{xy}, we get $r_{xy} = 0$. But X and Y are not independent and they are related by the non-linear relation $y = x^2$. Hence Proved.

4. Correlation co-efficient r is a pure number independent of unit of measurement.
5. Correlation co-efficient is symmetric.

7.7 ASSUMPTIONS OF KARL PEARSON'S CO-EFFICIENT OF CORRELATION.

There are three assumptions

(i) the variable x and y are linearly related.
(ii) there is a cause and effect relationship between factors affecting the values of the variable x and y.
(iii) the random variables x and y are normally distributed.

Example 1: Calculate the co-efficient of correlation r_{xy} from the following data

$\Sigma X = 71$, $\Sigma Y = 70$, $\Sigma X^2 = 555$,
$\Sigma Y^2 = 526$, $\Sigma XY = 527$, $n = 10$

(AHSEC 1997)

Solution:

$$r_{xy} = \frac{Cov(X,Y)}{\sigma_x \sigma_y}$$

$$= \frac{n\sum XY - (\sum X)(\sum Y)}{\sqrt{n\sum X^2 - (\sum X)^2}\sqrt{n\sum Y^2 - (\sum Y)^2}}$$

$$= \frac{10 \times 527 - 71 \times 70}{\sqrt{10 \times 555 - (71)^2}\sqrt{10 \times 526 - (70)^2}}$$

$$= \frac{5270 - 4970}{\sqrt{509}\sqrt{360}} = \frac{300}{427.96}$$

$$= .70$$

Example 2: Find the correlation co-efficient between X and Y from the following data:

X:	2	4	5	6	8	11
Y:	18	12	10	8	7	2

(AHSEC 1995)

Solution:

X	Y	$U = X - 6$	$V = Y - 8$	U^2	V^2	UV
2	18	–4	10	16	100	–40
4	12	–2	4	4	16	–8
5	10	–1	2	1	4	–2
6	8	0	0	0	0	0
8	7	2	–1	4	1	–2
11	2	5	–6	25	36	–30
		$\Sigma U = 0$	$\Sigma V = 9$	$\Sigma U^2 = 50$	$\Sigma V^2 = 157$	$\Sigma UV = -82$

$$r_{xy} = r_{uv} = \frac{n\sum UV - \left(\sum U\right)\left(\sum V\right)}{\sqrt{n\sum U^2 - \left(\sum U\right)^2}\sqrt{n\sum V^2 - \left(\sum V\right)^2}}$$

$$= \frac{6(-82) - 0 \times 9}{\sqrt{6 \times 50 - 0}\sqrt{6 \times 157 - 81}}$$

$$= \frac{-492}{508.23} = -.97$$

Example 3: Find the correlation co-efficient between X and Y from the following data and interpret the result.

(AHSEC 2001)

X :	16	20	24	28	32
Y :	30	40	25	35	45

Solution:

Here $\bar{X} = \frac{120}{5} = 24$

$\bar{Y} = \frac{175}{5} = 35$

Since $\bar{X}$ and $\bar{Y}$ are whole numbers, we can proceed as follows

X	Y	$X - \bar{X}$	$Y - \bar{Y}$	$(X - \bar{X})^2$	$(Y - \bar{Y})^2$	$(X - \bar{X})(Y - \bar{Y})$
16	30	–8	–5	64	25	40
20	40	–4	5	16	25	–20
24	25	0	–10	0	100	0
28	35	4	0	16	0	0
32	45	8	10	64	100	80
$\Sigma X = 120$	$\Sigma Y = 175$			$\Sigma(X - \bar{X})^2 = 160$	$\Sigma(Y - \bar{Y})^2 = 250$	$\Sigma(X - \bar{X})(Y - \bar{Y}) = 100$

$$r = \frac{\Sigma(X-\overline{X})(Y-\overline{Y})}{\sqrt{\Sigma\left(X-\overline{X}\right)^2 \Sigma\left(Y-\overline{Y}\right)^2}}$$

$$= \frac{100}{\sqrt{160 \times 250}}$$

$$= \frac{100}{200} = .5$$

Interpretation: Since $r = .5$ we find positive correlation between the variables X and Y.

Example 4: Calculate correlation co-efficient from the following data:

$$n = 10, \qquad \Sigma x = 140,$$
$$\Sigma y = 150, \qquad \Sigma(x-10)^2 = 180,$$
$$\Sigma(y-15)^2 = 215, \quad \Sigma(x-10)(y-15) = 60$$

Solution:

Let us take $u = x - 10$
$v = y - 15$

$$\begin{aligned}
\therefore\ \Sigma u &= \Sigma(x-10) &&= \Sigma x - n \times 10 = 140 - 100 &&= 40\\
\Sigma v &= \Sigma(y-15) &&= \Sigma y - n \times 15 = 150 - 150 &&= 0\\
\Sigma u^2 &= \Sigma(x-10)^2 &&= 180\\
\Sigma v^2 &= \Sigma(y-15)^2 &&= 215\\
\Sigma uv &= \Sigma(x-10)(y-15) &&= 60
\end{aligned}$$

$$r_{xy} = r_{uv} = \frac{n\sum uv - \left(\sum u\right)\left(\sum v\right)}{\sqrt{n\sum u^2 - \left(\sum u\right)^2}\sqrt{n\sum v^2 - \left(\sum v\right)^2}}$$

$$= \frac{10 \times 60 - 40 \times 0}{\sqrt{10 \times 180 - (40)^2}\sqrt{10 \times 215 - 0}}$$

$$= \frac{600}{\sqrt{200 \times 2150}}$$

$$= \frac{6}{6.557} = .91$$

Example 5: Show the correlation co-efficient between x and $a-x$ is -1

(AHSEC 1997, 2000)

Solution:

We know that

$$r_{xy} = \frac{Cov(x,y)}{\sigma_x \sigma_y}$$

$$\therefore \qquad r_{x(a-x)} = \frac{Cov(x,\, a-x)}{\sigma_x \sigma_{a-x}}$$

$$= \frac{\frac{1}{n}\sum\left(x-\bar{x}\right)\left(a-x-a+\bar{x}\right)}{\sqrt{\frac{1}{n}\sum\left(x-\bar{x}\right)^2}\sqrt{\frac{1}{n}\sum\left(a-x-a+\bar{x}\right)^2}}$$

$$= \frac{-\sum\left(x-\bar{x}\right)^2}{\sum\left(x-\bar{x}\right)^2} = -1$$

Example 6: If *a, b, c, d* are constants, then show that the co-efficient of correlation between *ax + b* and *cy + d* in numerically equal to that between *x* and *y*.

(AHSEC 1998)

Solution:

let $u = ax + b$ and $v = cy + d$

$\Rightarrow \quad \bar{u} = a\,\bar{x} + b \qquad \Rightarrow \quad \bar{v} = c\,\bar{y} + d$

$\therefore \quad u - \bar{u} = a\,(x - \bar{x}) \qquad v - \bar{v} = c(y - \bar{y})$

Now

$$r_{uv} = \frac{\sum\left(u-\bar{u}\right)\left(v-\bar{v}\right)}{\sqrt{\sum\left(u-\bar{u}\right)^2}\sqrt{\sum\left(v-\bar{v}\right)^2}}$$

$$= \frac{\sum a\left(x-\bar{x}\right)c\left(y-\bar{y}\right)}{\sqrt{a^2\sum\left(x-\bar{x}\right)^2}\sqrt{c^2\sum\left(y-\bar{y}\right)^2}}$$

$$= \frac{ac\sum\left(x-\bar{x}\right)\left(y-\bar{y}\right)}{ac\sqrt{\sum\left(x-\bar{x}\right)^2}\sqrt{\sum\left(y-\bar{y}\right)^2}}$$

$$= r_{xy}$$

Example 7: Given that r_{xy} = .6, cov (x, y) = 7.2, var (y) = 16, find the standard deviation of x.

(AHSEC 1998)

Solution:

We know that, $r_{xy} = \dfrac{Cov(x,y)}{\sigma_x.\sigma_y}$

$$\Rightarrow \quad .6 = \frac{7.2}{.\sigma_x\sqrt{16}} \text{ since SD} = \sqrt{\text{var}}$$

$$\Rightarrow \quad \sigma_x = \frac{7.2}{.6 \times 4}$$

$$\Rightarrow \quad \sigma_x = 3$$

Example 8: A computer while calculating correlation co-efficient between two variables X and Y from 25 pairs of observations obtained the following results.

$N =$ **25,**	$\Sigma X =$ **125,**	$\Sigma X^2 =$ **650,**
$\Sigma Y =$ **100,**	$\Sigma Y^2 =$ **460**	$\Sigma XY =$ **508**

It was however, discovered at the time of checking that two pairs of observations were not correctly copied. They were taken as (6, 14) and (8, 6) while the correct values were (8, 12) and (6, 8). What is the correct value of the correlation co-efficient.

Solution:

Corrected $\Sigma X = 125 - 6 - 8 + 8 + 6 = 125$

" $\Sigma Y = 100 - 14 - 6 + 12 + 8 = 100$

" $\Sigma X^2 = 650 - 6^2 - 8^2 + 8^2 + 6^2 = 650$

" $\Sigma Y^2 = 460 - 256 - 36 + 144 + 64 = 436$

" $\Sigma XY = 508 - 6 \times 14 - 8 \times 6 + 8 \times 12 + 6 \times 8 = 520$

$\therefore$ Corrected correlation co-efficient is

$$r = \frac{N\sum XY - (\sum X)(\sum Y)}{\sqrt{N\sum X^2 - (\sum X)^2}\sqrt{N\sum Y^2 - (\sum Y)^2}}$$

$$= \frac{25 \times 520 - 125 \times 100}{\sqrt{25 \times 650 - (125)^2}\sqrt{25 \times 436 - (100)^2}}$$

$$= \frac{500}{25 \times 30} = \frac{500}{750}$$

$$= \frac{2}{3} = .67$$

7.8 REGRESSION

The concept of regression was first used by Sir Francis Galton in his study of heredity. The general meaning of 'Regression' is to return or go back to the average value.

Regression indicates the average relationship between two variables and from this average relationship the average value of one variable is estimated corresponding to a given value of other variable. This process is known as simple regression. In regression analysis there are two types of variables-dependent variable and independent variable. A dependent variable is one whose value is to be predicted. It is also known as explained variable. An independent variable is the variable which influences the value of the other variable. It is also known as explanator.

In the words of *M.M. Blair*, "Regression analysis is a mathematical measure of the average relationship between two or more variables in terms of the original units of the data."

Lines of Regression

Line of regression is the line which gives the best estimate of one variable for any specific value of the other variable. For bivariate distribution we have two lines of regression.

(i) The regression line of Y on X – this gives the best estimated value of Y corresponding to a given value of X.

(ii) The regression line of X on Y – This gives the best estimated values of X corresponding to a given value of Y.

There are always two lines of regression since each variable may be treated as the dependent as well as the independent variable. When we consider X as independent variable and Y as the dependent variable we get the regression line of Y on X. Again, when we consider Y as independent variable and X as the dependent variable we get the regression line of X on Y.

7.9 REGRESSION EQUATIONS

The regression equations express the regression lines.

1. Regression equation of Y on X is

$$Y - \overline{Y} = b_{yx} (X - \overline{X})$$

where b_{yx} = Regression co-efficient of Y on X

$$= \frac{Cov(X,Y)}{\sigma_x^2} = r.\frac{\sigma_y}{\sigma_x}$$

$$= \frac{\sum(x-\bar{x})(y-\bar{y})}{\sum(x-\bar{x})^2}$$

$$= \frac{n\sum xy-(\sum x)(\sum y)}{n\sum x^2-(\sum x)^2}$$

2. Regression eq' ation of X on Y is

$$X - \bar{X} = b_{xy}(Y - \bar{Y})$$

where b_{xy} = Regression co-efficient of X on Y

$$= \frac{Cov(X,Y)}{\sigma_y^2} = r.\ \frac{\sigma_x}{\sigma_y}$$

$$= \frac{\sum(x-\bar{x})(y-\bar{y})}{\sum(y-\bar{y})^2}$$

$$= \frac{n\sum xy-(\sum x)(\sum y)}{n\sum y^2-(\sum y)^2}$$

Note: When actual mean is in decimal, we may take help of the formula given below:

$$b_{yx} = \frac{n\sum uv-(\sum u)(\sum v)}{n\sum u^2-(\sum u)^2};\quad \begin{aligned} u &= x-A \\ v &= y-B \end{aligned}$$

$$b_{xy} = \frac{n\sum uv-(\sum u)(\sum v)}{n\sum v^2-(\sum v)^2}$$

A and B are assumed means.

7.10 PROPERTIES OF REGRESSION CO-EFFICIENTS

1. The geometric mean of the regression co-efficients is the correlation co-efficient.

Proof: we have $b_{yx} = r\ \dfrac{\sigma_y}{\sigma_x}$

$$b_{xy} = r\ \frac{\sigma_x}{\sigma_y}$$

$$\therefore \quad b_{yx} \cdot b_{xy} = r\frac{\sigma_y}{\sigma_x} \cdot r\frac{\sigma_x}{\sigma_y}$$

$$\Rightarrow \quad b_{yx} \cdot b_{xy} = r^2$$

$$\Rightarrow \quad r = \pm\sqrt{b_{yx} . b_{xy}}$$

Remark

Both the regression co-efficients will have the same sign and that will be the sign of correlation co-efficient. If both b_{yx} and b_{xy} are positive then r is also positive and if both b_{yx} and b_{xy} are negative then r is also negative. Since $\sigma_x > 0$, $\sigma_y > 0$, the sign of each of r, b_{yx}, b_{xy} depends on Cov (X, Y).

2. If one of the regression co-efficients is >1, then the other must be <1

Proof: Let $b_{yx} > 1$ (say)

$$\Rightarrow \quad \frac{1}{b_{yx}} < 1$$

Also, $\quad b_{yx} . b_{xy} = r^2$

$$\Rightarrow \quad b_{yx} . b_{xy} \leq 1 \qquad \text{Since } -1 \leq r \leq 1$$

$$\Rightarrow \quad r^2 \leq 1$$

$$\Rightarrow \quad b_{xy} \leq \frac{1}{b_{yx}} < 1$$

Hence, $\quad b_{xy} < 1$

3. Regression co-efficients are independent of the change of origin but not of scale.

Proof: Let X and Y be the original variables and after changing origin and scale new variables

$$u = \frac{X-a}{h} \text{ and } v = \frac{Y-b}{k};$$

a, b, h, k are all constants. $h > o$, $k > o$.

Since correlation co-efficient are independent of change of origin and scale, $r_{xy} = r_{uv}$. Again SD is independent of change in origin but not of scale

$$\therefore \quad \sigma_x = h\sigma_u \text{ and } \sigma_y = k\sigma_v$$

Now $$b_{yx} = r_{xy}\frac{\sigma_y}{\sigma_x} = r_{uv}\frac{k\sigma_v}{h\sigma_u} = \frac{k}{h}b_{vu}$$

Also, $b_{xy} = r_{xy} \dfrac{\sigma_x}{\sigma_y} = r_{uv} \dfrac{h\sigma_u}{k\sigma_v} = \dfrac{h}{k}\, b_{uv}$. Hence proved.

4. The arithmetic mean of the regression co-efficients is $\geq$ correlation coefficient, provided $r > o$.

Proof: we want to show

$$\frac{b_{yx} + b_{xy}}{2} \geq r$$

$$\Rightarrow \quad b_{yx} + b_{xy} \geq 2r$$

$$\Rightarrow r\frac{\sigma_y}{\sigma_x} + r\frac{\sigma_x}{\sigma_y} \geq 2r$$

$$\Rightarrow \quad \frac{\sigma_y}{\sigma_x} + \frac{\sigma_x}{\sigma_y} \geq 2$$

$$\Rightarrow \quad \sigma_y^2 + \sigma_x^2 \geq 2\sigma_x\sigma_y$$

$$\Rightarrow \quad \sigma_y^2 + \sigma_x^2 - 2\sigma_x\sigma_y \geq 0$$

$\Rightarrow \quad (\sigma_y - \sigma_x)^2 \geq 0$ which is always true. Hence proved.

Some Important Remarks

1. The regression co-efficient of Y on X denoted by b_{yx} gives the change in Y for a unit change in the value of X. The regression co-efficient of X on Y denoted by b_{xy} gives the change in the value of X for a unit change in the value of Y.

2. Both the regression lines passes through the point $(\bar{x}, \bar{y})$

3. When two variables are uncorrelated ($r = o$), the lines of regression become perpendicular to each other. In case of perfect positive or perfect negative correlation ($r = \pm 1$), the lines of regression coincide since they can not be parallel.

Example 9: x and y are two variables for which 10 pairs of values are available. Further

$\Sigma x = 10$, $\Sigma y = 0$, $\Sigma x^2 = 148$,

$\Sigma y^2 = 164$, $\Sigma xy = 123$

Find the regression co-efficient of y on x (AHSEC 1992)

Solution:

The regression co-efficient of y on x

$$b_{yx} = \frac{Cov(x,y)}{\sigma_x^{\,2}}$$

$$= \frac{\frac{1}{n}\sum(x-\bar{x})(y-\bar{y})}{\frac{1}{n}\sum(x-\bar{x})^2}$$

$$= \frac{n\sum xy - (\sum x)(\sum y)}{n\sum x^2 - (\sum x)^2}$$

$$= \frac{10\times 123 - 0}{10\times 148 - 100}$$

$$= \frac{1230}{1380} = .89$$

Example 10: Given that b_{xy} = .25, var(x) = 4, var(y) = 36, find the correlation between x and y

(AHSEC 1993)

Solution:

Given $b_{xy} = .25$

$\sigma_x^{\,2} = 4 \qquad \Rightarrow \sigma_x = 2$

$\sigma_y^{\,2} = 36 \qquad \Rightarrow \sigma_y = 6$

we know that $b_{xy} = r\,\dfrac{\sigma_x}{\sigma_y}$

$\Rightarrow \quad .25 = r.\ \dfrac{2}{6}$

$\Rightarrow \quad r = .75$

Example 11: If r_{xy} = .6 and b_{yx} = .8, what is the value of b_{xy}? (AHSEC 1998)

Solution: We know that

$r^2 = b_{yx}\,.\,b_{xy}$

$\Rightarrow \quad (.6)^2 = .8 \times b_{xy}$

$\Rightarrow \quad \dfrac{.36}{.8} = b_{xy}$

$\Rightarrow \quad b_{xy} = .45$

Example 12: If two regression co-efficient are .8 and 1.2, what would be the value of the co-efficient of correlation?

(AHSEC 1997)

Solution: We know that

$$r = \sqrt{b_{yx}.b_{xy}}$$
$$= \sqrt{.8 \times 1.2}$$
$$= \sqrt{.96}$$
$$= .98$$

Example 13: The regression lines have the equations x + 2y –5 = 0 and 2x + 3y –8 = 0. Find $\bar{x}$ and $\bar{y}$. (AHSEC 1997)

Solution: Given the regression lines

$$x + 2y - 5 = 0$$
$$2x + 3y - 8 = 0$$

Since both the regression lines passes through the point $(\bar{x}, \bar{y})$, we have

$$\bar{x} + 2\bar{y} = 5 \qquad \text{(i)}$$

$$2\bar{x} + 3\bar{y} = 8 \qquad \text{(ii)}$$

$$\text{(i) } 2 \Rightarrow 2\bar{x} + 4\bar{y} = 10 \qquad \text{(iii)}$$

$$\text{(iii)} - \text{(ii)} \Rightarrow \bar{y} = 2$$

$$\therefore \text{(i)} \Rightarrow \bar{x} + 2 \times 2 = 5$$

$$\Rightarrow \bar{x} = 1$$

$$\left.\begin{array}{l} \bar{x} = 1 \\ \bar{y} = 2 \end{array}\right\}$$

Example 14: Given the following regression line of y on x is y = 10 –6x. Derive the condition under which the regression line of x on y can be written as $x = \frac{1}{6}(10 - y)$. (AHSEC 1998)

Solution: Given the regression line of y on x

$$y = 10 - 6x$$

$\therefore \qquad b_{yx} = -6$

If the regression line of y on x be

$$x = \frac{1}{6}(10 - y)$$

$$\Rightarrow \qquad x = \frac{10}{6} - \frac{1}{6}y$$

$$\therefore \qquad b_{xy} = -\frac{1}{6}$$

Now since both the regression co-efficient are negative

$$r = -\sqrt{(-6)\left(-\frac{1}{6}\right)}$$

$$= -\sqrt{1}$$

$$= -1$$

Hence the required condition is that there must be perfect negative correlation between x and y.

Example 15: Find the line of regression of y on x from the following data:

x :	**5**	**10**	**15**	**25**	**30**	**35**	**40**	**45**
y :	**25**	**32**	**44**	**32**	**39**	**49**	**55**	**60**

What will be the value of y for $x = 48$? **(AHSEC 2003)**

Solution:

x	y	$u = x - 30$	$v = y - 39$	u^2	v^2	uv
5	25	−25	−14	625	196	350
10	32	−20	−7	400	49	140
15	44	−15	5	225	25	−75
25	32	−5	−7	25	49	35
30	39	0	0	0	0	0
35	49	5	10	25	100	50
40	55	10	16	100	256	160
45	60	15	21	225	441	315
		$\Sigma u = -35$	$\Sigma v = 24$	$\Sigma u^2 = 1625$	$\Sigma v^2 = 1116$	$\Sigma uv = 975$

Now,

$$\bar{x} = \bar{u} + 30 \qquad \bar{y} = \bar{v} + 39$$

$$= \frac{-35}{8} + 30 \qquad\qquad = \frac{24}{8} + 39$$

$$= 25.63 \qquad\qquad = 42$$

$$b_{yx} = b_{vu} = \frac{n\sum uv - \left(\sum u\right)\left(\sum v\right)}{n\sum u^2 - \left(\sum u\right)^2}$$

$$= \frac{8 \times 975 - (-35) \times 24}{8 \times 1625 - (-35)^2}$$

$$= \frac{8640}{11775}$$

$$= .75$$

the regression line of y on x is

$$y - \bar{y} = b_{yx}\left(x - \bar{x}\right)$$

$\Rightarrow \qquad y - 42 = .75\ (x - 25.63)$

$\Rightarrow \qquad y = .75x + 22.78$

Now $\qquad x = 48,$

$$y = .75 \times 48 + 22.78$$

$$= 36 + 22.78$$

$$= 58.78$$

Example 16: The regression equation of x on y is $3y - 5x + 180 = 0$. Given that $\bar{y} = 4$, $\sigma_x^2 = \frac{9}{16}\sigma_y^2$ and $n = 4$. Find r and $\bar{x}$. (AHSEC 2005)

Solution:

Given the regression equation of x on y is

$$3y - 5x + 180 = 0$$

$\Rightarrow \qquad x = \frac{3}{5}y + 36 \qquad$ (i)

$\therefore \qquad b_{xy} = \frac{3}{5}$

$\Rightarrow \qquad r.\frac{\sigma_x}{\sigma_y} = \frac{3}{5}$

$\Rightarrow \qquad r = \frac{\sigma_y}{\sigma_x} \times \frac{3}{5}$

$$\Rightarrow \qquad r = \frac{4}{3} \times \frac{3}{5} \text{ since } \frac{\sigma_x^2}{\sigma_y^2} = \frac{9}{16} \text{ (given)}$$

$$\Rightarrow \qquad r = \frac{4}{5} = .8$$

From (i), $\bar{x} = \frac{3}{5}\bar{y} + 36$

$$= \frac{3}{5} \times 4 + 36$$

$$= \frac{12}{5} + 36$$

$$= 38.4$$

Example 17: Suppose b_{yx} is the regression co-efficient of y on x. What does it indicate? Interpret the meaning of the statement $b_{yx} = -.53$.
(AHSEC 2005)

Solution:

b_{yx} indicates the increment in the value of the dependent variable y for a unit change in the value of the independent variable x.

Interpretation: When $b_{yx} = -.53$, we mean that the change in the value of y is $-.53$ for the unit increase in the value of x.

Example 18: The Equation of two lines of regression are given below:

$$8x - 10y + 66 = 0$$

$$40x - 18y - 214 = 0$$

Find the co-efficient of correlation between x and y. **(AHSEC 1996)**

Solution: Let the equation of regression of y on x be $8x - 10y + 66 = 0$ and that of x on y be $40x - 18y - 214 = 0$

$8x - 10y + 66 = 0$

$$\Rightarrow \qquad 10y = 8x + 66$$

$$\Rightarrow \qquad y = \frac{4}{5}x + \frac{66}{10}$$

$$\therefore \qquad b_{yx} = \frac{4}{5}$$

$40x - 18y - 214 = 0$

$$\Rightarrow \qquad 40x = 18y + 214$$

$$\Rightarrow \qquad x = \frac{9}{20}y + \frac{214}{40}$$

$$\therefore \qquad b_{xy} = \frac{9}{20}$$

Co-efficient of correlation

$$r = \pm\sqrt{b_{yx}.b_{xy}}$$

$$= \pm\sqrt{\frac{4}{5}.\frac{9}{20}}$$

$$= \pm\frac{3}{5}$$

$$= \pm .6$$

Since both the regression co-efficients are positive, therefore $r = .6$

Note: Here $-1 \leq .6 \leq 1$, our supposition is correct. But if r goes outside the limits, we have to interchange the lines.

Example 19: Is the following statement correct? Give reasons.

'The regression co-efficient of x on y is 3.2 and that of y on x is .8'

(AHSEC 1996)

Solution: Given

$$b_{xy} = 3.2$$
$$b_{yx} = .8$$

Now,
$$r = \sqrt{b_{yx}.b_{xy}}$$

$$= \sqrt{3.2 \times .8}$$
$$= 1.6$$

Since r is out of the limit $-1 \leq r \leq 1$, the statement is incorrect.

Example 20: You are given the following data

	x	y
AM	**36**	**85**
SD	**11**	**8**

Correlation coefficient between x and y = .66

(i) Find two regression equations

(ii) Estimate the value of x when y = 75

Solution: Given,

$$\bar{x} = 36 \quad \bar{y} = 85$$

$$\sigma_x = 11 \quad \sigma_y = 8$$

$$r = .66$$

(i) Regression equation of y on x

$$y - \bar{y} = r\,\frac{\sigma_y}{\sigma_x}\left(x - \bar{x}\right)$$

$$\Rightarrow \quad y - 85 = .66.\frac{8}{11}(x - 36)$$

$$\Rightarrow \quad y - 85 = .48x - 17.28$$

$$\Rightarrow \quad y = .48x + 67.72$$

Regression equation of x on y

$$x - \bar{x} = r\,\frac{\sigma_x}{\sigma_y}(y - \bar{y})$$

$$\Rightarrow \quad x - 36 = .66.\frac{11}{8}(y - 85)$$

$$\Rightarrow \quad x - 36 = .908y - 77.14$$

$$\Rightarrow \quad x = .908y - 41.14$$

(ii) when $y = 75$, $x = .908 \times 75 - 41.14$

$$= 68.1 - 41.14$$

$$= 26.96$$

Example 21: If the two lines of regression are

$$\mathbf{4x - 5y + 30 = 0}$$

$$\mathbf{20x - 9y - 107 = 0}$$

which of these is the line of regression of x on y. Find r and σ_y when $\sigma_x = 3$.

Solution:

Let the regression equation of y on x be $20x - 9y - 107 = 0$ and that of x on y be $4x - 5y + 30 = 0$

$$20x - 9y - 107 = 0 \qquad\qquad 4x - 5y + 30 = 0$$

$$\Rightarrow \quad 9y = 20x - 107 \qquad\qquad \Rightarrow \quad 4x = 5y - 30$$

$$\Rightarrow \quad y = \frac{20}{9}x - \frac{107}{9} \qquad\qquad \Rightarrow \quad x = \frac{5}{4}y - \frac{30}{4}$$

$$\therefore \quad b_{yx} = \frac{20}{9} \qquad\qquad \therefore \quad b_{xy} = \frac{5}{4}$$

Now, $r = \sqrt{b_{yx}.b_{xy}}$, Since regression co-efficient are positive.

$$= \sqrt{\frac{20}{9} \cdot \frac{5}{4}}$$

$$= 1.67$$

Since, r goes outside $-1 \leq r \leq 1$, our supposition is wrong. So, the required regression line of x on y is $20x - 9y - 107 = 0$

Calculation of r

Let us interchange the regression lines

$20x - 9y - 107 = 0,$

$\Rightarrow \quad 20x = 9y + 107$

$\Rightarrow \quad x = \frac{9}{20}y + \frac{107}{20}$

$\therefore \quad b_{xy} = \frac{9}{20}$

$4x - 5y + 30=0$

$\Rightarrow \quad 5y = 4x + 30$

$\Rightarrow \quad y = \frac{4}{5}x + 6$

$\therefore \quad b_{yx} = \frac{4}{5}$

$$r = \sqrt{\frac{9}{20} \cdot \frac{4}{5}} = \frac{3}{5}$$

$$= .6$$

Now, $\quad b_{xy} = \frac{9}{20}$

$\Rightarrow \quad r\,\frac{\sigma_x}{\sigma_y} = \frac{9}{20}$

$\Rightarrow \quad \frac{3}{5} \cdot \frac{3}{\sigma_y} = \frac{9}{20}$

$\Rightarrow \quad \sigma_y = 4$

EXERCISE–7

1. Write a short note on scatter diagram. **(AHSEC 1998)**
2. Explain what do you mean by positive correlation and negative correlation. Give examples. **(AHSEC 1992)**
3. Can two uncorrelated variable be independent? **(AHSEC 1996)**
 [Ans. Need not]
4. If $r_{xy} = 0$, how are x and y related? **(AHSEC 1997)**
 [Ans. independent]
5. Define Karl Pearson's co-efficient of correlation. What does it measure? **(AHSEC 1999)**
6. If $r_{xy} = 0$, what is the value of cov (x, y) and how are x and y related? **(AHSEC 2000)**
 [Ans. o, independent]
7. Show that $-1 \leq r \leq 1$, where r is the correlation coefficient. **(AHSEC 2004, 2007)**
8. Show that correlation co-efficient is independent of change of origin and scale. **(AHSEC 2003, 2008)**
9. If the correlation co-efficient between two related variables x and y be 0.5, what will be the correlation co-efficient between y and x? **(AHSEC 1992)**
 [Ans. .5]
10. x and y are two independent variables, show that they are uncorrelated. **(AHSEC 1992, 2006)**
11. The correlation co-efficient between x and y be r. If a, b are constants, then show that correlation co-efficient between ax and by is numerically equal to r. **(AHSEC 1992)**
12. Suppose the correlation co-efficient ($r = 0$) between two variables x and y are zero. Does it mean that x and y are independent? Explain by means of an example. **(AHSEC 2004)**
13. Define correlation. Discuss positive, negative and zero correlation with the help of scatter diagram.
14. What do you mean by bi-variate data?
15. Write down the relationship between correlation and regression co-efficients.
16. Calculate correlation co-efficient from the following data.

x:	80	76	72	68	64	60
y:	73	59	66	45	52	38

 (AHSEC 1993)
 [Ans. .88]

17. The regression line of y on x is $y = 3 - x$, state the condition under which the regression line of x on y can be written as $x = 3 - y$. **(AHSEC 1992)**
18. Write down the equation of the regression line of y on x. For what purpose do we use it? **(AHSEC 1992)**
19. Explain what do you mean by the statement 'regression co-efficient of y on x is $-.63$'. **(AHSEC 1993)**
20. At what point do the regression lines intersect? **(AHSEC 1993)**

[Ans. $(\bar{x},\bar{y})$]

21. Why are there two regression lines? Explain clearly.

(AHSEC 1993, 1996, 1999, 2000)

22. If $r_{xy} = .6$ and $b_{yx} = .8$, what is the value of b_{xy}? **(AHSEC 1996)**

[Ans. .45]

23. Show that the correlation co-efficient is the G.M. of the co-efficients of regression. **(AHSEC 1996)**
24. Write down the equation of lines of regression of Y on X and X on Y.

(AHSEC 1997, 1999, 2006)

25. Write down the line of regression of Y on X. Explain why there are two lines of regression. Show that with usual notation,

$$r = \pm\sqrt{b_{xy}.b_{yx}}$$ **(AHSEC 2004)**

26. Fill up the gaps:
    (i) If $r = \pm 1$, the two lines of regression are mutually. ________.

    (AHSEC 1998)

    (ii) Two independent variables are __________.

    (AHSEC 2001)

    (iii) If b_{yx} and b_{xy} are two regression co-efficients then $b_{yx}.b_{xy} =$ ________.

    (AHSEC 2003)

    (iv) The correlation co-efficient is independent of change of _____ and _____.

    (AHSEC 1996, 2004, 2006)

    (v) When one regression co-efficient is positive then other would be _______.

    (AHSEC 1996, 2005)

    (vi) Correlation co-efficient lies between __________ and ___________.

    (AHSEC 2005)

    (vii) The lines of regression are _______ if $r = 0$ and they are _______ if r = $\pm$ 1.

    (AHSEC 1995)

    (viii) If one of the regression co-efficient is ______unity, the other must be _______ unity. **(AHSEC 1995)**

(ix) Regression co-efficients are independent of change in __________ but not of __________.

(x) Both the regression co-efficients can not __________ unity.

[Ans. (i) coincide, (ii) uncorrelated, (iii) r^2_{xy}, (iv) origin, scale, (v) positive, (vi) –1 and +1, (vii) perpendicular, coincide, (viii) greater than, less than, (ix) origin, scale, (x) exceed]

27. To study the relationship between the expenditure on lodging Rs X and on fooding Rs Y, an enquiry into 50 families revealed the following results

$\Sigma X = 8500$, $\Sigma Y = 9600$, $\sigma_x = 60$, $\sigma_y = 20$, $r = .6$

Estimate the expenditure on fooding when the expenditure on lodging is Rs. 200. **[Ans. 198]**

28. Find the regression equation of y on x and that of x on y. Also find the value of y when $x = 3$ and the vlaue of x when y = 5

x:	10	20	30	40	50	60
y:	4	12	20	24	32	38

[Ans. $y = 1.8 + .52x$, $x = 1 + 1.7y$, $y = 3.36$, $x = 9.5$]

29. Find the co-efficient of correlation of the heights of mothers and daughters from the following:

Height of Mothers:	65	66	67	68	69	70	71
Height of Daughters :	67	68	66	69	72	72	69

[Ans. .67 (approx)]

30.

(a) Given $b_{xy} = .85$, $b_{yx} = .89$ and $\sigma_x = 6$, find the value of r and σ_y.

[Ans. $r = .87$, $\sigma_y = 6.14$]

(b) You are given

$r_{xy} = .6$, $\bar{x} = 10$, $\bar{y} = 20$, $\sigma_x = 1.875$, $\sigma_y = 2.5$,

find the line of regression of x on y **[Ans. $x = 1 + .45\,y$]**

(c) Two lines of regression are $x + 2y - 5 = 0$ and $2x + 3y - 8 = 0$ and var(x) = 12. Calculate $\bar{x}$, $\bar{y}$, σ_y^2 and r. **[Ans. 1, 2, 35.998, –.87]**

31. What is the relationship between correlation coefficient and regression coefficients? **(AHSEC 2008)**

[Ans. r = $\pm\sqrt{b_{yx}.b_{xy}}$]

32. Define Karl Pearson's correlation coefficient. State its properties. **[AHSEC 2006]**

33. What are correlation and regression? **(AHSEC 2007)**

8

INDEX NUMBERS

1.1 INTRODUCTION

An index number is a statistical measure which is designed to express changes in a variable or a group of related variables under two different situations. They are usually expressed in percentage form. The comparisons may be between the periods of time, between places or other characteristics.

We can also say that an index number is a number which indicates the level of a certain phenomenon at any given time in comparison with the level of the same phenomenon at some standard time. They are constructed with reference to a base period and thus very useful in comparing changes over different time periods. For example if we say that index of a certain industry in 2004 was 125 with the base year 2000, it would imply that the production has increased by 25% in 2004 as compared to 2000. Such an index computed for only one item is called a univariate index. If we like to construct an index for the industrial sector as a whole, it will be a composite index because here a number of variables have to be combined.

A few definitions of index number are given below :

1. "A special type of average which provides a measurement of relative changes from time to time or from place to place".—Wesell and Willett.
2. "An index number is a stastical measure designed to show changes in a variable or group of related variables with respect to time, geographical location or other characteristics such as income, profession etc."—Spiegel

8.2 USES OF INDEX NUMBERS

The theory of Index Numbers is widely applied in economics and business. They serve the following purposes.

(i) Index numbers are very useful in studying trends of a series over a given period of time and thus facilitates forecasting.

(ii) Index numbers provide the basic criteria to make policies.

(iii) Index numbers help in measuring the purchasing power of money.

(iv) They are used in computing real wages through the process of deflating.

(v) Index numbers are Economic Barometers. Indices of prices, output, foreign exchange, bank deposits etc. act as barometers to find out ups and downs in the general economic conditions of a country. Barometers are used to measure atmospheric pressure and index numbers are called economic barometer as they measure the pressure of economic and business behaviour.

8.3 TYPES OF INDEX NUMBERS

Index number may be classified in terms of the variables that they measure. They are generally classified into three categories.

1. Price Index Number

The most common index numbers are the price index numbers. They study changes in the price level of commodities over a period of time. They are of two types:

(a) Wholesale Price Index Number

They depict changes in the general price level of the economy. The first series of the index number was constructed by the Government of India in 1947 with August 1939 as the base year.

(b) Retail Price Index Number

They reflect changes in the retail prices of different commodities. They are constructed for different classes of consumers.

2. Quantity Index Number

They reflect changes in the volume of goods produced or consumed. Some of the quantity index numbers are—index number of agricultural production, index numbers of industrial production, indices of exports and imports etc.

3. Value Index Number

They study changes in the total value (Price × quantity). For example, index number of profits or sales is a value index number.

8.4 PROBLEMS IN THE CONSTRUCTION OF INDEX NUMBERS

Since the basic approach in the construction of all types of index numbers is the same, we shall discuss the problems of constructing price index. The construction of index numbers involves the consideration of the following important points.

1. Purpose

There must be precise statement about the purpose of constructing index numbers. All index numbers will not serve the same purpose and there is no all purpose index number. The other steps in the construction of index number will mainly depend on the purpose. So, the purpose for which the index number is constructed should be clearly stated.

2. Selection of base period

The base period is a previous period with which comparison of some latter period is made. The index for the base period is taken as 100. The following points should be borne in mind while selecting a base period.

(a) Base period should be a normal period i.e., it should be free from all sorts of abnormalities such as floods, famines, earthquakes, strikes, epidemics etc. i.e., base period should be a period of economic stability. But in practice, it is rarely obtained. Usually the base period is of one year and is called the base year of the index number.

(b) Whether fixed base method or chain base method is adopted should be decided before constructing index numbers.

(c) Base period should not be a period of distant past as due to the passes of time new commodities enter the market and old one disappears.

(d) Base period should not be too short or too long.

3. Selection of Commodities

In constructing any index number it is not possible to include all the items or commodities. The commodities should have the representing capacity. A sufficiently large sample of commodities should be selected to obtain reliable index numbers but that would become costly. Again too small number of items will not reflect the information properly. As such number of commodities should neither be too large or too small. While selecting sample of commodities instead of random sampling, judgement or purposive sampling should be applied. Those commodities should be included which are stable in quality and preferably should be standardized because comparison of the relative change is possible only if the quality of the commodities does not change much.

4. Collection of Data

The raw data for the construction of index numbers are the prices of the selected commodities together with their quantities used. The data may be either primary or secondary depending upon the purpose. The price of a commodity varies from place to place, market to market, shop to shop and even within the same shop from customer to customer. It is not possible to obtain prices from all places where a particular commodity is sold. Even from selected places, prices cannot be collected from all shops. A selection of representative places and of representative persons has to be made. The task of collecting data can be done either by appointing staff or by giving to task to some individuals of that particular locality. If necessary, it is better to appoint more than one person in each selected place.

The information published in standard newspapers and journals or magazines about the prices can also be considered.

5. Selection of Average

The prices of various commodities have to be combined to arrive at a single index and this is done with the help of averages. For constructing an index number any average such as AM, median, mode, GM and HM can be used. Among all the averages, GM is the only average that gives appropriate results in case of relative measurement of changes. Thus, although GM is difficult to compute it gives quite satisfactory results. Hence GM is generally used in the construction of index numbers. Since each of the averages has its own merits and demerits, the selection of an average is also a problem for constructing index numbers.

6. Assignment of Weights

The assignment of weights to selected commodities is very important. 'Weight' means the relative importance of the items. All the commodities included in the index number are not of equal importance. Thus, different weights are to be assigned to different items in order that the index is true representative. For example, the importance of rice is no doubt more than cosmetics. There are two methods of constructing index numbers (i) Unweighted method and (ii) weighted method. In case of unweighted method the weights of all the commodities are equal and is taken as 1. There are two basis of weighting:

(i) Quantity of the items consumed.

(ii) Value of the items consumed.

When quantity forms the basis of weight it is called 'quantity weighting' and in case of value it is called 'value weighting'.

Weights may be either explicit or implicit. Explicit weighting implies that weights are laid down on the basis of importance of items. Implicit weighting implies the inclusion of a commodity in an index number more than once.

7. Selection of Appropriate Formula

There are various formulae for constructing index numbers and each formula has its own limitations. The selection of formula mainly depends on the purpose of constructing index number and the collected data. Hence selection of appropriate formula is also a problem in the construction of index number.

8.5 METHODS OF CONSTRUCTING INDEX NUMBERS

Generally base year is denoted by the symbol '0' and current year by '1'. So,

p_0 = Price of the commodity in the base year

p_1 = Price of the commodity in the current year

q_0 = Quantity in the base year

q_1 = Quantity in the current year

P_{01} = Price index number for the current year compared to the base year.

Index numbers may be classified as below:

- Index Number
  - Unweighted
    - Simple Aggregative method
    - Simple Average of Price Relatives
  - Weighted
    - Weighted Aggregative method
    - Weighted Average of Price Relatives

Unweighted Index Numbers

1. *Simple Aggregative Method*

In this method the sum of the prices of the commodities in the current year is divided by the sum of the prices of those commodities in the base year and is multiplied by 100. Symbolically,

$$P_{01} = \frac{\Sigma p_1}{\Sigma p_0} \times 100$$

2. *Simple Average of Price Relatives*

A price relative is the price of the current year expressed as a percentage of the price of the base year.

$$\text{Price Relative} = \frac{p_1}{p_0} \times 100$$

The construction of index number according to this method involves calculation of price relatives and then averaging these by applying AM or GM.

(i) If AM is applied, the index number formula is

$$P_{01} = \frac{\Sigma\left(\frac{p_1}{p_0} \times 100\right)}{N}$$

where N = Number of commodities.

(ii) If GM is applied,

$$P_{01} = \text{Antilog}\left[\frac{\Sigma \log\left(\frac{p_1}{p_0} \times 100\right)}{N}\right]$$

Weighted Index Numbers

1. ***Weighted Aggregative Method***

In this method, the index number is calculated by the formula of weighted arithmetic mean.

$$P_{01} = \frac{\Sigma p_1 W}{\Sigma p_0 W} \times 100 \qquad \text{..........(*)}$$

where W = Weight

Different formulae can be derived from (*) by giving different weights.

(a) *Laspeyre's Method*

Here W = base period quantity = q_0

$$\therefore \qquad P_{01} = \frac{\Sigma p_1 q_0}{\Sigma p_0 q_0} \times 100$$

(b) *Paasche's Method*

Here W = Current year quantity = q_1

$$\therefore \qquad P_{01} = \frac{\Sigma p_1 q_1}{\Sigma p_0 q_1} \times 100$$

(c) *Marshall-Edgeworth Method*

Here W = Average of the quantities in the base and current year = $\dfrac{q_0 + q_1}{2}$

$$\therefore \qquad P_{01} = \frac{\Sigma p_1 q_0 + \Sigma p_1 q_1}{\Sigma p_0 q_0 + \Sigma p_0 q_1} \times 100$$

(d) *Fisher's Method*

This is the GM of Laspeyre's and Paasche's index number.

$$\therefore \qquad P_{01} = \sqrt{\frac{\Sigma p_1 q_0}{\Sigma p_0 q_0} \times \frac{\Sigma p_1 q_1}{\Sigma p_0 q_1}} \times 100$$

This formula is commonly known as 'Fisher's Ideal index number'.

Besides all these there are various other index numbers also such as Dorbish-Bowley's index number, Walsch's index number, Kelly's index number etc.

2. *Weighted Average of Price Relatives*

Using AM the formula is

$$P_{01} = \frac{\Sigma IW}{\Sigma W}$$

where $I = \frac{p_1}{p_0} \times 100$, $W = p_0 q_0$

Using GM, the formula is

$$P_{01} = \text{Antilog}\left[\frac{\Sigma W.\log I}{\Sigma W}\right]$$

Quantity Index Numbers

The only difference in the formulae for quantity index numbers as compared to the formulae for price index numbers is that in place of p we write q and in place of q we write p

(a) *Laspeyre's Quantity Index*

$$Q_{01} = \frac{\Sigma q_1 p_0}{\Sigma q_0 p_0} \times 100$$

(b) *Paasche's Quantity Index*

$$Q_{01} = \frac{\Sigma q_1 p_1}{\Sigma q_0 p_1} \times 100$$

(c) *Fisher's Quantity Index*

$$Q_{01} = \sqrt{\frac{\Sigma q_1 p_0}{\Sigma q_0 p_0} \times \frac{\Sigma q_1 p_1}{\Sigma q_0 p_1}} \times 100$$

(d) *Marshall-Edgeworth Quantity Index*

$$Q_{01} = \frac{\Sigma q_1 p_0 + \Sigma q_1 p_1}{\Sigma q_0 p_0 + \Sigma q_0 p_1} \times 100$$

Value Index Numbers

Value of all commodities in the base year is $\Sigma p_0 q_0$ and value of all commodities in the current year is $\Sigma p_1 q_1$. The value index is defined as

$$V_{01} = \frac{\Sigma p_1 q_1}{\Sigma p_0 q_0} \times 100$$

8.6 TESTS FOR GOOD INDEX NUMBER

1. Time Reversal Test

According to Prof. Irving Fisher, an index number must satisfy two tests:

(a) Time Reversal Test

(b) Factor Reversal Test

Time Reversal Test

This test requires that if in an index number formula base and current year be interchanged i.e., P_{01} is converted by P_{10}, then one should be reciprocal of the other i.e., their product should be unity. (ignoring 100)

Symbolically, the test may be expressed as

$$P_{01} \times P_{10} = 1$$

This test is not satisfied by the Laspeyre's and Paasche's index numbers but is satisfied by Marshall-Edgeworth and Fisher's index number.

(a) Laspeyre's price index number, omitting 100

$$P_{01}^{L} = \frac{\Sigma p_1 q_0}{\Sigma p_0 q_0}$$

Interchanging base (0) and current (1) year

$$P_{10}^{L} = \frac{\Sigma p_0 q_1}{\Sigma p_1 q_1}$$

$$\therefore \quad P_{01}^{L} \times P_{10}^{L} \neq 1$$

Hence, the test is not satisfied by Laspeyre's index number.

(b) Paasche's price index number

$$P_{01}^{P} = \frac{\Sigma p_1 q_1}{\Sigma p_0 q_1}$$

$$\therefore \quad P_{10}^{P} = \frac{\Sigma p_0 q_0}{\Sigma p_1 q_0}$$

$$\therefore \quad P_{01}^{P} \times P_{10}^{P} \neq 1$$

Hence, the test is not satisfied by Paasche's index number.

(c) Fisher's index number

$$P_{01}^{F} = \sqrt{\frac{\Sigma p_1 q_0}{\Sigma p_0 q_0} \times \frac{\Sigma p_1 q_1}{\Sigma p_0 q_1}}$$

$$P_{10}^{F} = \sqrt{\frac{\Sigma p_0 q_1}{\Sigma p_1 q_1} \times \frac{\Sigma p_0 q_0}{\Sigma p_1 q_0}}$$

$$\therefore \quad P_{01}^{F} \times P_{10}^{F} = \sqrt{\frac{\Sigma p_1 q_0}{\Sigma p_0 q_0} \times \frac{\Sigma p_1 q_1}{\Sigma p_0 q_1} \times \frac{\Sigma p_0 q_1}{\Sigma p_1 q_1} \times \frac{\Sigma p_0 q_0}{\Sigma p_1 q_0}}$$

$$= \sqrt{1} = 1$$

Thus, Fisher's index number satisfies Time Reversal Test.

