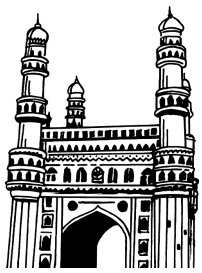
Rahul's ✔
Topper's Voice

AS PER
CBCS SYLLABUS



Latest 2022 Edition

B.Sc.

III Year V Sem

(All Universities in Telangana)

STATISTICS (PAPER-V)

SAMPLING THEORY, TIME SERIES INDEX NUMBERS AND DEMAND ANALYSIS

- ☞ Study Manual
- ☞ FAQ's & Important Questions
- ☞ Choose the Correct Answer
- ☞ Fill in the blanks
- ☞ Exercise Problems
- ☞ One Mark Answers
- ☞ Solved Previous Question Papers
- ☞ Solved Model Papers

Price
- 169/-

- by -

WELL EXPERIENCED LECTURER

Rahul Publications™
Hyderabad. Ph : 66550071, 9391018098

B.Sc.

III Year V Sem

(All Universities in Telangana)

STATISTICS

(PAPER -V)

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

Price ` . 169-00

Sole Distributors : ☎ : **665500071, Cell : 9391018098**

VASU BOOK CENTRE

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

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

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

STATISTICS
(PAPER -V)

CONTENTS

STUDY MANUAL

FAQ's & Important Questions	**V - VIII**
Unit - I	**1 - 44**
Unit - II	**45 - 116**
Unit - III	**117 - 204**
Practicals	**205 - 205**

SOLVED PREVIOUS QUESTION PAPERS

July - 2021	**206 - 207**
October / November - 2020	**208 - 208**
June / July - 2019	**209 - 210**
November / December - 2018	**211 - 212**
November / December - 2018(MGU)	**213 - 213**

SOLVED MODEL PAPERS

MODEL PAPER - I	**214 - 215**
MODEL PAPER - II	**216 - 217**
MODEL PAPER - III	**218 - 219**

SYLLABUS

UNIT - I

Sample Surveys: Concepts of Population, Sample, Sampling unit, Parameter, Statistic, Sample frame and Standard error. Principal steps in sample surveys - need for sampling, census versus sample surveys, sampling and non- sampling errors, sources and treatment of non-sampling errors, advantages and limitations of sampling.

Sampling Methods: Types of sampling: Subjective, probability and mixed sampling methods. Methods of drawing random samples using with and without replacement. Estimates of population mean, total, and proportion, their variances and the estimates of variances in Simple Random Sampling With and Without Replacement.

UNIT - II

Estimates of population mean, total, and proportion, their variances and the estimates of variances in the following methods.

(i) Stratified Random Sampling with Proportional and Neyman allocation, and

(ii) Systematic Sampling when N = nk.

Comparison of their relative efficiencies. Advantages and disadvantages of SRS, Stratified and Systematic sampling methods.

Time series: Time series and its components with illustrations, additive, multiplicative and mixed models.

Determination of trend by least squares and moving average methods. Growth curves and their fitting with reference to Modified exponential, Gompertz and Logistic curves. Determination of seasonal indices by Ratio to moving average, ratio to trend and link relative methods.

UNIT - III

Demand Analysis: Introduction. Demand and supply, price elasticity of supply and demand. Methods of determining demand and supply curves, Leontief's ,Pigous's methods of determining demand curve from time series data, limitations of these methods Pigou's method from time series data. Pareto law of income distribution curves of concentration.

Index Numbers: Concept, construction, uses and limitations of simple and weighted index numbers.

Laspeyer's, Paasche's and Fisher's index numbers, criterion of a good index numbers, problems involved in the construction of index numbers. Fisher's index as an ideal index number. Fixed and chain base index numbers. Cost of living index numbers and wholesale price index numbers. Base shifting, splicing and deflation of index numbers

Contents

UNIT - I

Topic **Page No.**

1.1 Sample Surveys 1

1.1.1 Concepts of Population, Sample, Sampling unit, Parameter 2

1.1.2 Statistic, Sample frame and Standard error 3

1.2 Principal Steps in Sample Surveys 4

1.2.1 Need for Sampling 6

1.2.2 Census Versus Sample Surveys 7

1.2.3 Sampling and Non-Sampling Errors, Sources and treatment of non-sampling errors 9

1.2.4 Advantages and Limitations of Sampling 12

1.3 Sampling Methods 13

1.3.1 Subjective Method 13

1.3.2 Probability Method 14

1.3.3 Mixed sampling Method 14

1.3.4 Methods of drawing random samples using with and without replacement 15

1.3.5 Estimates of population mean, total, and proportion, their variances 19

1.3.6 Estimates of variances in Simple Random Sampling With and Without Replacement 22

- **Short Question & Answers 37**
- **Choose the Correct Answers 42**
- **Fill in the Blanks 43**
- **One Mark Answers 44**

UNIT - II

2.1 Estimates of Population Mean, Total, and Proportion, their Variances and the Estimates of Variances in the following Methods 45

Topic **Page No.**

2.1.1 Stratified Random Sampling 45

2.1.2 Proportional and Neyman Allocation 47

2.2 Systematic Sampling when N = nk. 56

2.3 Comparison of their Relative Efficiencies 59

2.4 Advantages and Disadvantages of SRS, Stratified and Systematic Sampling Methods 67

2.5 Time Series 69

2.5.1 Components of Time Series with illustrations 71

2.6 Models of Time Series Analysis 73

2.6.1 Additive, Multiplicative and Mixed Models 73

2.7 Determination of Trend by Least Square Method 74

2.7.1 Determination of Trend by Moving Average Method 83

2.8 Growth Curves and their Fitting to Modified Exponential 86

2.9 Growth curves and their Fitting to Gompertz and Logistic curves 90

2.10 Determination of seasonal indices Ratio to moving average 93

2.10.1 Ratio to Moving Average 93

2.10.2 Ratio to Trend Method 96

2.10.3 Link Relative Method 99

- **Exercise Problems 102**
- **Short Question & Answers 104**
- **Choose the Correct Answers 112**
- **Fill in the Blanks 114**
- **One Mark Answers 116**

UNIT - III

3.1 Demand Analysis - Introduction 117

3.1.1 Law of Demand 119

3.2 Methods of Determining Demand 121

Topic		Page No.
3.3	Elasticity of Demand	126
	3.3.1 Measurement of Elasticity of Demand	131
3.4	Introduction to Supply - Methods of Determining Supply	136
	3.4.1 Price Elasticity of Supply	139
3.5	Leontief's Methods of Determining Demand Curve from Time Series Data and their Limitations	141
3.6	Pigou's Methods of Determining Demand Curve from Time Series Data and their Limitations	143
3.7	Pareto Law of Income	144
	4.7.1 Distribution Curves of Concentration	146
3.8	Index Numbers - Concept	149
	3.8.1 Construction of Index Number	151
	3.8.2 Uses and Limitations of Index Number	152
3.9	Methods of Constructing Index Numbers	154
	3.9.1 Simple and Weighted Index Numbers.	154
	3.9.2 Laspeyre's Method	156
	3.9.3 Paasche's Method	157
	3.9.4 Fisher's Ideal Index	157
	3.9.5 Criterion of Good Index Numbers	163
	3.9.6 Problems Involved in Construction of Index Numbers	169
3.10	Fixed and chain based index numbers	171
3.11	Cost of Living Index Numbers	176
3.12	Wholesale Price Index Numbers	179
3.13	Base Shifting	182

Topic **Page No.**

3.14 Spling of Index Number 184

3.15 Deflation of Index Number 186

- **Exercises Problems 189**
- **Short Question & Answers 191**
- **Choose the Correct Answers 201**
- **Fill in the Blanks 203**
- **One Mark Answers 204**

Frequently Asked & Important Questions

UNIT - I

1. What are the steps involved in sample survey ?

Ans : **(July-19)**

Refer Unit-I, Q.No. 7

2. Explain the various sources and treatment of non-sampling errors.

Ans : **(Dec.-18, July-21)**

Refer Unit-I, Q.No. 12

3. What are the benefits or advantages of sampling ?

Ans : **(Nov.-20)**

Refer Unit-I, Q.No. 13

4. What are the limitations of sampling ?

Ans : **(Nov.-20)**

Refer Unit-I, Q.No. 14

5. State and Explain subjective sampling (Or) Purposive Sampling (Or) Judge- ment Sampling.

Ans : **(July-19)**

Refer Unit-I, Q.No. 16

6. Explain the notation and terminology used in simple random sampling.

Ans : **(Dec.-18)**

Refer Unit-I, Q.No. 19

7. In a simple random sampling without replacement the sample mean is an unbiased estimate of the population mean.

$E(\bar{y}) = \bar{Y}$

Ans : **(Nov.-20, July-19, Dec.-18, Dec.-18(MGU)**

Refer Unit-I, Theorem. 1

8. In SRSWOR the variance of the sample mean is given by

$$\text{Var}(\bar{y}) = \frac{S^2}{n} \cdot \frac{N-n}{N}$$

Ans : **(July-21)**

Refer Unit-I, Theorem. 3

9. Explain Simple Random Sampling with Replacement (SRSWR).

Ans : **(July-19)**

Refer Unit-I, Q.No. 23

UNIT - II

1. What do you understand by stratified random sampling? Discuss its propor-tion means and its variance of Stratified Random Sampling.

Ans : **(July-21, July-19)**

Refer Unit-II, Q.No. 1

2. Explain briefly about proportional allocation.

Ans : **(July-19)**

Refer Unit-II, Q.No. 3

3. Explain briefly about Neyman allocation (or) optimum allocation.

Ans : **(July-19)**

Refer Unit-II, Q.No. 5

4. What do you understand by systematic sampling when n = nk.

Ans : **(Nov.-20)**

Refer Unit-II, Q.No. 8

5. Distinguish between relative efficiency of systematic sampling over stratified random sampling.

Ans : **(July-21)**

Refer Unit-II, Q.No. 10

6. What are the Components of Time Series ?

Ans : **(Nov.-20, Dec.-18(MGU)**

Refer Unit-II, Q.No. 17

7. What is a growth curve? Write in detail about modified exponential curve with example.

Ans : **(Nov.-20, July-19)**

Refer Unit-II, Q.No. 24

8. What do you understand by seasonal variation ?

Ans : **(Dec.-18)**

Refer Unit-II, Q.No. 27

9. What do you understand by link relative method (Or) Pearson's Method ?

Ans : **(July-21, Dec.-18(MGU)**

Refer Unit-II, Q.No. 30

UNIT - III

1. Explain the various methods of deter- ming demand.

Ans: **(July-21)**

Refer Unit-III, Q.No. 7

2. What do you understand by price elasticity of supply ? Explain different types of price elasticity of supply.

Ans : **(July-21)**

Refer Unit-III, Q.No. 21

3. Explain Leontief's Methods of Deter- mining Demand Curve from Time Series Data and their Limitations ?

Ans : **(Nov.-20, Dec.-18)**

Refer Unit-III, Q.No. 22

4. Explain Pigou's Methods of Determining Demand Curve from Time Series Data and their Limitations.

Ans : **(July-19)**

Refer Unit-III, Q.No. 23

5. Explain the criterion for testing the consistency of good index numbers.

Ans : **(July-19)**

Refer Unit-III, Q.No. 34

6. What are the problems Involved in Construction of Index Numbers

Ans : **(Dec.-18(MGU)**

Refer Unit-III, Q.No. 35

7. What do you understand by consumer price index ?

Ans : **(July-21, Dec.-18(MGU)**

Refer Unit-III, Q.No. 39

8. Define Wholesale Price Index Number. State various series of Wholesale Price Index Number.

Ans : **(Nov.-20)**

Refer Unit-III, Q.No. 41

9. What do you understand by base shifting of index numbers ?

Ans : **(Dec.-18)**

Refer Unit-III, Q.No. 42

10. What do you mean by Deflating with an example ?

Ans : **(Dec.-18)**

Refer Unit-III, Q.No. 45

UNIT I

Sample Surveys: Concepts of Population, Sample, Sampling unit, Parameter, Statistic, Sample frame and Standard error. Principal steps in sample surveys - need for sampling, census versus sample surveys, sampling and non- sampling errors, sources and treatment of non-sampling errors, advantages and limitations of sampling.

Sampling Methods: Types of sampling: Subjective, probability and mixed sampling methods. Methods of drawing random samples using with and without replacement. Estimates of population mean, total, and proportion, their variances and the estimates of variances in Simple Random Sampling With and Without Replacement.

1.1 SAMPLE SURVEYS

Q1. What do you understand by sampling ?

Ans :

Sampling may be defined as the selection of some part of an aggregate or totality on the basis of which a judgement or inference about the aggregate or totality is made. In other words, it is the process of obtaining information about an entire population by examining only a part of it.

In most of the research work and surveys, the usual approach happens to be to 'make generalizations' or to 'draw inferences' based on samples about the parameters of population from which the samples are taken. The researcher quite often selects only a few items from the universe for his study purpose. All this is done on the assumption that the sample data will enable him to estimate the population parameters.

The items so selected constitute what is technically called a sample, their selection process is called sample design and the survey conducted on the basis of sample is described a sample survey. Sample should be truly representative of population characteristics without any bias so that it may result in valid and reliable conclusions.

Q2. What do you understand by sample survey?

Ans : **(July-19)**

It is a term used in statistics and market research, indicating a survey conducted on a set of sample elements taken from a target population. A survey is a technique of observation to gather, measure and evaluate responses from people. Marketers often use sample surveys to conduct market research amongst target consumers for their brands, since it is not possible to go about asking the entire target population which may be huge in numbers. Hence, using sampling techniques, a representative set of the population viz. a sample is selected and used for conducting the required research, thus saving the cost of surveying the entire population. A survey conducted on the entire population is called a census.

Following points need to address for a sample survey:

(a) Sample unit: Whom to survey

(b) Sample size: How many people should be surveyed

(c) Sampling procedure: How should respondents for the survey be chosen

Example For understanding the consumer behavior in buying of luxury cars like Mercedes, BMW, Audi, the sample unit should consist of SEC A population of urban regions. Samples with large size give good results, but even 1% of the target population may give reliable results if proper sampling procedure is used. Probability sampling helps marketers determine confidence levels indicating the likelihood of a certain type of consumer being representative of the sample. In this case, the marketer may conclude that SEC A consumers having owned 1 luxury car in the past have a 95 out of 100 chance of being representative of the entire luxury car buyer population.

Q3. What is the purpose or importance of sampling ?

Ans :

The importance of sampling is that you can determine the adequate respondents from the total number of target population. Thus, it will be used in the research study which should be adequate to warrant generalization of the findings to the target population. And the sample size represents the characteristics of the whole population (representativeness of the sample).

The advantages of sampling are: it is economical and practical; faster and cheaper; it can yield more comprehensive information; it is more accurate; and because of savings it permits in time and money, the sample survey makes possible the use of much larger and much more varied populations than would be possible for the same expenditure if one were making a complete enumeration.

1.1.1 Concepts of Population, Sample, Sampling unit, Parameter

Q4. Define the terms

a) Population

b) Sample

c) Sampling Unit

d) Parameter

Ans : **(July-21)**

(a) Population

The collection of all units of a specified type in a given region at a particular point or period of time is termed as a population or universe. Thus, we may consider a population of persons, families, farms, cattle in a region or a population of trees or birds in a forest or a population of fish in a tank etc. depending on the nature of data required.

There are two different types of populations. They are,

(i) Finite population

(ii) Infinite population.

(i) Finite Population : A population whose values can be theoretically observed as it has fixed number is called a finite population. The size of the finite population is limited.

Example: Number of students in a school, number of rice plants in a field etc.

(ii) Infinite Population : A population whose values cannot be theoretically observed as it does not have a fixed number is called an infinite population. The size of the infinite population is unlimited.

Example: Number of White Blood Cells (WBC) in the human body.

The size of the population is always very large, even if it is finite or infinite. So, the collection of data from each of the individuals is very difficult. Because of this, samples are selected from the population and based on their results conclusions are made regarding that population.

(b) Sample

In biostatistical studies, the data collected depends on the observation of several individuals of a population. Here, sampling can also be performed i.e., few individuals can be selected from a population to represent the complete population. These selected individuals are together called as 'sample'.

Random samples is a sample selected from a population in such a way that each individual in the population has a fair chance of getting selected.

(c) Sampling Unit

Elementary units or group of such units which besides being clearly defined, identifiable and observable, are convenient for purpose of sampling are called sampling units. For instance, in a family budget enquiry, usually a family is considered as the sampling unit since it is found to be convenient for sampling and for ascertaining the required information. In a crop survey, a farm or a group of farms owned or operated by a household may be considered as the sampling unit.

(d) Parameter

The statistical constant of the population like mean (μ), variance (σ^2), skewness (β_1), kurtosis (β_2) moments (μ_r), correlation coefficient (ρ), etc., are known as parameters. We can compute similar statistical constants for the sample drawn from the given population. Prof. R.A. Fisher termed the statistical constants of the sample like mean $(\bar{x})$, variance (s^2), skewness (b_1), kurtosis (b_2), moments (m_r), correlation coefficient (r), etc., as statistics. Obviously, parameters are function of the population values while statistics are functions of the sample observations.

1.1.2 Statistic, Sample frame and Standard error

Q5. Define the terms

a) Statistics

b) Sample Frame

c) Standard Error

Ans : **(July-21)**

(a) Statistics

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

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

The term 'Statistics' has been defined differently by different authors. Some have defined the word in the sense of numerical data, whereas others have defined in the sense of statistical methods. A few definitions have been discussed here.

Statistics as numerical data (Plural Sense)

Bowel defined Statistics as 'Numerical statement of facts in any department of enquiry placed in relation to each other" According to Yule and Kendall "By statistics we mean quantitative data affected to a marked extent by multiplicity of causes'.

A comprehensive definition is given by Prof. Horace Secrets that points out all characteristics that numerical data must possess to be called statistics.

According to Horace Secrist "Statistics may be defined as the aggregates of facts affected to marked extent by multiplicity of causes, numerically expressed, enumerated or estimated according to reasonable standard of accuracy, collected in a systematic manner for a predetermined purpose and placed in relation to each other.''

(b) Sample Frame

Sampling frame is a collection of all the members/units of population. It should not contain any error and duplication of units. Each and every individual sampling unit is selected from a sampling frame.

It possess the following characteristics,

(i) Complete : A frame should consist of all legal units of the population.

(ii) Accuracy : A frame should not consist of any non-existing units of population.

(iii) Adequate : The structure of a frame must be sufficient enough to cover the total population.

(iv) Updated : The unit and its content present in a frame must be up to date.

(c) Standard Error

The standard deviation of the sampling distribution of a statistic is known as its Standard Error (S.E.). Thus, the Standard Error of the statistics is given by :

$$S.E.(t) = \sqrt{Var(t)} = \sqrt{\left[\frac{1}{k}\sum_{i=1}^{k}(t_i - t)^2\right]}$$

In particular the S.E. of the sampling distribution of the mean* is given by the standard deviation of the values $\bar{x}_1, \bar{x}_2, ..., \bar{x}_k$.

The derivation of the standard errors of the sampling distributions of various statistics is quite difficult and beyond the scope of the book. The standard errors of the sampling distributions of some of the well- known statistics, where n is the sample size, 0 is the population standard deviation, P is the population proportion and Q = 1 – P, and n_2 represent the sizes of two samples respectively, is given in the Table.

Standard Error of Important Statistics

S. No.	Statistic	Standard Error
1.	Sample mean $(\bar{x})$	$\sigma / \sqrt{n}$
2.	Sample proportion 'p'	$\sqrt{PQ / n}$
3.	Sample standard deviation (s)	$\sqrt{\sigma^2 / 2n}$
4.	Sample variance (s^2)	$\sigma^2\sqrt{2 / n}$
5.	Sample median	$1.25331\sigma / \sqrt{n}$
6.	'r' = sample correlation coefficient	$\left(1-\rho^2\right) / \sqrt{n}, \rho$ being the population correlation coefficient
7.	Difference of two independent sample means: $(\bar{x}_1 - \bar{x}_2)$	$\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}$
8.	Difference of two independent sample s.d.'s : $(s_1, - s_2)$	$\sqrt{\frac{\sigma_1^2}{2n_1} + \frac{\sigma_2^2}{2n_2}}$
9.	Difference of two independent sample proportions :	$\sqrt{\frac{P_1Q_1}{n_1} + \frac{P_2Q_2}{n_2}}$

1.2 PRINCIPAL STEPS IN SAMPLE SURVEYS

Q6. What are the principals in sample surveys ?

Ans : **(July-21)**

Three basic principles for the design of a sample survey are:

1. Principle of Optimization

The principle of optimization takes into account the factors of

(a) Efficiency and (b) cost.

(a) Efficiency : Efficiency is measured by the inverse of sampling variance of the estimator. The principle of optimization ensures that a given level of efficiency will be reached with the minimum possible resources and minimum cost.

(b) Cost : Cost is measured by expenditure incurred in terms of money or man powers. So, the term optimization means that, it is based on developing methods of sample selection and of estimation; these provide a given value of cost with the maximum possible efficiency.

2. Principle of Validity

By validity of a sample design, we mean that the sample should be so selected that the results could be interpreted objectively in terms of probability. According to this, sampling provides valid estimates about population parameters. This principle ensures that there is some definite and preassigned probability for each individual of the aggregate (population) to be included in the sample.

3. Principle of Statistical Regularity

According to the principle of statistical regularity we mean that a moderately large number of items chosen at random from a large group are almost sure on the average to possess the characteristics of the large group. This principle has also its origin in the law of large numbers of the theory of probability.

Q7. What are the steps involved in sample survey ?

Ans : **(July-19)**

The main steps involved in the planning and execution of a sample survey are under the following heads:

1. Objectives

The objective of the survey must be defined in clear and concrete terms. Generally, in survey a investigation team is not quite clear in mind as to what they want and how they are going to use the results. Some of the objectives may be immediate and some far-reaching. The investigator should take care of these objectives with the available resources in terms of money, manpower and the time limit required for the availability of the survey.

2. Defining the Population

The population from which sample is chosen should be defined in clear and unambiguous terms. The geographical, demographic and other boundaries of the population must be specified so that no ambiguity arises regarding the coverage of the survey.

3. Sampling Frame and Sampling Units

The sampling unit is the ultimate unit to be sampled for the purpose of the survey. The sampling units must cover the entire population and they must be distinct, unambiguous and non-overlapping in the sense that every element of the population belongs to one and only one sampling unit. In a Socio economic survey, whether a family or a member of a family is to be the ultimate sampling unit.

Once the sampling units are defined, one must see whether a sampling frame which is a list of all the units in the population, is available. The construction of the frame is often one of the major practical problem since it is the frame which determines the structure of the sample survey. The list of units have to be carefully scrutinized and examined to ensures that it is free from duplicity or incompleteness and are up-to-date. A good frame is hard to come by and only good experience helps to construct a good frame.

4. Selection of Proper Sampling Design

This is the most important step in planning a sample survey. There is a group of sampling designs (to be discussed later) and selection of the proper one is an important task. The design should take into account the available resources and the time-limit, if any, besides the degree of accuracy desired. The cost and precision should also be considered before the final selection of sampling design.

5. Method of Collection of Data

For collection of data, either the interview method or the mail questionnaire method is to be adopted. Although the later method is less costly but there is a large scope of non-response in it. In the cases, where the information is to be collected by observation they must decide upon the method of measurement.

6. Data to be Collected

Collection of data must be done in conformity with the objectives of the survey and the nature of the data. After it is decided upon, one must prepare a questionnaire or a schedule of enquiry. A schedule or a questionnaire contains a list of items of which information is sought, but the exact form of the questions to be asked is not standardized but left to the judgment of the investigators. A questionnaire should be in a specified order. The questions should be clear, brief, collaborative, non offending and unambiguous and to the point so that not much scope of speculation is left on the part of the respondent or interviewer.

7. Field Work Organization

Field work, itself has several stages and so it is to be well organized. The different stages include training the field workers, supervising the field workers, etc. It is absolutely essential that the personnel should be thoroughly trained in locating the sample units, the methods of collection of required data before starting the field word. The success of a survey to a great extent depends upon the reliable field work. Inspection after field work by the adequate supervisors should also be performed.

8. Summary and Analysis of Data

This is the last step wherein inference is to be made on the basis of collected data. This step again consists of the following steps:

(a) The filled in questionnaires should be carefully scrutinized to find out whether the data furnished are plausible and consistent;

(b) Depending upon the quantity of data, a hand-tabulation or machine tabulation is to be drawn;

(c) After the data has been properly scrutinized, edited and tabulated, a very careful statistical analysis is to be made; and

(d) Finally a report incorporating detailed statement of the different stages of the survey should be prepared. In the presentation of the result, it is advisable to report technical aspects of the design.

1.2.1 Need for Sampling

Q8. Explain the need for sampling.

Ans :

Sampling is used in practice for a variety of reasons such as:

1. Sampling can save time and money. A sample study is usually less expensive than a census study and produces results at a relatively faster speed.

2. Sampling may enable more accurate measurements for a sample study is generally conducted by trained and experienced investigators.

3. Sampling remains the only way when population contains infinitely many members.

4. Sampling remains the only choice when a test involves the destruction of the item under study.

5. Sampling usually enables to estimate the sampling errors and, thus, assists in obtaining information concerning some characteristic of the population.

1.2.2 Census Versus Sample Surveys

Q9. Discuss in detail about Census versus Sample Surveys.

Ans :

For any statistical enquiry in any field of humans activity, whether it is in business, economics or social sciences, the basic problem is to obtain adequate and reliable data relating to the particular phenomenon under study. There are two methods of collecting the data :

(i) The Census Method or Complete Enumeration.

(ii) The Sample Method or Partial Enumeration.

(i) Census Method. In the census method we resort to 100% inspection of the population and enumerate each and every unit of the population. In the sample method we inspect only a selected representative and adequate fraction (finite subset) of the population and after analyzing the results of the sample data we draw coclusions about the characteristics of the population.

The census method seems to provide more accurate and exact information as compared to sample enumeration as the information is collected from each and every unit of the population. Moreover, it affords more extensive and detailed study. For instance, the population census conducted by the Government of India every 10 years collects information not only about the population but also obtains data relating to age, marital status, occupation, religion, education, employment, income, property, etc. The census method has its obvious limitations and drawbacks given below :

(i) The complete enumeration of the population requires lot of time, money, manpower and administrative personnel. As such this method can be adopted only by the government and big organizations who have vast resources at their disposal.

(ii) Since the entire population is to be enumerated, the census method is usually very time consuming. If the population is sufficiently large, then it is possible that the processing and the analysis of the data might take so much time that when the results are available they are not of much use because of changed conditions.

(ii) Sample Method. The sample method has a number of distinct advantages over the complete enumeration method. Prof. R.A. Fisher sums up the advantages of sampling techniques over complete census in just four words : Speed, Economy, Adaptability and Scientific Approach. A properly designed and carefully executed sampling plan yields fairly good results, often better than those obtained by the census method. We now summarize the merits of the sample method over the census method.

1. Speed, i.e., less time

Since only a part of the population is to be inspected and examined, the sample method results in considerable amount of saving in time and labour. There is saving in time not only in conducting the sampling enquiry but also in the processing, editing and analyzing the data. This is a very sensitive and important point for the statistical investigations where the results are urgently and quickly needed.

2. Economy, i.e., Reduced Cost of the Enquiry

The sample method is much more economical than a complete census. In a sample enquiry, there is reduction in the cost of collection of the information, administration, transport, training and man hours. Although, the labour and the expenses of obtaining information per unit are generally large in a sample enquiry than in the census method, the overall expenses of a sample survey are relatively much less, since only a fraction of the population is to be enumerated. This is particularly significant in conducting socio-economic surveys in developing countries with budding economics who cannot afford a complete census because of lack of finances.

3. Administrative Convenience

A complete census requires a very huge administrative set up involving lot of personnel, trained investigators and above all

the co-ordination between the various operating agencies. On the other hand, the organization and administration of a sample survey is relatively much convenient as it requires less personnel staff and the field of enquiry is also limited.

4. Reliability.

In the census, the sampling errors are completely absent. If the non-sampling errors are also absent the results would be 100% accurate. On the other hand, a sample enquiry contains both sampling and non-sampling errors. In spite of this weakness, a carefully designed and scientifically executed sample survey gives results which are more reliable than those obtained from a complete census. This is because of the following reasons :

(a) It is always possible to ascertain the extent of sampling error and degree of reliability of the results. Even the desired degree of accuracy can be achieved through sampling using different devices.

(b) The non-sampling errors such as due to measuring and recording observations, inaccuracy or incompleteness of informations, location of units, non-response or incomplete response, training of investigators, interpretation of questions, bias of the investigators, etc., are of a more serious nature in a complete census. In a sample survey these errors can be effectively controlled and minimized by

(i) employing highly qualified, skilled and trained personnel,

(ii) imparting adequate training to the investigators for conducting the enquiry,

(iii) better supervision,

(iv) using more sophisticated equipment and statistical techniques for the processing and analysis of the relatively limited data.

5. Greater Scope

It appears that there is possibility of obtaining detailed information only in a complete census where each and every unit in the population is enumerated. But in practice because of our limitations in any statistical enquiry in terms of time, money and man hours and because of the fact that sampling procedure results in considerable savings in time, money and labour, it is possible to obtain more detailed and exhaustive information from the limited few units selected in the sample. Obviously, it is relatively easy to collect information on, say, 25 questions from each of 100 units selected in the sample than to obtain the information on, say, 10 questions from each of 1,000 units in the population. Moreover, complete enumeration is impracticable, rather inconceivable if the enquiry requires highly trained personnel and more sophisticated equipment for the collection, processing and analysis of the data. The sampling procedure is more readily adaptable than census for statistical investigations.

6. Infinite or / and Hypothetical Population

If the population is infinite or too large, then sampling procedure is the only way of estimating the parameters of a population. For instance, the number of fish in the sea or the number of wild elephants in a dense forest can be estimated only by sampling method.

Similarly, in case of hypothetical population, as for example in the problem of throwing a die or tossing a coin, where the process may continue large number of times or indefinitely, the sampling procedure is the only scientific technique of estimating the parameters of the population.

7. Destructive Testing

If the testing of units is destructive i.e.. if in the course of inspection the units are destroyed or affected adversely, then we are left with no other way but to resort to sampling. For example :

(i) to estimate the average life of the bulbs or tubes in a given consignment,

(ii) to determine the composition of a chemical salt,

(iii) to test the breaking strength of chalks manufactured in a factory or to estimate the tensile strength of the steel rods.

(iv) to test the quality of explosives, crackers, shells, etc., we have to inspect a representative sample, since complete census will destroy all the items.

1.2.3 Sampling and Non-Sampling Errors, Sources and treatment of non-sampling errors

Q10. State and explain sampling errors.

Ans :

In a sample survey, since only a small portion of the population is studied, its results are bound to differ from the census results and thus have a certain amount of error. This error would always be there, no matter that the sample is drawn at random and that it is highly representative. This error is attributed to fluctuations of sampling and is called sampling error. Sampling error is due to the fact that only a subset of the population (i.e., sample) has been used to estimate the population parameters and draw inferences about the population. Thus, sampling error is present only in a sample survey and is completely absent in census method.

Sampling errors are primarily due to the following reasons :

1. Faulty selection of the sample

Some of the bias is introduced by the use of defective sampling technique for the selection of a sample, e.g., purposive or judgment sampling in which the investigator deliberately selects a representative sample to obtain certain results. This bias can be overcome by strictly adhering to a simple random sample or by selecting a sample at random subject to restrictions which while improving the accuracy are of such nature that they do not introduce bias in the results.

2. Substitution

If difficulties arise in enumerating a particular sampling unit included in the random sample, the investigators usually substitute a convenient member of the population. This obviously leads to some bias since the characteristics possessed by the substituted unit will usually be different from those possessed by the unit originally included in the sample.

3. Faulty demarcation of sampling units

Bias due to defective demarcation of sampling units is particularly significant in area surveys such as agricultural experiments in the field or crop cutting surveys, etc. In such surveys, while dealing with border-line cases, it depends more or less on the discretion of the investigator whether to include them in the sample or not.

4. Error due to bias in the estimation method

Sampling method consists in estimating the parameters of the population by appropriate statistics computed from the sample. Improper choice of the estimation techniques might introduce the error. For example, in simple random sampling, if x_1, x_2, ..., x_n are observations on the n sampled units, then the sample variance.

5. Variability of the population

Sampling error also depends on the variability or heterogeneity of the population to be sample.

Q11. State and explain Non-sampling errors.

Ans :

Non-sampling errors are not attributed to chance and are a consequence of certain factors which arc within human control. In other words, they are due to certain causes which can be traced and may arise at any stage of the enquiry, viz., planning and execution of the survey and collection processing and analysis of the data. Non-sampling errors are thus present both in census surveys as well as sample surveys. Obviously, non-sampling errors will be of large magnitude in a census survey than in a sample survey because they increase with the increase in the number of units to be examined and enumerated. It is very difficult to prepare an

exhaustive list of all the sources of non-sampling errors. We enumerate below some of the important factors responsible for non-sampling errors in any survey (census or sample).

1. Faulty planning, including vague and faulty definitions of the population or the statistical units to be used, incomplete list of population-members (i.e., incomplete frame in case of sample survey).
2. Vague and imperfect questionnaire which might result in incomplete or wrong information.
3. Defective methods of interviewing and asking questions.
4. Vagueness about the type of the data to be collected.
5. Exaggerated or wrong answers to the questions which appeal to the pride or prestige or self-interest of the respondents. For example, a person may over-state his education or income or understate his age or he may give wrong statements to safeguard his self-interest.
6. Personal bias of the investigator.
7. Lack of trained and qualified investigators and lack of supervisory staff.
8. Failure of respondents' memory to recall the events or happenings in the past.
9. Non-response and Inadequate or Incomplete Response. Bias due to non-response results if in a house-to-house survey the respondent is not available in spite of repeated visits by the investigator or if the respondent refuses to furnish the information. Incomplete response error is introduced if the respondent is unable to furnish information on certain questions or if he is unwilling or even refuses to answer certain questions.
10. **Improper coverage.** If the objects of the survey are not precisely stated in clear cut terms, this may result in
    (i) the inclusion in the survey of certain units which are to be excluded, or
    (ii) the exclusion of certain units which were to be included in the survey under the objectives.

    For example, in a census to determine the number of individuals in the age group, say, 15 years to 55 years, more or less serious errors may occur in deciding whom to enumerate unless particular community or area is not specified and also the time at which the age is to be specified.
11. **Compiling Errors,** Wrong calculations or entries made during the processing and analysis of the data. Various operations of data processing such as editing and coding of the responses, punching of cards, tabulation and summarizing the original observations made in the survey arc a potential source of error. Compilation errors are subject to control through verification, consistency checks, etc.
12. **Publication Errors.** The errors committed during presentation and printing of tabulated results are basically due to two sources. The first refers to the mechanics of publication-the proofing error and the like. The other, which is of a more serious nature, lies in the failure of the survey organization to point out the limitations of the statistic

Q12. Explain the various sources and treatment of non-sampling errors.

Ans : **(Dec.-18, July-21)**

1. Errors in Planning or Definition

Planning mainly focuses on aims or objectives of the survey that are clearly stated in the planning phase. These objectives are converted into the following,

(a A set of definitions

(b) A set of specifications.

Non sampling errors in planning occur due to the fol-lowing reasons,

(i) Incomplete and inconsistent data specification.

(ii) Improper designing of questionnaires.

(iii) Errors in recording the observed measurements.

(iv) Errors in location of units.

(v) Insufficient skilled investigators and less number of supervisory staff.

2. Errors in Response

These errors occur while providing the responses given by the people (respondents) of certain category. Some of the other reasons are discussed below,

(i) Accidental Errors

These errors occur when a person misunderstands a ques-tion and gives irrelevant response to it. This irrelevant information which is unintentionally provided by the respondent leads to accidental errors.

(ii) Prestige Bias Errors

These errors are related to the prestige of a person who is interviewed. For instance, a person may provide incor-rect information regarding his/her education, income, age etc., in order to maintain their pride in the society.

(iii) Self-interest Errors

These errors occur when some people give wrong information in order to satisfy their self interests. For instance, a person may overrate his/her salary and under rate expenses.

(iv) Biasing Errors due to the Interviewer

An interviewer's beliefs, prejudices and way of ques-tioning or recording can influence the responses and its accuracy.

(v) Failure Error (Respondents Fail to Remember)

'Recall' is one of the most common source of error which is used for collecting information.

The questions in the survey are related to the past events and a responder may not remember the information and recall it at that particular instant.

3. Errors in Non Response Biases

These errors are usually found in house to house survey in the following cases,

(i) Respondent is unavailable.

