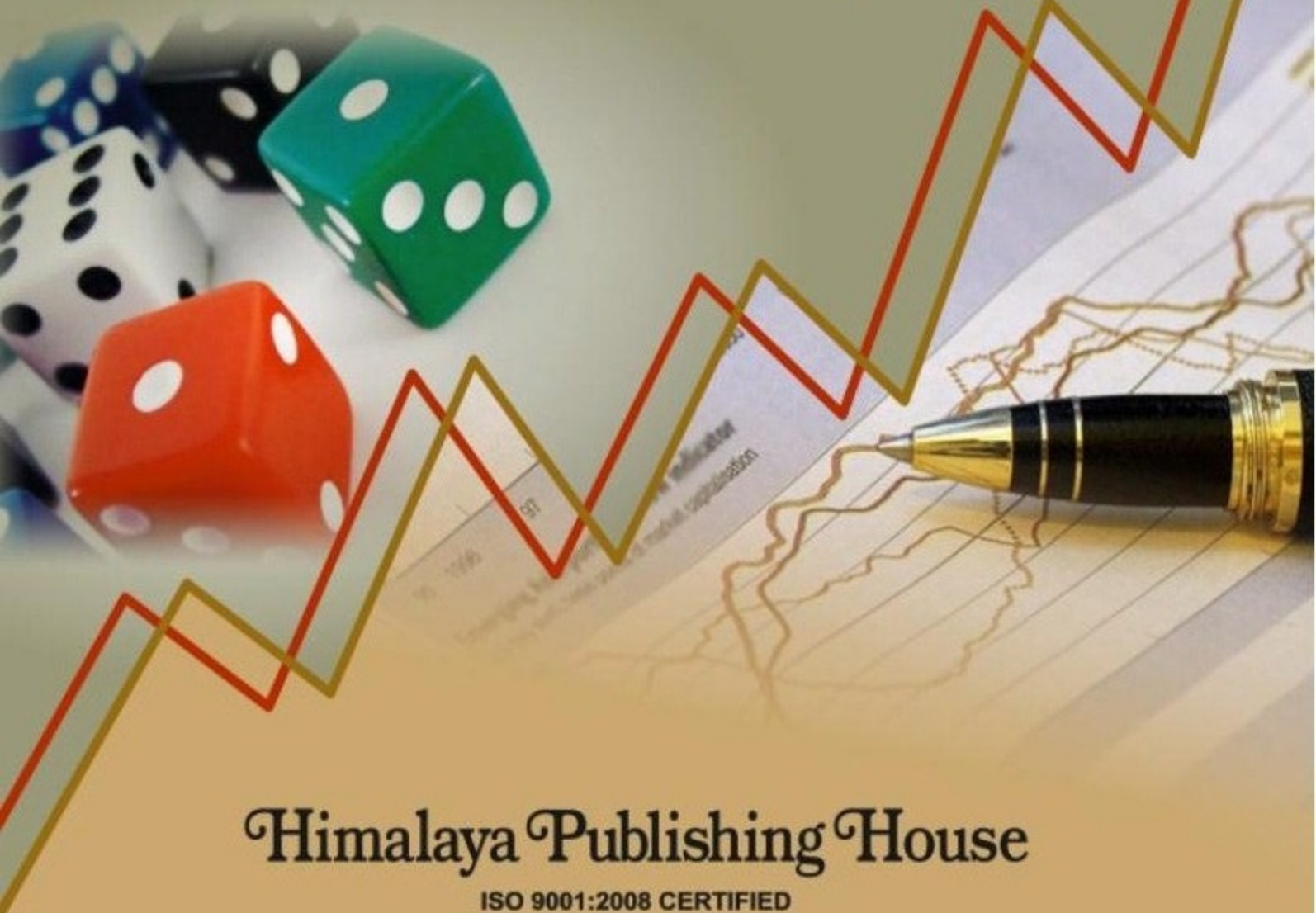


Business Statistics

Nazneen Khan Sarguroh
Dr. Abhilasha S. Magar



Himalaya Publishing House

ISO 9001:2008 CERTIFIED

BUSINESS STATISTICS

(As Per the Revised Syllabus 2016-17 of Mumbai University for
BMS, Semester - I)

Nazneen Khan Sarguroh

M.Sc. (Mathematics), B.Ed.

Lecturer,

Department of B.Sc. I.T. and Computer Science,
Rizvi College of Arts, Science and Commerce,
Bandra (W), Mumbai.

Dr. (Mrs.) Abhilasha S. Magar

M.Sc. (Maths), B.Ed. , Dip. C.P., Ph.D.

Department of Mathematics & Statistics,
Annasaheb Vartak College of Arts,
K.M. College of Commerce,
E.S. Andrades College of Science,
Vasai Raod (W), Thane – 401202.



Himalaya Publishing House

ISO 9001:2008 CERTIFIED

© **Author**

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.