(d) Marshall-Edgeworth index number

$$P_{01}^{ME} = \frac{\Sigma p_1 q_0 + \Sigma p_1 q_1}{\Sigma p_0 q_0 + \Sigma p_0 q_1}$$

$$P_{10}^{ME} = \frac{\Sigma p_0 q_1 + \Sigma p_0 q_0}{\Sigma p_1 q_1 + \Sigma p_1 q_0}$$

$$\therefore \quad P_{01}^{ME} \times P_{10}^{ME} = \frac{\Sigma p_1 q_0 + \Sigma p_1 q_1}{\Sigma p_0 q_0 + \Sigma p_0 q_1} \times \frac{\Sigma p_0 q_1 + \Sigma p_0 q_0}{\Sigma p_1 q_1 + \Sigma p_1 q_0} = 1$$

Hence, Marshall-Edgeworth formula satisfies this test.

Factor Reversal Test

This test requires that an index number formula should be such that it allows the interchange of the price and the quantity without giving inconsistent result i.e., the two results multiplied together should give the value index number. An index number should satisfy the following equation in order to satisfy Factor Reversal Test.

$$P_{01} \times Q_{01} = V_{01}$$

$$= \frac{\Sigma p_1 q_1}{\Sigma p_0 q_0} \text{ (ignoring 100)}$$

Most of the formulae of index number fail to satisfy this test. Fisher's index number satisfies the test.

(a) Laspeyre's price index number

$$P_{01}^{L} = \frac{\Sigma p_1 q_0}{\Sigma p_0 q_0}$$

Interchanging price and quantity

$$Q_{01}^{L} = \frac{\Sigma q_1 p_0}{\Sigma q_0 p_0}$$

$$\therefore \quad P_{01}^{L} \times Q_{01}^{L} = \frac{\Sigma p_1 q_0}{\Sigma p_0 q_0} \times \frac{\Sigma q_1 p_0}{\Sigma q_0 p_0} \neq \frac{\Sigma p_1 q_1}{\Sigma p_0 q_0}$$

Hence, Laspeyre's formula does not satisfy Time Reversal Test.

(b) Paasche's index number

$$P_{01}^{P} = \frac{\Sigma p_1 q_1}{\Sigma p_0 q_1}$$

$$Q_{01}^{P} = \frac{\Sigma q_1 p_1}{\Sigma q_0 p_1}$$

$$\therefore \quad P_{01}^{P} \times Q_{01}^{P} = \frac{\Sigma p_1 q_1}{\Sigma p_0 q_1} \times \frac{\Sigma q_1 p_1}{\Sigma q_0 p_1} \neq \frac{\Sigma p_1 q_1}{\Sigma p_0 q_0}$$

Thus, Paasche's formula does not satisfy Factor Reversal Test.

(c) Fisher's index number

$$P_{01}^{F} = \sqrt{\frac{\Sigma p_1 q_0}{\Sigma p_0 q_0} \times \frac{\Sigma p_1 q_1}{\Sigma p_0 q_1}}$$

$$Q_{01}^{F} = \sqrt{\frac{\Sigma q_1 p_0}{\Sigma q_0 p_0} \times \frac{\Sigma q_1 p_1}{\Sigma q_0 p_1}}$$

$$\therefore \quad P_{01}^{F} \times Q_{01}^{F} = \sqrt{\frac{\Sigma p_1 q_0}{\Sigma p_0 q_0} \times \frac{\Sigma p_1 q_1}{\Sigma p_0 q_1} \times \frac{\Sigma q_1 p_0}{\Sigma q_0 p_0} \times \frac{\Sigma q_1 p_1}{\Sigma q_0 p_1}}$$

$$= \frac{\Sigma p_1 q_1}{\Sigma p_0 q_0}$$

Hence, Fisher's formula satisfies Factor Reversal Test.

(d) Marshall-Edgeworth index number

$$P_{01}^{ME} = \frac{\Sigma p_1 q_0 + \Sigma p_1 q_1}{\Sigma p_0 q_0 + \Sigma p_0 q_1}$$

$$Q_{01}^{ME} = \frac{\Sigma q_1 p_0 + \Sigma q_1 p_1}{\Sigma q_1 p_1 + \Sigma q_0 p_1}$$

$$\therefore \quad P_{01}^{ME} \times Q_{01}^{ME} = \frac{\Sigma p_1 q_0 + \Sigma p_1 q_1}{\Sigma p_0 q_0 + \Sigma p_0 q_1} \times \frac{\Sigma q_1 p_0 + \Sigma q_1 p_1}{\Sigma q_1 p_1 + \Sigma q_0 p_1} \neq \frac{\Sigma p_1 q_1}{\Sigma p_0 q_0}$$

Hence, Marshall-Edgeworth formula does not satisfy Factor Reversal Test.

Example 1: Construct price index numbers from the following data using Laspeyre's, Paasche's, Marshall-Edgeworth and Fisher's method.

	1990		1993	
Commodity	**Price**	**Quality**	**Price**	**Quantity**
A	**6**	**50**	**10**	**56**
B	**2**	**100**	**2**	**120**
C	**4**	**60**	**6**	**60**
D	**10**	**30**	**12**	**24**

Solution:

Commodity	1990 p_0	1990 q_0	1993 p_1	1993 q_1	p_1q_0	p_0q_0	p_1q_1	p_0q_1
A	6	50	10	56	500	300	560	336
B	2	100	2	120	200	200	240	240
C	4	60	6	60	360	240	360	240
D	10	30	12	24	360	300	288	240
					Σp_1q_0 **= 1420**	Σp_0q_0 **= 1040**	Σp_1q_1 **= 1448**	Σp_0q_1 **= 1056**

(i) Laspeyre's price index $= \frac{\Sigma p_1 q_0}{\Sigma p_0 q_0} \times 100$

$$= \frac{1420}{1040} \times 100$$

$$= 136.54$$

(ii) Paasche's price index $= \frac{\Sigma p_1 q_1}{\Sigma p_0 q_1} \times 100$

$$= \frac{1448}{1056} \times 100$$

$$= 137.12$$

(iii) Marshall-Edgeworth price index $= \frac{\Sigma p_1 q_0 + \Sigma p_1 q_1}{\Sigma p_0 q_0 + \Sigma p_0 q_1} \times 100$

$$= \frac{1420+1448}{1040+1056} \times 100$$

$$= 136.83$$

(iv) Fisher's price index $= \sqrt{\frac{\Sigma p_1 q_0}{\Sigma p_0 q_0} \times \frac{\Sigma p_1 q_1}{\Sigma p_0 q_1}} \times 100$

$$= \sqrt{\frac{1420}{1040} \times \frac{1448}{1056}} \times 100$$

$$= \sqrt{1.3654 \times 1.3712} \times 100$$

$$= 1.3683 \times 100$$

$$= 136.83$$

Comment: Fisher's price index shows that price increases to 36.83% in 1993 as compared to 1990.

Example 2: Calculate Fisher's ideal index from the following data and show that it satisfies both the Time Reversal Test and Factor Reversal Test.

	2000		2004	
Commodity	**Price**	**Value**	**Price**	**Value**
A	**10**	**100**	**12**	**96**
B	**8**	**96**	**8**	**104**
C	**12**	**144**	**15**	**120**
D	**20**	**300**	**25**	**250**
E	**5**	**40**	**8**	**64**
F	**2**	**20**	**4**	**24**

Solution:

Before calculating the index first we divide value by price to get quantity

$$\text{Quantity} = \frac{\text{Value}}{\text{Price}}$$

Commodity	p_0	q_0	p_1	q_1	$p_0 q_0$	$p_0 q_1$	$p_1 q_0$	$p_1 q_1$
A	10	10	12	8	100	80	120	96
B	8	12	8	13	96	104	96	104
C	12	12	15	8	144	96	180	120
D	20	15	25	10	300	200	375	250
E	5	8	8	8	40	40	64	64
F	2	10	4	6	20	12	40	24
					$\Sigma p_0 q_0 = 700$	$\Sigma p_0 q_1 = 532$	$\Sigma p_1 q_0 = 875$	$\Sigma p_1 q_1 = 658$

Fisher's Ideal Index

$$P_{01}^{F} = \sqrt{\frac{\Sigma p_1 q_0}{\Sigma p_0 q_0} \times \frac{\Sigma p_1 q_1}{\Sigma p_0 q_1}} \times 100$$

$$= \sqrt{\frac{875}{700} \times \frac{658}{532}} \times 100$$

$$= 1.2434 \times 100$$

$$= 124.34$$

Time reversal test is satisfied when

$$P_{01} \times P_{10} = 1$$

Here $$P_{10}^{F} = \sqrt{\frac{\Sigma p_0 q_1}{\Sigma p_1 q_1} \times \frac{\Sigma p_0 q_0}{\Sigma p_1 q_0}}$$

$$= \sqrt{\frac{532}{658} \times \frac{700}{875}}$$

Also, $$P_{01}^{F} = \sqrt{\frac{\Sigma p_1 q_0}{\Sigma p_0 q_0} \times \frac{\Sigma p_1 q_1}{\Sigma p_0 q_1}}$$

$$= \sqrt{\frac{875}{700} \times \frac{658}{532}}$$

$$\therefore \quad P_{01}^{F} \times P_{01}^{F} = \sqrt{\frac{875}{700} \times \frac{658}{532} \times \frac{532}{658} \times \frac{700}{875}}$$

$$= \sqrt{1} = 1$$

Hence, time reversal test is satisfied.

Factor reversal test is satisfied when

$$P_{01} \times Q_{01} = \frac{\Sigma p_1 q_1}{\Sigma p_0 q_0}$$

$$Q_{01}^{F} = \sqrt{\frac{\Sigma q_1 p_0}{\Sigma q_0 p_0} \times \frac{\Sigma q_1 p_1}{\Sigma q_0 p_1}}$$

$$= \sqrt{\frac{532}{700} \times \frac{658}{875}}$$

$$\therefore \quad P_{01}^{F} \times Q_{01}^{F} = \sqrt{\frac{875}{700} \times \frac{658}{532} \times \frac{532}{700} \times \frac{658}{875}}$$

$$= \frac{658}{700}$$

$$= \frac{\Sigma p_1 q_1}{\Sigma p_0 q_0}$$

Hence, factor reversal test is also satisfied by the given data.

Example 3: From the data given below find the price index using

(i) Arithmetic mean of price relatives.

(ii) Geometric mean of price relatives.

Price (In Rs.)

Commodity	1995	2000
A	30	60
B	40	40
C	10	15
D	28	56

Solution:

Commodity	p_0	p_1	$\frac{p_1}{p_0} \times 100 = I$	$\log I$
A	30	60	200	2.30
B	40	40	100	2.00
C	10	15	150	2.17
D	28	56	200	2.30
			$\Sigma\left(\frac{p_1}{p_0}\right) \times 100 = 650$	$\Sigma \log I = 8.77$

(i) $$P_{01} = \frac{\Sigma\left(\frac{p_1}{p_0} \times 100\right)}{N} = \frac{650}{4} = 162.5$$

(ii) $$P_{01} = \text{Antilog}\left(\frac{\Sigma \log I}{N}\right)$$

$$= \text{Antilog}\left(\frac{8.77}{4}\right)$$

$$= \text{Antilog}\, 2.1925$$

$$= 155.78$$

Example 4: Given the following data, compute Fisher's price index number.
(AHSEC 1999)

	Base year		Current year	
Commodity	Price	Commodity	Price	Commodity
A	5	7	7	4
B	3	2	4	3
C	1	5	1	5
D	4	4	6	3

Solution:

Commodity	p_0	q_0	p_1	q_1	p_1q_0	p_0q_0	p_1q_1	p_0q_1
A	5	7	7	4	49	35	28	20
B	3	2	4	3	8	6	12	9
C	1	5	1	5	5	5	5	5
D	4	4	6	3	24	16	18	12
					$\Sigma p_1q_0 = 86$	$\Sigma p_0q_0 = 62$	$\Sigma p_1q_1 = 63$	$\Sigma p_0q_1 = 46$

Fisher's price index number

$$P_{01}^{F} = \sqrt{\frac{\Sigma p_1q_0}{\Sigma p_0q_0} \times \frac{\Sigma p_1q_1}{\Sigma p_0q_1}} \times 100$$

$$= \sqrt{\frac{86}{62} \times \frac{63}{46}} \times 100$$

$$= \sqrt{1.899} \times 100$$

$$= 1.3780 \times 100$$

$$= 137.80$$

$\therefore$ Price increases to 37.80% in the current year compared to the base year.

Example 5: Calculate price index number from the following data.
(AHSEC 1993)

Commodity	Price		Weight
	Base year	Current year	
A	200	225	64
B	120	100	24
C	90	80	9
D	27	30	3

Solution:

Commodity	p_0	p_1	$I = \frac{p_1}{p_0} \times 100$	W	IW
A	200	225	112.50	64	7200
B	120	100	83.33	24	1999.92
C	90	80	88.89	9	800.01
D	27	30	111.11	3	333.33
				$\Sigma W = 100$	$\Sigma IW = 10333.26$

$$\therefore \qquad P_{01} = \frac{\Sigma IW}{\Sigma W}$$

$$= \frac{10333.26}{100}$$

$$= 103.33$$

Example 6: Reconstruct the following indices using 1980 as base.

Year	**1976**	**1977**	**1978**	**1979**	**1980**	**1981**	**1982**
Index Number	**110**	**130**	**150**	**175**	**180**	**200**	**220**

Solution:

Year	Index Number	Index Number Base 1980 = 100
1976	110	$\frac{110}{180} \times 100 = 61.11$
1977	130	$\frac{130}{180} \times 100 = 72.22$
1978	150	$\frac{150}{180} \times 100 = 83.33$
1979	175	$\frac{175}{180} \times 100 = 97.22$
1980	180	$\frac{180}{180} \times 100 = 100$
1981	200	$\frac{200}{180} \times 100 = 111.11$
1982	220	$\frac{220}{180} \times 100 = 122.22$

Example 7: Calculate (i) Laspeyre's Price Index (ii) Laspeyre's Quantity Index from the given data, taking 2001 as base year:

	2001		2003	
Commodity	**Price**	**Quantity**	**Price**	**Quantity**
A	**4**	**10**	**5**	**12**
B	**6**	**8**	**7**	**10**
C	**10**	**5**	**12**	**4**
D	**3**	**12**	**4**	**15**
E	**5**	**7**	**5**	**8**

(AHSEC 2007)

Solution:

Laspeyre's price index, $P_{01}^{L} = \dfrac{\Sigma p_1 q_0}{\Sigma p_0 q_0} \times 100$

Laspeyre's quantity index, $Q_{01}^{L} = \dfrac{\Sigma q_1 p_0}{\Sigma q_0 p_0} \times 100$

Commodity	p_0	q_0	p_1	q_1	p_1q_0	q_1p_0	p_0q_0
A	4	10	5	12	50	48	40
B	6	8	7	10	56	60	48
C	10	5	12	4	60	40	50
D	3	12	4	15	48	45	36
E	5	7	5	8	35	40	35
					$\Sigma p_1q_0 = 249$	$\Sigma q_1p_0 = 233$	$\Sigma p_0q_0 = 209$

(i) $P_{01}^{L} = \dfrac{\Sigma p_1 q_0}{\Sigma p_0 q_0} \times 100$

$= \dfrac{249}{209} \times 100$

$= 119.12$

(ii) $Q_{01}^{L} = \dfrac{\Sigma q_1 p_0}{\Sigma q_0 p_0} \times 100$

$= \dfrac{233}{209} \times 100$

$= 111.48$

8.7 COST OF LIVING INDEX NUMBER (CLIN)

Cost of living index numbers are also known as consumer price index numbers are intended to measure the changes in the prices paid by the consumers for purchasing a basket of goods and services during the current year as compared to the base year. The general price index numbers fail to give an exact idea on the effect of change on the cost of living of different groups of population. The consumption pattern of different classes of people e.g., rich, poor, middle class etc. varies widely. Again within the same class of people, there is variation in consumption from place to place e.g., rural, urban, hills etc. Hence in the construction of CLIN it is necessary to specify the class of people and the geographical areas covered.

Construction of CLIN

The main steps in the construction of CLIN consists of:

1. *Scope and coverage*

The class of the people and the area to which the CLIN is related must be carefully determined. The class of people selected should form a homogeneous group i.e., people having similar consumption habits.

2. *Family Budget Enquiry*

The next step involves selecting a sample from this group for conducting family budget survey. The expenditure incurred in consuming various commodities and services received in the base and the current period are classified in the five groups—Food, Clothing, Fuel and Lighting, House rent and Miscellaneous. These groups may be further subdivided into various sub-groups. For example Food groups may be sub-divided into rice, wheat, milk, sugar etc.

3. *Collection of Retail Prices*

Another problem in the construction of CLIN is the collection of retail prices of the commodities in the index. The prices are to be collected both for the base period and the current period. The retail prices are to be collected from those markets where the particular class of people makes purchases. But this is not an easy job as prices differ from market to market.

The other steps in the construction of CLIN such as selection of base period, selection of weights and choice of formula etc. are similar to those which have already been discussed in the construction of price index numbers.

Methods of Construction of CLIN

The following are the two methods of constructing cost of living index:

(a) *Aggregate Expenditure Method*

In this method, quantities consumed in the base year are taken as weights. The CLIN is

$$\text{CLIN} = \frac{\Sigma p_1 q_0}{\Sigma p_0 q_0} \times 100$$

$$= \frac{\text{Total expenditure in the current year}}{\text{Total expenditure in the base year}} \times 100$$

This is, infact, the Laspeyre's index.

(b) *Family Budget Method*

In this method, the CLIN is calculated by taking the weighted average of price relatives. The weights are the values of the quantities consumed in the base year.

$$\text{CLIN} = \frac{\Sigma IW}{\Sigma W}$$

where $I = \frac{p_1}{p_0} \times 100$ for each item

$W = p_0 q_0$ = value weights

Uses of CLIN

The following are some uses of CLIN:

(i) CLIN are used for adjustment of dearness allowances (D.A.) to compensate the rise in the price level.

(ii) They are used in fixing wage policy, in various economic policies etc.

(iii) Reciprocal of CLIN is used as a measure of the purchasing power of money.

(iv) Real wages can be measured with the help of CLIN.

$$\text{Real wages} = \frac{\text{Actual wage}}{\text{CLIN}}$$

Example 8: Find the cost of living index number from the following data and interpret your result. (AHSEC 2005, 2008)

Article	Weight	Price Relative
Food	24.9	1.02
House Rent	27.5	1.28
Clothing	9.4	0.83
Transportation	10.8	1.14
Fuel	5.3	1.00
Miscellaneous	22.1	1.18

Solution:

Article	Weight (W)	$I = \frac{p_1}{p_0} \times 100$	IW
Food	24.9	102	2539.8
House Rent	27.5	128	3520.0
Clothing	9.4	83	780.2
Transportation	10.8	114	1231.2
Fuel	5.3	100	530.0
Misc.	22.1	118	2607.8
	$\Sigma W = 100$		$\Sigma IW = 11209$

$$\therefore \quad \text{CLIN} = \frac{\Sigma IW}{\Sigma W}$$

$$= \frac{11209}{100} = 112.09$$

Interpretation: Thus there is an increase of 12.09% in the cost of living in the current year compared to the base year.

Example 9: From the following data calculate consumer price index number.

Group	Weight	Group Indices
Food	**60**	**410**
Clothing	**5**	**450**
Fuel and Lighting	**7**	**300**
House rent	**10**	**370**
Miscellaneous	**18**	**280**

(AHSEC 1997)

Solution:

Group	Weight (W)	Group Indices (I)	IW
Food	60	410	24600
Clothing	5	450	2250
Fuel and Lighting	7	300	2100
House Rent	10	370	3700
Misc.	18	280	5040
	$\Sigma W = 100$		$\Sigma IW = 37690$

$$\therefore \quad \text{CLIN} = \frac{\Sigma IW}{\Sigma W}$$

$$= \frac{37690}{100} = 376.90$$

Example 10: Calculate the cost of living index number from the following data.

Items	Price		Weight
	Base year	Current year	
Food	39	47	4
Fuel	8	12	1
Clothing	14	18	3
House Rent	12	15	2
Misc.	25	30	1

Solution:

Items	p_0	p_1	W	$I = \frac{p_1}{p_0} \times 100$	IW
Food	39	47	4	120.51	482.04
Fuel	8	12	1	150.00	150.00
Clothing	14	18	3	128.57	385.71
House Rent	12	15	2	125.00	250.00
Misc.	25	30	1	120.00	120.00
			$\Sigma W = 11$		$\Sigma IW = 1387.75$

$$\therefore \quad \text{CLIN} = \frac{\Sigma IW}{\Sigma W}$$

$$= \frac{1387.75}{11}$$

$$= 126.16$$

Example 11: From some given data the cost of living index based on five groups, viz., Food, House rent, Fuel and Light, Clothing and Miscellaneous was calculated as 205. The percentage increase in price over the base period is given below.

House rent	**60**
Clothing	**210**
Fuel and Light	**120**
Misc.	**130**

Calculate the percentage increase in the food group, given that the weights of different groups were as follows:

Food 60, House Rent 16, Fuel and Light 12, Clothing 8, Misc. 4.

Solution:

Let the percentage increase in prices of the food group be x.

Commodity	**Percentage increase in price**	**Current index (I)**	**Weight (W)**	IW
Food	x	$100 + x$	60	$6000 + 60x$
House Rent	60	160	16	2560
Clothing	210	310	8	2480
Fuel and Light	120	220	12	2640
Misc.	130	230	4	920
			ΣW **= 100**	ΣIW **= 14600 + 60x**

$$\text{CLIN} = \frac{\Sigma IW}{\Sigma W}$$

$$\Rightarrow \quad 205 = \frac{60x + 14600}{100}$$

$$\Rightarrow \quad 60x = 5900$$

$$\Rightarrow \quad x = 98.33$$

Thus, percentage increase in price of food group is 98.33.

Example 12: A family budget enquiry gives the following information on average family expenditure on various items of food groups.

	Rs.
Cereals	**480.00**
Edible oil, Pulses, Spices	**180.00**
Vegetables	**160.00**
Egg, Meat and Fish	**150.00**
Milk, Tea and Sugar	**150.00**
Others	**80.00**
	1200.00

Find the weights of different items for constructing price-index for food group of the budget. (AHSEC 1992)

Solution:

Items	**Expenditure**	**Weights**
Cereals	480	$\frac{480}{1200} \times 100 = 40$
Edible oil, Pulses, Spices	180	$\frac{180}{1200} \times 100 = 15$
Vegetables	160	$\frac{160}{1200} \times 100 = 13.3$
Egg, Meat, Fish	150	$\frac{150}{1200} \times 100 = 12.5$
Milk, Tea, Sugar	150	$\frac{150}{1200} \times 100 = 12.5$
Others	80	$\frac{80}{1200} \times 100 = 6.7$
	1200	

Example 13: An employee of an industry earns Rs. 350 per month. The cost of living index for a particular month is given as 136. Using the following data find out the amounts he spent in house rent and clothing.

Group	**Food**	**Clothing**	**House rent**	**Fuel**	**Misc.**
Expenditure	**140**	**?**	**?**	**56**	**63**
Group Indices	**180**	**150**	**100**	**110**	**80**

Solution:

Let the expenditure on clothing and house rent be x_1 and x_2.

$\therefore$ $140 + x_1 + x_2 + 56 + 63 = 350$

$\Rightarrow \quad x_1 + x_2 = 350 - 259$

$\Rightarrow \quad x_1 + x_2 = 91 \qquad \text{...(i)}$

Again $\quad \text{CLIN} = \frac{\Sigma IW}{\Sigma W}$

$$\Rightarrow \quad 136 = \frac{140 \times 180 + 150x_1 + 100x_2 + 56 \times 110 + 63 \times 80}{350}$$

$\Rightarrow \quad 150x_1 + 100x_2 = 47600 - 36400$

$\Rightarrow \quad 150x_1 + 100x_2 = 11200$

$\Rightarrow \quad 3x_1 + 2x_2 = 224 \qquad \text{...(ii)}$

Solving (i) and (ii) we get,

$$x_1 = 42$$

and $$x_2 = 49$$

Hence, expenditure on clothing is 42 and expenditure on house rent is 49.

8.8 LIMITATIONS OF INDEX NUMBERS

The index numbers have their own limitations although they are very much useful. Some of them are:

(i) They are only approximate indicators since index numbers are based on sample data.

(ii) Each step in the construction of index number such as collection of data, selection of base period, selection of commodities, assignment of weight etc. is full of possibilities of errors of all types.

(iii) With the help of index numbers, same commodities are compared in two different situations which is quite unrealistic.

(iv) Index numbers are liable to be misinterpreted.

(v) Different formulae give different results which is impracticable.

EXERCISE–8

1. What do you mean by the term Index Number? Write down the name of the different indices. **(AHSEC 1999, 2000, 2001)**
2. What is price-index number? Mention the problems that are involved in the construction of an index number of prices. **(AHSEC 1995, 1996, 1999)**
3. Mention the different uses of index numbers. **(AHSEC 1999, 2000, 2001, 2004, 2006)**
4. Discuss the various steps in the construction of consumer price index. **(AHSEC 1995, 2000, 2006)**
5. Explain the time reversal and the factor reversal tests of an index number. **(AHSEC 1998, 2000)**
6. What are the limitations of index number?
7. What do you mean by consumer price index? What are its uses? **(AHSEC 1998, 1999, 2001)**
8. Why Fisher's index number formula is called an ideal index number formula? **(AHSEC 1999, 2003, 2008)**
9. "Index numbers are economic-barometers". Explain. **(AHSEC 1997, 2003)**
10. Write a note on selection of base period for construction of price index number. **(AHSEC 1992, 2005)**
11. Show that Fisher's index number formula satisfies both the Time Reversal and Factor Reversal Tests. **(AHSEC 1992, 1998, 2000, 2003, 2008)**
12. Describe aggregative method of construction of price index numbers. **(AHSEC 1993)**
13. What is cost of living index number? Discuss the role of family budget enquiry in construction of cost of living index number. **(AHSEC 1992, 1993, 1996, 1997)**
14. Give a brief description of the various types of index numbers. **(AHSEC 1996)**
15. Discuss briefly the problems of selection of base period and commmodities for construction of price index number. **(AHSEC 1997, 2000)**
16. Write down the formula for the following index numbers:
    (i) Laspeyre's price index number
    (ii) Paasche's price index number
    (iii) Fisher's ideal price index number
    (iv) Value index number

    (AHSEC 1997, 1999, 2005)

17. What are price relatives? How would you combine them to obtain price index number? **(AHSEC 1998)**
18. Explain the difficulties in the construction of cost of living index number. **(AHSEC 2000)**
19. Write any one formula for cost of living index number. For a place cost of living index number comes out to be 376.65. Now how will you interpret the result? **(AHSEC 2004)**
20. Explain why selection of weights are necessary for the construction of index numbers. Discuss how different index number formula can be obtained from the weighted aggregative index number. **(AHSEC 2004)**
21. Write a detailed note on family budget enquiry in connection with the construction of cost of living index number. **(AHSEC 2004)**
22. What does the cost of living index number measure? **(AHSEC 2003)**
23. Using Fisher's ideal index number formula calculate the price index number.

	1960		*1975*	
Commodities	*Price*	*Quantity*	*Price*	*Quantity*
A	4	50	10	40
B	3	10	9	12
C	2	5	4	3
D	5	20	6	8

(AHSEC 1996)
[Ans. 224.36]

24. Compute the consumer's price index, given the following information.

Groups	*Weights*	*Group indices*
Food	48	320
Fuel and Lighting	12	180
Clothing	10	210
House rent	20	250
Misc.	10	150

(AHSEC 1999)
[Ans. 261.20]

25. Fill in the blanks:
    (a) Fisher's index number is the ____________ of Laspeyre's and Paasche's indices.
    (b) If the price of a commodity in the current year be 50% in excess over the base year price, then the price index for the commodiy is __________.

[Ans. (a) Geometric mean (b) 150]

26. Construct cost of living index number for the year 1996 from the following data.

	1983		1996
Commodity	*Price*	*Quantity*	*Price*
A	25	16	35
B	36	7	48
C	12	3.5	16
D	6	2.5	10
E	28	4	28

[Ans. 151]

27. Compute quantity index for the following data using weighted aggregate of quantities.

	1990		2000
Commodity	*Price*	*Quantity*	*Price*
A	40	20	25
B	30	15	20
C	10	10	10
D	20	5	10

[Ans. 131]

28. What are the tests of a good index number? **(AHSEC 2006)**

9

VITAL STATISTICS

Demography is defined as the study of measurement of human population. It is the quantitative study of human population with respect to events such as birth, sex, age, caste, literacy, worker, income, marital status, natality and fecundity, migration, death etc. The word 'Demography' was introduced by a Belgian 'Achille Guillard'. Demography has many faces and vital statistics is one of them.

9.1 VITAL STATISTICS

The vital events are all those which may occur to an individual during his life time such as birth, death, marriage, separation, sickness, adoption etc. By the help of vital events, the health and growth of a community can be studied. Vital statistics refers to the data or the techniques applied for the analysis of vital events occuring in a community.

A few definitions of vital statistics are quoted below:

1. "Vital statistics are conventional numerical records of marriage, births, sickness and deaths by which the health and growth of a community may be studied"— Benjamin.
2. "The branch of Biometry which deals with data and the laws of human mortality, morbidity and demography"— Arthur Newsholme.

9.2 USES OF VITAL STATISTICS

Vital statistics are useful from different points of view. Some of them are

(i) Vital statistics help us in analysing demographic trends in the country.

(ii) Vital statistics fill the gap in the census method and supplement it.

(iii) Knowledge of vital events makes it easy to know the future population figures with the help of population projection.

(iv) The registration of vital events give us an idea of the requirements of the health service like hospital, maternity and child welfare centres etc.

(v) The study of the vital statistics enables the government to asses the impact of family planning on population growth.

(vi) Vital statistics are useful in disease eradication programmes.

(vii) Vital statistics are beneficial in studying the social situation prevalent in the society.

(viii) They are useful to Actuaries in framing insurance policies and in the calculation of premiums.

9.3 METHODS OF COLLECTING VITAL STATISTICS

The following are the two methods of collecting vital statistics.

1. Census Method

Census is the enumeration of the total population of a given region at a given time. Census is normally conducted after every 10 years in almost all the countries. Census usually yields data on age, sex, education, occupation, marital status, religion and other vital statistics. Data collected by this method fail to produce vital statistics for intercensal years. Also the data recorded in respect of births and deaths are incomplete even for the census year. There are two important census methods– De Facto method and De Jure method.

2. Registration Method

Under this method every birth, death and marriage is required to be registered. Unlike census, it continues throughout the year and as such no shortage of investigator is felt. Registration has legal importance as it provides information about the succession rights and settlement of dispute about birth and death. Registration is done with the proper authorities as appointed by the government of a country. Registration of birth supplies information with regard to the age, place, name of the parents, their religion, occupation, education, income etc. Whereas the death registration supplies informations regarding cause of death, sex, age, locality, religion, profession, place of death etc. But this method suffers from the drawback that a large number of births and deaths are not reported to the registration office.

9.4 VITAL STATISTICS: RATES AND RATIOS

A rate is a fraction such that the numerator is included within the denominator. The rate is a number between 0 and 1. In vital statistics these rates are normally expressed per some convenient base such as 1000 or 1,000,00, a large enough multiple of 10 being chosen so that the resulting rate is greater than 1. The rate describe the rapidity with which a given event is occuring.

$$\text{Rate of a vital event} = \frac{\text{No. of occurance of the vital event}}{\begin{array}{c}\text{Total no. of persons exposed to the risk}\\ \text{of the occurance of the event}\end{array}} \times 1000$$

The study of data that are qualitative in many ways simpler than is the study of measurement data. The count of the no. of individuals with a given qualitative characteristic is compared with the count of the no. having a second qualitative characteristic by forming a ratio. For example

$$\text{Sex Ratio} = \frac{\text{No. of females}}{\text{No. of males}}$$

9.5 MEASUREMENT OF MORTALITY

From demographic point of view, mortality denotes the fall in size of population on account of death. Study of mortality and its measurement is not only important from the demographic analysis but also from the point of health standard to achieve better health facilities.

The mortality conditions of a population are studied by measuring the following mortality rates:

(i) Crude Death Rate (CDR)

CDR is a measure of the fall in population which takes place as a result of death. "CDR may be defined as the ratio of the number of deaths which occur within a population during a specified year to the size of that population at mid year".– David M. Heer. The following formula is used to estimate the C.D.R.

$$\text{C.D.R.} = \frac{\sum D}{\sum P} \times 1000$$

where ΣD = Total death in a given region at a given time.

ΣP = Total mean population of the same region during the given time.

Merits:

It is very easy and as such can be understood easily by a common man. It is the simplest method of finding out death rate because it requires only the total population and number of deaths. It is a probability rate.

Demerits:

The most serious drawback of C.D.R. is that it does not consider the age and sex distribution of the population. We all know that mortality varies from age to age e.g. the death rate is high in infants and old persons, but C.D.R. completely ignores this fact. CD.R. is not suitable for comparing mortality conditions of two different regions since the population structure of two different regions are not generally alike.

Specific Death Rate

A death rate calculated on the basis of a specified section of the population is called a specific death rate. Death rate may be specific with respect to age, sex or some other characteristics.

$$\text{Specific Death Rate} = \frac{\text{No. of deaths in a specified section of the population of an area in a given period}}{\text{Mean population of that specific area in the same period}} \times 1000$$

Let

${}_nD_x$ = No. of deaths in the age group $(x, x + n)$ in the given region during a given period.

${}_nP_x$ = No. of persons in the age group $(x, x + n)$ in the given region during a given period.

${}_nm_x$ = Age specific Death Rate (ASDR) for the above age group.

Then, $${}_nm_x = \frac{{}_nD_x}{{}_nP_x} \times 1000$$

If we put $n = 1$, we get annual – ASDR.

Now ASDR for males and females can be formulated as

$${}_n^m m_x = \frac{{}_n^m D_x}{{}_n^m P_x} \times 1000$$

$${}_n^f D_x = \frac{{}_n^f D_x}{{}_n^f P_x} \times 1000$$

Merits:

Specific death rates are the true and best measure of mortality. These rates can be compared from one population to the other or from one time to another in the same population. They supply one of the essential components required for constructing life tables and net reproduction rate.

Demerits:

It does not eliminate other characteristics, such as race, occupation, locality of dwelling etc.

Standardised Death Rate

Sometimes it is not possible to compare the C.D.Rs. in two regions as the actual age distribution in one region may differ from another region. For the purpose of comparison the rates are brought on same standard basis by computing standardised death rate. The standard population for this purpose may be taken as the population of the country or some other population may be taken as the standard.

There are two methods of standardisation –Direct method and Indirect method.

Direct Method

Suppose we want to compare the mortality of two regions A and B. The C.D.R. in terms of annual–ASDR for region A and B are given by

$$m^a = \frac{\sum_x m_x^a P_x^a}{\sum_x P_x^a}$$

$$m^b = \frac{\sum_x m_x^b P_x^b}{\sum_x P_x^b}$$

Now if P_x^s is the number of persons at age x in the standard population, then

$$S_T DR \text{ for region } A = \frac{\sum_x m_x^a P_x^s}{\sum_x P_x^s}$$

$$S_T DR \text{ for region } B = \frac{\sum_x m_x^b P_x^s}{\sum_x P_x^s}$$

The direct method is very simple but the choice of the standard population affects the magnitude of the standardised death rate.

Indirect Method

Sometimes one comes across a situation in which the specific death rates of a population are not known but the distribution of population in various age groups is known. In such a case, an indirect method is adopted. Here approximate standardised death rate of the population is calculated by the relation

$$S_T DR_A = CDR_A \times F$$

where F = correction factor

$$= \frac{S_T DR_A}{CDR_A}$$

$$= \frac{\dfrac{\sum_x m_x^a P_x^s}{\sum_x P_x^s}}{\dfrac{\sum_x m_x^a P_x^a}{\sum P_x^a}}$$

Now replacing m_x^a by m_x^s (annual–ASDR for standard population) we get

$$\hat{F} = \frac{\dfrac{\sum_x m_x^s P_x^s}{\sum_x P_x^s}}{\dfrac{\sum_x m_x^s P_x^a}{\sum_x P_x^a}}$$

$$\therefore \quad S_T DR_A = CDR_A \times \hat{F}$$

Example 1 : From the following data calculate Crude Death Rate.

Age group : (in years)	**0-10**	**10-30**	**30-50**	**50-70**	**70 and above**
Population :	**10,000**	**20,000**	**30,000**	**20,000**	**5,000**
Deaths :	**250**	**40**	**80**	**400**	**2,000**

Solution:

$$CDR = \frac{\sum D}{\sum P} \times 1000$$

$$= \frac{2,770}{85,000} \times 1000$$

$$= 32.58 \text{ per thousand.}$$

Example 2 : On the basis of the figures given below calculate Age Specific Death Rate for all age groups. Also calculate Crude Death Rate on the basis of ASDR.

Age groups : (in years)	**0-10**	**10-30**	**30-50**	**50-70**	**70 and above**
Population :	**5,000**	**10,000**	**13,000**	**10,000**	**2,000**
Death :	**125**	**25**	**35**	**200**	**1,000**

Solution:

Age group	Population $_nP_x$	Deaths $_nD_x$	ASDR $_nm_x = \frac{_nD_x}{_nP_x} \times 1000$	$_nm_x\, _nP_x$
0-10	5,000	125	25	1,25,000
10-30	10,000	25	2.5	25,000
30-50	13,000	35	2.69	34,970
50-70	10,000	200	20	2,00,000
70 and above	2,000	1000	500	10,00,000
	Σ_nP_x = 40,000			$\sum_x {_nm_x}{_nP_x}$ = 13,84,970

$$\text{CDR in terms of ASDR} = \frac{\sum_x {_nm_x}{_nP_x}}{\sum_x {_nP_x}} = \frac{13,84,970}{40,000}$$

$$= 34.62 \text{ per thousand}$$

Example 3 : On the basis of following data, calculate Crude Death Rate for locality *A* and *B* and Standard Death Rate for *A* considering *B* as standard.

	Locality A		Locality B	
Age group	Population (in thousand)	Death	Population (in thousand)	Death
0-10	20	500	10	370
10-20	10	260	30	650
20-40	50	1140	62	1400
40-60	30	1050	15	700
Above 60	10	500	3	280

Solution:

	Locality A			Locality B			
Age group	Population (in '000) (P_x^a)	Death (D_x^a)	Death rate (m_x^a)	Population (in '000) (P_x^b)	Death (D_x^b)	Death rate (m_x^b)	$m_x^a\, p_x^b$
0-10	20	500	25	10	370	37	250
10-20	10	260	26	30	650	21.67	780
20-40	50	1140	22.8	62	1400	22.58	1413.6
40-60	30	1050	35	15	700	46.67	525
above 60	10	500	50	3	280	93.33	150
Total	**120**	**3450**		**120**	**3400**		**3118.6**

$$\text{C.D.R. for } A = \frac{\sum_x D_x^a}{\sum_x P_x^a} \times 1000$$

$$= \frac{3450}{1,20,000} \times 1000$$

$$= 28.75 \text{ per thousand}$$

$$\text{C.D.R. for } B = \frac{\sum_x D_x^b}{\sum_x P_x^b} \times 1000$$

$$= 28.33 \text{ per thousnad}$$

Since locality B is taken as standard

$$S_T DR \text{ for } A = \frac{\sum_x m_x^a P_x^b}{\sum_x P_x^b} \qquad [\therefore P_x^s = p_x^b]$$

$$= \frac{3118.6 \times 1000}{1,20,000} = 25.99 \text{ per thousand}$$

Example 4 : Calculate standardised death rate from the data given below

Age	Standard population ('000)	Population ('000)	Number of Death
< 10	20	12	300
10-20	12	20	600
20-40	60	64	1600
40-60	20	25	500
> 60	10	3	150

(AHSEC 2007)

Solution:

Age	P_x^s	P_x^a	D_x^a	Death rate m_x^a	$m_x^a\ P_x^s$
<10	20	12	300	25	500
10-20	12	20	600	30	360
20-40	60	64	1600	25	1500
40-60	20	25	500	20	400
>60	10	3	150	50	500
	122				**3260**

$$S_T DR = \frac{\sum_x m_x^a P_x^s}{\sum_x P_x^s}$$

$$= \frac{32,60,000}{1,22,000}$$

$= 26.72$ per thousand.

Example 5 : Calculate crude and standardised death rates of the local population.

Age	Standard Population	Local Populattion	No. of Deaths
0-10	600	400	16
10-20	1000	1500	6
20-60	3000	2400	24
60-100	400	700	21

(AHSEC 2004)

Solution:

Age	P_x^s	P_x^L	D_x^L	m_x^L	$m_x^L\ p_x^s$
0-10	600	400	16	40	24000
10-20	1000	1500	6	4	4000
20-60	3000	2400	24	10	30000
60-100	400	700	21	30	12000
Total	**5,000**	**5,000**	**67**		**70,000**

$$\text{CDR for Local Population} = \frac{\sum_x D_x^L}{\sum_x P_x^L} \times 1000$$

$$= \frac{67}{5000} \times 1000$$

$= 13.4$ per thousand

$$S_T DR = \frac{\sum_x m_x^L P_x^s}{\sum P_x^s}$$

$$= \frac{70,000}{5,000} = 14 \text{ per thousand}$$

9.6 MEASUREMENT OF FERTILITY

Fertility is a result of fecundity, the physiological capacity to reproduce. Fertility can exist only among fecund women. In demography, the term fertility refers to the actual production of children. Fertility must be distinguished from fecundity. In the measurement of population growth 'fertility' has a significant role. Fertility is a concept which is primarily concerned with the number of live births. Some important fertility rates are discussed below:

1. Crude Birth Rate (CBR)

The CBR for any given population is obtained by dividing the number of live births recorded during a period by its total population and multiplying this fraction by 1000. Thus CBR is the number of live births per thousand of population.

$$\text{CBR} = \frac{\sum B}{\sum P} \times 1000$$

where

ΣB = Total no. of live births registered during the calender year

ΣP = Total mean population of the same region during the same period.

This rate is called crude because all the differences in the composition between population are ignored in calculating it.

Merits:

It is easy to find CBR because figures which are needed for finding out it are easily available.

Demerits:

It is not a probability rate since in the denominator whole population is taken which is not exposed to the risk of producing children. Also the age and sex factors of the population are not considered here.

2. General Fertility Rate (GFR)

GFR is defined as the number of children born alive per thousand women of child bearing age, in a given area during a given period.

$$\text{GFR} = \frac{\text{No. of live births in a given area during a given period}}{\text{No. of women of child bearing age in the given region during the given period}} \times 1000$$

In our country child bearing age is taken as 15-49 years. Symbolically,

$$\text{GFR} = \frac{\sum B}{\sum_{x=15}^{49} {}^{f}P_x} \times 1000$$

Merits:

It is a probability rate. It takes into account the sex composition of the population and also the age composition to a certain extent.

Demerits:

The fecundity of women in all age groups is not the same. In our country fecundity is low in the age group 15-19 but it increases very rapidly in the age group 20-24 and only slightly in 25-29 after which it gradually declines. Also marital status of women has an impact on fertility rates. These facts are not considered in GFR.

3. Age-Specific Fertility Rate (ASFR)

Here fertility rates are calculated seperately for various age-groups of females of child bearing age. ASFR may be defined as the number of live births during a year, per thousand women of a particular age group. ASFR is given by

$$ {}_n i_x = \frac{{}_n B_x}{{}_n^f P_x} \times 1000 $$

where

${}_n B_x$ = No. of live births to females in the age group $(x, x + n)$ in the given region during the given period.

${}_n^f P_x$ = No. of females in the age group $(x, x + n)$ in the given region during the given period.

${}_n i_x$ = ASFR in the above age group

If we put $n = 1$ in the formula of ${}_n i_x$ we get annual ASFR

$$ i_x = \frac{B_x}{{}^f P_x} \times 1000 $$

Merits:

In ASFR it is considered that women in all ages do not have the same reproductive capacity and that changes with ages. With the help of ASFR, Total Fertility Rate (TFR) is possible to calculate.

Demerits:

It is very difficult to compare the overall fertility situations of two regions by ASFR because it is very likely that the rates will be higher or lower in one region than the other in some age groups.

4. Total Fertility Rate (TFR)

TFR is the total of ASFR and can be obtained by summing up of birth rates at each age group throughout the child bearing age. The TFR gives the total number of expected births to 1000 females in their entire child-bearing span of life if none of these 1000 females dies before passing through the child bearing age and all are subject to the observed fertility rates throughout the period,

$$\text{TFR} = \sum_{x=15}^{49} i_x$$

$$= \sum_{x=15}^{49} \frac{B_x}{{}^f P_x} \times 1000$$

Merits:

TFR is an index of the overall fertility of the community.

Demerits:

TFR is a hypothetical figure since the above mentioned assumptions are completely hypothetical.

9.7 MEASUREMENT OF POPULATION GROWTH

While calculating fertility rates, the sexes of the newly born babies are not taken care of. But the population growth depends mainly on the birth of female children who are the future mothers. Hence, population growth is mainly a function of fertility rate, restricted to female children. Reproduction rates are of two types.

1. Gross Reproduction Rate (GRR)

GRR gives the number of daughters that would be born by women arriving at and passing through the child bearing period. This rate is based on the following assumptions.

(i) A newly born female child is not subject to mortality till she attains the highest reproductive age i.e., 49 years.

(ii) A newly born female is subject to the given fertility conditions throughout her highest child bearing age.

(iii) There is no gain or loss due to migration.

GRR is calculated by summing annual female ASFR

$$\text{GRR} = \sum_{x=15}^{49} \frac{{}^f B_x}{{}^f P_x} \times 1000$$

$$= \sum_{x=15}^{49} {}^f i_x$$

Where ${}^f B_x$ = No. of live births (female) to females in the age group $(x, x+1)$ in the given region during the given period.

GRR may be calculated from the formula

$$\text{GRR} = \frac{\text{No. of female births}}{\text{Total birth}} \times \text{TFR}$$

If, for the child bearing age, the population of women and number of female children are given in 5-yearly age groups, then

$$\text{GRR} = 5\ \Sigma i_x$$

One serious limitation of GRR is that it does not take into account the current mortality factor. All the female children born do not survive till they reach the child bearing age. For this GRR is called a hypothetical figure.

2. Net Reproduction Rate (NRR)

In computing the GRR we observed that mortality was not taken into consideration. The GRR is computed on the assumption that all newly born females will live throughout the child bearing period. But it is possible that the newly born females die before attaining the age of child bearing or may not marry or become widows. So, the GRR is to be corrected. By taking the factor mortality we get another measure of population growth called NRR.

If ${}^fL_x$ is the number of females at age x out of the original group of females ${}^fl_o$, then

$$\text{NRR} = \sum_{x=15}^{49} {}^f i_x \frac{{}^fL_x}{{}^fl_o} \qquad \left(\frac{{}^fL_x}{{}^fl_o} \text{ are called survivorship values}\right)$$

If NRR = 1, then it shows that current fertility and mortality rates are such that the population will have tendency to remain constant. If NRR > 1, then it may be said to show a tendency to increase. If NRR < 1 then it may be said to show a tendency to decrease in the population.

NRR as a measure of replacement of population is not much relied upon because it does not take the factor migration into account.

Example 6 : You are given the following data.

Age group :	**15-19**	**20-24**	**25-29**	**30-34**	**35-39**	**40-44**	**45-49**
Female Population : (in '000)	**34**	**36**	**40**	**30**	**24**	**20**	**16**
No. of live births :	**680**	**3,960**	**5,800**	**3,000**	**1,680**	**800**	**80**

The total population in the city in 1991 was 7,00,000. Calculate Crude Birth Rate.

Solution:

$$\text{CBR} = \frac{\sum B}{\sum P} \times 1000$$

$$= \frac{16,000}{7,00,000} \times 1000 = 22.86 \text{ per thousand}$$

Example 7 : Calculate GFR, ASFR, TFR from the given data.

Age-group	No. of women (in '000)	No. of live Births (in '000)
15-19	55	20
20-24	70	180
25-29	65	200
30-34	64	170
35-39	60	120
40-44	58	60
45-49	50	10

Solution:

Age Group	No. of women (in '000) ${}_n^fP_x$	No. of live births to women of age group (x, x +n) (in '000) ${}_nB_x$	ASFR ${}_ni_x = \frac{{}_nB_x}{{}_n^fP_x}$
15-19	55	20	363.6
20-24	70	180	2571.4
25-29	65	200	3076.9
30-34	64	170	2656.3
35-39	60	120	2000.0
40-44	58	60	1034.5
45-49	50	10	200.0
Total	**422**	**760**	**11902.7**

$$\text{GFR} = \frac{\sum {}_nB_x}{\sum_x {}_n^fP_x} \times 1000$$

$$= \frac{760}{422} \times 1000$$

= 1800.95 per thousand women

ASFR (See the last column of the table)

$$\text{TFR} = 5 \sum_x {}_ni_x$$

$$= 5 \times 11902.7$$

= 59513.5 per thousand women

Example 8 : Show that NRR<GRR.