(ii) Respondent refuses to answer certain questions.

(iii) Respondent is unable to answer each and every question.

So, due to this non response, population belonging to a certain category is excluded and therefore it results in some bias.

4. Errors in Coverage

These errors occur when the objectives of the survey are not clearly mentioned which results in the following,

(i) Includes certain topics which are not necessary.

(ii) Excludes the topics which are needed.

Example

A survey is conducted to find the number of people be-longing to a age group of 10 to 30. However, if area and time are not defined errors occur.

5. Errors in Compilation

These errors occur during the data processing operations such as editing, coding, tabulating, summarizing the actual observation in the survey. They are key sources for the occur-rence of compilation errors. However, these type of errors can be reduced by conducting data verification and checking data consistency.

6. Errors in Publication

These errors occur during the presentation and printing of calculated results. The two main sources for such errors are,

(i) Mechanics of publication-proofing errors.

(ii) Failing to identify the limitations of the statistics during the organizational survey.

Treatment for Non-sampling Errors

Methods of Reducing/Controlling Non-sampling Error

1. Employing qualified and trained personnel for the survey.
2. Providing adequate supervisory checks on the field work.
3. Pretesting or conducting a pilot survey.
4. Thorough editing and scrutiny of the results.
5. Effective checking of all the steps in the processing and analysis of data.
6. More effective follow up of non response cases.

1.2.4 Advantages and Limitations of Sampling

Q13. What are the benefits or advantages of sampling ?

Ans : **(Nov.-20)**

Advantages of sampling

Sampling ensures convenience, collection of intensive and exhaustive data, suitability in limited resources and better rapport. In addition to this, sampling has the following advantages also.

1. Low cost of sampling

If data were to be collected for the entire population, the cost will be quite high. A sample is a small proportion of a population. So, the cost will be lower if data is collected for a sample of population which is a big advantage.

2. Less time consuming in sampling

Use of sampling takes less time also. It consumes less time than census technique. Tabulation, analysis etc., take much less time in the case of a sample than in the case of a population.

3. Scope of sampling is high

The investigator is concerned with the generalization of data. To study a whole population in order to arrive at generalizations would be impractical.

Some populations are so large that their characteristics could not be measured. Before the measurement has been completed, the population would have changed. But the process of sampling makes it possible to arrive at generalizations by studying the variables within a relatively small proportion of the population.

4. Accuracy of data is high

Having drawn a sample and computed the desired descriptive statistics, it is possible to determine the stability of the obtained sample value. A sample represents the population from which its is drawn. It permits a high degree of accuracy due to a limited area of operations. Moreover, careful execution of field work is possible. Ultimately, the results of sampling studies turn out to be sufficiently accurate.

5. Organization of convenience

Organizational problems involved in sampling are very few. Since sample is of a small size, vast facilities are not required. Sampling is therefore economical in respect of resources. Study of samples involves less space and equipment.

6. Intensive and exhaustive data

In sample studies, measurements or observations are made of a limited number. So, intensive and exhaustive data are collected.

7. Suitable in limited resources

The resources available within an organization may be limited. Studying the entire universe is not viable. The population can be satisfactorily covered through sampling. Where limited resources exist, use of sampling is an appropriate strategy while conducting marketing research.

8. Better rapport

An effective research study requires a good rapport between the researcher and the respondents. When the population of the study is large, the problem of rapport arises. But manageable samples permit the researcher to establish adequate rapport with the respondents.

Q14. What are the limitations of sampling ?

Ans : **(Nov.-20)**

1. Chances of bias

The serious limitation of the sampling method is that it involves biased selection and thereby leads us to draw erroneous conclusions. Bias arises when the method of selection of sample employed is faulty. Relative small samples properly selected may be much more reliable than large samples poorly selected.

2. Difficulties in selecting a truly representative sample

Difficulties in selecting a truly representative sample produces reliable and accurate results only when they are representative of the whole group. Selection of a truly representative sample is difficult when the phenomena under study are of a complex nature. Selecting good samples is difficult.

3. In adequate knowledge in the subject

Use of sampling method requires adequate subject specific knowledge in **sampling technique**. Sampling involves statistical

analysis and calculation of probable error. When the researcher lacks specialized knowledge in sampling, he may commit serious mistakes. Consequently, the results of the study will be misleading.

4. Changeability of units

When the units of the population are not in homogeneous, the sampling technique will be unscientific. In sampling, though the number of cases is small, it is not always easy to stick to the, selected cases. The units of sample may be widely dispersed.

Some of the cases of sample may not cooperate with the researcher and some others may be inaccessible. Because of these problems, all the cases may not be taken up. The selected cases may have to be replaced by other cases. Changeability of units stands in the way of results of the study.

5. Impossibility of sampling

Deriving a representative sample is dificult when the universe is too small or too heterogeneous. In this case, census study is the only alternative. Moreover, in studies requiring a very high standard of accuracy, the sampling method may be unsuitable. There will be chances of errors even if samples are drawn most carefully.

1.3 SAMPLING METHODS

Q15. Explain the different types of sampling methods.

Ans :

The choice of an appropriate sampling design is of paramount importance in the execution of a sample survey and is generally made keeping in view the objectives and scope of the enquiry and the type of the universe to be sampled. The sampling techniques may be broadly classified as follows :

(i) Purposive or Subjective or Judgment Sampling.

(ii) Probability Sampling.

(iii) Mixed Sampling.

Probability sampling provides a scientific technique of drawing samples from the population according to some laws of chance in which each unit in the universe has some definite pre-assigned probability of being selected in the sample. Different types of sampling are in which :

(i) Each sample unit has an equal chance of being selected.

(ii) Sampling units have varying probability of being selected.

(iii) Probability of selection of a unit is proportional to the sample size.

Sampling design in which the sample units are selected partly according to some probability laws, and partly according to a fixed sampling rule (no use of chance), is known as Mixed Sampling.

Some of the important types of sampling schemes covered are given below :

(i) Simple Random Sampling

(ii) Stratified Random Sampling

(iii) Systematic Sampling

(iv) Multistage Sampling

(v) Quasi Random Sampling

(vi) Area Sampling

(vii) Simple Cluster Sampling

(viii) Multistage Cluster Sampling

(ix) Quota Sampling

The selection of the sample based on the theory of probability is also known as random selection and sometimes the probability sampling is also called Random Sampling. It should be borne in mind that in ordinary language randomness means haphazardness or without any purpose or definite law but in Statistics randomness is a well defined concept. According to Simpson and Kafka, "Random samples are characterized by the way in which they are selected. Randomness is not used in the sense of haphazard or hit or miss",

1.3.1 Subjective Method

Q16. State and Explain subjective sampling (Or) Purposive Sampling (Or) Judgement Sampling.

Ans : **(July-19)**

Purposive sampling is one in which the sample units are selected with definite purpose in view. For example, if we want to give the picture that tire standard of living has increased in the city of New

Delhi, we may take individuals in the sample from rich and posh localities like Defence Colony, South Extension, Golf Links, Jor Bagh, Chanakyapuri, Greater Kailash etc. and ignore the localities where low income group and the middle class families live. This sampling suffers from the drawback of favouritism and nepotism and does not give a representative sample of the population.

1.3.2 Probability Method

Q17. Define Probability Sampling.

Ans : **(July-19)**

The choice of an appropriate sampling design is of paramount importance in the execution of a sample survey and is generally made keeping in view the objectives and scope of the enquiry and the type of the universe to be sampled. The sampling techniques may be broadly classified as follows :

(i) Purposive or Subjective or Judgment Sampling.

(ii) Probability Sampling.

(iii) Mixed Sampling.

Probability sampling provides a scientific technique of drawing samples from the population according to some laws of chance in which each unit in the universe has some definite pre-assigned probability of being selected in the sample. Different types of sampling are in which :

(i) Each sample unit has an equal chance of being selected.

(ii) Sampling units have varying probability of being selected.

(iii) Probability of selection of a unit is proportional to the sample size.

1.3.3 Mixed sampling Method

Q18. Define Mixed Sampling.

Ans : **(Nov.-20, July-19)**

Sampling design in which the sample units are selected partly according to some probability laws and partly according to a fixed sampling rule is known as Mixed Sampling.

(i) Simple Random Sampling

(ii) Stratified Random Sampling

(iii) Systematic Sampling

(iv) Multistage Sampling

(v) Quasi Random Sampling

(vi) Area Sampling

(vii) Simple Cluster Sampling

(viii) Multistage Cluster Sampling

(ix) Quota Sampling

The selection of the sample based on the theory of probability is also known as random selection and sometimes the probability sampling is also called Random Sampling. It should be borne in mind that in ordinary language randomness means haphazardness or without any purpose or definite law but in Statistics randomness is a well defined concept. According to Simpson and Kafka, "Random samples are characterized by the way in which they are selected. Randomness is not used in the sense of haphazard or hit or miss",

Q19. Explain the notation and terminology used in simple random sampling.

Ans : **(Dec.-18)**

The notation and terminology used in simple random sampling are as follows,

1. The characteristics of the population are represented by upper case letters and samples are represented by lower case let-ters. The number of units in a population is denoted by N.
2. The units of the population (N) are denoted by U_1 U_2, U_3..,U_N.
3. The units of the sample is denoted by u_1, u_2, u_3,...,u_n.
4. The value of the character for the unit of the population is denoted by i^{th} where, i = 1,2,3,...N.
5. The value of the character for the i^{th} unit of the sample is denoted by 'y_i' where, i = 1,2,3,...,n.

   By considering the above conventions, population mean, sample mean, population Mean square, sample mean square are defined as follows,

(a) Population Mean $\left(\bar{Y}_N\right) = \frac{1}{N}\sum_{i=1}^{N} Y_i$

(b) Sample Mean $\left(\bar{y}_n\right) = \frac{1}{n}\sum_{i=1}^{n} y_i$

OR

Sample Mean $\left(\bar{y}_n\right) = \frac{1}{n}\sum_{i=1}^{n} a_i Y_i$

Here,

a_i = 1, When the i^{th} unit is considered in the sample

a. = 0, When the i^{th} unit is considered in the sample

(c) Population Mean Square

$$S^2 = \frac{1}{N-1}\sum_{i=1}^{N}\left(Y_i - \bar{Y}_N\right)^2$$

$$= \frac{1}{N-1}\left[\sum_{i=1}^{N} Y_i^2 - N\bar{Y}_N^2\right]$$

(d) Sample Mean Square,

$$s^2 = \frac{1}{n-1}\sum_{i=1}^{n}\left(y_i - \bar{y}_n\right)^2$$

$$= \frac{1}{n-1}\left[\sum_{i=1}^{n} y_i^2 - n\bar{y}_n^2\right]$$

6. The upper case English or Greek letters are used to denote the population parameters. Where as the small letters or the caps $(\wedge)$ symbol are used to denote the estimates of these parameters. For instance, the estimate of the population mean $\bar{Y}_N$ is denoted by $\bar{Y}_N$.

1.3.4 Methods of drawing random samples using with and without replacement

Q20. Discuss the methods of drawing samples using with or without replacement.

Ans :

A random sample may be selected by :

(i) Lottery Method.

(ii) Use of Table of Random Numbers.

(i) Lottery Method

The simplest method of drawing a random sample is the lottery system. This consists in identifying each and every member or unit of the population with a distinct number which is recorded on a slip or a card. These slips should be as homogeneous as possible in shape, size, colour, etc., to avoid the human bias. The lot of these slips or cards is a kind of miniature of the population for sampling purposes. If the population is small, then these slips are put in a bag and thoroughly shuffled and then as many slips as units needed in the sample are drawn one by one, the slips being thoroughly shuffled after each draw. The sampling units corresponding to the numbers on the selected slips will constitute a random sample. For example, let us suppose that we want to draw a random sample of 10 individuals from a population of 100 individuals. We assign the numbers 1 to 100, one number to each individual of the population and prepare 100 identical slips bearing the numbers from 1 to 100. These slips are then placed in a bag or container and shuffled thoroughly. Finally, a sample of 10 slips is drawn out one by one. The individuals bearing the numbers on these selected slips will constitute the desired sample.

If the population to be sampled is fairly large, then we may adopt the lottery method in which all the slips or cards are placed in a metal cylinder which is thrown into a large rotating drum working under a mechanical system. The rotation of the drum results in through mixing or randomisation of the cards. Then a sample of desired size *n* is drawn out of the container mechanically and the corresponding *n* sample units constitute the desired random sample.

The lottery method gives a sample which is quite independent of the properties of the population. It is One of the best and most commonly used methods of selecting random samples. It is quite frequently used in the random draw of prizes, in the Tambola games and so on.

(ii) Use of Table of Random Numbers

The lottery method described above is quite time consuming and cumbersome to use if the population to be sampled is sufficiently large. Moreover, in this method, it is not humanly possible to make all the slips or cards exactly alike and as such some bias is likely to be introduced. Statisticians

have avoided this difficulty by considering the random sampling number series. Most of these series are the results of actual sampling operations recorded for future use. The most practical and inexpensive method of selecting a random sample consists in the use of 'Random Number Tables', which have been so constructed that each of the digits 0, 1, 2, ..., 9 appears with approximately the same frequency and independently of each other. If we have to select a sample from a population of size $N(\leq 99)$, then the numbers can be combined two by two to give pairs from 00 to 99. Similarly if $N \leq 999$ or $N \leq 9999$ and so on, then combining the digits three by three (or four by four and so on), we get numbers from 000 to 999 or 0000 to 9999 and so on. Since each of the digits 0,1,2,....9 occurs with approximately the same frequency and independently of each other, so does each of the pairs 00 to 99, triplets 000 to 999 or quadruplets 0000 to 9999 and so on.

The method of drawing a random sample comprises the following steps :

(i) Identify N units in the population with the numbers 1 to N.

(ii) Select at random, any page of the 'random number table' and pick up the numbers in any row, column or diagonal at random.

(iii) The population units corresponding to the numbers selected in step (ii) constitute the random sample.

We give below different sets of random numbers commonly used in practice. The numbers in these tables have been subjected to various statistical tests for randomness of a series and their randomness has been well established for all practical purposes.

1. **Tippet's (1927) Random Number Tables.** Tippet random number tables consist of 10,400 four-digited numbers, giving in all 10,400 × 4, i.e.. 41,600 digits selected at random from the British census reports.
2. **Fisher and Yates (1938)** Tables (in Statistical Tables for Biological. Agricultural and Medical Research) comprise 15,000 digits arranged in two's. Fisher and Yates obtained these tables by drawing numbers at random from the 10th to 19th digits of A.S. Thomson's 20-figures logarithmic tables.
3. **Kendall and Babington Smiths (1939)** random tables consist of 100,000 digits grouped into 25,000 sets of 4-digited random numbers (Tracts for computers, No. 24, Cambridge University Press).
4. **Rand Corporation (1955),** (Free Press, Illinois) random number tables consist of one million random digits consisting of 200,000 random numbers of 5 digits each.
5. **Table of Random Numbers** (The IS1 series, Calcutta) by C.R. Rao, Mitra and Mathai.

PROBLEMS

1. Draw a random sample (without replacement) of 15 students from a class of 450 students.

2952	6641	3992	9792	7979	5911	3170	5624
4167	9524	1545	1396	7203	5356	1300	2693
2370	7483	3408	2762	3563	1089	6913	7691
0560	5246	1112	6107	6008	8126	4233	8776
2754	9143	1405	9025	7002	6111	8816	6446

Sol :

First of all we identify the 450 students of the college with numbers from 1 to 450. Starting with the first number in the above extract from Tippett's random number tables and moving row-wise, we pick out, one by one, the three-digited numbers less than or equal to 450, till 15 numbers < 450 are obtained. In this process the numbers over 450 are discarded and the repeated numbers, if any, are taken only once.

The above numbers, grouped in three's are :

295, 266, 413, 992, 979, 279, 795, 911, 317, 056, 244, 167, 952,
415, 451, 396, 720, 353, 561, 300, 269, 323, 707, 483, 340 ...

Thus the students corresponding to the numbers

295, 266, 413, 279, 317, 56, 244, 167
415 396, 353, 300,

constitute the desired random sample of size I5.

2. The adjoining table of ten random numbers of two digits each is provided to the field investigator.

34	**96**	**61**	**85**	**49**	
78	**50**	**02**	**27**	**13**	**...(*)**

How should he use this table to make a random selection of 5 plots out of 40 ?

Sol :

In this case we shall first identify the 40 plots with the numbers 1 to 40. In the table in (*) there are only 3 numbers, *viz.,* 34, 02 and 13 which are less than 40 and accordingly we are not able to draw the desired sample of size 5 from this table as such. In this case, we shall assign more than one number to each of the sampling units, *i.e.,* plots. For example, the first plot will be assigned the numbers

Table

Number from Table (*)	Number of the Sampled Plot
34	34
96 = 16 + 2 × 40	16
61 = 21 + 40	21
85 = 5 + 2 × 40	5
49 = 9 + 40	9

01, 01 + 40, 01+2 × 40, 01+3 × 40, ...and so on, *i.e.,*1, 41, 81, 121, 161, 201,... and so on.

Similarly the second plot will be assigned the numbers

02, 02 + 40, 02 + 2 × 40, 02 + 3 × 40, ...

i.e., 02, 42, 82, 122, 162, 202, ..., and soon.

Finally, the last plot, *i.e.,* 40th plot can be assigned the numbers 0, 40, 80, 120, 160, ... and so on.

If we select the first number in (*) and move row-wise we get the adjoining Table.

Thus, the plot Nos. 5, 9, 16, 21 and 34 constitute the desired sample.

If we select the first number in (*) and move column wise, the desired sample consists of plot numbers :

34, 38, 16, 10 and 21,

because, 78 = 40 + 38, 96 = 2 × 40 + 16, 50 = 40+ 10 and 61=40 +21

3. Draw a random sample of size 5 from normal population with mean $\mu = 25$ and S.D, $\sigma = 15$.

Sol :

Given that,

Size of ihe sample, n = 5

Mean of the population, $\mu = 25$

Standard deviation of the population, $\sigma = 15$

Selecting randomly, any five 4-digited random numbers from the random numbers table as given below,

9792

1396

2762

6107

9025

Let the above observations be where, i = 1,2,3,4,5

We know that,

$F(x) = P(X \leq x) = 10^{-k} \times$ (k-digit random number)

Therefore,

$P(z \leq z_1) = 10^{-4} \times 9792 = 0.9792$ and $z_1 = 2.04$ ($\because$ from normal table)

$P(z \leq z_2) = 10^{-4} \times 1396 = 0.1396$ and $z_2 = -1.1$ ($\because$ from normal table)

$P(z \leq z_3) = 10^{-4} \times 2762 = 0.2762$ and $z_3 = -0.59$ ($\because$ from normal table)

$P(z \leq z_4) = 10^{-4} \times 6107 = 0.6107$ and $z_4 = 0.28$ ($\because$ from normal table)

$P(z < z_5) = 10^{-4} \times 9025 = 0.9025$ and $z_5 = 1.30$ ($\because$ from normal table)

We have,

$$z = \frac{X - \mu}{\sigma}$$

$\Rightarrow 2\sigma = X - \mu$

$\Rightarrow 2\sigma + \mu = X$

$x_1 = z_1\sigma + \mu$

$= 2.04 \times 15 + 25$

$= 30.60 + 25$

$= 55.60$

$x_2 = z_2 \times \sigma + \mu$

$= -1.1 \times 15 + 25$

$= -16.50 + 25$

$= 8.50$

$x_3 = z_3 \times \sigma + \mu$

$= -0.59 \times + 25$

$= -8.85 + 25$

$= 16.15$

$x_4 = z_4 \times \sigma + \mu$

$= 0.28 \times 15 + 25$

$= 4.20 + 25$

$= 29.20$

$x_5 = z_5 \times \sigma + \mu$

$= 1.30 \times 15 + 25$

$= 19.50 + 25$

$= 44.50$

Therefore, the random samples of size 5 from normal population with mean 25 and standard deviation 15 are 55.60,8.50, 16.15, 29.20 and 44.50 respectively.

1.3.5 Estimates of population mean, total, and proportion, their variances

Q21. How to estimate population mean, total, and proportion, their variances ?

Ans :

(i) Estimation of Population Mean

The population mean can be estimated by considering the sample mean.

Sample mean, $\bar{y} = \frac{1}{n}\sum_{i=1}^{n} y_i$ be an estimator of population mean

Population mean $\bar{Y}_N = \frac{1}{N}\sum_{i=1}^{N} Y_i$

Then, $\bar{y}$ is an unbiased estimator of $\bar{Y}$

(ii) Estimation of Population Total

Let T represents the population total. Then,

$$T = \sum_{i=1}^{N} Y_i$$

$$= N\bar{Y}$$

The estimated population total can be written as,

$$\hat{T} = N\hat{\bar{Y}}$$

$$= N\bar{Y}$$

Therefore,

$$E(\hat{T}) = NE(\bar{Y})$$

$$= N\bar{Y}$$

(iii) Estimation of Variance Population Total

The estimated population total is given by,

Where,

N - Population

$\bar{y}$ - Sample mean

Therefore, the variance of estimated total will be,

Variance of estimated total $= N^2 V(\bar{y})$

Where, $V(\bar{y})$ is the variance of sample mean $\bar{x}$.

The estimated sampling variance of total can then be written as,

$$\text{Estimated sampling variance of total} = N^2 \times \frac{N-n}{N} \times \frac{S^2}{n}$$

Where $\frac{(N-n)}{N} \times \frac{S^2}{n}$ is the estimated sampling variance of mean.

(iv) Estimation of Population Proportion

We know that, sample mean is an unbiased estimate of population mean,

i.e., $E(\bar{y}) = \bar{Y}$

$E(p) = P$

i.e., sample proportion p is an unbiased estimate of population proportion.

$E(p) = P$

$E(Np) = NP = A$

This, NP is an unbiased estimate of 'A' which can be expressed as,

$$\hat{A} = NP$$

(v) Estimation of Variance of Population Proportion

The population proportion with the attributes (P, 1 – P) is given by,

$$P = \frac{1}{N}\sum_{i=1}^{N} y_i = \mu$$

The population variance is given by,

$$\sigma^2 = \frac{\sum_{i=1}^{N}(y_i - \mu)^2}{N-1}$$

$$= \frac{\sum_{i=1}^{N}(y_i - P)^2}{N-1} \qquad [\because \text{From equation (1)}]$$

$$= \frac{\sum y^2 - NP^2}{N-1}$$

$$= \frac{NP - NP^2}{N-1}$$

$$= \frac{PN(1-P)}{N-1}$$

$$= \frac{N}{N-1}(1-P)P$$

Let $\hat{P}$ represents the sample proportion having attribute,

$$\hat{P} = \frac{1}{n}\sum_{i=1}^{n} y_i = \bar{y}$$

The sample variance is given by,

$$s^2 = \frac{\sum_{i=1}^{n}(y_i - \bar{y})^2}{n-1}$$

$$= \frac{\sum y_1^2 - n\bar{y}}{n-1}$$

$$= \frac{\sum y_1^2 - n\hat{P}}{n-1} \qquad [\because \text{From equation (2)}]$$

$$= \frac{n\hat{P}(1-\hat{P})}{n-1}$$

$$= \frac{n}{n-1}(1-\hat{P})$$

$$= \frac{n}{n-1}(1-\hat{P})\hat{P}$$

The sample proportion is an unbiased for the population proportion with the variance,

$$\text{Var}\ (\hat{P}) = \left(\frac{N-n}{n-1}\right)\frac{P(1-P)}{n}$$

And the unbiased estimator of this variance is given by,

$$\overline{\text{Var}}(\hat{P}) = \left(\frac{N-n}{N-1}\right)\frac{\hat{P}(1-P)}{n}$$

Therefore, the estimate of variance of population proportion is,

$$\overline{\text{Var}}(\hat{P}) = \left(\frac{N-n}{N}\right)\frac{\hat{P}(1-P)}{n-1}$$

1.3.6 Estimates of variances in Simple Random Sampling With and Without Replacement

Q22. Explain Simple Random Sampling without Replacement (SRSWOR).

Ans : **(July-19, Dec.-18)**

➢ **Notations**

Consider the population size 'N' with the observations are $Y_1, Y_2, ..., Y_n$.

Population mean $\bar{Y} = \frac{1}{N}\sum_{i=1}^{N} Y_i$

Consider the sample size 'n with the observations are $y_1, y_2, ..., y_n$.

Sample mean $\bar{y} = \frac{1}{n}\sum_{i=1}^{N} y_i$

Sample mean $\bar{y}$ may also be written alternatively as $\bar{y} = \frac{1}{n}\sum_{i=1}^{N} a_i y_i$ (sample mean is written in the form of population mean).

➢ **Population Variance**

$$S^2 = \frac{1}{N-1}\sum_{i=1}^{N}(Y_i - \bar{Y})^2$$

$$= \frac{1}{N-1}\sum_{i=1}^{N}(Y_i^2 + \bar{Y}^2 - 2Y_i\bar{Y})$$

$$= \frac{1}{N-1}\left\{\sum_{i=1}^{N} Y_i^2 + N\bar{Y}^2 - 2\bar{Y}\sum_{i=1}^{N} Y_i\right\} \qquad \left[\because \bar{Y} = \frac{1}{N}\sum_{i=1}^{N} Y_i \sum_{i=1}^{N} Y_i = N\bar{Y}\right]$$

$$= \frac{1}{N-1}\left\{\sum_{i=1}^{N} Y_i^2 + N\bar{Y}^2 - 2\bar{Y}\sum_{i=1}^{N} Y_i\right\}$$

$$= \frac{1}{N-1}\left\{\sum_{i=1}^{N} Y_i^2 + N\bar{Y}^2 - 2\bar{Y}N\bar{Y}\right\}$$

$$= \frac{1}{N-1}\left\{\sum_{i=1}^{N} Y_i^2 + N\bar{Y}^2 - 2N\bar{Y}^2\right\}$$

$$S^2 = \frac{1}{N-1}\left\{\sum_{i=1}^{N} Y_i^2 - N\bar{Y}^2\right\}$$

Sample Variance

$$s^2 = \frac{1}{n-1}\left\{\sum_{i-1}^{n} y_i^2 - n\bar{y}^2\right\}$$

$$\left(\sum_{i=1}^{3} y_i\right)^2 = (y_1 + y_2 + y_3)^2$$

$$= y_1^2 + y_2^2 + y_3^2 + 2y_1y_2 + 2y_2y_3 + 2y_1y_3$$

$$\left(\sum_{i=1}^{3} y_i\right)^2 = \sum_{i=1}^{3} y_i^2 + \sum_{i=1}^{3}\sum_{j=1}^{3} y_i y_j$$

$$\sum_{i=1}^{3}\sum_{j=1}^{3} y_i y_j = \sum_{j=2,3} y_i y_j + \sum_{j=1,3} y_2 y_j + \sum_{j=1,2} y_3 y_j$$

$$= y_1y_2 + y_1y_3 + y_2y_1 + y_2y_3 + y_3y_1 + y_3y_2$$

$$= 2y_1y_2 + 2y_1y_3 + 2y_3y_2$$

$$\left(\sum_{i=1}^{n} y_i\right)^2 = \sum_{i=1}^{n} y_1^2 + \sum_{i=1}^{n}\sum_{j=1}^{n} y_i y_j$$

$$\sum_{i=1}^{n}\sum_{j=1}^{n} y_i y_j = \left(\sum_{i=1}^{n} y_i\right)^2 - \sum_{i=1}^{n} y_1^2$$

$$E(a_i) = \frac{n}{N}$$

$$E(a_i a_j) = \frac{n}{N}\,\frac{(n-1)}{N-1}$$

THEOREMS :

Theorem 1

In a simple random sampling without replacement the sample mean is an unbiased estimate of the population mean.

$$E(\bar{y}) = \bar{Y}$$

Proof : **(Nov.-20, July-19, Dec.-18, Dec.-18(MGU)**

Consider sample mean

$$\bar{y} = \frac{1}{n}\sum_{i=1}^{n} y_i \quad \text{(or)} \quad \bar{y} = \frac{1}{n}\sum_{i=1}^{N} a_i Y_i$$

Taking expectations on both sides

$$E(\bar{y}) = E\left[\frac{1}{n}\sum_{i=1}^{N} a_i Y_i\right]$$

when we are using a_i we should write y_i's are in Y_i.

$$= \frac{1}{n} \sum_{i=1}^{n} Y_i \, E(a_i)$$

$$= \frac{1}{n} \sum_{i=1}^{n} Y_i \cdot \left(\frac{n}{N}\right) \qquad [\because \text{ from notation (7)}]$$

$$= \frac{1}{N} \sum_{i=1}^{n} Y_i$$

$$= \bar{Y}$$

$$\therefore \quad E(\bar{y}) = \bar{Y}$$

$\therefore$ Sample mean is an unbiased estimate of the population mean.

Theorem 2

In simple random sampling without replacement the sample mean square is an unbiased estimate of the population mean square.

$$E(s^2) = S^2$$

Proof :

$$s^2 = \frac{1}{n-1} \left[\sum_{i=1}^{n} y_i^2 - n\bar{y}^2\right]$$

$$= \frac{1}{n-1} \left[\sum_{i=1}^{n} y_i^2 - n\left[\frac{1}{n}\sum_{i=1}^{n} y_i\right]^2\right]$$

$$= \frac{1}{n-1} \left[\sum_{i=1}^{n} y_i^2 - \frac{1}{n}\left[\sum_{i=1}^{n} y_i\right]^2\right]$$

$$= \left\{\frac{1}{n-1}\left[\sum_{i=1}^{n} y_i^2 - \frac{1}{n}\left[\sum_{i=1}^{n} y_i^2 + \sum_{i=1}^{n}\sum_{j=1}^{n} y_i y_j\right]\right]\right\} \qquad [\because \text{ from notation (6)}]$$

$$= \frac{1}{n-1} \left[\sum_{i=1}^{n} y_i^2 - \frac{1}{n}\sum_{i=1}^{n} y_i^2 - \frac{1}{n}\sum_{i=1}^{n}\sum_{j=1}^{n} y_i y_j\right]$$

$$= \frac{1}{n-1} \left[\sum_{i=1}^{n} y_i^2\left[1 - \frac{1}{n}\right] - \frac{1}{n}\sum_{i=1}^{n}\sum_{j=1}^{n} y_i y_j\right]$$

$$= \frac{1}{n-1} \left[\sum_{i=1}^{n} y_i^2\left[\frac{n-1}{n}\right] - \frac{1}{n}\sum_{i=1}^{n}\sum_{j=1}^{n} y_i y_j\right]$$

$$= \frac{1}{n-1} \frac{n-1}{n} \sum_{i=1}^{n} y_i^2 - \frac{1}{n(n-1)} \sum_{i=1}^{n}\sum_{j=1}^{n} y_i \, y_j$$

$$s^2 = \frac{1}{n}\sum_{i=1}^{n} y_i^2 - \frac{1}{n(n-1)}\sum_{i=1}^{n}\sum_{j=1}^{n} y_i y_j$$

when we use 'a' we should write y_i's are in Y_i's

$$s^2 = \frac{1}{n}\sum_{i=1}^{n} a_i Y_i^2 - \frac{1}{n(n-1)}\sum_{i=1}^{N}\sum_{j=1}^{N} a_i a_j Y_i Y_j$$

Taking expectation on both sides

$$E(s^2) = E\left[\frac{1}{n}\sum_{i=1}^{N} a_i Y_i^2 - \frac{1}{n(n-1)}\sum_{i=1}^{N}\sum_{j=1}^{N} a_i a_j Y_i Y_j\right]$$

$$= \frac{1}{n}\sum_{i=1}^{N} Y_i^2 E(a_i) - \frac{1}{n(n-1)}\sum_{i=1}^{N}\sum_{j=1}^{N} Y_i Y_j E(a_i a_j)$$

$$= \frac{1}{n}\sum_{i=1}^{N} Y_i^2 \frac{n}{N} - \frac{1}{n(n-1)}\sum_{i=1}^{N}\sum_{j=1}^{N} Y_i Y_j \frac{n(n-1)}{N(N-1)}$$

$$= \frac{1}{N}\sum_{i=1}^{N} Y_i^2 - \frac{1}{N(N-1)}\sum_{i=1}^{N}\sum_{j=1}^{N} Y_i Y_j$$

$$= \frac{1}{N}\sum_{i=1}^{N} Y_i^2 - \frac{1}{N(N-1)}\left[\left[\sum_{i=1}^{N} Y_i\right]^2 - \sum_{i=1}^{N} Y_i^2\right] \quad [\because \text{ from notation (6)}]$$

$$= \frac{1}{N}\sum_{i=1}^{N} Y_i^2 + \frac{1}{N(N-1)}\sum_{i=1}^{N} Y_i^2 - \frac{1}{N(N-1)}\left(\sum_{i=1}^{N} Y_i\right)^2$$

$$= \frac{1}{N}\sum_{i=1}^{N} Y_i^2\left[1 + \frac{1}{(N-1)}\right] - \frac{1}{N(N-1)}\left(\sum_{i=1}^{N} Y_i\right)^2$$

$$= \frac{1}{N}\sum_{i=1}^{N} Y_i^2\left[\frac{N-1+1}{N-1}\right] - \frac{1}{N(N-1)}\left(\sum_{i=1}^{N} Y_i\right)^2$$

$$= \frac{1}{N}\sum_{i=1}^{N} Y_i^2 \cdot \frac{N}{N-1} - \frac{1}{N(N-1)}\left(\sum_{i=1}^{N} Y_i\right)^2$$

$$= \frac{1}{N}\sum_{i=1}^{N} Y_i^2 - \frac{1}{N(N-1)}\left(\sum_{i=1}^{N} Y_i\right)^2 \qquad \left[\because \frac{1}{N}\sum_{i=1}^{N} Y_i = \bar{Y}\ \sum_{i=1}^{N} Y_i = N\bar{Y}\right]$$

$$= \frac{1}{N}\sum_{i=1}^{N} Y_i^2 - \frac{1}{N(N-1)}(N\bar{Y})^2$$

$$= \frac{1}{N}\sum_{i=1}^{N} Y_i^2 - \frac{N^2}{N(N-1)}\bar{Y}^2$$

$$= \frac{1}{N}\sum_{i=1}^{N} Y_i^2 - \frac{N}{N-1}\bar{Y}^2$$

$$= \frac{1}{N}\left[\sum_{i=1}^{N} Y_i^2 - N\bar{Y}^2\right]$$

$E(s^2) = S^2$

$\therefore$ Sample mean square is an unbiased estimate of population mean square.