First Edition : 2010

Second Edition : 2012

Third Edition : 2013

Fourth Revised Edition : 2014

Reprint : 2015

Fifth Edition : 2016

(as per Revised Syllabus)

Reprint : 2017

Edition : 2018

Published by	: Mrs. Meena Pandey for Himalaya Publishing House Pvt. Ltd. , Ramdoot, Dr. Bhalerao Marg, Girgaon, Mumbai - 400 004 Phone: 022-23860170/23863863; Fax: 022-23877178 E-mail: himpub@vsnl.com; Website: www.himpub.com
Branch Offices	:
New Delhi	: Pooja Apartments, 4-B, Murari Lal Street, Ansari Road, Darya Ganj, New Delhi - 110 002. Phone: 011-23270392, 23278631; Fax: 011-23256286
Nagpur	: Kundanlal Chandak Industrial Estate, Ghat Road, Nagpur - 440 018. Phone: 0712-2738731, 3296733; Telefax: 0712-2721216
Bengaluru	: Plot No. 91-33, 2nd Main Road, Seshadripuram, Behind Nataraja Theatre, Bengaluru - 560 020. Phone: 080-41138821; Mobile: 09379847017, 09379847005
Hyderabad	: No. 3-4-184, Lingampally, Besides Raghavendra Swamy Matham, Kachiguda, Hyderabad - 500 027. Phone: 040-27560041, 27550139
Chennai	: New No. 48/2, Old No. 28/2, Ground Floor, Sarangapani Street, T. Nagar, Chennai-600 012. Mobile: 09380460419
Pune	: First Floor, Laksha Apartment, No. 527, Mehunpura, Shaniwarpath (Near Prabhat Theatre), Pune - 411 030. Phone: 020-24496323, 24496333; Mobile: 09370579333
Lucknow	: House No. 731, Shekhupura Colony, Near B.D. Convent School, Aliganj, Lucknow - 226 022. Phone: 0522-4012353; Mobile: 09307501549
Ahmedabad	: 114, SHAIL, 1st Floor, Opp. Madhu Sudan House, C.G. Road, Navrang Pura, Ahmedabad - 380 009. Phone: 079-26560126; Mobile: 09377088847
Ernakulam	: 39/176 (New No. 60/251), 1st Floor, Karikkamuri Road, Ernakulam, Kochi - 682011. Phone: 0484-2378012, 2378016; Mobile: 09387122121
Bhubaneswar	: Plot No. 214/1342, Budheswari Colony, Behind Durga Mandap, Bhubaneswar - 751 006. Phone: 0674-2575129; Mobile: 09338746007
Kolkata	: 108/4, Beliaghata Main Road, Near ID Hospital, Opp. SBI Bank, Kolkata - 700 010, Phone: 033-32449649; Mobile: 07439040301
DTP by	: HPH, Editorial Office, Bhandup (Rajani Tambe)
Printed at	: Rose Fine Art, Mumbai. On behalf of HPH.

PREFACE

Almighty God has bestowed upon me opportunity to present this book namely "Business Statistics" for F.Y.B.M.S, Semester - I. It was my honest endeavour to equip this book within a framework of Mumbai Universities prescribed syllabus.

The obvious reason for writing this book is to cater to the needs of non-Statistical background of majority of students. Special attention was given while writing this book to explain the topics in simpler form. Every chapter concludes with exercise for the students to test their understanding of the topic and to gain confidence in the subject.

I dedicate this book to my beloved mother Sarguroh Saeeda Khan for being pillar of strength throughout my life. I owe much to my brother Sarguroh Yusuf Khan and my uncle Sarguroh Nadir Khan who were source of constant encouragement. I would like to thank my colleagues and friends for their support.

I was fortunate to have chosen Himalaya Publishing House to publish this book. Mr. S. K. Srivastava and others involved in editing and publishing this book have been meticulous, friendly and responsive also deserve my thanks.

I welcome any comment or suggestion for improvements or changes that could be implemented in possible future editions of this book. The email address for gathering feedback is sargurohnazneen@yahoo.co.in

Authors

SYLLABUS

Sr. No.	Modules	No. of Lectures
1.	Introduction to Statistics	15
2.	Measures of Dispersion, Co-Relation and Linear Regression	15
3.	Time Series and Index Number	15
4.	Probability and Decision Theory	15
	Total	60

Sr. No.	Modules/Units
1	<p>Introduction to Statistics</p> <ul style="list-style-type: none"> • Introduction: Functions/Scope, Importance, Limitations. • Data: Relevance of Data (Current Scenario), Type of Data (Primary and Secondary), Primary (Census vs. Samples, Method of Collection (in Brief), Secondary (Merits, Limitations and Sources) (in Brief). • Presentation of Data: Classification – Frequency Distribution – Discrete and Continuous, Tabulation, Graph (Frequency, Bar Diagram, Pie Chart, Histogram, Ogives). • Measures of Central Tendency: Mean (AM, Weighted, Combined), Median (Calculation and Graphical using Ogives), Mode (Calculation and Graphical using Histogram), Comparative Analysis of All Measures of Central Tendency.
2	<p>Measures of Dispersion, Correlation and Linear Regression</p> <ul style="list-style-type: none"> • Measures of Dispersion: Range with CR (Coefficient of Range), Quartiles and Quartile Deviation with CQ (Coefficient of Quartile), Mean Deviation from mean with CMD (Coefficient of Mean Deviation), Standard Deviation with CV (Coefficient of Variance), Skewness and Kurtosis (Only Concept). • Correlation: Karl Pearson, Rank Correlation. • Linear Regression: Least Square Method.
3	<p>Time Series and Index Number</p> <ul style="list-style-type: none"> • Time Series: Least Square Method, Moving Average Method, Determination of Season.