Solution: If ${}^fL_x$ is the number of females at age x out of the original group of females ${}^fl_o$ (cohort) then

$${}^fL_x < {}^fl_o, \quad x \neq 0$$

$$\Rightarrow \qquad \frac{{}^fL_x}{{}^fl_o} < 1$$

Multiplying both sides by ${}^fi_x$ (ASFR) and summing both sides from 15-49, we get

$$\sum_{x=15}^{49} {}^fi_x \frac{{}^fL_x}{{}^fl_o} < \sum_{x=15}^{49} {}^fi_x$$

$$\Rightarrow \qquad \text{NRR} < \text{GRR}$$

Example 9 : Given TFR = 1070.75.

No. of female live birth = 100

Total no. of male live births = 105

GRR= ? **(AHSEC 1999)**

Solution:

$$\text{T.F.R. per woman} = \frac{1070.75}{1000}$$

$$\text{GRR per woman} = \frac{\text{No. of female birth}}{\text{Total birth}} \times \text{TFR}$$

$$= \frac{100}{100+105} \times \frac{1070.75}{1000}$$

$$= \frac{100}{205} \times \frac{1070.75}{1000}$$

$$= 0.52$$

Meaning of the result

If GRR = 1, this means that on the average each female will produce sufficient female offspring to replace herself only if such is not subject to mortality. In the given problem the rate shows that for one present mother there will be .52 in future and therefore the population will decrease.

Example 10 : You are given TFR = 5000 per thousand, male and female ratio 60 : 40. Find out GRR.

Solution: $\text{TFR per woman} = \dfrac{5000}{1000} = 5$

$$GRR = \frac{\text{No. of female birth}}{\text{Total birth}} \times TFR$$

$$= \frac{40}{100} \times 5$$

$$= 2 \text{ per woman}$$

Example 11 : Calculate GFR, ASFR, TFR and GRR from the data given below.

Age group :	**15-19**	**20-24**	**25-29**	**30-34**	**35-39**	**40-44**	**45-49**
No. of women : ('000)	**15.0**	**16.2**	**15.8**	**15.2**	**14.8**	**15.0**	**14.0**
Total birth :	**250**	**2243**	**1897**	**1320**	**915**	**280**	**145**

Assume that the proportion of female birth is 45.2 percent.

Solution:

Age group	No. of women ($_n^f P_x$)	No. of live birth to women in the age group (x, x +n) ($_nB_x$)	ASFR mix
15-19	15.0	250	16.67
20-24	16.2	2243	138.46
25-29	15.8	1897	120.06
30-34	15.2	1320	86.84
35-39	14.8	915	61.82
40-44	15.0	280	18.67
45-49	14.0	145	10.36
Total	**106**	**7050**	**452.88**

$$GFR = \frac{\sum {}_nB_x}{\sum {}_n^f P_x} \times 1000$$

$$= \frac{7050}{1,06,000} \times 1000$$

$$= 66.51 \text{ per thousand women}$$

$$TFR = 5 \times \Sigma \, {}_ni_x$$

$$= 5 \times 452.88$$

$$= 2264.4 \text{ per thousand women}$$

$$GRR = \frac{\text{No. of female birth}}{\text{Total birth}} \times TFR$$

$$= \frac{45.2}{100} \times 2264.4$$

$$= 1023.51 \text{ per thousand women}$$

Example 12 : From the following figures calculate the gross and net reproduction rates.

Age group (year)	No. of female children born to 1000 women	percent survival rate
15 – 19	200	85
20 – 24	365	80
25 – 29	300	70
30 – 34	188	65
35 – 39	115	60
40 – 44	42	50
45 – 49	5	45

Solution:

The gross reproduction rate is given by

$$\text{G.R.R.} = 5 \sum_x {}_5^f i_x$$

where $${}_5^f i_x = \frac{{}_5^f B_x}{{}_5^f P_x} \times 1000$$

${}_5^f B_x$ = female children born to the females in the age group $(x, x + 5)$

${}_5^f P_x$ = females in the age group $(x, x + 5)$

Thus the second column of the table gives us ${}_5^f i_x$

Again net reproduction rate is given by

$$\text{NRR} = 5 \sum_x {}_5^f i_x \text{ . Survival rate per female.}$$

Now we shall prepare the following table.

${}_5^f i_x$	survival rate per female	${}_5^f i_x$ × survival rate per female
200	0.85	170
365	0.80	292
300	0.70	210
188	0.65	122.2
115	0.60	69
42	0.50	21
5	0.45	2.25
1235		886.45

$$\therefore \quad \text{GRR} = 5 \times \sum_x {}_5^f i_x$$

$$= 5 \times 1235$$

$$= 6175 \text{ per thousand women}$$

$$\therefore \quad \text{NRR} = 5 \times 886.45$$

$$= 4432.25 \text{ per thousand women}$$

EXERCISE–9

1. What do you mean by vital statistics? What are the important sources of vital statistics? **(AHSEC 2000, 2007)**
2. Define the following terms:
   (i) Crude Birth Rate **(AHSEC 2000, 2001, 2004, 2006)**
   (ii) General Fertility Rate **(AHSEC 2000, 2002, 2005, 2007)**
   (iii) Age-Specific Fertility Rate. **(AHSEC 2000, 2006)**
   (iv) Standardised Death Rate. **(AHSEC 2001, 2003, 2006)**
   (v) Vital Statistics Rate. **(AHSEC 2001, 2006, 2007, 2008)**
   (vi) Total Fertility Rate. **(AHSEC 2002, 2003, 2005, 2007)**
   (vii) Vital Statistics. **(AHSEC 2003, 2006)**
   (viii) Age Specific Death Rate. **(AHSEC 2003)**
   (ix) Gross Reproduction Rate. **(AHSEC 1999)**
   (x) Net Reproduction Rate. **(AHSEC 1999)**
   (xi) Fertility and Fecundity.
   (xii) Crude Death Rate. **(AHSEC 2007)**
   (xiii) Specific Death Rate.
   (xiv) Vital event **(AHSEC 2008)**
   (xv) Vital ratio **(AHSEC 2006, 2007, 2008)**
3. Define Crude Death Rate. Why it is not an adequate measure of mortality? **(AHSEC 1999)**
4. Write short notes on:
   (i) Standardised Death Rate and its uses. **(AHSEC 1999)**
   (ii) Measurement of population growth.
5. Define NRR. Show that NRR cannot exceed unity. **(AHSEC 2001)**
6. Define General Fertility Rate and Total Fertility Rate. Which is better and why? **(AHSEC 2002)**
7. Define NRR. Interpret the results NRR <=> 1. **(AHSEC 2003, 2005, 2007)**
8. Explain why Standard Death Rates of different population are comparable while Crude Death Rates are not. **(AHSEC 2003)**
9. "GRR is a hypothetical figure"– Discuss. **(AHSEC 2004)**
10. Is CDR an accurate measure of mortality of population of a country? Give reasons for your answer. **(AHSEC 2005)**
11. Explain the gross and net reproduction rate. Discuss their suitability as measures of population growth.
12. Are there any defects in comparing mortality situations between two regions by means of their CDR's? If yes, explain the defects. How would you eliminate such type of defects?

13. Discuss the indirect method of standardisation of death rate.
14. "Specific Death Rates are the best measure of mortality"–Explain.
15. Differentiate between the following:
    (i) Curde Death Rate and Specific Death Rate.
    (ii) GRR and NRR.
    (iii) General Fertility Rate and Total Fertility Rate.
    (iv) Crude Death Rate and Standardised Death Rate.
    (v) General Fertility Rate and Gross Reproduction Rate.
16. From the following data compute GRR:

Age group of mother	No. of women ('000)	Total Births
15-19	16.0	260
20-24	16.5	2200
25-29	15.8	1895
30-34	15.0	1320
35-39	14.8	915
40-44	15.0	280
45-49	14.5	145

Assume that the proportion of female births is 46%. **[Ans. 1030.42]**

17. Given Total Fertility Rate = 2251 per thousand. Number of male births = 105, Number of female births = 100. Compute the value of GRR.

(AHSEC 2001)

[Ans. 1.09]

18. Find:
    (i) Crude Death Rate
    (ii) Standardised Death Rate for the population *A* from the following data.

Age group	Standard Population		Population A	
	Population ('000)	Specific Death Rate	Population ('000)	Specific Death Rate
0-5	8	50	7	48
5-15	10	15	12	14
15-50	27	10	25	9
>50	5	60	4	59

(AHSEC 2002)

[Ans. (i) 20.10, (ii) 21.24]

19. Explain the meaning of the statement 'NRR is 1105 per thousand'.

(AHSEC 2003)

20. Write down the formula for crude Death Rate and Standardised Death Rate. **(AHSEC 2004)**
21. Compute GRR and NRR from the data given below:

Age (years)	No. of women ('000)	No. of female births	Survival Rate
15-19	16.0	140	.969
20-24	16.4	1130	.967
25-29	15.8	980	.963
30-34	15.2	670	.958
35-39	14.2	460	.952
40-44	15.0	150	.942
45-49	14.5	80	.928

[Ans. GRR = 1.15834, NRR = 1.11195]

22. Calculate GRR from the following data:

Age group : (years)	15-19	20-24	25-29	30-34	35-39	40-44	45-49
No. of Children : born to 1000 women	180	1800	2200	1000	600	100	80

Given sex ratio of children born is Male : Female = 52 : 48.

[Ans. 14.30 per thousand women]

23. Calculate GFR, ASFR, TFR from the follwoing data

Age group :	15-19	20-24	25-29	30-34	35-39	40-44	45-49
Female Population : ('000)	32	30	28	26	24	22	18
Birth :	800	3420	4200	2860	1920	660	72

[Ans. G.F.R = 77.4, T.F.R. = 2565]

24. Briefly describe various measures of mortality. **(AHSEC 2008)**
25. Show that NRR cannot exceed GRR. **(AHSEC 2006)**

10

TIME SERIES

10.1 INTRODUCTION

The time series is an arrangement of statistical data in accordance with the time of its occurance. Time series refers to any statistical data collected at regular intervals of time. Time may be in terms of years, months, weeks, days and so on.

Mathematically, a time series may be defined as $y = f(t)$, where y is the value of the variable at time t. Thus, a time series gives relationship between two variables– one is independent variable time (t) and the other, dependent variable (y). For example the price of a commodity (y) in different years (t), the production (y) of tea in Assam in different years (t), the sales (y) of a store in different months (t) etc.

10.2 IMPORTANCE OF TIME-SERIES ANALYSIS

The analysis of Time Series is important to businessmen, economists, social scientists and also to people working in various other fields due to the following reasons.

(i) It helps in understanding the past behaviour of the variable.

(ii) It enables to isolate the various forces affecting the time series.

(iii) Analysis of time series helps in forecasting the future value of a variable.

(iv) A time series provides a scientific basis for making comparison between two or more sets of data.

(v) Time series helps in planning future operations.

10.3 COMPONENTS OF TIME SERIES

The changes in the values of a variable with respect to time can be the resultant effect of a variety of factors. These factors can be classified into four categories which are called the components of time series. Some or all of the components may be present in any time series with varying degrees. The components are as follows:

(a) Secular Trend or Trend

(b) Seasonal Variations

(c) Cyclical Variations

(d) Random or Irregular Variations.

Trend

The trend is the general long term movement in the time series value of the variable over a fairly long period of time. The trend of a series is generally either upward or downward in nature. For instance, the data relating to population, production, literacy etc have upward trend while the data relating to death rate, illiteracy, epidemic etc. have downward trend. However, such a trend may not always hold good. It is to be noted that depending on the nature of increase and decrease, trend may be linear as well as non-linear. Again long period of time can not be defined exactly and it would depend on the nature of the problem.

Uses

The study of trend helps a businessman in forecasting and planning his future business activities. It helps an economist in formulating economic policies. It is used in making comparison between two or more time series. By isolating trend from the given time series we can study the short-term and irregular variations.

Seasonal Variations

Seasonal variations refer those pattern of change in a time series that repeat over a period of one year or less and they repeat from year to year. They can be studied when the data are recorded at half yearly, quarterly, monthly, weekly etc. They cannot be studied when the data are recorded annually. Seasonal variations occur regularly season after season and therefore they are definite and can be foreseen with a little effort. They are due to the following two reasons.

(i) *Natural Cause*

Changes in the climate and weather conditions have a profound effect on seasonal variations. For example, the sales of umbrella pick up very fast in rainy season, the demand for cold drinks goes up during summer etc.

(ii) *Man-made Cause*

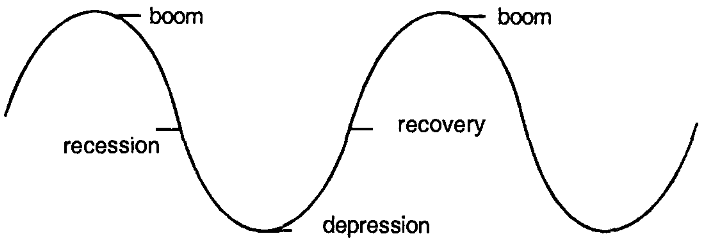
The social customs, traditions and conventions play an important role in fixing seasonal variations. For example, during festivals like Bihu, Puja, Id etc. the sales of many commodities go high, the price of ornaments go high during marriage seasons etc.

Uses

An accurate assessment of seasonal variations is an aid in business planning and scheduling such as in the area of production, inventory control, personnel, advertising and so on. A knowledge of seasonal variation helps a consumer in purchasing the articles at least prices during the slack seasons.

Cyclical Variations

The cyclical variations refer to the movements which occur after time intervals of more than one year. These are the changes that take place due to economic booms or depressions. The cyclical variations, though more or less regular are not uniformly periodic. One complete period normally mean 7 to 9 years. Series relating to prices, production, wages etc. are all affected by 'business cycles'. Business cycle is composed of four phases, namely-boom, recession, depression and recovery. Each phase change gradually into the next phase



Business cycle

Fig. 101

Uses

The knowledge of cyclical variations is of great importance to businessmen in the formulation of policies for stabilizing the business activity. With a knowledge of cyclical variation a businessman can face the challenges of recession and depression by taking the appropriate decisions in advance.

Random or Irregular Variations

As the name suggests, these variations are random, accidental or simply due to chance factors. These variations may be due to such isolated incidents as floods, famines, wars, earthquakes etc. which are completely irregular in nature. Thus, they are wholly unpredictable. This includes all types of variations in a time series which are not attributed to trend, seasonal or cyclical variations.

10.4 MODELS FOR TIME SERIES ANALYSIS

(i) Additive Model

This model assumes that all the component of the time series are independent of one another and each part can be measured independently. Here

$$Y = T + S + C + I$$

Where Y is the time series value and T, C, S and I are Trend, Seasonal, Cyclical and Irregular variations.

(ii) Multiplicative Model

This model assumes that the four components of the time series are due to different causes but they are not necessarily independent. The traditional time series analysis model is characterised by the multiplicative relationship.

$$Y = T \times S \times C \times I$$

10.5 MEASUREMENT OF TREND

To measure the secular trend, the short-term variations should be removed and irregularities should be smoothed out. The following are the methods of measuring trend.

1. Graphic (or free hand curve) method
2. Semi-average method
3. Moving average method
4. Least squares method

GRAPHIC METHOD

The values of the time series are plotted on a graph paper with the time (t) along x-axis and the values of the variable (y) along y-axis. A freehand curve is drawn through these points in such a manner that it may show a general trend. A free hand curve removes the short-term variations and irregular movements.

It is the simplest method. Time and labour is saved. It is very flexible method as it represents both linear and non-linear trends.

The main drawback of this method is that it is highly subjective as different persons will draw different free hand curves. Because of its subjective nature it is useless in forecasting.

Example 1 : The following are the figures for sales for last six years. Determine the trend by free hand curve method.

Year	Sales (Lakh units)
1992	50
1993	80
1994	105
1995	90
1996	150
1997	140

Solution:

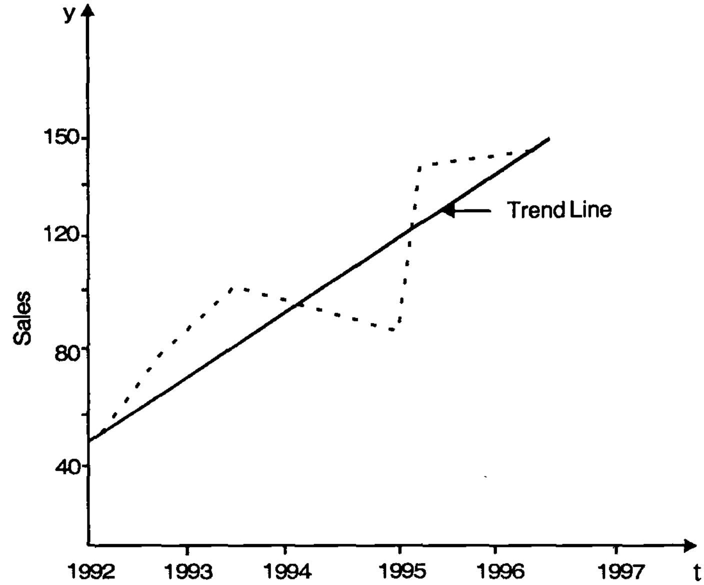


Fig. 10.2

Semi-Average Method

This method is sometimes used when a straight line appears to be an adequate expression of trend. In this method, the original data are divided into two equal parts. The averages of each part are then calculated. The average of each part is centred in the period of the time of the part from which it has been computed and then plotted on the graph paper. In this way, a line may be drawn to pass through the plotted points which gives the trend line. In case of odd number of years, the mid-year is eliminated while dividing the data into two equal parts.

This method is not subjective and everyone gets the same trend line. It is possible to extend the trend line both the ways to estimate future or past values. But the method assumes the presence of linear trend which may not exist.

Example 2 : The following are production of a company, fit a trend line by semi-average method.

Year :	**1960**	**1961**	**1962**	**1963**	**1964**	**1965**	**1966**
Production : ('000 units)	**40**	**60**	**50**	**70**	**55**	**75**	**50**

Solution:

Year	Production ('000 units)	Totals	Averages
1960	40		
1961	60	150	50
1962	50		
1963	70		
1964	55		
1965	75	180	60
1966	50		

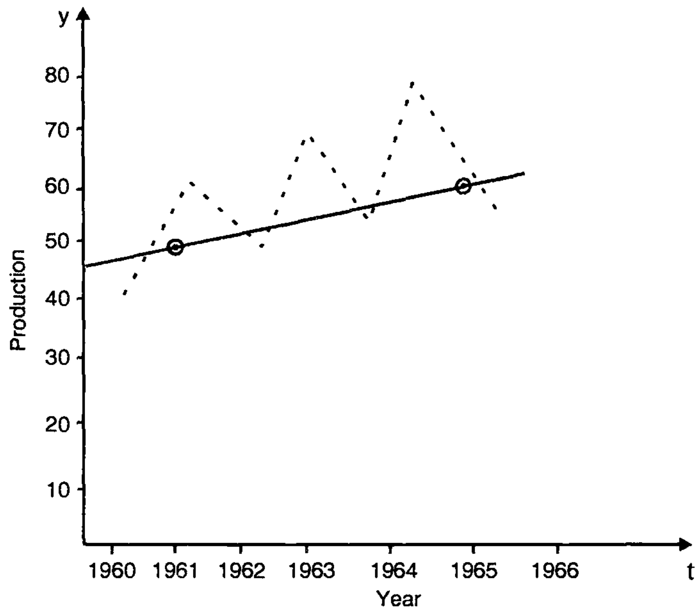


Fig. 10.3

Moving Average Method

Method of moving average consists in calculating a series of arithmetic means of some overlapping groups of the time series. There is no hard and fast rule for decision regarding period of moving average and the selection of period depends upon the objective that has to be attained from the figures of trend. The averaging process removes ups and downs in the data. If $y_1, y_2, y_3, \ldots$ are the values of the time series for time periods $t_1, t_2, t_3 \ldots$ then

$$\text{1st moving average of period } m = \frac{1}{m}(y_1 + y_2 + \ldots + y_m)$$

2nd moving average of period m $= \frac{1}{m} (y_2 + y_3 + ... + y_{m+1})$

3rd moving average of period m $= \frac{1}{m} (y_3 + y_4 + ... + y_{m+2})$ etc.

Now, if the period m is odd, then the moving average values are placed against the middle values of the time periods. Again if the period m is even, then the moving average values are placed in between two middle time periods. After that, method of centering is applied to adjust the data with the original time periods.

The moving average values plotted against time give the trend line.

Merits and Demerits of Moving Average Method

Merits:

(i) This method is easy to understand and calculate.

(ii) This method is not subjective.

(iii) This method is flexible in the sense that a few more observations may be included without affecting the previous results.

(iv) If the period of moving average is equal to or multiple of the cycle of a cyclical variation, then the effect of cyclical variations can be removed.

Demerits:

(i) Moving average can not be obtained for all the years.

(ii) There is no definite rule for fixing the period of moving average.

(iii) This method is suitable only when the trend is linear.

(iv) This method is not suitable for forecasting.

Example 3 : Calculate 3 yearly moving average from the following data:

Year :	t_1	t_2	t_3	t_4	t_5	t_6	t_7	t_8	t_9
Values :	y_1	y_2	y_3	y_4	y_5	y_6	y_7	y_8	y_9

Solution:

Year	Values	3 yearly moving totals	3 yearly moving average
t_1	y_1		—
t_2	y_2	$y_1 + y_2 + y_3$	$(y_1 + y_2 + y_3)/3$
t_3	y_3	$y_2 + y_3 + y_4$	$(y_2 + y_3 + y_4)/3$
t_4	y_4	$y_3 + y_4 + y_5$	$(y_3 + y_4 + y_5)/3$
t_5	y_5	$y_4 + y_5 + y_6$	$(y_4 + y_5 + y_6)/3$
t_6	y_6	$y_5 + y_6 + y_7$	$(y_5 + y_6 + y_7)/3$
t_7	y_7	$y_6 + y_7 + y_8$	$(y_6 + y_7 + y_8)/3$
t_8	y_8	$y_7 + y_8 + y_9$	$(y_7 + y_8 + y_9)/3$
t_9	y_9		—

Example 4 : Calculate 5 yearly moving average from the following data:

Year :	**1995**	**1996**	**1997**	**1998**	**1999**	**2000**	**2001**	**2002**	**2003**	**2004**
Production : (thousand tonnes)	**52**	**79**	**76**	**66**	**68**	**93**	**87**	**79**	**90**	**95**

Solution:

Year	Production (000 tonnes)	5 yearly moving total	5 yearly moving average
1995	52		—
1996	79		—
1997	76	341	68.2
1998	66	382	76.4
1999	68	390	78.0
2000	93	393	78.6
2001	87	417	83.4
2002	79	444	88.8
2003	90		—
2004	95		—

Example 5 : Assuming a four yearly cycle, calculate the trend by method of moving averages for the following data:

Year :	**1991**	**1992**	**1993**	**1994**	**1995**	**1996**	**1997**	**1998**	**1999**	**2000**
Sales : (in lakhs of Rs.)	**50**	**52**	**55**	**47**	**51**	**54**	**56**	**57**	**59**	**61**

Solution: Since we are given a four yearly cycle, we shall compute 4 yearly moving average to get the trend values.

Year	Sales (in lakhs of Rs.)	4 yearly moving totals	4 yearly moving average	Centred moving average
1991	50			—
1992	52			—
	....	204	51	
1993	55			
	....	51.13		
		....	205	51.25
1994	47		....	51.50
	....	207	.... 51.75	
1995	51		....	51.88
	....	208	 52	
1996	54		....	53.25
	....	218	.... 54.50	
1997	56		....	55.50
	....	226	.... 56.50	
1998	57		....	57.38
	....	233	.... 58.25	
1999	59			—
2000	61			—

Determination of Period of Moving Average

If a series contains cyclical variations, moving average with a period equal to the period of the cycle or its multiple will eliminate the fluctuations. If a series contains cyclical variations the period of which is not uniform, the average duration of the cycle should be calculated and that should be taken as the period of moving average. For determining the average duration, the data should be plotted and the distances between the various peak points should be noticed. The average of these time distances should be taken as the cycle.

Example 6 : Find the period of moving average from the given data:

Year :	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**
Sales : (in pieces)	**55**	**34**	**30**	**48**	**62**	**42**	**37**	**31**	**40**	**66**

Solution: Since the peak points of the series is noticed at the years 1, 5 and 10, the series gives a regular cycle of 5. Hence the period of moving average can be taken as 5.

Example 7 : A study of demand (d_i) for last 12 years (t = 1, 2, ..., 12) has indicated the following:

d_i **= 100; (t = 1, 2, ..., 5)**
= 20 ; (t = 6) **(AHSEC 2005)**
= 100 ; (t = 7, 8, ..., 12)

Compute 5 yearly moving average.

Solution:

t	Demand (d_i)	5 yearly moving totals	5 yearly moving average
1	100		—
2	100		—
3	100	500	100
4	100	420	84
5	100	420	84
6	20	420	84
7	100	420	84
8	100	420	84
9	100	500	100
10	100	500	100
11	100		—
12	100		—

Weighted Moving Average

Here each item is multiplied by its respective weight and the moving average is obtained by dividing the weighted total by the total weights.

$$\text{Weighted moving average} = \frac{\sum YW}{\sum W}$$

Exmaple 8 : From the data given below calculate 3 yearly weighted moving averages, the weights being 1, 2, 2.

Year :	**1998**	**1999**	**2000**	**2001**	**2002**	**2003**	**2004**
Value :	**5**	**7**	**3**	**2**	**4**	**6**	**3**

Solution:

Year	**Value**	**3 yearly weighted moving total**	**3 yearly weighted moving average**
1998	5		—
1999	7	$5 \times 1 + 7 \times 2 + 3 \times 2 = 25$	$\frac{25}{5} = 5.0$
2000	3	$7 \times 1 + 3 \times 2 + 2 \times 2 = 17$	$\frac{17}{5} = 3.4$
2001	2	$3 \times 1 + 2 \times 2 + 4 \times 2 = 15$	$\frac{15}{5} = 3.0$
2002	4	$2 \times 1 + 4 \times 2 + 6 \times 2 = 22$	$\frac{22}{5} = 4.4$
2003	6	$4 \times 1 + 6 \times 2 + 3 \times 2 = 22$	$\frac{22}{5} = 4.4$
2004	3		----

Least Squares Method

The principle of least square is one of the best ways of obtaining the trend values. The method of least squares is a mathematical method which can be applied to both linear and non linear trends. This method places a line through a series of plotted points in such a way that the sum of the squares of deviations of the actual points above and below the trend line is at the minimum. Here a straight line trend is obtained which is known as the line of best fit. From this line, the sum of the deviations of various points on either side is zero. Thus in this method, the sum of the squares of these deviations would be the least as compared to the sums of squares of the deviations obtained by using any other line. That is why the name the method of least squares.

Fitting of Straight Line Trend

Let the equation of the straight line trend be

$$y = a + bt \qquad \ldots(1)$$

where y is the required trend value

t is the unit of time

a and b are constants which indicates y intercept and slope respectively.

For a fitted trend line, the values of a and b must be determined from the given data. The values of a and b are determined by two normal equations.

$$\Sigma y = na + b\,\Sigma t$$

$$\Sigma yt = a\Sigma t + b\Sigma t^2$$

For simplicity of calculations we can take the deviations of the time variable as stated below but if there is a break in the time series, the simplification would not be of much help.

(i) When n is odd we can take

$$t = \frac{\text{time -mid point of time}}{\text{interval in time}}$$

(ii) When n is even we can take

$$t = \frac{\text{2 (time -Average of two middle points of time)}}{\text{interval in time}}$$

Under these substitutions we get $\Sigma t = 0$ and two normal equations give.

$$a = \frac{\sum y}{n}$$

$$b = \frac{\sum yt}{\sum t^2}$$

Substituting the values of a and b in (1) we get the required trend line.

Fitting of Second degree Parabolic Trend

Let the equation of the second degree parabola be

$$y = a + bt + ct^2 \qquad \ldots(2)$$

The three constants a, b and c are determined by solving three normal equations.

$$\Sigma y = na + b\Sigma t + c\Sigma t^2$$

$$\Sigma yt = a\Sigma t + b\Sigma t^2 + c\Sigma t^3$$

$$\Sigma yt^2 = a\Sigma t^2 + b\Sigma t^3 + c\Sigma t^4$$

If deviations are taken from middle point (or points) of time, as in the case of straight line, the three equations reduce to

$$\Sigma y = na + c\Sigma t^2$$
$$\Sigma yt = b\Sigma t^2$$
$$\Sigma yt^2 = a\Sigma t^2 + c\Sigma t^4$$

Solving these equations we get a, b and c and putting these values in equation (2) we get the required trend line.

Merits and Demerits of Least Squares Method

Merits:

(i) This method is mathematical and is objective in nature i.e., there is not question of subjectivity.

(ii) Unlike moving average method, this method provides trend values for all the years.

(iii) This method is suitable for forecasting since there exists a functional relationship between two variables.

(iv) This method gives us the most satisfactory results because the sum of the positive and negative deviations of the actual plotted points from the trend line is zero and the sum of the squares of these deviations is the least.

Demerits:

(i) Out of all the methods of measuring trend this is the most difficult to understand.

(ii) Like moving average method this method is not flexible because a few new observations can not be added without affecting the previous results.

(iii) This method needs due care for the determination of the type of the trend curve to be fitted.

(iv) This method is based on the assumption that the data follow a trend that can be expressed by a mathematical equation.

Example 9 : Using least squares method find the trend values from the following data:

year :	**1990**	**1991**	**1992**	**1993**	**1994**	**1995**	**1996**
Production :	**83**	**60**	**54**	**21**	**22**	**13**	**23**

Solution: Let the equation of the straight line trend be

$$y = a + bt$$

Since n is odd, $t = \dfrac{year - 1993}{1}$ = year – 1993

Year	Production (y)	t = year – 1993	t^2	yt	Trend in Estimated Production $y = 39.43 - 10.93\ t$
1990	83	– 3	9	– 249	72.22
1991	60	– 2	4	– 120	61.29
1992	54	– 1	1	–54	50.36
1993	21	0	0	0	39.43
1994	22	1	1	22	28.50
1995	13	2	4	26	17.57
1996	23	3	9	69	6.64
	$\Sigma y = 276$	$\Sigma t = 0$	$\Sigma t^2 = 28$	$\Sigma yt = -306$	

$$\therefore \qquad a = \frac{\sum y}{n} = \frac{276}{7} = 39.43$$

$$b = \frac{\sum yt}{\sum t^2} = \frac{-306}{28} = -10.93$$

Thus, the equation of the straight line trend becomes

$$y = 39.43 - 10.93\ t$$

Example 10 : Fit a straight line trend to the following data and estimate sales for the year 1988.

Year :	**1980**	**1981**	**1982**	**1983**	**1984**	**1985**	**1986**	**1987**
Sales :	**12**	**13**	**13**	**16**	**19**	**23**	**21**	**23**

Solution: Let the equation of the straight line trend be

$$y = a + bt$$

Since n is even, $t = \dfrac{\text{year} - 1983.5}{\frac{1}{2}} = 2\ (\text{year} - 1983.5)$

Year	Sales (y)	t=2 (year–1983.5)	t^2	yt	Trend in estimated sales $y = 17.5 + 0.01$ t
1980	12	– 7	49	– 84	17.43
1981	13	– 5	25	– 65	17.45
1982	13	– 3	9	– 39	17.47
1983	16	– 1	1	– 16	17.49
1984	19	1	1	19	17.51
1985	23	3	9	69	17.53
1986	21	5	25	105	17.55
1987	23	7	49	161	17.57
	$\Sigma y = 140$	$\Sigma t = 0$	$\Sigma t^2 = 168$	$\Sigma yt = 150$	

$$\therefore \qquad a = \frac{\sum y}{n} = \frac{140}{8} = 17.5$$

$$b = \frac{\sum yt}{\sum t^2} = \frac{150}{168} = .89$$

$\therefore$ The equation of the straight line trend becomes

$$y = 17.5 + .89\ t$$

Now, for the year 1988, the value of t is 9, and hence the estimated value is

$$y = 17.5 + .89 \times 9$$
$$= 25.51$$

Note: When n is odd and time interval is 1, b can be taken as yearly increment and when n is even and time interval is 1, b can be taken as half yearly increment.

Example 11 : State the components of a time series with which the following can be associated.

(i) An era of prosperity.

(ii) Bihu sales in a cloth store.

(iii) Decrease in death rate due to advancement in medical science.

(iv) Fire in a factory.

(v) An increase in employment of labour during harvest time.

(vi) Recession.

Solution:

(i) Cyclical	(ii) Seasonal	(iii) Trend
(iv) Random	(v) Seasonal	(vi) Cyclical

Example 12 : Enumerate the objective of analysis of time-series.

(AHSEC 2004)

Solution: The chief objectives of time series analysis are

(i) to identify the causes of variation of the time series data.

(ii) to understand, interpret and evaluate the changes in time series and to forecast the future aspects.

EXERCISE–10

1. Define a Time Series. What are its components? Explain any one of them. **(AHSEC 2004)**
2. What are the components of time series? Discuss them with examples. **(AHSEC 2002, 2003, 2005)**
3. Enumerate the objective of analysis of time-series. **(AHSEC 2004)**
4. What is time-series? What are its main components? State the various methods of studying trend in a time-series. **(AHSEC 2001)**
5. Which component of time series is mainly applicable in the following cases.
   (i) Fire in a factory.
   (ii) An era of prosperity.
   (iii) The traffic density of a particular road at different hours of the day. **(AHSEC 2001)**
   (iv) Weekly sales of cold drinks.
6. Discuss the necessities of the time series data.
7. What do you mean by trend in a time series? Name the mehtods of determining the trend. Determine trend in the following time series taking 3 yearly moving average.

Year:	1969	1970	1971	1972	1973	1974	1975	1976
Sales: (thousand unit)	5	7	9	12	11	10	8	12

Year:	1977	1978	1979	1980	1981	1982
Sales: (thousand unit)	13	17	19	14	13	12

(AHSEC 1999)

8. Discuss the chief demerits of studying trend by moving average method. **(AHSEC 2001)**
9. Determine 3 yearly moving average from the following data:

Year:	1973	1974	1975	1976	1977	1978	1979
Production: (in , 000 tones)	20	22	23	25	24	22	25

Year:	1980	1981	1982
Production: (in , 000 tones)	25	27	30

(AHSEC 2001)

10. What do you mean by trend in a time series? Estimate the trend values by using the data given below by taking a four yearly moving average.

Year:	1964	1965	1966	1967	1968	1969	1970
Values:	12	25	39	54	70	87	105
Year:	1971	1972	1973	1974	1975	1976	1977
Values:	100	82	65	49	34	20	7

(AHSEC 2003)

11. Define moving averages. What are the merits and demerits of the method of moving averages? **(AHSEC 2002, 2007)**

12. Calculate trend values by method of least squares from the data given below:

Year :	1976	1977	1978	1979	1980
Sales :	70	74	80	86	90

(Rs. in Lakh) **(AHSEC 2002)**

[Ans. Trend line is $y = 80 + 5.2\ t$]

13. From the following data, determine linear trend by least squares method or moving average method.

Year :	1975	1976	1977	1978	1979	1980	1981
Production :	50	36	43	45	39	38	33
Year :	1982	1983	1984	1985			
Production :	42	41	34	40			

(AHSEC 2000)

[Ans. The trend line is $y = 40.09 - 0.69\ t$]

14. Give the idea of seasonal and cyclical components of a time series.

(AHSEC 2002)

15. Determine trend by 3 yearly moving average for the following data:

Year :	1973	1974	1975	1976	1977	1978	1979
Production :	15	21	30	36	42	46	50
Year :	1980	1981	1982				
Production :	56	63	70				

(AHSEC 2004)

16. Applying the method of least squares fit a straight line trend from the following data and find the trend values. Estimate the trend value for the year 2005.

Year :	1997	1998	1999	2000	2001	2002	2003
Values :	30	45	39	41	42	46	46
Year :	2004						
Values :	49						

[Ans. The trend line is $y = 42.25 + 0.95\ t$, 50.8]

17. Apply the method of semi average of measuring trend from the following data:

Year :	1985	1986	1987	1988	1989	1990
Sales :	20	26	21	30	27	34

(' 000 units)

18. Define trend of a time series. Determine trend of the following time series by the method of semiaverages

Year :	1943	1944	1945	1946	1947	1948	1949
Sales (Rs.):	38	41	45	48	52	50	60
Year:	1950	1951	1952	1953			
Sales (Rs.):	64	68	73	79			

(AHSEC 2008)

19. Give example of
i) Trend, ii) Seasonal variation, iii) Cyclical variation and iv) Irregularity.
(AHSEC 2008)

20. Calculate trend values by the method of least squares from the data given below:

Year :	1986	1987	1988	1989	1990
Sales (Rs.) :	80	84	90	96	100

(AHSEC 2007)

21. State the various methods of studying trend in a time series. Estimate the trend values by using the following date by taking a 4 yearly moving average.

Year :	1970	1971	1972	1973	1974	1975	1976
Values :	60.0	46.5	53.0	54.5	48.9	48.2	42.6
Year :	1977	1978	1979				
Values :	51.7	51.1	43.8				

(AHSEC 2006)

22. What is a time series? State its uses.
(AHSEC 2006)

"This page is Intentionally Left Blank"

UNIT II

BASIC MATHEMATICS

"This page is Intentionally Left Blank"

1

LAWS OF INDICES

If a be any number, then $a \times a$ is denoted by a^2, $a \times a \times a \times a \times a$ is denoted by a^5. It can be said as that if a is multiplied '2' times, it can be written as a^2, if a is multipled '5' times it can be written as a^5. In general, $a \times a \times \ldots a \times a$ (for m times) can be written as a^m, m is called index and a is called base.

FUNDAMENTAL LAWS OF INDICES

If a and b be two real numbers and m and n be any two rational numbers then

(i) $\boxed{a^m \times a^n = a^{m+n}}$

(ii) $\boxed{a^m \div a^n = a^{m-n}}$

(iii) $\boxed{(a^m)^n = a^{mn}}$

(iv) $\boxed{(a.b)^n = a^n b^n}$

(v) $\boxed{\left(\dfrac{a}{b}\right)^n = \dfrac{a^n}{b^n}; b \neq 0}$

Proof: (i)

$$\begin{aligned} & a^m \times a^n \\ &= \underbrace{(a \times a \ldots \times a)}_{m \text{ times}} \times \underbrace{(a \times a \ldots \times a)}_{n \text{ times}} \\ &= \underbrace{(a \times a \ldots\ldots\ldots \times a)}_{(m+n) \text{ times}} \\ &= a^{m+n} \end{aligned}$$

(ii) $\dfrac{a^m}{a^n}$ $(m > n)$

$$= \frac{\overbrace{a \times a \ldots \times a}^{m \text{ times}}}{\underbrace{a \times a \ldots \times a}_{n \text{ times}}}$$

$$= \overbrace{a \times a \ldots \times a}^{(m-n) \text{ times}}$$

$$= a^{m-n}$$

(iii) $(a^m)^n$

$$= \left(\overbrace{a\times a\times\ldots\times a}^{m\text{ times}}\right)^n$$

$$= \underbrace{\left(\overbrace{a\times a\times\ldots\times a}^{m\text{ times}}\right)\left(\overbrace{a\times a\times\ldots\times a}^{m\text{ times}}\right)\times\left(\overbrace{a\times a\times\ldots\times a}^{m\text{ times}}\right)\times\ldots\times\left(\overbrace{a\times a\times\ldots\times a}^{m\text{ times}}\right)}_{n\text{ times}}$$

$$= \overbrace{a\times a\times\ldots\times a}^{m.n\text{ times}} = a^{m.n}$$

(iv) $(a.b)^n$

$$= \overbrace{(a.b)\times(a.b)\ldots\times(ab)}^{n\text{ times}}$$

$$= \left(\overbrace{a\times a\ldots\times a}^{n\text{ times}}\right)\times\left(\overbrace{b\times b\ldots\times b}^{n\text{ times}}\right)$$

$$= a^n\times b^n$$

(v) $\left(\frac{a}{b}\right)^n$

$$= \left(\overbrace{\left(\frac{a}{b}\right)\times\left(\frac{a}{b}\right)\ldots\times\left(\frac{a}{b}\right)}^{n\text{ times}}\right) = \frac{\overbrace{a\times a\times\ldots\times a}^{n\text{ times}}}{\underbrace{b\times b\times\ldots b}_{n\text{ times}}}$$

$$= \frac{a^n}{b^n}$$

Important notes

(1) $a^{1/n}$ is called n th root of a $(a > o)$

(2) $a^{m/n} = \left(a^{\frac{1}{n}}\right)^m = (a^m)^{\frac{1}{n}}$, is called nth root of a^m

(3) $a^0 = 1$

(4) $a^{-m} = a^{0-m} = \frac{a^0}{a^m} = \frac{1}{a^m}$

SOLVED EXAMPLES

1.1 If $a = x.y^{p-1}$, $b = x.y^{q-1}$, $c = x.y^{r-1}$ then show that $a^{q-r}\, b^{r-p}\, c^{p-q} = 1$

Solution:

L.H.S. $= a^{q-r}\, b^{r-p}.c^{p-q}$

$$= \left(x.y^{p-1}\right)^{q-r} \left(x.y^{q-1}\right)^{r-p} \left(x.y^{r-1}\right)^{p-q}$$
$$= x^{q-r} y^{(p-1)(q-r)}\ x^{r-p} y^{(q-1)(r-p)}\ x^{p-q} y^{(r-1)(p-q)}$$
$$= x^{q-r+r-p+p-q}\ y^{pq-pr-q+r}\ y^{qr-qp-r+p}\ y^{rp-rq-p+q}$$
$$= x^0 y^{pq-pr-q+r+qr-qp-r+p+rp-rq-p+q}$$
$$= x^0 y^0$$
$$= 1 = \text{R.H.S.}$$

1.2 Prove that, $\left(a^{\frac{1}{x-y}}\right)^{\frac{1}{x-z}} \left(a^{\frac{1}{y-z}}\right)^{\frac{1}{y-x}} \left(a^{\frac{1}{z-x}}\right)^{\frac{1}{z-y}} = 1$

Solution:

$$\text{L.H.S.} = \left(a^{\frac{1}{x-y}}\right)^{\frac{1}{x-z}} \left(a^{\frac{1}{y-z}}\right)^{\frac{1}{y-x}} \left(a^{\frac{1}{z-x}}\right)^{\frac{1}{z-y}}$$
$$= a^{\frac{1}{(x-y)(x-z)}}\ a^{\frac{1}{(y-z)(y-x)}}\ a^{\frac{1}{(z-x)(z-y)}}$$
$$= a^{\frac{1}{(x-y)\{-(z-x)\}}}\ a^{\frac{1}{(y-z)\{-(x-y)\}}}\ a^{\frac{1}{(z-x)\{-(y-z)\}}}$$
$$= a^{-\frac{1}{(x-y)(z-x)}}\ a^{-\frac{1}{(y-z)(x-y)}}\ a^{-\frac{1}{(z-x)(y-z)}}$$
$$= a^{\frac{-(y-z)-(z-x)-(x-y)}{(x-y)(y-z)(z-x)}}$$
$$= a^{\frac{-y+z-z+x-x+y}{(x-y)(y-z)(z-x)}} = a^0$$
$$= 1 = \text{R.H.S.}$$

1.3 Show that, $\left(\frac{x^{b^n}}{x^{c^n}}\right)^{\frac{1}{b^n.c^n}} \left(\frac{x^{c^n}}{x^{a^n}}\right)^{\frac{1}{c^n.a^n}} \left(\frac{x^{a^n}}{x^{b^n}}\right)^{\frac{1}{a^n.b^n}} = 1$

Solution:

$$\text{L.H.S.} = \left(\frac{x^{b^n}}{x^{c^n}}\right)^{\frac{1}{b^n.c^n}} \left(\frac{x^{c^n}}{x^{a^n}}\right)^{\frac{1}{c^n.a^n}} \left(\frac{x^{a^n}}{x^{b^n}}\right)^{\frac{1}{a^n.b^n}}$$
$$= \frac{\left(x^{b^n}\right)^{\frac{1}{b^n.c^n}}}{\left(x^{c^n}\right)^{\frac{1}{b^n.c^n}}}\ \frac{\left(x^{c^n}\right)^{\frac{1}{c^n.a^n}}}{\left(x^{a^n}\right)^{\frac{1}{c^n.a^n}}}\ \frac{\left(x^{a^n}\right)^{\frac{1}{a^n.b^n}}}{\left(x^{b^n}\right)^{\frac{1}{a^n.b^n}}}$$
$$= \frac{x^{\frac{b^n}{b^n.c^n}}}{x^{\frac{c^n}{b^n.c^n}}}\ \frac{x^{\frac{c^n}{c^n.a^n}}}{x^{\frac{a^n}{c^n.a^n}}}\ \frac{x^{\frac{a^n}{a^n.b^n}}}{x^{\frac{b^n}{a^n.b^n}}}$$

$$= \frac{\frac{1}{x^{c^n}}}{\frac{1}{x^{b^n}}} \frac{\frac{1}{x^{a^n}}}{\frac{1}{x^{c^n}}} \frac{\frac{1}{x^{b^n}}}{\frac{1}{x^{a^n}}}$$

$= 1 =$ R.H.S.

1.4 Prove that, $a^{1/2} - \dfrac{a-a^{-2}}{a^{\frac{1}{2}} - a^{-\frac{1}{2}}} + \dfrac{1-a^{-2}}{a^{\frac{1}{2}} + a^{-\frac{1}{2}}} + \dfrac{2}{a^{3/2}} = 0$

Solution:

$$\text{L.H.S.} = a^{1/2} - \frac{a-a^{-2}}{a^{\frac{1}{2}} - a^{-\frac{1}{2}}} + \frac{1-a^{-2}}{a^{\frac{1}{2}} + a^{-\frac{1}{2}}} + \frac{2}{a^{3/2}}$$

$$= \frac{a^{\frac{1}{2}}.a^{\frac{3}{2}} + 2}{a^{\frac{3}{2}}} - \frac{a - \frac{1}{a^2}}{a^{\frac{1}{2}} - \frac{1}{a^{\frac{1}{2}}}} + \frac{1 - \frac{1}{a^2}}{a^{\frac{1}{2}} + \frac{1}{a^{\frac{1}{2}}}}$$

$$= \frac{a^2+2}{a^{\frac{3}{2}}} - \frac{\frac{a.a^2-1}{a^2}}{\frac{a^{\frac{1}{2}}.a^{\frac{1}{2}}-1}{a^{\frac{1}{2}}}} + \frac{\frac{a^2-1}{a^2}}{\frac{a^{\frac{1}{2}}.a^{\frac{1}{2}}+1}{a^{\frac{1}{2}}}}$$

$$= \frac{a^2+2}{a^{\frac{3}{2}}} - \frac{\left(a^3-1\right)/a^2}{\left(a-1\right)/a^{\frac{1}{2}}} + \frac{\left(a^2-1\right)/a^2}{\left(a+1\right)/a^{\frac{1}{2}}}$$

$$= \frac{a^2+2}{a^{\frac{3}{2}}} - \frac{\left(a^3-1\right)}{\left(a-1\right)}.\frac{1}{a^{2-\frac{1}{2}}} + \frac{\left(a-1\right)\left(a+1\right)}{\left(a+1\right)}.\frac{1}{a^{2-\frac{1}{2}}}$$

$$= \frac{a^2+2}{a^{\frac{3}{2}}} - \frac{\left(a-1\right)\left(a^2+a+1\right)}{\left(a-1\right)a^{\frac{3}{2}}} + \frac{\left(a-1\right)\left(a+1\right)}{\left(a+1\right)a^{\frac{3}{2}}}$$

$$= \frac{a^2+2}{a^{\frac{3}{2}}} - \frac{a^2+a+1}{a^{\frac{3}{2}}} + \frac{a-1}{a^{\frac{3}{2}}}$$

$$= \frac{a^2+2-a^2-a-1+a-1}{a^{\frac{3}{2}}}$$

$= 0$

$=$ R.H.S., proved

1.5 Prove that,

$$y = \frac{2xz}{x+z}$$

if $a^x = b^y = c^z$ and $b^2 = ac$ **[AHSEC, 02]**

Solution:

$$a^x = b^y = c^z$$

$\therefore \quad a^x = b^y$

$\Rightarrow \quad (a^x)^{\frac{1}{x}} = (b^y)^{\frac{1}{x}}$

$\Rightarrow \quad a = b^{y/x}$ (1)

and $\quad b^y = c^z \Rightarrow (b^y)^{\frac{1}{z}} = (c^z)^{\frac{1}{z}}$

$\Rightarrow \quad b^{\frac{y}{z}} = c$ (2)

again $\quad b^2 = a.c$

$\Rightarrow \quad b^2 = b^{\frac{y}{x}}.b^{\frac{y}{z}}$ by (1) & (2)

$\Rightarrow \quad b^2 = b^{\frac{y}{x}+\frac{y}{z}}$

$\Rightarrow \quad 2 = \frac{y}{x} + \frac{y}{z}$

$\Rightarrow \quad 2 = y\left(\frac{1}{x} + \frac{1}{z}\right)$

$\Rightarrow \quad 2 = y\left(\frac{x+z}{xz}\right)$

$\Rightarrow \quad y = \frac{2xz}{x+z}$, proved.