Theorem 3

In SRSWOR the variance of the sample mean is given by

$$\textbf{Var}(\bar{y}) = \frac{S^2}{n} \cdot \frac{N-n}{N}$$

Proof : **(July-21)**

$$\text{Var}(\bar{y}) = E(\bar{y}^2) - [E(\bar{y})]^2$$

$$= E(\bar{y}^2) - \bar{y}^2$$

$$= E\left[\frac{1}{n}\sum_{i-1}^{n} y_i\right]^2 - \bar{y}^2$$

$$= \frac{1}{n^2} E\left[\sum_{i=1}^{n} y_i\right]^2 - \bar{y}^2$$

$$= \frac{1}{n^2} E\left[\sum_{i=1}^{n} y_i^2 + \sum_{i=-1}^{n}\sum_{j=1}^{n} y_i y_j\right] - \bar{y}^2$$

When we use 'a' we should write y_i's are in Y_i's

$$= \frac{1}{n^2} E\left[\sum_{i=1}^{N} a_i Y_i^2 + \sum_{i=1}^{N}\sum_{j=1}^{N} a_i a_j Y_i Y_j\right] - \bar{Y}^2$$

$$= \frac{1}{n^2}\sum_{i=1}^{N} Y_i^2 E(a_i) + \frac{1}{n^2}\sum_{i=1}^{N}\sum_{j=1}^{N} Y_i Y_j E(a_i a_j) - \bar{Y}^2$$

$$= \frac{1}{n^2}\sum_{i=1}^{N} Y_i^2 \cdot \frac{n}{N} + \frac{1}{n^2}\sum_{i=1}^{N}\sum_{j=1}^{N} y_i y_j \frac{n(n-1)}{N(N-1)} - \bar{Y}^2$$

$$= \frac{1}{nN}\sum_{i=1}^{N} Y_i^2 + \frac{(n-1)}{nN(N-1)}\sum_{i=1}^{N}\sum_{j=1}^{N} Y_i Y_j - \bar{Y}^2$$

$$= \frac{1}{nN}\sum_{i=1}^{N} Y_i^2 + \frac{(n-1)}{nN(N-1)}\left[\left(\sum_{i=1}^{N} y_i\right)^2 - \sum_{i=1}^{N} y_i^2\right] - \bar{Y}^2$$

[$\because$ from notation (6)]

$$= \frac{1}{nN}\sum_{i=1}^{N} Y_i^2 + \frac{(n-1)}{nN(N-1)}\left(\sum_{i=1}^{N} y_i\right)^2 - \frac{(n-1)}{nN(N-1)}\sum_{i=1}^{N} Y_i^2 - \bar{Y}^2$$

$$= \frac{1}{nN}\sum_{i=1}^{N} Y_i^2\left[1-\frac{(n-1)}{N-1}\right] + \frac{(n-1)}{nN(N-1)}\sum_{i=1}^{N} Y_i^2 - \bar{Y}^2$$

$$= \frac{1}{nN}\sum_{i=1}^{N} Y_i^2\left[\frac{N-1-n+1}{N-1}\right] + \frac{(n-1)}{nN(N-1)}\left(\sum_{i=1}^{N} y_i\right)^2 - \bar{Y}^2$$

$$= \frac{1}{nN}\sum_{i=1}^{N} Y_i^2\left[\frac{N-n}{N-1}\right] + \frac{(n-1)}{nN(N-1)}\left(\sum_{i=1}^{N} Y_i\right)^2 - \bar{Y}^2$$

$$= \frac{1}{nN}\sum_{i=1}^{N} Y_i^2\left[\frac{N-n}{N-1}\right] + \frac{n-1}{nN(N-1)}\left(N\bar{Y}\right)^2 - \bar{Y}^2 \quad [\because \Sigma Y_i = N\bar{Y}]$$

$$= \frac{N-n}{nN(N-1)}\sum_{i=1}^{N} Y_i^2 + \frac{n-1}{nN(N-1)}N^2\bar{Y}^2 - \bar{Y}^2$$

$$= \frac{N-n}{nN(N-1)}\sum_{i=1}^{N} Y_i^2 - \bar{Y}^2\left[1-\frac{N(n-1)}{n(N-1)}\right]$$

$$= \frac{N-n}{nN(N-1)}\sum_{i=1}^{N} Y_i^2 - \bar{Y}^2\left[\frac{n(N-1)-N(n-1)}{n(N-1)}\right]$$

$$= \frac{N-n}{nN(N-1)}\sum_{i=1}^{N} Y_i^2 - \bar{Y}^2\left[\frac{nN-n-Nn+N}{n(N-1)}\right]$$

$$= \frac{N-n}{nN(N-1)}\sum_{i=1}^{N} Y_i^2 - \bar{Y}^2\left[\frac{N-n}{n(N-1)}\right]$$

$$= \frac{N-n}{nN(N-1)}\left[\sum_{i-1}^{N} Y_i^2 - N\bar{Y}^2\right]$$

$$= \frac{N-n}{nN}\left[\frac{1}{N-1}\left[\sum_{i=1}^{N} Y_i^2 - N\bar{Y}^2\right]\right]$$

$$\text{Var}(\bar{y}) = \frac{N-n}{nN}S^2$$

$\therefore$ L.H.S = R.H.S

Theorem 4

If X_i, Y_i are the pair of variates define for every unit (i = 1, 2, ..., N) of the population and $\bar{x}$ and $\bar{y}$ are the corresponding sample means of a simple random sample of size 'n' taken without replacement then. Prove that cov $(\bar{X}, \bar{Y}) = \frac{N-n}{nN} \cdot \frac{1}{N-1} \sum_{i=1}^{N} (X_i - \bar{X})(Y_i - \bar{Y})$.

Proof :

Let $U_i = X_i + Y_i$

$$\bar{U} = \frac{1}{N} \sum_{i=1}^{N} (U_i)$$

$$= \frac{1}{N} \sum_{i=1}^{N} (X_i + Y_i)$$

$$= \frac{1}{N} \left[\sum_{i=1}^{N} X_i + \sum_{i=1}^{N} Y_i \right]$$

$$= \frac{1}{N} \sum_{i=1}^{N} X_i + \frac{1}{N} \sum_{i=1}^{N} Y_i$$

$$= \bar{X} + \bar{Y}$$

$$\text{Var}(\bar{U}) = \text{Var}(\bar{X} + \bar{Y})$$

$$= \text{Var}(\bar{X}) + \text{Var}(\bar{Y}) + 2\,\text{Cov}(\bar{X}, \bar{Y})$$

$$2\,\text{Cov}(\bar{X}, \bar{Y}) = \text{Var}(\bar{U}) - \text{Var}(\bar{X}) - \text{Var}(\bar{Y}) \quad \ldots (1)$$

$$\text{Var}(\bar{U}) = \frac{N-n}{Nn} S^2 \rightarrow (*) \qquad [\because \text{ from theorem (3)}]$$

$$S^2 = \frac{1}{N-1} \sum_{i=1}^{N} (U_i - \bar{U})^2 \quad \ldots (2)$$

$$= \frac{1}{N-1} \sum_{i=1}^{N} (X_i + Y_i - (\bar{X} + \bar{Y})]^2$$

$$= \frac{1}{N-1} \sum_{i=1}^{N} [(X_i - \bar{X}) + (Y_i - \bar{Y})]^2$$

$$= \frac{1}{N-1} \sum_{i=1}^{N} [(X_i - \bar{X})^2 + (Y_i - \bar{Y})^2 + 2(X_i - \bar{X})(Y_i - \bar{Y})]$$

$$= \frac{1}{N-1} \sum_{i=1}^{N} (X_i - \bar{X})^2 + \frac{1}{N-1} \sum_{i=1}^{N} (Y_i - \bar{Y}) + \frac{1}{N-1} 2 \sum_{i=1}^{N} (X_i - \bar{X})(Y_i - \bar{Y})$$

$$= S_X^2 + S_Y^2 + \frac{2}{N-1}\sum_{i=1}^{N}(X_i - \bar{X})(Y_i - \bar{Y}) \quad \ldots (3)$$

Substitute (3) in (*)

$$\text{Var}(\bar{U}) = \frac{N-n}{Nn}\left[S_X^2 + S_Y^2 + 2.\frac{1}{N-1}\sum_{i=1}^{N}(X_i - \bar{X})(Y_i - \bar{Y})\right]$$

$$= \frac{N-n}{Nn}S_X^2 + \frac{N-n}{Nn}S_Y^2 + 2.\frac{N-n}{Nn(N-1)}\sum_{i=1}^{N}(X_i - \bar{X})(Y_i - \bar{Y})$$

$$= \text{Var}(\bar{X}) + \text{Var}(\bar{Y}) + 2.\frac{N-n}{Nn(N-1)}\sum_{i=1}^{N}(X_i - \bar{X})(Y_i - \bar{Y}) \quad \ldots (4)$$

Substitute (4) in (1), we get $\left[\because \text{Var}(\bar{X}) = \frac{N-n}{Nn}S_x^2\right]$

$$2\text{Cov}(\bar{X}, \bar{Y}) = \text{Var}(\bar{X}) + \text{Var}(\bar{Y}) + 2.\frac{N-n}{Nn(N-1)}\sum_{i=1}^{N}(X_i - \bar{X})(Y_i - \bar{Y})$$

$$- \text{Var}(\bar{X}) + \text{Var}(\bar{Y})$$

$$\therefore \quad \text{Cov}(\bar{X}, \bar{Y}) = \frac{N-n}{Nn(N-1)}\sum_{i=1}^{N}(X_i - \bar{X})(Y_i - \bar{Y})$$

$\therefore$ Hence proved.

Q23. Explain Simple Random Sampling with Replacement (SRSWR).

Ans : **(July-19)**

Notations

1. Consider b_i i^{th} unit of observation in the sample.

   The probability of each unit of the i^{th} observation in each drawn is $\frac{1}{N}$.

   $$E(bi) = \frac{n}{N}$$

2. Variance of b_i

   $$V(b_i) = npq$$

   $$= n\left(\frac{1}{N}\right)\left(1 - \frac{1}{N}\right)$$

   $$= \frac{n}{N}\left(\frac{N-1}{N}\right)$$

   $$= \frac{n(N-1)}{N^2}$$

3. $\text{Cov}(b_i, b_j) = \frac{-n}{N^2}$

4. $\text{Var}\left(\sum_{i=1}^{N} b_i y_i\right) = \sum_{i=1}^{N} y_i^2 \text{ Var}(b_i) + \sum_{i=1}^{N}\sum_{j=1}^{N} y_i y_j \text{ Cov}(b_i, b_j)$

5. $E(\bar{y}^2) = \text{Var}(\bar{Y}) + [E(\bar{Y})]^2$

Theorem 5

Sample mean is an unbiased estimator of population mean $E(\bar{y}) = \bar{Y}$.

Proof :

Consider sample mean

$$\bar{y} = \frac{1}{n}\sum_{i=1}^{N} y_i$$

(or)

$$\bar{y} = \frac{1}{n}\sum_{i=1}^{N} b_i Y_i$$

Taking expectation on both sides

$$E(\bar{y}) = E\left[\frac{1}{n}\sum_{i=1}^{N} b_i Y_i\right]$$

$$= \frac{1}{n}\sum_{i=1}^{N} Y_i \cdot E(b_i)$$

$$= \frac{1}{n}\sum_{i=1}^{N} Y_i \cdot \frac{n}{N}$$

$$= \frac{1}{N}\sum_{i=1}^{N} Y_i$$

$$= \bar{Y}$$

$\therefore$ $E(\bar{y}) = \bar{Y}$

Hence proved.

Theorem 6

Simple random sample with replacement prove that $\text{Var}(\bar{y}) = \frac{\sigma^2}{n}$

Proof :

$$\text{Var}(\bar{y}) = \text{Var}\left(\frac{1}{n}\sum_{i=1}^{N} y_i\right)$$

$$= \text{Var}\left(\frac{1}{n}\sum_{i=1}^{N} b_i\, y_i\right)$$

$$= \frac{1}{n^2}\,\text{Var}\left(\sum_{i=1}^{N} b_i\, y_i\right)$$

$$= \frac{1}{n^2}\left[\sum_{i=1}^{N} y_i^{\,2}\,\text{Var}(b_i) + \sum_{i=1}^{N}\sum_{j=1}^{N}\text{Cov}(b_i b_j)\right] \qquad [\because \text{ from notation (4)}]$$

$$= \frac{1}{n^2}\left[\sum_{i=1}^{N} y_i^{\,2}.\frac{n(N-1)}{N^2} + \sum_{i=1}^{N}\sum_{i=1}^{N} y_i y_j\left(\frac{-n}{N^2}\right)\right]$$

$$= \frac{1}{n^2}\left[\frac{n(N-1)}{N^2}\sum_{i=1}^{N} y_i^{\,2} - \frac{n}{N^2}\left[\left(\sum_{i=1}^{N} y_i\right)^2 - \sum_{i=1}^{N} y_i^{\,2}\right]\right]$$

$$= \frac{n}{n^2}\left[\frac{(N-1)}{N^2}\sum_{i=1}^{N} y_i^{\,2} - \frac{1}{N^2}\left(\sum_{i=1}^{N} y_i\right)^2 + \frac{1}{N^2}\sum_{i=1}^{N} y_i^{\,2}\right]$$

$$= \frac{1}{n}\left[\frac{1}{N^2}\sum_{i=1}^{N} y_i^{\,2}\,[N-1+1] - \frac{1}{N^2}\left(\sum_{i=1}^{N} y_i\right)^2\right]$$

$$= \frac{1}{n}\left[\frac{1}{N^2}\sum_{i=1}^{N} y_i^{\,2}(N) - \frac{1}{N^2}\left(\sum_{i=1}^{N} y_i\right)^2\right] \qquad \left[\because \sum_{i=1}^{N} y_i - N\bar{y}\right]$$

$$= \frac{1}{n}\left[\frac{1}{N}\sum_{i=1}^{N} y_i^{\,2} - \frac{1}{N^2}(N\bar{y})^2\right]$$

$$= \frac{1}{n}\left[\frac{1}{N}\sum_{i=1}^{N} y_i^{\,2} - \frac{1}{N^2}\,N^2\,\bar{y}^2\right]$$

$$= \frac{1}{n}\left[\frac{1}{N}\sum_{i=1}^{N} y_i^{\,2} - \bar{y}^2\right] \qquad \left[\because \sigma^2 = \frac{1}{N}\sum_{i=1}^{N} y_i^{\,2} - \bar{y}^2 = E(y^2) - [E(y)]^2\right]$$

$$= \frac{\sigma^2}{n}$$

$\therefore$ $\text{Var}(\bar{y}) = \dfrac{\sigma^2}{n}$

Theorem 7

Sample mean square is an unbiased estimator of population mean square $E(s^2) = \sigma^2$.

Proof :

$$s^2 = \frac{1}{n-1}\left[\sum_{i=1}^{n} y_i^2 - n\bar{y}^2\right]$$

$$= \frac{1}{n-1}\left[\sum_{i=1}^{N} b_i y_i^2 - n\bar{y}^2\right]$$

Taking expectation on both sides

$$E(s^2) = \frac{1}{n-1}\left[\sum_{i=1}^{N} y_i^2\, E(b_i) - nE(\bar{y}^2)\right]$$

$$= \frac{1}{n-1}\left[\sum_{i=1}^{N} y_i^2\left(\frac{n}{N}\right) - n\,[Var(\bar{y}) + [E(\bar{y})]^2]\right]$$

$$= \frac{1}{n-1}\left[\frac{n}{N}\sum_{i-1}^{N} y_i^2 - n\,\frac{\sigma^2}{n} - n\bar{y}^2\right]$$

$$= \frac{1}{n-1}\left[n\left[\frac{1}{N}\sum_{i=1}^{N} y_i^2 - \bar{y}^2\right] - \sigma^2\right]$$

$$= \frac{1}{n-1}\,[n\,\sigma^2 - \sigma^2]$$

$$= \frac{1}{n-1}\,[\sigma^2\,(n-1)]$$

$$= \sigma^2$$

$\therefore \quad E(s^2) = \sigma^2$

Q24. What are the advantages and limitations of simple random sampling.

Ans :

Merits

1. Since it is a probability sampling, it eliminates the bias due to the personal judgment or discretion of the investigator. Accordingly, the sample selected is more representative of the population than in the case of judgment sampling.

2. Because of its random character, it is possible to ascertain the efficiency of the estimates by considering the standard errors of their sampling distributions. As pointed out earlier [See Remark 2 above], $\bar{x}$ as an estimate of $\bar{y}$ = p can be made more efficient by taking large samples. Moreover, large

sample will be more representative of the population according to the Principle of Statistical Regularity and the Principle of Inertia of Large Numbers and thus provide better results.

3. The theory of random sampling is highly developed so that it enables us to obtain the most reliable and maximum information at the least cost, and results in savings in time, money and labour.

Demerits

1. Simple random sampling requires an up-to-date frame, i.e., a complete and up-to-date list of the population units to be sampled. In practice, since this is not readily available in many inquiries, it restricts the use of this sampling design.

2. In field surveys if the area of coverage is fairly large, then the units selected in the random sample are expected to be scattered widely geographically and thus it may be quite time consuming and costly to collect the requisite information or data.

3. If the sample is not sufficiently large, then it may not be representive of the population and thus may not reflect the true characteristics of the population

4. The numbering of the population units and the preparation of the slips is quite time consuming and uneconomical particularly if the population is large. Accordingly, this method can't be used effectively to collect most of the data in social sciences.

5. For given degree of accuracy, simple random sampling usually requires larger sample as compared to stratified random sampling discussed below.

6. Sometimes, simple random sample gives results which are highly improbabilistic in nature, i.e., whose probability is very small. For example, a random selection of 13 cards from a pack of 52 cards might give all thirteen cards of spades, say. The probability of the happening of such an event in practice is very very small.

PROBLEMS

4. Consider a population of 5 units with values 3, 5, 7, 9, 11 write the sample values.

Sol :

(i) $E(\bar{y}) = \bar{Y}$

Here $N = 5, \quad n = 2, \quad N_{c_n} = 5_{c_2} = 10$

$$\bar{Y} = \frac{1}{N} \sum_{i=1}^{N} Y_i$$

$$= \frac{1}{5} (3 + 5 + 7 + 9 + 11)$$

$$= \frac{1}{5} (35)$$

$$\bar{Y} = 7$$

K = No. of samples

Sample No's	Sample Values	Sample mean ($\bar{y}$)
1	(3, 5)	$\frac{3+5}{2} = 4$
2	(3, 7)	5
3	(3, 9)	6
4	(3, 11)	7
5	(5, 7)	6
6	(5, 9)	7
7	(5, 11)	8
8	(7, 9)	8
9	(7, 11)	9
10	(9, 11)	10
		$\sum_{i=1}^{n} \bar{y} = 70$

$$E(\bar{y}) = \frac{1}{K}\sum_{i=1}^{N}(\bar{y})$$

$$= \frac{1}{10}(70)$$

$$E(\bar{y}) = 7$$

$\therefore$ $E(\bar{y}) = \bar{y}$

(ii) SRSWR

$\bar{y} = 7$

No. of samples $N^n = 5^2 = 25$

Sample No's	Sample Values	Sample means
1	(3, 3)	3
2	(3, 5)	4
3	(3, 7)	5
4	(3, 9)	6
5	(3, 11)	7
6	(5, 3)	4
7	(5, 5)	5
8	(5, 7)	6
9	(5, 9)	7
10	(5, 11)	8
11	(7, 3)	5
12	(7, 5)	6

13	(7, 7)	7
14	(7, 9)	8
15	(7, 11)	9
16	(9, 3)	6
17	(9, 5)	7
18	(9, 7)	8
19	(9, 9)	9
20	(9, 11)	10
21	(11, 3)	7
22	(11, 5)	8
23	(11, 7)	9
24	(11, 9)	10
25	(11, 11)	11
		$\sum_{i=1}^{n} \bar{y} = 175$

$$E(\bar{y}) = \frac{1}{K} \cdot \sum_{i=1}^{N} \bar{y}$$

$$= \frac{1}{25} (175)$$

$$E(\bar{y}) = 7$$

$\therefore$ The sample mean is an unbiased estimator of population mean.

5. Consider a population of 4 observations where values are 1, 2, 3, 4 write all possible sample size 2 without replacement verify the sample mean is an unbiased estimator of population mean and also find sample variance.

Sol :

SRSWOR

$N_{c_n} = 4_{c_2} = 6, \quad N = 4, \quad n = 2$

No. of Samples	Sample Values	Sample mean ($\bar{y}$)
1	(1, 2)	1.5
2	(1,3)	2
3	(1, 4)	2.5
4	(2, 3)	2.5
5	(2, 4)	3
6	(3, 4)	3.5
		$\sum_{i=1}^{n} (\bar{y}) = 15$

$$E(\bar{y}) = \bar{y}$$

$$E(\bar{y}) = \frac{1}{K} \sum_{i=1}^{N} \bar{y}$$

$$= \frac{1}{6} (15) = 2.5$$

$$s^2 = \frac{1}{n-1} \left[\sum_{i=1}^{n} y_i^2 - n\bar{y}^2 \right]$$

$$= \frac{1}{2-1} [(30) - 2(2.5)^2] = [30 - 12.5]$$

$s^2 = 17.5$ is the sample variance

$$\bar{y} = \frac{1}{N} \sum_{i=1}^{N} (y_i)$$

$$\bar{y} = \frac{1}{4} (10) = 2.5$$

$\therefore$ The sample mean is an unbiased estimator of population mean.

Short Question & Answers

1. Sampling

Ans :

Sampling may be defined as the selection of some part of an aggregate or totality on the basis of which a judgement or inference about the aggregate or totality is made. In other words, it is the process of obtaining information about an entire population by examining only a part of it.

In most of the research work and surveys, the usual approach happens to be to 'make generalizations' or to 'draw inferences' based on samples about the parameters of population from which the samples are taken. The researcher quite often selects only a few items from the universe for his study purpose. All this is done on the assumption that the sample data will enable him to estimate the population parameters.

2. Sample survey

Ans :

It is a term used in statistics and market research, indicating a survey conducted on a set of sample elements taken from a target population. A survey is a technique of observation to gather, measure and evaluate responses from people. Marketers often use sample surveys to conduct market research amongst target consumers for their brands, since it is not possible to go about asking the entire target population which may be huge in numbers. Hence, using sampling techniques, a representative set of the population viz. a sample is selected and used for conducting the required research, thus saving the cost of surveying the entire population. A survey conducted on the entire population is called a census.

Following points need to address for a sample survey:

(a) Sample unit: Whom to survey

(b) Sample size: How many people should be surveyed

(c) Sampling procedure: How should respondents for the survey be chosen

3. Define Population

Ans :

The collection of all units of a specified type in a given region at a particular point or period of time is termed as a population or universe. Thus, we may consider a population of persons, families, farms, cattle in a region or a population of trees or birds in a forest or a population of fish in a tank etc. depending on the nature of data required.

There are two different types of populations. They are,

(i) Finite population

(ii) Infinite population.

(i) Finite Population

A population whose values can be theoretically observed as it has fixed number is called a finite population. The size of the finite population is limited.

Example: Number of students in a school, number of rice plants in a field etc.

(ii) Infinite Population

A population whose values cannot be theoretically observed as it does not have a fixed number is called an infinite population. The size of the infinite population is unlimited.

Example: Number of White Blood Cells (WBC) in the human body.

The size of the population is always very large, even if it is finite or infinite. So, the collection of data from each of the individuals is very difficult. Because of this, samples are selected from the population and based on their results conclusions are made regarding that population.

4. Sample Frame

Ans :

Sampling frame is a collection of all the members/units of population. It should not contain any error and duplication of units. Each and every individual sampling unit is selected from a sampling frame.

It possess the following chafacteristics,

(i) **Complete :** A frame should consist of all legal units of the population.

(ii) **Accuracy :** A frame should not consist of any non-existing units of population.

(iii) **Adequate :** The structure of a frame must be sufficient enough to cover the total population.

(iv) **Updated :** The unit and its content present in a frame must be up to date.

5. Standard Error

Ans :

The standard deviation of the sampling distribution of a statistic is known as its Standard Error (S.E.). Thus, the Standard Error of the statistic t is given by :

$$\text{S.E.(t)} = \sqrt{\text{Var}(t)} = \sqrt{\left[\frac{1}{k}\sum_{i=1}^{k}(t_i - t)^2\right]}$$

In particular the S.E. of the sampling distribution of the mean* is given by the standard deviation of the values $\bar{x}_1, \bar{x}_2, ..., \bar{x}_k$.

6. Principals in sample surveys

Ans :

Three basic principles for the design of a sample survey are:

1. Principle of Optimization

The principle of optimization takes into account the factors of

(a) Efficiency and (b) cost.

(a) **Efficiency :** Efficiency is measured by the inverse of sampling variance of the estimator. The principle of optimization ensures that a given level of efficiency will be reached with the minimum possible resources and minimum cost.

(b) **Cost :** Cost is measured by expenditure incurred in terms of money or man powers. So, the term optimization means that, it is based on developing methods of sample selection and of estimation; these provide a given value of cost with the maximum possible efficiency.

2. Principle of Validity

By validity of a sample design, we mean that the sample should be so selected that the results could be interpreted objectively in terms of probability. According to this, sampling provides valid estimates about population parameters. This principle ensures that there is some definite and preassigned probability for each individual of the aggregate (population) to be included in the sample.

3. Principle of Statistical Regularity

According to the principle of statistical regularity we mean that a moderately large number of items chosen at random from a large group are almost sure on the average to possess the characteristics of the large group. This principle has also its origin in the law of large numbers of the theory of probability.

7. Need for sampling

Ans :

1. Sampling can save time and money. A sample study is usually less expensive than a census study and produces results at a relatively faster speed.
2. Sampling may enable more accurate measurements for a sample study is generally conducted by trained and experienced investigators.
3. Sampling remains the only way when population contains infinitely many members.
4. Sampling remains the only choice when a test involves the destruction of the item under study.
5. Sampling usually enables to estimate the sampling errors and, thus, assists in obtaining information concerning some characteristic of the population.

8. Census Method

Ans :

In the census method we resort to 100% inspection of the population and enumerate each and every unit of the population. In the sample method we inspect only a selected representative and adequate fraction (finite subset) of the population and after analysing the results of the sample data we draw coclusions about the characteristics of the population.

The census method seems to provide more accurate and exact information as compared to sample enumeration as the information is collected from each and every unit of the population. Moreover, it affords more extensive and detailed study. For instance, the population census conducted by the Government of India every 10 years collects information not only about the population but also obtains data relating to age, marital status, occupation, religion, education, employment, income, property, etc. The census method has its obvious limitations and drawbacks given below :

(i) The complete enumeration of the population requires lot of time, money, manpower and administrative personnel. As such this method can be adopted only by the government and big organisations who have vast resources at their disposal.

(ii) Since the entire population is to be enumerated, the census method is usually very time consuming. If the population is sufficiently large, then it is possible that the processing and the analysis of the data might take so much time that when the results are available they are not of much use because of changed conditions.

9. Advantages of sampling.

Ans :

1. Low cost of sampling

If data were to be collected for the entire population, the cost will be quite high. A sample is a small proportion of a population. So, the cost will be lower if data is collected for a sample of population which is a big advantage.

2. Less time consuming in sampling

Use of sampling takes less time also. It consumes less time than census technique. Tabulation, analysis etc., take much less time in the case of a sample than in the case of a population.

3. Scope of sampling is high

The investigator is concerned with the generalization of data. To study a whole population in order to arrive at generalizations would be impractical.

Some populations are so large that their characteristics could not be measured. Before the measurement has been completed, the population would have changed. But the process of sampling makes it possible to arrive at generalizations by studying the variables within a relatively small proportion of the population.

10. Limitations of sampling.

Ans :

1. Chances of bias

The serious limitation of the sampling method is that it involves biased selection and thereby leads us to draw erroneous conclusions. Bias arises when the method of selection of sample employed is faulty. Relative small samples properly selected may be much more reliable than large samples poorly selected.

2. Difficulties in selecting a truly representative sample

Difficulties in selecting a truly representative sample produces reliable and accurate results only when they are representative of the whole group. Selection of a truly representative sample is difficult when the phenomena under study are of a complex nature. Selecting good samples is difficult.

3. In adequate knowledge in the subject

Use of sampling method requires adequate subject specific knowledge in **sampling technique**. Sampling involves statistical analysis and calculation of probable error. When the researcher lacks specialized knowledge in sampling, he may commit serious mistakes. Consequently, the results of the study will be misleading.

4. Changeability of units

When the units of the population are not in homogeneous, the sampling technique will be unscientific. In sampling, though the number of cases is small, it is not always easy to stick to the, selected cases. The units of sample may be widely dispersed.

Some of the cases of sample may not cooperate with the researcher and some others may be inaccessible. Because of these problems, all the cases may not be taken up. The selected cases may have to be replaced by other cases. Changeability of units stands in the way of results of the study.

5. Impossibility of sampling

Deriving a representative sample is di6icult, when the universe is too small or too heterogeneous. In this case, census study is the only alternative. Moreover, in studies requiring a very high standard of accuracy, the sampling method may be unsuitable. There will be chances of errors even if samples are drawn most carefully.

---

11. Define Probability Sampling.

Ans :

The choice of an appropriate sampling design is of paramount importance in the execution of a sample survey and is generally made keeping in view the objectives and scope of the enquiry and the type of the universe to be sampled. The sampling techniques may be broadly classified as follows :

(i) Purposive or Subjective or Judgment Sampling.

(ii) Probability Sampling.

(iii) Mixed Sampling.

Probability sampling provides a scientific technique of drawing samples from the population according ursome laws of chance in which each unit in the universe has some definite pre-assigned probability of being selected in the sample. Different types of sampling are in which :

(i) Each sample unit has an equal chance of being selected.

(ii) Sampling units have varying probability of being selected.

(iii) Probability of selection of a unit is proportional to the sample size.

12. Sampling Unit.

Ans :

Elementary units or group of such units which besides being clearly defined, identifiable and observable, are convenient for purpose of sampling are called sampling units. For instance, in a family budget enquiry, usually a family is considered as the sampling unit since it is found to be convenient for sampling and for ascertaining the required information. In a crop survey, a farm or a group of farms owned or operated by a household may be considered as the sampling unit.

Choose the Correct Answers

1. A small part taken from the population is called _____ [b]

   (a) Population (b) Sample
   (c) Bias (d) None

2. Sample mean is an unbiased estimate of population [a]

   (a) Mean (b) Variance
   (c) S.E (d) Bias

3. _____ errors are subject to control though verification. [c]

   (a) Prestige biase (b) Non-response
   (c) Compiling (d) Publication

4. _____ method is very popular method of taking random samples [b]

   (a) Random (b) Lottery
   (c) Census (d) Sample

5. The method of selecting the sample observation by the help of _____ number table [a]

   (a) Random (b) Lottery
   (c) Census (d) Sample

6. $E(\bar{y})$ = _____ [c]

   (a) σ^2 (b) s^2
   (c) $\bar{y}$ (d) None

7. In simple random sample with replacement _____ = $\frac{\sigma^2}{n}$. [a]

   (a) $var(\bar{y})$ (b) $E(\bar{y})$
   (c) a & b (d) All

8. An _____ is a qualitative characteristic which cannot be measured quantitatively [c]

   (a) Sample (b) Population
   (c) Attribute (d) Median

9. Sample mean square is an _____ estimator of population mean square [d]

   (a) Consistency (b) Sufficiency
   (c) Efficiency (d) Unbiassed

10. A group of _____ is known as population [d]

    (a) Individuals (b) Group
    (c) People (d) a & c

Fill in the Blanks

1. A _____ is an entire collection of people, animals, plants from which we may collect data.
2. A population is _____ if it contains infinite number of individuals.
3. Any _____ measure computed from population data is known as parameter.
4. A sampling frame is the source material (or) device from which a _____ is drawn.
5. The standard error sample correlation coefficient is _____.
6. The data should be collected keeping in view the detectives of the _____.
7. In _____ method the investigator goes from house to house and interviews the individuals personally.
8. _____ errors have their origin in sampling.
9. _____ biases occur if full information is not obtained on all the sampling units.
10. _____ sampling is the scientific method of selecting sample according to some laws of chance.

ANSWERS

1. Population
2. Infinite
3. Statistical
4. Sample
5. $(1 - \rho^2)/\sqrt{n}$
6. Survey
7. Interview
8. Sampling
9. Non-response
10. Probability

One Mark Answers

1. Define population.

Ans :

A population is an entire collection of people, animals, plants (or) things from which we may collect data.

2. Define parameter.

Ans :

Any statistical measure computed from population data is known as parameter.

3. Define sample frame.

Ans :

A sampling frame is the source material (or) device from which a sample is drawn.

4. Define sample.

Ans :

A small part taken from the population is known as sample.

5. Define standard error.

Ans :

The standard deviation of sampling distribution of a statistic is known as standard error.

UNIT II

Estimates of population mean, total, and proportion, their variances and the estimates of variances in the following methods.
(i) Stratified Random Sampling with Proportional and Neyman allocation, and
(ii) Systematic Sampling when N = nk.
Comparison of their relative efficiencies. Advantages and disadvantages of SRS, Stratified and Systematic sampling methods.

Time series: Time series and its components with illustrations, additive, multiplicative and mixed models. Determination of trend by least squares and moving average methods. Growth curves and their fitting with reference to Modified exponential, Gompertz and Logistic curves. Determination of seasonal indices by Ratio to moving average, ratio to trend and link relative methods.

2.1 Estimates of Population Mean, Total, and Proportion, Their Variances and the Estimates of Variances in the Following Methods

2.1.1 Stratified Random Sampling

Q1. What do you understand by stratified random sampling? Discuss its proportion means and its variance of Stratified Random Sampling.

Ans : **(July-21, July-19)**

Stratified Random Sampling

This process divides the population into homogenous groups or classes called 'strata'. A sample is taken from each group by simple random method and the resulting sample is called a stratified sample.

A stratified sample may be either proportionate or disproportionate. In a proportionate stratified sampling plan, the number of items drawn from each stratum is proportional to the size of the strata. While in a disproportionate stratified sampling, equal number of items are taken from each stratum irrespective of the size of the stratum.

Example

People interested in watching movies can be divided into groups or strata of people watching Hindi movies and English movies. After division, these groups can be selected randomly which is called stratified random sampling.

This method yields the most accurate results compared to other methods when population under study is heterogeneous. In this method population is divided into strata or subsets and random samples are chosen from each strata. This division of population into strata is based on the relevant similar characteristics within individual units. But the individual units present within each strata are similar.

Precision and reliability depends primarily on the variability within the strata and this method of sampling is highly beneficial when individual unit variability within each stratum is small compared to variability between individual units in different strata.

Stratified sampling is usually performed to reduce sampling errors especially in case of heterogeneous population.

Estimate of Population Mean and its Variance

Assume,

N_i as the number of sampling units present in the i^{th} stratum.

N as the number of sampling units present in the overall population which can be given as $\sum_{i=1}^{k} N_i$ where k is the number of strata i.e., $N = \sum_{i=1}^{k} N_i$.

n_i as the selected sampling units with simple random sampling without replacement (srswor) from i^{th} stratum n as the total sample size which is given as $\sum_{i=1}^{k} n_i$.

If Y_{ij} represents the j^{th} unit of i^{th} stratum, then the population mean of i^{th} stratum can be given as

$$\bar{Y}_{N_i} = \frac{1}{N_i} \sum_{i=1}^{k} Y_{ij}$$

Similarly, the population mean for overall population can be given as,

$$\bar{Y}_N = \frac{1}{N}\sum_i\sum_j Y_{ij} = \frac{1}{N}\sum_i^k N_i \bar{Y}_{N_i}$$

$$= \sum_{i=1}^k P_i \bar{Y}_{N_i}$$

Where P_i is the weight of i^{th} stratum which is given as $P_i = \frac{N_i}{N}$

For the i^{th} stratum, the population mean square S_1^2 can be given as,

$$S_i^2 = \frac{1}{N_i - 1}\sum_{j=1}^{N_i}(Y_{ij} - \bar{Y}_{N_i})^2$$

If y_{ij} represents j^{th} sample unit $\bar{y}_{n_i}$ represents the mean of sample from ith stratum, then the sample mean square of this stratum can be given as,

$$S_i^2 = \frac{1}{n_i - 1}\sum_{j=1}^{n_i}(y_{ij} - \bar{y}_{n_i})^2$$

Consider the two estimates of population mean $\bar{Y}_N$ as,

1. $\bar{y}_n = \frac{1}{n}\sum_{i=1}^k n_i \bar{y}_{n_i}$

2. $\bar{y}_{st} = \frac{1}{N}\sum_{i=1}^k N_i \bar{y}_{n_i}$

$$= \sum_{i=1}^k P_i \bar{y}_{n_i}$$

Where, P_i is the mean of weights of strata sample. These weights are nothing but the sizes of strata.

Theorem

$\bar{y}_{st}$ is considered as unbiased estimate of $\bar{Y}_N$ which means $E(\bar{y}_{st}) = \bar{Y}_N$.

Proof

Every stratum consists of simple random samples and therefore, the estimate,

$$E(\bar{y}_{ni}) = \bar{Y}_{N_i}$$

$$E(\bar{y}_{st}) = \frac{1}{N}\sum_{i=1}^k N_i\, E(\bar{y}_{n_i}) \quad [\because \text{From Estimate 2}]$$

$$= \frac{1}{N}\sum_{i=1}^k N_i\, \bar{Y}_{N_i} \quad [\text{From equation (1)}]$$

$$\Rightarrow E(\bar{y}_{st}) = \bar{Y}_N \left[\because \frac{1}{N}\sum_{i=1}^k N_i \bar{Y}_{N_i} = \bar{Y}_N\right]$$

The variance of mean $\bar{y}_{st}$ can be given as,

$$Var(\bar{y}_{st}) = \frac{1}{N^2}\sum_{i=1}^k N_i(N_i - n_i)\frac{S_i^2}{n_i} = \sum_{i=1}^k\left(\frac{1}{n_i} - \frac{1}{N_i}\right)p_i^2 S_i^2$$

Q2. Define:

(a) Population total

(b) Population proportion

(c) Estimates of variance

Ans :

(a) Population Total

Population total can be defined as the product of population mean and the number of sampling units in the population. It can be represented in stratified random sampling as,

$$N\bar{y}_{st} = N\,\bar{Y}_N$$

There are three estimations possible for population total while using stratified sampling. They are as follows,

1. The population total using an unbiased estimation procedure can be estimated as,

$$N\bar{y}_{st} = \sum_{i=1}^k N_i\,\bar{y}_i$$

Where N_i refers to the sampling unit within strata of population containing n_i samples.