	<ul style="list-style-type: none"> • Index Number: Simple (Unweighted) Aggregate Method, Weighted Aggregate Method, Simple Average of Price Relatives, Weighted Average of Price Relatives, Chain Base Index Numbers, Base Shifting, Splicing and Deflating, Cost of Living Index Number.
4	<p>Probability and Decision Theory</p> <ul style="list-style-type: none"> • Probability: Concept of Sample Space, Concept of Event, Definition of Probability, Addition and Multiplication Laws of Probability, Conditional Probability, Bayes' Theorem (Concept Only), Expectation and Variance, Concept of Probability Distribution (Only Concept). • Decision Theory: Acts, State of Nature Events, Payoffs, Opportunity Loss, Decision Making under Certainty, Decision Making under Uncertainty. • Non-probability: Maximax, Maximin, Minimax, Regret, Laplace and Hurwicz). • Probabilitistics (Decision Making under Risk): EMV, EOL, EVPI. • Decision Tree



PAPER PATTERN

Maximum Marks: 75

Question to be Set: 05

Duration: $2 \frac{1}{2}$ Hours

All questions are compulsory carrying 15 Marks each.

Question No.	Particular	Marks
Q.1	Objective Questions (A) Sub-questions to be asked (10) and to be answered (any 08) (B) Sub-questions to be asked (10) and to be answered (any 07) (* Multiple Choice/True or False/Match the Columns, Fill in the Blanks)	15 Marks
Q.2	Full Length Practical Question OR	15 Marks
Q.2	Full Length Practical Question	15 Marks
Q.3	Full Length Practical Question OR	15 Marks
Q.3	Full Length Practical Question	15 Marks
Q.4	Full Length Practical Question OR	15 Marks
Q.4	Full Length Practical Question	15 Marks
Q.5	(A) Theory Questions (B) Theory Questions OR	08 Marks 07 Marks
Q.5	Short Notes To be asked (05) To be answered (03)	15 Marks

Note: Practical question of 15 marks may be divided into two sub-questions of 7/8 and 10/5 marks. If the topic demands, instead of practical questions, appropriate theory question may be asked.

CONTENTS

Chapter 1: Introduction to Statistics	1 – 12
1.1 Introduction	
1.2 Definitions of Statistics	
1.3 Applications of Statistics	
1.4 Functions of Statistics	
1.5 Limitations of Statistics	
1.6 Types of Data and Data Collection Methods	
1.6.1 Data	
1.6.2 Continuous and Discrete Data	
1.6.3 Collection of Data	
1.7 Census and Sample Survey	
1.7.1 Sample Survey	
1.7.2 Census Survey	
1.7.3 The Decision	
Chapter 2. Presentation of Data	13 – 27
2.1 Introduction	
2.2 Frequency Distribution	
2.3 Graphical Presentation of Data	
2.3.1 Histogram	
2.3.2 Frequency Polygon	
2.3.3 Frequency Curve	
2.3.4 Cumulative Frequency Curve (Ogive)	
Chapter 3. Measures of Central Tendency	28 – 102
3.1 Introduction	
3.2 Requisite of Good Measure of Central Tendency	
3.3 Types of Measures of Central Tendency	
3.4 Arithmetic Mean (A.M.)	
3.4.1 Simple Arithmetic Mean for Individual Series	
3.4.2 Arithmetic Mean for Frequency Distribution	
3.4.3 Weighted Arithmetic Mean	

- 3.5 Arithmetic Mean of Combined Groups
- 3.6 Correcting Incorrect Values
- 3.7 Calculation of Missing Frequencies when A.M is known
- 3.8 Merits and Demerits of Arithmetic Mean
- 3.9 Geometric Mean
 - 3.9.1 Simple Geometric Mean
 - 3.9.2 Compound Interest Formula
 - 3.9.3 Geometric Mean for Frequency Distribution
 - 3.9.4 Weighted Geometric Mean
 - 3.9.5 Combined Geometric Mean
- 3.10 Harmonic Mean
 - 3.10.1 Harmonic Mean for Ungrouped Data
 - 3.10.2 Harmonic Mean for Frequency Distribution
- 3.11 Median
 - 3.11.1 Median for Individual Series
 - 3.11.2 Median for Frequency Distribution
- 3.12 Calculation of Missing Frequencies when Median is Known
- 3.13 Merits and Demerits of Median
- 3.14 Quartiles, Deciles and Percentiles
- 3.15 Median, Quartiles, etc., from Ogives