1.6 If $m = a^x$, $n = a^y$ and $a^2 = (m^y.n^x)^z$, *prove that* $x\,y\,z = 1$. [AHSEC, 00, 06]

Solution:

$\therefore \quad a^2 = (m^y.n^x)^z,\ m = a^x,\ n = a^y$

$\Rightarrow \quad a^2 = \left\{(a^x)^y.(a^y)^x\right\}^z$

$\Rightarrow \quad a^2 = (a^{xy}.a^{yx})^z$

$\Rightarrow \quad a^2 = (a^{xy+xy})^z$

$\Rightarrow \quad a^2 = (a^{2xy})^z$

$\Rightarrow \quad a^2 = a^{2xyz}$

$\Rightarrow \quad 2 = 2xyz$

$\Rightarrow \quad 1 = xyz$, proved

1.7 If $\sqrt[3]{x}+\sqrt[3]{y}+\sqrt[3]{z}=0$

prove that $(x+y+z)^3 = 27\,xyz$ **[AHSEC, 98]**

Solution:

$$\sqrt[3]{x}+\sqrt[3]{y}+\sqrt[3]{z}=0$$

$$\Rightarrow \quad \sqrt[3]{x}+\sqrt[3]{y} = -\sqrt[3]{z} \qquad (1)$$

$$\Rightarrow \left(\sqrt[3]{x}+\sqrt[3]{y}\right)^3 = \left(-\sqrt[3]{z}\right)^3$$

$$\Rightarrow \left(x^{\frac{1}{3}}\right)^3 + \left(y^{\frac{1}{3}}\right)^3 + 3x^{\frac{1}{3}}.y^{\frac{1}{3}}\left(x^{\frac{1}{3}}+y^{\frac{1}{3}}\right) = \left(-1.z^{\frac{1}{3}}\right)^3$$

$$\Rightarrow \quad x+y+3x^{\frac{1}{3}}y^{\frac{1}{3}}\left(-z^{\frac{1}{3}}\right) = -z \qquad \text{from (1)}$$

$$\Rightarrow \quad x+y+z = 3x^{\frac{1}{3}}.y^{\frac{1}{3}}.z^{\frac{1}{3}}$$

$$\Rightarrow (x+y+z)^3 = \left(3x^{\frac{1}{3}}.y^{\frac{1}{3}}.z^{\frac{1}{3}}\right)^3$$

$$\Rightarrow (x+y+z)^3 = 3^3.\left(x^{\frac{1}{3}}\right)^3\left(y^{\frac{1}{3}}\right)^3\left(z^{\frac{1}{3}}\right)^3$$

$$\Rightarrow (x+y+z)^3 = 27.x.y.z.$$

1.8 If $x^{\frac{1}{a}} = y^{\frac{1}{b}} = z^{\frac{1}{c}}$ and $xyz = 1$, show that $a+b+c=0$ **(AHSEC, 94)**

Solution:

$$\Rightarrow \quad \text{Let } x^{\frac{1}{a}} = y^{\frac{1}{b}} = z^{\frac{1}{c}} = k$$

$$\therefore \quad x^{\frac{1}{a}} = k$$

$$\Rightarrow \quad \left(x^{\frac{1}{a}}\right)^a = k^a$$

$$\Rightarrow \quad x = k^a \qquad ...(1)$$

$$\therefore \quad y^{\frac{1}{b}} = k$$

$$\Rightarrow \quad \left(y^{\frac{1}{b}}\right)^b = k^b$$

$$\Rightarrow \quad y = k^b \qquad ...(2)$$

$$z^{\frac{1}{c}} = k$$

$\Rightarrow \quad \left(z^{\frac{1}{c}}\right)^c = k^c$

$\Rightarrow \quad z = k^c$ ...(3)

Now as, $x.\ y.\ z=1$

$\Rightarrow \quad k^a.k^b.k^c = 1$ from (1),(2), (3)

$\Rightarrow \quad k^{a+b+c} = k^0$

$\Rightarrow \quad a + b + c = 0$, proved

1.9 Solve $2^{2x} - 5.2^x + 4 = 0$, **[AHSEC, 95]**

Solution:

$\Rightarrow \quad 2^{2x} - 5.2^x + 4 = 0$

$\Rightarrow \quad (2^x)^2 - 5.(2^x) + 4 = 0$

$\Rightarrow \quad a^2 - 5\ a + 4 = 0$, put $2^x = a$

$\Rightarrow \quad a^2 - 4a - a + 4 = 0$

$\Rightarrow \quad a(a-4) - (a-4) = 0$

$\Rightarrow \quad (a-1)\ (a-4) = 0$

$\Rightarrow \quad a = 1, 4$

$\Rightarrow \quad 2^x = 1;\ 2^x = 4$

$\Rightarrow \quad 2^x = 2^0;\ 2^x = 2^2$

$\Rightarrow \quad x = 0;\ x = 2$

1.10 Simplify

$$\frac{(b.c)^{b-c}\,(c.a)^{c-a}\,(a.b)^{a-b}}{\left(a^{b-c}.b^{c-a}c^{a-b}\right)^{-1}}$$ **(AHSEC; 01)**

Solution:

$$\frac{(bc)^{b-c}\,(c.a)^{c-a}\,(ab)^{a-b}}{\left(a^{b-c}\,b^{c-a}.c^{a-b}\right)^{-1}}$$

$$= \frac{b^{b-c}\,.c^{b-c}\,.c^{c-a}\,.a^{c-a}\,.a^{a-b}\,.b^{a-b}}{a^{(b-c)(-1)}\,.b^{(c-a)(-1)}\,.c^{(a-b)(-1)}}$$

$$= \frac{b^{b-c+a-b}\,.c^{b-c+c-a}\,.a^{c-a+a-b}}{a^{c-b}b^{a-c}\,.c^{b-a}}$$

$$= \frac{b^{a-c}\,.c^{b-a}\,.a^{c-b}}{b^{a-c}\,.c^{b-a}\,.a^{c-b}} = 1$$

1.11Simplify $\left(\sqrt{x} - \frac{\sqrt{xy}+y}{\sqrt{x}+\sqrt{y}}\right)\left(\frac{\sqrt{x}}{\sqrt{x}+\sqrt{y}} + \frac{\sqrt{y}}{\sqrt{x}-\sqrt{y}} + \frac{2\sqrt{xy}}{x-y}\right)$

Solution:

$$\left(\sqrt{x} - \frac{\sqrt{xy}+y}{\sqrt{x}+\sqrt{y}}\right)\left(\frac{\sqrt{x}}{\sqrt{x}+\sqrt{y}} + \frac{\sqrt{y}}{\sqrt{x}-\sqrt{y}} + \frac{2\sqrt{xy}}{x-y}\right)$$

$$=\left(\frac{\sqrt{x}\left(\sqrt{x}+\sqrt{y}\right)-\sqrt{xy}-y}{\sqrt{x}+\sqrt{y}}\right)\left(\frac{\sqrt{x}\left(\sqrt{x}-\sqrt{y}\right)+\sqrt{y}\left(\sqrt{x}+\sqrt{y}\right)}{\left(\sqrt{x}+\sqrt{y}\right)\left(\sqrt{x}-\sqrt{y}\right)}+\frac{2\sqrt{xy}}{x-y}\right)$$

$$=\left(\frac{x+\sqrt{xy}-\sqrt{xy}-y}{\sqrt{x}+\sqrt{y}}\right)\left(\frac{x-\sqrt{xy}+\sqrt{yx}+y}{\left(\sqrt{x}\right)^2-\left(\sqrt{y}\right)^2}+\frac{2\sqrt{xy}}{x-y}\right)$$

$$=\left(\frac{x-y}{\sqrt{x}+\sqrt{y}}\right)\left(\frac{x+2\sqrt{xy}+y}{x-y}\right)$$

$$=\left(\frac{x-y}{\sqrt{x}+\sqrt{y}}\right)\left(\frac{\left(\sqrt{x}\right)^2+2\sqrt{x}\sqrt{y}+\left(\sqrt{y}\right)^2}{x-y}\right)$$

$$=\frac{\left(\sqrt{x}+\sqrt{y}\right)^2}{\left(\sqrt{x}+\sqrt{y}\right)}=\left(\sqrt{x}+\sqrt{y}\right)$$

1.12 Simplify $\dfrac{\sqrt{4+2\sqrt{3}}}{\sqrt{4-2\sqrt{3}}}$

Solution:

$$\frac{\sqrt{4+2\sqrt{3}}}{\sqrt{4-2\sqrt{3}}}=\frac{\left(\sqrt{4+2\sqrt{3}}\right)\left(\sqrt{4+2\sqrt{3}}\right)}{\left(\sqrt{4-2\sqrt{3}}\right)\left(\sqrt{4+2\sqrt{3}}\right)}$$

$$=\frac{\left(\sqrt{4+2\sqrt{3}}\right)^2}{\left(4-2\sqrt{3}\right)^{\frac{1}{2}}\left(4+2\sqrt{3}\right)^{\frac{1}{2}}}$$

$$=\frac{4+2\sqrt{3}}{\left\{\left(4-2\sqrt{3}\right)\left(4+2\sqrt{3}\right)\right\}^{\frac{1}{2}}}$$

$$=\frac{4+2\sqrt{3}}{\left\{16-\left(2\sqrt{3}\right)^2\right\}^{\frac{1}{2}}}=\frac{4+2\sqrt{3}}{(16-12)^{\frac{1}{2}}}$$

$$=\frac{4+2\sqrt{3}}{2}=2+\sqrt{3}$$

1.13 Find the value of x and y if $x^y=y^2$; $y^{2y}=x^4$

Solution:

$$y^{2y}=x^4$$

$$\Rightarrow\qquad \left(y^2\right)^y=x^4$$

$\Rightarrow \quad (x^y)^y = x^4 \left(\because \; x^y = y^2\right)$

$\Rightarrow \quad x^{y^2} = x^4$

$\Rightarrow \quad y^2 = 4$

$\Rightarrow \quad y = \pm\sqrt{4}$

$\Rightarrow \quad y = \pm 2$

Now $\quad y = 2$

$\Rightarrow \quad 2^{2.2} = x^4$

$\Rightarrow \quad 2^4 = x^4$

$\Rightarrow \quad x = \pm 2$

$\quad \& \; y = -2$

$\Rightarrow \quad (-2)^{-4} = x^4$

$\Rightarrow \quad \{(-1).2\}^{-4} = x^4$

$\Rightarrow \quad (-1)^{-4}.2^{-4} = x^4$

$\Rightarrow \quad \dfrac{1}{2^4} = x^4$

$\Rightarrow \quad x^4 = \left(\dfrac{1}{2}\right)^4$

$\Rightarrow \quad x = \pm\dfrac{1}{2}$

So the solutions are–

$$\left(\frac{1}{2},-2\right),\left(-\frac{1}{2},-2\right),(2,2)(-2,2)$$

1.14 If. $a^x + b^y = a + b, \; a^{x-1} + b^{y-1} = 2$

find the values of x and y

Solution:

$a^{x-1} + b^{y-1} = 2$

$\Rightarrow (a + b)\,(a^{x-1} + b^{y-1}) = (a + b).2$

$\Rightarrow a^x + ab^{y-1} + ba^{x-1} + b^y \qquad = 2(a^x + b^y) \; (\because \; a + b = a^x + b^y)$

$\Rightarrow \quad ab^{y-1} + ba^{x-1} = a^x + b^y$

$\Rightarrow \quad ab^{y-1} - b^y = a^x - ba^{x-1}$

$\Rightarrow \quad b^{y-1}(a - b) = a^{x-1}(a - b)$

$\Rightarrow \quad b^{y-1} = a^{x-1} \qquad \ldots(1)$

$\therefore \quad a^{x-1} + b^{x-1} = 2$

$\Rightarrow \quad 2a^{x-1} = 2 \qquad$ by (1)

$\Rightarrow \qquad a^{x-1} = 1$

$\Rightarrow \qquad a^{x-1} = a^0$

$\Rightarrow \qquad x-1 = 0$

$\Rightarrow \qquad x = 1$ ...(2)

$b^{y-1} = a^{1-1}$ by (1), (2)

$\Rightarrow \qquad b^{y-1} = a^0$

$\Rightarrow \qquad b^{y-1} = 1$

$\Rightarrow \qquad b^{y-1} = b^0$

$\Rightarrow \qquad y - 1 = 0$

$\Rightarrow \qquad y = 1$

1.15 Simplify $\sqrt{\frac{a}{b}\sqrt{\frac{a}{b}\sqrt[3]{\frac{a}{b}}}}.a^{-\frac{1}{3}}.\frac{1}{b^3}$

Solution: $\sqrt{\frac{a}{b}\sqrt{\frac{a}{b}\sqrt[3]{\frac{a}{b}}}}.a^{-\frac{1}{3}}.\frac{1}{b^3}$

$$= \sqrt{\frac{a}{b}\sqrt{\frac{a}{b}\left(\frac{a}{b}\right)^{\frac{1}{3}}}}.a^{-\frac{1}{3}}.\frac{1}{b^3}$$

$$= \sqrt{\frac{a}{b}\sqrt{\left(\frac{a}{b}\right)^{1+\frac{1}{3}}}}.a^{-\frac{1}{3}}.\frac{1}{b^3}$$

$$= \sqrt{\frac{a}{b}\sqrt{\left(\frac{a}{b}\right)^{\frac{4}{3}}}}\,a^{-\frac{1}{3}}.\frac{1}{b^3}$$

$$= \sqrt{\frac{a}{b}\left\{\left(\frac{a}{b}\right)^{\frac{4}{3}}\right\}^{\frac{1}{2}}}.a^{-\frac{1}{3}}.\frac{1}{b^3} = \sqrt{\left(\frac{a}{b}\right)^{1+\frac{4}{6}}}.a^{-\frac{1}{3}}.\frac{1}{b^3}$$

$$= \left(\frac{a}{b}\right)^{\frac{10}{6}\cdot\frac{1}{2}}.a^{-\frac{1}{3}}.\frac{1}{b^3} = \frac{a^{\frac{10}{12}}}{b^{\frac{10}{12}}}.\frac{a^{-\frac{1}{3}}}{b^3}$$

$$= \frac{a^{\frac{10}{12}-\frac{1}{3}}}{b^{\frac{10}{12}+3}} = \frac{a^{\frac{6}{12}}}{b^{\frac{46}{12}}} = \frac{a^{\frac{3}{6}}}{b^{\frac{23}{6}}} = a^{\frac{3}{6}}.b^{-\frac{23}{6}} = \left(a^3.b^{-23}\right)^{\frac{1}{6}}$$

$$= \frac{a^{\frac{1}{2}}}{b^{\frac{23}{6}}}$$

1.16 Solve the equations

(i) $5^{x+2} + 6 = 631$

(ii) $5^{x-1}\, 3^{x-2} = 5$

(iii) $4^x - 3.2^{x+2} + 2^5 = 0$

Solution:

(i) $5^{x+2} + 6 = 631$

$\Rightarrow \quad 5^x \,.\, 5^2 = 631 - 6$

$\Rightarrow \quad 5^x \,.\, 25 = 625$

$\Rightarrow \quad 5^x = 25$

$\Rightarrow \quad 5^x = 5^2$

$\Rightarrow \quad x = 2$

(ii) $5^{x-1}\,.3^{x-2} = 5$

$\Rightarrow \quad 5^x \,.\, 5^{-1}\, 3^x \,.\, 3^{-2} = 5$

$\Rightarrow \quad 5^x \,.\, 3^x = 5^2 \,.\, 3^2$

$\Rightarrow \quad (5.3)^x = (5.3)^2$

$\Rightarrow \quad x = 2$

(iii) $4^x - 3.2^{x+2} + 2^5 = 0$

$\Rightarrow \quad (2^2)^x - 3.2^x \,.\, 2^2 + 32 = 0$

$\Rightarrow (2^x)^2 - 3.2^x \,.\, 4 + 32 = 0$

$\Rightarrow (2^x)^2 - 12.2^x + 32 = 0$

$\Rightarrow a^2 - 12a + 32 = 0$ (Let $2^x = a$)

$\Rightarrow a^2 - 8a - 4a + 32 = 0$

$\Rightarrow a(a - 8) - 4(a - 8) = 0$

$\Rightarrow (a - 4)\,(a - 8) = 0$

Now $a = 4$

$\Rightarrow \quad 2^x = 4$

$\Rightarrow \quad 2^x = 2^2$

$\Rightarrow \quad x = 2$

and $\quad a = 8$

$\Rightarrow \quad 2^x = 2^3$

$\Rightarrow \quad x = 3$

EXERCISE–1

1. (a) If $a = \dfrac{x}{y^p}$, $b = \dfrac{x}{y^q}$, $e = \dfrac{x}{y^r}$ then show that

$a^{q-r} . b^{r-p} . e^{p-q} = 1$

(b) Show that $\left(\dfrac{x^b}{x^e}\right)^a \cdot \left(\dfrac{x^e}{x^a}\right)^b \cdot \left(\dfrac{x^a}{x^b}\right)^e = 1$

2. If $x = \dfrac{\sqrt{3}+1}{\sqrt{3}-1}$, $y = \dfrac{\sqrt{3}-1}{\sqrt{3}+1}$, find the value of $x^2 + xy + y^2$ **(Ans. 15)**

3. If $x = \dfrac{\sqrt{a+2b}+\sqrt{a-2b}}{\sqrt{a+2b}-\sqrt{a-2b}}$, find the value of $b.x^2 - a.x + b$ **(Ans. 0)**

4. If $\sqrt[3]{a+b} + \sqrt[3]{b+c} + \sqrt[3]{c+a} = 0$, show that

$(a+b+c)^3 = \dfrac{27}{8}(a+b)(b+c)(c+a)$.

5. If $x = \left[a+\sqrt{a^2+b^3}\right]^{\frac{1}{3}} + \left[a-\sqrt{a^2+b^3}\right]^{\frac{1}{3}}$ show that $x^3 + 3b^3x - 2a = 0$.

6. If $x = 2 + 2^{\frac{2}{3}} + 2^{\frac{1}{3}}$, show that $x^3 - 6x^2 + 6x - 2 = 0$.

7. If $x = 5 + 2\sqrt{6}$, find the value of $\sqrt{x} - \dfrac{1}{\sqrt{x}}$. **(Ans. 8)**

8. Find the value of $\dfrac{3+\sqrt{6}}{5\sqrt{3}-2\sqrt{12}-\sqrt{32}+\sqrt{50}}$. **(Ans. $\sqrt{3}$)**

9. If $x^y = y^x$ and $y = 2x$, find the values of x and y. **(Ans. 2, 4)**

10. If $a^x = b^y = c^z$ and $b^2 = ac$ prove that $\dfrac{1}{x} + \dfrac{1}{z} = \dfrac{z}{y}$

[AHSEC 2005, 2007]

2

ARITHMETICAL AND GEOMETRICAL PROGRESSIONS

ARITHMETICAL PROGRESSION

When the terms of a sequence increase or decrease with a common difference then that sequence is called Arithmetical Progression (A.P.) e.g.

1, 3, 5, 7, 9, 11....

3, 6, 9, 12, 15....

100, 97, 94, 91....

are some Arithmetical Progressions where common differences are respectively 2, 3 and (–3)

The n th term of an Arithmetical Progression, having the first term 'a' and common diference 'd' is

$$\boxed{t_n = a + (n-1).d}$$

The sum of a finite number of terms of an Arithmetical Progression:

The sum of first 'n' terms of an Arithmetical Progression having the first term 'a' and common difference 'd' is

$$\boxed{s_n = \frac{n}{2}\{2.a + (n-1).d\}}$$

Proof: $s_n = a + (a + d) + (a + 2d) + ... + \left(a + \overline{n-1}d\right)$ (1)

$\Rightarrow$ $s_n = \left(a + \overline{n-1}d\right) + \left(a - \overline{n-2}d\right) + \left(a + \overline{n-3}d\right) + ... + a$ (2)

(writting (1) in reverse order)

$\therefore$ (1) + (2) $\Rightarrow$

$$2s_n = \overbrace{\left(2a + \overline{n-1}d\right) + \left(2a + \overline{n-1}d\right) + ... + \left(2a + \overline{n-1}d\right)}^{\text{n times}}$$

$\Rightarrow$ $2s_n = n\left(2a + \overline{n-1}d\right)$

$\Rightarrow$ $s_n = \frac{n}{2}\left(2a + \overline{n-1}d\right)$

SOLVED EXAMPLES

2.1.1 Find the 20th term of the A.P. 3, 5, 7,... **(AHSEC, 07)**

Solution:

Here Ist term, $a = 3$

common difference, $d = 2$

$\therefore \quad t_{20} = 3 + (20 - 1).2$

$= 3 + 19.2$

$= 41$

2.1.2 Find the sum of first 10 terms of the A.P. 2, 6, 10, ...

Solution:

Here Ist term, $a = 2$

common difference, $d = 4$

$$\therefore \quad s_{10} = \frac{10}{2}\{2.2+(10-1).4\}$$

$= 5\ (4 + 36) = 200$

2.1.3 If a^2, b^2 and c^2 be the first three terms of an A.P. then prove that $\frac{1}{b+c}, \frac{1}{c+a}, \frac{1}{a+b}$ will also be in A.P.

Solution:

Since a^2, b^2, and c^2 are in A.P.

$\therefore \quad a^2 - b^2 = b^2 - c^2$

$\Rightarrow (a + b)\ (a - b) = (b + c)\ (b - c)$ (1)

Now $\frac{1}{b+c} - \frac{1}{c+a}$

$$= \frac{(c+a)-(b+c)}{(b+c)(c+a)}$$

$$= \frac{(a-b)}{(b+c)(c+a)}$$

$$= \frac{(b+c)(b-c)}{(a+b)(b+c)(c+a)} \quad \text{from (1)}$$

$$= \frac{(a+b)-(c-a)}{(a+b)(c+a)}$$

$$= \frac{1}{c+a} - \frac{1}{a+b}$$

$\therefore \frac{1}{b+c} - \frac{1}{c+a} = \frac{1}{c+a} - \frac{1}{a+b}$, which shows that $\frac{1}{b+c}, \frac{1}{c+a}, \frac{1}{a+b}$ are in A.P.

2.1.4 If the positive terms a, b, c are in A.P. then prove that $\frac{1}{\sqrt{b}+\sqrt{c}}, \frac{1}{\sqrt{a}+\sqrt{c}}$ and $\frac{1}{\sqrt{a}+\sqrt{b}}$ will also be in A.P.

Solution: Since *a, b, c* are in A.P.

$$\therefore \quad a - b = b - c$$

$$\Rightarrow \quad \left(\sqrt{a}\right)^2 - \left(\sqrt{b}\right)^2 = \left(\sqrt{b}\right)^2 - \left(\sqrt{c}\right)^2$$

$$\Rightarrow \quad \left(\sqrt{a}+\sqrt{b}\right)\left(\sqrt{a}-\sqrt{b}\right) = \left(\sqrt{b}+\sqrt{c}\right)\left(\sqrt{b}-\sqrt{c}\right) \qquad \text{...(1)}$$

Now $\frac{1}{\sqrt{b}+\sqrt{c}} - \frac{1}{\sqrt{c}+\sqrt{a}}$

$$= \frac{\left(\sqrt{c}+\sqrt{a}\right) - \left(\sqrt{b}+\sqrt{c}\right)}{\left(\sqrt{b}+\sqrt{c}\right)\left(\sqrt{c}+\sqrt{a}\right)}$$

$$= \frac{\sqrt{a}-\sqrt{b}}{\left(\sqrt{b}+\sqrt{c}\right)\left(\sqrt{c}+\sqrt{a}\right)}$$

$$= \frac{\left(\sqrt{b}+\sqrt{c}\right)\left(\sqrt{b}-\sqrt{c}\right)}{\left(\sqrt{a}+\sqrt{b}\right)\left(\sqrt{b}+\sqrt{c}\right)\left(\sqrt{c}+\sqrt{a}\right)} \qquad \text{from (1)}$$

$$= \frac{\left(\sqrt{a}+\sqrt{b}\right) - \left(\sqrt{c}-\sqrt{a}\right)}{\left(\sqrt{a}+\sqrt{b}\right)\left(\sqrt{c}+\sqrt{a}\right)}$$

$$= \frac{1}{\sqrt{c}+\sqrt{a}} - \frac{1}{\sqrt{a}+\sqrt{b}} \qquad \text{...(2)}$$

$$\therefore \quad \frac{1}{\sqrt{b}+\sqrt{c}} - \frac{1}{\sqrt{c}+\sqrt{a}} = \frac{1}{\sqrt{c}+\sqrt{a}} - \frac{1}{\sqrt{a}+\sqrt{b}} \qquad \text{[from (1) and (2)]}$$

$\therefore \frac{1}{\sqrt{b}+\sqrt{c}}, \frac{1}{\sqrt{c}+\sqrt{a}}, \frac{1}{\sqrt{a}+\sqrt{b}}$ are also in A.P.

2.1.5 If nth term of a sequence be $a_n = n^2 - 1$, prove that the sequence is not an A.P.

Solution:

$$n\text{th term, } a_n = n^2 - 1$$

$$\therefore \quad (n\text{-}1)\text{ th term, } a_{n-1} = (n-1)^2 - 1$$

$$\therefore \quad (n+1)\text{ th term, } a_{n+1} = (n+1)^2 - 1$$

Now $$a_{n+1} - a_n = \{(n+1)^2 - 1\} - \{n^2 - 1\}$$

$$= 2n + 1 \qquad \text{...(1)}$$

and $$a_n - a_{n-1} = \{n^2 - 1\} - \{(n-1)^2 - 1\}$$

$$= 2n - 1 \qquad \text{...(2)}$$

$$\therefore \quad a_{n+1} - a_n \neq a_n - a_{n-1} \text{ from (1), (2)}$$

so, a_{n+1}, a_n, a_{n-1} are not in A.P. and hence the sequence is not an A.P.

2.1.6 If the sum of first six terms of an A.P. is five times, the sum of next six terms and if first term be 100, then find the common difference.

Solution:

The sum of first six terms,

$$s_1 = \frac{6}{2}\{2.100 + (6-1)d\}$$

(Let, common difference = d)

$$= 3(200 + 5d) \qquad (1)$$

The first term of the next six terms would evidently be the seventh term; which is

$$t_7 = 100 + (7-1)d$$

$$= 100 + 6d$$

Now sum of these terms would be

$$s_2 = \frac{6}{2}\{2(100 + 6d) + (6-1)d\}$$

$$= 3\ (200 + 17d) \qquad \text{...(2)}$$

Now according to given condition,

$$3(200 + 5d) = 5.3(200 + 17d)$$

$$\Rightarrow \quad 600 + 15d = 3000 + 255d$$

$$\Rightarrow \quad -2400 = 240d$$

$$\Rightarrow \quad d = -10$$

2.1.7 If the nth term of an A.P. be $2n+3$ then find the sum of the A.P. for the first 'n' terms.

Solution:

n th term, $t_n = 2n + 3$

$\therefore$ 1st term, $t_1 = 2.1 + 3 = 5$

$\therefore$ 2nd term, $t_2 = 2.2 + 3 = 7$

$\therefore$ common difference, $d = 2$

$$\therefore \quad s_n = \frac{n}{2}\{2.5+(n-1).2\}$$

$$= \frac{n}{2}\{10+2n-2\}$$

$$= \frac{n}{2}(8+2n)$$

$$= n(n+4)$$

2.1.8 If s_1, s_2 and s_3 be the sums of n terms of three A.P., the first term of each series being 1 and the respective common differences being 1, 2 and 3, prove that

$$s_1 + s_3 = 2s_2$$ **(AHSEC, 02)**

Solution:

$$s_1 = \frac{n}{2}\{2.1+(n-1)\}$$

$$= \frac{n}{2}(2+n-1)$$

$$= \frac{n(n+1)}{2} \qquad \text{...(1)}$$

$$s_2 = \frac{n}{2}\{2.1+(n-1).2\}$$

$$= \frac{n}{2}(2+2n-2)$$

$$= n^2 \qquad \text{...(2)}$$

$$s_3 = \frac{n}{2}\{2.1+(n-1).3\}$$

$$= \frac{n}{2}(2+3n-3)$$

$$= \frac{n(3n-1)}{2} \qquad \text{...(3)}$$

$$\therefore \qquad s_1 + s_3 = \frac{n(n+1)}{2} + \frac{n(3n-1)}{2} \text{ from (1) and (3)}$$

$$= \frac{n}{2}(4n)$$

$$= 2.n^2$$

$$= 2s_2 \text{ by (2)}$$

$$\therefore \qquad s_1 + s_3 = 2s_2$$

2.1.9 If s_1, s_2, s_3 be respectively the sum of the first n, $2n$, $3n$ terms of an A.P. show that $s_3 = 3(s_2 - s_1)$ (AHSEC, 03)

Solution:

$$s_1 = \frac{n}{2}\{2a + (n-1)d\}$$

$$s_2 = \frac{2n}{2}\{2a + (2n-1)d\}$$

$$s_3 = \frac{3n}{2}\{2a + (3n-1)d\}$$

Now $\qquad 3(s_2 - s_1)$

$$= 3.\left[\frac{2n}{2}\{2a + (2n-1)d\} - \frac{n}{2}\{2a + (n-1)d\}\right]$$

$$= \frac{3.n}{2}[4a + 4nd - 2d - 2a - nd + d]$$

$$= \frac{3n}{2}[2a + (3n-1)d]$$

$$= s_3$$

2.1.10 If 2 and 26 are the first and 9th terms of an A.P. then find the middle five terms of the A.P.

Solution:

The middle terms between 1st and 9th term are

t_3, t_4, t_5, t_6 and t_7

Now $\qquad t_1 = 2$

$\Rightarrow \qquad a = 2$

$\qquad t_9 = 26$

$\Rightarrow \qquad a + (9-1)d = 26$

$\Rightarrow \qquad 2 + 8d = 26$

$\Rightarrow \qquad d = 3$

$\therefore \qquad t_3 = a + (3-1)d$

$\qquad = 2 + 2.3 = 8$

$\therefore \quad t_4 = t_3 + d = 8 + 3 = 11$
$t_5 = t_4 + 3 = 11 + 3 = 14$
$t_6 = t_5 + 3 = 14 + 3 = 17$
$t_7 = t_6 + 3 = 17 + 3 = 20$

2.1.11 The sum of the three numbers in A.P. is 27 and the sum of their squares is 293, find them. (AHSEC, 96)

Solution: Let the terms be a, $a + d$, $a + 2d$

$\therefore \quad a + a + d + a + 2d = 27$

$\Rightarrow \quad 3a + 3d = 27$

$\Rightarrow \quad a + d = 9 \qquad ...(1)$

$\therefore \quad a^2 + (a + d)^2 + (a + 2d)^2 = 293$

$\Rightarrow \quad 3a^2 + 6ad + 5d^2 = 293$

$\Rightarrow \quad 3(9 - d)^2 + 6(9-d).d + 5d^2 = 293 \qquad$ [from (1)]

$\Rightarrow \quad 243 - 54d + 3d^2 + 54d - 6d^2 + 5d^2 = 293$

$\Rightarrow \quad 2d^2 = 50 \Rightarrow d = \pm 5$

Now if $d = 5$, from (1)

$a + d = 9 \Rightarrow a = 4$

$\therefore$ terms are 4, 9, 14

& if $\quad d = -5, a + d = 9 \Rightarrow a = 14$

$\therefore$ terms are 14, 9, 4

2.1.12 Find the sum $\sum_{r=1}^{n} r$

Solution:

$$\sum_{r=1}^{n} r = 1 + 2 + 3 + ... + n$$

$$= \frac{n}{2}\{2.1 + (n-1).1\}$$

$$= \frac{n(n+1)}{2}$$

2.1.13 Find the sum $\sum_{r=1}^{n} r^2$

Solution:

$$\sum_{r=1}^{n} r^2 = 1^2 + n^2 + ... + n^2$$

But $\quad r^3-(r-1)^3 = 3r^2 - 3r + 1$

Putting $\quad r = 1, 2, \ldots n$

$$1^3 - 0^3 = 3.1^2 - 3.1 + 1$$
$$2^3 - 1^3 = 3.2^2 - 3.2 + 1$$

$$n^3-(n-1)^3 = 3.n^3 - 3.n + 1$$

Adding all, $\quad n^3 = 3(1^2+2^2+\ldots+n^2) - 3(1+2\ldots+n) + n$

$$\Rightarrow \quad n^3 + \frac{3n(n+1)}{2} - n = 3\sum_{r=1}^{n} r^2$$

$$\Rightarrow \quad \sum_{r=1}^{n} r^2 = \frac{1}{3}\left(n^3 + \frac{3n(n+1)}{2} - n\right)$$

$$= \frac{1}{3}\left(\frac{2n^3 + 3n(n+1) - 2n}{2}\right)$$

$$= \frac{n}{6}(n+1)(2n+1)$$

2.1.14 If the sum of three consecutive terms of an A.P. is 27 and their product is 693, find the terms. (AHSEC 2006)

Solution: Let the terms be a, $a + d$, $a + 2d$

According to the question,

$\Rightarrow \quad a + a + d + a + 2d = 27$

$\Rightarrow \quad 3(a + d) = 27$

$\Rightarrow \quad a + d = 27/3$

$\Rightarrow \quad a + d = 9 \quad \ldots(1)$

Also, $a\,(a + d)\,(a + 2d) = 693$

$\Rightarrow \quad (9 - d)\,.\,9\,.\,(a + d + d) = 693$

$\Rightarrow \quad (9 - d)\,9\,(9 + d) = 693$

$\Rightarrow \quad (9^2 - d^2) = 77$

$\Rightarrow \quad d^2 = 4$

$\Rightarrow \quad d = \pm 2 \quad \ldots(2)$

So the terms are either 7, 9, 11 or 11, 9, 7, (from (1), (2))

2.1.15 A man saved Rs. 16,500 in ten years. In each year after the first he saved Rs. 100 more than he did in the preceeding year. How much did he save in the first year?

Solution: Here we are given

$$S_{10} = 16500, \text{ d} = 100$$

Now $$S_{10} = \frac{10}{2}\{2.a+(10-1)d\}$$

$\Rightarrow$ $$16500 = 5\{2a + 9.100\}$$

$\Rightarrow$ $$3300 = 2a + 900$$

$\Rightarrow$ $$2400 = 2a$$

$\Rightarrow$ $$a = 1200$$

So, in the first year he saved Rs. 1200

2.1.16 A man borrows Rs. 1000 and agrees to pay with a total interest of Rs. 140 in 12 instalments, each instalment being less than the immediately preceeding one by Rs. 10. Find the value of first instalment.

Solution: Here

$$S_{12} = 1140, d = (-10)$$

Now $$S_{12} = \frac{12}{2}\{2 . a+(12-1)(-10)\}$$

$\Rightarrow$ $$1140 = 6\{2a+11(-10)\}$$

$\Rightarrow$ $$190 = 2a-110$$

$\Rightarrow$ $$300 = 2a$$

$\Rightarrow$ $$a = 150$$

So, the value of Ist instalment is Rs. 150.

EXERCISE–2.1

1. Prove that if, $\frac{a}{b+c}, \frac{b}{c+a}, \frac{c}{a+b}$ are in A.P. then $\frac{1}{b+c}, \frac{1}{c+a}, \frac{1}{a+b}$ will also be in A.P. $(a+b+c \neq o)$
2. Find the five arithrmetic means between 5 and 35

   (Hints: Here first term = 5 & seventh term = 35, find second, third, fourth, fifth and sixth terms) **(Ans. 10, 15, 20, 25, 30)**
3. The *n*th term of a sequence be $a_n = 10 - 3n$, prove that sequence is an A.P.
4. Prove that sum of first 19 terms of the A.P. 18, 16, 14... is zero.
5. Find the sum upto first '*n*' terms of $\frac{x^2-1}{x} + x + \frac{x^2+1}{x} ...$

   $$\left(\textbf{Ans. } nx + \frac{n(n-3)}{2x}\right)$$
6. Show that sum of the digits from 20 to 50 is 1085.
7. If the sum upto first *p, q* and *r* terms of an A.P. be *a, b* and *c*, prove that

   $$\frac{a}{p}(q-r) + \frac{b}{q}(r-p) + \frac{c}{r}(p-q) = 0$$
8. Is –450, a term of the A.P. 8, 5, 2.... ? **(Ans. no)**
9. In an A.P., 10^{th} term is 23 and 32nd term is 67, find the first term and common difference. **(Ans. 5, 2)**
10. If 27 + 24 + 21 + ... to *n* terms = 132, find *n*. **(Ans. 8or 11)**

GEOMETRICAL PROGRESSION

A sequence $t_1, t_2, t_3...$ is said to be a Geometrical Progression (G.P.)

if $$\frac{t_2}{t_1} = \frac{t_3}{t_2} = \frac{t_4}{t_3} = ... = r$$

The first term is denoted as '*a*' and common ratio is denoted as '*r*'

Thus, G.P. is $a, ar, ar^2...$

Examples of G.P. are ...

1, 2, 4, 8 ....

2, 6, 18, 54...

200, 100, 50, 25...

The *n*th term of a G.P. having the first term '*a*' and common ratio '*r*' is

$$\boxed{t_n = a.r^{n-1}}$$

The sum of finite number of terms upto nth terms of a G.P. having the first term 'a' and common ratio 'r' is

$$\boxed{s_n = \frac{a\left(1-r^n\right)}{1-r}}\ ; r \neq 1$$

Proof: $s_n = a + ar + ... + ar^{n-1}$

$\Rightarrow \quad s_n + a.r^n = a + a.r + ... + a.r^{n-1} + a.r^n$

$\Rightarrow \quad s_n + a.r^n = a + r\ (a + ar + ... + a.r^{n-1})$

$\Rightarrow \quad s_n + a.r^n = a + r.s_n$

$\Rightarrow \quad s_n\ (1-r) = a(1-r^n)$

$$\Rightarrow \quad s_n = \frac{a\left(1-r^n\right)}{1-r}$$

On the otherhand, sum of an infinite G.P. and when $|r| < 1$, is

$$\boxed{s_\infty = \frac{a}{1-r}} \qquad \therefore\ \pi^n \rightarrow o \text{ as } n \rightarrow \alpha$$

SOLVED EXAMPLES

2.2.1 Find the 5th term of G.P.,

3, 12, 48...

Solution:

Here $\quad a = 3,\ r = 4$

$\therefore \quad t_5 = 3.4^{5-1}$

$= 3.4^4$

$= 768$

2.2.2 Find the sum upto first 6th terms of G.P. 1, 2, 4, 8...

Solution:

Here $\quad a = 1,\ r = 2$

$$\therefore \quad s_6 = \frac{1\left(1-2^6\right)}{1-2}$$

$= 63$

2.2.3 The sum of three terms in G.P. is 73 and their product is 512, find the terms.

Solution:

Let the terms be a, ar, ar^2

$\therefore \quad a + ar + ar^2 = 73$ ...(1)

and $\quad a.ar.ar^2 = 512$

$\Rightarrow \quad (ar)^3 = 8^3$

$\Rightarrow \quad ar = 8$

$\Rightarrow \quad a = \frac{8}{r}$ ...(2)

$\therefore \quad \frac{8}{r} + 8 + 8r = 73$ from (1), (2)

$\Rightarrow \quad \frac{8}{r}(1 + r + r^2) = 73$

$\Rightarrow \quad 8r^2 - 73r + 8r + 8 = 0$

$\Rightarrow \quad 8r^2 - 64r - r + 8 = 0$

$\Rightarrow 8r(r - 8) - 1(r - 8) = 0$

$\Rightarrow \quad (8r - 1)(r - 8) = 0$

$\Rightarrow \quad r = \frac{1}{8}$ or 8

$\therefore \quad a = 64$ or 1

So, the terms are 64, 8, 1 or 1, 8, 64

2.2.4 If a, b, c, d are in G.P. show that $\frac{1}{a^2+b^2}, \frac{1}{b^2+c^2}, \frac{1}{c^2+d^2}$ will also be in G.P.

Solution:

Since a, b, c, d are in G.P.

So, $\quad \frac{a}{b} = \frac{b}{c} = \frac{c}{d}$

$\Rightarrow \quad \frac{a}{b} = \frac{b}{c}$

$\Rightarrow \quad b^2 = a.c$ ...(1)

also, $\quad \frac{b}{c} = \frac{c}{d}$

$\Rightarrow \quad c^2 = bd$ ...(2)

Now $\quad \frac{a^2+b^2}{b^2+c^2} = \frac{a^2+ac}{ac+c^2} = \frac{a}{c}$ (3)

Also, $$\frac{b^2+c^2}{c^2+d^2} = \frac{b^2+bd}{bd+d^2} = \frac{b}{d} \quad (4)$$

again, $$\frac{a}{b} = \frac{c}{d} \Rightarrow \frac{a}{c} = \frac{b}{d} \quad ...(5)$$

From; (3), (4), (5)

$$\frac{a^2+b^2}{b^2+c^2} = \frac{b^2+c^2}{c^2+d^2}$$

Hence, a^2+b^2, b^2+c^2, c^2+d^2 are in G.P.

Now, $$\frac{\frac{1}{a^2+b^2}}{\frac{1}{b^2+c^2}} = \frac{b^2+c^2}{a^2+b^2}$$

$$= \frac{c^2+d^2}{b^2+c^2} = \frac{\frac{1}{b^2+c^2}}{\frac{1}{c^2+d^2}}$$

$$\therefore \quad \frac{\frac{1}{a^2+b^2}}{\frac{1}{b^2+c^2}} = \frac{\frac{1}{b^2+c^2}}{\frac{1}{c^2+d^2}}$$

So, $\frac{1}{a^2+b^2}, \frac{1}{b^2+c^2}, \frac{1}{c^2+d^2}$ will also be in G.P.

2.2.4 If *a, b, c* be in A.P. and *x, y, z* in G.P. prove that:

$$x^{b-c}.y^{c-a}.z^{a-b} = 1 \quad \textbf{(AHSEC, 01)}$$

Solution:

$\because$ a, b, c be in A.P.

$$\therefore \quad a-b = b-c \quad ...(1)$$

$\because$ x, y, z be in G.P.

$$\therefore \quad \frac{x}{y} = \frac{y}{z} = \Rightarrow y^2 = xz \Rightarrow y = \sqrt{x.z} \quad ...(2)$$

$$\therefore \quad x^{b-c}.y^{c-a}.z^{a-b}$$

$$= x^{b-c}.\left(\sqrt{xz}\right)^{c-a}.z^{b-c} \quad \textbf{[applying (2)]}$$

$$= x^{b-c}.x^{\frac{c-a}{2}}.z^{\frac{c-a}{2}}.z^{b-c}$$

$$= x^{b-c+\frac{c-a}{2}}.z^{\frac{c-a}{2}+b-c}$$

$$= x^{\frac{a+c}{2}-c+\frac{c-a}{2}} . z^{\frac{c-a}{2}+\frac{a+c}{2}-c} \quad (\because b = \frac{a+c}{2}, \text{ from (1)})$$

$$= x^{\frac{a+c-2c+c-a}{2}} z^{\frac{c-a+a+c-2c}{2}}$$

$$= x^0 . z^0$$

$$= 1$$

2.2.5 In a G.P. the sum of first n terms is 'S' the product is 'P' and sum of the reciprocals of the terms is 'R'. Show that $P^2 = \left(\frac{S}{R}\right)^n$ (AHSEC, 00)

Solution:

Let the terms be $a, ar \ldots ar^{n-1}$

Now, $S = a + ar + ar^2 + \ldots + ar^{n-1}$

$\Rightarrow$ $S = \dfrac{a\left(1-r^n\right)}{1-r}$ ...(1)

$\therefore$ $P = a.ar.ar^2 \ldots ar^{n-1}$

$\Rightarrow$ $P = a^n r^{1+2+3+ \quad +(n-1)}$

$\Rightarrow$ $P = a^n r^{\frac{n-1}{2}\left[2.1+\{(n-1)-1\}1\right]}$

$\Rightarrow$ $P = a^n . r^{\frac{n-1}{2}.n}$ ...(2)

$\therefore$ $R = \dfrac{1}{a} + \dfrac{1}{ar} + \ldots + \dfrac{1}{ar^{n-1}}$

$\Rightarrow$ $R = \dfrac{r^{n-1} + r^{n-2} + \ldots + r + 1}{a.r^{n-1}}$

$\Rightarrow$ $R = \dfrac{\frac{1\left(1-r^n\right)}{(1-r)}}{ar^{n-1}}$

$\Rightarrow$ $R = \dfrac{\left(1-r^n\right)}{a.r^{n-1}\left(1-r\right)}$ ...(3)

Now, $P^2 = a^{2n}.r^{n(n-1)}$ from (2)

and $\left(\dfrac{S}{R}\right)^n = \left(\dfrac{a\left(1-r^n\right)/\left(1-r\right)}{\left(1-r^n\right)/\left(1-r\right)a.r^{n-1}}\right)^n$ from (1), (3)

$$= a^{2n} r^{n(n-1)}$$
$$= P^2$$

2.2.6 If a, b, c are in A.P. and a, $b-a$, $c-a$ are in G.P. show that

$$a = \frac{b}{3} = \frac{c}{5}$$

Solution: $\because$ a, b, c are in A.P.

$\Rightarrow$ $a-b = b-c$

$\Rightarrow$ $b = \dfrac{a+c}{2}$ ...(1)

$\because$ a, $b-a$, $c-a$ are in G.P.

$\therefore$ $\dfrac{a}{b-a} = \dfrac{b-a}{c-a}$

$\Rightarrow$ $a(c-a) = (b-a)^2$

$\Rightarrow$ $(b-a) = \sqrt{a(c-a)}$

$\Rightarrow$ $b = a + \sqrt{a(c-a)}$ ...(2)

$\therefore$ From (1), (2)

$$\frac{a+c}{2} = a + \sqrt{a(c-a)}$$

$\Rightarrow$ $a + c - 2a = 2\sqrt{a(c-a)}$

$\Rightarrow$ $\sqrt{(c-a)} = 2\sqrt{a}$

$\Rightarrow$ $c-a = 4a$

$\Rightarrow$ $c = 5a$ ...(3)

From (1), (3)

$$b = \frac{a+5a}{2} = 3a \qquad ...(4)$$

From (3), (4)

$$a = \frac{c}{5} \Rightarrow \frac{b}{3}$$

2.2.7 Difference between G.M. and A.M. of two numbers is 16. If the ratio of the numbers is 1:25 find the numbers. **(AHSEC, 97)**

Solution: Let the numbers be a, b

$$\left.\begin{aligned} \text{Now} \qquad \text{G.M.} &= \sqrt{a.b} \\ \& \qquad \text{A.M.} &= \frac{a+b}{2} \end{aligned}\right\} \qquad ...(1)$$

Again, $a:b = 1:25$

$\Rightarrow \quad \frac{a}{b} = \frac{1}{25}$

$\Rightarrow \quad b = 25a \qquad ...(2)$

$\because \quad AM \geq G.M.$

$\Rightarrow \quad \frac{a+b}{2} - \sqrt{ab} = 16$

$\Rightarrow \quad \frac{a+25a}{2} - \sqrt{25a^2} = 16$

$\Rightarrow \quad 13a - 5a = 16$

$\Rightarrow \quad 8a = 16$

$\therefore \quad a = 2$

from (2), b = 50

$\therefore$ The numbers are 2 and 50.