2. The variance estimator of population using an unbiased procedure can be obtained as,

$$N^2 Var(\bar{y}_{st}) = \sum_{i=1}^{k} N_i^2 \ Var\ (\bar{y}_{n_i})$$

3. If the size of sample is large, the estimation of confidence interval $100(1 - \alpha)\%$ for population total using stratified random sampling can be obtained as,

$$N\bar{y}_{st} + z_{\alpha/2}\ NVar(\bar{y}_{st}).$$

(b) Population Proportion

Population proportion can be defined as sum of all the samples in the population divided by the total number of sampling units in the population. It can be represented as,

$$P = \frac{N_1P_1 + N_2P_2 + N_3P_3 + ... + N_kP_k}{N}$$

$$= \frac{1}{N}\sum_{i=1}^{k} N_iP_i$$

The estimations possible for population while using stratified random sampling are as follows,

1. If P represents the overall population proportion then its unbiased estimation obtains,

$$\hat{P}_{st} = \frac{1}{N}\sum_{i=1}^{k} N_iP_i$$

Where P_i represents the population proportion in i^{th} stratum.

2. The estimator of population proportion while using an unbiased estimation procedure for variance is obtained as,

$$Vâr\ (\hat{P}_{st}) = \frac{1}{N^2}\sum_{i=1}^{k} N_i^2\ Vâr(\hat{P}_i)^2$$

3. If the size of sample is large, the estimation of confidence interval $100(1 - \alpha)\%$ for population proportion using stratified random sampling can be obtained as,

$$\hat{P}_{st} + Z_{\alpha/2} Vâr\ (\hat{P}_{st})$$

(c) Estimates of Variance

The estimate of variance in stratified random sampling in two different cases is obtained as follows,

Case 1

If samples are drawn using Simple Random Sampling Without Replacement (SRSWOR), then,

$$E(s_t)^2 = S_i^2$$

Here,

$$S_i^2 = \frac{1}{n_i - 1}\sum_{j=1}^{n_i}(y_{ij} - \bar{y}_i)^2$$

The variance in such case in given as,

$$Vâr\ (\bar{y}_i) = \frac{N_i - n_i}{N_i n_i} S_i^2 \quad ...(1)$$

The variance of stratified random sampling can be estimated as,

$$Vâr(\bar{y}_{st}) = \sum_{i=1}^{k} w_i^2\ Vâr\ (\bar{y}_i) \quad ...(2)$$

Where, w_i is the weight of i^{th} stratum.

Substituting equation (1) in equation (2).

$$Vâr(\bar{y}_{st}) = \sum_{i=1}^{k} w_i^2 \left(\frac{N_i - n_i}{N_i n_i}\right) S_i^2$$

Case 2

If samples are drawn using Simple Random Sampling With Replacement (SRSWR), then

$E(\bar{y}_{st}) = \bar{y}$,

$$Vâr(\bar{y}_{st}) = \sum_{i=1}^{k} \frac{w_i^2 S_i^2}{n_i}.$$

2.1.2 Proportional and Neyman Allocation

Q3. Explain briefly about proportional allocation.

Ans : **(July-19)**

Proportional Allocation

In this, the items are selected from each stratum in the same proportion as they exist in the population. The allocation of sample sizes is termed

as proportional if the sample fraction, i.e, the ratio of the sample size to the population size, remains the same in all the strata, Mathematically, the principle of proportional allocation gives :

$$\frac{n_1}{N_1} = \frac{n_2}{N_2} = ... = \frac{n_k}{N_k} \quad ... (1)$$

By the property of ratio and proportions, each of these ratios is equal to the ratio of the sum of numerators to the sum of denominators, i.e.,

$$\frac{n_1}{N_1} = \frac{n_2}{N_2} = ... = \frac{n_k}{N_k} = \frac{n_1 + n_2 + ... + n_k}{N_1 + N_2 + ... + N_k} = \frac{n}{N} = c, \text{ (Constant)} \quad ... (2)$$

since the total sample size n, and the population size N are fixed. Hence,

$$n_1 = N_1\left(\frac{n}{N}\right), \qquad n_2 = N_2\left(\frac{n}{N}\right), ... , n_k = N_k\left(\frac{n}{N}\right)$$

$$\Rightarrow \quad n_i = N_i\left(\frac{n}{N}\right), (i = 1, 2, ..., k)$$

$$\Rightarrow \quad n_i \propto N_i,$$

where c = n/N, is taken as the constant of proportionality.

Q4. Define cost function.

Ans : **(Dec.-18)**

In stratified random sampling the cost function means a function which indicates the cost of the survey and is given by

$$C = a + \sum_{i=1}^{k} C_i n_i$$

Where,

a = Overhead cost and

C_i is the cost per unit in the i^{th} stratum.

C is the total survey cost.

Q5. Explain briefly about Neyman allocation (or) optimum allocation.

Ans : **(July-19)**

The size of the samples to be drawn from the various strata is determined by the principle of optimization, i.e., obtaining best results at minimum possible cost. In optimum allocation, n_i s, (i = 1, 2, ..., k) are determined so that:

(i) Variance of sample estimate of the population mean is minimum (i.e., its precision is maximum) for fixed total sample size n. (Neynian's Allocation).

(ii) Variance of the estimate is minimum for a fixed cost of the plan.

(iii) Total cost of the sampling design is minimum for fixed desired precision, i.e., total cost is minimum for a fixed value of the variance of the sample estimate.

Let us denote by $\bar{x}_{st}$ the estimate of the population mean obtained by the technique of stratified random sampling. Wc state below the results concerning the above three modes of optimum allocation.

Result 1

Neyman Allocation. The variance of the estimate of the population mean i.e., Var (x_{st}) is minimum for a fixed total size of the sample if,

$$n_i = n N_i S_i / \left(\sum_{i=1}^{k} N_i S_i \right) \Rightarrow n_i \propto N_i S_i$$

This is known as Neyman's formula for optimum allocation. This suggest that greater the value of the product N_iS_i, i.e., product of size and the variability of the stratum, the larger is the number of units to be sampled from it. Thus, if there are two strata of the same size then a larger sample is to be drawn from the stratum having greater variability, i.e., standard deviation.

Result 2

In stratified random sampling the variance of the sampling estimate of the population mean is minimum for a fixed given cost of the sampling design if

$$n_i \propto \frac{N_i S_i}{\sqrt{C_i}}, \ (i = 1, 2, ..., k)$$

where C_i, (i = 1, 2, ..., k) is the cost per unit for the ith stratum. This principle of allocation suggests that to obtain the most precise estimates for a fixed cost a large sample would be needed from a stratum if

(i) Stratum size (N_i) is large.

(ii) Stratum variability (σ_i) is large.

(iii) Sampling cost per unit is low in the stratum.

Result 3

In this case we want to minimize the total cost of the design to achieve desired precision of the estimate, i.e., for fixed value of Var ($\bar{x}_{st}$) = V_0, (say). Using Neyman's principle of optimum allocation, i.e., taking

$$n_i \propto N_i S_i \Rightarrow n_i = n N_i S_i / \left(\sum_{i=1}^{k} N_i S_i \right)$$

the cost of design is minimized for fixed value of V_0 if n is chosen such that

$$n = \frac{\left(\sum_{i=1}^{k} N_i S_i \right)^2}{N^2 V_0 + \sum_{i=1}^{k} N_i S_i^2}$$

Q6. In stratified random sampling with given cost function of the form $C = a + \sum_{i=1}^{k} C_i n_i$, $V(\bar{y}_{st})$ is minimum of $n_i \propto \frac{N_i S_i}{\sqrt{C_i}}$.

Ans : **(Nov.-20, Dec.-18)**

Given

$$C = a + \sum_{i=1}^{k} C_i n_i \qquad ... (1)$$

Let

$$\phi = \text{var}(\bar{y}_{st}) + \lambda \left(a + \sum_{i=1}^{k} C_i n_i - C \right)$$

$$= \frac{1}{N^2} \left[\sum_{i=1}^{k} \left(\frac{N_i}{n_i} - 1 \right) N_i S_i^2 \right] + \lambda \left[a + \sum_{i=1}^{k} C_i n_i - C \right]$$

Differentiate with respect to n_i and equating to '0'.

$$-\frac{1}{N^2} \frac{N_i^2}{n_i^2} S_i^2 + \lambda C_i = 0$$

$$-\frac{N_i^2 S_i^2}{N^2 n_i^2} = -\lambda C_i$$

$$n_i^2 = \frac{N_i^2 . S_i^2}{N^2 C_i \lambda}$$

Taking square root on both sides

$$n_i = \frac{N_i S_i}{N \sqrt{C_i \lambda}} \qquad ... (2)$$

$\sum_{i=1}^{k} n_i = n$ [from optimum method] ... (3)

Substitute (2) in (3)

$$\sum_{i=1}^{k} n_i = n$$

$$\sum_{i=1}^{k} \frac{N_i S_i}{N\sqrt{C_i \lambda}} = n$$

$$\sqrt{\lambda} = \frac{\sum_{i=1}^{k} N_i S_i}{nN\sqrt{C_i}}$$

Substitute $\sqrt{\lambda}$ value in (2)

$$n_i = \frac{N_i S_i}{N\sqrt{C_i}\left[\frac{\sum_{i=1}^{k} N_i S_i}{nN\sqrt{C_i}}\right]}$$

$$= \frac{nN_i S_i / \sqrt{C_i}}{\sum_{i=1}^{k} N_i S_i / \sqrt{C_i}}$$

$$n_i = \frac{n}{\sum_{i=1}^{k} N_i S_i / \sqrt{C_i}} \times N_i S_i / \sqrt{C_i}$$

$$n_i = C \times N_i S_i \sqrt{C_i}$$

Where

$$C = \frac{n}{\sum_{i=1}^{k} N_i S_i / \sqrt{C_i}}$$

$$n_i \propto N_i S_i / \sqrt{C_i}$$

Q7. Comparison between simple Random sampling and stratified Random Sampling.

(or)

$var(\bar{y}_{st})_{op} \le (\bar{y}_{st})_{pr} \le var(\bar{y})_R$

Ans :

$$var(\bar{y}_{st})_{pr} = \frac{N-n}{Nn} \sum_{i=1}^{k} P_i S_i^2$$

$$var(\bar{y}_{st})_{op} = \frac{1}{n}\left(\sum_{i=1}^{k} P_i S_i\right)^2 - \frac{1}{N}\sum_{i=1}^{k} P_i S_i^2$$

We assume that

$$var(\bar{y}_{st})_{pr} - var(\bar{y}_{st})_{op} \ge 0 \quad \text{... (1)}$$

$$\frac{N-n}{nN}\sum_{i=1}^{k} P_i S_i^2 - \left[\frac{1}{n}\left(\sum_{i=1}^{k} P_i S_i\right)^2 - \frac{1}{N}\sum_{i=1}^{k} P_i S_i^2\right] \ge 0$$

$$\left[\frac{1}{n} - \frac{1}{N}\right]\sum_{i=1}^{k} P_i S_i^2 - \frac{1}{n}\left(\sum_{i=1}^{k} P_i S_i\right)^2 + \frac{1}{N}\sum_{i=1}^{k} P_i S_i^2 \ge 0$$

$$\frac{1}{n}\sum_{i=1}^{k} P_i S_i^2 - \frac{1}{N}\sum_{i=1}^{k} P_i S_i^2 - \frac{1}{n}\left(\sum_{i=1}^{k} P_i S_i\right)^2 + \frac{1}{N}\sum_{i=1}^{k} P_i S_i^2 \ge 0$$

$$\frac{1}{n}\sum_{i=1}^{k} P_i S_i^2 - \frac{1}{n}\left(\sum_{i=1}^{k} P_i S_i\right)^2 \ge 0$$

$$\frac{1}{n}\left[\sum_{i=1}^{k} P_i S_i^2 - \left(\sum_{i=1}^{k} P_i S_i\right)^2\right] \ge 0$$

$$\frac{1}{n}\left[\sum_{i=1}^{k} \frac{N_i}{N} S_i^2 - \left(\sum_{i=1}^{k} \frac{N_i}{N} S_i\right)^2\right] \ge 0 \quad \left[\because P_i = \frac{N_i}{N}\right]$$

$$\frac{1}{n}\left[\frac{1}{N}\sum_{i=1}^{k} N_i S_i^2 - \left(\frac{1}{N}\sum_{i=1}^{k} N_i S_i\right)^2\right] \ge 0 \quad \text{... (2)}$$

From (1) and (2)

$$var(\bar{y}_{st})_{pr} \ge (\bar{y}_{st})_{op} \quad \text{... (I)}$$

Now, population variance

$$S^2 = \frac{1}{N}\sum_{i=1}^{k}\sum_{j=1}^{N_i} (y_{ij} - \bar{y})^2$$

Adding and substracting $\bar{y}_i$

$$= \frac{1}{N-1}\sum_{i=1}^{k}\sum_{j=1}^{N_i} (\bar{y}_{ij} - \bar{y}_i + \bar{y}_i - \bar{y})^2$$

$$= \frac{1}{N-1}\sum_{i=1}^{k}\sum_{j=1}^{N_i}[(y_{ij}-\bar{y}_i)+(\bar{y}_i-\bar{y})]^2$$

$$= \frac{1}{N-1}\sum_{i=1}^{k}\sum_{j=1}^{N_i}[(y_{ij}-\bar{y}_i)^2+(\bar{y}_i-\bar{y})^2+2(y_{ij}-\bar{y}_i)(\bar{y}_i-\bar{y})]$$

$$(N-1)S^2 = \sum_{i=1}^{k}\sum_{j=1}^{N_i}(y_{ij}-\bar{y}_i)^2+\sum_{i=1}^{k}\sum_{j=1}^{N_i}(\bar{y}_i-\bar{y})^2+2\sum_{i=1}^{k}\sum_{j=1}^{N_i}(y_{ij}-\bar{y}_i)(\bar{y}_i-\bar{y})$$

$$= \sum_{i=1}^{k}(N_i-1)S_i^2+\sum_{i=1}^{k}\sum_{j=1}^{N_i}(\bar{y}_i-\bar{y})^2+2\sum_{i=1}^{k}\sum_{j=1}^{N_i}(y_{ij}-\bar{y}_i)(\bar{y}_i-\bar{y}) \quad \left[\because \frac{1}{N_i-1}\sum_{i=1}^{k}(y_{ij}-\bar{y}_i)^2 = S_i^2\right]$$

$$= \sum_{i=1}^{k}(N_i-1)S_i^2+\sum_{i=1}^{k}N_i(\bar{y}_i-\bar{y})^2+2\sum_{i=1}^{k}\sum_{j=1}^{N_i}(y_{ij}-\bar{y}_i)(\bar{y}_i-\bar{y})$$

Sum of the algebraic deviations of the mean is zero. N_i consecutive value of 'N' is large so we take $N_i - 1 = N_i$, $N - 1 = N$.

$NS^2 = N_i$, $N - 1 = N$

$$NS^2 = \sum_{i=1}^{k}N_iS_i^2 + \sum_{i=1}^{k}N_i(\bar{y}_i-\bar{y})^2$$

$$S^2 = \sum_{i=1}^{k}\frac{N_i}{N}S_i^2+\sum_{i=1}^{k}\frac{N_i}{N}(\bar{y}_i-\bar{y})^2$$

$$S^2 = \sum_{i=1}^{k}P_iS_i^2+\sum_{i=1}^{k}P_i(\bar{y}_i-\bar{y})^2 \qquad \ldots (3)$$

Now $\text{var}(\bar{y})_R = \frac{N-n}{Nn}S^2$ ... (4)

Substitute S^2 in (4)

$$= \frac{N-n}{Nn}\left[\sum_{i=1}^{k}P_iS_i^2+\sum_{i=1}^{k}P_i(\bar{y}_i-\bar{y})^2\right]$$

$$= \frac{N-n}{Nn}\sum_{i=1}^{k}P_iS_i^2+\frac{N-n}{Nn}\sum_{i=1}^{k}P_i(\bar{y}_i-\bar{y})^2$$

$$= \text{var}(\bar{y})_{prp} + \frac{N-n}{Nn}\sum_{i=1}^{k}P_i(\bar{y}_i-\bar{y})^2$$

$\text{var}(\bar{y})_R \geq \text{var}(\bar{y}_{sat})pr$ ... (II)

From (I) & (II)

$$\text{var}(\bar{y}_{st})_{op} \leq \text{var}(\bar{y}_{st})_{pr} \leq \text{var}(\bar{y})_R$$

$\therefore$ Proportional method is best.

PROBLEMS

1. Population of size is 660 is divided into 3 strata as follows.

No.	I	II	III
Population size	150	250	260
S.D.	5	7	6

A stratified random sampling size of 100 is to be selected from the population. Find the sample size of proportional and optimum method.

Sol :

$N = 660,, n = 100$

$N_1 = 150, N_2 = 250, N_3 = 260$

$S_1 = 5, S_2 = 7, S_3 = 6, K = 3$

$$n_i = \frac{n}{N} \times N_i$$

$$n_1 = \frac{n}{N} \times N_1 = \frac{100}{660} \times 150 = 22.727$$

$$n_2 = \frac{n}{N} \times N_2 = \frac{100}{660} \times 250 = 37.87$$

$$n_3 = \frac{n}{N} \times N_3 = \frac{100}{660} \times 260 = 39.39$$

Sample size in optimum method

$$n_i = \frac{n_i N_i S_i}{\sum_{i=1}^{k} N_i S_i}$$

$$\sum_{i=1}^{k} N_i S_i = N_1 S_1 + N_2 S_2 + N_3 S_3$$

$$= 150 \times 5 + 250 \times 7 + 260 \times 6$$

$$= 4060$$

$$n_1 = \frac{nN_1 S_1}{\sum_{i=1}^{k} N_i S_i} = \frac{100 \times 150 \times 5}{4060}$$

$$n_1 = 18.47$$

$$n_2 = \frac{nN_2 S_2}{\sum_{i=1}^{k} N_i S_i} = \frac{100 \times 250 \times 7}{4060}$$

$n_2 = 43.10$

$$n_3 = \frac{nN_3S_3}{\sum_{i=1}^{k} N_iS_i} = \frac{100 \times 260 \times 6}{4060}$$

$n_3 = 38.42$

2. A stratified sample of 50 items gives the following data

Strata No.	Ni	ni	$\bar{y}_i$	S_i^2
1	30	5	35	40
2	50	10	40	55
3	60	15	40	80
4	60	20	55	140

Find the variance

Sol :

$$\text{var}(\bar{y}_{st}) = \frac{1}{N^2}\sum_{i=1}^{k}\left(\frac{N_i}{n_i} - 1\right)N_iS_i^2$$

$$\text{var}(\bar{y}_{st})_{pr} = \frac{N-n}{nN}\sum_{i=1}^{k} P_iS_i^2$$

$$\text{var}(\bar{y}_{st})_{opt} = \frac{1}{n}\left(\sum_{i=1}^{k} P_iS_i\right)^2 - \frac{1}{N}\sum_{i=1}^{k} P_iS_i^2$$

$$\sum_{i=1}^{k}\frac{N_i}{n_i} = \frac{N_1}{n_1} + \frac{N_2}{n_2} + \frac{N_3}{n_3} + \frac{N_4}{n_4} = 17.5$$

$$\text{var}(\bar{y}_{st}) = \frac{1}{N^2}\sum_{i=1}^{k}\left(\frac{N_i}{n_i} - 1\right)N_iS_i^2$$

$$= \frac{1}{200^2} \times 48200 = 1.205$$

$$P_i = \frac{N_i}{N} = \frac{30}{200} = 0.15,$$

$$\frac{N_2}{N} = 0.25,\ \frac{N_3}{N} = 0.3,\ \frac{N_4}{N} = 0.3$$

$$\sum_{i=1}^{k} P_i = 1$$

Start	N_i	n_i	$\bar{y}_i$	S_i^2	N_i/n_i	$\left(\frac{N_i}{n_i}-1\right)$	$\left(\frac{N_i}{n_i}-1\right)N_iS_i^2$	$P_i = \frac{N_i}{N}$	$P_iS_i^2$	S_i	P_iS_i
1	30	5	35	40	6	5	6000	0.15	6	6.32	0.948
2	50	10	40	55	5	4	11000	0.25	13.75	7.41	1.852
3	60	15	40	80	4	3	14400	0.3	24	8.94	2.683
4	60	20	55	140	3	2	16800	0.3	42	11.83	3.549
					18	14	48200		85.75		9.032

$$\text{var}(\bar{y}_{st}) = \frac{1}{N^2}\sum_{i=1}^{k}\left(\frac{N_i}{n_i}-1\right)N_iS_i^2$$

$$\text{var}(\bar{y}_{st})_{pr} = \frac{N-n}{nN}\sum_{i=1}^{k}P_iS_i^2$$

$$= \frac{200-50}{50\times 200} \times 85.75$$

$$\text{var}(\bar{y}_{st})_{pr} = 1.28625$$

$$\text{var}(\bar{y}_{st})_{op} = \frac{1}{n}\left(\sum_{i=1}^{k}P_iS_i\right)^2 - \frac{1}{N}\sum_{i=1}^{k}P_iS_i^2$$

$$= \frac{1}{50}(81.577) - \frac{1}{200}(85.75)$$

$$\text{var}(\bar{y}_{st})_{op} = 1.202.$$

3. The publisher of a business statistics text wants an estimate of the total number of students taking business statistics courses in all U.S. colleges. Suppose that there are 1,395 colleges in the United States, 364 have 2-years programs and 1,031 are 4-years schools. A simple random sample of 40 two- years schools and an independent simple random sample of 60 four-year schools were taken. The sample means and standard deviations of numbers of students enrolled in the past year in business statistics courses are given in the table. Estimate the total annual enrollment in business statistics courses.

	2 - Years Schools	4 - Years Schools
Mean	**154.3**	**411.8**
Standard deviation	**87.3**	**219.9**

Sol.:

Given data,

$N_1 = 364$

$N_2 = 1{,}031$

$n_1 = 40$

$n_2 = 60$

$\bar{y}_1 = 154.3$

$\bar{y}_2 = 411.8$

$s_1 = 87.3$

$s_2 = 219.9$

$\bar{y}_1 = 154.3$

$y_2 = 411.8$

The population total from the above estimates can be obtained as,

$$N\,\bar{y}_{st} = \sum_{i=1}^{n} N_i \bar{y}_i$$

$= (364 \times 154.3) + (1{,}031 \times 411.8)$

$= 56{,}165.2 + 4{,}24{,}565.8$

$= 4{,}80{,}731$

The variance estimates can be obtained as,

$$\text{Var}\ (\bar{y}_i) = \frac{s_i^2}{n_i} \times \frac{N_i - n_i}{N_i - 1}$$

$$\Rightarrow \quad \text{Var}(\bar{y}_i) = \frac{(87.3)^2}{40} \times \frac{364 - 40}{364 - 1}$$

$$= \frac{7621.29}{40} \times \frac{324}{363}$$

$= 190.532 \times 0.8925$

$= 170.06$

$$\Rightarrow \quad \text{Var}(\bar{y}_2) = \frac{(219.9)^2}{60} \times \frac{1{,}031 - 60}{1{,}031 - 1}$$

$$= \frac{48{,}356.01}{60} \times \frac{971}{1{,}030}$$

$= 805.9335 \times 0.9427$

$= 759.75$

Now the population proportion formula is applied as,

$$N^2 \text{Var}(\bar{y}_{st}) = \sum_{i=1}^{k} N_i^2\ \text{Var}\ (\bar{y}_{st})$$

$= (364)^2 \times 170.06 + (1{,}031)^2 \times 759.77$

$= 1{,}32{,}496 \times 170.06 + 80{,}76{,}05{,}879$

$= 83{,}01{,}38{,}148.8$

$\sqrt{(N^2 Var(\bar{y}_{st}))} = \sqrt{83,01,38,148.73} = 28,812$

95% confidence intervals can be obtained as,

$Z_{\alpha/z} = Z_{0,\,0.25} = 1.96$

$\therefore$ The required interval is,

$= 4,80,731 \pm 1.96 \times 28,812$

$= 4,80,731 \pm 56471.52$

$\Rightarrow$ 4,80,731 – 56471.52 to 4,80,731 + 56471.52

$\Rightarrow$ 4,24,259 to 5,37,203 are the required intervals which indicates the number of students enrolled.

2.2 SYSTEMATIC SAMPLING WHEN N = nK.

Q8. What do you understand by systematic sampling when n = nk.

Ans : **(Nov.-20)**

Systematic sampling is slight variation of the simple random sampling in which only the first sample unit is selected at random and the remaining units are automatically selected in a definite sequence at equal spacing from one another. This technique of drawing samples is usually recommended if the complete and up-to-date list of the sampling units, i.e., the frame is available and the units are arranged in some systematic order such as alphabetical, chronological, geographical order, etc. This requires the sampling units in the population to be ordered in such a way that each item in the population is uniquely identified by its order, for example the names of persons in a telephone directory, the list of voters, etc.

Let us suppose that N sampling units in the population are arranged in some systematic order and serially numbered from I to N and we want to draw a sample of size n from it such that

$$N = nk \Rightarrow k = \frac{N}{n}$$

where k is usually called the sample interval.

Systematic sampling consists in selecting any unit at random from the first k units numbered from 1 to k and then selecting every kth unit in secession subsequently. Thus, it the first unit selected at random is i^{th} unit, then the systematic sample of size n will consist of the units numbered.

$$i,\ i + k,\ i + 2k,\ ...,\ i + (n - 1)k$$

The random number 'i' is called the random start and its value, in fact, determines the whole sample. As an example, let us suppose that we want to select 50 voters from a list of voters containing 1,000 names arranged systematically, Here

$$n = 50 \quad \text{and} \quad N = 1,000 \Rightarrow k = \frac{N}{n} = \frac{1,000}{50} = 20.$$

Q9. Discuss about the following concepts with respect to systematic sampling.

(a) Population total and its variance

(b) Population proportion and its variance.

(c) Estimation of variance.

Ans :

(a) Population Total and its Variance

Population total in systematic sampling is represented as $\bar{y}_{sy}$ which can be estimated as,

$$\bar{y}_{sy} = \frac{N}{n}\sum_{i=1}^{n} y_i$$

Its variance estimate can be given as,

$$V(\bar{y}_{sy}) = \frac{N(N-n)}{2n(n-1)}\sum_{i=1}^{n-1}(y_{i+1} - y_i)^2$$

(b) Population Proportion and its Variance

Population proportion in systematic sampling is represented as 'P' which can be estimated as,

$$P_{sy} = \frac{n_1}{n}$$

Its estimate of variance ($\hat{P}$) which can be estimated as,

$$\hat{Var}(\hat{P}_{sy}) = \left(\frac{N-n}{N}\right)\frac{\hat{P}(1-\hat{P})}{n-1}$$

(c) Estimation of Variance Estimates

The estimation of variance estimates can not be performed using a direct method as it is impossible to get the unbiased estimate for $V(\bar{y}_{sys})$ when only a single sample exists. In such case, the following methods are adopted.

1. Consider that the ith systematic sample is drawn using SRS where the size of sample is n units then the estimate of variance estimate can be given as,

$$\hat{V}(\bar{y}_{sys}) = \left(\frac{1}{n} - \frac{1}{N}\right)S_i^2$$

Here, the value of S_i^2 can be given as,

$$S_i^2 = \frac{1}{n-1}\sum_{j=1}^{n}(y_{ij} - \bar{y}_i)^2$$

2. Another method is to consider the successive differences between the values of samples which results in,

$$\hat{V}(\bar{y}_{sys}) = \left(\frac{1}{n} - \frac{1}{N}\right)\sum_{i=1}^{n-1}(y_{i+1} + y_i)^2$$

3. The final method is a little complicated as it considers the selection of sample of size n which is in the form of multiple systematic independent sub samples of size (n/m). These samples should be inter-penetrating and selected randomly. The unbiased estimator of variance where $\hat{\bar{y}}$, i = 1, 2, 3, ..., m are considered as estimates of population mean can be given as,

$$\hat{V}(\hat{\bar{y}}_c) = \frac{1}{m(m-1)}\sum_{i=1}^{m}(\hat{\bar{y}}_i - \hat{\bar{y}}_c)^2$$

4. In a class there are 96 students. It is desired to take sample of a students. Use the systematic sample method to determine the sample size.

Sol :

N = 96, n = 10

$$k = \frac{96}{10} = 9.6 \simeq 10$$

If we select overall class students i^{th} sample is 6, 16, 26, 36, 46 .... 96.

Notations

1. $\bar{y}_i$ = mean of the i^{th} systematic sample

$$\bar{y}_i = \frac{1}{n}\sum_{j=1}^{n} y_{ij}, \ (i = 1, 2, \ldots k)$$

2. $\bar{y}_{..}$ = the population mean

$$\bar{y}_{..} = \frac{1}{nk}\sum_{i=1}^{n}\sum_{j=1}^{n} y_{ij} = \frac{1}{nk}\sum_{i=1}^{n} n.y_{ij}$$

$$\bar{y}_{..} = \frac{1}{k}\sum_{i=1}^{n} y_{ij}$$

3. S^2 = Population mean square

$$S^2 = \frac{1}{N-1}\sum_{i=1}^{k}\sum_{j=1}^{n}(y_{ij} - \bar{y}_{..})^2$$

(or)

$$S^2 = \frac{1}{nk-1}\sum_{i=1}^{k}\sum_{j=1}^{n}(y_{ij} - \bar{y}_{..})^2$$

4. $$\text{var}(\bar{y}_{sys}) = \frac{1}{k}\sum_{i=1}^{k}(\bar{y}_i - \bar{y}_{..})^2$$

5. On public complaint that some gas cylinders supplied for domestic use were underweight, an inquiry committee was set up. The committee decided to examine 1-in-50 cylinders from the 8000 cylinders stored in a warehouse, arranged in rows by the gas company. The committee found 18 cylinders to be underweight from the 160 sampled cylinders. Estimate the total number N_1 and also the proportion of underweight cylinders in warehouse.

Sol :

Given that,

N = 8,000

n = 160

n_1 = 18

The estimate for the population proportion for under weight cylinder is given by,

$$P_{sy} = \frac{n_1}{n} = \frac{18}{160}$$

$$= 0.1125$$

The estimated number of such cylinders can be obtained as,

$$\hat{N}_1 = NP_{sy} = 8000 \times 0.1125$$

$$= 900$$

Let the gas cylinders be arranged in random order when drawing a sample. In such case the variance can be obtained as follows,

$$\text{Var}(\hat{N}_1) = \frac{N(N-n)P_{sy}(1-P_{sy})}{n-1}$$

$$= \frac{8000(8000-160)(0.1125)(1-0.1125)}{160-1}$$

$$= \frac{8000 \times 7840 \times 0.1125 \times 0.8875}{159}$$

$$= \frac{62,62,200}{159}$$

= 39,384,9056.

2.3 COMPARISON OF THEIR RELATIVE EFFICIENCIES

Q10. Distinguish between relative efficiency of systematic sampling over stratified random sampling.

Ans : **(July-21)**

A systematic sample can be considered as stratified sample in certain specific cases. For instance, a population satisfying the condition of systematic sampling i.e., N = nk units can be divided into n number of strata where the first unit is chosen randomly. Now, the given sample is converted to stratified random sample having size 'n'.

The mean of a stratum is stratified sampling having k strata is given as,

$$\bar{y}_{st} = \frac{1}{N}\sum_{i=1}^{k} N_i \bar{y}_i$$

This mean acts as unbiased estimator of population mean.

Consider the same sample for systematic sampling with n number of strata k size and size of sample can be obtained from every stratum as 1. In this cas, the above equation becomes,

$$\bar{y}_{st} = \frac{1}{nk}\sum_{i=1}^{k} N_i \bar{y}_i = \frac{1}{nk}\, k\sum_{i=1}^{k} \bar{y}_i$$

$$= \frac{1}{n}\sum_{i=1}^{n} \bar{y}_i$$

Similarly, the variance is computed as,

$$Var(\bar{y}_{st}) = \frac{1}{n^2}\sum_{i=1}^{n} Var(\bar{y}_i)$$

$$= \frac{1}{n^2}\sum_{i=1}^{n} \frac{k-1}{k.1} S_i^2 \qquad \left[\because Var(\bar{y}_{SRS}) = \frac{N-n}{N_n} S^2\right]$$

$$= \frac{k-1}{kn^2}\sum_{i=1}^{n} S_i^2$$

$$\therefore \quad S_{wst}^2 = \frac{1}{n}\sum_{i=1}^{n} S_i^2$$

$$= \frac{1}{n(k-1)}\sum_{i=1}^{k}\sum_{j=1}^{n} (y_{ij} - \bar{y}_j)^2$$

This called the mean sum of squares within the strata.

Now, the variance of systematic sample mean can be given as,

$$\text{Var}(\bar{y}_{sy}) = \frac{1}{k}\sum_{i=1}^{k}(\bar{y}_i - \bar{Y})^2 = \frac{1}{k}\sum_{i=1}^{k}\left[\frac{1}{n}\sum_{j=1}^{n}y_{ij} - \frac{1}{n}\sum_{j=1}^{n}\bar{y}_j\right]^2$$

$$= \frac{1}{n^2k}\sum_{i=1}^{k}\left[\sum_{j=1}^{n}(y_{ij} - \bar{y}_j)\right]^2$$

$$= \frac{1}{n^2k}\left[\sum_{i=1}^{k}\sum_{j=1}^{n}(y_{ij} - \bar{y}_j)^2 + \sum_{i=1}^{k}\sum_{j\neq 1=1}^{n}(y_{ij} - \bar{y}_j)(y_{i1} - \bar{y}_l)\right]$$

For a pair of units existing in the same systematic sample, the interaclass correlation coefficient can be given as,

$$\rho_{wst} = \frac{E(y_{ij} - \bar{Y})(y_{il} - \bar{Y})}{E(y_{ij} - \bar{Y})^2}$$

This can also be written as,

$$\rho_{wst} = \frac{\sum_{i=1}^{k}\sum_{j\neq l=1}^{n}(y_{ij} - \bar{y}_j)(y_{il} - \bar{y}_l)}{(N-n)(n-1)S^2_{wst}}$$

$$\Rightarrow (N-n)(n-1)\ S^2_{wst\,\rho\,wst} = \sum_{i=1}^{k}\sum_{j\neq l=1}^{n}(y_{ij} - \bar{y}_j)(y_{il} - \bar{y}_l)$$

$$\text{Moreover, } S^2_{wst} = \frac{1}{N-n}\sum_{i=1}^{k}\sum_{j=1}^{n}(y_{ij} - \bar{y}_j)^2$$

$$\Rightarrow (N-n)\ S^2_{wst} = \sum_{i=1}^{k}\sum_{j=1}^{n}(y_{ij} - \bar{y}_j)^2$$

Substituting the above values in $\text{Var}(\bar{y}_{sy})$

$$\Rightarrow \text{Var}(\bar{y}_{sy}) = \frac{1}{n^2k}\left[(N-n)S^2_{wst} + (N-n)(n-1)\rho_{wst}S^2_{wst}\right]$$

$$= \frac{(N-n)}{n^2k}S^2_w[1 + (n-1)\rho_{wst}]$$

In systematic sample when N = nk, the above equation becomes,

$$\text{Var}(\bar{y}_{sy}) = \frac{N-n}{Nn}S^2_{wst}[1 + (n-1)\rho_{wst}]$$

$$= \frac{N-n}{Nn}S^2_{wst} + \frac{N-n}{Nn}(n-1)\rho_{wst}S^2_{wst}$$

$$= \text{Var}(\bar{y}_{st}) + \frac{N-n}{Nn}(n-1)\,\rho_{wst}S^2_{wst} \qquad \left[\because \text{Var}(\bar{y}_{st}) + \frac{N-n}{Nn}(n-1)S^2_{wst}\right]$$

$$\Rightarrow \text{Var}(\bar{y}_{sy}) - \text{Var}(\bar{y}_{st}) = \frac{N-n}{Nn}(n-1)\ \rho_{wst}S^2_{wst}$$

Therefore, it can be concluded that,

1. Systematic sampling is considered as efficient over stratified random sampling when the value of rwst is greater than zero.
2. Stratified random sampling is considered as efficient over systematic random sampling when the value of ρ_{wst} is less than zero.
3. Stratified and systematic sampling are equally efficient when the value of ρ_{wst} is equal to zero.