Chapter 4. Measures of Dispersion

103 – 140

- 4.1 Introduction
- 4.2 Requisite of Good Measure of Dispersion
- 4.3 Types of Measures of Dispersion
- 4.4 Range
 - 4.4.1 Absolute Measure of Range and Relative Measure of Range
 - 4.4.2 Merits and Demerits of Range
- 4.5 Quartiles-based Measures of Dispersion
 - 4.5.1 For Individual Series
 - 4.5.2 For Frequency Distribution

- 4.6 Mean Deviation
 - 4.6.1 Mean Deviation for Individual Series
 - 4.6.2 Mean Deviation for Frequency Distribution
 - 4.6.3 Merits and Demerits of Mean Deviation
- 4.7 Standard Deviation and Variance
 - 4.7.1 Standard Deviation for Individual Series
 - 4.7.2 Standard Deviation for Frequency Distribution
- 4.8 Combined Standard Deviation
- 4.9 Corrected Standard Deviation
- 4.10 Relative Dispersion: The Coefficient of Variation
- 4.11 Merits and Demerits of Standard Deviation

Chapter 5. Correlation Analysis

141 – 168

- 5.1 Introduction
- 5.2 Types of Correlation
- 5.3 Properties of Correlation Coefficient
- 5.4 Methods of Studying Correlation
- 5.5 Scatter Diagram
- 5.6 Karl Pearson's Coefficient of Correlation
 - 5.6.1 Calculation of Karl Pearson's Coefficient of Correlation by Actual Mean Method
 - 5.6.2 Calculation of Karl Pearson's Coefficient of Correlation by Direct Method (When No Deviations are Taken)
 - 5.6.3 Calculation of Karl Pearson's Coefficient of Correlation by Assumed Mean Method
 - 5.6.4 Coefficient of Correlation for Bivariate Grouped Data
 - 5.6.5 Merits and Demerits of Karl Pearson's Coefficient of Correlation
- 5.7 Spearman's Rank Correlation Coefficient
 - 5.7.1 When Actual Ranks are given
 - 5.7.2 When Ranks are not given
 - 5.7.3 Merits and Demerits of Rank Correlation Coefficient
- 5.8 Correlation and Causation

Chapter 6. Regression Analysis**169 – 191**

- 6.1 Introduction
- 6.2 Uses of Regression Analysis
- 6.3 Methods of Studying Regression
- 6.4 Regression Lines
- 6.5 Regression Equations
- 6.6 Least Square Method
- 6.7 Regression Equations Using Regression Coefficients
 - 6.7.1 Direct Method
 - 6.7.2 Deviations taken from Arithmetic Mean
 - 6.7.3 Deviations taken from the Assumed Mean
- 6.8 Properties of the Regression Coefficient

Chapter 7. Time Series Analysis**192 – 211**

- 7.1 Introduction
- 7.2 Significance of Time Series Analysis
- 7.3 Components of Time Series
- 7.4 Time Series Models
- 7.5 Estimation of Trend
 - 7.5.1 Method of Moving Averages
 - 7.5.2 Curve Fitting and Method of Least Squares
- 7.6 Measurement of Seasonal Variations
(By the Method of Simple Average)

Chapter 8. Index Numbers**212 – 239**

- 8.1 Introduction
- 8.2 Meaning
- 8.3 Uses
- 8.4 Types of Index Numbers
- 8.5 Methods of Calculating Index Numbers
 - 8.5.1 Unweighted Index Numbers
 - 8.5.2 Weighted Index Numbers
- 8.6 Cost of Living Index Number

- 8.7 Chain Base Index Numbers
 - 8.7.1 Conversion of Chain Base to Fixed Base Index
 - 8.7.2 Conversion of Fix Base Index to Chain Base Index
- 8.8 Base Shifting, Splicing and Deflating the Index Numbers
 - 8.8.1 Base Shifting
 - 8.8.2 Splicing
 - 8.8.3 Deflating
- 8.9 Problems in the Construction of Index Numbers
- 8.10 Limitations of Index Numbers

Chapter 9. Probability

240 – 268

- 9.1 Introduction
- 9.2 Prerequisites
 - 9.2.1 Fundamental Principle of Counting
 - 9.2.2 Permutation
 - 9.2.3 Combination
- 9.3 Basic Concepts in Probability
- 9.4 Mathematical Definition of Probability
- 9.5 Addition Theorem of Probability
- 9.6 Conditional Probability
 - 9.6.1 Probability in Case of Independent Events
 - 9.6.2 Conditional Probability in Various Case
- 9.7 Bayes' Theorem on Inverse probability

Chapter 10. Probability Distribution

269 – 316

- 10.1 Decision making process involves the following step
 - 10.1.1 Identification of possible events which is also called state of nature E_i . Events are beyond the control of decision maker.
 - 10.1.2 Identify possible course of action A_i . These are controllable and decision maker has control over choice of this.
 - 10.1.3 Determine the pay-off function which describes the possible outcome for every action against the possible event.
 - 10.1.4 Choose the best among the various alternatives based on the criterion.
- 10.2 Features
- 10.3 Types of Decision Making Environment



1.1 Introduction

The word 'Statistics' have been derived from the Latin word 'Status' or the Italian word 'Statista' or the German word 'Statistik', each of which means a 'political state'. In ancient days government used to collect data regarding the population and the wealth – the former to have an idea of the manpower of the country and the latter for providing a basis for introducing new taxes. Indian history also reveals the use of statistics in our ancient kingdoms.