2.2.8 Find four G.Ms between 2 and 64.

Solution:

Since there are four G.Ms. between 2 and 64, 2 is the 1st term of the G.P. and 64 is the 6th term of that G.P.

i.e. $a = 2, t_6 = 64$

Again, $t_6 = ar^{6-1}$

$\Rightarrow \quad 2.r^5 = 64$

$\Rightarrow \quad r^5 = 32 = 2^5$

$\Rightarrow \quad r = 2$

So, the G.Ms are

$a.r, a.r^2, ar^3, ar^4$

or, $2.2, 2.2^2, 2.2^3, 2.2^4 = 4, 8, 16, 32$

(i.e. 2, 4, 8, 16, 32, 64 is a G.P. with 6 terms or between 2 and 64 there are four G.M.s which are 4, 8, 16, 32)

2.2.9 Find the sum

$1 + (1 + x) + (1 + x + x^2) + ... +$ to n terms

Solution: The general term of the given series

$$t_n = 1 + x + x^2 + ... + x^{n-1}$$

$$= \frac{1(1-x^n)}{1-x}$$

$$\therefore \qquad s_n = t_1 + t_2 + \ldots + t_n$$

$$= \frac{1}{(1-x)}\left\{(1-x)+(1-x^2)+\ldots+(1-x^n)\right\}$$

$$= \frac{1}{(1-x)}\left\{n-(x+x^2+\ldots+x^n)\right\}$$

$$= \frac{1}{(1-x)}\left\{n-\frac{x(1-x^n)}{1-x}\right\}$$

2.2.10 Find the sum

$1 + 2x + 3x^2 + 4x^3 + \ldots + n$ th term

Solution:

Let $\qquad S = 1 + 2x + 3x^2 + \ldots + nx^{n-1}$ ...(1)

(1) × x, $\qquad x.S = x + 2x^2 + \ldots + (n-1)x^{n-1} + nx^n$ ...(2)

(1) − (2) $\Rightarrow$ $S(1-x) = 1 + x + x^2 + \ldots + x^{n-1} - nx^n$

$$\Rightarrow \qquad S(1-x) = \frac{1(1-x^n)}{1-x} - n.x^n$$

$$\Rightarrow \qquad S = \frac{1-x^n}{(1-x)^2} - \frac{nx^n}{(1-x)}$$

2.2.11 Find the sum

$1 + 3 + 7 + 15 + \ldots + t_n$

Solution:

$1 + (2^2-1) + (2^3-1) + (2^4-1) + \ldots + (2^n-1)$

$$= 2(1 + 2 + 2^2 + \ldots + 2^{n-1}) - n$$

$$= 2\left(\frac{1-2^n}{1-2}\right) - n$$

$$= 2(2^n-1) - n$$

2.2.12 Three numbers form a G.P. If the third term be reduced by 64, the resulting numbers would be in AP. Again if second term be reduced by 8 then resulting terms would form a G.P., find them.

Solution: Let the terms be a, b, c

Now, $\qquad \frac{a}{b} = \frac{b}{c} \Rightarrow b^2 = ac$ ...(1)

Again, $a-b = b-(c-64)$

$$\Rightarrow \quad b = \frac{a+c-64}{2} \qquad ...(2)$$

Given, $\dfrac{a}{b-8} = \dfrac{b-8}{c-64}$

$\Rightarrow \quad (b-8)^2 = ac-64a$

$\Rightarrow \quad b^2-16b + 64 = b^2-64a$ by (1)

$\Rightarrow \quad b = 4 + 4a \qquad ...(3)$

From (2) and (3), $2(4 + 4a) = a + c-64 \Rightarrow 7a = c-72 \qquad ...(4)$

Also, $(4 + 4a)^2 = a(7a + 72)$ [From (1), (3) and (4)],

$\Rightarrow 16 + 32a + 16a^2 = 7a^2 + 72a$

$\Rightarrow \quad 9a^2-40a + 16 = 0$

$\Rightarrow 9a^2-36a - 4a + 16= 0$

$\Rightarrow 9a(a-4) -4(a-4) = 0$

$\Rightarrow \quad (9a-4)(a-4) = 0$

$$\Rightarrow \quad a = 4 \text{ or } \frac{4}{9} \qquad ...(5)$$

From (3), (5) $\quad b = 4 + 4.4$ or $4 + 4.\dfrac{4}{9} = 20$ or $\dfrac{52}{9}$

From (4), (5), $\quad c = 7.4 + 72$ or $7.\dfrac{4}{9} + 72 = 100$ or $\dfrac{676}{9}$

So the terms are, 4, 20, 100 or $\dfrac{4}{9}, \dfrac{52}{9}, \dfrac{676}{9}$

2.2.13 Find the sum

$$1+\frac{1}{\sqrt{3}}+\frac{1}{3}+\frac{1}{3\sqrt{3}}+...$$

Solution:

Here $\quad r = \dfrac{1}{\sqrt{3}}$ and $\left|\dfrac{1}{\sqrt{3}}\right| < 1$

Also, the series is infinite

$\therefore$ Sum, $\quad S = \dfrac{1}{1-\frac{1}{\sqrt{3}}}$

$$= \frac{\sqrt{3}}{\sqrt{3}-1}$$

EXERCISE–2.2

1. If a, b, c, be the sums of first 'n' terms, the next n terms and the next n terms of a G.P. respectively. Show that a, b, c are in G.P.
2. Find the sum of first 'n' terms of a G.P. whose 3rd term is 4 and 12th term is 64.

$$\left(\textbf{Ans. } \frac{2\left\{\left(\sqrt{2}\right)^{n}-1\right\}}{\sqrt{2}-1}\right)$$

3. If 2, 7, 9 and 5 are respectively added to four terms which are in G.P., then the resulting terms would be in A.P. Find the terms.

(Ans. 5, 13, 21, 29)

4. The sum of three successive terms in G.P. is 21 and their product is 216, find the terms. (Ans. 3, 6, 12 or 12, 6, 3)
5. Find the sum

(a) $1+\frac{1}{2}+\frac{1}{4}+\frac{1}{8}+\ldots$

(b) $1-\frac{1}{2}+\frac{1}{4}-\frac{1}{8}+\ldots$ **(Ans. (a) 2, (b)** $\frac{2}{3}$**)**

6. If a, b, c, d be in G.P. then show that:
$(b-c)^2+(c-a)^2+(d-b)^2=(a-d)^2$
7. If p, q, r be in A.P. then show that p th, q th and r th term of a G.P. will also be in G.P.
8. If a, b, c are in both A.P. and G.P. then show that $a=b=c$
9. If a, b, c, d be in G.P. then show that a^2+b^2, b^2+c^2, c^2+d^2 are also in G.P.
10. If a, b, c be in G.P. show that

$\frac{1}{a+b},\frac{1}{2b},\frac{1}{b+c}$ will be in A.P.

3

PERMUTATIONS AND COMBINATIONS

Given a number of things, we can arrange them by taking one or more of them at a time in different orders and each of these arrangements is called a permutation. On the other hand the different groups or teams or selections that can be made from a given number of things by taking some or all of them at a time are called their combinations. It should be noted that combinations are independent of order.

For illustration, let us consider the following cases:

(a) Suppose we have to form two digit numbers from the three digits 1, 2, 3. Now the possible numbers are—
12, 13, 23, 21, 31, 32 (no digit is repeated).

(b) Suppose from three boys Ram, Shyam and Gopal, two boys are to be selected as representative. Then possible groups would be—
(Ram, Shyam), (Ram, Gopal), (Shyam, Gopal).

Here, in both the cases two are to be selected from a total of three. And in the first case six possibilities are found whereas in the second case three possibilities are found. Now the first case is permutation and second case is combination. And as we have seen, in the first case, the order has been considered (12 and 21 both are different) but in the second case, the order has not been considered ((Ram, Shyam) and (Shyam, Ram) are the same group).

Factorial Notation

The product of first 'n' natural numbers in any order (usually in ascending or descending order) is called factorial 'n' and is denoted by $\underline{|n}$ or $n!$. Thus

$$\boxed{\underline{|n} = 1.2\ldots(n-1).n}$$

In particular $\underline{|0} = 1$, $\underline{|1} = 1$, $\underline{|n} = n\underline{|n-1}$

Permutations of all different things

The number of permutations of n different things taking r ($\leq n$) of them at a time is denoted and defined as

$$\boxed{{}^nP_r = \frac{\underline{|n}}{\underline{|n-r}}}$$

Permutation of things not all different

If out of n given things, p things are alike of 1st kind, q things are alike of 2nd kind, r things are alike of 3rd kind and so on then the number of permutation of n given things taking all at a time is

$$\boxed{\frac{\underline{|n}}{\underline{|p}\,\underline{|q}\,\underline{|r}\ldots\ldots}};\quad p+q+r+\ldots=n$$

Combination of all different things

The number of combinations of n different things taking r ($\leq n$) of them at a time is denoted and defined as

$$\boxed{{}^{n}C_{r}=\frac{\underline{|n}}{\underline{|r}\,\underline{|n-r}}}$$

In particular—

(a) $${}^{n}P_{0}=\frac{\underline{|n}}{\underline{|n-0}}=1$$

(b) $${}^{n}P_{n}=\frac{\underline{|n}}{\underline{|n-n}}=\frac{\underline{|n}}{\underline{|0}}=\underline{|n}$$

(c) $${}^{n}C_{r}=\frac{\underline{|n}}{\underline{|n-r}\,\underline{|r}}=\frac{\underline{|n}}{\underline{|n-r}\,\underline{|n-(n-r)}}={}^{n}C_{n-r}$$

(d) $${}^{n}C_{o}={}^{n}C_{n}=1$$

(e) $${}^{n}C_{1}=\frac{\underline{|n}}{\underline{|n-1}\,\underline{|1}}=\frac{n\underline{|n-1}}{\underline{|n-1}}=n$$

Fundamental Counting Principles

(1) Multiplication principle
(2) Addition principle

Now let us discuss these principles—

Multiplication principle

If an event can occur in m different ways and if for each occurance of first event a second event can occur in n different ways then the two events in succession can occur in $m.n$ different ways.

Addition principle

If two events can occur independently in exactly m and n ways respectively, then either of the two events can occur in $m + n$ ways.

For illustration, let us consider the following two cases—

(a) There are 3 boys—Mrinal, Gopal, Kripal and two girls—Rita and Sita. In how many ways a group of one boy and one girl can be selected?

Sol. The possible groups are—(Mrinal, Sita), (Mrinal, Rita), (Gopal, Sita), (Gopal, Rita), (Kripal, Sita) and (Kripal, Rita). Thus the number of possible groups are—6(=3×2).

(b) From the above 5 persons, only '1' is to be selected, then how many possible selections can be made?

Sol. Here the possible selections are—Mrimal, Kripal, Gopal, Rita and Sita. Thus the total possibilities is—5(=3+2).

Now the first case is the example of application of multiplication principle and the second case is the example of application of addition principle.

Though the above statement is made for only two events, but the principles can be generalised for any number of events.

Solved examples

3.1. Show that the number of different triangles which can be formed by joining the angular points of a polygon of n sides is $\frac{1}{6}n.(n-1)(n-2)$. Show also that the figure has $\frac{1}{2}n(n-3)$ diagonals. (AHSEC, 04)

Solution:

To form a triangle, three angular points are to be joined which can be done in ${}^nC_3$ ways.

Now $${}^nC_3 = \frac{\underline{|n}}{\underline{|3}\,\underline{|n-3}}$$

$$= \frac{n(n-1)(n-2)\underline{|n-3}}{3.2.1\underline{|n-3}}$$

$$= \frac{n(n-1)(n-2)}{6}$$

Again, a diagonal can be formed by joining any two angular points which can be done in ${}^nC_2$ ways. But among these there are 'n' sides. So, the possible diagonals are ${}^nC_2 - n$ i.e.,

$$^nC_2 - n = \frac{\lfloor n}{\lfloor n-2 \lfloor 2} - n$$

$$= \frac{n(n-1)\lfloor n-2}{2.1.\lfloor n-2} - n$$

$$= n\left(\frac{n-1}{2} - 1\right)$$

$$= n\left(\frac{n-1-2}{2}\right)$$

$$= \frac{n(n-3)}{2}$$

3.2. In how many ways can the letters of the word ACTION be arranged so that vowels remain together?
(AHSEC, 02)

Solution:

In the word there are six letters, keeping vowels together there would be four, again among the three vowels themselves, arrangement would be possible. So number of arrangements would be

$$^4P_4 . {}^3P_3 = \lfloor 4 . \lfloor 3$$
$$= 4.3.2.1.3.2.1$$
$$= 144$$

3.3. A candidate would be declared passed if the candidate passes all the five subjects? In how many ways a candidate may fail?

Solution:

A candidate may fail if candidate fails one subject and the possible cases would be $^5C_1$. Or,

A candidate may fail if the candidate fails two subjects and the possible cases would be $^5C_2$. Or,

A candidate may fail if the candidate fails three subjects and the possible cases would be $^5C_3$. Or,

A candidate may fail if the candidate fails four subjects and the possible cases would be $^5C_4$. Or,

A candidate may fail if the candidate fails all the five subjects and the possible cases would be $^5C_5$.

Thus a candidate fails in the following way—

$$^5C_1 + {}^5C_2 + {}^5C_3 + {}^5C_4 + {}^5C_5 = 5 + \frac{\lfloor 5}{\lfloor 3\lfloor 2} + \frac{\lfloor 5}{\lfloor 3\lfloor 2} + 5 + 1$$

$$= 5 + \frac{5.4\lfloor 3}{\lfloor 3.2.1} + \frac{5.4\lfloor 3}{\lfloor 3.2.1} + 5 + 1$$

$$= 11 + 10 + 10$$

$$= 31$$

3.4. Prove that $^nC_x + {}^nC_{x-1} = {}^{n+1}C_x$. **(AHSEC 2006)**

Solution:

$$^nC_x + {}^nC_{x-1} = \frac{\lfloor n}{\lfloor n-x\lfloor x} + \frac{\lfloor n}{\lfloor x-1\lfloor n-x+1}$$

$$= \frac{\lfloor n}{\lfloor n-x.x\lfloor x-1} + \frac{\lfloor n}{\lfloor x-1(n-x+1)\lfloor n-x}.$$

$$= \frac{\lfloor n}{\lfloor n-x\,\lfloor x-1}\left(\frac{1}{x} + \frac{1}{n-x+1}\right)$$

$$= \frac{\lfloor n}{\lfloor x-1\lfloor n-x}\left(\frac{n-x+1+x}{x(n-x+1)}\right)$$

$$= \frac{\lfloor n+1}{\lfloor x\lfloor n-x+1}$$

$$= {}^{n+1}C_x$$

3.5. In how many ways 15 balls may be kept in 5 boxes, if the first box contains 4 balls?

Solution:

The 4 balls which are to be kept in 1st box, may be selected in $^{15}C_4$ ways.

Now the 1st of the remaining 11 balls may be kept in any one of the four boxes and this can be done in $^4C_1$ ways.

The 2nd of the remaining 11 balls may be kept in any one of the 4 boxes in $^4C_1$ ways.

....

The 11th of the remaining 11 balls may be kept in any one of the 4 boxes in $^4C_1$ ways.

So the total ways—

$$^{15}C_4.\underbrace{^4C_1.{}^4C_1 \ldots\ldots\ldots\ldots {}^4C_1}_{\text{11 Times}}$$

$= {}^{15}C_4.({}^4C_1)^{11}$
$= {}^{15}C_4.4^{11}$

3.6. In how many ways x heads can be achieved from 'n' throws of a coin?

Solution:

From n outcomes x heads are required and this can be achieved in ${}^nC_x$ ways.

3.7. In how many ways can the letters of the word INSURANCE be arranged such that vowels remain together? (AHSEC, 03)

Solution:

In the word, there are 9 letters among whom '4' are vowels and '2' N's. Now keeping vowels together there would be 6 groups but in these groups there are 2 Ns. So possible arrangements are—

$$\frac{\lfloor 6}{\lfloor 2}$$

But in these there are 4 vowels and the possible arrangements would be—

$${}^4P_4$$

So total arrangements are—

$$\frac{\lfloor 6}{\lfloor 2}.{}^4P_4 = \frac{\lfloor 6}{\lfloor 2}.\lfloor 4$$

$$= \frac{6.5.4.3.2.1.4.3.2.1}{2.1}$$

$$= 8640$$

3.8. From 7 gentlemen and 4 ladies a committee of 5 is to be formed. In how many ways can this be done to include at least one lady? (AHSEC, 97)

Solution:

The committee can be formed by the following possible ways

— 1 lady and 4 gentlemen in ${}^4C_1.{}^7C_4$ ways.

— 2 ladies and 3 gentlemen in ${}^4C_2.{}^7C_3$ ways.

— 3 ladies and 2 gentlemen in ${}^4C_3.{}^7C_2$ ways.

— 4 ladies and 1 gentleman in ${}^4C_4.{}^7C_1$ ways.

So total ways ${}^4C_1.{}^7C_4 + {}^4C_2.{}^7C_3 + {}^4C_3.{}^7C_2 + {}^4C_4.{}^7C_1$

$$= 4.\frac{\lfloor 7}{\lfloor 4\lfloor 3} + \frac{\lfloor 4}{\lfloor 2\lfloor 2}.\frac{\lfloor 7}{\lfloor 3\lfloor 4} + \frac{\lfloor 4}{\lfloor 3\lfloor 1}.\frac{\lfloor 7}{\lfloor 2\lfloor 5} + 1.7$$

$$= \frac{4.7.6.5.\lfloor 4}{\lfloor 4.3.2} + \frac{7.6.5.4\lfloor 3}{2.1.2.1\lfloor 3} + \frac{\lfloor 4.7.6.5.4.\lfloor 3}{\lfloor 3.5.\lfloor 4.2} + 7$$

$$= 140 + 210 + 84 + 7$$

$$= 441$$

3.9. A student is to answer 7 out of 10 questions in an examination.

(i) How many choice has he?

(ii) How many if he must answer the first 3 questions?

(iii) How many if he must answer atleast 4 of the first 5 questions?

(AHSEC, 95)

Solution:

(i) He may select 7 out of 10 in ${}^{10}C_7$ ways.

$$\frac{\underline{|10}}{\underline{|7}\,\underline{|3}} = \frac{10.9.8\underline{|7}}{\underline{|7}.3.2.1} = 120$$

(ii) Here total questions would be reduced to '7' and the questions to be answered would be reduced to '4'. So total ways—

$${}^{7}C_4 = \frac{\underline{|7}}{\underline{|3}\,\underline{|4}} = \frac{7.6.5\underline{|4}}{\underline{|4}.3.2} = 35$$

(iii) He may do in the following ways—

— 4 from the 1st five and 3 from rest five i.e., ${}^5C_4.{}^5C_3$.

— 5 from the 1st five and 2 from rest five i.e., ${}^5C_5.{}^5C_2$.

So total ways—

$${}^5C_4.{}^5C_3 + {}^5C_5.{}^5C_2 = 5.\frac{\underline{|5}}{\underline{|3}\,\underline{|2}} + 1.\frac{\underline{|5}}{\underline{|2}\,\underline{|3}}$$

$$= \frac{5.5.4\underline{|3}}{\underline{|3}.2.1} + \frac{5.4\underline{|3}}{\underline{|3}.2.1}$$

$$= 50 + 10$$

$$= 60$$

3.10. In a gettogether, each gentleman shakes hands once with one another in the meet. The total number of hand shakes is 28. Find the number of gentlemen in the meet. (AHSEC, 98)

Solution:

Let the number of gentlemen be n, then total handshakes is ${}^nC_2$.

Now according to the question

$${}^nC_2 = 28$$

$$\Rightarrow \qquad \frac{\underline{|n}}{\underline{|2}\,\underline{|n-2}} = 28$$

$$\Rightarrow \qquad \frac{n(n-1)\underline{|n-2}}{\underline{|n-2}} = 56$$

$\Rightarrow \quad n^2 - n - 56 = 0$

$\Rightarrow \quad n^2 - 8n + 7n - 56 = 0$

$\Rightarrow \quad n(n-8) + 7(n-8) = 0$

$\Rightarrow \quad (n+7)(n-8) = 0$

$\Rightarrow \quad n = 8 \text{ or } -7$

But number of persons cannot be negative, so the number of persons is 8.

3.11. Given $^nC_2 = 21$, find $^nP_2$. **(AHSEC, 00)**

Solution:

We know $^nC_2 = \dfrac{\lfloor n}{\lfloor n-2 \lfloor n}$ and $^nP_2 = \dfrac{\lfloor n}{\lfloor n-2}$

$\Rightarrow \quad \lfloor 2 . {}^nC_2 = \dfrac{\lfloor n}{\lfloor n-2}$

$\Rightarrow \quad {}^nP_2 = 2.{}^nC_2$

Now $\quad {}^nC_2 = 21$

$\therefore \quad {}^nP_2 = 2.21 = 42$

3.12. Given that $^{12}C_r = {}^{12}C_3$. Find r ($r \neq 3$). **(AHSEC, 00, 07)**

Solution:

We know

$$^nC_r = {}^nc_{n-r}$$

$\Rightarrow \quad {}^{12}C_r = {}^{12}C_{12-r}$

Also, $\quad 12_{c_r} = 12_{c_3}$

$\therefore \quad 12 - r = 3$

$\Rightarrow \quad r = 9$

3.13. Without using any formulla, show that $^nC_r = \dfrac{n}{r} . {}^{n-1}C_{r-1}$

Solution:

Let there be 'n' distinct numbers—a_1, a_2.............a_n. From these we have to find combination of r, then possible combination is ^nC_r, in which total letters would be—$r.{}^nC_r$, because in one combination total letters is 'r'.

This can also be thought of as follows—

Consider all combinations which contain a specific letter (say a_1) the possible number of such combinations is $^{n-1}C_{r-1}$ because then the total numbers from which selection is to be made would be reduced to '$n - 1$' and the numbers to be selected would be '$r - 1$'. Again as there are n distinct letters so total letters in all combinations would be—$n.{}^{n-1}C_{r-1}$, hence $r.{}^nC_r = n.{}^{n-1}C_{r-1}$ or ${}^nC_r = \frac{n}{r}.{}^{n-1}C_{r-1}$

3.14. From the letters of the word BUBUL, how many words can be formed?

Solution:

In the word, there are 5 letters, but 2 are of one type (B) and 2 are of one type (U). So total words—

$$\frac{\lfloor 5}{\lfloor 2 \lfloor 2} = \frac{5.4.3.\lfloor 2}{\lfloor 2.2.1} = 30$$

3.15. If ${}^np_r = 336$, ${}^nc_r = 56$, then find the value of r. (AHSCE, 08)

Solution:

Given ${}^np_r = 366$

$$\Rightarrow \qquad \frac{\lfloor n}{\lfloor n-r} = 336 \qquad (1)$$

$${}^nC_x = 56$$

$$\Rightarrow \qquad \frac{\lfloor n}{\lfloor n-r \lfloor r} = 56 \qquad (2)$$

$\therefore$ From (1) and (2)

$$\frac{\lfloor n}{\lfloor n-r} / \frac{\lfloor n}{\lfloor n-r \lfloor r} = \frac{336}{56}$$

$\Rightarrow$ $\lfloor r = 6$

$\Rightarrow$ r = 3

3.16. How many words can be formed out of the letters of ARTICLE, so that vowels always occupy the even places? (AHSCE, 08)

Solution:

This is a seven letter word and so there are three even places - 2nd, 4th and 6th. Again there are three vowels in the word. So these can occupy the even positions in ${}^3p_3 = \lfloor 3$ ways and the rest four positions would be occupied by four consonents in ${}^4p_4$ $\lfloor 4$ ways. So total ways (words) would be

$$\lfloor 3 . \lfloor 4$$

$$= 6.24$$

$$= 144$$

3.17. In how many ways can 10 men stand in a row so that

(a) two particular men always stand together?

(b) two particular men don't stand together?

Solution:

(a) Considering those particular to be one single unit the total positions reduces to 9 so the total possible arrangements would be ${}^9p_9$ i.e. $\underline{|9}$. Again the particular two persons can interchange their positions in ${}^2p_2$ i.e. $\underline{|2}$ ways.

Hence ultimate arrangements would be

$$\underline{|2} \cdot \underline{|9} \text{ i.e. } 2\,\underline{|9}$$

(b) If there would have been no condition, then there would be ${}^{10}p_{10}$ i.e. $\underline{|10}$ arrangements. But among these $2\,\underline{|9}$ cases would be there, where two particular persons always stand together. So, total cases where two particular persons cannot stand together is

$$\underline{|10} - 2\,\underline{|9}$$

$$= \underline{|10} \cdot \underline{|9} - \underline{|2} \cdot \underline{|9} = 8\,\underline{|9}$$

EXERCISE–3

1. Prove that
   ${}^{n}C_{x} + {}^{n-1}C_{x-1} + {}^{n-1}C_{x-2} = {}^{n+1}C_{x}$
2. In how many ways 15 balls may be kept in 5 boxes? **(Ans. 5^{15})**
3. How many times x heads can be achieved from n throws of a coin?
   (Ans. ${}^{n}c_{r}$)
4. From 8 persons a group of 5 or more persons is to be made, in how many ways this can be done?
   (Ans. 93)
5. A student has to answer 6 questions from two groups A and B, each comprising of 5 questions. In how many ways this can be done if at most 4 questions can be done from each group? **(Ans. 200)**
6. Find the number of sides of a polynomial if it has 90 diagonals. **(Ans. 15)**
7. If ${}^{n}C_{12} = {}^{n}C_{8}$, find n. **(Ans. 20)**
8. In how many ways one can invite 3 or more of his 6 friends?**(Ans. 42)**
9. Show that ${}^{n}P_{r} = {}^{n-1}P_{r} + r.{}^{n-1}P_{r-1}$.
10. If ${}^{n}P_{5} : {}^{n}P_{3} = n : 1$, find n. **(Ans. 6)**
11. How many arrangements are possible with the letters of the word ENGINEERING? **(Ans. 277200)**
12. In how many ways can the letters of the word SUNDAY be arranged?
    (Ans. 720)
13. If ${}^{n}P_{r} = {}^{n}P_{r+1}$ and ${}^{n}C_{r} = {}^{n}C_{r-1}$, find n and r. **(Ans. 3, 2)**
14. Prove that:
    (a) ${}^{n}C_{r} + 2.{}^{n}C_{r-1} + {}^{n}C_{r-2} = {}^{n+2}C_{r}$
    (b) ${}^{n}C_{r} + {}^{n-1}C_{r-1} + {}^{n-1}C_{r-2} = {}^{n+1}C_{r}$
15. Given ${}^{n}P_{2} = 42$, find ${}^{n}C_{2}$ **(AHSEC 2007)**
    (Ans. 21)

4

BINOMIAL THEOREM

An expression involving two terms a and x is called a binomial expression and the theorem which gives us the general rule for expanding $(a + x)^n$ is known as the binomial theorem. The binomial theorem for positive integral index is stated as follows:

If n be any positive integer then

$$
\begin{aligned}
(a + x)^n &= {}^nC_0a^n + {}^nC_1a^{n-1}.x + {}^nC_2a^{n-2}x^2 \\
&\quad + \ldots + {}^nC_ra^{n-r}x^r + \ldots + {}^nC_nx^n \\
&= a^n + na^{n-1}x + \frac{n(n-1)}{\underline{|2}}a^{n-2}x^n \\
&\quad + \ldots \frac{n(n-1)\ldots(n-r+1)}{\underline{|r}}a^{n-r}x^r + \ldots + x^n
\end{aligned}
$$

for all values of a and x.

Proof: (By the method of mathematical induction) Let us consider that the theorem holds good for $n = k$ i.e.,

$$(a + x)^k = {}^kC_0a^k + {}^kC_1a^{k-1}x + {}^kC_2a^{k-2}x^2 + \ldots + {}^kC_ra^{k-r}x^r + \ldots + x^K$$

Multiplying both sides by $(a + x)$

$$
\begin{aligned}
(a + x)^{k+1} &= (a + x)\,({}^kC_0a^k + {}^kC_1a^{k-1}.x + \ldots + {}^kC_ra^{k-r}x^r + \ldots + x^k) \\
&= ({}^kC_0a^{k+1} + {}^kC_1a^kx + {}^kC_2a^{k-1}x^2 + \ldots \\
&\quad + {}^kC_{r-1}a^{k-r+2}x^{r-1} + {}^kC_ra^{k-r+1}x^r + \ldots + ax^k) \\
&\quad + ({}^kC_0a^kx + {}^kC_1a^{k-1}x^2 + {}^kC_2a^{k-2}x^3 + \ldots \\
&\quad + {}^kC_{r-1}a^{k-r+1}x^r + {}^kC_ra^{k-r}x^{r+1} + \ldots + x^{k+1}) \\
&= a^{k+1} + ({}^kC_0 + {}^kC_1)a^kx + ({}^kC_2 + {}^kC_1)a^{k-1}x^2 \\
&\quad + \ldots + ({}^kC_r + {}^kC_{r-1})a^{k-r+1}x^r + \ldots + x^{k+1} \\
&= a^{k+1} + {}^{k+1}C_1a^{(k+1)-1}x + {}^{k+1}C_2a^{(k+1)-2}x^2 \\
&\quad + \ldots + {}^{k+1}C_ra^{k+1-r}x^r + \ldots + x^{k+1}
\end{aligned}
$$

Thus the theorem is true for $n = k + 1$ too, so by the principle of mathematical induction it is true for any $n \in N$.

Remarks: (1) The total number of terms is $(n + 1)$ in $(a + x)^n$.

(2) The $(r + 1)$th term of $(a + x)^n$ i.e., ${}^nC_ra^{n-r}x^r$ is called the general term.

(3) The co-efficients ${}^nC_0$, ${}^nC_1$ ... ${}^nC_n$ are called binomial co-efficients.

(4) *Middle term:* If 'n' is even then in the expanstion of $(a + x)^n$ there is only

one middle term which is $\left(\frac{n}{2}+1\right)$th term i.e., ${}^{n}C_{\frac{n}{2}}a^{n-\frac{n}{2}}x^{\frac{n}{2}}$.

But if '*n*' is odd, then there are two middle terms—$\left(\frac{n-1}{2}+1\right)^{th}$ and $\left(\frac{n+1}{2}+1\right)^{th}$ term, i.e., ${}^{n}C_{\frac{n-1}{2}}a^{n-\frac{n-1}{2}}x^{\frac{n-1}{2}}$ and ${}^{n}C_{\frac{n+1}{2}}a^{n-\frac{n+1}{2}}x^{\frac{n+1}{2}}$.

General Binomial Theorem

If n be any rational number then

$$(1+x)^n = 1 + nx + \frac{n(n-1)}{\underline{|n}}x^2 + \ldots + \frac{n(n-1)\ldots(n-r+1)}{\underline{|r}}x^r + \ldots..$$

provided $|x| < 1$, in case of n being a negative integer or fraction.

SOLVED EXAMPLES

4.1. Prove that (a) ${}^{n}C_0 + {}^{n}C_1 + \ldots {}^{n}c_n = 2^n$

(b) ${}^{n}C_1 + {}^{n}C_3 + \ldots = {}^{n}C_0 + {}^{n}C_2 + \ldots = 2^{n-1}$

Solution:

(a) $(1+x)^n = {}^{n}C_0 + {}^{n}C_1x + {}^{n}C_2x^n + \ldots + {}^{n}C_nx^n$

Putting x = 1, we get—

$\Rightarrow \quad 2^n = {}^{n}C_0 + {}^{n}C_1 + {}^{n}C_2 + \ldots + {}^{n}C_n$

(b) $(1+x)^n = {}^{n}C_0 + {}^{n}C_1x + {}^{n}C_2x^n + \ldots + {}^{n}C_nx^n$

Putting $x = -1$, we get—

$0 = {}^{n}C_0 - {}^{n}C_1 + {}^{n}C_2 - {}^{n}C_3 + \ldots$

$\Rightarrow \quad {}^{n}C_0 + {}^{n}C_2 + {}^{n}C_4 + \ldots = {}^{n}C_1 + {}^{n}C_3 + \ldots$

But $\quad {}^{n}C_0 + {}^{n}C_1 + {}^{n}C_2 + {}^{n}C_3 + \ldots = 2^n$

$\Rightarrow \quad ({}^{n}C_0 + {}^{n}C_2 + {}^{n}C_4 + \ldots) + ({}^{n}C_1 + {}^{n}C_3 + \ldots) = 2^n$

$\Rightarrow \quad 2({}^{n}C_0 + {}^{n}C_2 + \ldots) = 2^n$

$\Rightarrow \quad 2({}^{n}C_1 + {}^{n}C_3 + \ldots) = 2^n$

$\Rightarrow \quad {}^{n}C_0 + {}^{n}C_2 + {}^{n}C_4 + \ldots = {}^{n}C_1 + {}^{n}C_3 + \ldots$

$= \frac{2^n}{2} = 2^{n-1}$

4.2. Find the general term of $\left(x^2+\dfrac{3}{x}\right)^6$. **(AHSEC, 07)**

Solution:

$$\text{General term} = t_{r+1} = {}^6C_r(x^2)^{6-r}\left(\frac{3}{x}\right)^r$$
$$= {}^6C_r x^{12-2r}\frac{3^r}{x^r}$$
$$= {}^6C_r.3^r x^{12-3r}$$

4.3. Find the co-efficient of x^7 in the expansion of $\left(x-\dfrac{1}{x^2}\right)^{13}$.

Solution:

Let general term contains x^7, now-general term

$$= {}^{13}C_r x^{13-r}\left(-\frac{1}{x^2}\right)^r$$
$$= {}^{13}C_r x^{13-r}(-1)^r x^{-2r}$$
$$= {}^{13}C_r{}^{13-3r}(-1)^r$$

Now to contain x^7, evidiently

$$x^{13-3r} = x^7$$
$$\Rightarrow \quad 13-7 = 3r$$
$$\Rightarrow \quad r = 2$$

$\therefore$ Co-efficient is ${}^{13}C_2(-1)^2$

$$= \frac{|\underline{13}}{|\underline{11}\,|\underline{2}}$$
$$= \frac{13.12.|\underline{11}}{2.|\underline{11}}$$
$$= 78$$

4.4. Show that there will be no term containing x^9 in the expansion of $\left(2x^2-\dfrac{1}{x}\right)^{20}$ **(AHSEC, 04)**

Solution:

Let the general term contains x^9 (if possible). Now general term is

$${}^{20}C_r(2x^2)^{20-r}\left(-\frac{1}{x}\right)^r$$
$$= {}^{20}C_r 2^{20-r}x^{40-2r}(-1)^r x^{-r}$$

$$= {}^{20}C_r 2^{20-r} x^{40-3r} (-1)^r$$

To contain x^9,

$$x^{40-3r} = x^9$$

$$\Rightarrow \quad 40 - 3r = 9$$

$$\Rightarrow \quad 31 = 3r$$

$$\Rightarrow \quad r = \frac{31}{3}$$

which is not an integer, but evidently r is integer and hence $40 - 3r \neq 9$ or there is no term containing x^9.

4.5. Find the middle term/terms in the expansion of $\left(a^4 - \frac{1}{a^3}\right)^{15}$.

(AHSEC, 02)

Solution:

Here $n = 15$ is odd and so there will be two middle terms—$\left(\frac{15-1}{2}+1\right)$th and $\left(\frac{15+1}{2}+1\right)$th terms i.e. (7 + 1)th and (8 + 1)th term. And these are—

$$t_{(7+1)} = {}^{15}C_7 (a^4)^{15-7}\left(-\frac{1}{a^3}\right)^7$$

$$= {}^{15}C_7 a^{60-28}(-1)^7 a^{-21}$$

$$= {}^{15}C_7 a^{60-49}(-1)^7$$

$$= -{}^{15}C_7 a^{11}$$

and

$$t_{(8+1)} = {}^{15}C_8 (a^4)^{15-8}\left(-\frac{1}{a^3}\right)^8$$

$$= {}^{15}C_8 a^{60-32}(-1)^8 (a)^{-24}$$

$$= {}^{15}C_8 a^4$$

4.6. In the expansion of $\left(x^2 - \frac{a}{x}\right)^9$ the term free from x is 5376. Find *a*.

(AHSEC, 00)

Solution:

Let the x free term be t_{r+1}.

Now $\quad t_{r+1} = {}^9C_r(x^2)^{9-r}\left(-\dfrac{a}{x}\right)^r = 5376$

$\Rightarrow \quad {}^9C_r\, x^{18-2r}(-a)^r x^{-r} = 5376$

Now to be x free, $\quad x^{18-3r} = x^0$

$\Rightarrow \quad 18 - 3r = 0$

$\Rightarrow \quad r = 6$

So $\quad {}^9C_6\, x^0\, (-a)^6 = 5376$

$$\Rightarrow \quad \frac{\underline{|9}}{\underline{|6}\,\underline{|3}}a^6 = 5376$$

$$\Rightarrow \quad \frac{9.8.7.\underline{|6}}{\underline{|6}.3.2}\, a^6 = 5376$$

$\Rightarrow \quad a^6 = 64 = 2^6$

$\Rightarrow \quad a = 2$

4.7. Find the co-efficient of x^7 in the expansion of $(x + x^2 + x^3 + \ldots + x^6)^2$.

Solution:

$$x + x^2 + x^3 + \ldots + x^6 = \left(\frac{x(1-x^5)}{1-x}\right) \quad \because \text{This is a G.P. with } a = x,\ r = x$$

$$\therefore \quad (x + x^2 + \ldots + x^6)^2 = \left\{\frac{x(1-x^5)}{(1-x)}\right\}^2$$

$$= x^2(1 - x^5)^2 (1 - x)^{-2}$$

$$= x^2(1 - 2x^5 + x^{25})\,(1 + (-2)(-x) + \frac{(-2)(-2-1)}{\underline{|2}}(-x)^2 + \ldots)$$

$$= (x^2 - 2x^7 + x^{27})\,(1 + 2x + 3x^2 + 4x^3 + 5x^4 + 6x^5 + \ldots)$$

$\therefore$ Co-efficient of x^7 is $6 - 2 = 4$.

4.8. Find the value of

(i) $(1 - x)^{-1}$ (ii) $(1 - x)^{-2}$ (iii) $(1 + x)^{-1}$ (iv) $(1 + x)^{-2}$

Solution:

(i) $$(1 - x)^{-1} = 1 + (-1)(-x) + \frac{(-1)(-1-1)}{\underline{|2}}(-x)^2 + \frac{(-1)(-1-1)(-1-2)}{\underline{|3}}(-x)^3 + \ldots\ldots\ldots$$

$$= 1 + x + \frac{1.2}{\lfloor 2}x^2 + \frac{1.2.3}{\lfloor 3}x^3 + ...$$

$$= 1 + x + x^2 + x^3 + ...$$

(ii) $$(1 - x)^{-2} = 1 + (-2)(-x) + \frac{(-2)(-2-1)}{\lfloor 2}(-x)^2$$

$$+ \frac{(-2)(-2-1)(-2-2)}{\lfloor 3}(-x)^3 + ...$$

$$= 1 + 2x + 3x^2 + 4x^3 + ...$$

(iii) $$(1 + x)^{-1} = 1 + (-1)x + \frac{(-1)(-1-1)}{\lfloor 2}x^2 + \frac{(-1)(-1-1)(-1-2)}{\lfloor 3}x^3 + ...$$

$$= 1 - x + x^2 - x^3 + ...$$

(iv) $$(1 + x)^{-2} = 1 + (-2)x + \frac{(-2)(-2-1)}{\lfloor 2}x^2 + \frac{(-2)(-2-1)(-2-2)}{\lfloor 3}x^3 + ...$$

$$= 1 - 2x + 3x^2 - 4x^3 + ...$$

4.9. Find the co-efficient of x^8 in the expansion of $\left(\frac{1-x^6}{1-x}\right)^2$.

(AHSEC, 94)

Solution:

$$\left(\frac{1-x^6}{1-x}\right)^2 = (1 - x^6)^2 (1 - x)^{-2}$$

$$= (1 - 2x^6 + x^{12})(1 + 2x + 3x^2 + 4x^3 + 5x^4 + 6x^5 + 7x^6 + 8x^7 + 9x^8 + ...)$$

$\therefore$ Co-efficient of x^8—

$9 - 6 = 3$

4.10. Prove that—

$$^nC_0 + \frac{^nC_1}{2} + \frac{^nC_2}{3} + ... + \frac{^nC_n}{n+1} = \frac{2^{n+1}-1}{n+1}.$$

(AHSEC, 94)

Solution:

$$\text{L.H.S.} = {}^nC_0 + \frac{\lfloor n}{\lfloor n-1.2\lfloor 1} + \frac{\lfloor n}{\lfloor n-2.3.\lfloor 2} + ... + \frac{\lfloor n}{\lfloor n-n(n+1)\lfloor n}$$

$$= 1 + \frac{\lfloor n}{\lfloor n-1 \lfloor 2} + \frac{\lfloor n}{\lfloor n-2 \lfloor 3} + \ldots\ldots\ldots + \frac{\lfloor n}{\lfloor n-n \lfloor n+1}$$

$$= \frac{1}{n+1}\left(n+1+\frac{(n+1)\lfloor n}{\lfloor 2 \lfloor (n+1)-2} + \frac{(n+1)\lfloor n}{\lfloor 3 \lfloor (n+1)-3}\right.$$

$$\left. + \ldots\ldots\ldots + \frac{(n+1)\lfloor n}{\lfloor n+1 \lfloor (n+1)-(n+1)}\right)$$

$$= \frac{1}{n+1}\left({}^{n+1}C_1 + {}^{n+1}C_2 + {}^{n+1}C_3 + \ldots\ldots\ldots + {}^{n+1}C_{n+1}\right)$$

$$= \frac{1}{n+1}\left({}^{n+1}C_0 + {}^{n+1}C_1 + {}^{n+1}C_2 + \ldots\ldots\ldots + {}^{n+1}C_{n+1} - {}^{n+1}C_0\right)$$

$$= \frac{1}{n+1}\left(2^{n+1} - 1\right) \quad \text{(See example 4.1(a) } (n+1) \text{ in place of } n\text{)}$$

$$= \frac{2^{n+1} - 1}{n+1}$$

4.11. Prove that—

$$1 + \frac{2n}{3} + \frac{2n(2n+2)}{3.6} + \frac{2n(2n+2)(2n+4)}{3.6.9} + \ldots\ldots\ldots$$

$$= 2^n \left\{1 + \frac{n}{3} + \frac{n(n+1)}{3.6} + \frac{n(n+1)(n+2)}{3.6.9} + \ldots\ldots\ldots\ldots\ldots\ldots\right\}$$

Solution:

$$\text{L.H.S.} = 1 + n.\frac{2}{3} + \frac{n(n+1)}{2.1}.\frac{2^2}{3^2} + \frac{n(n+1)(n+2)}{3.2.1}.\frac{2^3}{3^3} + \ldots\ldots\ldots$$

$$= 1 + (-n)\left(-\frac{2}{3}\right) + \frac{(-n)(-n-1)}{\lfloor 2}\left(-\frac{2}{3}\right)^2$$

$$+ \frac{(-n)(-n-1)(-n-2)}{\lfloor 3}\left(-\frac{2}{3}\right)^3 + \ldots\ldots\ldots$$

$$= \left(1 - \frac{2}{3}\right)^{-n}$$

$$= \left(\frac{1}{3}\right)^{-n} \qquad \ldots(1)$$

$$\text{R.H.S.} = 2^n\left\{1+(-n)\left(-\frac{1}{3}\right)+\frac{(-n)(-n-1)}{2.1}\left(-\frac{1}{3}\right)^2\right.$$

$$\left.+\frac{(-n)(-n-1)(-n-n)}{3.2.1}\left(-\frac{1}{3}\right)^3+\ldots\ldots\ldots\right\}$$

$$= 2^n\left(1-\frac{1}{3}\right)^{-n}$$

$$= 2^n\left(\frac{2}{3}\right)^{-n}$$

$$= \frac{2^n.2^{-n}}{3^{-n}} \qquad \ldots(2)$$

$\therefore$ From (1), (2); L.H.S. = R.H.S.

4.12. Expand $\left(x^2+\frac{y}{2}\right)^6$ **(AHSEC, 08)**

Solution:

$$\left(x^2+\frac{y}{2}\right)^6$$

$$= \left(x^2\right)^6 + {}^6C_1\left(x^2\right)^{6-1}\left(\frac{y}{2}\right)$$

$$+ {}^6C_2\,(x^2)^{6-2}\left(\frac{y}{2}\right)^2$$

$$+ {}^6C_3\,(x^2)^{6-3}\left(\frac{y}{2}\right)^3$$

$$+ {}^6C_4\,(x^2)^{6-4}\left(\frac{y}{2}\right)^4$$

$$+ {}^6C_5\,(x^2)^{6-5}\left(\frac{y}{2}\right)^5$$

$$+ \left(\frac{y}{2}\right)^6$$

$$= x^{12} + 6.x^{10}.\frac{y}{2} + 15.x^8.\frac{y^2}{4}$$

$$+ 20.x^6.\frac{y^3}{8} + 15.x^4.\frac{y^4}{16}$$

$$+ 6.x^2.\frac{y^5}{32} + \frac{y^6}{64}$$

$$= x^{12} + 3.x^{10}.y + \frac{15}{4}.x^8.y^2$$

$$+ \frac{5}{2}.x^6.y^3. + \frac{15}{10}.x^4.y^4$$

$$+ \frac{3}{16}.x^2.y^5 + \frac{y^6}{64}$$

4.13. Find the term independent of x in the expansion of $\left(x+\frac{1}{x}\right)^{10}$

(AHSEC, 06)

Solution: Let $(r + 1)$th term is independent of x

Now $(r + 1)$th term i.e. t_{r+1} is

$$t_{r+1} = {}^{10}C_r\, x^{10-r}\left(\frac{1}{x}\right)^r$$

$$= {}^{10}C_r x^{10-r} x^{-r}$$

$$= {}^{10}C_r\, x^{10-2r}$$

For this term to be independent of x

$$x^{10-2r} = x^0$$

$$\Rightarrow \quad 10-2r = 0$$

$$\Rightarrow \quad r = 5$$

So the term independent of x is t_{5+1} i.e. t_6 which is

$$t_6 = t_{5+1}$$

$$= {}^{10}C_5$$

$$= \frac{10.9.8.7.6.\underline{|5}}{54.3.2.\underline{|5}}$$

$$= 252$$

EXERCISE–4

1. Prove that—

$${}^nC_0.{}^nC_r + {}^nC_1.{}^nC_{r-1} + ... + {}^nC_{n-r}.{}^nC_n = \frac{\underline{|2n}}{\underline{|n-r}\underline{|n+r}}$$

(*Hints* : $(1+x)^{2n} = (1+x)^n (1+x)^n$,expand both sides and then compare the co-efficient of x^{n-r})

2. If n be an even positive number then prove that—

$$\frac{1}{\underline{|1}\underline{|n-1}} + \frac{1}{\underline{|3}\underline{|n-3}} + ... + \frac{1}{\underline{|n-1}\underline{|1}} = \frac{2^{n-1}}{\underline{|n}}$$

3. Prove that—
$(1 - x + x^2 - ....) (1 + x + x^2 + ...) = (1 + x^2 + x^4 + ...)$

4. Find the term independent of x in $\left(2x - \frac{1}{x^2}\right)^9$. **(Ans. –5376)**

5. Find the co-efficient of

(i) x^7 in $\left(x - \frac{1}{x^2}\right)^{13}$

(ii) x^9 in $\left(2x^2 + \frac{1}{x}\right)^{20}$ **(Ans. (1) 78 (ii) 0)**

6. Find the middle term in

(i) $\left(2x - \frac{1}{y}\right)^8$

(ii) $\left(x + \frac{1}{x}\right)^{21}$ **(Ans. (i) $1120\frac{x^4}{y^4}$ (ii) ${}^{21}C_{10}.x$, ${}^{21}C_{11}\frac{1}{x}$**

7. Show that the middle term in $\left(x + \frac{1}{x}\right)^{2n}$ is ${}^{2n}C_n$.

8. Write down the 4th term in $\left(2x^2 - \frac{1}{4x}\right)^{11}$ **(Ans. $-660x^{13}$)**

9. In the expansion of $(1 + x)^{12}$; coefficients of $(r + 1)$th term : coefficients rth term = 6 : 7; find the co-efficient of rth term. **(Ans. ${}^{12}C_6$)**

10. Find the value of r if the co-efficients of x^r and x^{r+1} in $(3x + 2)^{19}$ are equal. **(Ans. 11)**

11. Write the expansion of $(1 + x)^2$. **(AHSEC 2006, 07)**

5

LOGARITHMS, EXPONENTIAL AND LOGARITHMIC SERIES

DEFINITION OF LOGARITHM

If $a^x = n$; $a, x, n \in R$ and $a > 0$, but $a \neq 1$ then the index x is defined as the logarithm (log) of the number n w.r.t. the base 'a' and it is written as

$$x = \log_a n$$

For example logarithm of 16 base 2 is '4' which can be found as–

$$\log_2 16 = x \text{ (say)}$$
$$\Rightarrow \quad 2^x = 16 = 2^4$$
$$\Rightarrow \quad x = 4$$
$$\therefore \quad \log_2 16 = 4$$

Properties

(1) $\log_a 1 = 0$

proof: $a^0 = 1$

$\Rightarrow \quad \log_a 1 = 0$

(2) $\log_a a = 1$

proof: $a^1 = a$

$\Rightarrow \quad \log_a a = 1$

(3) $\log_a m.n = \log_a m + \log_a n$

proof: Let $a^x = m$

$\Rightarrow \quad \log_a m = x$

also, $a^y = n$

$\Rightarrow \quad \log_a n = y$

Again $a^x.a^y = a^{x+y}$

$\Rightarrow \quad m.n = a^{x+y}$

$\Rightarrow \quad \log_a m.n = x + y$

$\Rightarrow \quad \log_a m.n = \log_a m + \log_a n$

(4) $\log_a \left(\frac{m}{n}\right) = \log_a m - \log_a n$

proof: Let $a^x = m$

$\Rightarrow \quad log_a m = x$

also $\quad a^y = n$

$\Rightarrow \quad \log_a n = y$

also, $\quad \dfrac{a^x}{a^y} = a^{x-y}$

$\Rightarrow \quad \dfrac{m}{n} = a^{x-y}$

$\Rightarrow \quad a^{x-y} = \dfrac{m}{n}$

$\Rightarrow \log_a\left(\dfrac{m}{n}\right) = x - y$

$\Rightarrow \log_a\left(\dfrac{m}{n}\right) = \log_a m - \log_a n$

(5) $\quad \log_a m^n = n \log_a m$

proof: Let $a^x = m$

$\Rightarrow \quad (a^x)^n = m^n \Rightarrow a^{x.n} = m^n$

$\Rightarrow \quad \log_a m^n = x.n \Rightarrow \log_a m^n = (\log_a m)\, n \; (\because\; a^x = m \Rightarrow \log_a m = x)$

$\Rightarrow \quad \log_a m^n = n \log_a m$

(6) $\quad \log_a m = \log_b m . \log_a b$

proof: Let $a^x = b^y = m$

$\therefore \quad \log_a m = x \quad$ and $\quad \log_b m = y$

Again $\quad x = \log_a m = \log_a b^y = y \log_a b$

$\Rightarrow \quad \log_a m = y. \log_a b \quad \Rightarrow \quad \log_a m = \log_b m . \log_a b$

(7) $\quad \log_a 0 = -\infty;\ a > 1$

proof: $-\quad a^{-\infty} = 0;\ a > 1$

$\Rightarrow \quad \log_a 0 = -\infty$

Common and Natural logarithm

Logarithm with base 10 are called common logarithms and logarithms with base e are called natural (Naperian) logarithms.