Q11. Comparison between stratified, systematic and random simplifying

$$\mathbf{var(\bar{y}_{st}) \le var(\bar{y}_{sys}) \le var(\bar{y}_n)_R}$$

Ans :

Let us suppose that the population has a linear trend given by the model.

$$y_i = i \; ; (i = 1, 2 \ldots N)$$

then $$\sum_{i=1}^{N} y_i = \sum_{i=1}^{N} i = \frac{N(N+1)}{2}$$

$$\sum_{i=1}^{N} y_i^2 = \sum_{i=1}^{N} i^2 = \frac{N(N+1)(2N+1)}{6}$$

$$\bar{y}_N = \frac{1}{N}\sum_{i=1}^{N} y_i = \frac{N+1}{2}$$

$$S^2 = \frac{1}{N-1}\sum_{i=1}^{N}(y_i - y_N)^2 = \frac{1}{N-1}\left[\sum_{i=1}^{N} y_i^2 - N\bar{y}_N^2\right]$$

$$= \frac{1}{N-1}\left[\frac{N(NH)(2N+1)}{6} - \frac{N(N+1)^2}{4}\right]$$

$$= \frac{N(N+1)}{12} \text{ (on simplification)}$$

$$\therefore \quad var(\bar{y}_n)_R = \left(\frac{1}{n} - \frac{1}{N}\right)s^2 \frac{N-n}{Nn}s^2$$

$$= \frac{n(k-1)}{n^2k} \cdot \frac{nk(nk+1)}{12}$$

$$= \frac{(k-1)(nk+1)}{12} \qquad [\because N = nk]$$

$$var(\bar{y}_{st}) = \frac{k-1}{n^2k}\sum_{j=1}^{n} S_j^2$$

We have $$S^2 = \frac{N(N+1)}{12}$$

For population of N units since j^{th} stratum consists of k units, we have,

$$S_j^2 = \frac{K(K+1)}{12}$$

$$\therefore \quad Var(\bar{y}_{st}) = \frac{K-1}{n^2k}.\frac{nk(k+1)}{12} = \frac{k^2-1}{12n} \quad \ldots (2)$$

For finding out $var(\bar{y}_{sys})$, we have

$\bar{y}_{i.}$ = mean of the values of i^{th} sample

= mean of the i^{th} row

$$= \frac{1}{n}\sum_{i=1}^{n} y_{ij}$$

$$= \frac{1}{n}[i + (i + k) + (i + 2k) \ldots + i + (n - 1)k]$$

$$= \frac{1}{n}[(ni + \{1 + 2 + \ldots + (n - 1)\}k]$$

$$= i + \frac{(n-1)}{2}k \quad \ldots (3)$$

Also $\bar{y}_{..} = \bar{y}_N = \frac{N+1}{2} = \frac{nk+1}{2} \quad \ldots (4)$

$$\bar{y}_{i.} = \bar{y}_{..} = i - \frac{k+1}{2}$$

$$\therefore \quad var(\bar{y}_{sys}) = \frac{1}{k}\sum_{i=1}^{k}(\bar{y}_{i.} - \bar{y}_{..})^2$$

$$= \frac{1}{k}\sum_{i=1}^{k}\left(i - \frac{k+1}{2}\right)^2$$

$$= \frac{1}{k}\sum_{i=1}^{k}\left[i^2 + \left(\frac{k+1}{2}\right)^2 - 2\left(\frac{k+1}{2}\right)i\right]$$

$$= \frac{1}{k}.\sum_{i=1}^{k} i^2 + \left(\frac{k+1}{2}\right)^2 - \left(\frac{k+1}{2}\right).\sum_{i=1}^{k} i$$

$$= \frac{(k+1)(2k+1)}{6} + \frac{(k+1)^2}{4} - \frac{(k+1)^2}{2}$$

$$= \frac{k^2-1}{12} \text{ (on simplification)} \quad \ldots (5)$$

From (1), (2) and (5) we get,

$$\text{var}(\bar{y}_{st}) : \text{var}(\bar{y}_{st}) : \text{var}(\bar{y}_n) :: \frac{k+1}{n} : (k+1) : (nk + 1)$$

$$\cong \frac{1}{n} : 1 : n(\text{approximation})$$

$$\therefore \quad \text{var}(\bar{y}_{st}) \leq \text{var}(\bar{y}_{sys}) \leq \text{var}(\bar{y}_n)$$

PROBLEMS

6. Find the variance of the estimates of the mean under systematic sampling, stratified sampling and SRSWOR and compare them where the size of the sample in each case is 4.

S.No. / Sample Values	1	2	3	4	5	6	7	8	9	10
I	0	1	1	2	5	4	7	7	8	6
II	6	8	9	10	13	12	15	16	16	17
III	18	19	20	20	24	22	25	28	29	27
IV	26	30	31	31	33	32	35	37	38	38

Sol :

$$\text{var}(\bar{y}_{sys}) = \frac{1}{n^2 k} \Sigma T_i^2 - \frac{G^2}{N^2}$$

Where

$T_i = \Sigma y_{ij}$, G = Grand total

$$G = \sum_{i=1}^{K} \sum_{j=1}^{n} y_{ij}$$

$$\text{var}(\bar{y}_{SRSWOR}) = \frac{N-n}{Nn} S^2$$

$$S^2 = \frac{\text{Total sum of squares(T, S.S)}}{N-1}$$

$$\text{Total sum of squares T.S.S.} = \sum_{i=1}^{k} \sum_{j=1}^{n} y_{ij}^2 - \frac{G^2}{N}$$

$$\text{var}(\bar{y}_{sys}) = \frac{N-n}{Nn} S_{wsy}^2$$

$$S_{wsy}^2 = \frac{\text{With in the strata}}{N-n}$$

With in the strata is also know as "error sum of squares". (ESS)

E.S.S = T.S.S – B.S.S

Between sum of squares

$$B.S.S. = \frac{1}{k}\sum_{j=1}^{k} T_j^2 - \frac{G^2}{N}$$

$$T.j = \sum_{j=1}^{k} y_{ij}$$

S.No. / Sample Values	1	2	3	4	5	6	7	8	9	10	$T_{.j}$	$T_{.j}^2$
I	0	1	1	2	5	4	7	7	8	6	41	16181
II	6	8	9	10	13	12	15	16	16	17	122	14884
III	18	19	20	20	24	22	25	28	29	27	232	53824
IV	26	30	31	31	33	32	35	37	38	38	331	109561
$T_{i.}$	50	58	61	63	75	70	82	88	91	88	$\Sigma T.j$ 726	$\Sigma T.j^2$ 179950
$T_{i.}^2$	2500	3364	3721	3969	5625	4900	6724	7744	8281	7744		

Now

$$v(v(\bar{y}_{sys}) = \frac{1}{n^2 k}\Sigma T_{i.}^2 - \frac{G^2}{N^2}$$

G = 726, N = 40, K = 10, n = 4. $ST_{i.}^2$ = 54572

$$v(\bar{y}_{sys}) = \frac{1}{4^2 \times 10} \times 54572 - \frac{(726)^2}{(40)^2}$$

$$v(\bar{y}_{sys}) = 11.6525$$

$$T.S.S = \Sigma\Sigma y_{ij}^{\ 2} - \frac{G^2}{N}, \ \Sigma\Sigma y_{ij}^{\ 2} = 18482$$

$$T.S.S = 18482 - \frac{(726)^2}{40}$$

$$T.S.S = 5305.1$$

$$S^2 = \frac{T.S.S}{N-1} = \frac{5305.1}{40-1} = 136.0282$$

$$v(\bar{y}_{SRSNOR}) = \frac{N-n}{Nn}.S^2$$

$$= \frac{40-4}{40 \times 4}(136.028)$$

$$v(\bar{y}_{SRSNOR}) = 30.6063$$

$$B.S.S = \frac{1}{k}\Sigma T_{.j}^2 - \frac{G^2}{N}$$

$$= \frac{1}{10}(179950) - \frac{(726)^2}{40}$$

$$B.S.S = 4818.1$$

$$E.S.S = T.S.S - B.S.S$$

$$E.S.S = 487 = W.S.S$$

$$v(\bar{y}_{st}) = \frac{N-n}{Nn} S_{wsy}^2$$

$$S_{wsy}^2 = \frac{W.S.S}{N-n}$$

$$= 13.527$$

$$v(\bar{y}_{st}) = \frac{40-4}{40(4)} \times 13.527$$

$$v(\bar{y}_{st}) = 3.0437$$

$$\therefore \quad v(\bar{y}_{st}) < v(\bar{y}_{sys}) < v(\bar{y})_{SRSWOR}$$

7. Compare the variance of SRSWOR, stratified Random sampling and systematic random sampling from the given data.

Sol :

S.No.	1	2	3	4	5	6	7	8	9	10
I	2	6	3	4	4	2	4	6	4	3
II	4	2	8	6	3	1	3	3	8	5
III	3	3	2	5	7	7	5	5	9	6
IV	5	4	7	2	8	5	2	4	1	5

$$var(\bar{y}_{sys}) = \frac{1}{n^2K}\Sigma T_i^2 - \frac{G^2}{N^2}$$

Where

$$T_i = \Sigma y_{ij}$$

$$G = \text{Grand total} = \sum_{i=1}^{k}\sum_{j=1}^{n} y_{ij}$$

$$\text{var}(\bar{y})_{SRSWOR} = \frac{N-n}{Nn}S^2$$

$$S^2 = \frac{T.S.S}{N-1}$$

$$T.S.S = \sum_{i=1}^{k}\sum_{j=1}^{n} y_{ij} - \frac{G^2}{N}$$

$$\text{var}(\bar{y}_{st}) = \frac{N-n}{Nn}S^2_{wsy}$$

$$S^2_{wsy} = \frac{\text{With in the strata}}{N-n} \text{ (error sum squares)}$$

E.S.S = T.S.S – B.S.S (Between sum of squares)

$$B.S.S = \frac{1}{K}\Sigma T_{.j}^2 - \frac{G^2}{N}$$

$$T_{.j} = \sum_{j=1}^{k} y_{ij}$$

S.No. / Sample values	1	2	3	4	5	6	7	8	9	10	$T_{.j}$	$T_{.j}^2$
I	2	6	3	4	4	2	4	6	4	3	38	1444
II	4	2	8	6	3	1	3	3	8	5	43	1849
III	3	3	2	5	7	7	5	5	9	6	52	2704
IV	5	4	7	2	8	5	2	4	1	5	43	1849
$T_{i.}$	14	15	20	17	22	15	14	18	22	19	$\Sigma T_{.j}$ 196 $\Sigma T_{i.}$	7846 $\Sigma T_{.j}^2$
$T_{i.}^2$	196	225	400	289	484	225	196	324	484	361	3184 $\Sigma T_{.}^2$	

Stratum No.	N_i	$\bar{y}_{Ni}$	S_i	nN_iS_i	N_iS_i	$(N_i - 1)S_i^2$	$N_i\bar{y}_{Ni}^2$	$\bar{y}_N^2 = \Sigma P_i\bar{y}_{N_i}$	$N\bar{y}^2$
1	394	5.4	8.3	327020	3270.2	27073.77	11489.04	7.271544	11489.04
2	461	16.3	13.3	613130	6131.3	81369.4	122483.09	77.52094	122483.09
3	391	24.3	15.1	590410	5904.1	88923.9	230881.59	146.12758	230881.5
4	334	34.5	19.8	661320 2191880	6613.2 21918.8	130941.36 328308.4	397543.5 762397.22	251.6098101 482.5298861	3975435 762397.2

$$v(\bar{y}_{st})_{prop} = \left[\frac{1}{100} - \frac{1}{1580}\right] \times 208.0900127$$

$$= 1.949197587$$

$$(\bar{y}_{st})_{opt} = \frac{1}{100} \times 50.8251317$$

$$- \frac{1}{1580} \times 208.0900$$

$$= 0.3765487$$

$$\therefore (\bar{y}_{st})_{opt} < v(\bar{y})_{SRSWOR} < (\bar{y}_{st})_{prop}$$

2.4 Advantages and Disadvantages of SRS, Stratified and Systematic Sampling Methods

Q12. What are the advantages and disadvantages of Stratified Random Sampling?

Ans :

Advantages

1. More Representative Sample

A properly constructed and executed stratified random sampling plan overcomes the drawbacks of purposive sampling and random sampling and still enjoys the virtues of both these methods by dividing the given universe into a number of homogeneous subgroups with respect to purposive characteristic and then using the technique of random sampling in drawing samples from each stratum. A stratified random sample gives adequate representation to each strata or important section of the population and eliminates the possibility of any important group of the population being completely ignored. The stratified random sampling provides a more representative sample of the population and accordingly results in less variability as compared with other sampling designs.

2. Greater Precision

As a consequence of the reduction in the variability within each stratum, stratified random sampling provides more efficient estimates as compared with simple random sampling. For instance, the sample estimate of the population mean is more efficient in both proportional and Neyman's allocation of the samples to different strata in stratified random sampling as compared with the corresponding estimate obtained in simple random sampling.

3. Administrative Convenience

The division of the population into relatively homogeneous subgroups brings administrative convenience. Unlike random samples, the stratified samples are expected to be localised geographically. This ultimately results in reduction in cost and saving in time in terms of collection of the data, interviewing the respondents and supervision of the field work.

4. Sometimes it is desired to achieve different degrees of accuracy for different segments of the population. Stratified random sampling is the only sampling plan which enables us to obtain the results of known precision for each of the stratum.

5. Quite often, the sampling problems differ quite significantly in different segments of the population. In such a situation, the problem can be tackled effectively through stratified sampling by regarding each segment of the population as a different strata and approaching upon them independently during sampling.

Disadvantages

1. As already pointed out the success of stratified random sampling depends on :

   (i) Effective stratification of the universe into homogeneous strata and

   (ii) Appropriate size of the samples to be drawn from each of the stratum.

   If stratification is faulty, the results will be biased. The error due to wrong stratification cannot be compensated even by taking large samples.

The allocation of the sample sizes to different strata requires an accurate knowledge of the population size in each stratum N_i, i = 1, 2, ..., k, [c.f., Proportional Allocation $n_i \propto N_i$]. Further Neyman's principle of optimum allocation, *viz*., $n_i \propto N_i S_i$, requires an additional knowledge of the variability or standard deviation of each strata. N_i and S_i (i = 1, 2, ..., k) are usually unknown and are a serious limitation to the effective use of stratified random sampling.

2. Disproportional stratified sampling requires the assignment of weights to different strata and if the weights assigned are faulty, the resulting sample will not be representative and might give biased results.

Q13. Explain merits and demerits and systematic sampling.

Ans :

Merits

1. Systematic sampling is very easy to operate and checking can also be done quickly. Accordingly, it results in considerable saving in time and labour relative to simple random sampling or stratified random sampling.
2. Systematic sampling may be more efficient than simple random sampling provided the frame is complete and up-to-date and the units are arranged serially in a random order like the names in a telephone directory where the units are arranged in alphabetical order. However, even in alphabetical arrangement, certain amount of non-random character may persist.

Demerits

1. Systematic sampling works well only if the complete and up-to-date frame is available and if the units are randomly arranged. However, these requirements are not generally fulfilled.
2. Systematic sampling gives biased results if there are periodic features in the frame and the sampling interval (k) is equal to or a multiple of the periods.

Q14. Discuss the advantages and disadvantages of Simple Random Sampling (SRS).

Ans :

Advantages

Simple Random Sampling (SRS) offers the following advantages,

1. The samples obtained using SRS are more appropriate and representative than the samples obtained using judgement sampling.
2. Simple random sampling is economical as it saves time, labour and money.
3. Sampling errors can be measured.
4. It is useful for finite as well as growing population.
5. Every item of the population has the same chance of being selected, hence it is more scientific and significant in biostatistical research.

Disadvantages

Following are the disadvantages of using simple random sampling,

1. It needs a complete and update list of all elements which sometimes may not be feasible.
2. It is not applicable in case if the elements of population are spread over large area thereby involving increased cost and time.
3. Simple random sampling cannot be a perfect and correct representative of the population when the sample size is small.
4. It is in efficient in case of studies involving data in social sciences.
5. It requires larger samples to achieve greater accuracy.
6. It can some times produce results that are not probabilistic.

2.5 Time Series

Q15. Define Time Series. Explain the applications of time series analysis.

Ans :

An arrangement of statistical data in accordance with time of occurrence or in a chronological order is called a time series. The numerical data which we get at different points of time-the set of observations is known as time series.

Examples:

i) The Annual Production of Steel in India over the last 10 years;

ii) The Monthly Sales of a Chemical Industry for the last 6 months;

iii) The daily closing price of a share in the Calcutta Stock Exchange;

iv) Hourly temperature recorded by the Meteorological office in a city;

v) Yearly Price or Quantity Index Numbers.

Mathematically, a Time Series is defined by the values Y_1, Y_2, —, Y_n of a variable Y at times t_1, t_2,....t_n

Here Y is a function of time t and Y, denotes the value of the variable Y at time t. The basic assumption is that changes witnessed over time in a sample group will be extrapolated to population.

i) According to Morris Hamburg, "A time series is a set of observations arranged in chronological order".

ii) According to Kenny and Keeping, "A set of data depending on the time is called time series".

iii) According to Croxton and Cowden, "A time series consists of data arranged chronologically".

iii) According to Patterson, "Time series consists of statistical data which are collected, recorded or observed over successive increments".

Application of Time Series Analysis

1) **In Financial Decision Making:** Econometrics is vital in finance and in financial time series analysis. A key element of financial planning and financial forecasting is the ability to construct models showing the interrelatedness of financial data. Models showing correlation or causation between variables can be used to improve financial decision-making. **For example,** one would be more concerned about the consequences on the domestic stock market of a downturn in another economy, if it can be shown that there is a mathematically provable causative impact of that nation's economy and the domestic stock market.

2) **In Inventory Decision Making:** Better forecasting, physical control, supplier reliability, and an ultimate reduction in safety stock and inventory investment can all result from perfect analysis of the data from the time series technique. Without them, managers cannot make precise decisions about ordering, scheduling and shipping. The inventory related decision making relates to ensuring accuracy, incoming and outgoing record keeping must be good, and well-organized inventory storage will have limited access, good housekeeping, and storage areas that hold fixed amounts of inventory.

3) **To Understand Past Behavior of Business:** The analysis of a time series

enables us to understand the past behavior or performance of business. We can know how the data have changed over time and find out the probable reasons responsible for such changes. If the past performance of a company has been poor, it can lake corrective measures to arrest the poor performance.

4) **Helpful in Business Planning:** A time series analysis helps directly in business planning. A firm can know the long-term trend in the sale of its products. It can find out at what rate sales have been increasing over the years. This may help it in making projections of its sales for the next few years and plan the procurement of raw material, equipment and manpower accordingly.

5) **To Understand Cyclical Fluctuations:** A time series analysis enables one to study such movements as cycles that fluctuate around the trend. Knowledge of cyclical pattern in certain series of data will be helpful in making generalizations in the concerned business or industry.

6) **To Analyze Meaningful Comparisons:** A time series analysis enables one to make meaningful comparisons in two or more series regarding the rate or type of growth. **For example,** growth in consumption at the national level can be compared with that in the national income over specified period. Such comparisons are of considerable importance to business and industry.

7) **Marketing and Modeling Advertising Campaign:** Marketing department does quite a expenditure hence the data keeping, analysis and interpretation acquires prime importance. The expenditure with the time and the corresponding output can be analyzed by the time series.

Q16. Explain the Utility of Time Series Analysis

Ans :

i) It Helps in the Analysis of past Behaviour of a Variable

Analysis of past data discloses the effect of various factors on the variable under study. These studies isolate and analyze the effects of various sets of homogenous factors on the problem under study.

ii) It Helps in Forecasting :

The analysis of past condition is the basis of forecasting the future behaviour of the variable under study. For example, the analysis of a time series relating to income or wages or production would be the basis for forecasting future income or wages or production. This helps in making future plan of action.

iii) It helps in Evaluation of Current Achievement :

The review and evaluation of progress made on the basis of a plan are done on the basis of time series data. For example the progress of our Five-year plans is judged by the annual growth rates in the gross national product.

iv) It Helps in Making Comparative studies :

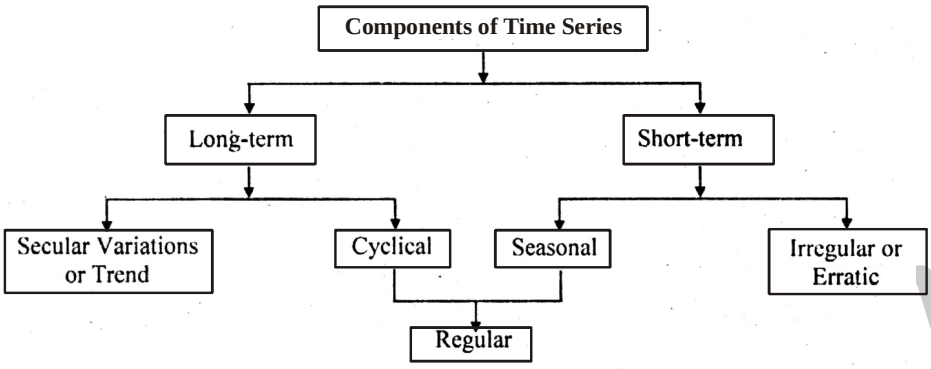
Once the data are arranged chronologically comparison between one time period and another is facilitated. It provides a scientific basis for making comparisons by studying and isolating the effects of various components of a time of series. It also helps in making regional comparison amongst data collected on the basis of time.

2.5.1 Components of Time Series with illustrations

Q17. What are the Components of Time Series ?

Ans : **(Nov.-20, Dec.-18(MGU)**

A lime series is a historical series of statistical data. Since these statistical data are historical, they are subject to the influences of all the changes that may occur over any period of time. The usual classification of the influences affecting statistical data recording over time is one based upon the nature of the influence. Classified in this manner, these influences are as follows :



1) **Secular Variations or Trend (T):** The general tendency of the time series data to increase or decrease or stagnate during a long-period of time is called the secular trend or simply trend.

According to Simpson and Kafka: "Trend, also called secular or long-term trend, is the basic tendency of a series to grow or decline over a period of time. The concept of trend does not include short-range oscillations, but rather the steady movements over a long time".

The trend thus refers to the general direction and the movement of time series considering a fairly long period of time. Most business and economic series are influenced by some secular forces of charge in which the underlying tendency is one of growth or decline, for example, in series concerning population, national income, agricultural production, currency in circulation, bank deposits, etc., an upward tendency is usually observed while a downward tendency is noticed in time series relating 10 birth and death rates specially in deaths by epidemics, tuberculosis, etc.

2) **Seasonal Variations (S) :** Seasonal variations refers to rhythmic forces of change inherent in most time series showing a regular or a periodic pattern of movement over a span of less than a year and has the same or almost the same pattern year after year.

According to Hirsch "a recurrent pattern of charge within the period that results from the operation of forces connected with climate or custom at different times of the period."

Reasons for Seasonal Variations

i) **Climate and Natural Forces:** The result of natural forces like climate is causing seasonal variation. Umbrellas are sold more in rainy season. In winter season, sale of the woolen clothes will increase. In hot season, the sales of ice, ice-cream, fruit salad, etc., will increase. Thus climate and weather play an important role in seasonal movement. Agricultural production depends upon the monsoon.

ii) **Customs and Habits:** Man-made conventions are the customs, habits, fashion, etc. There is the custom of wearing new clothes, preparing sweets and buying crackers for Deepavali, Onam, Christmas, etc. At that time, there is more demand for cloth, sweets and crackers. It will happen every year. In marriage season, the price of gold will increase.

Purposes of Studying Seasonal Variations

i) To analyze seasonal pattern in a short-period time series.

ii) Once the seasonal factor is known it can be used for separating cyclical and irregular forces from the residual forces of changes.

iii) The seasonal factor can be used for adjustment in the value projected on the basis of trend. Short-term forecasts are always affected by seasonal factors.

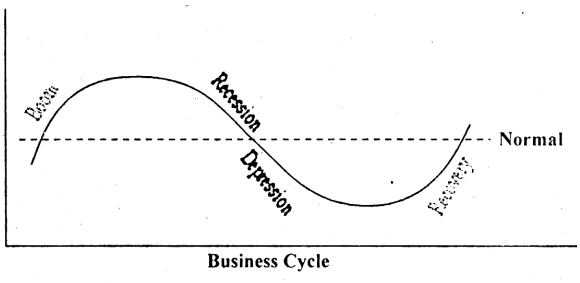
iv) The observed value can be appraised in terms of seasonal factor and adjusted so as to get correct idea of the trend, if any, in an economic phenomenon. The study of seasonal variations is, therefore, necessary for proper appraisal of business facts influenced by a seasonal variation.

v) The seasonal factor can also be used for pricing of articles and services so as to level up the seasonal variations in demand.

vi) A study of the seasonal patterns is extremely useful in planning future operations and in formulation of policy decisions regarding purchase, production, inventory control, personnel requirement, selling and advertising programme, etc.

3) **Cyclical Variations (C)**

As the economy expands during a period of boom, we would expect to find that such data as sales, output or consumer expenditure also show a rising trend; and during a period of slump, we would expect them to show a downward trend. Thus a wavelike motion may be observed in the pattern of our data. They are prosperity (boom) recession, depression and recovery; and are shown below:



The study of cyclical variation is helpful to businessmen and economist for framing suitable policies and stabilizing the level of business activities.

Purposes of Studying Cyclical Variation

i) One can easily study the character of business fluctuations. Good policies can be formulated at stabilizing the level of business activity.

ii) Businessman can take timely steps in maintaining business during booms and depressions.

iii) A careful study of cyclical variations facilitates a businessman to face the recession period and make them ready to reap the benefits during booms.

4) **Irregular Variations (I)**

Irregular variations are the effect of random factors. These are generally mixed up with seasonal and cyclical variations and are caused by irregular and accidental factors like floods, famines, wars, strikes, lockouts, etc. There is no regular period or time of their occurrence and that is why they are called random or chance fluctuations. Sometimes these factors are very effective and may even give rise to cyclical fluctuations. However since these are not regular factors, no advance preparation can be done to meet their consequences. Their effects are also unpredictable and irregular.

2.6 Models of Time Series Analysis

2.6.1 Additive, Multiplicative and Mixed Models

Q18. Explain the various models are used in Time Series Analysis.

Ans :

There are two mathematical models which are commonly used for the decomposition of a time series into its components viz.

(a) Decomposition by Additive model

(b) Decomposition by Multiplicative model.

(c) Decomposition by Mixed Model

(a) Decomposition by Additive Model : According to the additive mode, a time series can be expressed as

$$Y_t = T_t + S_t + C_t + R_t$$

where Y_i = is the time series value at time t,

T_t = represent the trend value.

S_t = represents the seasonal variations.

C_t = represents the cyclic movements and

R_t = represents the random fluctuations at time *t*.

Obviously, the term S_t will not appear in a series of annual data. The additive model implicitly implies that seasonal forces (in different years), cyclical forces (in different cycles) and irregular forces (in different long time periods) operate with equal absolute effect irrespective of the long trend value. As such C_t (and S_t) will haves positive or negative values according as whether we are in an above normal or below normal phase of the cycle (and year) and the total of positive and negative values for any cycle (and any year) will be zero. R_t will also have positive or negative values and in the long run R_t will be zero. Occasionally, there may be a few isolated occurrences of extreme R_t of episodic nature.

(b) Decomposition by Multiplicative Model : On the other hand if we have reasons to assume that the various components in a time series operate proportionately to the general level of the series, the traditional or classical multiplicative model is appropriate. According to the multiplicative model

$$Y_t = T_t \times S_t \times C_t \times R_t$$

where S_t, C_t and R_t instead of assuming positive and negative values are indices fluctuating above or below unity and the geometric means of S_t in a year, C_t in a cycle

and R_i in a long-term period are unity. In a time series with both positive and negative values the multiplicative model can not be applied unless the time series is translated by adding a suitable positive value. It may be pointed out that the multiplicative decomposition of a time series is same as the additive decomposition of the logarithmic values of the original time series.

In practice, most of the series relating to economic data confirm to the multiplicative model.

(c) Decomposition by Mixed Models : In addition to the additive and multiplicative models discussed above, the components in a time series may be combined in large number of ways. The different models, defined under different assumptions, will yield different results. Some of the defined models can be:

$$Y_t = (T_t \times S_t \times C_t) + R_t$$

$$Y_t = (T_t \times C_t) + (S_t \times R_t)$$

$$Y_t = T_t + (S_t \times C_t \times R_t)$$

$$Y_t = T_t + S_t + (C_t \times R_t)$$

$$Y_t = T_t \times (S_t + C_t) \times R_t$$

2.7 Determination of Trend by Least Square Method

Q19. Define Trend Analysis. Explain the purpose of measuring trend.

Ans :

The term trend analysis refers to the concept of collecting information and attempting to spot a pattern, or trend, in the information. In some fields of study, the term "trend analysis" has more formally-defined meanings.

Although trend analysis is often used to predict future events, it could be used to estimate uncertain events in the past, such as how many ancient kings probably ruled between two dates, based on data such as the average years which other known kings reigned.

Trend analysis uses a technique called least squares to fit a trend line to a set of time series data and then project the line into the future for a forecast.

Purposes of Measuring Trend

There are three basic purposes of measuring secular trend :

1) The first purpose is to study the past growth or decline of a series. The secular trend describes the basic growth tendency ignoring short-term fluctuations.

2) The second and most important purpose of measuring secular trend is to project the curve into the future as a long-term forecast. If the past growth has been steady and if the conditions that determine this growth may reasonably be expected to persist in the future, a trend curve may be projected over five to ten years into the future as a preliminary forecast.

3) The third purpose of measuring secular trend is to eliminate it, in order to clarify the cycles and other short-term movements in the data. A steep trend may observe minor cycles. Dividing the data, by the trend values yield ratios which make the curve fluctuate around a horizontal line, thus bringing the cycles into clear relief.

Q20. Explain the uses of trend analysis.

Ans :

Uses of Trend

1) The trend describes the basic growth tendency ignoring short term fluctuations.

2) It describes the pattern of behavior which has characterized the series in the past.

3) Future behavior can be forecasted in the assumption that past behavior will continue in the future also.

4) Trend analysis facilitate us to compare two or more time series over different period of time and this helps to draw conclusions about them,.

Q21. Explain the various methods for measuring trend.

Ans :

Methods / Measurement of Trend

There are Your methods for determining trend in time series :

1) Freehand (or Graphical) Method,

2) Semi-Average Method,

3) Moving Average Method.

4) Least Squares Method;

Q22. Explain the method of least square method.

Ans :

Least square method is the most widely used method and provides us with a mathematical device to obtain an objective fit to the trend of a given time series. This method is so called because a trend line computed by this method is such that the sum of the squares of the deviation between the original data and the corresponding computed trend values is minimum. This method can be used to fit either a straight line trend or a parabolic trend.

The straight line trend equation is in the form of $T = a + bX$

Where, Y denotes the trend value of the dependent variable

X denotes the independent variable.

a and b are constants.

The values of a and b are obtained by solving the following normal equations.

$\Sigma Y = Na + b \Sigma X$

$\Sigma XY = a\Sigma X + b\Sigma X^2$

Where, N represents the number of years in the series.

When $\Sigma X = 0$ the above normal equations are simplified to

$$a = \frac{\sum Y}{N}$$

$$b = \frac{\sum XY}{\sum X^2}$$

By substituting a and b values in straight line trend equation Y = a + bX, we get the straight line equation which can be used for estimation of future values.

Merits of Least Squares

The following are the merits of least squares method,

1. The method of least squares is a mathematical method of measuring trend and is free from subjectiveness.
2. This method provides the line of best fit since it is this line from where the sum of positive and negative deviations is zero and the sum of square of deviations is the least.
3. This method enables us to compute the trend values for all the given time periods in the series.
4. The trend equation can be used to estimate the values of the variable for any given time period 'f' in future and the forecasted values are quite reliable.
5. This method is the only technique which enables us to obtain the rate of growth per annum for yearly data in case of linear trend.

Demerits of Least Squares

Some of the demerits of least squares are as follows,

1. Fresh calculations become necessary even if a single new observation is added.
2. Calculations required in this method are quite tedious and time consuming as compared with other methods.
3. Future predictions based on this method completely ignore the cyclical, seasonal and erratic fluctuations.
4. This method cannot be used to fit growth curves, gomper t_Z curve, logistic curve etc. to which most of the business and economic time series conform.

PROBLEMS ON LEAST SQUARE METHOD

8. Given below is the data of production of a certain company in lakhs of units.

Year	1995	1996	1997	1998	1999	2000	2001
Production	15	14	18	20	17	24	27

Compute the Linear trend by the method of least squares. predict the production for the year "2004".

Sol :

The Normal Equation are,

$$\Sigma Y = Na + b\Sigma X$$

$$\Sigma XY = a\Sigma X + b\Sigma X^2$$

By solving these equations, we get a and b. This method is called the method of least squares.

Year	Production Y	Deviation from 1998 X	X^2	XY
1995	15	– 3	9	– 45
1996	14	– 2	4	– 28
1997	18	– 1	1	– 18
1998	20	0	0	0
1999	17	1	1	17
2000	24	2	4	48
2001	27	3	9	81
Total	**135**	**0**	**28**	**55**

Since $\Sigma X = 0$, the above equations simplify to,

$$a = \frac{\Sigma Y}{N} = \frac{135}{7}$$

$$= 19.29$$

$$b = \frac{\Sigma XY}{\Sigma X^2} = \frac{55}{28}$$

$$= 1.96$$

$\therefore$ The straight line trend, $Y_e = 19.29 + 1.96X$.

Prediction of Production for the Year 2004

For year 2004,

$X = 6$ (deviation from mid year 1998).

$\therefore$ $Y_c = 19.29 + 1.96\ (6) = 19.29 + 11.76 = 31.05$

The production of the company for the year 2004 is predicated to be 31.05 lakhs of units.

9. The sales of a company in lakhs of rupees for the years 1994-2001 are given below:

Year	1994	1995	1996	1997	1998	1999	2000	2001
Production	550	560	555	585	540	525	545	585

Compute the Linear trend by the method of least squares. predict the production for the year "2002".

Sol :

The model of least squares helps in solving this problem.

The straight line trend is $y_c = a + bX$

By solving these normal equations, we get a and b.

Normal equation are,

$$\Sigma Y = Na + b\Sigma X$$

$$\Sigma XY = a\Sigma X + b\,\Sigma X^2$$

Since there are even number of observations, deviations are taken from both of the middle periods i.e., 1997 and 1998.

Year	Sales Y	Deviations From mid years X	X^2	XY
1994	550	–7	49	– 3850
1995	560	–5	25	–2800
1996	555	–3	9	–1665
1997	585	–1	1	–585
1998	540	1	1	540
1999	525	3	9	1575
2000	545	5	25	2725
2001	585	7	49	4095
N = 8	**ΣY = 4445**	**ΣX = 0**	**Σx^2 = 168**	**ΣXY = 35**

Since $\Sigma Y = 0$, the normal equation simplify to,

$$a = \frac{\Sigma Y}{N} = \frac{4445}{8} = 555.63$$

$$b = \frac{\Sigma XY}{\Sigma X^2} = \frac{35}{168} = 0.21$$

$\therefore$ The straight line trend is $Y_e = 555.63 + 0.21X$

Estimation of Sales for the Year 2002

For year 2002, deviation X = 9,

$\therefore$ $Y = 555.63 + 0.21(9)$

$= 555.63 + 1.89$

$= 557.52$

For year 2002, the sales of the company is estimated to be ` 557.52 lakhs.

10. Below table shows the figures of production (in thousand tons) of a sugar factory.

Year	1979	1980	1981	1982	1983	1984	1985
Production	77	88	94	85	91	98	90

Fit a straight line trend by the method of least squares and show trend values.

Sol :

Computation of Straight Line Trend

Year (Y)	Production	X	XY	x^2	Trend $Y_c = a+bx$
1979	77	-3	-231	9	83
1980	88	-2	-176	4	85
1981	94	-1	-94	1	87
1982	85	0	0	0	89
1983	91	+1	+91	1	91
1984	98	+ 2	+196	4	93
1985	90	+ 3	+ 270	9	95
	623	0	56	28	

Since

$$\Sigma X = 0;\ a = \frac{\Sigma Y}{N} = \frac{623}{7} = 89$$

$$b = \frac{\Sigma XY}{\Sigma X^2} = \frac{56}{28} = 2$$

Σ The equation of Trend line is $Y_c = 89 + 2x$

On the basis of the above equation, we can calculate the value of Y_c for all the seven years. For example, Y_C for 1979 = 89+2(–3) = 89 – 6 = 83.