In the present age, statistics is regarded as one of the most important tools for taking decisions. Statistics is defined as the science of collecting, organizing, presenting, analyzing and interpreting numerical data for the purpose of assisting in making a more effective decision.

Kinds or Branches of Statistics

Statistics may be divided into two main branches:

(1) Descriptive Statistics (2) Inferential Statistics

- (1) **Descriptive Statistics:** In descriptive statistics, it deals with collection of data, its presentation in various forms, such as tables, graphs and diagrams and findings averages and other measures which would describe the data.

Example: Industrial statistics, population statistics, trade statistics etc. Businessman make use of descriptive statistics in presenting their annual reports, final accounts, bank statements.

- (2) **Inferential Statistics:** It is a method used to generalize from a sample to a population. Inferential statistics, deals with techniques used for analysis of data, making the estimates and drawing conclusions from limited information taken on sample basis and testing the reliability of the estimates.

Example:

- (i) Suppose we want to have an idea about the percentage of illiterates in our country. We take a sample from the population and find the proportion of illiterates in the sample. This sample proportion with the help of probability enables us to make some inferences about the population proportion. This study belongs to inferential statistics.
- (ii) The average income of all families (the population) in India can be estimated from figures obtained from a few hundred (the sample) families.

1.2 Definitions of Statistics

Statistics have been defined by different authors from time to time. We will describe here the definition of Croxton, Horace Secrist and Ya Lun Chou.

- (i) "Statistics is the science that deals with collection, analysis and interpretation of numerical data."

- Croxton

- (ii) "By Statistics, we mean aggregate of facts affected to a marked extent by multiplicity of causes numerically expressed, enumerated or estimated according to reasonable standards of accuracy, collected in a systematic manner for a predetermined purpose and placed in relation to each other."

- Prof Horace Secrist.

- (iii) "Statistics is a method of decision making in the face of uncertainty on the basis of numerical data and calculated risk."

- Ya – Lun – Chu

1.3 Applications of Statistics

Statistics plays a vital role in every fields of human activity. It is used in virtually every sector of business and government, including medicine, insurance, agriculture, computer science. Now statistics holds a central position in almost every field like Industry, Commerce, Trade, Physics, Chemistry, Economics, Mathematics, Biology, Botany, Psychology, Astronomy etc., so application of statistics is very wide. Now we discuss some important fields in which statistics is commonly applied.

(1) Business

The following are the main uses of statistics in various business activities.

- (i) By using the techniques of time series analysis which are based on statistical methods, the businessman can predict the effect of the large number of variables with a fair degree of accuracy.
- (ii) With the help of statistical methods, a businessman can make correct decision about the location of business, marketing of the products, financial resources etc.

- (iii) Statistics helps businessman to plan production according to the taste of the customers, the quality of the products can also be checked more efficiently by using statistical methods.

(2) In Economics

Statistics play an important role in economics. National income accounts are multipurpose indicators for the economists and administrators. Statistical methods are used for preparation of these accounts. In economics research, statistical methods are used for collecting and analysis the data and testing hypothesis. The relationship between supply and demands is studies by statistical methods, the imports and exports, the inflation rate, the per capita income are the problems which require good knowledge of statistics.

(3) In Natural Sciences

Statistics plays a vital role in almost all the natural sciences. Statistical methods are commonly used for analyzing the experiments results, testing their significance in Biology, medicine, zoology, Meteorology etc .

For example:

- (i) statistics relating to pulse rate, body temperature, patient's blood pressure, heart beats, weight etc. of the patient helps the doctor to diagnose the disease.
- (ii) An important application of statistics lies in using the test of significance for testing the efficacy of a particular drug or injection meant to cure a specific disease.
- (ii) In metrology, how well can we predict weather patterns? How fast is the ozone layer depleting?

(4) In Social Sciences

Statistical data provides information on mortality, marriage, population and growth. It can be used to find how have drinking patterns changed this century? Are marriages failing more often?

(5) In Astronomy

Astronomy is one of the oldest branch of statistical study, it deals with the measurement of distance, sizes, masses and densities of heavenly bodies by means of observations. During these measurements, errors are unavoidable so most probable measurements are founded by using statistical methods.

For example: Since old days the astronomers have been using statistical methods like method of least squares for finding the movements of stars.

(6) In State Management (Administration)

The government requires periodical data about population, agriculture, education, crime, health etc. so that necessary policies may be framed depending on such statistics. Suppose if the government wants to revise the pay scales of employees in view of an increase in the living cost, statistical methods will be used to determine the rise in the cost of living.