Use of Common Logarithms for Calculations

Logarithm of any number always has two parts, one is integral and the other is decimal. The integral part is termed as *characteristic* whereas the decimal part is termed as mantissa. The characteristic can be found by observation whereas mantissa can be found by using a standard table called log table.

How to find Characteristic

(a) Numbers greater than one

Numbers having one digit in the integral part always lie between 1 and 10 and therefore logarithom of such numbers always will be between $\log_{10} 1$ and $\log_{10} 10$. For illustration 3.78 is such a number which is between 1 and 10 and so $\log_{10} 3.78$ will be between $\log_{10} 1$ and $\log_{10} 10$. Again as $\log_{10} 1 = 0$ ($\because$ $10^0 = 1$) and $\log_{10} 10 = 1$, so $\log_{10} 3.78$ will be between 0 and 1or $\log_{10} 3.78 = 0 + a$ decimal. Hence characteristic of $\log_{10} 3.78 = 0$ (by definition of characteristic).

Similarly, numbers having two digits in the integral part always lie between 10 and 100, so, logarithm of such a number, say 32.5 will be between $\log_{10} 10$ (= 1) and $\log_{10} 100$ (= 2), or $\log_{10} 32.5 = 1 + a$ decimal and hence characteristic of $\log_{10} 32.5 = 1$.

Similarly, for numbers having three digits in the integral part, logarithm is 2 + *a* decimal. So characteristic of logarithm of such a number say 563.52 is 2.

In brief we can say that characteristic of logarithm of a number greater than one is postive and is one less than the number of digits in the integral part of the number.

Thus, characteristic of $\log_{10} 1.53 = 0$, $\log_{10} 23.67 = 1$, $\log_{10} 345.23 = 2$, $\log_{10} 3758.91 = 3$ etc.

(b) Numbers less than one but greater than zero

Numbers having no zero immediately after the decimal point always lie between 0.1 and 1, so logarithm of such a number say 0.785 will be between $\log_{10} 0.1$ (= – 1, since $10^{-1} = 0.1$) and $\log_{10} 1$ (= 0) or $-1 < \log_{10} 0.785 < 0$, hence $\log_{10} 0.785 = -1 + a$ decimal, therefore characteristic of $\log_{10} 0.785 = -1$

Similarly, numbers having one zero immediately after the decimal point always lie between 0.01 and 0.1, so logarithm of such a number say 0.095 will be between $\log_{10} 0.01$ ($= -2$ since $10^{-2} = \frac{1}{10^2} = \frac{1}{100} = 0.01$ or $10^{-2} = 0.01$ or $\log_{10} 0.01 = -2$) and $\log_{10} 0.1$ ($= -1$) or $-2 < \log_{10} 0.095 < -1$ or $\log_{10} 0.095 = -2 + a$ decimal. Therefore characteristic of $\log_{10} 0.095 = -2$.

Similary, characteristic of logarithm of numbers having two zeros immediately after the decimal point is '– 3'.

In brief we can say that characteristic of logarithm of a positive number less than one is negative and is one more than the number of zeros immediately after the decimal point.

Note: In finding logarithm we often use BAR form, for illustration :- $-1 + 0.75$ is written as $\bar{1}.75$ (read as Bar one point seven five), here it is to be noted that $\bar{1}.75 \neq -1.75$ because $-1.75 = -1 - 7.5$ but $\bar{1}.75 = -1 + .75$

How to Find Mantissa

Before proceeding further, it may be noted that– mantissa of logarithm of all numbers having the same digits in the same order is the same i.e. mantissa is independent of the position of the decimal point. Thus mantissa of $\log_{10} 3725 = \log_{10} 3.725 = \log_{10} 37.25 = \log_{10} 0.00\,3725$

The log table provides mantissa of common logarithm of any positive number. The main body gives mantissa of logarithm of numbers having 3 digits or less. While mean difference part of the table provides increment for fourth digit.

For illustration

(i) If we have to find mantissa of $\log_{10} 300$ then first consider 30 of 1st column and then 0 of 1st row, from table we see that 4771 corresponds to both 30 and 0 as it is right to 30 (of 1st column) and below 0 (of 1st row) so mantissa of $\log_{10} 300 = 0.4771$

(ii) Similarly mantissa of $\log_{10} 301 = 0.4786$ (it is right to 30 of 1st column) and below 1 of 1st row).

(iii) The mantissa of $\log_{10} 0.03$ is again 0.4771 as mentioned above.

(iv) To find mantissa of $\log_{10} 3751$, first consider 37 of 1st column and then 5 of 1st row, the value is 5740, now consider 37 of 1st column and 1 of 1st row (of mean difference part), the value is '1' so, add 5740 and 1, the result is '5741', which is the mantissa of $\log_{10} 3751$.

Now we can find logarithm of any number,

For illustration

(i) If we have to find $\log_{10} 3751$, then here characteristic is '3' and mantissa is '5741', therefore

$$\log_{10} 3751 = 3.5741$$

(ii) If we have to find $\log_{10} 37.51$, then here characteristic is '1' and mantissa is '5741' threrefore

$$\log_{10} 37.51 = 1.5741$$

(iii) If we have to find $\log_{10} 0.3751$, then here, characteristic is '-1' and mantissa is '5741' therefore

$$\log_{10} 0.3751 = -1 + 0.5741 = \bar{1}.5741$$

Antilogarithm and use of Antilogarithm

If $\log_{10} x = y$, then x is called antilogarithm of 'y' and written as Antilog $y = x$.

Therefore since $\log_{10} 37.51 = 1.5741$ so, Antilog $(1.5741) = 37.51$. It is used for conversion of logarithm of a number to that number.

How to find Antilogarithm

While finding antilogarithm we have to consider both characteristic and mantissa part separately. The characteristic is required for finding the position of decimal point. Mantissa is required to obtain the table value. For illustration.

(i) ***To obtain Antilog (1.5741)***

The characteristic is '1' and mantissa is '0.5741'. Now from antilogarithm table first consider '0.57' of 1st column and '4' of 1st row the value corresponding to both these is '3750'. Again consider '.57' of 1st column and 1 of 1st row (of mean difference part) the corresponding value is '1' so, now add both 3750 and 1, the result is 3751. Again as characteristic is '1' so there would be two digits left to the decimal point. Thus

$$\text{Antilog}(1.5741) = 37.51$$

(ii) ***To obtain Antilog (2.5741)***

The characteristic is '2' and mantissa is same as above so

$$\text{Antilog}(2.5741) = 375.1$$

(iii) ***To obtain Antilog (0.05741)***

The characteristic is '0' and mantissa is same as above so

$$\text{Antilog}(0.5741) = 3.751$$

(iv) ***To obtain Antilog ($\bar{1}.5741$)***

The characteristic is $\bar{1}$ or -1 and mantissa is same as above so

$$\text{Antilog}(\bar{1}.5741) = 0.3751$$

(v) ***Similarly Antilog ($\bar{2}.5741$) = 0.03751***

Logarithmic series:–

The series $\log_e(1+x) = x - \dfrac{x^2}{2} + \dfrac{x^3}{3} - \dfrac{x^4}{4} + \ldots$ is called logarithmic series where $-1 < x < 1$

Exponential series:–

The series

$$e^x = 1 + \frac{x}{\underline{|1}} + \frac{x^2}{\underline{|2}} + \ldots$$

is called exponential series

Important deductions

(1) $$\log_e(1-x) = (-x) - \frac{(-x)^2}{2} + \frac{(-x)^3}{3} - \ldots$$

$$= -x - \frac{x^2}{2} - \frac{x^3}{3} - ...$$

where $-1 < x < 1$

(2) $\log_e(1+x) + \log_e(1-x) = (x - \frac{x^2}{2} + \frac{x^2}{2} - ...) + (-x - \frac{x^2}{2} - \frac{x^3}{3} - ...)$

$$\Rightarrow \log_e(1+x) + \log_e(1-x) = (-2)(\frac{x^2}{2} + \frac{x^4}{4} + ...)$$

$$\Rightarrow \frac{\log(1-x^2)}{(-2)} = \frac{x^2}{2} + \frac{x^4}{4} + \frac{x^6}{6} + ...$$

$$\Rightarrow \left(-\frac{1}{2}\right)\log(1-x^2) = \frac{x^2}{2} + \frac{x^4}{4} + \frac{x^6}{6} + ...$$

$$\Rightarrow \log\frac{1}{\sqrt{1-x^2}} = \frac{x^2}{2} + \frac{x^4}{4} + \frac{x^6}{6} + ...$$

(iii) $\log_e(1+x) - \log_e(1-x) = (x - \frac{x^2}{2} + \frac{x^3}{3} + \frac{x^4}{4} + ...)$

$$-(-x - \frac{x^2}{2} - \frac{x^3}{3} - \frac{x^4}{4} - ...)$$

$$\Rightarrow \log_e\left(\frac{1+x}{1-x}\right) = 2(x + \frac{x^3}{3} + \frac{x^5}{5} + ...)$$

$$\Rightarrow \frac{1}{2}\log_e\left(\frac{1+x}{1-x}\right) = x + \frac{x^3}{3} + \frac{x^5}{5} + ...$$

$$\Rightarrow \log_e\sqrt{\frac{1+x}{1-x}} = x + \frac{x^3}{3} + \frac{x^5}{5} + ...$$

(iv) $e^{-x} = 1 + \frac{(-x)}{\underline{|1}} + \frac{(-x)^2}{\underline{|2}} + \frac{(-x)^3}{\underline{|3}} + ...$

$$= 1 - \frac{x}{\underline{|1}} + \frac{x^2}{\underline{|2}} - \frac{x^3}{\underline{|3}} + ...$$

(v) $e^x + e^{-x} = (1 + \frac{x}{\underline{|1}} + \frac{x^2}{\underline{|2}} + ...) + (1 - \frac{x}{\underline{|1}} + \frac{x^2}{\underline{|2}} - ......)$

$\Rightarrow \qquad e^x + e^{-x} = 2\left(1 + \frac{x^2}{\lfloor 2} + \frac{x^4}{\lfloor 4} + \ldots\right)$

$\Rightarrow \qquad \frac{e^x + e^{-x}}{2} = 1 + \frac{x^2}{\lfloor 2} + \frac{x^4}{\lfloor 4} + \ldots$

(vi) $\qquad e^x - e^{-x} = \left(1 + \frac{x}{\lfloor 1} + \frac{x^2}{\lfloor 2} + \frac{x^3}{\lfloor 3} + \ldots\right) - \left(1 - \frac{x}{\lfloor 1} + \frac{x^2}{\lfloor 2} - \frac{x^3}{\lfloor 3}\right.$
$\left. + \ldots\right)$

$\Rightarrow \qquad e^x - e^{-x} = 2\left(x + \frac{x^3}{\lfloor 3} + \frac{x^5}{\lfloor 5} + \ldots\right)$

$\Rightarrow \qquad \frac{e^x - e^{-x}}{2} = \left(x + \frac{x^3}{\lfloor 3} + \frac{x^5}{\lfloor 5} + \ldots\right)$

SOLVED EXAMPLES

5.1 Prove that

$$3 + \frac{5}{\lfloor 1} + \frac{7}{\lfloor 2} + \frac{9}{\lfloor 3} + \ldots = 5e$$

Solution: Here $t_n = \frac{3 + (n-1)2}{\lfloor n-1}$

$$= \frac{3 + 2n - 2}{\lfloor n-1} = \frac{2n+1}{\lfloor n-1}$$

$$= \frac{2n}{\lfloor n-1} + \frac{1}{\lfloor n-1} = \frac{2(n-1+1)}{\lfloor n-1} + \frac{1}{\lfloor n-1}$$

$$= \frac{2(n-1)}{\lfloor n-1} + \frac{2}{\lfloor n-1} + \frac{1}{\lfloor n-1}$$

$$= \frac{2}{\lfloor n-2} + \frac{2}{\lfloor n-1} + \frac{1}{\lfloor n-1}$$

Now putting $n = 1, 2, 3\ldots$

L.H.S. $= t_1 + t_2 + t_3 + \ldots$

$$= 2\left(\frac{1}{\lfloor 0} + \frac{1}{\lfloor 1} + \frac{1}{\lfloor 2} + \ldots\right) + 2\left(\frac{1}{\lfloor 0} + \frac{1}{\lfloor 1} + \frac{1}{\lfloor 2} + \ldots\right) + \left(\frac{1}{\lfloor 0} + \frac{1}{\lfloor 1} + \ldots\right)$$

$= 2.e^1 + 2.e^1 + e^1$

$= 5\,e =$ R.H.S.

5.2. Prove that

$$\frac{1}{\lfloor 1} + \frac{1+3}{\lfloor 2} x + \frac{1+3+5}{\lfloor 3} x^2 + \ldots = (x+1)e^x$$

Solution: Here $t_n = \dfrac{1+3+5+\ldots+(2n-1)}{\lfloor n} x^{n-1}$

$$= \frac{\frac{n}{2}\{2.1+(n-1).2\}}{\lfloor n} . x^{n-1} = \frac{n^2}{\lfloor n} . x^{n-1}$$

$$= \frac{\{n(n-1)+n\}}{\lfloor n} x^{n-1} = \left(\frac{n(n-1)}{\lfloor n} + \frac{n}{\lfloor n}\right) x^{n-1}$$

$$= \left(\frac{1}{\lfloor n-2} + \frac{1}{\lfloor n-1}\right) x^{n-1}$$

$\therefore \quad t_1 = 0 + 1$

$\therefore \quad t_2 = x + \dfrac{x}{\lfloor 1} \qquad \therefore \quad t_3 = \dfrac{x^2}{\lfloor 1} + \dfrac{x^2}{\lfloor 2}$

$\therefore \quad t_4 = \dfrac{x^3}{\lfloor 2} + \dfrac{x^3}{\lfloor 3}$

Now adding all these we get

L.H.S. $= t_1 + t_2 + t_3 + \ldots$

$$= (0+1) + (x + \frac{x}{\lfloor 1}) + \left(\frac{x^2}{\lfloor 1} + \frac{x^2}{\lfloor 2}\right) + \left(\frac{x^3}{\lfloor 2} + \frac{x^3}{\lfloor 3}\right) + \ldots$$

$$= (1 + \frac{x}{\lfloor 1} + \frac{x^2}{\lfloor 1} + \ldots) + x\,(1 + \frac{x}{\lfloor 1} + \frac{x^2}{\lfloor 2} + \ldots)$$

$$= e^x + x.e^x = (1+x)\,e^x = \text{R.H.S}$$

5.3 Find the co-efficient of x^n in e^{e^x}

Solution: $\quad e^{e^x} = 1 + \dfrac{e^x}{\lfloor 1} + \dfrac{(e^x)^2}{\lfloor 2} + \dfrac{(e^x)^3}{\lfloor 3} + \ldots$

$$= 1 + \frac{e^x}{\lfloor 1} + \frac{e^{2x}}{\lfloor 2} + \frac{e^{3x}}{\lfloor 3} + \ldots$$

$$= 1 + \frac{1}{\lfloor 1}(1 + \frac{x}{\lfloor 1} + \frac{x^2}{\lfloor 2} + \ldots + \frac{x^n}{\lfloor n} + \ldots) + \frac{1}{\lfloor 2}(1 + \frac{2x}{\lfloor 1} + \frac{2^2x^2}{\lfloor 2}$$

$+ \ldots$

$$+ \frac{2^n x^n}{\lfloor n} + \ldots) + \frac{1}{\lfloor 3}(1 + \frac{3.x}{\lfloor 1} + \frac{3^2 x^2}{\lfloor 2} + \ldots + \frac{3^n x^n}{\lfloor n}$$

So co-efficient of x^n is $\frac{1}{\lfloor 1} \cdot \frac{1}{\lfloor n} + \frac{2^n}{\lfloor n \lfloor 2} + \frac{3^n}{\lfloor n \lfloor 3} + \frac{4^n}{\lfloor n \lfloor 4} + \ldots$

$$= \frac{1}{\lfloor n}\left(\frac{1^n}{\lfloor 1} + \frac{2^n}{\lfloor 2} + \frac{3^n}{\lfloor 3} + \ldots\right)$$

5.4. Find the co-efficient of x^r in

$$\frac{(x+1)}{\lfloor 1} + \frac{(x+1)}{\lfloor 1} + \frac{(x+1)^3}{\lfloor 3} + \ldots$$

Solution: $\frac{(x+1)}{\lfloor 1} + \frac{(x+1)^2}{\lfloor 2} + \frac{(x+1)^3}{\lfloor 3} + \ldots$

$$= 1 + \frac{(x+1)}{\lfloor 1} + \frac{(x+1)^2}{\lfloor 2} + \frac{(x+1)^3}{\lfloor 3} + \ldots - 1$$

$$= e^{(x+1)} - 1 = e.e^x - 1$$

$$= e\,(1 + \frac{x}{\lfloor 1} + \frac{x^2}{\lfloor 2} + \ldots + \frac{x^r}{\lfloor r} + \ldots + \ldots) - 1$$

$\therefore$ Co-efficient of x^r is $\frac{e}{\lfloor r}$

5.5. Write down the expansion of

(i) $-\log_e (1-x)$ **[AHSEC, 01]**

(ii) $\sqrt{e}$ **[AHSEC, 01]**

Solution: $-\log_e (1-x)$

$$= -(-x - \frac{x^2}{2} - \frac{x^2}{2} - \frac{x^4}{4} - \ldots)$$

$$= x + \frac{x^2}{2} + \frac{x^3}{3} + \frac{x^4}{4} + \ldots$$

(ii) $\sqrt{e} \quad = e^{\frac{1}{2}}$

$$= 1 + \frac{\frac{1}{2}}{\underline{|1}} + \frac{\left(\frac{1}{2}\right)^2}{\underline{|2}} + \frac{\left(\frac{1}{2}\right)^3}{\underline{|3}} + \ldots$$

$$= 1 + \frac{1}{2\underline{|1}} + \frac{1}{4\underline{|2}} + \frac{1}{8\underline{|3}} + \ldots$$

5.6. Find the value of $\sum_{i=1}^{\alpha} \frac{(-1)^i}{i}$ **[AHSEC, 98]**

Solution: $\sum_{i=1}^{\alpha} \frac{(-1)^i}{i}$

$$= (-1) + \frac{(-1)^2}{2} + \frac{(-1)^3}{3} + \frac{(-1)^4}{4} + \ldots$$

$$= -1 + \frac{1}{2} - \frac{1}{3} + \frac{1}{4} - \frac{1}{5} + \ldots$$

$$= -\left(1 - \frac{1}{5} + \frac{1}{3} - \frac{1}{4} + \frac{1}{5} - \ldots\right)$$

$$= -\{\log_e (1 + 1)\}$$

$$= -\log_e 2$$

5.7. If $y = x - \frac{x^2}{2} + \frac{x^3}{3} - \frac{x^4}{4} + \ldots$

then prove that

$$x = y + \frac{y^2}{\underline{|2}} + \frac{y^3}{\underline{|3}} + \ldots$$

Solution: $y = x - \frac{x^2}{2} + \frac{x^3}{3} - \ldots$

$$\Rightarrow \quad y = \log_e (1 + x)$$

$$\Rightarrow \quad e^y = 1 + x$$

$$\Rightarrow 1 + \frac{y}{\underline{|1}} + \frac{y^2}{\underline{|2}} + \ldots = 1 + x$$

$$\Rightarrow \quad x = \frac{y}{\underline{|1}} + \frac{y^2}{\underline{|2}} + \frac{y^3}{\underline{|3}} + \ldots$$

5.8. If $\frac{1}{x}, \frac{1}{y}, \frac{1}{z}$, are in A.P. prove that

$\log(x+z) + \log(x+z-2y) = 2\log(x-z)$

Solution: $\therefore \frac{1}{x}, \frac{1}{y}, \frac{1}{z}$ are in A.P.

$$\therefore \quad \frac{1}{x} - \frac{1}{y} = \frac{1}{y} - \frac{1}{z}$$

$$\Rightarrow \quad \frac{2}{y} = \frac{1}{x} + \frac{1}{z}$$

$$\Rightarrow \quad \frac{2}{y} = \frac{x+z}{xz}$$

$$\Rightarrow \quad \frac{2xz}{x+z} = y \rightarrow \quad (1)$$

Now, $\log(x+z) + \log(x+z-2y)$

$= \log(x+z) + \log(x+z \frac{4xz}{x+z})$

$= \log(x+z) + \log \frac{(x+z)^2 - 4xz}{x+z}$

$= \log(x+z) + \log(x-z)^2 - \log(x+z)$

$= 2\log(x-z)$

5.9. Write down the expansion of e^{-ax} **[AHSEC, 07]**

Solution: e^{-ax}

$$= 1 + \frac{(-ax)}{\underline{|1}} + \frac{(-ax)^2}{\underline{|2}} + \frac{(-ax)^3}{\underline{|3}} +$$

$$= 1 - \frac{ax}{\underline{|1}} + \frac{a^2x^2}{\underline{|2}} - \frac{a^3x^3}{\underline{|3}} +$$

5.10. Find the value of $10^{-2\log_{10}3}$ **[AHSEC, 08]**

Solution:let $10^{-2\log}10^3 = x$

$$\Rightarrow \quad \log_{10} x = -2\log_{10}3$$

$$\Rightarrow \quad \log_{10}x = \log_{10}3$$

$$\Rightarrow \quad x = 3^{-2}$$

$$\Rightarrow \quad x = \frac{1}{3^2}$$

$$\Rightarrow \quad x = \frac{1}{9}$$

5.11. Write down the expansion of e^{-1}. **[AHSEC, 08]**

Solution: e^{-1}

$$= 1 + \frac{(-1)}{\lfloor 1} + \frac{(-1)^2}{\lfloor 2} + \frac{(-1)^3}{\lfloor 3} +$$

$$= 1 - \frac{1}{\lfloor 1} + \frac{1}{\lfloor 2} + \frac{1)^3}{\lfloor 3} +$$

5.12. Prove that

$$\left(\frac{a-b}{a}\right) + \frac{1}{2}\left(\frac{a-b}{a}\right)^2 + \left(\frac{a-b}{a}\right)^3 + \ \ldots$$

$$= \log_e a - \log_e b$$ **[AHSEC, 08]**

Solution:

$$\textbf{LHS} \quad = \left(\frac{a-b}{a}\right) + \frac{1}{2}\left(\frac{a-b}{a}\right)^2 + \left(\frac{a-b}{a}\right)^3 +$$

$$= -\left\{-\left(\frac{a-b}{a}\right) - \frac{1}{2}\left(\frac{a-b}{a}\right)^2 - \frac{1}{3}\left(\frac{a-b}{a}\right)^3 + \ldots\right\}$$

$$= -\left[\log_e\left(1-\frac{a-b}{a}\right)\right]$$

$$= -\log_e\left(\frac{a-a+b}{a}\right)$$

$$= -\log_e\frac{b}{a}$$

$$= \log_e\left(\frac{b}{a}\right)^{-1}$$

$$= \log_e\left(\frac{a}{b}\right)$$

$= \log_e a - \log_e b$

= R.H.S., proved.

5.13. Obtain the sum

$$\frac{1}{\lfloor 2}-\frac{1}{\lfloor 3}+\frac{1}{\lfloor 4}-\frac{1}{\lfloor 5}+$$ **[AHSEC, 08]**

Solution:

$$\frac{1}{\lfloor 2}-\frac{1}{\lfloor 3}+\frac{1}{\lfloor 4}-\frac{1}{\lfloor 5}+$$

$$= 1-\frac{1}{\lfloor 1}+\frac{1}{\lfloor 2}-\frac{1}{\lfloor 3}+\frac{1}{\lfloor 4}-\frac{1}{\lfloor 5}+$$

$$= 1-\frac{(-1)}{\lfloor 1}+\frac{(-1)^2}{\lfloor 2}-\frac{(-1)^3}{\lfloor 3}+\frac{(-1)^4}{\lfloor 4}-\frac{(-1)^5}{\lfloor 5}+$$

$= e^{-1}$ (a, b, c, d, e, f)

EXERCISE–5

1. Prove that $\log_e(1 + 3x + 2x^2)$

$$= 3x - \frac{5x^2}{2} + \frac{9x^3}{3} - \frac{17x^4}{4} + \ldots$$

2. Prove that

$$\frac{1}{n^2} + \frac{1}{2n^4} + \frac{1}{3n^6} + \ldots$$
$$= 2\log_e n - \log_e(n+1) - \log_e(n-1)$$

3. Prove that

$$1 + \frac{2^3}{\underline{|2}} + \frac{3^3}{\underline{|3}} + \frac{4^3}{\underline{|4}} + \ldots = 5.e$$

4. Prove that

$$\frac{1 + \frac{2^2}{\underline{|2}} + \frac{2^4}{\underline{|3}} + \frac{2^6}{\underline{|4}} + \ldots}{1 + \frac{1}{\underline{|2}} + \frac{2}{\underline{|3}} + \frac{2^2}{\underline{|4}} + \ldots} = e^2 - 1$$

5. If $y = \dfrac{x}{1+x}$; $0 < x < 1$

Prove that $x - \dfrac{x^2}{2} + \dfrac{x^3}{3} - \ldots = y + \dfrac{y^2}{2} + \dfrac{y^3}{3} + \ldots\ldots.$

6. Find the co-efficient of x^r in $\log_e(1 + x + x^2)$

(Hint : $\log_e(1 + x + x^2)$

$$= \log_e\left(\frac{1-x^3}{1-x}\right)$$ (**Ans.** $\frac{1}{r}$)

7. Show that

$$(1 + \frac{x^2}{\underline{|2}} + \frac{x^4}{\underline{|4}} + \ldots)^2 - (x + \frac{x^3}{\underline{|3}} + \frac{x^5}{\underline{|5}} + \ldots)^2 = 1$$

8. Prove that

$$n + \frac{1}{n} = 2\left[1 + \frac{(\log_e n)^2}{\underline{|2}} + \frac{(\log_e n)^4}{\underline{|4}} + \ldots\ldots\right]$$

9. Show that

$$\log_e \frac{1+x}{1-x} = 2\left(x + \frac{x^3}{3} + \frac{x^5}{5} + \ldots\right)$$ **(AHSEC, 02)**

10. Prove that

$$\frac{e^1 + e^{-1}}{2} = 1 + \frac{1}{\underline{|2}} + \frac{1}{\underline{|4}} + \ldots$$ **(AHSEC, 96)**

6

Sets

The famous German mathematician G. Cantor (1845-1918) first originated the idea of set theory and he gave a mathematical shape to the set theory. Cantor said—a set is a collection of definite and distinct objects called elements or members or points of the set. By the word definite it is tried to mean that for a given specific set and specific object it must be possible to verify if that object is a member of that set or not. By the word distinct, it is tried to mean that no two members of a set are same. For denoting the sets usually the capital letters of English alphabets are used.

If each element of a set is a set itself then the set is called set of sets or class of sets.

There are two ways of representation of sets. These are—

(i) Tabular or roster method

(ii) Set builder or rule method

According to the first method, all the elements are listed within brackets e.g., set of major rivers of Assam can be written as—

$R = \{$Brahmaputra, Borak$\}$

Similarly the set of state capitals of north-east

$C = \{$Guwahati, Kohima, Imphal, Itanagar, Shillong, Aizwal, Agartala$\}$

Again by the second method, x is used to represent the elements of the set which satisfy a definite property e.g.,

Set of vowels, $V = \{x : x$ is a vowel$\}$

Set of even numbers less than 10,

$B = \{x : x$ is an even number less than $10\}$

Null and Singleton set

The set containing no element is called a null set. On the other hand set containing only one element is called singleton set. A null set is usually denoted by ϕ. Example of a null set is set of even numbers (integer) between 2 and 4.

Subset and Superset

If each element of a set A is an element of another set B. Then set A is called a subset of B and denoted by $A \subseteq B$. On the other hand the set B is called superset of A and denoted as $B \supseteq A$.

Again, if all the elements of B are the elements of A, then A is called improper subset of B. On the other hand if there exists atleast one element in B which is not an element of A then A is called a proper subset of B.

Universal Set

The largest possible set in a particular study is called universal set.

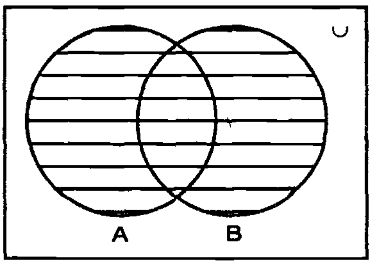
Venn Diagram

For the representation of sets diagrammatically geometric shapes—rectangle and circle are used. These diagrams are called venn diagrams.

Operations on Set

(a) *Union of Sets*

The union of two set A and B and denoted by $A \cup B$ is the set of all elements which belong to either A or B or both. In venn diagram the shaded area is $A \cup B$.

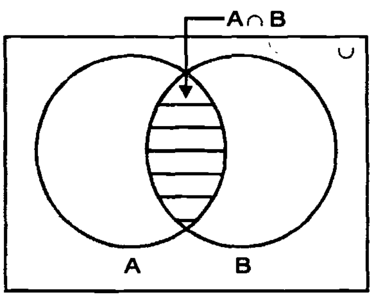


A∪B

For numerical illustration, let $A = \{1, 2, 3\}$, $B = \{3, 4, 5\}$ then $A \cup B = \{1, 2, 3, 4, 5\}$

(b) *Intersection of Sets*

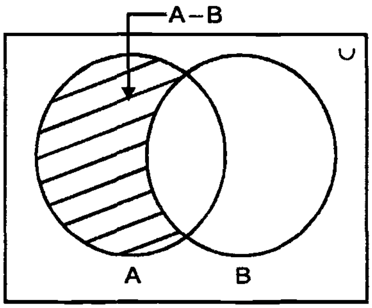
The set of all elements which are present in both A and B is called intersection and denoted by $A \cap$ B. In venn diagram—



For numerical illustration—$A = \{1, 2, 3, 4\}$, $B = \{3, 4, 5, 6\}$ then $A \cap B = \{3, 4\}$

(c) *Difference of Sets*

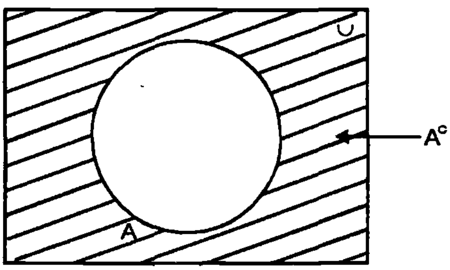
The difference of two sets A and B denoted by $A - B$ is the set of all those elements of A which are not elements of B and denoted by $A - B$. In venn diagram—



For numerical illustration—$A = \{1, 2, 3, 4\}$, B = $\{3, 4, 5, 6\}$ then $A - B = \{1, 2)$.

(d) *Complement*

Complement of a set A, denoted by A^C or A', is the set of all points which are elements of universal set but not in A.



For numerical illustration : if $u = \{1, 2, 3, 4, 5, 6, 7, 8, 9\}$ and $A = \{1, 2, 3, 4\}$ then $A^C = \{5, 6, 7, 8, 9\}$.

Some Important Laws

(1) *Idempotent laws*
(i) $A \cup A = A$ (ii) $A \cap A = A$

(2) *Associative laws*
(i) $(A \cup B) \cup C = A \cup (B \cup C)$ (ii) $(A \cap B) \cap C = A \cap (B \cap C)$

(3) *Commutative laws*

(i) $A \cup B = B \cup A$ (ii) $A \cap B = B \cap A$

(4) Distributive laws

(i) $A \cup (B \cap C) = (A \cup B) \cap (A \cup C)$ (ii) $A \cap (B \cup C) = (A \cap B) \cup (A \cap C)$

(5) *Identity laws*

(i) $A \cup \phi = A$ (ii) $A \cap \phi = \phi$ (iii) $A \cup U = U$ (iv) $A \cap U = A$

(6) *Complement laws*

(i) $A \cup A^C = \text{U}$ (ii) $A \cap A^C = \phi$ (iii) $(A^C)^C = A$ (iv) $U^C = \phi$, $\phi^C = U$

(7) De Morgan's law

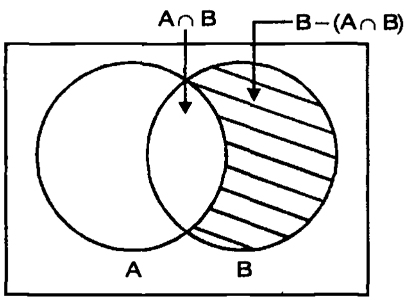
(i) $(A \cup B)^C = A^C \cap B^C$ (ii) $(A \cap B)^C = A^C \cup B^C$

Number of elements in a set

In a finite set 'A' the number of elements in the set is denoted by $n(A)$.

Again if A and B be any two sets then

$n(A \cup B) = n(A) + n(B) - n(A \cap B)$, this can be proved by venn diagram as—



From the diagram it is clear that—

$$A \cup B = A \cup \{B - (B \cap A)\}$$

Hence $n(A \cup B) = n[A \cup \{B - (B \cap A)\}]$

$$= n(A) + n(B) - n(A \cap B)$$

But if A and B are null sets then

$$A \cap B = \phi$$

and $n(A \cup B) = n(A) + n(B)$ $\quad \because n(\phi) = 0$

For any three events A, B, C

$$n(A \cup B \cup C) = n(A) + n(B) + n(C)$$
$$-n(A \cap B) - n(A \cap C) - n(B \cap C) + n(A \cap B \cap C)$$

This can be proved as—

$$A \cup B \cup C = A \cup (B \cup C)$$

Now

$$n(A \cup B \cup C) = n\{A \cup (B \cup C)\}$$
$$= n(A) + n(B \cup C) - n(A \cap (B \cup C)\}$$
$$= n(A) + n(B) + n(C) - n(B \cap C) - n\{(A \cap B) \cup (A \cap C)\}$$

(By distributive laws)

$$= n(A) + n(B) + n(C) - n(B \cap C)$$
$$- \{n(A \cap B) + n(A \cap C) - n(A \cap B) \cap (A \cap C)\}$$
$$= n(A) + n(B) + n(C) - n(B \cap C)$$
$$- n(A \cap B) - n(A \cap C) + n(A \cap B \cap C)$$
$$\because (A \cap B) \cap (A \cap C) = A \cap B \cap C$$

Again if A, B, C are disjoint then

$$n(A \cup B \cup C) = n(A) + n(B) + n(C)$$

SOLVED EXAMPLES

6.1. In a class of 30 students, 10 students have taken English but not French, 15 students have taken English. Find the number of students who have taken (1) French, (2) French but not English.

Solution:

(1) Let the number of students who have taken English be $n(E)$.

Number of students who have taken French $n(F)$.

Now according to question—

$n(E \cup F) = 30,\ n(E) = 15,\ n(E \cap F^C) = 10$

but $n(E \cap F^C) = 10$

$\Rightarrow n\{(E \cap (U - F)\} = 10$

$\Rightarrow n\{(E \cap U) - (E \cap F)\} = 10$

$\Rightarrow n(E) - n(E \cap F) = 10$

$\Rightarrow n(E \cap F) = 5$

We have $n(E \cup F) = n(E) + n(F) - n(E \cap F)$

$\Rightarrow 30 = 15 + n(F) - 5$

$\therefore$ Number of students taking French = 20.

(2) Again number of students taking French but not English is n$\{F \cap E^C\}$

$$= n\{F \cap (U - E)\}$$
$$= n(F \cap U) - n(F \cap E)$$
$$= n(F) - n(\mathrm{F} \cap E)$$
$$= 20 - 5$$
$$= 15$$

6.2. Out of 440 boys, in a college, 112 boys read German, 120 read French and 168 Spanish. Of these 32 read French and Spanish, 40 read German and Spanish, 20 read German and French while 12 read all the three languages. How many boys did not read any language?

Solution:

Number of students who read German be $n(G)$.

Number of students who read French be $n(F)$.

Number of students who read Spanish be $n(S)$.

According to question—

$$n(G) = 112,\ n(F) = 120,\ n(S) = 168,$$
$$n(F \cap S) = 32,\ n(G \cap S) = 40,\ n(G \cap F) = 20$$
$$n(F \cap S \cap G) = 12$$

Now number of boys who did not study any be

$$\begin{aligned} n(F \cup S \cup G)^C &= n\{\cup - (F \cup S \cup G)\} \\ &= n(\cup) - n(F \cup S \cup G) \\ &= n(\cup) - n(F) - n(S) - n(G) + n(F \cap S) \\ &\quad + n(S \cap G) + n(F \cap G) - n(F \cap S \cap G) \\ &= 440 - 120 - 168 - 112 + 32 + 40 + 20 - 12 \\ &= 120 \end{aligned}$$

6.3. In a class test of 45 students 23 students passed in 1st paper, 15 passed in 1st paper but not in 2nd, then find the number of students who passed in both and who passed in 2nd but not in 1st.

Solution:

Let the number of students who passed in 1st paper be n (F).

Number of students who passed in 2nd paper be n (S).

Now according to question—

$$n(F) = 23$$
$$n(F \cap S^C) = 15$$
$$n(U) = 45$$

Again $\quad n(F \cap S^C) = 15$

$\Rightarrow \quad n\{F \cap (U - S)\} = 15$

$\Rightarrow \quad n\{(F \cap U) - (F \cap S)\} = 15$

$\Rightarrow \quad n(F \cap U) - n(F \cap S) = 15$

$\Rightarrow \quad n(F) - n(F \cap S) = 15$

$\Rightarrow \quad n(F \cap S) = 8$

Therefore number of students who passed in both is '8'.

Now number of students who passed in 2nd but not in 1st is—

$$\begin{aligned} n(S \cap F^C) &= n\{S \cap (U - F)\} \\ &= n\{(S \cap U) - (S \cap F)\} \\ &= n(S \cap U) - n(S \cap F) = n(S) - n(S \cap F) \end{aligned} \quad \text{...(1)}$$

Again $\quad n(S \cup F) = n(U)$

$$\Rightarrow \quad n(S) + n(F) - n(S \cap F) = n(U)$$
$$\Rightarrow \quad n(S) - n(S \cap F) = n(U) - n(F)$$
$$\Rightarrow \quad n(S) - n(S \cap F) = 45 - 23$$
$$\Rightarrow \quad n(S) - n(S \cap F) = 22 \qquad ...(2)$$

$\therefore$ From (1), (2);

$$n(S \cap F^C) = 22$$

6.4. In a city, three daily newspapers A, B, C are published. 42% of the people read A, 51% read B, 68% read C, 30% read A and B, 28% read B and C, 36% read A and C, 8% don't read any of the three newspapers, find the percentage of persons who read all the three papers.

Solution:

Let the percentage of people who read A be $n(A)$.

Who read B be $n(B)$.

Who read C be $n(C)$.

Now according to question—

$$n(A) = 42$$
$$n(B) = 51$$
$$n(C) = 68$$
$$n(A \cap B) = 30$$
$$n(B \cap C) = 28$$
$$n(A \cap C) = 36$$
$$n(A \cup B \cup C)' = 8$$

But $\quad n(A \cup B \cup C)' = 8$

$$\Rightarrow n\{U - (A \cup B \cup C)\} = 8$$
$$\Rightarrow n(U) - n(A \cup B \cup C) = 8$$
$$\Rightarrow \quad 100 - 8 = n(A \cup B \cup C)$$
$$\Rightarrow \quad 92 = n(A \cup B \cup C)$$

Again, percentage of persons who read all the three newspapers would be

$$n(A \cap \mathrm{B} \cap \mathrm{C}) = n(\mathrm{A} \cup \mathrm{B} \cup \mathrm{C}) - n(A) - n(B) - n(C) + n(A \cap B) + n(A \cap C) + n(B \cap C)$$
$$\Rightarrow \quad n(A \cap B \cap C) = 92 - 42 - 51 - 68 + 30 + 28 + 36$$
$$\Rightarrow \quad n(A \cap B \cap C) = 92 - 67 = 25$$

6.5. State true or false with reasons:

(i) $\{1\} \in \{1, 2, 3\}$ (ii) $1 \in \{1, 2, 3\}$ (iii) $\{1\} \subset \{1, 2, 3\}$

Solution:

(i) False, because the set $\{1\}$ is not an element of the set $\{1, 2, 3\}$.

(ii) True, because 1 is an element of the set $[1, 2, 3\}$.

(iii) True, because the set $\{1\}$ is a proper subset of $\{1, 2, 3\}$.

EXERCISE–6

1. What is the difference between ϕ and $\{\phi\}$?

   (Ans. ϕ, a null set; $\{\phi\}$ a singleton set)

2. Verify with numerical example:

   (i) $(A \cup B)' = A' \cap B'$

   (ii) $(A \cap B)' = A' \cup B'$

3. If $A = \{0, 1, 2, 3, 4, 5\}$, $B = \{2, 3, 4, 5\}$, $C = \{4, 5, 6, 7\}$ find $A \cap (B \cup C)$, $(A \cap B) \cup (A \cap C)$. **(Ans. $\{2, 3, 4, 5\}$, $\{2, 3, 4, 5\}$)**

4. State true or false:

   (i) $\phi \in \{1, 2, 3\}$ (ii) $\{2, 4, 5\} \subset \{5, 4, 2\}$ **(Ans. (i) False (ii) False)**

5. In a survey of 320 persons, number of persons taking tea is 210, milk is 100 and coffee is 70, tea and milk is 50, milk and coffee is 30 and tea and coffee is 50. Also number of persons taking all is 20. Find the number of persons who take neither of these. **(Ans. 50)**

7

DIFFERENTIAL CALCULAS

PRELIMINARIES

Constant

A quantity which retains the same value in a mathematical study is called a constant.

Variable

A symbol which assumes different values from a given set is called a variable.

Discrete Variable

A variable which can attain only some values from a set (given) is called a discrete variable.

Continuous Variable

A variable which can attain any value from a given set is called a continuous variable.
(Marks attained by examinnes in any examination are discrete whereas height of the examinnees are continuous variable)

Functions as a Mapping

Given two non empty sets (distinct or not) of real numbers, if there exists a rule 'f' by which each element x in A is associated with a unique element y in B, then f is called a function (real valued) from A to B which is denoted by $f : A \to B$, $y = f(x)$.

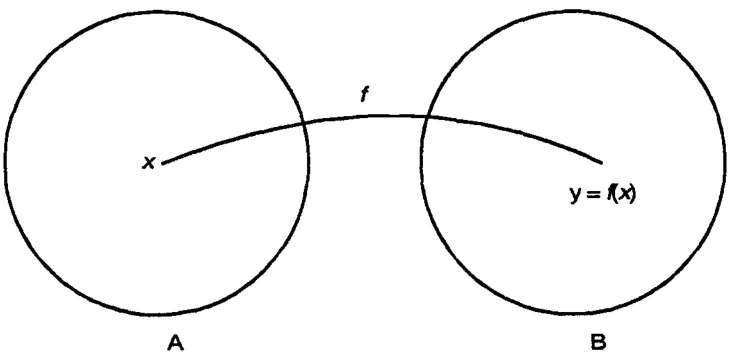


Fig. 7.1

The set A is called domain of f and set B is called co-domain of f.

$y = f(x)$ is called the image of x and x is called the preimage of y under f, the set of all images of x (in B) is called range of f. Thus, range is a subset of co-domain.

Graph

Graphs of a function 'f' is the geometrical representation of the equation $y = f(x)$. The definition of function considers thus a set of ordered pair of numbers (x, y). The set of points corresponding to these ordered pairs is called a graph. Thus a graph can completely explain or identify the function.

Limit

Let 'x' be a variable in a function $f(x)$ and 'a' be a constant. If the function gets nearer and nearer to some quantity l as x gets nearer and nearer to a, we say, limit of $f(x)$ as x approaches 'a' is l and write it as $\lim\limits_{x \to a} f(x) = l$.

Left Hand Limit

If x approaches 'a' by taking values less than 'a', we say that x approches a from left and write it as $x \to a^-$ and the corresponding limit is called left hand limit and write it as

$$\lim_{x \to a^-} f(x)$$

Right Hand Limit

If x approaches 'a' by taking values greater than 'a', we say that x approaches a from the right and write it as $x \to a^+$ and the corresponding limit is called right hand limit and write it as

$$\lim_{x \to a^+} f(x)$$

For illustration, let $f(x) = x^3$ and $a = 2$. The numerical approach to guess $\lim\limits_{x \to 2} f(x)$ is as following:

x	$f(x)$	x	$f(x)$
1.5	3.375	2.5	15.625
1.9	6.859	2.1	9.261
1.99	7.880	2.01	8.1206
1.999	7.988	2.001	8.012
1.9999	7.9988	2.0001	8.00012

We now infer that "for x sufficiently close" to 2, x^3 can be made sufficiently close to 8. That is

$$\lim_{x \to 2} x^3 = 8$$

Theorems on limit (wtihout proof):

(1) $\lim_{x \to a}[f(x) + g(x)] = \lim_{x \to a} f(x) + \lim_{x \to a} g(x)$

(2) $\lim_{x \to a}[f(x) - g(x)] = \lim_{x \to a} f(x) - \lim_{x \to a} g(x)$

(3) $\lim_{x \to a}[k.f(x)] = k \lim_{x \to a} f(x)$ (k is a constants)

(4) $\lim_{x \to a} \dfrac{f(x)}{g(x)} = \dfrac{\lim_{x \to a} f(x)}{\lim_{x \to a} g(x)}$ $(\lim_{x \to a} g(x)\, 0)$

Continuity

Let f be a function and 'a' be a point in the domain of f. The function f is said to be continuous at 'a' if the limit of the function at 'a' exists and is same as the value of the function at 'a'. In otherwords a function 'f' is said to be continuous at a point of its domain if

$$\lim_{x \to a} f(x) = f(a)$$

or $\lim_{x \to a^+} f(x) = \lim_{x \to a^-} f(x) = f(a)$

Properties of continuous function: If f and g are two continuous functions then

(i) $f + g$ is continuous

(ii) $f - g$ is continuous

(iii) $f.g$ is continuous

(iv) $\dfrac{f}{g}$ is continuous

(v) $c.f$ is continuous where c is any constant number.

Again when we say that a real function is continuous (without mentioning where) we mean that it is continuous at every point of its domain.