1980 = 89 + 2 (–2) = 89 – 4 = 85

1981 = 89 + 2 (–2) = 89 – 2 = 87

1982 = 89 + 2 (0) = 89 + 0 = 89

1983 = 89 + 2 (1) = 89 + 2 = 91

1984 = 89 + 2 (2) = 89 + 4 = 93

1985 = 89 + 2 (3) = 89 + 6 = 95

11. Fit a straight line trend for the following series. Estimate the value for 1997.

Year	1990	1991	1992	1993	1994	1995	1996
Production of steel (in tonnes)	60	72	75	60	80	85	95

Sol. :

FITTING STRAIGHT LINE TREND

Year	Production of steel (m. tonnes) Y	Deviations from 1993 X	XY	X^2
1990	60	–3	–180	9
1991	72	–2	–144	4
1992	75	–1	–75	1
1993	65	0	0	0
1994	80	+1	+80	1
1995	85	+2	+170	4
1996	95	+3	+285	9
N = 7	**$\Sigma Y = 532$**	**$\Sigma X = 0$**	**$\Sigma XY = 136$**	**$\Sigma X^2 = 28$**

The equation of the straight line trend is

$$\boxed{Y_C = a + b X}$$

$\Sigma X = 0$, $a = \dfrac{\sum Y}{\sum N}$

$= \dfrac{532}{7} = 76$

and $b = \dfrac{\sum XY}{\sum X^2} = \dfrac{136}{28}$

$= 4.857$

Hence, $Y_C = 76 + 4.857 X$

For 1997, X will be $Y_{1997} = 76 + 4.857 (4)$

$= 95.428$

Thus, the likely sales for the year 1997 is 95.428 m tonnes

12. Fit a straight line trend by the method of least squares to the following data. Assuming that the same rate of change continues, what would be the predicted earning for the year 1998 ?

Year	1989	1990	1991	1992	1993	1994	1995	1996
Earnings (Rs.lakhs)	38	40	65	72	69	60	87	95

Sol. :

FITTING OF STRAIGHT LINE TREND BY THE METHOD OF LEAST SQUARES

Year	Earnings (Rs.lakhs) Y	Deviations from 1992.5	Deviations multiplied by 2 X	XY	X^2
1989	38	-3.5	-7	-266	49
1990	40	-2.5	-5	-200	25
1991	65	-1.5	-3	-195	9
1992	72	-0.5	1	-72	1
1993	69	+0.5	+1	+69	1
1994	60	+1.5	+3	+180	9
1995	87	+2.5	+5	+435	25
1996	95	+3.5	+7	+665	49
N = 8	**ΣY=526**		**ΣX = 0**	**ΣXY=616**	**ΣX^2 = 168**

The equation of the straight line trend is

$$a = \frac{\sum Y}{N} = \frac{526}{8} = 65.75$$

$$b = \frac{\sum XY}{\sum X^2} = \frac{616}{168} = 3.667 \quad \boxed{Y_c = 65.75 + 3.667\,(X)}$$

For 1998, X will be +11.

When X is 11, Y will be

$Y_C = 65.75 + 3.667(11) = 65.75 + 40.337 = 106.087.$

Thus, the estimated earnings for the year 1998 are Rs.106.087 lakhs.

13. The following data relate to the number of passenger cars (in millions) sold from 2004 to 2011.

Years	2004	2005	2006	2007	2008	2009	2010	2011
Number	6.7	5.3	4.3	6.1	5.6	7.9	5.8	6.1

a) Fit a straight line trend to the data through 2009.

b) Use your result in, (a) To estimate production in 2011 and compare with the actual production.

Sol :

Let, the equation of the straight line trend be $Y_C = a + b_x$, then,

$\Sigma Y = Na + b\Sigma X$

$\Sigma XY = a\Sigma X + b\Sigma X$

a) Calculations to fit a straight line

Year	No. of Passenger cars (in millions) (Y)	Deviations from 2009	X^2 (X)	XY
2004	6.7	-5	25	-33.5
2005	5.3	-4	16	-21.2
2006	4.3	-3	9	-12.9
2007	6.1	-2	4	-12.2
2008	5.6	-1	1	-5.6
2009	7.9	0	0	0
2010	5.8	1	1	5.8
2011	6.1	2	4	12.2
N = 8	**ΣY = 47.8**	**ΣX = -12**	**ΣX^2 = 60**	**ΣXY = –67.4**

$\Sigma Y = Na + b\Sigma X$ or $47.8 = 8a - 126$ ... (1)

$\Sigma XY = a\Sigma X + b\Sigma X^2$ or $-67.4 = -12a + 60\,b$ ... (2)

Simplifying both the equations by multiplying equation(l) by 3 and equation (2) by 2, we get.

$$143.4 = 24a - 36\,b$$
$$-134.8 = -24a + 120\,b$$
$$8.6 = 84\,b$$

i.e., $b = \frac{8.6}{84} = 0.102$

Substituting the value of b in equation (1), we get,

$47.8 = 8a - 12\,b$

$47.8 = 8a - 12\,(0.102)$

$47.8 = 8a - 1.224$

$8a = 49.024$

$a = \frac{49.024}{8} = 6.128$

Hence, the required straight line equation is $\boxed{Y_C = 6.128 + 0.102\,X}$

(b) Estimating the sale of passenger cars in 2011 by substituting the value of X = 2.

$Y_C = 6.128 + 0.102\,(2)$

$= 6.128 + 0.204 = 6.332$

Thus, the estimated sale for 2011 is 6.332 million passenger cars whereas actual sales are 6.1 million passenger cars. There is some difference in actual figure and estimated figure because estimates depend on few assumptions.

Note : As number of passenger cars sold is given in question. So, estimated sales and actual sales are calculated instead of estimated production are calculated instead of estimated production and actual production.

2.7.1 Determination of Trend by Moving Average Method

Q23. Discuss the method of moving averages in measuring trend. What are its merits and limitations ?

Ans :

Moving Average Method

In moving average method, the average value for a number of years (month or weeks) is secured and this average is taken as the normal or trend value for the unit of time falling at the middle of the period covered in the calculation of the average.

The effect of averaging is to give a smoother curve, lessening the influence of the fluctuations that pull the annual figures away from the general trend.

The period of moving average is decided in the light of the length of the cycle. More applicable to data with cyclical movements.

Formula for 3 yearly moving average will be,

$$\frac{a+b+c}{3}, \frac{b+c+d}{3}, \frac{c+d+e}{3}...$$

Formula for 5 yearly moving average will be,

$$\frac{a+b+c+d+e}{5}, \frac{b+c+d+e+f}{5}, \frac{c+d+e+f+g}{5}....$$

Methods of Moving Average

The following two methods are followed in moving averages,

a) Odd Yearly Method

i) Calculate 3/5...yearly totals

ii) Now compute 3/5 yearly average by dividing the totals calculated in step (i) by the respective number of years, i 3/5/...

iii) Short term oscillations are calculated using the formula, $Y - Y_C$

Where, Y - Actual value and Y - Estimated value.

b) Even Yearly Method

Example : 4 years

i) Calculate 4 yearly moving totals and place at the centre of middle two years of the four years considered.

ii) Divide 4 yearly moving totals by 4 to get 4 yearly average.

iii) Take a 2 period moving average of the moving average which gives the 4 yearly moving average centered.

Merits of Moving Average

The merits of moving average are as follows,

1. Of all the mathematical methods of fitting a trend, this method is the simplest.«
2. The method is flexible so that even if a few more observations are to be added, the entire calculations are not changed.
3. If the period of the moving average happens to coincide with the period of the cycle, the cyclical fluctuations are automatically eliminated.
4. The shape of the curve in case of moving average method is determined by the data rather than the statisticians choice of mathematical function.

Limitations of Moving Average

The following are the limitations of moving averages,

1. Trend values cannot be computed for all the years. For example, in a 5 yearly moving we cannot compute trend values for the first two and the last two years.
2. It is difficult to decide the period of moving average since there is no hard and fast rule for the purpose.
3. Moving average cannot be used in forecasting as it is not represented by any mathematical function.
4. When the trend is not linear, the moving average lies either above or below the true sweep of the data.

PROBLEMS ON MOVING AVERAGE METHOD

14. Calculate the 3-yearly moving averages of the production figures given below and draw the trend:

Year	Production (in met. tonnes)	Year	Production (in met. tonnes)
1985	15	1993	63
1986	21	1994	70
1987	30	1995	74
1988	36	1996	82
1989	42	1997	90
1990	46	1998	95
1991	50	1999	102
1992	56	–	–

Sol .

CALCULATION OF 3-YEARLY MOVING AVERAGES

Year	Production Y	3-Yearly totals	3-Yearly Moving average
1985	15	—	—
1986	21	66	22.00
1987	30	87	29.00
1988	36	108	36.00
1989	42	124	41.33
1990	46	138	46.00
1991	50	152	50.67
1992	56	169	56.33
1993	63	189	63.00
1994	70	207	69.00
1995	74	226	75.33
1996	82	246	82.00
1997	90	267	89.00
1998	95	287	95.67
1999	102	—	—

15. Construct 5-yearly moving averages of the number of students studying in a college shown below :

Year	No.of students	Year	No. of students
1990	**332**	**1995**	**405**
1991	**317**	**1996**	**410**
1992	**357**	**1997**	**427**
1993	**392**	**1998**	**405**
1994	**402**	**1999**	**438**

Sol :

CALCULATION OF 5-YEARLY MOVING AVERAGES

Year	No.of Students	5-Yearly Total	5-Yearly Moving average
1990	332	—	—
1991	317	—	—
1992	357	1800	360.0
1993	392	1873	374.6
1994	402	1966	393.2
1995	405	2036	407.2
1996	410	2049	409.8
1997	427	2085	417.0
1998	405	—	—
1999	438	—	–

16. Assume a four-yearly cycle and calculate the trend by the method of moving averages from the following data relating to the production of tea in India :

Year	Production (mn.lbs.)	Year	Production (mn.lbs.)
1987	464	1992	540
1988	515	1993	557
1989	518	1994	571
1990	467	1995	586
1991	502	1996	612

Sol :

CALCULATION OF TREND BY THE MOVING AVERAGE METHOD

Year	Production (mn.lbs)	4-yearly moving totals	4-yearly moving average	4-yearly moving average centered
1987	464	—	—	—
1988	515	1964	491.00	—
		←	←	
1989	518	2002	500.50	495.75
		←	←	
1990	467	2027	506.75	503.62
		←	←	
1991	502	2066	516.50	511.62
		←	←	
1992	540	2170	542.50	529.50
		←	←	
1993	557	2254	563.50	553.00
		←	←	
1994	571	2326	581.50	572.50
		←	←	
1995	586	← —	—	—
1996	612	— —	—	—

2.8 GROWTH CURVES AND THEIR FITTING TO MODIFIED EXPONENTIAL

Q24. What is a growth curve? Write in detail about modified exponential curve with example.

Ans : **(Nov.-20, July-19)**

The graphical representation of information related to any kind of growth or development (like-population) as how quantity increases over a time period is known as growth curve. Mostly, growth curves are used as a statistical tool for determining the type of growth pattern of quantity. It could be either linear, exponential or cubic. Growth curves can be used for different purpose which may includes,

a) Growth of organizational sales or profit.

b) Growth of country's population.

c) Growth of bacterial population in atmosphere etc.

I) Modified Exponential Curve

The modified exponential curve is defined as "a number of asymptotic growth curve which recognize the population and employment of a region and also approach an lower or upper growth limit".

This equation can be equivalent to the geometric curve i.e.,y_c = abx with the help of an parameter 'c'.

Following figure shows the nature of modified exponential curve,

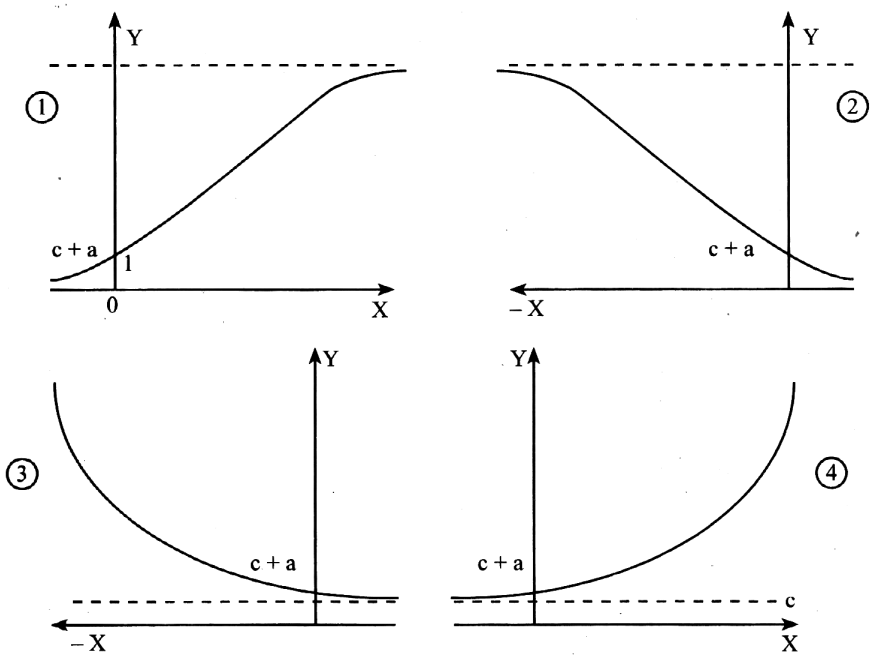


Figure: Forms of Modified Exponential Curve

From the above figure it can be seen or assumed that there are four different shapes which depends upon the parameters 'a' and 'b' In graph-1 parameter 'a' is negative and parameter 'b' is in between 'O' and 'l'. In this situation, the parameter 'c' becomes upper asymptotic. As a result v approaches as x increases i.e., which x increase the value of y_t becomes closer but does not equals to the asymptotic value of 'c'. Hence, this is one of the form of modified exponential curve. However, to explain in more detail, considering the following example of modified exponential curve.

Example

y_C — 115 — 65 $(0.75)^X$.

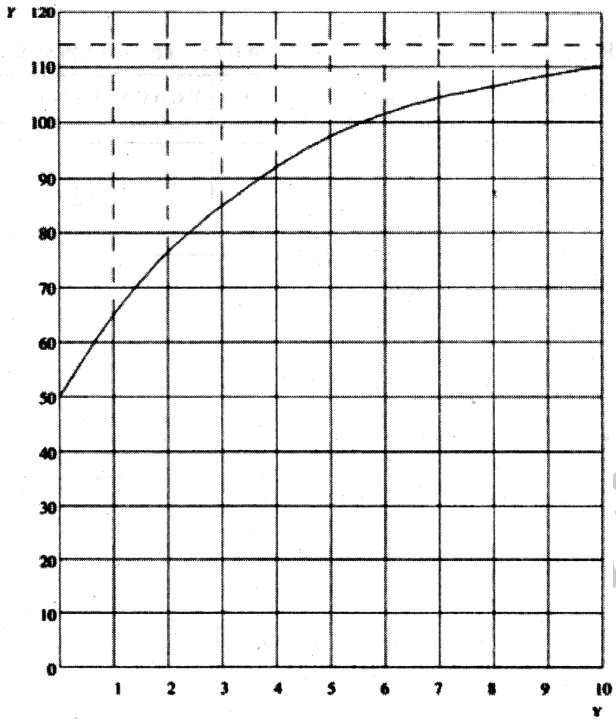
Here the parameter 'c' has the value of 115 which would be the upper limit through which approaches as x increases. Whereas, y intercept 50 is equal to the upper limit 115 plus parameter 'a' - 65. The parameter '6' (0.75) is equal to the ratio of successive growth increments. For instance,

i) Ifx = 0, y_c = 115 – 65 $(0.75)^o$ is equal to 50.

ii) Ifx = 1, y_c — 115 – 65 $(0.75)^1$ is equals to 66.25.

iii) If $x = 2$, $y_2 = 115 - 65\ (0.75)^2$ is equals to 78.4.

The growth increment for the first interval is 66.25 - 50 = 16.25. Whereas, the growth increment second interval is 78.4 – 66.25 = 12.15.



Graph

The ratio of the increments will be,

$$\therefore \quad \frac{12.15}{16.25} = 0.747 \text{ or } 0.75$$

Thus, the 0.75 ratio is equal to the parameter 'b'.

Procedure for Computing Growth Limits

The procedures used to calculate the growth limit 'c' and the a and b parameters for the modified exponential curve which could be the best fit to group of observations. In order to do this, these are two conditions which need to be satisfied. The total observations should be divisible by three and such observations can be odd or even.

The first observation should be assigned as a index value of zero and later observations should be assigned as continuous positive values (i.e., 1, 2, 3 ....).

After this, the parameters of the modified exponential curve can be determined by applying the following formula,

$$b^n = \frac{\sum_3 y - \sum_2 y}{\sum_2 y - \sum_1 y} \qquad \ldots(1)$$

$$a = (\sum_2 y - \sum_1 y)\left[\frac{b-1}{(b^x-1)^2}\right] \quad ...(2)$$

$$c = \frac{1}{n}\left[\frac{(\Sigma_1 y)(\Sigma_3 y)-(\Sigma_2 y)^2}{\Sigma_1 y + \Sigma_3 y - 2\Sigma_2 y}\right] \quad ... (3)$$

Where,

n = Number of observations in each partial sum.

$\Sigma_1 y$ = Partial sum of first third of observations.

$\Sigma_2 y$ = Partial sum of second third of observations.

$\Sigma_3 y$ = Partial sum of final third of observations.

Example :

Modified Exponential Curve Computation for Computing Upper Limit

Year (1)	Index value (x) (2)	Observed Values (y) (3) ('000')	Partial Sum of Observed Values ($\Sigma_n y$) (4) ('000')
1970	0	18	
1975	1	24	72
1980	2	30	
1985	3	36	
1990	4	44	133
1995	5	53	
2000	6	64	
2005	7	78	230
2010	8	88	

In the above table, the observations are assigned with applicable index values and the partial sums are computed for every third of the observations.

The amount of observations within every partial total, 'n' is adequate to the whole variety of observations, N, divided by 3 i.e.,

$$n = \frac{N}{3} = \frac{9}{3} = 3$$

The value of b^n should be first computed by applying Equation (1)

$$b^n = \frac{\Sigma_3 y - \Sigma_2 y}{\Sigma_2 y - \Sigma_1 y}$$

$$= \frac{230-133}{133-72}$$

$$= \frac{97}{61} = 1.590$$

Taking the n^{th} root of this value yields the 'b' parameter i.e.,

$b = (1.590)^{1/3} = 1.167$

The parameter 'a' can be computed by applying equation (2)

$$a = (\Sigma_2 y - \Sigma_1 y)\left[\frac{b-1}{(b^n-1)^2}\right]$$

$$= (133 - 72)\left[\frac{1.167-1}{(1.590-1)^2}\right]$$

$$= 61\left[\frac{1670}{3481}\right]$$

$$= 61\ [\ 0.480\ ] = 29.28$$

Parameter 'c' is computed by applying equation (3)

$$c = \frac{1}{x}\left[\frac{(\Sigma_1 y)(\Sigma_3 y)-(\Sigma_2 y)^2}{\Sigma_1 y + \Sigma_3 y - 2\Sigma_2 y}\right]$$

$$= \frac{1}{3}\left[\frac{(72)(230)-(133)^2}{72+230-2(133)}\right]$$

$$= -10.45$$

$\therefore \quad y_c = -10.45 + 29.28\ (1.167)^X$

Hence, the parameter 'a' is positive, parameter 'b' is greater than 1 and parameter 'c' is lower limit which indicates the curve approaches large negative values of x.

2.9 GROWTH CURVES AND THEIR FITTING TO GOMPERTZ AND LOGISTIC CURVES

Q25. Write about Gompertz curve and explain it with the help of figure.

Ans :

Gompertz Curve

Gompertz curve is developed by English actuary and mathematician "Benjamin Gompertz". It is expressed as, $Y = Z.\ x\ \exp(y^h)$ which can be converted to modified exponential form as $\log Y_c = \log z + (\log x)\ y^h$.

Gompertz curve equation is similar to modified exponential curve equation which is used for expecting various shapes based on the values of 'x' and 'y' parameters. The foremost interesting version is graph (x), in which the logarithm of the parameter (x) is negative and parameter 'y' is between zero and one. During this case, the Gompertz curve is 'S' shaped with associate higher limit of 'z' for big positive values of B and x lower limit of zero for big negative values of A i.e., If 'y' is less than 1, 'y^h' meets zero as 'A' increases. Through this the gompatz curve meets zx°, or the upper limit of 'z' for large positive values of A. If 'y' is less than 1, y^h develops quickly to increase larger negative values of A. Therefore, the curve meets zero for larger negative values of A.

In this situation, the Gompertz curve describes a development model which is very slow in initial stages but starts increasing after a period of time and then decreases so that dependent variable reach an upper limit. The upper limit may be assumed from the observed information to provide an autonomous audit on the Gompertz projections.

Linear Transformation for Assumed Growth Limits

In modified exponential curve logarithms can be used to convert the transformed gompertz curve, $\log\_v_c = \log z + \log x\ (y^h)$, into the linear form, $y_c = x + yA$. This enables the linear curve-fitting process to be used to fit a gompertz curve or a set of observation of data with imaginary upper or lower growth limits. The process is similar to modified exponential curve except that log z replace the term 'z', log y_c replaces the term Y^C and log x replace the term 'x'. For the situation in the graph (x) of figure: 1 following is the transformed gompertz curve for an expected upper limit,

$$\log (\log z - \log y_c) = \log (\log x) + \log y\ (A)$$

For the situation shown ip the graph (z) of figure-1 the transformed gompertz curve for an expected lower limit 'z' is as follows,

$$\log (\log y_c - \log z) = \log (\log x) + \log y\ (A).$$

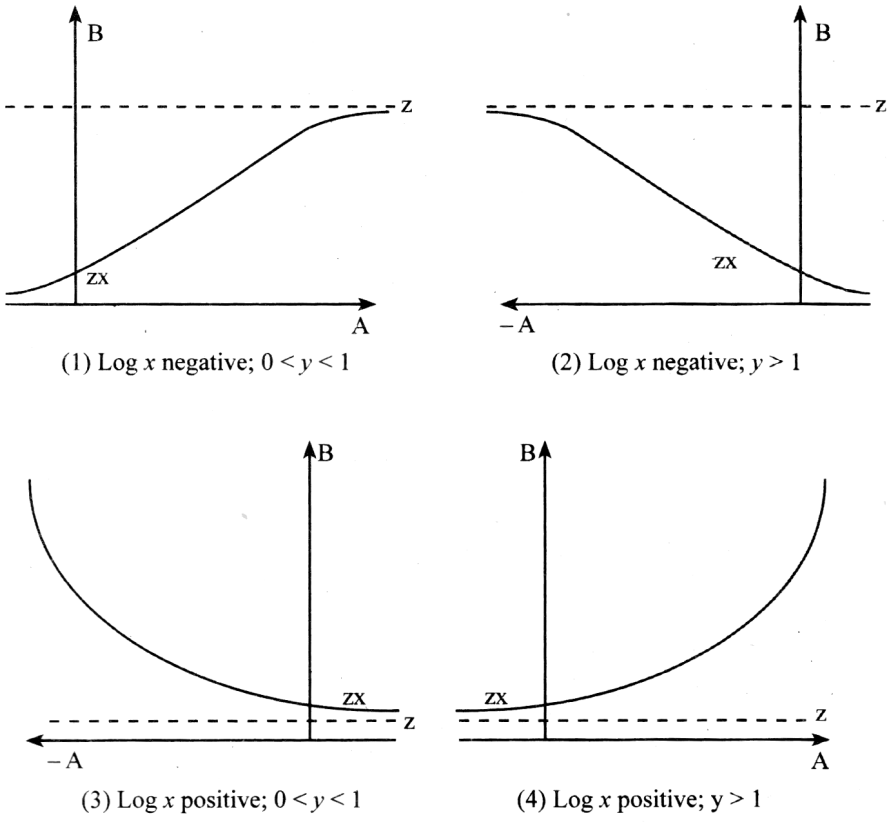


Figure: Alternative Forms of Gompertz Curve

Q26. What is logistic curve ? How it is prepared ?

Ans : **(July-19)**

Logistic Curve

Logistic curve was proposed by the Belgian mathematician P.F.Verhurst. It is the most popular equation for describing the growth of population. This curve was re-proposed by Raymond Pearl and Lowell J.Reed in the year 1920. As a result of this, it is also known as Pearl-Reed curve or pearl curve. The equation for the logistic curve is as follows,

$$\frac{1}{y_c} = z + xy^h \quad \text{i.e.,} \quad y_c^{-1} = z + xy^h$$

Or

$$y_c = \frac{1}{z + xy^h} \quad \text{i.e.,} \quad y_c = (z + xy^h)^{-1}$$

Preparation of Logistic Curve

The logistic curve is similar to the gompertz curve and modified exponential curve. However in logistic curve, the observed values of modified exponential curve are substituted or replaced by reciprocals of the observed values. As a result, the ratio of continuous increments growth for the reciprocals of y_c values in the logistic curve would be equal to a constant.

Logistic curves adopts various shapes based on the values of the parameters of a curve. Logistic curve comprises of two parameters namely *x* and *y,* which lies between zero and 1. In this circumstances, the logistic curve assumes "S" Shaped curve same as Gompertz curve. However in that log V is negative and *'y'* lies between zero and 1. For those values, the logistic curve meets an upper limit of 1/z. For high positive values of'X' and a lower limit of zero for high negative values of 'X'. Here, assume that A is equal to zero, than B deflects is equal to l/(z + *xy°)* or l/(z + x).

Following figure shows the logistic curve functioning,

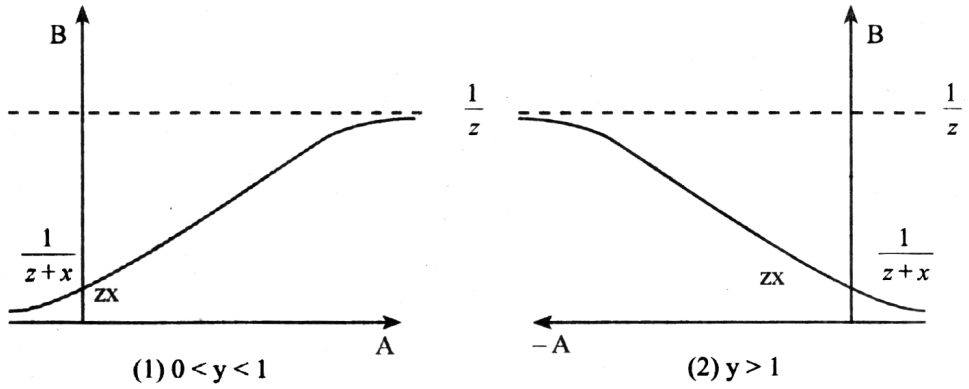


Fig. : Alternate Forms of the Logistic Curve

Thus, the logistic curve is an effective way of appealing in order to describe the phenomena which may have small initial growth increments, rapid growth and increasing at slower rate which continues till

the curve approaches upper limit. The studies conducted by Pearl and Reed explain that the logistic curve provides wide range of accurate phenomena which also includes the growth of albino rates, the tails of tadpoles and the various fruit flies in a bottle with some limited supply of food.

2.10 Determination of Seasonal Indices Ratio to Moving Average

Q27. What do you understand by seasonal variation ?

Ans : **(Dec.-18)**

Most of the phenomena in economics and business show seasonal pattern. "When data are expressed annually there is no seasonal variation. However, monthly or quarterly data frequently exhibit strong seasonal movements and considerable interest attaches to devising a pattern of *average seasonal variation.*

For example, if we observe the sales of a bookseller we find that for the quarter July-September (when most of the students purchase books), sales are maximum. If we know by how much the sales of this quarter are usually above or below the previous quarter for seasonal reasons, we shall be able to answer a very basic question, namely, was this due to an underlying upward tendency or simply because this quarter is usually seasonally higher than the previous quarter.

The method should be designed to meet the following criteria : '

1) It should measure only the seasonal forces in the data. It should not be influenced by the forces of trend or cycle that may be present.

2) It should modify the erratic fluctuations in the data with an acceptable system of averaging.

3) It should recognize slowly changing seasonal patterns that may be present and modify the index to keep up with those changes.

2.10.1 Ratio to Moving Average

Q28. Explain briefly about ratio to moving average method.

Ans :

The ratio-to-moving average method also known as the percentages of moving average method is the most widely used method of measuring seasonal variations. The steps necessary for determining seasonal pattern by this method are:

1. **Eliminate seasonality from the data by ironing it out of the original data.** Since seasonal variations recur every year—that is, since the fluctuations have a time span of 12 months—a centered 12-month moving average tends to eliminate these fluctuations (In case of quarterly data, a centered 4-quarter moving average must be used). The centered 12-month moving average which aims to eliminate seasonal and irregular fluctuations (S and I) represents the remaining elements of the original data, namely, trend and cycles. Thus, the centered 12-month moving average approximates T. C.
2. Express the original data for each month as a percentage of the centered 12-month moving average corresponding to it.
3. Divide each monthly item of the original data by the corresponding 12-month moving average, and list the quotients as 'Percent of Moving Average'. We have now succeeded in eliminating from the original data to a considerable extent the disturbing influences of trend and cycles. It remains to rid the data of irregular variations. By averaging these percentages for a given month (step 4) the irregular factors tend to cancel out and the average itself reflects the seasonal influence alone.

4. **The purpose of this step is to average, and—in the process of averaging— to eliminate the irregular factors.** We assume that the relatively high or extremely low values of seasonal relatives for any month are caused by irregular factors. The elimination of extremes may be achieved while we are averaging all Januarys, Februarys and the like. We do this by using an appropriate type of average. The median is appropriate since it is not affected by extremes. Thus, by using the median as an average we can obtain the typical seasonal relative for each month which will not be affected by irregular factors.

   Sometimes a so-called modified mean is used as an average for each month. Here, extreme values are omitted before the arithmetic mean is taken. In any array of seasonal relatives for each month, a value or several values on one end or both ends may be dropped and then the arithmetic mean of the remaining seasonal relatives is taken. A separate table is prepared in which the calculations involved in this step are shown. These means are preliminary seasonal indexes. They should average 100 per cent or total 1,200 for 12 months by definition.

5. If the total is not equal to 1,200 or 100 per cent, an adjustment is made to eliminate the discrepancy. The adjustment consists of multiplying average of each month obtained in step 4 by

$$\frac{1200}{\text{the total of the modified mean for 12 months}}$$

This adjustment is made not only to achieve accuracy, but also because when we come to eliminate seasonality from the original data we do not wish to raise or lower the level of the data unduly. Thus, if a seasonal index aggregates more than 1,200 (or averages more than 100) then the original data adjusted in terms of it will total less than the unadjusted original data. If it totals less than 1,200. The opposite would be true.

The logical reasoning behind this method follows from the Tact that 12-month moving average can be considered to represent 'the influence of cycle and trend CxT. If the actual value for any month is divided by the 12-month moving average centred to that month, presumably cycle and trend are removed. This may be represented by the following expression :

$$\frac{T \times S \times C \times I}{T \times C} = S \times I.$$

Thus the ratio to the moving average, from which this method gets its name, represents irregular and seasonal influences. If the ratios for each worked over a period of years are then averaged most random influences will usually be eliminated. Hence, in effect.

$$\frac{S \times I}{I} = S.$$

PROBLEM ON RATIO TO MOVING AVERAGE METHOD

9. Calculate seasonal indices by the ratio to moving average method in the following data:

Year	I Quarter	II Quarter	III Quarter	IV Quarter
1971	68	62	51	63
1972	65	58	66	61
1973	68	63	63	67

Sol. :

Year	Values	Moving Average	Moving Average	Centered Totals	Centered Moving Average	Adjusted Seasonal Variations
1971-Q1	68					
Q2	62					
		254	63.50			
Q3	61			126.25	63.125	61/63.125x100 = 96.63
		257	62.75			
Q4	63			124.50	62.250	101.20
1972-		247	61.75			
Q_1	65			124.75	62.375	104.21
		252	63.00			
Q_2	58			125.50	62.750	92.43
		250	62.50			
Q_3	66			125.75	62.875	104.97
		253	63.25			
Q_4	61			125.75	63.875	95.50
1973-		258	64.50			
Q_1	68			128.25	64.125	106.04
		255	63.75			
Q_2	63			129.00	64.500	97.67
		261	65.25			
Q_3	63					
Q_4	67					

Computation of Seasonal Indices

Year	Q1	Q2	Q3	Q4	Total
1971	–	–	96-63	101.20	
1972	104.21	92.43	104.97	95.20	
1973	106.40	97.67	–	–	
Total	210.25	190.10	201.60	196.70	
Average Adjusted	105.125	95.05	100.80	98.35	399.325
Seasonal Index : $\frac{400}{299.325} \times 105.125$					
	105.30	95.21	100.97	98.52	

Since most business and economic data has a multiplicative model, the 'Percentage to Moving Averages' method is most widely used. It eliminates both Trend and cyclical fluctuations from the time series. A major problem with this method is that there is massive data loss.

2.10.2 Ratio to Trend Method

Q29. What do you understand by ratio to trend ?

Ans : **(Dec.-18)**

This method of calculating a seasonal index (also known as the percentage-to-trend method) is relatively simple and yet an improvement over the method of simple averages explained in the preceding section. This method assumes that seasonal variation for a given month is constant fraction of trend. The ratio-to-trend method presumably isolates the seasonal factor in the following manner. Trend is eliminated when the ratios are computed, in effect.

$$\frac{T \times S \times C \times I}{T} = S \times C \times I.$$

Random elements are supposed to disappear when the ratios are averaged. A careful selection of the period of years used in the computation is expected to cause die influences of prosperity or depression to offset each other and thus remove the cycle. For series that are not subject to pronounced cyclical or random influences and for which trend can be computed accurately, this method may suffice. The steps in the com putation of seasonal index by this method are :

1. Trend values are obtained by applying the method of least squares.
2. The next step is to divide the original data month by month by the corresponding trend values and multiply these ratios by 100. The values so obtained are now free from trend and the problem that remains is to free them also of irregular and cyclical movements.
3. In order to free the values from irregular and cyclical movements the figures given for the various years for January, February, etc., are averaged with any one of the usual measures of central value, for instance, the *median* or the *mean.* If the data are examined month by month, it is sometimes possible to ascribe a definite cause to usually high or low values.

When such causes are found to be associated with irregular variations (extremely bad weather, an earthquake, famine and the like) they may be cast out and the mean of the remaining items is referred to as a *modified mean.* Since such scrutiny of the data requires considerable knowledge of prevailing conditions and is to a large extent subjective, it is often desirable to use the *median* which is generally not affected by very high or very low values.

4. The seasonal index for each month is expressed as a percentage of the average month. The sum of 12 values must equal 1,200 or 100 per cent. If it is not, an adjustment is made by multiplying each index by a suitable factor $\left(\frac{1200}{\text{the sum of the 12 values}}\right)$. This gives the final seasonal index.

PROBLEM ON RATIO TO TREND

17. Find seasonal variations by the 'ratio-to-trend' method from the data given below:

Year	1st Quarter	2nd Quarter	3rd Quarter	4th Quarter
1973	30	40	36	34
1974	34	52	50	44
1975	40	58	54	48
1976	54	76	68	62
1977	80	92	86	82

Sol :

Step -1

Yearly Total	Yearly Average
Total of 1973 = 30+40+36+34 = 140	140/4 = 35
Total of 1974 = 34+52+50+44 = 180	180/4 = 45
Total of 1975 = 40+58+54+48 = 200	200/4 = 50
Total of 1976 = 54+76+68+62 = 260	260/4 = 65
Total of 1977 = 80+92+86+82 = 340	340/4 = 85

Calculation of Trend by Least Squares Method

Year	Yearly Average	Deviation from Mid-year (x)	XY	X^2	Estimated Trend (Yc)
1973	35	-2	-70	4	56+12(-2) = 32
1974	45	-1	-45	1	56+12(-l) = 44
1975	50	0	0	0	56+12(0) = 56
1976	56	1	65	1	56+12(1) = 68
1977	85	2	120	4	56+12(2).80
	$\Sigma Y = 280$	$\Sigma x = 0$	$\Sigma XY = 120$	$\Sigma X^2 = 10$	

Therefore Trend Equation: Yx=56+26x, Yc values can now be calculated. Yearly increment = b=12, Hence, quarterly increment = 72/4 = 3. Calculation of Quarterly Trend Values: The first year 1973. The estimated trend value is 32. This represents the value for the middle of the year i.e., average of 2nd and 3rd. The value of second and third quarter is calculated as under.

$$\frac{Q^2 + Q^3}{2} = 32 \Rightarrow Q2 + Q3 = 64$$

Since quarterly increment is 3, $Q_3 = Q2 + 3$

Therefore $Q_2 + (Q_2 + 3) = 64 \Rightarrow 2\,Q_2 = 64\text{-}3 = 61$; Q,= 61.2 = 30.5

$Q_3 = Q_2 + 3 = 30.5 + 3 = 33.5$

Since quarterly increment is 3, value of $Q_1 = Q_2 - 3 = 27.5$

Value of $Q_4 = Q_a + 3 = 33.5 + 3 = 36.5$

The value of Q_1 of 1974 will be Q_4 of 1973 + 3 = 36.5 + 3 = 39.5

Similarly; all quarterly trend values can be calculated. The quarterly trend values are :

Year	Q_1	Q_2	Q_3	Q_4
1973	27.50	30.50	33.50	36.50
1974	39.50	42.50	45.50	48.50
1975	51.50	54.50	57.50	60.50
1976	63.50	66.50	69.50	72.50
1977	75.50	78.50	81.50	84.50

Step 2 :

The original data available to us has to be divided with the quarterly trend values obtained in step 1. For example, the first quarter of 1973 had an original figure of 30. The quarterly trend value obtained is 27.50. Thus, the 'trend eliminated' quarterly value is obtained by dividing 30 with 27.5 and multiplying it with 100. The resultant figure is 30 x 100/275 = 109.09. Similarly the second quarter would be 40 x 100/30.5 = 131.15.