1.4 Functions of Statistics

The functions of statistics are as follows:

- (1) **Simplifies complex data:** Human mind cannot understand a large number of facts and figures at any one time. For example it is really difficult for one to develop a precise idea about the income of the people of India from a record of individual income of the entire population. However, one can easily remember the figures of per capita income.
- (2) **Helps to compare:** Comparison is one of the main function of Statistics. It helps in comparing the data with respect to time and location. Two or more characteristics can be compared using measures like averages, graphs or coefficient of variation.
- (3) **Helps in formulating policies:** Statistics also helps in the formulation of various economic, business and other policies at state, national or global level. For example framing of government policies in education, taxation, pollution, law and order, import and export, social welfare, wages etc. are formulated on the basis of statistical data and inferences drawn from their analysis.

Business organizations also make use of statistics to design their policies in areas of finance, marketing and personnel.

- (4) **Helps in business forecasting:** Statistic helps for projecting the future with the help of present and past data. Statistics facilitates business forecasting through various techniques as time series, extrapolation etc. Forecasting is used in various industries to predict the future sales, demand so that adjustments can be made accordingly.
- (5) **Enlarges Individual experience:** It enlarges human experience and knowledge by making it easier for man to understand, describe and measure the effect of his own actions.
- (6) **Tests Hypothesis:** Hypothesis is an important concept in research studies. Statistics provides various methods for testing the hypothesis. The important methods for hypothesis tests are chi square test, z- test, t-test.

1.5 Limitations of Statistics

Although Statistics has its application in almost all sciences — social, physical and natural – it has certain limitations as well.

- (1) **Does not study qualitative phenomena:** Statistical methods are best applicable on quantitative data. As a matter of fact, qualitative phenomenon like honesty, poverty, beauty, intelligence etc., cannot be expressed numerically and any statistical analysis cannot be directly applied on these qualitative characteristics. It is possible to do so only by assigning suitable quantitative measures to such a characteristics. For example, the intelligence of a group of students can be studied on the basis of their marks in a particular examination.
- (2) **Does not study individual items:** Statistics does not give any specific importance to the individual items, in fact it deals with an aggregate of objects. Individual items, when they are taken individually do not constitute any statistical data and do not serve any purpose for any statistical enquiry.
- (3) **It can be misused:** Statistical methods are most dangerous tools in the hands of inexperience. The use of statistical tools by the inexperienced persons might lead to wrong conclusions. Statistics can be easily misused by quoting wrong figures of data. Only a person who has an expert knowledge of statistics can handle statistical data efficiently.
- (4) **Statistical laws are not exact:** It is well known that mathematical and physical sciences are exact. But statistical laws are not exact and statistical laws are only approximations. Statistical conclusions are not universally true. They are true only on an average.

1.6 Types of Data and Data Collection Methods

1.6.1 Data

Data are collection of any number of related observations. We can collect the number of telephones that workers install on a given day or that one worker install per day over a period of several days, and we can call the results our data. Information can be collected in statistics using **qualitative and quantitative data**.

Qualitative data, such as eye color of a group of individuals, is not computable by arithmetic relations. They are labels that advise in which category or class an individual, object, or process fall. They are called categorical variables.

Quantitative data sets consist of measures that take numerical values like the ages of the students.

1.6.2 Continuous and Discrete Data

The Quantitative data can be classified as:

(a) Discrete data: It consists of values which are separated from each other.

Examples of discrete data are:

- (i) The marks obtained in a test by 30 students.
- (ii) Number of patients in a hospital.
- (iii) Number of members in a family.
- (iv) Number of times 100 motorists had to take their driving test before they passed.

(b) Continuous data: It consists of numbers which can take any value within a given range.

Examples of discrete data are:

- (i) The height of 500 females.
- (ii) The time taken by 1000 people to run a marathon race.
- (iii) The temperature at noon at 50 seaside resorts.

Now, for better understanding compare the following items:

- A box of pencils, which contains an average of 12 pencils.
- A bottle of soft drink, which contains an average of 300 ml of liquid.

There could be 11 or 13 pencils in the pencil box, but there could not be 12.23. However many boxes we open and count the result must be whole numbers. The number of pencils inside the box are discrete data.

In a bottle of soft drink, there could be 302.45 ml, or 297.72 ml. These does not have to be a whole number of millilitres. However many bottles we open and measure the content, the result will always come from a continuous range of numbers. Therefore, the volumes of the bottles of soft drink are continuous data.

1.6.3 Collection on data

This is the first step in any statistical enquiry. We require to collect the data for a given problem or research. Collection of data is very important as it is used for analysis. The types of sources through which data is collected are of two types, primary sources and secondary sources.

(1) Primary data: When data is collected from first hand sources, it is called primary data collection. For primary data collection, we adopt various methods:

(a) Observation method: In this method, the investigator records his observations from the experiment or in a specific situation. This method is applied in Science, Medical Sciences, Social Sciences.