Differentiability

Let $f(x)$ be a real function, when x increases to $x + \Delta x$, $f(x)$ increases to $f(x + \Delta x)$. Then derivative of function $f(x)$ with respect to x is defined as

$$f'(x) = \frac{df(x)}{dx} = \lim_{\Delta x \to 0} \frac{f(x+\Delta x) - f(x)}{(x+\Delta x) - x}$$

$$= \lim_{\Delta x \to 0} \frac{f(x+\Delta x) - f(x)}{\Delta x},$$

provided the limit exist. However if the function is expressed by y i.e. $y = f(x)$, and Δy is the increase in value of y corresponding to an increase Δx in value of x then the derivative of y w.r.t. x is defined as

$$\frac{dy}{dx} = \lim_{\Delta x \to 0} \frac{\Delta y}{\Delta x}$$

Theorems on Derivative

(1) $\frac{d}{dx}\{f(x) + g(x)\} = \frac{d}{dx} f(x) + \frac{d}{dx} g(x)$

proof: $\frac{d}{dx}\{f(x) + g(x)\}$

$$= \lim_{\Delta x \to 0} \frac{[\{f(x+\Delta x) + g(x+\Delta x)\} - \{f(x) + g(x)\}]}{\Delta x}$$

$$= \lim_{\Delta x \to 0} \frac{[\{f(x+\Delta x) + f(x)\} - \{g(x+\Delta x) - g(x)\}]}{\Delta x}$$

$$= \lim_{\Delta x \to 0} \left\{ \frac{f(x+\Delta x) - f(x)}{\Delta x} + \frac{g(x+\Delta x) - g(x)}{\Delta x} \right\}$$

$$= \lim_{\Delta x \to 0} \frac{f(x+\Delta x) - f(x)}{\Delta x} + \lim_{\Delta x \to 0} \frac{g(x+\Delta x) - g(x)}{\Delta x}$$

$$= \frac{d}{dx} f(x) + \frac{d}{dx} g(x)$$

(2) $\frac{d}{dx}\{k.f(x)\} = k.\frac{d}{dx} f(x)$ where k is any constant.

proof: $\frac{d}{dx}\{k\, f(x)\}$

$$= \lim_{\Delta x \to 0} \left\{ \frac{kf(x+\Delta x) - kf(x)}{\Delta x} \right\}$$

$$= \lim_{\Delta x \to 0} \frac{k\{f(x+\Delta x) - f(x)\}}{\Delta x}$$

$$= k \lim_{\Delta x \to 0} \frac{f(x+\Delta x) - f(x)}{\Delta x}$$

$$= k.\frac{d}{dx} f(x)$$

(3) $\frac{d}{dx}\{f(x) - g(x)\} = \frac{d}{dx} f(x) - \frac{d}{dx} g(x)$

proof: Just as that of (1) above

(4) $\frac{d}{dx}\{f(x).g(x)\} = f(x)\frac{d}{dx} g(x) + g(x)\frac{d}{dx} f(x)$

proof: $\frac{d}{dx}\{f(x).g(x)\}$

$$= \lim_{\Delta x \to 0} \frac{f(x+\Delta x)g(x+\Delta x) - f(x).g(x)}{\Delta x}$$

$$= \lim_{\Delta x \to 0} \frac{f(x+\Delta x)g(x+\Delta x) - f(x).g(x+\Delta x) + f(x)g(x+\Delta x) - f(x)g(x)}{\Delta x}$$

$$= \lim_{\Delta x \to 0} \frac{g(x+\Delta x)\{f(x+\Delta x) - f(x)\} + f(x)\{g(x+\Delta x) - g(x)\}}{\Delta x}$$

$$= \lim_{\Delta x \to 0} \frac{g(x+\Delta x)\{f(x+\Delta x) - f(x)\}}{\Delta x}$$

$$+ \lim_{\Delta x \to 0} \frac{f(x)\{g(x+\Delta x) - g(x)\}}{\Delta x}$$

$$= \lim_{\Delta x \to 0} g(x+\Delta x) \lim_{\Delta x \to 0} \frac{f(x+\Delta x) - f(x)}{\Delta x}$$

$$+ \lim_{\Delta x \to 0} \frac{f(x)\{g(x+\Delta x) - g(x)\}}{\Delta x}$$

$$= g(x).\frac{d}{dx} f(x) + f(x) \lim_{\Delta x \to 0} \frac{g(x+\Delta x) - g(x)}{\Delta x}$$

$$= g(x)\frac{d}{dx} f(x) + f(x).\frac{d}{dx} g(x)$$

(5) $\frac{d}{dx}\left\{\frac{f(x)}{g(x)}\right\} = \frac{g(x)\frac{d}{dx} f(x) - f(x)\frac{d}{dx} g(x)}{\{g(x)\}^2}$

proof:

$$\frac{d}{dx}\frac{f(x)}{g(x)}$$

$$= \lim_{\Delta x \to 0} \frac{\dfrac{f(x+\Delta x)}{g(x+\Delta x)} - \dfrac{f(x)}{g(x)}}{\Delta x}$$

$$= \lim_{\Delta x \to 0} \frac{g(x)f(x+\Delta x) - f(x)g(x+\Delta x)}{g(x)g(x+\Delta x)\Delta x}$$

$$= \lim_{\Delta x \to 0} \frac{g(x)f(x+\Delta x) - f(x)g(x) - f(x)g(x+\Delta x) + f(x)g(x)}{g(x)g(x+\Delta x)\Delta x}$$

$$= \lim_{\Delta x \to 0} \frac{\{g(x)f(x+\Delta x) - f(x)g(x)\} - \{f(x)g(x+\Delta x) - f(x)g(x)\}}{g(x)g(x+\Delta x)\Delta x}$$

$$= \lim_{\Delta x \to 0} \left\{\frac{g(x)f(x+\Delta x) - f(x)g(x)}{g(x)g(x+\Delta x)\Delta x} - \frac{f(x)g(x+\Delta x) - f(x)g(x)}{g(x)g(x+\Delta x)\Delta x}\right\}$$

$$= \lim_{\Delta x \to 0} \left\{\frac{g(x)f(x+\Delta x) - f(x)g(x)}{g(x)g(x+\Delta x)\Delta x}\right\} -$$

$$\lim_{\Delta x \to 0} \left\{\frac{f(x)g(x+\Delta x) - f(x)g(x)}{g(x)g(x+\Delta x)\Delta x}\right\}$$

$$= \lim_{\Delta x \to 0} \frac{g(x)\{f(x+\Delta x) - f(x)\}}{g(x)g(x+\Delta x)\Delta x} - \lim_{\Delta x \to 0} \frac{f(x)\{g(x+\Delta x) - g(x)\}}{g(x)g(x+\Delta x)\Delta x}$$

$$= \lim_{\Delta x \to 0} \left\{\frac{g(x)}{g(x)g(x+\Delta x)}\right\} \bullet \lim_{\Delta x \to 0} \left\{\frac{f(x+\Delta x) - f(x)}{\Delta x}\right\}$$

$$- \lim_{\Delta x \to 0} \left\{\frac{f(x)}{g(x)g(x+\Delta x)}\right\} \lim_{\Delta x \to 0} \left\{\frac{g(x+\Delta x) - g(x)}{\Delta x}\right\}$$

$$= \lim_{\Delta x \to 0} \frac{1}{g(x+\Delta x)} \lim_{\Delta x \to 0} \frac{f(x+\Delta x) - f(x)}{\Delta x} - \lim_{\Delta x \to 0} \frac{f(x)}{g(x)g(x+\Delta x)}$$

$$\lim_{\Delta x \to 0} \frac{g(x+\Delta x) - g(x)}{\Delta x}$$

$$= \frac{1}{g(x)} \frac{d}{dx} f(x) - \frac{f(x)}{\{g(x)\}^2} \cdot \frac{d}{dx} g(x)$$

$$= \frac{g(x)\frac{d}{dx}f(x) - f(x)\frac{d}{dx}g(x)}{\{g(x)\}^2}$$

Some Important Derivatives

(1) $\frac{d}{dx}x^n = n.x^{n-1}$; n is real number

(2) $\frac{d}{dx}\ k = 0$; k is a constant

(3) $\frac{d}{dx}\ \log_e x = \frac{1}{x}$

(4) $\frac{d}{dx}\ e^{ax} = a.e^{ax}$; a is a constant

(5) $\frac{d}{dx}\ a^x = a^x \log_e a$; a is a constant

SOLVED EXAMPLES

7.1. Find $\lim_{x\to 1}\frac{x^2 - 1}{x - 1}$

Solution: $\lim_{x\to 1}\frac{(x+1)(x-1)}{(x-1)}$

$= (1 + 1) = 2$

7.2. Find $\lim_{x\to 1}(x^2 + 2x + 2)$

Solution: $\lim_{x\to 2}(x^2 + 2x + 2)$

$= 2^2 + 2.2 + 2 = 10$

7.3. Find $\lim_{x\to 2}\frac{x^2 - 5x + 6}{x^2 - 9x + 18}$

Solution: $\lim_{x\to 3}\frac{x^2 - 5x + 6}{x^2 - 9x + 18}$

$$= \lim_{x\to 3}\frac{x^2-3x-2x+6}{x^2-6x-3x+18}$$

$$= \lim_{x\to 3}\frac{x(x-3)-2(x-3)}{x(x-6)-3(x-6)}$$

$$= \lim_{x\to 3}\frac{\cancel{(x-3)}(x-2)}{\cancel{(x-3)}(x-6)} = \frac{1}{-3}$$

7.4. Find $\lim_{x\to 3}\frac{x^3-27}{x^2-9}$

Solution: $$= \lim_{x\to 3}\frac{\cancel{(x-3)}(x^2+3x+9)}{\cancel{(x-3)}(x+3)}$$

$$= \frac{3^2+3.3+9}{3+3} = \frac{27}{6}$$

7.5. Find $\lim_{x\to 0}\frac{\sqrt{1+x}-\sqrt{1-x}}{x}$

Solution: $$\lim_{x\to 0}\frac{\sqrt{1+x}-\sqrt{1-x}}{x}$$

$$= \lim_{x\to 0}\frac{(\sqrt{1+x}-\sqrt{1-x})(\sqrt{1+x}+\sqrt{1-x})}{x(\sqrt{1+x}+\sqrt{1-x})}$$

$$= \lim_{x\to 0}\frac{(1+x)-(1-x)}{x(\sqrt{1+x}+\sqrt{1-x})}$$

$$= \lim_{x\to 0}\frac{2\cancel{x}}{\cancel{x}(\sqrt{1+x}+\sqrt{1-x})}$$

$$= \frac{2}{\sqrt{1+0}+\sqrt{1-0}} = \frac{2}{2\sqrt{1}} = \frac{2}{2} = 1$$

7.6. Find $\lim_{x\to 5}\frac{1-\sqrt{x-4}}{x-5}$

Solution: $\lim\limits_{x\to 5}\dfrac{1-\sqrt{x-4}}{x-5}$

$$=\lim_{x\to 5}\frac{\left(1-\sqrt{x-4}\right)\left(1+\sqrt{1-4}\right)}{(x-5)\left(1+\sqrt{x-4}\right)}$$

$$=\lim_{x\to 5}\frac{1-(x-4)}{(x-5)(1+\sqrt{x-4})}$$

$$=\lim_{x\to 5}\frac{-(x-5)}{(x-5)\left(1+\sqrt{x-4}\right)}$$

$$=\lim_{x\to 5}\frac{(-1)}{1+\sqrt{x-4}}$$

$$=\frac{(-1)}{1+\sqrt{1}}=-\frac{1}{2}$$

7.7. Find $\lim\limits_{x\to 2}\dfrac{x^2-4}{\sqrt{3x-2}-\sqrt{x+2}}$

Solution: $\lim\limits_{x\to 2}\dfrac{(x-2)(x-2)(\sqrt{3x-2}+\sqrt{x+2})}{(\sqrt{3x-2}-\sqrt{x+2})(\sqrt{3x-2}+\sqrt{x+2})}$

$$=\lim_{x\to 2}\frac{(x-2)(x-2)(\sqrt{3x-2}+\sqrt{x+2})}{\{(3x-2)-(x+2)\}}$$

$$=\lim_{x\to 2}\frac{\cancel{(x-2)}(x+2)(\sqrt{3x-2}+\sqrt{x+2})}{2\cancel{(x-2)}}$$

$$=\frac{(2+2)(\sqrt{3.2-2}+\sqrt{2+2})}{2}=8$$

7.8. Find $\lim\limits_{x\to\sqrt{2}}\dfrac{x^4-4}{x^2+3x\sqrt{2}-8}$

Solution: $\lim\limits_{x\to\sqrt{2}}\dfrac{(x^2-2)(x^2+2)}{x^2+4\sqrt{2}x-\sqrt{2}x-8}$

$$= \lim_{x\to\sqrt{2}} \frac{(x+\sqrt{2})(x-\sqrt{2})(x^2+2)}{x(x+4\sqrt{2})-\sqrt{2}(x+4\sqrt{2})}$$

$$= \lim_{x\to\sqrt{2}} \frac{(x+\sqrt{2})\cancel{(x-\sqrt{2})}(x^2+2)}{\cancel{(x-\sqrt{2})}(x+4\sqrt{2})}$$

$$= \frac{(\sqrt{2}+\sqrt{2})((\sqrt{2})^2+2)}{\sqrt{2}+4\sqrt{2}}$$

$$= \frac{2\sqrt{2}.4}{5\sqrt{2}} = \frac{8}{5}$$

7.9. Find $\lim_{n\to\infty} \frac{(n+1)^2+(n+2)^2}{(n+3)^2}$

Solution: $$\lim_{n\to\infty} \frac{n^2+2n+1+n^2+4n+4}{n^2+6n+9}$$

$$= \lim_{n\to\infty} \frac{2n^2+6n+5}{n^2+6n+9}$$

$$= \lim_{n\to\infty} \frac{\frac{2n^2+6n+5}{n^2}}{\frac{n^2+6n+9}{n^2}} = \lim_{n\to\infty} \frac{2+\frac{6}{n}+\frac{5}{n^2}}{1+\frac{6}{n}+\frac{9}{n^2}}$$

$$= \frac{2}{1} \quad (\because \lim_{n\to\infty} \frac{a}{n} = 0)$$

$$= 2$$

[Dividing both Numberator and Dinominator by highest power of x]

7.10 Find $\lim_{n\to\infty}(3n - \sqrt{9n^2-6n+2})$

Solution: $$\lim_{n\to\infty} \frac{(3n-\sqrt{9n^2-6n+2})(3n+\sqrt{9n^2-6n+2})}{(3n+\sqrt{9n^2-6n+2})}$$

$$= \lim_{n\to\infty} \frac{9n^2-(9n^2-6n+2)}{3n+\sqrt{9n^2-6n+2}}$$

$$= \lim_{n\to\infty} \frac{6n-2}{3n+\sqrt{9n^2-6n+2}}$$

$$= \lim_{n\to\infty} \frac{\frac{6n-2}{n}}{\frac{3n+\sqrt{9n^2-6n+2}}{n}} = \lim_{n\to\infty} \frac{6-\frac{2}{n}}{3+\sqrt{\frac{9n^2-6n+2}{n^2}}}$$

$$= \lim_{n\to\infty} \frac{6-\frac{2}{n}}{3+\sqrt{9-\frac{6}{n}+\frac{2}{n^2}}} = \frac{6}{3+\sqrt{9}} = 1$$

7.11. Find $\lim_{x\to 0} \frac{\sqrt{1+ax}-\sqrt{1-ax}}{x}$ **[AHSEC, 04]**

Solution: $\lim_{x\to 0} \frac{(\sqrt{1+ax}-\sqrt{1-ax})(\sqrt{1+ax}+\sqrt{1-ax})}{x(\sqrt{1+ax}+\sqrt{1-ax})}$

$$= \lim_{x\to 0} \frac{(1+ax)-(1-ax)}{x(\sqrt{1+ax}+\sqrt{1-ax})}$$

$$= \lim_{x\to 0} \frac{2ax}{x(\sqrt{1+ax}+\sqrt{1-ax})}$$

$$= \frac{2a}{\sqrt{1}+\sqrt{1}} = a$$

7.12. Find $\lim_{x\to\infty} \frac{3x^2+5x-11}{5x^2+6x+7}$ **[AHSEC, 02]**

Solution: $\lim_{x\to\infty} \frac{3x^2+5x-11}{5x^2+6x+7}$

$$= \lim_{x\to\infty} \frac{\frac{3x^2+5x-11}{x^2}}{\frac{5x^2+6x+7}{x^2}} = \lim_{x\to\infty} \frac{3+\frac{5}{x}-\frac{11}{x^2}}{5+\frac{6}{x}+\frac{7}{x^2}}$$

$$= \frac{3+0-0}{5+0+0} = \frac{3}{5}$$ [As $x\to\infty$, $\frac{1}{x}\to 0$]

7.13. Find $\lim\limits_{x\to 0}\dfrac{a-\sqrt{a^2-x^2}}{x^2}$

Solution:

$$\lim_{x\to 0}\frac{a-\sqrt{a^2-x^2}}{x^2}$$

$$=\lim_{x\to 0}\frac{(a-\sqrt{a^2-x^2})(a+\sqrt{a^2-x^2})}{x^2(a+\sqrt{a^2-x^2})}$$

$$=\lim_{x\to 0}\frac{a^2-(a^2-x^2)}{x^2(a+\sqrt{a^2-x^2})}$$

$$=\lim_{x\to 0}\frac{x^2}{x^2(a+\sqrt{a^2-x^2})}$$

$$=\lim_{x\to 0}\frac{1}{a+\sqrt{a^2-x^2}}$$

$$=\frac{1}{a+\sqrt{a^2-0}}$$

$$=\frac{1}{a+\sqrt{a^2}}=\frac{1}{2a}$$ **[AHSEC, 07]**

7.14. Find $\lim\limits_{x\to 4}\dfrac{\sqrt{x}-2}{x-4}$ **[AHSEC, 01]**

Solution:

$$\lim_{x\to 4}\frac{(\sqrt{x}-2)(\sqrt{x}+2)}{(x-4)(\sqrt{x}+2)}$$

$$=\lim_{x\to 4}\frac{\cancel{(x-4)}}{\cancel{(x-4)}(\sqrt{x}+2)}$$

$$=\lim_{x\to 4}\frac{1}{\sqrt{x+2}}$$

$$=\frac{1}{4}$$

7.15. Find $\lim\limits_{x\to\infty}(\sqrt{x+1}-\sqrt{x})$ **[AHSEC, 06]**

Solution: $\lim\limits_{x\to\infty}(\sqrt{x+1}-\sqrt{x})$

$$= \lim_{x\to\infty} \frac{(\sqrt{x+1}-\sqrt{x})(\sqrt{x+1}+\sqrt{x})}{(\sqrt{x+1}+\sqrt{x})}$$

$$= \lim_{x\to\infty} \frac{(x+1)-x}{(\sqrt{x+1}+\sqrt{x})}$$

$$= \lim_{x\to\infty} \frac{1}{\sqrt{x+1}+\sqrt{x}} = 0$$

7.16. A function is defined as follows

$$\begin{aligned} f(x) &= x^2;\ x < 1 \\ &= 2.5;\ x = 1 \\ &= x^2 + 2;\ x > 1 \end{aligned}$$

Does $\lim_{x\to 1} f(x)$ exist? **[AHSEC, 95]**

Solution: Left hand limit at $x = 1$,

$$\lim_{x\to 1^-} f(x) = \lim_{x\to 1^-} x^2 = 1$$

Right hand limit at $x = 1$

$$\lim_{x\to 1^+} f(x)$$

$$= \lim_{x\to 1^+} (x^2+2)$$

$$= 3$$

The value of functionl at $x = 1$

$$f(1) = 2.5$$

Thus $\lim_{x\to 1^-} f(x) \neq \lim_{x\to 1^+} f(x) \neq f(1)$

So the limit does not exist

7.17. If $f(x) = |x|$; for all x ; discuss the continuity at $x = 0$ [AHSEC, 96]

Solution: Left hand limit, (at $x = 0$)	right hand limit (at x = 0)
$= \lim_{x\to 0^-} f(x)$	$= \lim_{x\to o^+} f(x)$
$= \lim_{x\to 0^-} \lvert x \rvert$	$= \lim_{x\to 0+} \lvert x \rvert$
$= 0$	$= 0$

Value of the function at $x = 0$

$$f(0)$$
$$= |0|$$
$$= 0$$

$\therefore \lim_{x\to 0^-} f(x) = \lim_{x\to 0^+} f(x) = f(0)$

So the function is continuous at $x = 0$

7.18. What is the difference between $f(a)$ and $\lim_{x\to a} f(x)$?

Solution: $f(a)$ is the value of function at $x = a$ but $\lim_{x\to a} f(x)$ is the limit of the function when x takes value in the neighbourhood of a but $x \neq a$.

7.19. Draw the graph of the function

$$f(x) = \frac{|x-2|}{x-2}; \; x \neq 2$$
$$0; \; x = 2$$

[AHSEC, 96]

Solution: Let us first prepare the following table

x:	(– 1)	0	1	1.5	1.9	2	2.1	3	4
$f(x)$:	(–1)	(– 1)	(– 1)	(– 1)	(– 1)	0	1	1	1

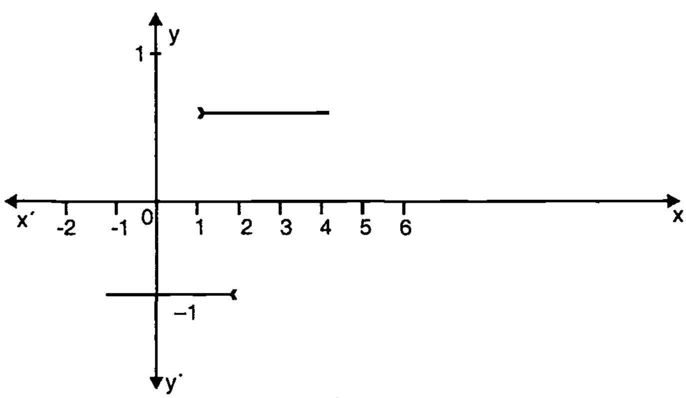


Fig. 7.2

7.20. Draw the graph of the function.

$$f(x) = 0; \; |x| > 1$$
$$= 1 + x; \; -1 \leq x \leq 0$$
$$= 1 - x; \; 0 < x \leq 1$$

[AHSEC, 95]

Solution: Let us first prepare the following table

x :	(– 1)	(– .5)	(– .2)	(– .1)	(– .01)	0
$f(x)$:	0	.5	.8	.9	.99	1
	.01	.1	.2	.5	1	
	.99	.9	.8	.5	0	

Now let us draw the graph

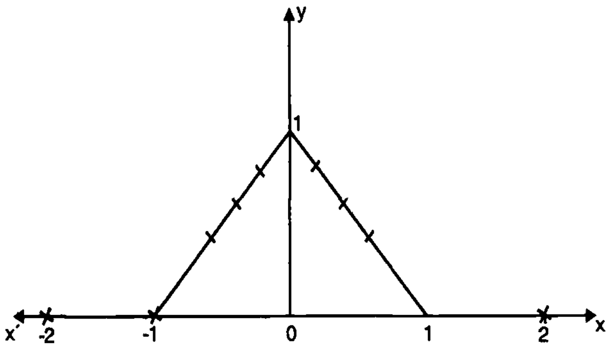


Fig. 7.3

7.21. Find derivative of

(i) x^5 (ii) $7x^5$ (iii) x^{-7} (iv) $\sqrt{x}$ (v) $\frac{1}{x}$

Solution:(i) $\frac{d}{dx} x^5$

$= 5.x^{5-1}$ $\left(\because \frac{d}{dx} x^n = n.x^{n-1}\right)$

$= 5.x^4$

(ii) $\frac{d}{dx}(7.x^5)$

$= 7. \frac{d}{dx}(x^5)$ $\left[\left(\because \frac{d}{dx}\{kf(x)\} = k\frac{d}{dx}\{f(x)\}\right]\right.$

$= 7.5.\ x^{5-1} = 35\ x^4$

(iii) $\frac{d}{dx} x^{-7}$

$= (-7)x^{-7-1}$

$= -7\,x^{-8}$

(iv) $\frac{d}{dx}\ \sqrt{x}$

$$\frac{d}{dx}\ x^{\frac{1}{2}} = \frac{1}{2}\ x^{\frac{1}{2}-1}$$

$$= \frac{1}{2\sqrt{x}}$$

(v) $\frac{d}{dx}\left(\frac{1}{\sqrt{x}}\right)$

$$= \frac{d}{dx}\ x^{-½} = \left(-\frac{1}{2}\right)x^{-\frac{1}{2}-1} = -\frac{1}{2}x^{-\frac{3}{2}}$$

7.22. Find derivative of $\log_e\left(\frac{1}{\sqrt{x}}\right) - 7\ e^x + \frac{5}{3}\ a^x$

Solution: $\frac{d}{dx}\left\{\log_e\left(\frac{1}{\sqrt{x}}\right) - 7e^x + \frac{5}{3}a^x\right\}$

$$= \frac{d}{dx}\ \log_e \frac{1}{\sqrt{x}} - \frac{d}{dx}(7.e^x) + \frac{d}{dx}\left(\frac{5}{3}a^x\right)$$

$$= \frac{d}{dx}\ \log_e x^{½} - 7.\frac{d}{dx}e^x + \frac{5}{3}\frac{d}{dx}a^x$$

$$= \frac{d}{dx}\left(-\frac{1}{2}\log_e x\right) - 7.e^x + \frac{5}{3}\ a^x \log_e a$$

$$= -\frac{1}{2}.\frac{1}{x} - 7.e^x + \frac{5}{3}\ a^x \log_e a$$

7.23. Find derivative of

(i) $\frac{x^2+3x+5}{x^2+2}$ (ii) $\frac{1}{x^2+5}$ (iii) $\frac{e^x}{x^3}$

Solution: (i) $\frac{d}{dx}\left(\frac{x^2+3x+5}{x^2+2}\right)$

$$= \frac{(x^2+2)\frac{d}{dx}(x^2+3x+5) - (x^2+3x+5)\frac{d}{dx}(x^2+2)}{(x^2+2)^2}$$

$$= \frac{(x^2+2)(2x+3)-(x^2+3x+5)(2x)}{(x^2+2)^2}$$

$$= \frac{2x^3+3x^2+4x+6-2x^3-6x^2-10x}{(x^2+2)^2}$$

$$= \frac{-3x^2-6x+6}{(x^2+2)^2}$$

(ii) $$\frac{d}{dx}\left(\frac{1}{x^2+5}\right)$$

$$= \frac{\left(x^2+5\right)\frac{d}{dx}(1)-1.\frac{d}{dx}(x^2+5)}{(x^2+5)^2} = \frac{-2x}{(x^2+5)^2}$$

(iii) $$\frac{d}{dx}\frac{e^x}{x^3}$$

$$= \frac{x^3\frac{d}{dx}(e^x)-e^x\frac{d}{dx}x^3}{(x^3)^2} = \frac{x^3.e^x-e^x.3.x^2}{x^6}$$

$$= \frac{x^3e^x-3e^x.x^2}{x^6}$$

7.24. Find $\frac{dy}{dx}$ if

$$x^3+y^3-4xy=0$$

Solution: Differentiating w.r.t. x we get

$$\frac{d}{dx}(x^3)+\frac{d}{dx}(y^3)-\frac{d}{dx}(4xy)=0$$

$$\Rightarrow 3x^2+\frac{d}{dx}(y^3)\frac{dy}{dx}-4y\frac{d}{dx}x-4x\frac{d}{dx}y= 0$$

$$\Rightarrow \quad 3x^2+3y^2\frac{dy}{dx}-4y-4x\frac{dy}{dx}=0$$

$$\Rightarrow \quad (3y^2-4x)\frac{dy}{dx}=4y-3x^2$$

$$\Rightarrow \qquad \frac{dy}{dx} = \frac{4y-3x^2}{3y^2-4x}$$

7.25. Find $\frac{dy}{dx}$; **if** $x = \frac{3at}{1+t^2}$, $y = \frac{3at^2}{1+t^2}$

Solution: $x = \frac{3at}{1+t^2}$

$$\Rightarrow \qquad \frac{dx}{dt} = \frac{d}{dt}\left(\frac{3at}{1+t^2}\right)$$

$$= \frac{(1+t^2)\frac{d}{dt}(3\ at) - 3at\frac{d}{dt}(1+t^2)}{(1+t^2)^2}$$

$$= \frac{(1+t^2)\frac{d}{dt}(3at) - 3at.2t}{(1+t^2)^2} = \frac{3a+3at^2-6at^2}{(1+t^2)^2} \qquad ...(1)$$

$$= \frac{3a-3at^2}{(1+t^2)^2}$$

Also $y = \frac{3at^2}{1+t^2}$

$$\Rightarrow \qquad \frac{dy}{dt} = \frac{d}{dt}\left(\frac{3at^2}{1+t^2}\right)$$

$$= \frac{(1+t^2)\frac{d}{dt}(3\ at^2) - 3at^2\frac{d}{dt}(1+t^2)}{(1+t^2)^2}$$

$$= \frac{(1+t^2)\frac{d}{dt}(3at^2) - 3at^2.2t}{(1+t^2)^2}$$

$$= \frac{6at+6at^3-6at^3}{(1+t^2)^2}$$

$$= \frac{6at}{(1+t^2)^2}$$

$$\therefore \qquad \frac{dy}{dx} = \frac{\frac{dy}{dt}}{\frac{dx}{dt}}$$

$$= \frac{\frac{6at}{(1+t^2)^2}}{\frac{(3a-3at^2)}{(1+t^2)^2}} = \frac{2t}{1-t^2}$$

7.26. Find derivative of

(i) $\log_e(\log_e x)$ **(ii)** e^{x^2}

Solution: (i) $\frac{d}{dx}(\log_e(\log_e x))$

$$= \frac{d}{d\log_e x}\log_e(\log_e x)\frac{d\log_e x}{dx} = \frac{1}{\log_e x}\cdot\frac{1}{x}$$

(ii) $\frac{d}{dx}e^{x^2}$

$$= \frac{d}{dx^2}e^{x^2}\frac{dx^2}{dx} = e^{x^2}.2x = 2\,x.\,e^{x^2}$$

7.27. Find derivative.

(i) $x.2^x$ **[AHSEC, 04]**

(ii) e^{ax^2+bx} **[AHSEC, 02]**

(iii) $xe^x\log_e x$ **[AHSEC, 01]**

(iv) $\frac{1}{\sqrt{2\pi}}e^{-\frac{x^2}{2}}$ **[AHSEC, 00]**

(v) $\frac{1+\sqrt{x}}{1-\sqrt{x}}$ **[AHSEC, 97]**

Solution: (i) $\frac{d}{dx}(x.2^x)$

$$= x\frac{d}{dx}2^x + 2^x\frac{d}{dx}x$$

$= x\, 2^x \log_e 2 + 2^x.1$

$= 2^x (x \log_e 2 + 1)$

(ii) $\dfrac{d}{dx}\left(e^{ax^2+bx}\right)$

$$= \frac{de^{ax^2+bx}}{d(ax^2+bx)}\ \frac{d\left(ax^2+bx\right)}{dx} = e^{ax^2+bx}.(2ax+b)$$

(iii) $\dfrac{d}{dx}\ (x.e^x \log_e x)$

$$= x\ \frac{d}{dx}\ (e^x.\log_e x) + (e^x \log_e x)\frac{d}{dx}\ x$$

$$= x\left[e^x\frac{d}{dx}\log_e x + (\log_e x)\frac{d}{dx}e^x\right] + e^x \log_e x.1$$

$$= x\left[e^x\frac{1}{x} + (\log_e x)e^x\right] + e^x.\ \log_e x = e^x + e^x (\log_e x).x + e^x \log_e x$$

(iv) $$\frac{d}{dx}\left(\frac{1}{\sqrt{2\pi}}e^{-\frac{x^2}{2}}\right) = \frac{1}{\sqrt{2\pi}}\ \frac{d}{dx}\left(e^{-\frac{x^2}{2}}\right)$$

$$= \frac{1}{\sqrt{2\pi}}\ \frac{d\,e^{-\frac{x^2}{2}}}{d\left(-\frac{x^2}{2}\right)}\ \frac{d\left(-\frac{x^2}{2}\right)}{dx}$$

$$= \frac{1}{\sqrt{2\pi}}\ e^{-\frac{x^2}{2}}\left(-\frac{1}{2}\right).\ 2x = -\ \frac{xe^{-\frac{x^2}{2}}}{\sqrt{2\pi}}$$

(v) $$\frac{d}{dx}\left(\frac{1+\sqrt{x}}{1-\sqrt{x}}\right)$$

$$= \frac{(1-\sqrt{x})\frac{d}{dx}(1+\sqrt{x}) - (1+\sqrt{x})\frac{d}{dx}(1-\sqrt{x})}{(1-\sqrt{x})^2}$$

$$= \frac{(1-\sqrt{x})\frac{1}{2}x^{-\frac{1}{2}} - (1+\sqrt{x})\left(-\frac{1}{2}\right)x^{-\frac{1}{2}}}{(1-\sqrt{x})^2}$$

$$= \frac{\frac{1}{2}\left\{(1-\sqrt{x})+(1+\sqrt{x})\right\}}{\sqrt{x}(1-\sqrt{x})^2} = \frac{1}{\sqrt{x}(1-\sqrt{x})^2}$$

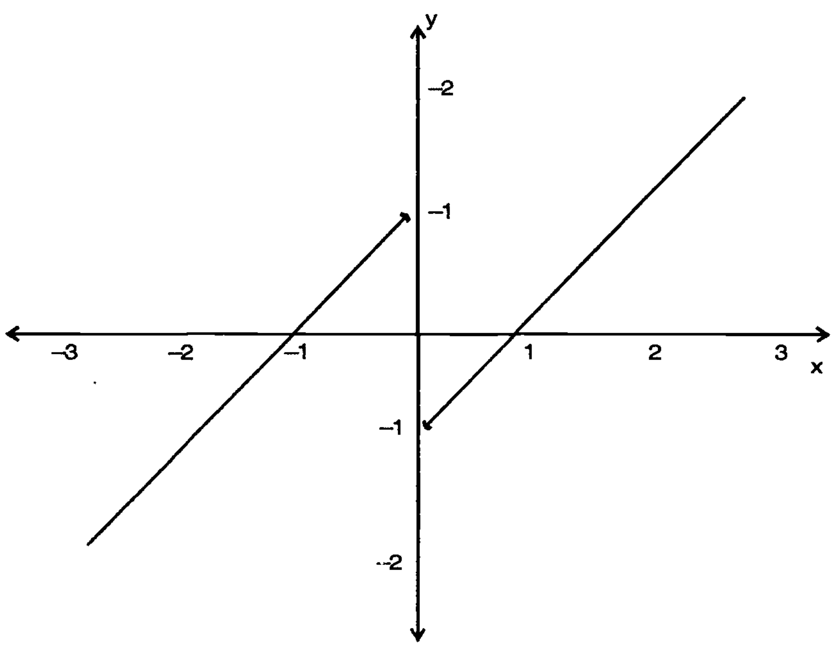
7.28. Draw the graph of

$$\begin{aligned} f(x) &= x-1 \text{ when } x>0 \\ &= \frac{1}{x} \text{ when } x=0 \\ &= x+1 \text{ when } x<0 \end{aligned}$$ **(AHSEC, 2008)**

Solution: Let us first prepare the following table

x :	3	2	1	.5	.1	0
$f(x)$:	2	1	0	(–.5)	(–.9)	(–.5)
x :	(–.1)	(–.5)	(–1)	(–2)	(–3)	
$f(x)$:	.9	.5	0	(–1)	(–2)	

Now let us draw the graph



7.29. Find $\lim\limits_{x\to 1}\dfrac{x^3-3x^2+2}{x^2+5x-6}$ **(AHSEC 2008)**

Solution: $\lim\limits_{x\to 1}\dfrac{x^3-3x^2+2}{x^2+5x-6}$

$$=\lim_{x\to 1}\frac{x^2(x-1)-2x(x-1)-2(x-1)}{x^2+6x-x-6}$$

$$=\lim_{x\to 1}\frac{\cancel{(x-1)}(x^2-2x-2)}{(x+6)\cancel{(x-1)}}$$

$$=\frac{1-2-2}{1+6}$$

$$=\frac{-3}{7}$$

$$=\left(-\frac{3}{7}\right)$$

7.30. Find derivative of

(i) $\log_e\sqrt{\dfrac{1-x^2}{1+x^n}}$ **(AHSEC, 2008)**

(ii) $e^x \,.\, x^n$ **(AHSEC, 2007)**

(iii) $x\, x^2\, (x-1)$ **(AHSEC, 2006)**

Solution: (i)

$$\frac{d}{dx}\log_e\sqrt{\frac{1-x^2}{1+x^2}}$$

$$=\frac{d\log_e\sqrt{\frac{1-x^2}{1+x^2}}}{d\left(\sqrt{\frac{1-x^2}{1+x^2}}\right)}\cdot\frac{d\sqrt{\frac{1-x^2}{1+x^2}}}{dx}$$

$$= \frac{1}{\sqrt{\frac{1-x^2}{1+x^2}}} \frac{d\left(\sqrt{\frac{1-x^2}{1+x^2}}\right)}{d\left(\frac{1-x^2}{1+x^2}\right)} \frac{d}{dx}\left(\frac{1-x^2}{x+x^2}\right)$$

$$= \sqrt{\frac{1+x^2}{1-x^2}} \frac{d\left(\frac{1-x^2}{1+x^2}\right)^{\frac{1}{2}}}{d\left(\frac{1-x^2}{1+x^2}\right)} \cdot \frac{d}{dx}\left(\frac{1-x^2}{1+x^2}\right)$$

$$= \sqrt{\frac{1+x^2}{1-x^2}} \cdot \frac{1}{2}\left(\frac{1-x^2}{1+x^2}\right)^{\frac{1}{2}-1} \cdot \frac{d}{dx}\left(\frac{1-x^2}{1+x^2}\right)$$

$$= \sqrt{\frac{1+x^2}{1-x^2}} \cdot \frac{1}{2} \cdot \frac{1}{(1-x^2)/(1+x^2)}$$

$$\left[\frac{(1+x^2)\frac{d}{dx}(1-x^2) - (1-x^2)\frac{d}{dx}(1+x^2)}{(1+x^2)^2}\right]$$

$$= \frac{1}{2} \frac{1+x^2}{x-x^2} \left[\frac{(1+x^2)(0-2x) - (1-x^2)(2x)}{(1+x^2)^2}\right]$$

$$= \frac{1}{2} \cdot \frac{1}{(1-x^2)(1+x^2)} [-2x - 2x^3 - 2x + 2x^3]$$

$$= -\frac{2x}{(1-x^4)}$$

(ii) $\frac{d}{dx}\left(e^x . x^n\right)$

$$= e^x.\frac{d}{dx}.x^n + x^n.\frac{d}{dx}e^x$$

$$= e^x \,.\, n\,x^{n-1} + x^n \,.\, e^x$$

$$= e^x \,.\, x^{n-1}\,(\,n + x)$$

(iii) $\frac{d}{dx}\left[2x^2(x-1)\right]$

$$= \frac{d}{dx}\left[2x^3 - 2x^2\right]$$

$$= \frac{d}{dx}\left(2x^3\right) - \frac{d}{dx}\left(2x^2\right)$$

$$= 2\,.\,3\,x^{3-1} - 2.2\,x^{2-1}$$

$$= 6x^2 - 4x$$

EXERCISE–7

1. Differentiate the following

(i) a^x (ii) 2^x (iii) e^{2x} (iv) $\frac{1}{\sqrt{x}}$ (v) $\frac{ax+b}{cx+d}$ (vi) $e^{\sqrt{x}}$

(vii) $\sqrt{2x^2+6x+5}$ (viii) $\frac{1+e^x}{1-e^x}$ (ix) $e^x \log_e (1+x^2)$

(Ans. (i) $a^x \log_e a$ (ii) $2x \log_e 2$ (iii) $2e^{2x}$ (iv) $\left(-\frac{1}{2}\right) x^{-\frac{3}{2}}$ (v) $\frac{ad-bc}{(cx+d)^2}$

(vi) $\frac{e^{\sqrt{x}}}{2\sqrt{x}}$ (vii) $\frac{2x+3}{\sqrt{2x^2+6x+5}}$ (viii) $\frac{2e^x}{(1-e^x)^2}$ (ix) $e^x \left[\log_e(1+x^2)+\frac{2x}{1+x^2}\right]$

2. Find $\frac{dy}{dx}$ of the following

(i) $x = t^2, y = t^3$ (ii) $x = \log t, y = e^t$

(Ans. (i) $\frac{3t}{2}$ (ii) te^t)

3. Find $\frac{dy}{dx}$ of

(i) $x^3 + y^3 = 3axy$ (ii) $x^2 + y^2 = xy$

(iii) $x\sqrt{y} + y\sqrt{x} = 1$

(Ans. (i) $\frac{ay-x^2}{y^2-ax}$ (ii) $\frac{2x-y}{x-2y}$ (iii) $\frac{-\left(\sqrt{y}+\frac{y}{2\sqrt{x}}\right)}{\frac{x}{2y}+\sqrt{x}}$

4. Define $y = f(x)$. State the difference between $f(a)$ and $\underset{x\to a}{Lt} f(x)$.

(AHSEC, 2006)

5. Evaluate :

(i) $\underset{x\to 3}{Lt} \frac{x^2-9}{x-3}$ **(AHSEC, 2006)**

(ii) $\underset{x\to 0}{lt} \frac{2-\sqrt{4-x^2}}{x^2}$

(iii) $\underset{x\to\infty}{lt} \frac{2x^8-7x^2+5}{9x^4+5x+12}$

Ans. (i) 6,(ii) $\frac{1}{4}$, (iii) ∞

"This page is Intentionally Left Blank"

8

INTEGRAL CALCULAS

In differential calculas, we differentiate a funtion $f(x)$ and obtain another function $F(x)$ Now definitely there must be some way to obtain $f(x)$ from $F(x)$. This reverse process of differentiation is known as integration. In notation we write it as–

$$\frac{d}{dx} f(x) = F(x)$$

$$\Rightarrow \quad \int F(x)dx = f(x)$$

Some Important intergrals

(1) $\int x^n dn = \frac{x^{n+1}}{n+1} + c$

(2) $\int \frac{1}{x} dx = \log_e x + c$

(3) $\int e^x dx = e^x + c$

(4) $\int a^x dx = \frac{a^x}{\log_e a} + c$

(5) $\int \frac{1}{x^2 - a^2} dx = \frac{1}{2a} \log_e \frac{x-a}{x+a} + c$

(6) $\int \frac{1}{a^2 - x^2} dx = \frac{1}{2a} \log_e \frac{a+x}{a-x} + c$

where c is called constant of integration. We may verify if integration is the reverse of differentiation:–

we have said earlier–

$$\frac{d}{dx} \cdot f(x) = F(x)$$

then $\int F(x)dx = f(x)$ and this reverse process of differentiation is the integration. Now let us verify this for the above integrals–

(1) $$\frac{d}{dx}\left(\frac{x^{n+1}}{n+1}+c\right)$$

$$= \frac{d}{dx}\left(\frac{x^{n+1}}{n+1}\right)+\frac{d}{dx}(c)$$

$$= \frac{1}{(n+1)}(n+1)x^{n+1-1}+0 \; = x^n$$

and we stated $\int x^n dx = \frac{x^{n+1}}{n+1}+c$

(2) $$\frac{d}{dx}\ (\log_e x + c)$$

$$= \frac{d}{dx}\ (\log_e x) + \frac{d}{dx}\ (c) = \frac{1}{x}+0$$

and we have stated earlier–

$$\int \frac{1}{x} dx \ = \log_e x + c$$

(3) $$\frac{d}{dx}(e^x + c)$$

$$= \frac{d}{dx}(e^x)+\frac{d}{dx}(c)$$

$$= e^x + 0 = e^x$$

and we have stated earlier

$$\int e^x dx \ = e^x + c$$

(4) $$\frac{d}{dx}\left(\frac{a^x}{\log_e a}+c\right)$$

$$= \frac{d}{dx}\left(\frac{a^x}{\log_e a}\right)+\frac{d}{dx}(c)$$

$$= \frac{1}{\log_e a}.a^x.\log_e a + 0 = a^x$$

and we have stated earlier

$$\int a^x dx = \frac{a^x}{\log_e a} + c$$

(5)

$$\frac{d}{dx}\left(\frac{1}{2a}\log_e \frac{x-a}{x+a} + c\right)$$

$$= \frac{d}{dx}\left(\frac{1}{2a}\log_e \frac{x-a}{x+a}\right) + \frac{d}{dx}(c)$$

$$= \frac{1}{2a}\frac{d}{dx}\{\log_e(x-a) - \log_e(x+a)\} + 0$$

$$= \frac{1}{2a}\left\{\frac{d}{d(x-a)}\log_e(x-a).\frac{d(x-a)}{dx} - \right.$$

$$\left.\frac{d\log_e(x+a)}{d(x+a)}\frac{d(x+a)}{dx}\right\}$$

$$= \frac{1}{2a}\left\{\frac{1}{(x-a)} - \frac{1}{(x+a)}\right\} = \frac{1}{x^2 - a^2}$$

and we have stated earlier

$$\int \frac{1}{x^2 - a^2} dx = \frac{1}{2a}\log_e \frac{x-a}{x+a} + c$$

(6)

$$\frac{d}{dx}\left(\frac{1}{2a}\log_e \frac{a+x}{a-x} + c\right)$$

$$= \frac{1}{2a}\frac{d}{dx}\left\{\log_e \frac{a+x}{a-x}\right\} + \frac{d}{dx}(c)$$

$$= \frac{1}{2a}\frac{d}{dx}\{\log_e(x+a) - \log_e(a-x)\} + 0$$

$$= \frac{1}{2a}\left[\left\{\frac{d}{dx}\log_e(a+x)\right\} - \left\{\frac{d}{dx}\log_e(a-x)\right\}\right]$$

$$= \frac{1}{2a}\left[\frac{d\log_e(a+x)}{d(a+x)}\frac{d(a+x)}{dx} - \frac{d\log_e(a-x)}{d(a+x)}\frac{d(a-x)}{dx}\right]$$

$$= \frac{1}{2a}\left[\frac{1}{(a+x)} - \frac{1}{(a-x)}.(-1)\right] = \frac{1}{2a}\left[\frac{a-x+a+x}{(a+x)(a-x)}\right]$$

$$= \frac{1}{a^2 - x^2}$$

and we have stated earlier

$$\int \frac{1}{a^2 - x^2} dx = \frac{1}{2a} \log_e \frac{a+x}{a-x} + c$$

thus from these results it is clear that integration is the reverse process of differentiation.