Trend Eliminated Quarterly values as percentage of 'Trend' values.

Year	Q_1	Q_2	Q_3	Q_4
1973	109.09	131.15	107.46	93.15
1974	86.08	122.35	109.89	90.72
1975	77.67	106.42	93.91	79.34
1976	85.04	114.29	97.84	85.52
1977	105.96	117.20	105.52	97.04

Step 3 :

Total	463.84	591.41	514.63	445.77	
Average	92.77	118.28	102.93	89.15	403.12

Since the total of the averages is greater than 400, each average has to be multiplied with 400/403.12. Thus,

Adjusted Seasonal Index = (400/403.12) × 92.77

Index	92.05	117.36	102.13	88.46

The 'Ratio to trend' method eliminates the trend components from the time series values. Hence, it is more representative of the seasonal swings in the values. However, it does not eliminate the cyclical effects. Thus, the ratio to trend method is appropriate if the cyclical movements are either present, or if present, their effect is not very significant.

2.10.3 Link Relative Method

Q30. What do you understand by link relative method (Or) Pearson's Method ?

Ans : **(July-21, Dec.-18(MGU)**

Amongst all the methods of measuring seasonal variation, link relative 'method is the most difficult one. When this method is adopted the following steps are taken to calculate the seasonal variation indices:

1. Calculate the link relatives of the seasonal figures. Link relatives are calculated by dividing the figure of each season by the figure of the immediately preceding season and multiplying it by 100.

$$\frac{\text{Current season's figure}}{\text{Previous season's figure}} \times 100$$

These percentages are called link relative since they link each month (or quarter or other time period) to the preceding one.

2. Calculate the average of the link relatives for each season. While calculating average we might take arithmetic average but the median is probably better. The arithmetic average would give undue weight to extreme cases which were not due primarily to seasonal influences.

3. Convert these averages into chain relatives on the base of the first season.

4. Calculate the chain relatives of the first season on the base of the last season. There will be some difference between the chain relative of the first season and the chain relative calculated by the previous method. This difference will be due to the effect of long-term changes. It is, therefore, necessary to correct these chain relatives.

5. For correction, the chain relative of the first season calculated by the first method is deducted from the chain relative (of the first season) calculated by the second method. The difference is divided by the number of seasons. The resulting figure multiplied by 1, 2, 3 (and so on) is deducted respectively from the chain relatives of the 2nd, 3rd, 4th (and so on) seasons. These are correct chain relatives.

6. Express the corrected chain relatives as percentages of their averages. These provide the required seasonal indices by the method of link relatives. The following example will illustrate the process.

PROBLEM ON LINK RELATIVE METHOD

18. Apply the method of link relatives to the following data and calculate seasonal indices,

Quarterly figures					
Quarter	**2005**	**2006**	**2007**	**2008**	**2009**
1st	**6.0**	**5.4**	**6.8**	**7.2**	**6.6**
2nd	**6.5**	**7.9**	**6.5**	**5.8**	**7.3**
3rd	**7.8**	**8.4**	**9.3**	**7.5**	**8.0**
4th	**8.7**	**7.3**	**6.4**	**8.5**	**7.1**

Sol. :

Computation of Seasonal Indices by Link Relative Method

Year	1st	2nd	3rd	4th
2005	–	6.5/6.0×100 108.3	7.8/6.5 × 100 120	8.7/7.8 × 100 111.5
2006	5.4/8.7×100 62.1	7.9/5.4×100 146.3	8.4/7.9×100 106.3	7.3/8.4×100 86.9
2007	6.8/7.3 × 100 93.2	6.5/6.8 × 100 95.6	9.3/6.5 × 100 143.1	6.4/9.3×100 68.8
2008	7.2/6.4 × 100 112.5	5.8/7.2 × 100 80.6	7.5/5.8 × 100 129.3	8.5/7.5 x 100 113.3
2009	6.6/8.5 x 100 77.6	7.3/6.6 x 100 110.6	8.0/7.3 x 100 109.6	7.1/8.0 x 100 88.8
Arithmetic	$\frac{345.4}{4}$	$\frac{541.4}{5}$	$\frac{608.3}{5}$	$\frac{469.3}{5}$
Average	86.35	108.28	121.66	93.86
Chain Relatives	100	$\frac{100\times 08.28}{100}$ = 108.28	$\frac{121.66\times 108.28}{100}$ = 131.73	$\frac{93.86\times 131.73}{100}$ = 123.64
Corrected Chain Relatives	100	108.28 – 1.675 = 106.605	131.73 – 3.35 = 128.38	123.64 – 5.025 = 118.615
Seasonal Indices	$\frac{100}{113.4\times 100}$ = 88.18	$\frac{106.605}{113.4}\times 100$ = 94.01	$\frac{128.38}{113.4}\times 100$ = 113.21	$\frac{118.615}{113.4}\times 100$ = 104.60

Working Notes :

Chain relative of 1st quarter = 100

Chain relative of 1st quarter (on the basis of last quarter)

$$= \frac{86.35 \times 123.64}{100} = 106.7$$

$\therefore$ Difference between chain relatives = 106.7- 100 = 6.7

Difference per quarter = $\frac{6.7}{4} = 1.675$

Adjusted chain relatives are obtained,

1 x 1.675 = 1.675

2 x 1.675 = 3.35

3 x 1.675 = 5.025

Average of corrected chain relatives

$$= \frac{100 + 106.605 + 128.38 + 118.615}{4}$$

$$= \frac{453.6}{4} = 113.4$$

Seasonal Variation Index = $\frac{\text{Correct chain relatives}}{113.4} \times 100$

1st quarter = $\frac{100 \times 100}{113.4} = 88.18$.

Exercise Problems

1. Calculate the trend values by the method of least squares from the data given below and estimate the sales for the year 1985.

Year :	1976	1977	1978	1979	1980
Sales of T.V. ('000) :	9	12	14	16	20

(Ans: Yc = 20 + 5x, 44.5)

2. Fit a straight line trend by the method of least of least squares from the following data and estimate 1986 sales.

Year :	1979	1980	1981	1982	1983	1984	1985
Sales :	100	105	96	102	108	109	112

(Ans : Y=104.57 + 2x) Sales = 112.57

3. Fit a trend equation Y= a +bx:

Years	1997	1998	1999	2000	2001	2002	2003	2004
Production (M.T)	20	37	46	55	62	57	60	82

(Ans : 69.83 - 7.01x)

4. Calculate the Seasonal Index from the following data using the average method:

Year	Quarter I	Quarter II	Quarter III	Quarter IV
1992	72	68	80	70
1993	76	70	82	74
1994	74	66	84	80
1995	76	74	84	78
1996	78	74	86	82

(Ans: Q-I: 98.43, Q-II: 92-15, Q-III: 108.9, Q-IV: 100-52)

5. Calculate the seasonal index by the 'ratio to moving average' method from the following data :

Year	Quarter	Y	4 Quarterly Moving Average
1972	I	75	—
	II	60	—
	III	54	63.375
	IV	59	65.375
1973	I	86	67.125

	II	65	70.875
	III	63	74.000
	IV	80	75.375
1974	I	90	76.625
	II	72	77.625
	III	66	79.5000
	IV	85	81.5000
1975	I	90	76.625
	II	72	77.625
	III	72	—
	IV	93	—

(Ans: Q-I: 122.37; Q-II: 92.43; Q-III: 84.69 Q-IV: 100.51)

6. Calculate Seasonal Index by "ratio to Moving Average" method from the following data :

Year	QI	QII	QIII	QIV
1996	290	280	285	310
1997	320	305	310	330
1998	340	321	320	340
1999	370	360	362	380

(Ans : Q1 : 104.25, Q2 : 97.94, Q3 : 96.52, Q4 :101.29)

7. From the following relatives, find the link relatives:

Year	1998	1999	2000	2001	2002	2003
Price relatives	150	210	220	250	260	280

(Ans : 140, 104.76, 113.64, 104, 107.69)

Short Question and Answers

1. Stratified Random Sampling

Ans :

This process divides the population into homogenous groups or classes called 'strata'. A sample is taken from each group by simple random method and the resulting sample is called a stratified sample.

A stratified sample may be either proportionate or disproportionate. In a proportionate stratified sampling plan, the number of items drawn from each stratum is proportional to the size of the strata. While in a disproportionate stratified sampling, equal number of items are taken from each stratum irrespective of the size of the stratum.

Example

People interested in watching movies can be divided into groups or strata of people watching Hindi movies and English movies. After division, these groups can be selected randomly which is called stratified random sampling.

This method yields the most accurate results compared to other methods when population under study is heterogeneous. In this method population is divided into strata or subsets and random samples are chosen from each strata. This division of population into strata is based on the relevant similar characteristics within individual units. But the individual units present within each strata are similar.

2. Population Total

Ans :

Population total can be defined as the product of population mean and the number of sampling units in the population. It can be represented in stratified random sampling as,

$$N\bar{y}_{st} = N\,\bar{Y}_N$$

There are three estimations possible for population total while using stratified sampling. They are as follows,

i) The population total using an unbiased estimation procedure can be estimated as,

$$N\bar{y}_{st} = \sum_{i=1}^{k} N_i\,\bar{y}_i$$

Where N_i refers to the sampling unit within strata of population containing n_i samples.

ii) The variance estimator of population using an unbiased procedure can be obtained as,

$$N^2 Var(\bar{y}_{st}) = \sum_{i=1}^{k} N_i^2\ Var\ (\bar{y}_{n_i})$$

iii) If the size of sample is large, the estimation of confidence interval $100(1-\alpha)\%$ for population total using stratified random sampling can be obtained as,

$$N\bar{y}_{st} + z_{\alpha/2}\ NVar(\bar{y}_{st}).$$

3. Population Proportion

Ans :

Population proportion can be defined as sum of all the samples in the population divided by the total number of sampling units in the population. It can be represented as,

$$P = \frac{N_1P_1 + N_2P_2 + N_3P_3 + ... + N_kP_k}{N}$$

$$= \frac{1}{N}\sum_{i=1}^{k} N_iP_i$$

The estimations possible for population while using stratified random sampling are as follows,

i) If P represents the overall population proportion then its unbiased estimation obtains,

$$\hat{P}_{st} = \frac{1}{N}\sum_{i=1}^{k} N_iP_i$$

Where P_i represents the population proportion in i^{th} stratum.

ii) The estimator of population proportion while using an unbiased estimation procedure for variance is obtained as,

$$\hat{Var}(\hat{P}_{st}) = \frac{1}{N^2}\sum_{i=1}^{k} N_i^2 \, \hat{Var}(\hat{P}_i)^2$$

iii) If the size of sample is large, the estimation of confidence interval $100(1 - \alpha)\%$ for population proportion using stratified random sampling can be obtained as,

$$\hat{P}_{st} + Z_{\alpha/2}\hat{Var}(\hat{P}_{st}).$$

4. Define cost function.

Ans :

In stratified random sampling the cost function means a function which indicates the cost of the survey and is given by

$$C = a + \sum_{i=1}^{k} C_i n_i$$

Where,

a = Overhead cost and

C_i is the cost per unit in the i^{th} stratum.

C is the total survey cost.

5. Optimum allocation.

Ans :

The size of the samples to be drawn from the various strata is determined by the principle of optimization, i.e., obtaining best results at minimum possible cost. In optimum allocation, n_i s, (i = 1, 2, ..., k) are determined so that:

(i) Variance of sample estimate of the population mean is minimum (i.e., its precision is maximum) for fixed total sample size n. (Neynian's Allocation).

(ii) Variance of the estimate is minimum for a fixed cost of the plan.

(iii) Total cost of the sampling design is minimum for fixed desired precision, i.e., total cost is minimum for a fixed value of the variance of the sample estimate.

6. Advantages of Stratified Random Sampling

Ans :

i) More Representative Sample

A properly constructed and executed stratified random sampling plan overcomes the drawbacks of purposive sampling and random sampling and still enjoys the virtues of both these methods by dividing the given universe into a number of homogeneous subgroups with respect to purposive characteristic and then using the technique of random sampling in drawing samples from each stratum. A stratified random sample gives adequate representation to each strata or important section of the population and eliminates the possibility of any important group of the population being completely ignored. The stratified random sampling provides a more representative sample of the population and accordingly results in less variability as compared with other sampling designs.

ii) Greater Precision

As a consequence of the reduction in the variability within each stratum, stratified random sampling provides more efficient estimates as compared with simple random sampling. For instance, the sample estimate of the population mean is more efficient in both proportional and Neyman's allocation of the samples to different strata in stratified random sampling as compared with the corresponding estimate obtained in simple random sampling.

iii) Administrative Convenience

The division of the population into relatively homogeneous subgroups brings administrative convenience. Unlike random samples, the stratified samples are expected to be localised geographically. This ultimately results in reduction in cost and saving in time in terms of collection of the data, interviewing the respondents and supervision of the field work.

7. Disadvantages of Stratified Random Sampling

Ans :

i) As already pointed out the success of stratified random sampling depends on :

(a) Effective stratification of the universe into homogeneous strata and

(b) Appropriate size of the samples to be drawn from each of the stratum.

If stratification is faulty, the results will be biased. The error due to wrong stratification cannot be compensated even by taking large samples.

The allocation of the sample sizes to different strata requires an accurate knowledge of the population size in each stratum N_i, i = 1, 2, ..., k, [c.f., Proportional Allocation $n_i \propto N_i$]. Further Neyman's principle of optimum allocation, *viz*., $n_i \propto N_i S_i$, requires an additional knowledge of the variability or standard deviation of each strata. N_i and S_i (i = 1, 2, ..., k) are usually unknown and are a serious limitation to the effective use of stratified random sampling.

ii) Disproportional stratified sampling requires the assignment of weights to different strata and if the weights assigned are faulty, the resulting sample will not be representative and might give biased results.

8. Merits of Systematic Sampling

Ans :

i) Systematic sampling is very easy to operate and checking can also be done quickly. Accordingly, it results in considerable saving in time and labour relative to simple random sampling or stratified random sampling.

ii) Systematic sampling may be more efficient than simple random sampling provided the frame is complete and up-to-date and the units are arranged serially in a random order like the names in a telephone directory where the units are arranged in alphabetical order. However, even in alphabetical arrangement, certain amount of non-random character may persist.

9. Advantages of Simple Random Sampling

Ans :

Simple Random Sampling (SRS) offers the following advantages,

i) The samples obtained using SRS are more appropriate and representative than the samples obtained using judgement sampling.

ii) Simple random sampling is economical as it saves time, labour and money.

iii) Sampling errors can be measured.

iv) It is useful for finite as well as growing population.

v) Every item of the population has the same chance of being selected, hence it is more scientific and significant in biostatistical research.

10. Demerits of Systematic Sampling

Ans :

i) Systematic sampling works well only if the complete and up-to-date frame is available and if the units are randomly arranged. However, these requirements are not generally fulfilled.

ii) Systematic sampling gives biased results if there are periodic features in the frame and the sampling interval (k) is equal to or a multiple of the periods.

11. Define Time Series

Ans :

An arrangement of statistical data in accordance with time of occurrence or in a chronological order is called a time series. The numerical data which we get at different points of time-the set of observations is known as time series.

Examples :

i) The Annual Production of Steel in India over the last 10 years;

ii) The Monthly Sales of a Chemical Industry for the last 6 months;

iii) The daily closing price of a share in the Calcutta Stock Exchange;

iv) Hourly temperature recorded by the Meteorological office in a city;

v) Yearly Price or Quantity Index Numbers.

Mathematically, a Time Series is defined by the values Y_1, Y_2, —, Y_n of a variable Y at times t_1, t_2,....t_n

Here Y is a function of time t and Y, denotes the value of the variable Y at time t. The basic assumption is that changes witnessed over time in a sample group will be extrapolated to population.

i) According to Morris Hamburg, "A time series is a set of observations arranged in chronological order".

ii) According to Kenny and Keeping, "A set of data depending on the time is called time series".

iii) According to Croxton and Cowden, "A time series consists of data arranged chronologically".

iii) According to Patterson, "Time series consists of statistical data which are collected, recorded or observed over successive increments".

12. Utility of Time Series Analysis.

Ans :

i) It Helps in the Analysis of past Behaviour of a Variable

Analysis of past data discloses the effect of various factors on the variable under study. These studies isolate and analyze the effects of various sets of homogenous factors on the problem under study.

ii) It Helps in Forecasting :

The analysis of past condition is the basis of forecasting the future behaviour of the variable under study. For example, the analysis of a time series relating to income or wages or production would be the basis for forecasting future income or wages or production. This helps in making future plan of action.

iii) It helps in Evaluation of Current Achievement :

The review and evaluation of progress made on the basis of a plan are done on the basis of time series data. For example the progress of our Five-year plans is judged by the annual growth rates in the gross national product.

iv) It Helps in Making Comparative studies :

Once the data are arranged chronologically comparison between one time period and another is facilitated. It provides a scientific basis for making comparisons by studying and isolating the effects of various components of a time of series. It also helps in making regional comparison amongst data collected on the basis of time.

13. Secular Variations or Trend (T)

Ans :

The general tendency of the time series data to increase or decrease or stagnate during a long-period of time is called the secular trend or simply trend.

According to Simpson and Kafka: "Trend, also called secular or long-term trend, is the basic tendency of a series to grow or decline over a period of time. The concept of trend does not include short-range oscillations, but rather the steady movements over a long time".

The trend thus refers to the general direction and the movement of time series considering a fairly long period of time. Most business and economic series are influenced by some secular forces of charge in which the underlying tendency is one of growth or decline, for example, in series concerning population, national income, agricultural production, currency in circulation, bank deposits, etc., an upward tendency is usually observed while a downward tendency is noticed in time series relating 10 birth and death rates specially in deaths by epidemics, tuberculosis, etc.

14. Seasonal Variations (S)

Ans :

Seasonal variations refers to rhythmic forces of change inherent in most time series showing a regular or a periodic pattern of movement over a span of less than a year and has the same or almost the same pattern year after year.

According to Hirsch "a recurrent pattern of charge within the period that results from the operation of forces connected with climate or custom at different times of the period."

Reasons for Seasonal Variations

i) **Climate and Natural Forces:** The result of natural forces like climate is causing seasonal variation. Umbrellas are sold more in rainy season. In winter season, sale of the woolen clothes will increase. In hot season, the sales of ice, ice-cream, fruit salad, etc., will increase. Thus climate and weather play an important role in seasonal movement. Agricultural production depends upon the monsoon.

ii) **Customs and Habits:** Man-made conventions are the customs, habits, fashion, etc. There is the custom of wearing new clothes, preparing sweets and buying crackers for Deepavali, Onam, Christmas, etc. At that time, there is more demand for cloth, sweets and crackers. It will happen every year. In marriage season, the price of gold will increase.

15. Irregular Variations (I)

Ans :

Irregular variations are the effect of random factors. These are generally mixed up with seasonal and cyclical variations and are caused by irregular and accidental factors like floods, famines, wars, strikes, lockouts, etc. There is no regular period or time of their occurrence and that is why they are called random or chance fluctuations. Sometimes these factors are very effective and may even give rise to cyclical fluctuations. However since these are not regular factors, no advance preparation can be done to meet their consequences. Their effects are also unpredictable and irregular.

16. Uses of trend analysis

Ans :

Uses of Trend

1) The trend describes the basic growth tendency ignoring short term fluctuations.

2) It describes the pattern of behavior which has characterized the series in the past.

3) Future behavior can be forecasted in the assumption that past behavior will continue in the future also.

4) Trend analysis facilitate us to compare two or more time series over different period of time and this helps to draw conclusions about them,.

17. Merits of Least Squares

Ans :

The following are the merits of least squares method,

1. The method of least squares is a mathematical method of measuring trend and is free from subjectiveness.
2. This method provides the line of best fit since it is this line from where the sum of positive and negative deviations is zero and the sum of square of deviations is the least.
3. This method enables us to compute the trend values for all the given time periods in the series.
4. The trend equation can be used to estimate the values of the variable for any given time period 'f' in future and the forecasted values are quite reliable.
5. This method is the only technique which enables us to obtain the rate of growth per annum for yearly data in case of linear trend.

18. Demerits of Least Squares

Ans :

Some of the demerits of least squares are as follows,

1. Fresh calculations become necessary even if a single new observation is added.
2. Calculations required in this method are quite tedious and time consuming as compared with other methods.
3. Future predictions based on this method completely ignore the cyclical, seasonal and erratic fluctuations.
4. This method cannot be used to fit growth curves, gomper t_Z curve, logistic curve etc. to which most of the business and economic time series conform.

19. Moving Average Method

Ans :

In moving average method, the average value for a number of years (month or weeks) is secured and this average is taken as the normal or trend value for the unit of time falling at the middle of the period covered in the calculation of the average.

The effect of averaging is to give a smoother curve, lessening the influence of the fluctuations that pull the annual figures away from the general trend.

The period of moving average is decided in the light of the length of the cycle. More applicable to data with cyclical movements.

Formula for 3 yearly moving average will be,

$$\frac{a+b+c}{3}, \frac{b+c+d}{3}, \frac{c+d+e}{3} \ldots$$

Formula for 5 yearly moving average will be,

$$\frac{a+b+c+d+e}{5}, \frac{b+c+d+e+f}{5}, \frac{c+d+e+f+g}{5} \ldots.$$

20. Merits of Moving Average

Ans :

The merits of moving average are as follows,

1. Of all the mathematical methods of fitting a trend, this method is the simplest.«
2. The method is flexible so that even if a few more observations are to be added, the entire calculations are not changed.
3. If the period of the moving average happens to coincide with the period of the cycle, the cyclical fluctuations are automatically eliminated.
4. The shape of the curve in case of moving average method is determined by the data rather than the statisticians choice of mathematical function.

21. Limitations of Moving Average

Ans :

The following are the limitations of moving averages,

1. Trend values cannot be computed for all the years. For example, in a 5 yearly moving we cannot compute trend values for the first two and the last two years.
2. It is difficult to decide the period of moving average since there is no hard and fast rule for the purpose.
3. Moving average cannot be used in forecasting as it is not represented by any mathematical function.
4. When the trend is not linear, the moving average lies either above or below the true sweep of the data.

22. What is a growth curve ?

Ans :

The graphical representation of information related to any kind of growth or development (like-population) as how quantity increases over a time period is known as growth curve. Mostly, growth curves are used as a statistical tool for determining the type of growth pattern of quantity. It could be either linear, exponential or cubic. Growth curves can be used for different purpose which may includes,

a) Growth of organizational sales or profit.

b) Growth of country's population.

c) Growth of bacterial population in atmosphere etc.

23. Write about Gompertz curve.

Ans :

Gompertz curve is developed by English actuary and mathematician "Benjamin Gompertz". It is expressed as, $Y = Z. x \exp (y^h)$ which can be converted to modified exponential form as $\log Y_c = \log z + (\log x) y^h$.

Gompertz curve equation is similar to modified exponential curve equation which is used for expecting various shapes based on the values of 'x' and 'y' parameters. The foremost interesting version is graph (x), in which the logarithm of the parameter (x) is negative and parameter 'y' is between zero and one. During this case, the Gompertz curve is 'S' shaped with associate higher limit of 'z' for big positive values of B and x lower limit of zero for big negative values of A i.e., If 'y' is less than 1, 'y^h' meets zero as 'A' increases. Through this the gompatz curve meets zx°, or the upper limit of 'z' for large positive values of A. If 'y' is less than 1, y^h develops quickly to increase larger negative values of A. Therefore, the curve meets zero for larger negative values of A.

In this situation, the Gompertz curve describes a development model which is very slow in initial stages but starts increasing after a period of time and then decreases so that dependent variable reach an upper limit. The upper limit may be assumed from the observed information to provide an autonomous audit on the Gompertz projections.

24. Additive Model

Ans :

According to the additive mode, a time series can be expressed as

$$Y_t = T_t + S_t + C_t + R_t$$

where Y_i = is the time series value at time t,

T_t = represent the trend value.

S_t = represents the seasonal variations.

C_t = represents the cyclic movements and

R_t = represents the random fluctuations at time *t*.

Obviously, the term S_t will not appear in a series of annual data. The additive model implicitly implies that seasonal forces (in different years), cyclical forces (in different cycles) and irregular forces (in different long time periods) operate with equal absolute effect irrespective of the long trend value. As such C_t (and S_t) will haves positive or negative values according as whether we are in an above

normal or below normal phase of the cycle (and year) and the total of positive and negative values for any cycle (and any year) will be zero. R_t will also have positive or negative values and in the long run R_t will be zero. Occasionally, there may be a few isolated occurrences of extreme R_t of episodic nature.

25. Multiplicative Model

Ans :

On the other hand if we have reasons to assume that the various components in a time series operate proportionately to the general level of the series, the traditional or classical multiplicative model is appropriate. According to the multiplicative model

$$Y_t = T_t \times S_t \times C_t \times R_t$$

where S_t, C_t and R_t instead of assuming positive and negative values are indices fluctuating above or below unity and the geometric means of S_t in a year, C_t in a cycle and R_i in a long-term period are unity. In a time series with both positive and negative values the multiplicative model can not be applied unless the time series is translated by adding a suitable positive value. It may be pointed out that the multiplicative decomposition of a time series is same as the additive decomposition of the logarithmic values of the original time series.

26. Mixed Models

Ans :

In addition to the additive and multiplicative models discussed above, the components in a time series may be combined in large number of ways. The different models, defined under different assumptions, will yield different results. Some of the defined models can be:

$$Y_t = (T_t \times S_t \times C_t) + R_t$$

$$Y_t = (T_t \times C_t) + (S_t \times R_t)$$

$$Y_t = T_t + (S_t \times C_t \times R_t)$$

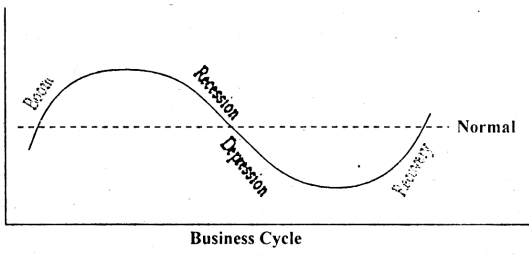
$$Y_t = T_t + S_t + (C_t \times R_t)$$

$$Y_t = T_t \times (S_t + C_t) \times R_t$$

27. Cyclical Variations

Ans :

As the economy expands during a period of boom, we would expect to find that such data as sales, output or consumer expenditure also show a rising trend; and during a period of slump, we would expect them to show a downward trend. Thus a wavelike motion may be observed in the pattern of our data. They are prosperity (boom) recession, depression and recovery; and are shown below:



The study of cyclical variation is helpful to businessmen and economist for framing suitable policies and stabilizing the level of business activities.

28. Explain about random fluctuations in time series data.

Ans :

Some kind of variation which are purely, random, unpredictable, accidental are called irregular variations. These variations do not exhibit any pattern and do not occur in regular period of time. These variations are due to floods, waves, earth quakes, strikes etc which cannot be predictable and beyond the controle of human hand.

Choose the Correct Answer

1. _____ means division into layers [a]
   (a) Stratification (b) Systematic
   (c) Simple (d) None

2. _____ of population is said to be strata. [b]
   (a) Individual (b) Group
   (c) Things (d) People

3. $Var(\bar{y}_{st})$ is _____ for fixed total size of the sample if $n_i \alpha N_i S_i$ [a]
   (a) Minimum (b) Maximum
   (c) Less (d) Greater

4. $n = \sum_{i=1}^{k} n_i$, total sample _____ from all the strata [c]
   (a) Cost (b) Function
   (c) Size (d) None

5. S_1^2 is the population mean square of the _____ stratum [c]
   (a) j^{th} (b) k^{th}
   (c) i^{th} (d) None

6. $\bar{y}_{n_i}$ is the _____ of sample selected from i^{th} stratum [a]
   (a) Mean (b) Variance
   (c) Median (d) None

7. Sample mean is an unbiased estimator of _____ mean. [b]
   (a) Sample (b) Population
   (c) a & b (d) None

8. $\bar{y}_{..}$ is the _____ mean. [a]
   (a) Population (b) Sample
   (c) Mean (d) Median

9. _____ is the inter class correlation coefficient. [d]
   (a) r (b) q
   (c) p (d) ρ

10. K is the _____ [b]
   (a) Systematic (b) Strata
   (c) a & b (d) None

11. A trend is said to be positive if both the variable in _____ direction. [a]
(a) Same (b) Increasing
(c) Decreasing (d) b & c

12. Where _____ are classified into climate and man mode. [c]
(a) Climate (b) Periodic
(c) Seasons (d) None

13. An arrangement of statistical data in accordance with time of occurrence or in a chronological order is called _____ [a]
(a) time series (b) trend value
(c) secular trend (d) least square

14. Moving averages gives the correct _____ of the long term trend of the series. [d]
(a) Item (b) Product
(c) Trend (d) Picture

15. _____ curve cannot be determined by the principle of least square. [c]
(a) Exponential (b) Straight line
(c) Logistic (d) Parabola

16. $y = a \cdot b^{c^t}$ is called _____ curve [a]
(a) Gompertz (b) Exponential
(c) Logistic (d) None

17. $y_t = a + bc^t$ is called _____ curve [a]
(a) Modified exponential (b) Logistic
(c) Exponential (d) Gompertz

18. $y_t = ab^t$ is called _____ curve [c]
(a) Logistic (b) Gompertz
(c) Exponential (d) None

19. A trend is said to be non-linear if the variables in time series are related _____ positive (or) negative. [a]
(a) nor (b) or
(c) and (d) All the three

20. A trend is said to be negative if both the variables in the _____ direction. [b]
(a) Positive (b) Negative
(c) a & b (d) Increasing

Fill in the blanks

1. Stratification means division into _____.
2. The process of selecting the sample from each group of the population is said to be _____.
3. Group of population is said to be _____.
4. $var(\bar{y}_{st})$ is minimum for fixed total size of the sample if $n_i \propto$ _____.
5. The process of selecting the sample systematically from the population said to be _____.
6. $var(\bar{y}_{st})$ _____ $var(\bar{y}_{sys}) \leq var(\bar{y}_n)_R$.
7. $var(\bar{y}_{sys})$ = _____.
8. Each sample selecting from strata randomly without giving any _____.
9. _____ means a function which indicates the cost of the survey.
10. C is the total survey _____.
11. _____ gives an idea of the establishing the organisation.
12. Different factors effecting the value of the variable is known as _____ of ime series.
13. The value of the variable either increasing (or) decreasing during the long period of time is known as _____ .
14. A trend is said to be _____ if both the variables in same direction.
15. A trend is said to be negative if both the variables in _____ direction.
16. The straight line trend is _____ .
17. The sum of the squares of the deviation of the observation and its _____ value should be minimum.
18. The Gompertz curve is _____
19. Periodic variation effecting the value of the variable by seasons is known as _____ .
20. In ratio to trend method, trend eliminated values of IQR = _____

ANSWERS

1. Layers
2. Stratified random sampling
3. Strata
4. $N_i S_i$
5. Systematic sample
6. $\leq$
7. $\frac{nk-1}{nk}\frac{S^2}{n}\{(1+(n-1)\rho)\}$
8. Important
9. Cost function
10. Cost
11. Time series
12. Components
13. Secular trend
14. Positive
15. Opposite
16. $y_t = a + bt$
17. Expected
18. $y_t = a \cdot b^{c^t}$
19. Seasonal variation
20. $\frac{\text{Original data}}{\text{Trend value}} \times 100$

One Mark Answers

1. Define stratified random sample.

Ans :

The process of selecting the sample from each group of the population is said to be stratified random sample.

2. Define strata.

Ans :

Group of population is said to be strata.

3. Define stratification.

Ans :

Stratification means division into layers.

4. Define cost function.

Ans :

Cost function means a function which indicates the cost of the survey.

5. Define systematic sampling.

Ans :

The process of selecting the sample systematically from the population said to be systematic sample.

6. Define positive trend ?

Ans :

A trend is said to be positive if both the variables time series are related in same direction.

7. Define secular trend ?

Ans :

The value of the variable either increasing (or) decreasing during the long period of time is known as secular trend.

8. Define Gompertz curve ?

Ans :

The Gompertz curve is given by $y = a \cdot b^{c^t}$

9. Define seasonal variation ?

Ans :

Periodic variation effecting the values of variable by seasons is known as seasonal variation.

10. Write the formula for link relative method.

Ans :

$$L.R = \frac{\text{Current month figure}}{\text{Previous month figure}} \times 100$$

UNIT III

Demand Analysis: Introduction. Demand and supply, price elasticity of supply and demand. Methods of determining demand and supply curves, Leontief's ,Pigous's methods of determining demand curve from time series data, limitations of these methods Pigou's method from time series data. Pareto law of income distribution curves of concentration.
Index Numbers: Concept, construction, uses and limitations of simple and weighted index numbers. Laspeyer's, Paasche's and Fisher's index numbers, criterion of a good index numbers, problems involved in the construction of index numbers. Fisher's index as an ideal index number. Fixed and chain base index numbers. Cost of living index numbers and wholesale price index numbers. Base shifting, splicing and deflation of index numbers

3.1 DEMAND ANALYSIS - INTRODUCTION

Q1. Define Demand? Explain the features of Demand.

Ans:

Introduction

It is necessary to estimate the demand for the goods or services before they are produced and provided. The producers, for this purpose, heavily depend upon the data relating to the pattern of consumption of these goods and services. The demand analysis provides them the basis to take decisions relating to volume of production (How many products required to produce), capital to be invested (How much amount to be invested) and so on.

Demand

Demand for a commodity refers to the quantity of the commodity which an individual consumer is willing to purchase at a particular time at a particular price.

A product or service is said to have demand when three conditions are satisfied.

(a) Desire to acquire - Desire of the consumer to buy the + Product

(b) Willingness to pay - His willingness to buy the product and

(c) Ability to pay - Ability to pay the specified price for it.

Definitions

(i) According to Ferguson, Demand refers to the quantities of commodity that the consumers are able to buy at each possible price during a given period of time, other things being equal.

(ii) According toB.R.Schiller, Demand is the ability and the willingness to buy a specific quantity of a good at alternatives prices in a given time period

Features of Demand

The various features of demand are:

(**a) Difference between Desire and Demand:** Demand is the amount of commodity for which a consumer has the willingness and the ability to buy .These is difference between need and demand. Demand is not only the need, it also implies that the consumer has the money to purchase it.

(b) Relationship between Demand and price: Demand is always at a price .unless price is stated, the amount demanded has no meaning. The consumer must know both the price and the commodity and he will tell his amount demanded.

(c) Demanded at a point of time: The amount demanded must refer to some period of time such as 10 quintals of wheat per year or six shirts per year of five kilos of sugar per month .not only this, the amount demanded and the price must refer to a particular data.

Q2. Explain Demand Function (or) State Demand function ?

Ans :

Demand function is a function which describes a relationship between one variable and its determinants, describes how much quantity of goods is bought at alternative prices of good and related goods, alternative income levels, and alternative values of other variables affecting demand. Thus, the demand function for a good relates the quantity of a good which consumers demand during a given period to the factors which influence the demand. The above factors can be built up into a demand function. Mathematically, the demand function for a product A can be expressed as follows:

$$Qd = f(P, I, T, P_R, E_P, E_I, S_P, D_C, A, O)$$

Where

Qd	-	refers to quantity of demand and it is a function of the following variables
P	-	refers to price of the product
I	-	refers to Income level of the consumer
T	-	refers to tastes and preferences of the consumer
P_R	-	refers to prices of related goods (substitutes/complementary)
E_P	-	refers to expectations about the prices in future
E_I	-	refers to expectations about the incomes in future
S_P	-	refers to size of population
D_c	-	refers to distribution of consumers over different regions
A	-	refers to advertising efforts and
O	-	refers to any other factors capable of affecting the demand .

Q3. What is demand schedule ? Explain the characteristics of demand schedule.