- (b) **Questionnaire method:** Questionnaire is a series of questions asked for enquiry. Generally these questions are of objective nature. Many times the possible answers are also given. There are certain guidelines about length of questionnaire, which types of questions should be asked, which types of questions should not be asked, how we can check the information by asking cross questions etc. The information through questionnaire may be collected by sending an investigator or it may be collected by post which is a postal enquiry or by e-mail or fax.
- (c) **Interview method:** We may collect the information through interview also. Generally, if we require opinions of experts, or opinions on a specific topic in detail then this method is adopted. This is comparatively informal method. We require to prepare an interview schedule for interview. Most of the commissions and committees appointed by government to carry on investigations make use of this method.
- (2) **Secondary data:** Secondary data means data that are already available, i.e., they refer to the data which have already been collected and analysed by someone else. While collecting such type of information, we should observe authenticity of such sources. Some such sources are:
- (1) Census report, (2) Government report, (3) Well established magazines,
 - (3) Annual reports of companies, (4) Technical and trade journals.

1.7 Census and Sample Survey

Population means the entire universe. In statistics, all units under a given situation is called population, so population is a relative term in statistics. If we study colleges in Mumbai then each and every college is included in the population. If we consider Maharashtra state, then each and every college in Maharashtra is the population. When we study the entire population, it is complete enumeration. This is also called **census survey**. In case of census survey, the results are more accurate and reliable. It is an extensive study.

Sample is a part of population. It is a small lot, representative of the whole population. A sample is not studied for its own sake. The basic objective of its study is to draw inference about the population.

1.7.1 Sample Survey

The following are the merits and/or necessities for sampling in statistical decision making:

Merits of Sample Surveys

1. Cost Reduction

In most cases, conducting a sample survey costs less than a census survey. If fewer people are surveyed, fewer surveys need to be produced, printed, shipped, administered, and analyzed. Further, fewer data reports are often required, thus the amount of time and expense required to analyze and distribute the results is reduced.

2. Generalizability of Results

If conducted properly, the results of a sample survey can still be generalized to the entire population, meaning that the sample results can be considered representative of the views of the entire target population.

3. Timeliness

Sample surveys can typically be printed, distributed, administered, and analyzed more quickly than census surveys. As a result, a shorter turnaround time for results is often achieved.

4. Identification of Strengths and Opportunities

Results from a properly conducted sample survey can be used to identify strengths and opportunities and develop plans for meaningful change.

Demerits of Sample Surveys

The following are the demerits of sampling surveys:

1. Determining the Correct Sample Frame

Determining the correct sample frame (i.e., determining who should participate) can require extensive work. Laypersons may be left to engage in guesswork to determine the correct sample frame. This guesswork, if incorrect, will yield survey results that inaccurately reflect the sentiments of the population, and therefore compromise any conclusions drawn based upon the results.

Many factors should be considered when determining a suitable sample frame, such as the total population size, appropriate confidence level, anticipated participation level, anticipated standard deviation etc. Individuals who are not confident in making these considerations are well advised to consult an experienced survey researcher or statistician.

It should be noted, however, that even sampling experts may inadvertently overlook organizational or employee-relations issues that impact sampling effectiveness. For example, a well-calculated organizational sampling strategy may require 10 out of 50 employees from one plant or work group

to complete a survey. If, however, this group is experiencing severe absenteeism or morale problems that is not known to the sampling expert, actual participation rates may be lower than anticipated, which could prevent the desired confidence level and confidence interval from being attained. These types of issues are difficult to incorporate into a sampling strategy.

2. Possibility of Creating Feelings of Exclusion

While employees may intuitively understand the sampling process, some may feel left out or have ill feelings toward management due to being excluded from the sample frame. This may be particularly likely when unequal treatment or inclusion has been perceived as an issue in the past. In these cases, inviting all employees to participate is best from a humanistic and employee morale perspective.

3. More Inclination to Discount Results and Less Inclination to Take Action

When survey results are favorable there tends to be a general feeling that the sampling strategy was effective. However, when survey results are unfavorable there is a general tendency to “discount” the results and claim that the results were based on a poor sample. Leaders may argue that the results are less favorable because the sample only selected those persons with more negative attitudes. While great lengths may have been taken to ensure a representative sample, some leaders may still question the selection process and therefore be resistant to taking action.

1.7.2 Census Survey

The Census survey has following merits:

Merits of Census Surveys

1. Everyone Has an Opportunity to Participate

One of the greatest advantages of a census survey is that all employees have the same opportunity to participate. Some employees may still choose not to participate, but at least the opportunity to do so is presented. No one person or group can feel slighted or left out.

2. Accuracy Concerns are Reduced

Sample surveys, when conducted properly, are certainly capable of yielding representative results, however, census surveys tend to enhance feelings of security surrounding the accuracy of the results.

3. Easier to Administer

Administration of sample surveys is more complicated. A census survey is easier to administer, because it includes all persons. The volume of surveys that need to be distributed may increase with a census survey, but figuring out who receives a survey is clear — everyone.