Theorems on integration

If $f(x)$, $g(x)$
be any two functions of x and k be any constant then

(1) $\int \{f(x) + g(x)\}\, dx = \int f(x)dx + \int g(x)dx$

(2) $\int \{f(x) - g(x)\}\, dx = \int f(x)dx - \int g(x)dx$

(3) $\int \{f(x).g(x)\}\, dx = f(x)\int g(x)dx - \int\left\{\int g(x)dx\right\}\frac{d}{dx}.f(x).dx$

(4) $\int k.f(x)dx = k\int f(x)dx$

SOLVED EXAMPLES

8.1. Integrate

(a) $\sqrt{x}$ **(b)** $(x+b)^2$

Solution:

(a) $\int \sqrt{x}dx$

$$= \int x^{\frac{1}{2}}\, dx = \frac{x^{\frac{1}{2}+1}}{\frac{1}{2}+1} + c$$

$$= \frac{x^{\frac{3}{2}}}{\frac{3}{2}} + c = \frac{2}{3}x^{\frac{3}{2}} + c$$

(b) $\int (x+b)^2 dx$

$$= \int (x^2 + 2bx + b^2)dx$$

$$= \int x^2 dx + 2b\int x dx + b^2 \int x^0 dx$$

$$= \frac{x^{2+1}}{2+1} + 2b.\frac{x^{1+1}}{1+1} + b^2.\frac{x^{0+1}}{0+1} + c$$

$$= \frac{x^3}{3} + \frac{2bx^2}{2} + b^2.x + c$$

$$= \frac{x^3}{3} + bx^2 + b^2x + c$$

8.2. Integrate

(a) e^{ax+b} **(b)** $x\sqrt{x+3}$

Solution:

(a)	$\int e^{ax+b}dx$	(Let $ax + b = z$)
	$= \int e^z \frac{dz}{a}$	$\Rightarrow \frac{d}{dx}(ax+b) = \frac{dz}{dx}$
	$= \frac{1}{a}\int e^z dz$	$\Rightarrow dx = \frac{dz}{a}$

$$= \frac{1}{a}e^z + c = \frac{1}{a}e^{ax+b} + c$$

(This method is known as method of substitution)

(b) $\int x\sqrt{x+3}dx$;

Let $x + 3 = z \Rightarrow dx = dz$

$$= \int (z-3)z^{\frac{1}{2}}dz$$

$$= \int\left(z^{\frac{3}{2}} - 3.z^{\frac{1}{2}}\right)dz$$

$$= \int z^{\frac{3}{2}}dz - 3\int z^{\frac{1}{2}}dz$$

$$= \frac{z^{\frac{3}{2}+1}}{\frac{3}{2}+1} - 3.\frac{.z^{\frac{1}{2}+1}}{\frac{1}{2}+1} + c$$

$$= \frac{2}{5}z^{\frac{5}{2}} - 2z^{\frac{3}{2}} + c$$

$$= \frac{2}{5}(x+3)^{\frac{5}{2}} - 2.(x+3)^{\frac{3}{2}} + c$$

8.3. Integrate

(a) $(6x + 5)\sqrt{3x^2+5x+9}$ (b) $\frac{2x+5}{x^2+5x+3}$

Solution:

(a) Let $3x^2 + 5x + 9 = z$

$$\Rightarrow \quad \frac{d}{dx}\left(3x^2+5x+a\right) = \frac{d}{dx}z$$

$$\Rightarrow \quad (6x + 5)dx = dz$$

$$\therefore \int(6x+5)\sqrt{3x^2+5x+9}dx$$

$$= \int\sqrt{z}.dz = \int z^{\frac{1}{2}}dz$$

$$= \frac{z^{\frac{1}{2}+1}}{\frac{1}{2}+1} + c = \frac{2}{3}z^{\frac{3}{2}} + c$$

$$= \frac{2}{3}\left(3x^2 + 5x + 9\right)^{\frac{3}{2}} + c$$

(b) Let $x^2 + 5x + 3 = z$

$\Rightarrow$ $$\frac{d}{dx}\left(x^2 + 5x + 3\right) = \frac{dz}{dx}$$

$\Rightarrow$ $$(2x + 5)dx = dz$$

$\therefore$ $$\int \frac{2x+5}{x^2 + 5x + 3}dx$$

$$= \int \frac{dz}{z} = \log_e z + c$$

$$= \log_e\left(x^2 + 5x + 3\right) + c$$

8.4. Integrate $\frac{1}{e^x + 1}$ **(AHSEC, 04)**

Solution: $$\int \frac{1}{e^x + 1}dx$$

$$= \int \frac{1}{e^x\left(1 + e^{-x}\right)}dx$$

$$= \int \frac{e^{-x}}{\left(1 + e^{-x}\right)}dx \quad [\text{Let } 1 + e^{-x} = z - e^{-x}dx = dz]$$

$$= \int \frac{-dz}{z} = -\log_e z + c$$

$$= -\log_e\left(1 + e^{-x}\right) + c$$

8.5. Evaluate

(a) $\int \frac{1 + 3x + 7x^2 - 2x^3}{x^2}dx$ **(AHSEC, 02)**

(b) $\int \frac{1}{x^2 - 4}dx$ **(AHSEC, 02, 07)**

Solution:

(a)
$$\int\frac{1+3x+7x^2-2x^3}{x^2}dx$$

$$=\int\left(\frac{1}{x^2}+\frac{3}{x}+7-2.x\right)dx$$

$$=\int\frac{1}{x^2}dx+\int\frac{3}{x}dx+\int 7.dx-\int 2xdx$$

$$=\frac{x^{-2+1}}{-2+1}+3\log_e x+7.x-2\frac{x^{1+1}}{1+1}+c$$

$$=3\ \log_e x-\frac{1}{x}+7x-\frac{2x^2}{2}+c$$

$$=3\ \log_e x-\frac{1}{x}+7x-x^2+c$$

(b)
$$\int\frac{1}{x^2-4}dx=\int\frac{1}{x^2-2^2}dx$$

$$=\frac{1}{2.2}\log_e\frac{x-2}{x+2}+c=\frac{1}{4}\log_e\frac{x-2}{x+2}+c$$

8.6. Evaluate

(a) $\int x^2e^x dx$ **(AHSEC, 01, 07)**

(b) $\int\frac{x^2}{x^2-4}dx$ **(AHSEC, 01)**

Solution:

(a)
$$\int x^2e^xdx=x^2\int e^xdx-\int\left\{\frac{d}{dx}x^2\int e^xdx\right\}dx$$

$$=x^2e^x-\int 2xe^xdx$$

$$=x^2e^x-2\int xe^xdx$$

$$=x^2e^x-2\left[x\int e^xdx-\int\left\{\frac{d}{dx}x\int e^xdx\right\}dx\right.$$

$$=x^2e^x-2\left[xe^x-\int e^xdx\right]$$

$$= x^2e^x - 2xe^x + 2e^x + c$$
$$= e^x(x^2 - 2x + 2) + c$$

Note that if we would have proceeded like–

$$e^x \int x^2dx - \int\left\{\frac{d}{dx}e^x \int x^2dx\right\}dx$$

$$= \frac{e^x.x^3}{3} - \tfrac{1}{3}\int e^x.x^3dx$$

then the powers of x within the integral sign would go on increasing and termination wouldnot be possible.

(b)
$$\int\frac{x^2}{x^2-4}dx = \int\frac{x^2-4+4}{x^2-4}dx$$
$$= \int 1dx + \int 4.\frac{1}{x^2-4}dx$$
$$= x + 4\int\frac{1}{x^2-2^2}dx$$
$$= x + 4.\frac{1}{2.2}\log_e\frac{x-2}{x+2} + c$$
$$= x + \log_e\frac{x-2}{x+2} + c$$

Example 8.7: Evaluate

(a) $\int\frac{x}{x^2-9}dx$ **(AHSEC, 00)**

(b) $\int\frac{1}{16-x^2}dx$ **(AHSEC, 00)**

(c) $\int\frac{x}{x-2}dx$ **(AHSEC, 00)**

Solution:

(a) Let $x^2 - 9 = u$

$\Rightarrow$ $$\frac{d}{dx}\left(x^2-9\right) = \frac{du}{dx}$$

$\Rightarrow$ $$2xdx = du$$

$\therefore$ $$\int\frac{xdx}{x^2-9} = \tfrac{1}{2}\int\frac{du}{u}$$

$$= \frac{1}{2}\log_e u + c$$

$$= \frac{1}{2}\log_e\left(x^2-9\right)+c$$

(b) $$\int\frac{dx}{16-x^2} = \int\frac{1}{4^2-x^2}dx$$

$$= \frac{1}{2.4}\log_e\frac{4+x}{4-x}+c$$

$$= \frac{1}{8}\log_e\frac{4+x}{4-x}+c$$

(c) $$\int\frac{x}{x-2}dx = \int\frac{x-2+2}{x-2}dx$$

$$= \int 1dx + 2\int\frac{1}{x-2}dx$$

$$= x+2\int\frac{1}{z}dz \quad \text{Let } x-2=z \Rightarrow dx = dz$$

$$= x + 2\log_e z + c$$

$$= x + 2\log_e(x-2) + c$$

8.8. Evaluate

(a) $\int\frac{1}{x\log x}dx$ **(AHSEC, 99)**

(b) $\int\frac{(1+x)^3}{x}dx$ **(AHSEC, 99)**

Solution: (a) Let $\log_e x = u$

$$\Rightarrow \frac{du}{dx} = \frac{1}{x}$$

$$\Rightarrow du = \frac{1}{x}.dx$$

$$\therefore \quad \int\frac{1}{x\log_e x}dx = \int\frac{1}{u}.du$$

$$= \log_e u + c$$

$$= \log_e(\log_e x) + c$$

(b) $$\int\frac{(1+x)^3}{x}dx$$

$$= \int\frac{1+3x+3x^2+x^3}{x}dx$$

$$= \int \frac{1}{x} dx + 3\int dx + 3\int x\, dx + \int x^2 dx$$

$$= \log_e x + 3x + 3\frac{x^2}{2} + \frac{x^3}{3} + c$$

Difinite Integral

Let $f(x)$ be a continuous function defined in the closed interval (a, b) also let $\int f(x)dx = F(x)$. Then the definite integral of $f(x)$ over $[a, b]$ is defined as –

$\int_a^b f(x)dx = F(b) - F(a)$ where a is called lower limit and b is called upper limit of the integral.

Some Important Properties of Definite Integral

(i) $$\int_a^b f(x)dx = \int_a^b f(z)dz$$

Proof: Let $x = z \Rightarrow dx = dz$

when $x = a, z = a, x = b, z = b$

$$\therefore \quad \int_a^b f(x)dx = \int_a^b f(z)dz$$

(ii) $$\int_a^b f(x)dx = -\int_b^a f(x)dx$$

Proof:

$$\int_a^b f(x)dx = F(b) - F(a)$$

$$= -\{F(a) - F(b)\}$$

$$= -\left\{\int_b^a f(x)dx\right\}$$

$$= -\int_b^a f(x)dx$$

(iii) $$\int_a^b f(x)dx + \int_b^c f(x)dx = \int_a^c f(x)dx$$

Proof: $$\int_a^b f(x)dx + \int_b^c f(x)dx = \{F(b) - F(a)\} + \{F(c) - F(b)\}$$

$$= F(c) - F(a)$$

$$= \int_a^c f(x)dx$$

(iv) $$\int_0^a f(x)dx = \int_0^a f(a-x)dx$$

Proof: Let $y = a - x \Rightarrow dy = -dx$
when $x = a, y = o, x = o, y = a$

$\therefore$ $$\int_0^a f(a-x)dx$$

$$= \int_a^0 f(y)(-dy)$$

$$= -\int_a^0 f(y)dy = \int_0^a f(y)dy$$

$$= \int_0^a f(x)dx \text{ (by theorem 1)}$$

(v) $$\int_{-a}^a f(x)dx = 0 \text{ if } f(x) \text{ is an odd function}$$

$$= 2\int_0^a f(x)dx \text{ if } f(x) \text{ is an even function}$$

Proof: $$\int_{-a}^a f(x)dx$$

$$= \int_{-a}^o f(x)dx + \int_0^a f(x)dx \qquad **$$

Now consider $\int_{-a}^0 f(x)dx$; let $x = -y$

$\Rightarrow$ $dy = -dx$

Also $x = -a \Rightarrow y = a; x = 0 \Rightarrow y = 0$

$\therefore$ $$\int_{-a}^0 f(x)dx = +\int_{+a}^0 f(-y)(-dy)$$

$$= -\int_a^0 f(-y)dy$$

But if $f(x)$ is an even function then $f(-x) = f(x)$ so,

$$\int_a^0 f(-y)dy = \int_a^0 f(y)dy$$

$$= \int_a^0 f(x)dx \text{ (by theorem (i))}$$

$$= -\int_0^a f(x)dx \text{ (by theorem (ii))}$$

$$\therefore \quad \int_{-a}^0 f(x)dx = \int_0^a f(x)dx$$

But if $f(x)$ is an odd function then $f(-x) = -f(x)$ so,

$$\int_a^0 f(-y)dy = -\int_a^0 f(y)dy$$

$$= -\int_a^0 f(x)dx \text{ (by theorem (i))}$$

$$= \int_0^a f(x)dx \text{ (by theorem (ii))}$$

$$\therefore \quad \int_{-a}^0 f(x)dx = -\int_0^a f(x)dx \qquad ***$$

$\therefore$ From * and **, for even function

$$\int_{-a}^a f(x)dx = \int_{-a}^0 f(x)dx + \int_0^a f(x)dx$$

$$= \int_0^a f(x)dx + \int_0^a f(x)dx$$

$$= 2\int_0^a f(x)dx$$

and from * and ***, for odd function

$$\int_{-a}^a f(x)dx = \int_{-a}^0 f(x)dx + \int_0^a f(x)dx$$

$$= -\int_o^a f(x)dx + \int_o^a f(x)dx$$

$$= 0$$

SOLVED EXAMPLES

8.9. Evaluate

$$\int_{-1}^{1} \frac{e^x}{1+e^x} dx$$

Solution:

Let $1 + e^x = z \Rightarrow e^x dx = dz$

Here $$\int \frac{e^x}{1+e^x} dx = \int \frac{1}{z} dz = \log_e z + c$$

$$= \log_e \left(1+e^x\right) + c$$

$\therefore$ $$\int_{-1}^{1} \frac{e^x}{1+e^x} dx = \left[\log_e \left(1+e^x\right) + c \Big|_{-1}^{1}\right]$$

$$= \log_e \left(1+e^1\right) - \log_e \left(1+e^{-1}\right)$$

$$= \log_e (1+e) - \log_e \left(\frac{e+1}{e}\right)$$

$$= \log_e \left(\frac{1+e}{1+e}\right).e = 1$$

8.10. Evaluate

$$\int_{-1}^{1} \frac{1}{4-3x} dx$$

Solution:

Here $\frac{1}{4-3x} dx$ Let $4\text{-}3x = z \Rightarrow -3dx = dz$

Here $$\int \frac{1}{4-3x} dx$$

$$= \int \frac{1}{z} \cdot \left(\frac{dz}{-3}\right) = -\frac{1}{3} \int \frac{1}{z} dz$$

$$= -\frac{1}{3} \log_e z + c$$

$$= -\frac{1}{3}\log_e(4-3x)+c$$

$$\therefore \qquad \int_{-1}^{1}\frac{1}{4-3x}dx = -\frac{1}{3}\log_e(4-3x)+c\Big|_{-1}^{1}$$

$$= \left(-\frac{1}{3}\right)\left[\log_e(4-3.1)-\log_e\{4-3(-1)\}\right]$$

$$= \left(-\frac{1}{3}\right)\left[\log_e(1)-\log_e(7)\right]$$

$$= \left(-\frac{1}{3}\right)\log_e\left(\frac{1}{7}\right)$$

8.11. Evaluate

$$\int_{-1}^{1} x^7.e^{\frac{x^2}{2}}dx$$

Solution:

Here $f(x) = x^7.e^{\frac{x^2}{2}}$ is an odd function as $f(-x) = (-x)^7 \; e^{\frac{(-x)^2}{2}}$

$$= -x^7 e^{\frac{x^2}{2}} = -f(x)$$

$$\therefore \qquad \int_{-1}^{1} f(x)dx = \int_{-1}^{1} x^7 e^{\frac{x^2}{2}}dx = 0$$

8.12: Find y_0; if $\int_0^{\infty} y_o.\overline{e}^{\frac{x}{\sigma}}dx = 1$ (AHSEC, 97)

Solution:

$$\int_0^{\infty} y_o.e^{-\frac{x}{\sigma}}dx = y_0\int_0^{\infty} e^{-\frac{x}{\sigma}}dx$$

$$= y_0\left(\frac{e^{-\frac{x}{\sigma}}}{-\frac{1}{\sigma}}\Big|_0^{\infty}\right) = \sigma y_0\left(-e^{-\infty}+e^0\right) = \sigma.y_0$$

$$\therefore \quad \int_0^{\infty} y_o.e^{-\frac{x}{\sigma}}dx = 1$$

$$\Rightarrow \quad \sigma.y_0 = 1$$

$$\Rightarrow \quad y_0 = \frac{1}{\sigma}$$

8.13. Evaluate $\int_{-\alpha}^{\alpha} y_o.e^{-|x|}dx$ **(AHSEC, 94)**

Solution: $\int_{-\alpha}^{\alpha} y_o.e^{-|x|}dx$

$$= \int_{-\alpha}^{0} y_o.e^{-|x|}dx + \int_{0}^{\alpha} y_o.e^{-|x|}dx$$

Now for the Ist part; $-\alpha < x < 0; so, -|x| = $ x. For the 2nd part $0 < x < \alpha$; so, $-|x| = -x$. Hence

$$\int_{-\alpha}^{0} e^{-|x|}dx + \int_{0}^{\alpha} e^{-|x|}dx$$

$$= \int_{-\alpha}^{0} e^{x}dx + \int_{0}^{\alpha} e^{-x}dx$$

$$= \left(e^{x}\Big|_{-\alpha}^{0}\right) + \left(-e^{-x}\Big|_{0}^{\alpha}\right)$$

$$= \left(e^{0} - e^{-\alpha}\right) + \left(-e^{-\alpha} + e^{0}\right) = 2$$

8.14. Evaluate $\int_0^1 \frac{x^2}{\sqrt{1-x^3}}dx$ **(AHSEC, 95)**

Solution: Here let $1-x^3 = z$

$$\Rightarrow \quad 3x^2dx = dz$$

$$\therefore \quad \int \frac{x^2}{\sqrt{1-x^3}}dx = \int\left(-\frac{1}{3}\right)\frac{dz}{\sqrt{z}}$$

$$= -\frac{1}{3}\int z^{-\frac{1}{2}}dx$$

$$= \left(-\frac{1}{3}\right) \cdot \frac{z^{-\frac{1}{2}+1}}{-\frac{1}{2}+1} + c$$

$$= \left(-\frac{1}{3}\right) 2.\sqrt{z} + c$$

$$= \left(-\frac{2}{3}\right) \sqrt{1-x^3} + c$$

$$\therefore \quad \int_0^1 \frac{x^2}{\sqrt{1-x^3}} dx$$

$$= \left(-\frac{2}{3}\right) \sqrt{1-x^3} + c \Big|_0^1$$

$$= \left(-\frac{2}{3}\right) \left(\sqrt{0} - \sqrt{1}\right) = \frac{2}{3}$$

8.15. Evaluate $\int_3^9 \frac{e^{\log_e x}}{x} dx$ **(AHSEC, 00)**

Solution : Let $\log_e x = u$

$$\Rightarrow \quad \frac{1}{x} dx = du$$

$$\therefore \quad \int \frac{e^{\log_e x}}{x} dx = \int e^u du$$

$$= e^u + c = e^{\log_e x} + c$$

$$\therefore \quad \int_3^9 \frac{e^{\log_e x}}{x} dx = e^{\log_e x} + c \Big|_3^9$$

$$= \left(x + c \Big|_3^9\right) = 9 - 3 = 6$$

8.16. Find K if $\int_0^\alpha K_e^{-\frac{x}{2}} dx = 1$ **(AHSEC, 06)**

Solution: $\int_0^\alpha K_e^{-\frac{x}{2}} dx = 1$

$$\Rightarrow K\left[\frac{e^{-\frac{x}{2}}}{-\frac{1}{2}}\right]_0^{\infty} = 1$$

$$\Rightarrow (-2K)\left(e^{-\frac{\alpha}{2}} - e^{o}\right) = 1$$

$$\Rightarrow (-2K)(-1) = 1$$

$$\Rightarrow K = \frac{1}{2}$$

8.17. Evaluate $\int_0^1 x^2(1-x)dx$ **(AHSEC, 06)**

Solution: $\int_0^1 x^2(1-x)dx$

$$= \int_0^1 x^2 dx - \int_0^1 x^3 dx$$

$$= \left.\frac{x^{2+1}}{2+1}\right|_0^1 - \left.\frac{x^{3+1}}{3+1}\right|_0^1$$

$$= \frac{1^3 - 0^3}{3} - \frac{1^4 - 0^4}{4}$$

$$= \frac{4-3}{12}$$

$$= \frac{1}{12}$$

EXERCISE–8

Evaluate

1. $\int \frac{e^x - 1}{e^x + 1} dx$ **(Ans.** $2\log_e\left(e^{\frac{x}{2}} + e^{-\frac{x}{2}}\right) + c$ **)**

2. $\int \frac{x^2}{x^2 - 25} dx$ **(Ans.** $x + \frac{5}{2}\log_e \frac{x-5}{x+5} + e$ **)**

3. $\int \log_e x dx$ **(Ans.** $x\log_e x - x + c$ **)**

4. $\int (3x^2 + 4x + 5) dx$ **(Ans.** $x^3 + 2x^2 + 5x + c$ **)**

5. $\int \left(x + \frac{1}{x^2}\right) dx$ **(Ans.** $\frac{x^2}{2} - \frac{1}{x} + c$ **)**

6. $\int \frac{x^3 + 2x^2 + 4}{x} dx$ **(Ans.** $\frac{x^3}{3} + x^2 + 4\log_e x + c$ **)**

7. $\int \left(\sqrt{x} + \frac{1}{\sqrt{x}}\right)^2 dx$ **(Ans.** $\frac{x^2}{2} + \log_e x + 2x + c$ **)**

8. $\int \frac{\log x^2}{x} dx$ **(Ans.** $(\log x)^2 + c$ **)**

9. $\int \frac{16}{16 - x^2} dx$ **(Ans.** $2\log_e \frac{4+x}{4-x} + c$ **)**

10. $\int \frac{x^2}{\sqrt{x^3 - 1}} dx$ **(Ans.** $\frac{2}{3}\sqrt{x^3 - 1} + e$ **)**

11. $\int_2^3 (5x + 4) dx$ **(Ans.** $\frac{33}{2}$ **)**

12. $\int_0^\infty x e^{-ax} dx$ **(Ans.** $\frac{1}{a^2}$ **)**

13. $\int_0^\infty x^2 e^{-ax} dx$ **(Ans.** $\frac{2}{a^3}$ **)**

14. $\int_{-a}^{a} x e^{x^2} dx$ **(Ans. 0)**

15. $\frac{1}{\sqrt{2\pi}}\int_{-\alpha}^{0} e^{-\frac{x^2}{2}}dx$ **(Ans.** $\frac{1}{2}$**)**

16. $\int\frac{2x^2+3x+4}{\sqrt{x}}\ dx$ **(AHSEC, 08)**

(Ans. $\frac{4}{5}x^{5/2}2x^{3/2}+8x^{1/2}+e$**)**

APPENDIX

AHSEC QUESTION PAPER – 2003
STATISTICS

1. Write down the first three terms in the expansion of $\sqrt{e}$. 1

2. If $\frac{d}{dx}[F(x)] = f(x)$, write down the expression for $\int f(x)dx$. 1

3. Under what condition does $P(A|B) = P(A)$ holds? 1

4. If ϕ is the empty set, then show that $P(\phi) = 0$. 1

5. Which component of time series is mainly applicable in the following case?

   Weekly sales of cold drinks. 1

6. State the order in which mean, median and mode of a distribution lie for a positively skew frequency distribution. 1

7. What is the difference between $f(a)$ and $\underset{x \to a}{Lt} f(x)$? 2

8. Draw the graph of the following function: 2

$$f(x) = \frac{|x-2|}{x-2}, \qquad \text{when } x \neq 2$$
$$= 0, \qquad \text{when } x = 2$$

9. If S_1, S_2, S_3 be respectively the sum of the first n, $2n$, $3n$ terms of an AP, show that $S_3 = 3(S_2 - S_1)$. 2

10. Define sample point and sample space. 2

11. A coin is tossed three times. Find the probability of getting two or more heads consecutively. 2

12. If $x_i | f_i (i = 1, 2, ..., n)$ is the frequency distribution, prove that $\Sigma f_i (x_i - \overline{x}) = 0$, $\overline{x}$ being the mean of the distribution. 2

13. Show that

    AM $\geq$ GM 2

14. Explain the meaning of the statement—

    'Net reproduction rate is 1105 per thousand'. 2

15. Find the term independent of x in the expansion of $\left(x - \frac{1}{x^2}\right)^{3n}$. 4

16. Evaluate: **2+2=4**

(i) $\underset{x \to 1}{Lt} \dfrac{x^2 - 3x + 2}{x^2 - 4x + 3}$

(ii) $\dfrac{d}{dx}\left(\dfrac{1+\sqrt{x}}{1-\sqrt{x}}\right)$

17. Evaluate: **2+2=4**

(i) $\int \dfrac{dx}{x^2 - a^2}$

(ii) $\int_0^1 \dfrac{x\,dx}{x \log x}$

18. What is secondary data? Name the sources from which this type of data can be collected? **1+3=4**

19. A number is chosen from each of the two sets—

1, 2, 3, 4, 5, 6, 7, 8, 9

1, 2, 3, 4, 5, 6, 7, 8, 9

If p_1 denotes the probability that the sum of the two numbers be 10 and p_2 be the probability that their sum be 8, find $p_1 + p_2$. 4

20. A committee of 5 is to be selected out of 6 gentlemen and 4 ladies. What is the probability that the committee contains at least 2 ladies? 4

21. State the multiplication theorem of probability.
Let A and B be two events with $P(A) = 1/2$, $P(B) = 1/3$ and $P(A \cap B) = 1/4$. Find $P(A|B)$ and $P(A^C|B^C)$. **1+3=4**

22. (a) If $\dfrac{\log a}{y-z} = \dfrac{\log b}{z-x} = \dfrac{\log c}{x-y}$, then show that

$a^x.b^y.c^z = 1$ **3**

(b) In how many ways can the letters of the word INSURANCE be arranged such that the vowels remain together? **3**

23. (a) (i) Represent the following data by a suitable diagram:

Population of four towns in three censuses

Town	*Population* 1961	*Population* 1971	*Population* 1981
A	20,480	30,500	50,870
B	25,680	32,800	60,730
C	30,470	40,500	80,300
D	32,600	45,000	1,20,000

(ii) The AM and standard deviation of a series of 20 items were calculated by a student as 20 cm and 5 cm respectively. But while calculating them one item 13 was misread as 30. Find the correct AM and standard deviation. **2+2=4**

(b) (i) Draw a histogram and frequency polygon from the following data.

(ii) Find the modal wage from the graph and check the value by direct calculation. **2+2+3=7**

Wages (in Rs.)	*No. of Workers*
10–15	50
15–20	140
20–25	110
25–30	150
30–35	120
35–40	100
40–45	80

(c) (i) Define standard deviation. Why is standard deviation called ideal measure of dispersion? **3**

(ii) Calculate coefficient of variation from the following data: **4**

Marks	*No. of Students*
Below 20	8
Below 40	20
Below 60	50
Below 80	70
Below 100	80

Or

The first four moments of a distribution about $x = 4$ are 1, 4, 10 and 45. Obtain the various characteristics of the distribution on the basis of the information given. Comment upon the nature of the distribution.

24. (a) What do you mean by trend in a time series? Estimate the trend values by using the data given below by taking a four-yearly moving average: **2+4=6**

Year	*Values*	*Year*	*Values*
1964	12	1971	100
1965	25	1972	82
1966	39	1973	65
1967	54	1974	49
1968	70	1975	34
1969	87	1976	20
1970	105	1977	7

(b) What are the components of time series? Discuss them with examples. **2+4=6**

25. (a) Define the following terms: **1×4=4**

(i) Vital statistics

(ii) Age specific death rate

(iii) Total fertility rate

(iv) Standardised death rate

(b) Define NRR. Interpret the results

NRR $< = > 1$ **2+3=5**

(c) Explain why standardised death rates of different populations are comparable while crude death rates are not. **3**

QUESTION PAPER – 2004
STATISTICS

1. Answer the following: **1×10=10**

(a) Write down the expansion of $\log_e (1+x)$ mentioning the condition of validity.

(b) Explain the meaning of the statement

$$P(A) = \frac{1}{2}$$

(c) State symbolically the relation between the probability of an event and the probability of its complementary event.

(d) Under what condition the equality $P(A + B) = P(A) + P(B)$ holds?

(e) Define Geometric Mean.

(f) Which component of time series is mainly applicable in the following case?

An era of prosperity.

(g) Can two or more histograms be shown on the same graph?

(h) Interpret the result

NRR = 1

(i) Define $y = f(x)$.

(j) Fill in the blank:

For a mesokurtic curve, β_2 equal to _____.

2. If a, b, c are in AP and x, y, z in GP, prove that

$x^{b-c}.y^{c-a}.z^{a-b} = 1$ **2**

3. Draw the graph of the following function: **2**

$$f(x) = \frac{|x|}{x}; \quad x \neq 0$$

4. A coin and a die are tossed together. Find the chance of getting either 'a head and 5' or 'a tail and 6'. **2**

5. Define conditional probability. **2**

6. Distinguish between population and sample. **2**

7. Write a note on coefficient of variation. **2**

8. What do you mean by vital statistics? Explain. **2**

9. Show that

$$2 < e < 3$$ **2**

10. Write down the formulae for crude death rate and standardised death rate.

2

11. Distinguish between primary and secondary data. **2**

12. Evaluate: **2+2=4**

(i) $\underset{x \to 0}{Lt} \dfrac{\sqrt{1+ax} - \sqrt{1-ax}}{x}$

(ii) $\dfrac{d}{dx}[x.2^x]$

13. (a) Show that the number of different triangles which can be formed by joining the angular points of a polygon on n sides is $\frac{1}{6}n(n-1)(n-2)$. Show also that the figure has $\frac{1}{2}n(n-3)$ diagonals. **3+1=4**

(b) Show that there will be ɔ term containing x^9 in the expansion of

$$\left(2x^2 - \frac{1}{x}\right)^{20}.$$ **4**

14. (a) If A and B are any two events, then show that

$$P(A \cup B) = P(A) + P(B) - P(A \cap B)$$ **4**

(b) If A and B are two events such that $P(A) = 0.4$, $P(A \cup B) = 0.7$ and $P(B) = P$. For what value of P are A and B (i) mutually exclusive and (ii) independent? **2+2=4**

15. (a) Explain positive and negative association between two attributes.

Given:

$N = 100$, $(A) = 470$

$(B) = 620$, (AB) $= 320$

Examine the nature of association. **1+1+2=4**

(b) Represent the following data by a suitable diagram: **4**

Items of Expenditure	*% of Total Expenditure*
Food	60
Clothing	15
Rent	12
Fuel and lighting	5
Miscellaneous	8

16. (a) Prepare a blank table to show the distribution of population of a State of India according to age, sex and literacy. **3**

(b) "GRR is a hypothetical figure." Discuss. **3**

17. (a) Enumerate the objective of analysis of time series. **3**

(b) Determine trend by 3-yearly moving average for the following data: **3**

Year	*Production (in metric tonnes)*
1973	12
1974	21
1975	30
1976	36
1977	42
1978	46
1979	50
1980	56
1981	63
1982	70

18. Evaluate: **2+3=5**

(i) $\int \frac{dx}{e^x + 1}$

(ii) $\int_a^b \frac{\log x}{x} dx$

19. Arrange $P(A)$, $P(A \cap B)$, $P(A \cup B)$, $P(A) + P(B)$ in ascending order of magnitude. **5**

20. Define a time series. What are its components? Explain any one of them. **2+1+2=5**

21. The first three moments of a distribution about the value 2 of the variable are 1, 16 and –40. Show that the mean is 3, the variance is 15 and the third moment about the mean, μ_3 is -86. Comment on the skewness of the distribution. **3+2=5**

22. Find the mean and variance of the following distribution: **2+3=5**

Measurement	*No. of articles*
More than 80	5
More than 70	14
More than 60	34
More than 50	65
More than 40	110
More than 30	150
More than 20	170
More than 10	176
More than 0	180

23. Calculate the crude and standardised death rates of the local population: **2+3=5**

Age	*Standard Population*	*Local Population*	*No. of deaths*
0–10	600	400	16
10–20	1000	1500	6
20–60	3000	2400	24
60–100	400	700	21

Or

Define crude birth rate. Discuss whether it is an adequate measure of fertility. **2+3=5**

QUESTION PAPER – 2005
STATISTICS

1. Answer the following: **1×10=10**

(a) State the relation between ${}^nC_r$ and ${}^{n-1}C_{r-1}$.

(b) Examine the correctness of the statement:
Standard deviation is not affected by the change of origin.

(c) For a moderately skew distribution, what relation between mean, median and mode exists?

(d) When will the conditional probability P(A|B) be equal to P(A)?

(e) Find the range of variation of the following values:
Weight (kg) : 40, 51, 47, 39, 60, 48, 64, 51, 57

(f) Which component of time series is mainly associated in the following case?
The traffic density on a particular road at different hours of the day.

(g) Name a measure of location which is not a measure of central tendency.

(h) What is meant by an impossible event?

(i) If $\frac{d}{dx}[F(x)] = f(x)$, write down the expression for $\int f(x)dx$.

(j) Define vital statistics rate.

2. If $a^x = b^y = c^z$, $b^2 = ac$, prove that

$$\frac{1}{x} + \frac{1}{z} = \frac{2}{y}$$ **2**

3. Draw the graph of the following function: **2**

$$f(x) = \frac{|x-2|}{x-2}; \quad x \neq 2$$
$$= 0; \quad x = 2$$

4. Evaluate: **2**

$$\underset{x\to 1}{Lt} \frac{\sqrt{1+x} - \sqrt{1-x}}{\sqrt{1+x} + \sqrt{1-x}}$$

5. Distinguish between questionnaire and schedule. **2**

6. What are the merits and demerits of the method of moving averages?

7. State the limitations of statistics. **2**

8. With both P(A) > 0 and P(B) > 0, can two mutually exclusive events A an B be independent?
2

9. Give the statistical definition of probability of an event.

10. The first three moments of a distribution about the value 2 of the variable 1, 16 and –40. Find μ_2 and μ_3 of the distribution. **2**

11. Prove that for the n observations $x_1, x_2, ..., x_n$ with arithmetic mean $\bar{x}$

$$\sum_{i=1}^{n}(x_i - \bar{x}) = 0$$ **2**

12. If the constant term in the expansion of $\left(\sqrt{x} + \frac{k}{x}\right)^6$ is 135, then find the value of k. **4**

13. Find $\frac{dy}{dx}$:

$$\frac{1+\sqrt{x}}{1-\sqrt{x}}$$ **4**

14. Evaluate:

$$\int_0^2 \frac{2x+5}{x^2+5x+7}\,dx$$ **4**

15. If $A_1, A_2, ..., A_n$ are n mutually exclusive and equally likely events, then show that

$$P(A_1 \cup A_2 \cup ... \cup A_n) = P(A_1) + P(A_2) + ... + P(A_n)$$ **4**

16. Two unbiased dice are thrown. Find the probability that—

(a) both the dice show the same number;

(b) the total of the numbers on the dice is greater than 8. **2+2=4**

17. Represent the following data by a suitable diagram: **4**

Number of engineering students graduated from a college during 1980 to 1983

Year	*Civil*	*Electrical*	*Mechanical*	*Total*
1980	20	20	10	50
1981	20	25	10	55
1982	24	24	12	60
1983	25	25	20	70

18. What is meant by attribute? When two attributes A and B are said to be independent? Examine the consistency of the following data:

$N = 1000$, $(A) = 600$, $(B) = 500$, $(AB) = 50$ **1+1+2=4**

19. Prove that

$$\text{AM} \geq \text{GM} \geq \text{HM}$$ **4**

20. A study of demand (d_t) for the past 12 years (t = 1, 2, 3, ..., 12) has indicated the following:

$$\begin{aligned} d_t &= 100; \ (t = 1, 2, ..., 5) \\ &= 20; \ (t = 6) \\ &= 100; \ (t = 7, 8, ..., 12) \end{aligned}$$

Compute a 5-year moving average. **3**

21. Define NRR and interpret the results. **1+2=3**

(a) NRR = 1

(b) NRR < 1

(c) NRR > 1

22. From the following data, compute gross reproduction rata: **3**

Age group of mother	*No. of women ('000)*	*Total Births*
15–19	16.0	260
20–24	16.5	2200
25–29	15.8	1895
30–34	15.0	1320

35–39	14.8	915
40–44	15.0	280
45–49	14.5	145

Assume that the proportion of female births is 46 per cent.

23. If the coefficient of three consecutive terms in the expansion of $(1+x)^n$ are 462, 330 and 165 respectively, find the value of n.

24. (a) State the multiplicative theorem of probability. **2**

(b) Given, $P(A) = p$, $P(A|B) = q$, $P(B|A) = r$.

Find the relations between the numbers p, q, r for the following cases:

(i) Events A and B are mutually exclusive.

(ii) $\overline{A}$ and $\overline{B}$ are mutually exclusive. **4**

25. Find the mean, standard deviation and skewness of the distributionb: 6

Variable :	0–5	5–10	10–15	15–20	20–25	25–30	30–35	35–40
Frequency :	2	5	7	13	21	16	88	3

26. What are the components of time series? Explain them with examples. **2+4=6**

27. Define General fertility rate and total fertility rate. **2+2=4**

28. Is CDR an accurate measure of mortality of population of a country? Give reasons for your answer. **1+1=2**

QUESTION PAPER – 2006
STATISTICS

1. Answer/Fill up the gaps of the following: **1×10=10**

(a) Write the expansion of $(1+x)^n$.

(b) Which component of time series is mainly applicable in the following case?

"Monthly sales of cold drink and warm clothes."

(c) Write down the expansionof $-\log(1-x)$.

(d) If $\overline{x}$ is the AM of observations $x_1, x_2, ..., x_n$ then find the value of

$$\sum_{i=1}^{n}\left(x_i - \overline{x}\right)$$

(e) What is the value of Geometric Mean between the regression coefficients?

(f) What are the best measures of location and dispersion?

(g) In drawing histograms, the class interval should be____.

(h) The standard deviation is affected by the change of ___.

(i) The correlation coefficient is not affected by the changes of ___ and ___.

(j) The log of a number to itself as base is ____.

2. (a) If

$m = a^x, n = a^y, a^2 = (m^y n^x)^x$

show that $xyz = 1$. **2**

$m = a^x, n = a^y, a^2 = (m^y n^x)^z$

Or

Show that

$${}^nC_r + {}^nC_{r-1} = {}^{n+1}C_r$$

(b) If the sum of three consecutive terms of an AP is 27 and their product is 693, find the terms.

(c) Sum to infinity: **2**

$1 + 2a + 3a^2 + 4a^3 + \ldots \infty$

(d) Find the term independent of x in the expansion of

$$\left(x+\frac{1}{x}\right)^{10}$$

3. (a) Define $y = f(x)$. State the difference between

$f(a)$ and $\underset{x\to a}{Lt} f(x)$

(b) Evaluate : **1 ½ × = 3**

(i) $\underset{x\to 3}{Lt}\frac{x^2-9}{x-3}$

(ii) $\underset{x\to\infty}{Lt}\left(\sqrt{x+1}-\sqrt{x}\right)$

(c) Find $\frac{dy}{dx}$ (any *two*) :

(i) $y = \log x$

(ii) $2x^2(x-1)$

(iii) $y = e^{-x}$

(d) Find k if

$\int_0^\infty ke^{-x/2}dx = 1$ **2**

(e) Find the value of (any *one*) :

(i) $\int xe^x dx$

(ii) $\int_0^\infty x^2 (1-x)dx$

4. (a) What are the Statistics and Statistical data? Define sample and population. **2 + 2 = 4**

(b) (i) Write the characteristics of an ideal average. **2**

(ii) Find the mean and variance of the following distribution : **4**

Marks	No. of Students
Less than 10	6
Less than 20	20
Less than 30	35
Less than 40	40

(c) What are the different measures of dispersion? **2**

"Standard deviation is the best measure of dispersion." Explain.

5. (a) What is secondary data? Name the sources from which this type of data can be collected. 1 + 3 = 4

(b) Draw a histogram and frequency polygon from the following data:

Wages (in Rs.)	No. of workers
10 – 15	40
15 – 20	120
20 – 25	90
25 – 30	140
30 – 35	120
35 – 40	80

6. (a) Explain the following : 2 × 3 = 6

(i) Coefficient of variation

(ii) Skewness

(iii) Kurtosis

(b) Define Karl Pearson's correlation coefficient. State its properties. 1 + 3 = 4

(c) Show that for independent variables correlation coefficient is zero. Write down the equations of two regression lines. 2 + 2 = 4

7. Answer any *two* questions :

(a) Discuss briefly the various steps involved in the construction of consumer price index number. **6**

(i) Random experiment

(ii) Event

(iii) Sample space

(iv) Probability

(b) Define an index number and state its uses. 2 + 4 = 6

(c) What are the tests of a good index number? Why is Fisher's index number? Why is Fisher's index number called an ideal? 3 + 3 = 6

8. (a) Define (any *three*) : 2 × 3 = 6

(i) Vital statistics

(ii) Standardised death rate

(iii) Crude birthrate

(iv) Age-specific death rate

(b) What are vital rate and ratio? Show that NRR cannot exceed GRR. 3 + 3 =6

9. (a) What is Time Series? State its uses. 2 + 3 = 5

(b) State the various methods of studying trend in a time series. Estimate the trend values by using the following data by taking a 4-yearly moving average: 2 + 4 = 6

Year	Values	Year	Values
1970	60.0	1975	48.2
1971	46.5	1976	42.6
1972	53.0	1977	51.7
1973	54.5	1978	51.1
1974	48.9	1979	43.8

QUESTION PAPER – 2007
STATISTICS

1. Answer/Fill up the gaps of the following : 1 × 10 = 10

(a) Write down the expansion of the following :

(i) $\log_e (1 - x)$ Or (ii) $(1 + x)^n$

(b) Given that ${}^{12}C_r = {}^{12}C_3$, find if r $\neq$ 3.

(c) Given ${}^nC_2 = 42$, find ${}^nC_2$.

(d) Write down the expansion of e^{-ax}.

(e) Write the value of $\sqrt{AM \times HM}$, where AM is the Arithmetic Mean and HM is the Harmonic Mean of two values x_1 and x_2.

(f) Standard deviation is a measure of ___.

(g) In which case is mean = median = mode?

(h) What is Index number?

Or

(i) Define vital rate or vital ratio.

(j) Write the names of components of a Time Series.

2. (a) Find the 20th term of the AP series 3, 5, 7, ... 2

(b) Prove that

$$^nC_0 + ^nC_1 + \ldots + ^nC_n = 2^n$$ 2

(c) If

$a^x = b^y c^z$ and $b^2 = ac$

Prove that

$$\frac{1}{x} + \frac{1}{z} = \frac{2}{y}$$ 2

Or

Find the general term of $\left(x^2 + \frac{3}{x}\right)^6$.

(d) In how many ways can the letters of the word 'ACTION' be arranged so that vowels remain together?

(e) The function $f(x)$ is defined as below: 3

$$f(x) = \frac{|x-2|}{x-2}, \quad \text{when } x \neq 2$$
$$= 0, \quad \text{when } x = 2$$

Draw the graph of the function.

(f) Evaluate :

$$\underset{x \to 0}{Lt} \frac{a - \sqrt{a^2 - x^2}}{x^2}$$

3. (a) Find $\frac{dy}{dx}$ (any *one*) :

(i) $y = e^x x^n$

(ii) $y = x \cdot 2^x$

(b) Find the value (any *one*) :

(i) $\int \frac{dx}{x^2 - 4}$

(ii) $\int x^2 e^x dx$

(c) Distinguish between (any *one*) : 3

(i) Primary and Secondary data

(ii) Quantitative data and Qualitative data

(d) What is mean? Show that for a frequency distribution

$$\sum_{i-1}^{n} f_i(x_i - \bar{x}) = 0, \text{ where } f_i \text{ is frequency.}$$

(e) Which is the best measure of location? Give reasons. 3

(f) Find the mean and GM of 1, 2, 8, 16. 3

4. Answer the following :

(a) Prove that

AM $\geq$ Gm $\geq$ HM 4

wages (in Rs.) :	10 – 15	15 – 20	20 – 25	25 – 30
No. of Workers :	40	120	190	150

(d) Explain (any *two*) : 4

Quartiles, Ogive, Declies, Percentiles.

(e) Explain : 4

Skewness and Kurtosis

5. (a) What are correlation and regression? Show that $-1 \leq r \leq 1$, r = correlation coefficient. 5

(b) Give the idea of an index number. Define simple and weighted index number. 5

6. (a) Answer the following :

(i) What are the important sources of vital statistics? 2

(ii) Define—General fertility rate, total fertility rate, crude death rate. 3

(b) (i) Calculate standardised death rate from the data given below : 4

Age	Standard Population ('000)	Population A: Population ('000)	Population A: No. of Deaths
0–10	20	12	300
10–20	12	20	600
20–40	60	64	1600
40–60	20	25	500
Above 60	10	3	150

(ii) Interpret the results : 2

NRR = 1, NRR < 1, NRR > 1

(c) Define moving averages. What are the merits and demerits of the method of moving averages? 5

(d) Calculate trend values by the method of least square from the data given below: 6

Year :	1986	1987	1988	1989	1990
Sales (Rs.) :	80	84	90	96	100

Calculate (i) Laspeyre's Price Index and (ii) Laspeyre's Quantity Index Numbers from the given data, taking 2001 as base year :

Commodity	2001		2003	
	Price	Quantity	Price	Quantity
A	4	10	5	12
B	6	8	7	10
C	10	5	12	4
D	3	12	4	15
E	5	7	5	8

QUESTION PAPER – 2008
STATISTICS

1. Answer/Fill up the gaps of the following : $1 \times 10 = 10$

(a) Find the value of :

$$10^{-2\log_{10}3}$$

(b) Write down the expansion of :

$$e^{-1}$$

(c) If ${}^nP_r = 336$ and ${}^nC_r = 56$, then find the value of r.

(d) The value of variance is always ____.

(e) If $U = \dfrac{x-a}{h}$, then $\bar{x}$ = ____.

(f) What do you mean by frequency?

(g) The sum of deviations of the values from their mean is always ____.

(h) Why Fisher's index number is known as ideal index number?

(i) What is the relationship between correlation coefficient and regression coefficients?

(j) In case of a skewed distribution, what happens to mean, median and mode?

2. (a) How many words can be formed out of the letters ARTICLE, so that the vowels always occupy the even places? 2

(b) Expand :

$$\left(x^2 + \frac{y}{2}\right)^6$$

(c) Prove that

$$\frac{a-b}{a} + \frac{1}{2}\left(\frac{a-b}{a}\right)^2 + \frac{1}{3}\left(\frac{a-b}{a}\right)^3 + \ldots\infty$$

$= \log_e a - \log_e b$ 2

(d) Obtain the sum of the series : 2

$$\frac{1}{2!} - \frac{1}{3!} + \frac{1}{4!} - \frac{1}{5!} + \frac{1}{6!} - \dots$$

(e) Draw the graph of the function : 3

$$f(x) = x - 1 \text{ when } x > 0$$

$$= \frac{1}{x} \text{ when } x = 0$$

$$= x + 1 \text{ when } x < 0$$

(f) Evaluate :

$$\underset{x \to 1}{Lt} \frac{x^3 - 3x^2 + 2}{x^2 + 5x - 6}$$

3. (a) Find the derivative of : 3

$$\log \sqrt{\frac{1 - x^2}{1 + x^2}}$$

(b) Evaluate :

$$\int \frac{2x^2 + 3x + 4}{\sqrt{x}} dx$$

(c) Distinguish between (any *one*) :

(i) Time series data and Spatial data

(ii) Measures of central tendency and Measures of dispersion

(d) What are the main limitations of statistics? 3

(e) If $U = \frac{x - a}{h}$, then prove that $\sigma_x = h \cdot \sigma_v$. 3

(f) The AM and SD of 100 observations are 50 and 10 respectively. Find the new AM and SD if 2 is added to each observation. 3

4. (a) Briefly discuss various methods of collection of primary data. 4

(b) Explain the method of locating median graphically. 4

(c) Write a note on skewness and kurtosis. 4

(d) Explain :

(i) Coefficient of variation

(ii) Time series

(e) Draw a histogram and frequency curve from the following data: 4

Marks	No. of Students
0–10	4

10–20	8
20–30	11
30–40	15
40–50	12
50–60	7
60–70	3

5. (a) Show that correlation coefficient is independent of change of origin and scale. 5

(b) Show that Fisher's index number satisfies both time reversal and factor reversal tests. 5

(c) Define vital event, vital rate and vital ratio with examples. 5

(d) Briefly describe vasrious measures of mortality.

6. (a) Define tend of a time series. Determine trend of the following time series by the mehtod of semi-averages : $2 + 4 = 6$

Year	Sales (Rs.)
1943	38
1944	41
1945	45
1946	48
1947	52
1948	56
1949	60
1950	64
1951	68
1952	73
1953	79

(b) "Cost of living index measures the purchasing power of money." Explain. Calculate cost of living index number from the following data : $2 + 4 = 6$

Items	Weight	Price Relatives
Food	24.9	1.02
House rent	27.5	1.28
Clothing	9.4	0.83
Fuel	10.8	1.14
Transportation	5.3	1.0
Miscellaneous	22.1	1.18

(c) Give examples each of :

(i) Trend

(ii) Seasonal variation

(iii) Cyclical variation
(iv) Irregularity
(v) Fertility rate
(vi) Reproduction rate

ANTILOGARITHMS.

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

ANTILOGARITHMS.

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

LOGARITHMS

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						5	9	13	17	21	26	30	34	38
						0212	0253	0294	0334	0374	4	8	12	16	20	24	28	32	36
11	0414	0453	0492	0531	0569						4	8	12	16	20	23	27	31	35
						0607	0645	0682	0719	0755	4	7	11	15	18	22	26	29	33
12	0792	0828	0864	0899	0934						3	7	11	14	18	21	25	28	32
						0969	1004	1038	1072	1106	3	7	10	14	17	20	24	27	31
13	1139	1173	1206	1239	1271						3	6	10	13	16	19	23	26	29
						1303	1335	1367	1399	1430	3	7	10	13	16	19	22	25	29
14	1461	1492	1523	1553	1584						3	6	9	12	15	19	22	25	28
						1614	1644	1673	1703	1732	3	6	9	12	14	17	20	23	26
15	1761	1790	1818	1847	1875						3	6	9	11	14	17	20	23	26
						1903	1931	1959	1987	2014	3	6	8	11	14	17	19	22	25
16	2041	2068	2095	2122	2148						3	6	8	11	14	16	19	22	24
						2175	2201	2227	2253	2279	3	5	8	10	13	16	18	21	23
17	2304	2330	2355	2380	2405						3	5	8	10	13	15	18	20	23
						2430	2455	2480	2504	2529	3	5	8	10	12	15	17	20	22
18	2553	2577	2601	2625	2648						2	5	7	9	12	14	17	19	21
						2672	2695	2718	2742	2765	2	4	7	9	11	14	16	18	21
19	2788	2810	2833	2856	2878						2	4	7	9	11	13	16	18	20
						2900	2923	2945	2967	2989	2	4	6	8	11	13	15	17	19
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	4166	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

Mean Differences.

	0	1	2	3	4	5	6	7	8	9	1	2	3	4	5	6	7	8	9
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
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	4	4	5	6	6
63	7993	8000	8007	8014	8021	8028	8035	8041	8048	8055	1	1	2	3	3	4	5	6	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	9652	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