4. Obtains Better Demographic Data

A great advantage to conducting census surveys is the ability to collect better demographic data across the population. Demographic items such as age, tenure, location, function, ethnicity, education and intent to leave can be used to create subsets of the population. Demographic data can be used to enhance the depth of clarity surrounding critical organizational issues. For example, demographic data can help determine if all employees are unsatisfied with a particular issue or just one demographic group (e.g., longer tenured employees). It may be important to know that employees who have been with the organization for five to 10 years are the ones most intent upon leaving. After all, employees comprising this demographic group may represent the most knowledgeable employees. If they intend to leave, then they will take their knowledge, skills and talent with them. Demographic data not only builds better insight, it also enables more effective action once the survey results are delivered.

5. All Work Groups Receive Results

Because all employees participate, there is a greater chance of obtaining responses that are representative of all sub-groups within the organizational structure. Managers at all levels of an organization can then review their results, compare these to the results of the overall company, and establish **action plans**. Armed with this knowledge, managers at all levels of the organization can then be held accountable for implementing actions that deliver effective change.

6. Less Inclination to Discount Results and More Inclination to Take Action

While it is a fact that a well-designed sample survey can provide accurate results, most managers tend to believe their results more readily and be more inclined to take action when all employees have an opportunity to respond. Belief in the results and inclination to take action go hand in hand.

Demerits of Census Surveys

This method suffers from certain limitations which are:

1. Higher Cost and More Time

Census surveys, generally, cost more to administer than sample surveys. Printing, distribution, and data collection costs all increase with a census survey due to the increased volume of materials and labor required.

Larger volumes of data require more time for data entry, data cleaning/verification, data analysis and report production. Data coding and report dissemination may also become more complicated. The bottom line is that the more data there is, the more complicated a survey process becomes.

2. Longer Administration Time

Including the entire population of interest in a survey process tends to require additional administration time. After all, coordinating schedules with a few persons or groups is much easier than coordinating schedules across an entire organization. Production schedules, delivery deadlines, vacation and holiday schedules, and travel schedules make it more difficult to gather feedback.

3. Increases Expectations

Expectations for **action planning** and improvements increase when all employees participate. In return for their input, employees now want you to take action. Employees expect to see the link between what they said in the survey and the changes implemented by the organization. It's not good enough to say something will improve; they want to see proof.

4. Requires More Training and Support

Not all managers are survey experts. Many managers are used to looking at numbers, spreadsheets, and facts and figures. But reviewing the statistical results from a survey and making sense of the percentages, mean scores, rankings, or statistical significance values requires a different skill set. When survey results are dispersed to all managers across an organization, more assistance, training and guidance are needed to improve understanding and interpretation of the results. When managers are held accountable for creating and implementing **action plans**, more assistance is also needed to sustain managerial involvement and motivation. Increased training and assistance may increase overall survey costs.

1.7.3 The Decision

Sample and census surveys both provide value, and when implemented properly, produce valid results. Four primary factors are important to keep in mind when determining which strategy is most suitable:

1. Desired Outcomes

What is the survey vision and goal? What do you intend to do with the results? How are **action plans** going to be implemented? Who will be held accountable for results? Knowing the end result will help you determine the appropriate path to take. If the outcome is organization-wide **action planning** and accountability at the work group level, then a census survey is required. If the outcome is the collection of high-level results with **action planning** at the highest levels of the organization, then a sample survey will provide adequate data.

2. Purpose

What are the issues you want to address? Are you losing key talent and in need of understanding why people are leaving? Are you losing your customers to competitors? Do you just feel like it is a good time to solicit feedback from your employees? While this goes hand in hand with the survey vision and goal, it is important to understand the driving reason or reasons for collecting the data in the first place. Understanding these key drivers will help you select the appropriate strategy for gathering feedback.

3. Budget

What is your budget? Budgetary constraints often dictate which survey process to implement. You may feel the value of conducting a census survey is greater, but not have sufficient funds to devote to a large-scale survey. Instead of eliminating the survey altogether, it is more advantageous to conduct a sample survey instead.

4. Perceptions

How will your employees perceive the survey process? Will employees feel left out if not included in the sample? Do employees value surveys and look forward to providing input? Have your employees been surveyed to death without seeing beneficial improvements? Perception is reality so no matter how you rationalize the survey process, it's what matters most to your employees that counts.

Exercise

- A**
1. Define business statistics.
 2. Explain some important functions of statistics.
 3. Discuss briefly the possible applications of statistical methods in business, pointing out the limitations, if any.
 4. What are the limitations of statistics?
- B**
1. Define primary data. What are the methods of collecting primary data?
 2. What is secondary data? State its chief sources.
 3. Distinguish between primary and secondary data.
- C**
1. Explain the terms 'population' and 'sample'. Explain why it is sometimes necessary and often desirable to collect information about the population by conducting a sample survey instead of complete enumeration.
 2. Distinguish between census and sample methods of data collection. Point out the special advantages of sampling techniques and also its limitations if any.
 3. Explain briefly the limitations and advantages of census survey.