Ans:

A demand schedule is a tabular representation of the relationship between the amount demanded of a commodity and different price levels of that commodity. In other words, a demand schedule is a tabular statement of price and quantity relationship. It relates to the amount of the commodity the consumer is willing to purchase corresponding to the given price of that commodity per unit of time.

The table below is an example of a demand schedule of product x.

Price of the Commodity x(`)	Quantity Demanded of Commodity x (kg)
5	15
8	14
10	12
12	10
15	8
20	5

Characteristics of Demand Schedule

The following are the characteristics of a demand schedule,

- A demand schedule shows variations in demand of in commodity at its varying prices.
- It indicates the behaviour of an individual consumer in purchasing the commodity at alternative prices.
- It shows the inverse relation between the quantity demanded and the price of the commodity.

3.1.1 Law of Demand

Q4. Define Law of Demand ?

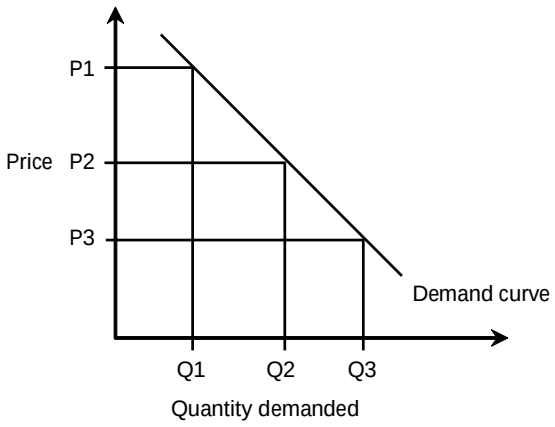
Ans :

Definition

The law of demand states that other factors being constant (cetris peribus), price and quantity demand of any good and service are inversely related to each other. When the price of a product increases, the demand for the same product will fall.

Description

Law of demand explains consumer choice behavior when the price changes. In the market, assuming other factors affecting demand being constant, when the price of a good rises, it leads to a fall in the demand of that good. This is the natural consumer choice behavior. This happens because a consumer hesitates to spend more for the good with the fear of going out of cash.



The above diagram shows the demand curve which is downward sloping. Clearly when the price of the commodity increases from price p3 to p2, then its quantity demand comes down from Q3 to Q2 and then to Q3 and vice versa.

Q5. Explain the assumptions of law of demand.

Ans :

The statement of the law of demand, demonstrates that that this law operates only when all other things remain constant. These are then the assumptions of the law of demand. We can state the assumptions of the law of demand as follows:

1. Income Level Should Remain Constant

The law of demand operates only when the income level of the buyer remains constant. If the income rises while the price of the commodity does not fall, it is quite likely that the demand may increase. Therefore, stability in income is an essential condition for the operation of the law of demand.

2. Tastes Of The Buyer Should Not Alter

Any alteration that takes place in the taste of the consumers will in all probability thwart the working of the law of demand. It often happens that when tastes or fashions change people revise their preferences. As a consequence, the demand for the commodity which goes down the preference scale of the consumers declines even though its price does not change.

3. Prices Of Other Goods Should Remain Constant

Changes in the prices of other goods often impinge on the demand for a particular commodity. If prices of commodities for which demand is inelastic rise, the demand for a commodity other than these in all probability will decline even though there may not be any change in its price. Therefore, for the law of demand to operate it is imperative that prices of other goods do not change.

4. No New Substitutes For The Commodity

If some new substitutes for a commodity appear in the market, its demand generally declines. This is quite natural, because with the availability of new substitutes some buyers will be attracted towards new products and the demand for the older product will fall even though price remains unchanged. Hence, the law of demand operates only when the market for a commodity is not threatened by new substitutes.

5. Price Rise In Future Should Not Be Expected

If the buyers of a commodity expect that its price will rise in future they raise its demand in response to an initial price rise. This behavior of buyers violates the law of demand. Therefore, for the operation of the law of demand it is necessary that there must not be any expectations of price rise in the future.

6. Advertising Expenditure Should Remain The Same

If the advertising expenditure of a firm increases, the consumers may be tempted to buy more of its product. Therefore, the advertising expenditure on the good under consideration is taken to be constant.

Q6. Explain the exceptions to the Law of Demand.

Ans :

The law of demand does not apply in every case and situation. The circumstances when the law of demand becomes ineffective are known as exceptions of the law. Some of these important exceptions are as under.

1. Giffen Goods

Some special varieties of inferior goods are termed as Giffen goods. Cheaper varieties of this category like bajra, cheaper vegetable like potato come under this category. Sir Robert Giffen or Ireland first observed that people

used to spend more their income on inferior goods like potato and less of their income on meat. But potatoes constitute their staple food. When the price of potato increased, after purchasing potato they did not have so many surpluses to buy meat. So the rise in price of potato compelled people to buy more potato and thus raised the demand for potato. This is against the law of demand. This is also known as Giffen paradox.

2. Conspicuous Consumption

This exception to the law of demand is associated with the doctrine propounded by Thorsten Veblen. A few goods like diamonds etc are purchased by the rich and wealthy sections of the society. The prices of these goods are so high that they are beyond the reach of the common man. The higher the price of the diamond the higher the prestige value of it. So when price of these goods falls, the consumers think that the prestige value of these goods comes down. So quantity demanded of these goods falls with fall in their price. So the law of demand does not hold good here.

3. Conspicuous Necessities

Certain things become the necessities of modern life. So we have to purchase them despite their high price. The demand for T.V. sets, automobiles and refrigerators etc. has not gone down in spite of the increase in their price. These things have become the symbol of status. So they are purchased despite their rising price. These can be termed as "U" sector goods.

4. Ignorance

A consumer's ignorance is another factor that at times induces him to purchase more of the commodity at a higher price. This is especially so when the consumer is haunted by the phobia that a high-priced commodity is better in quality than a low-priced one.

5. Emergencies

Emergencies like war, famine etc. negate the operation of the law of demand. At such times, households behave in an abnormal way. Households accentuate scarcities and induce further price rises by making increased purchases even at higher prices during such periods. During depression, on the other hand, no fall in price is a sufficient inducement for consumers to demand more.

6. Future Changes in Prices

Households also act speculators. When the prices are rising households tend to purchase large quantities of the commodity out of the apprehension that prices may still go up. When prices are expected to fall further, they wait to buy goods in future at still lower prices. So quantity demanded falls when prices are falling.

7. Change in Fashion

A change in fashion and tastes affects the market for a commodity. When a broad toe shoe replaces a narrow toe, no amount of reduction in the price of the latter is sufficient to clear the stocks. Broad toe on the other hand, will have more customers even though its price may be going up. The law of demand becomes ineffective.

3.2 METHODS OF DETERMINING DEMAND

Q7. Explain the various methods of determing demand.

Ans: **(July-21)**

There are different types of demand, they are as follows,

1. Price Demand

Price demand refers to the various quantities of a commodity purchased by a consumer at different prices in a given market. In this context, it is assumed that price of the commodity is the only one determinant of demand and the other determinants like consumer's income, individual tastes and preferences, prices of the related goods etc., remain constant. The relationship between price of the commodity and quantity demanded is given by,

$$D_x = f(P_x)$$

Where,

D_x - Demand for the commodity x

P_x - Price of the commodity x

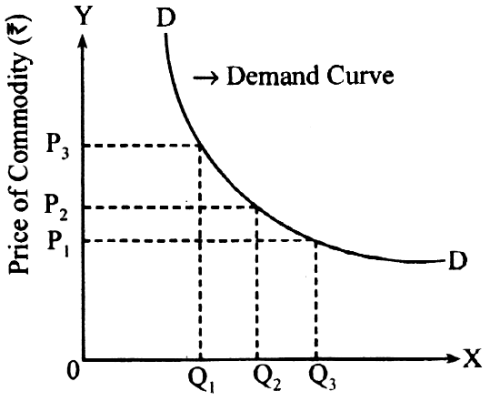


Fig. Quantity Demanded of the Commodity

2. Income Demand

Income demand refers to the various quantities of a good or a commodity that a consumer purchases at different levels of his income. In this context, the determinants of demand other than consumer's income are assumed to be constant. The income demand is given by,

$$D = f(I)$$

Where,

D = Demand for the commodity

I = Consumer's income.

The income effect varies for different goods, which are as follows,

(i) Normal Goods

For normal goods, the demand increases as the income increases and the demand decreases as the income level decreases. The demand curve for normal goods is as follows,

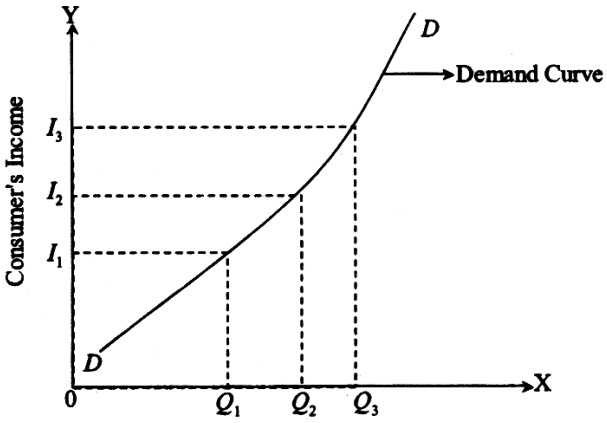


Fig : Quantity Demanded of the Commodity

The income demand curve for normal goods has a positive slope.

(ii) Inferior Goods

For inferior goods, the demand increases as income decreases and the demand decreases as the income increases. The demand curve for inferior goods is as follows,

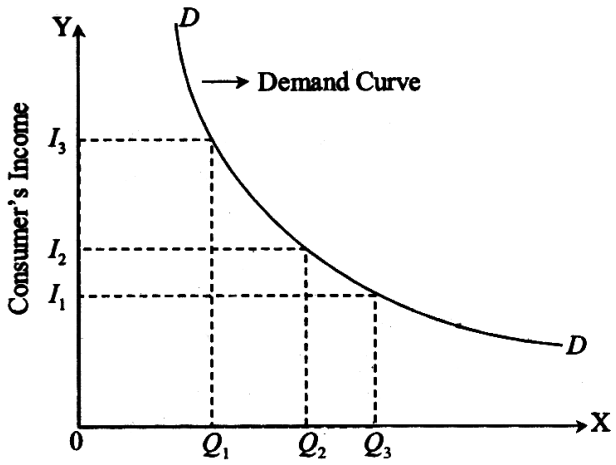


Fig : Quantity Demanded of the Commodity

The income demand curve for inferior goods has negative slope.

3. Cross Demand

Cross demand refers to the changes in the quantity demanded of a commodity with respect to the changes in the prices of the related goods (substitutes and comple - mentary goods). In this context, the determinants other than the prices of the related goods are assumed to be constant. Cross demand is expressed as,

$$D = f(P_1, P_2, ...,P_n)$$

Where,

D = Demand for the commodity

$P_1, P_2,.....,P_n$ = Prices of the related goods.

The cross demand is different for substitutes and complementary goods.

(i) Substitutes

When the price of a given commodity increases, the demand for its substitutes increases and vice-versa. For example, a decrease in the price of coffee decreases the demand for tea. The demand curve for substitute goods is as follows,

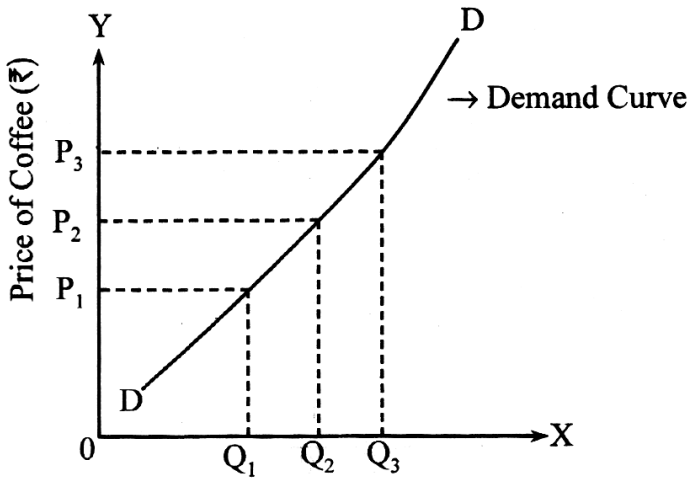


Fig. Quantity Demanded of Tea

The demand curve for substitute goods slopes upwards to right and has a positive slope.

(ii) Complementary Goods

When the price of a given commodity increases, the quantity demanded of its complementary goods decreases and vice-versa. For example, milk and sugar are complementary goods. An increase in the price of sugar decreases the demand for milk. The demand curve for complementary goods is as follows,

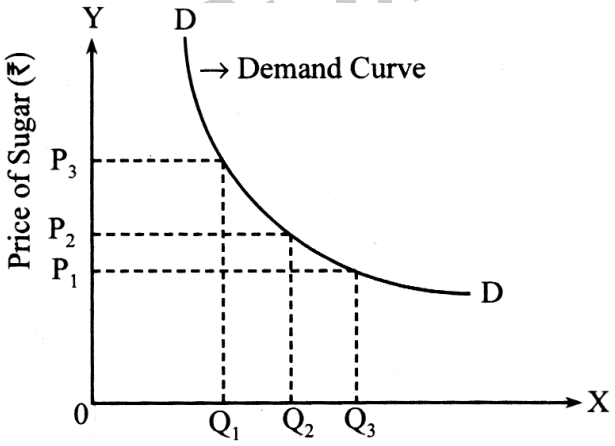


Fig. Quantity Demanded of Milk

The cross demand curve for complementary goods slopes downwards to the right and has negative slope.

4. Autonomous/Direct Demand and Induced/Derived Demand

Autonomous demand is a spontaneous (natural) demand with an urge to satisfy some wants directly. In other words, it is a demand where a commodity is demanded because it is needed for direct consumption. Autonomous demand is also called as direct demand because the goods are demanded directly from the natural urge to satisfy needs. The demand for consumer goods is autonomous. Autonomous demand is direct and it is the final demand.

5. Company/Firm Demand and Industry Demand

Company demand refers to the demand for a particular product to be produced by a particular firm. Example: Pens produced by cello. When the demand faced by all the companies producing a particular kind of product is added together, we get the industry demand. In other words, industry demand is the total demand for a product produced by a particular industry.

6. Total Market and Market Segment Demand

The demand for a product can be identified for the total market or for a given market segment. For example, for manipulating the product prices, promoting the product and improving it, increasing product competitiveness, a market segment is considered, whereas for sales forecasting the total market is considered. Demand for a market segment is obtained by breaking down the total demand into different segments like geographical areas, sub-products, product use, distribution channels, size of customer group etc. The market segments are also divided such that, each market segment has its own demand characteristics and each market segment has separate significance in price delivery, net profit margins, competition. number of substitutes etc.

7. Individual Demand

It is the demand or reaction of a single customer or household towards a product at different price levels.

8. Market Demand

It is the demand or reaction of multiple customers or household towards particular product at different price levels.

9. Joint Demand and Composite Demand

When goods are demanded jointly with one another in order to satisfy a single want is called as Joint Demand. For example: Vehicle, Petrol, Pen, Ink etc on other hand, A commodity is said to be in composite demand when it is wanted for several different uses. For example : Steel is needed for manu- facturing cars, buildings, construction of railway etc.

10. Demand for Consumers Goods and Producer's Goods

When goods are demanded by consumers for the direct satisfaction of their wants, they are called demand for consumer goods. For example: Food Items, Readymade Clothes etc on the other hand, When goods are demanded by producers for production of other goods including consumer goods, they are called demand for producers goods. For example : Machines, Tools, Equipment etc.

11. Demand for Perishable Goods and Durable Goods

Perishable goods are those which can one consumed only once, while durable goods are those which can be used more than once over a period of time. The examples of Perishable goods are Bread, Sweet, Milk and the example of Durable goods are T.V, Refrigerator, Air Cooler and Washing Machine etc.

12. Short Run and Long Run Demand

Short run demand refers to that demand which changes immediately due to reaction in price changes and income fluctuation. Whereas, Long run demand refers to that demand which does not react immediately due to price change. It will take some time for change in demand.

Q8. Explain the factors Determining of Demand.

Ans :

1. Price of the Given Commodity

Other things remaining constant, the rise in price of the commodity, the demand for the commodity contracts, and with the fall in price, its demand increases.

2. Price of Related Goods

Demand for the given commodity is affected by price of the related goods, which is called cross price demand.

3. Income of the Individual Consumer

Change in consumer's level of income also influences their demand for different commodities. Normally, the demand for certain goods increase with the increasing level of income and vice versa.

4. Tastes and Preferences

The taste and preferences of individuals also determine the demand made for certain goods and services. Factors such as climate, fashion, advertisement, innovation, etc. affect the taste and preference of the consumers.

5. Expectation of Change in Price in the Future

If the price of the commodity is expected to rise in the future, the consumer will be willing to purchase more of the commodity at the existing price. However, if the future price is expected to fall, the demand for that commodity decreases at present.

6. Size and Composition of Population

The market demand for a commodity increases with the increase in the size and composition of the total population. For instance, with the increase in total population size, there is an increase in the number of buyers. Likewise, with an increase in the male composition of the population, the demand for goods meant for male increases.

7. Season and Weather

The market demand for a certain commodity is also affected by the current weather conditions. For instance, the demand for cold beverages increases during summer season.

8. Distribution of Income

In case of equal distribution of income in the economy, the market demand for a commodity remains less. With an increase in the unequal distribution of income, the demand for certain goods increase as most people will have the ability to buy certain goods and commodities, especially luxury goods.

3.3 Elasticity of Demand

Q9. Define Elasticity of Demand?

Ans :

The law of demand simply explains the inverse relationship between price and quantity demanded. It doesn't specify how much more is purchased when price falls and how much less is purchased when price rises. In order to understand the rate of change in price and consequent changes in demand, elasticity of demand concept is used.

Elasticity is one of the most important concepts in necoclassical economic theory. It is useful in understanding the incidence of indirect taxation, marginal concepts as they relate to the theory of the firm and distribution of wealth and different types of goods. Elasticity is also crucially important in any discussion of welfare distribution, in particular consumer surplus, producer surplus or government surplus.

Meaning of Elasticity of Demand

Elasticity of demand is the responsiveness of demand for a commodity to changes in its determinants.

$$\text{Elasticity of Demand} = \frac{\text{Percentage change in quantity demanded of commodity}}{\text{Percentage change in its price}}$$

Definitions

(i) According to in the words of Dr.Marshall, "Elasticity of Demand may be defined as the percentage change in the quantity demanded divided by the percentage change in the price."

(ii) According to Boulding, "Price elasticity of demand measures the responsiveness of the quantity demanded to the change in price."

(iii) According to in the words of Dooley, "The price elasticity of demand measures the responsiveness of the quantity demanded to a change in its price."

(iv) According to Antol Murad, "Elasticity of demand is the ratio of relative change in quantity to relative change in price."

Thus, price elasticity of demand is a device to measure the rate of change in the quantity of a product demanded in response to a small change in its price.

Q10. Explain different types of elasticity of demand.

Ans :

There are 4 types of Elasticity of Demand

1. Price Elasticity of Demand
2. Income Elasticity of Demand
3. Cross Elasticity of Demand
4. Advertisement Elasticity of Demand

Q11. Explain different types of price elasticity demand.

Ans :

Elasticity of demand in general refers to price elasticity of demand. In other words, it refers to the quantity demanded of a commodity in response to a given change in price. Price elasticity is always negative which indicates that the customer tends to buy more with every fall in the price. The relationship between the price and the demand is inverse.

It is measured as follows:

$$\text{Price elasticity of demand} = \frac{\text{Proportionate change in the quantity demanded for product X}}{\text{Proportionate change in the price of X}}$$

The same is expressed as

$$\text{Edp} = \frac{(Q_2 - Q_1)/Q_1}{(P_2 - P_1)/P_1}$$

Where Q_1 is the quantity demanded before price change, Q_2 is quantity demanded after price change, P_1 is the price before change and P_2 is the price after change.

The price is said to be elastic, when the proportionate change in quantity demanded is more than the proportionate change in price. For instance, a 5 percent fall in the price results in an increase of 20 percent in the quantity demanded, the price is said to be elastic, which implies that the elasticity is more than one ($e > 1$)

Types of Price Elasticity of Demands :

a) Perfectly Elastic

b) Perfectly Inelastic

c) Relatively Elastic

d) Relatively Inelastic

e) Unit Elasticity

(a) Perfectly Elastic Demand

When any quantity can be sold at given price, and when there is no need to reduce price, the demand is said to be *perfectly elastic*. In such cases, even a small increase in price will lead to complete fall in demand. This is illustrated in fig.below.

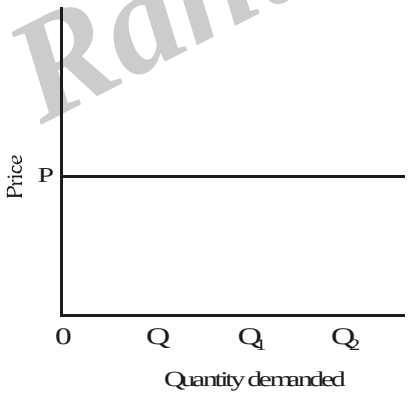


Figure : Perfectly Elastic Demand

(b) Perfectly Inelastic Demand

When a significant degree of change in price leads to little or no change in the quantity demanded, then the elasticity is said to be perfectly inelastic.

In other words, the demand is said to be perfectly inelastic when there is no change in the quantity demanded even though there is a big change (increase or decrease) in price.

Figure below reveals that there is no change in the quantity demanded though there is change in price, say increase or decrease. In other words, despite the increase in price from OP to OP_1, the quantity demanded has not fallen down. Similarly, though there is a fall in the price from OP_3 to OP_2, the quantity demanded remains unchanged.

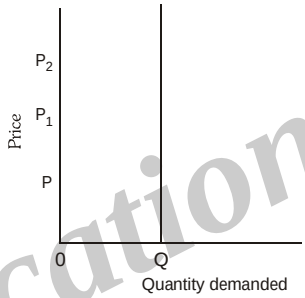


Figure : Perfectly Inelastic Demand

The concepts of perfectly elastic and perfectly inelastic demand do not manifest in real life.

(c) Relatively Elastic Demand

The demand is said to be relatively elastic when the change in demand is more than the change in price. Figure below reveals that the quantity demanded increases from OQ_1 to OQ_2 because of a decrease in price from OP_1 to OP_2. The extent of increase in the quantity demanded is greater than the extent of fall in the price.

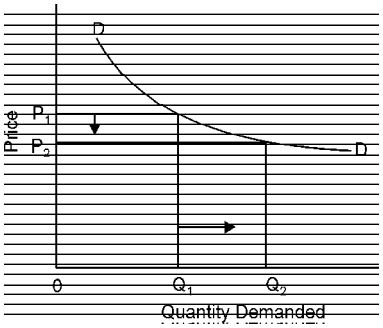


Figure : Relatively Elastic Demand

(d) Relatively Inelastic Demand

The demand is said to be relatively inelastic when the change in demand is less than the change in the price. This is illustrated in fig.below.

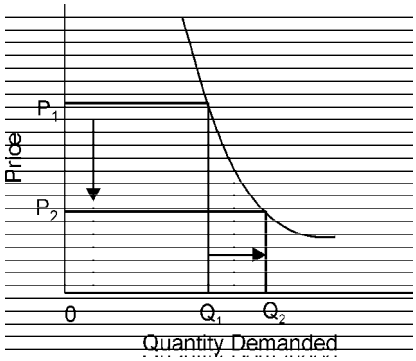


Figure : Relatively Inelastic Demand

Figure above reveals that the quantity demanded increases from OQ_1 to OQ_2 because of a degree in price from OP_1 to OP_2. The extent of increase in the quantity demanded is lesser than the extent of fall in the price.

(e) Unity Elasticity

The elasticity in demand is said to be unity when the change in demand is equal to the change in price. This is illustrated in fig.below.

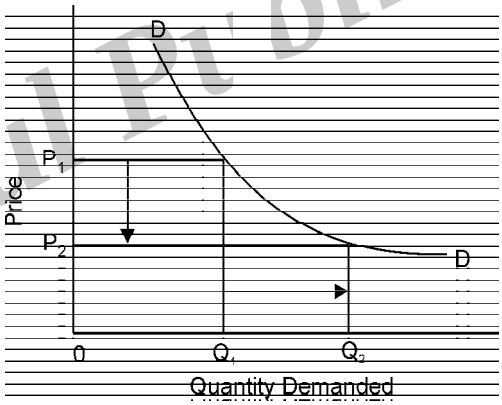


Figure: Unity Elasticity

Figure above reveals that the quantity demanded increases from OQ_1 to OQ_2 because of a decrease in price from OP_1 to OP_2. The extent of increase in the quantity demanded is equal to the extent of fall in the price.

Q12. Explain income elasticity demand.

Ans :

Income elasticity of demand refers to the quantity demanded of a commodity in response to a given change in income of the consumer.

Income elasticity is normally positive, which indicates that the consumer tends to buy more and more with every increase in income.

It is measured as follows:

$$\text{Income elasticity of demand} = \frac{\text{Proportionate change in quantity demanded for product X}}{\text{Proportionate change in income}}$$

The same is expressed as

$$\text{Edi} = \frac{(Q_2 - Q_1)/Q_1}{(I_2 - I_1)/I_1}$$

Where Q_1 is the quantity demanded before change, Q_2 is quantity demanded after change I_1 is income before change and I_2 is the income after change.

A positive income elasticity indicates that the demand for the product rises more quickly than the rise in disposable income. In other words, the demand is more responsive to a change in income.

Q13. Explain Cross Elasticity of Demand.

Ans :

Cross elasticity of demand refers to the quantity demanded of a commodity in response to a change in the price of a related good, which may be substitute or complement.

It is measured as follows :

$$\text{Cross elasticity of demand} = \frac{\text{Proportionate change in quantity demanded for product X}}{\text{Proportionate change in price of product Y}}$$

The same is expressed as

$$\text{Edc} = \frac{(Q_2 - Q_1)/Q_1}{(P_2y - P_1y)/P_1y}$$

Where Q_1 is the quantity demanded before change, Q_2 is quantity demanded after change, P_1y is the price before change and P_2 y is the price after change in the case of product Y.

Cross elasticityis always positive for substitutes (which means that the demand for tea goes up if there is an increase in the price of coffee) and negative for complements (which menas that if there is an increase in the price of sugar, the demand for coffee tends to fall).

Q14. Explain advertising elasticity of Demand.

Ans :

It refers to increase in the sales revenue because of change in the advertising expenditure. In other words there is a direct relationship between the amount of money spent on advertising and its impact on sales.

$$\text{Advertising elasticity} = \frac{\text{Proportionate change in quantity demanded for product X}}{\text{Proportionate change in advertisement costs}}$$

The same is expressed as

$$\text{Eda} = \frac{(Q_2 - Q_1)/Q_1}{(A_2 - A_1)/A_1}$$

Where Q, is the quantity demanded before change, Q_2 is quantity demanded after change A, is the amount spent on advertisement before change and A_2 is the amount spent on advertisement after change

The advertising elasticity is said to be high when even a small percentage change in the advertising expenditure results in a large percentage of change in the level of quantity demanded or sales.

3.3.1 Measurement of Elasticity of Demand

Q15. Explain different Measurement of Elasticity of Demand.

Ans :

The proportionate changes in quantity of demand and the proportionate changes prices of commodity functional relation is called price elasticity of demand. It can be derived the following equation.

$$\therefore \quad \boxed{\eta d = \frac{\Delta Q}{\Delta P}}$$

$\therefore$ ηd = Demand elasticity

AQ = Changes in quantity of demand

ΔP = Changes in prices of commodity

There are three types methods are available for estimating the elasticity of demand. They are

1) Total expenditure method
2) Point method
3) Arc method

1. Total Expenditure Method

It has been proposed by "Marshall based on price of commodity, quantity of unit and total expenditure base, he can analyses to estimated greater than 1, equal to 1, less than –1 elasticities of demand is being determined it can illustrated here under schedule.

Schedule

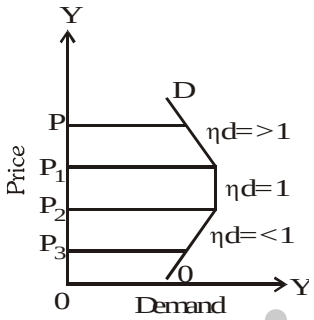
Price	Qty	Total expenditure	Determinants of yd
10	100	1000	Relative price yd = >1%
9	120	1080	
8	140	1120	Oxitary price yd = 1
7	160	1120	
6	180	1080	Relative price in yd = <1
5	200	1000	

In the above schedule if the price at Rs.10/- the purchased units are 100/- and the incurring total expenditure is 1000 rupees, if the price is comedown at Rs. 9/- the purchased units are raised at 120 units in order to incurred the total expenditure 1080 rupees which is more than to previous expenditure.

Therefore it is equalent to greater than 1 price elasticity of demand.

If the price is at Rs. 18/- the purchased units are 140 and the incurring total expenditure is 1120 rupees, if the price is comedown at Rs. 7/- the purchased units are raised at 160 units in order to incurred the total expenditure 1120 rupees which remains constant. Therefore it is equalent to price elasticity of demand.

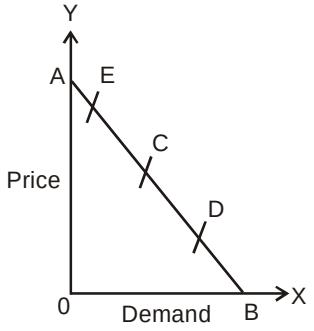
If the price is at rupees 6/- the purchased units. So units in order to incurred the total expenditure is 1080 is the price is comedown at Rs. 5/- the purchased units are raised 200 units in order to incurred. The total expenditure. Therefore it is equalent less than/price elasticity of demand. Based on the schedule we can illustrated here under diagram.



In the above diagram on y axis we are showing a price and on x-axis quantity of demand, the changes of prices OP to P_1 shows greater than 1 elasticity of demand, the changes of prices of P_1 to P_2 shows equal to 1 price elasticity of demand and the changes of price P_2 to P_3 shows less than 1 elasticity of demand.

2. Point Method

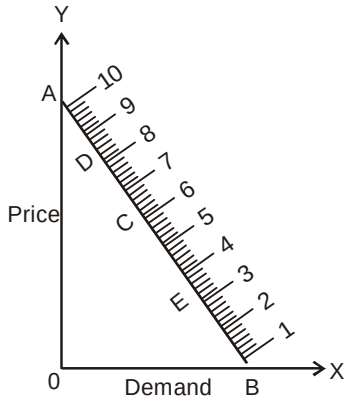
Based on this method on any point of the demand line we can traceout the nature of elasticity of demand it, can illustrated here under example :



On the above diagram the left to right downwards a and b and the demand line, and a, b demand line are plotted a, e, c, d, b points are mentioned in respective point the nature of elasticity of demand can we traceout with the help of point method.

Let we know that a, b demand line length gpr.

Example : If the a, b demand line length is above 10 centimeters. Let us assume based on it we can fixed here under a scale of demand line.



The point method of elasticity of demand the following formula.

(Lower segment of a point)

Point method of $\eta d = \frac{L}{U}$

(Upper segment of a point)

Based on above formula (or) equation, for

Example :

At the point of C yd $= \frac{L}{U} = \frac{CB}{CA} = \frac{5}{5} = 1$, so it equal to = 1 ηd

At the point of D $\eta d = \frac{L}{U} = \frac{DB}{DA} = \frac{75}{75} = 3$, so it equal to = Z1 ηd

At the point of C $\eta d = \frac{L}{U} = \frac{EB}{EA} = \frac{2.5}{7.5} = \frac{1}{3} = 0.33$, so it equal to ηd

At the point of A $\eta d = \frac{L}{U} = \frac{AB}{A} = \frac{10}{0} = a = 0.33$, so it equal to ηd

At the point of B $\eta d = \frac{L}{U} = \frac{B}{BA} = \frac{0}{10} = a = 0.33$, so it equal to $\eta d = 0$

3. Arc Method : The method is depends upon the following formula.

$$\text{Arc method } \eta d = \frac{Q_1 + Q_2}{Q_1 - Q_2} - \frac{P_1 + P_2}{P_1 - P_2} \Big/ 2$$

$\therefore$ Q_1 = old demand

Q_2 = new demand

P_1 = old price

P_2 = new price.

Q16. Explain Significance of Elasticity of Demand.

Ans :

Elasticity of demand is a crucial concept in the spheres of trade, commerce and finance. The following are some of the important advantages of elasticity of demand:

1. Price Determination

The doctrine of elasticity of demand plays a vital role in price determination. The sellers increase prices if the demand is less elastic and lower prices if the demand is elastic.

2. Monopoly Market

For a monopolist to optimize his profits, he must know the elasticity of demand for his products. In other words, the degree of monopoly can be measured with the help of elasticity of demand. A monopolist can perform price discrimination only when he is aware of price elasticity for his commodities. A prudent monopolist increases prices in the inelastic market and lowers prices in the elastic market. In addition, the concept of price elasticity of demand plays a vital role in dumping practice as well.

Example

Oil producing countries tend to increase price by cutting down oil production. When the oil production is reduced, an artificial scarcity is created; consequently, the price is increased. This is possible as long as the oil products are price inelastic.

3. Pricing Public Utilities

Many of the public utilities are necessities. For instance, supply of water, electricity, transport and so on is essential for our everyday activities. Therefore, the demand for these utilities is price inelastic. The concept of elasticity of demand helps the government to rationalize prices for these important utilities. Otherwise, prices for these utilities will be very high, if they are provided by private entities.

4. Prosperity Versus Poverty

Do you agree that even an overwhelming prosperity can cause poverty? It is true in some special cases. One of such special cases is bumper crop. In agriculture when there is bumper crop, the price of the commodity falls because of excessive supply. In this case, if the demand for the commodity is inelastic, it may create disaster because, the farmers will get very low prices for their goods. Therefore, not all properties may alleviate poverty. This scenario can be studied well with the help of the concept of elasticity of demand.

5. Currency Devaluation

The application of elasticity of demand can be extended to the analysis of currency devaluation. Devaluation helps to increase exports. It is possible only when the demand for exported goods is highly elastic. If the demand is inelastic, there will not be any use in currency devaluation.

6. Taxation

Suppose you are the finance minister of your country. On which commodities you levy taxes – the one that is price elastic or the one that is price inelastic? Obviously, you have to choose the commodities that are price inelastic. The reason is that if the commodities are price elastic, the demand will come down drastically when their prices increase. Taxes increase the price of the commodities. If the demand falls, revenue also will fall. Because of this reason, you have to choose those commodities that are price inelastic.

7. Wage Determination

The idea of elasticity of demand helps to wage determination. You know that wages of laborers are closely associated with the price of products they produce. What is the way to determine the right price that gives maximum profits? The concept of elasticity of demand answers this question. Once the right price is determined, the profit will be maximized. Subsequently, wages of laborers can also be determined.

Q17. Explain Factors affecting Elasticity of Demand.

Ans :

Elasticity is governed by a number of factors. Change in any one of these factors is likely to affect the elasticity of demand. The factors are:

(a) Nature of Product

Based on their nature, the products and services are classified into necessities, comforts and luxuries. Necessaries imply the absolute or basic necessities such as food, clothing, hous-ing. Comforts refer to TV, refrigerator and so on. By luxuries, we mean sofa sets, marble flooring in a house and such others. The meaning and definition of these necessaries, luxuries and comforts change from person to person, time to time and place to place. For example, a scooter may be a comfort or luxury for a student but when he does a part-time job, it may be a necessity for him.

The nature of product has a significant impact on the elasticity of demand. For instance, if there is an increase in the price of rice, we still buy it because it is a necessity for us. This means that the demand is inelastic to price. Though there is an increase in price, we tend to buy the necessaries such as petrol, diesel and so on. In other words, the demand does not fall because of increase in price. From this, we can say that the necessaries have inelastic demand. For comforts and luxuries, the demand is relatively elastic. It means that any increase in the price of comforts or luxuries will lead to moderate to significant fall in their demand.

(b) Time Frame

The more the time available for the customer, the demand for a particular product may be elastic and vice versa. Take the case of vegetables. When you do not have time, you go to a nearby shop and buy whatever you want at the given price. Had you had little free time, you would have preferred to get the same from a vegetable market at lesser price.

(c) Degree of Postponement

Where the product consumption can be postponed, the product is said to have elastic demand and where it cannot be postponed, it is said to have inelastic demand. The consump-tion of necessaries cannot be postponed and hence they have inelastic demand.

(d) Number of Alternative uses

If the number of alternative uses are more, the demand is said to be highly inelastic and vice versa. Take the case of power or electricity. It is used for a number of alternative uses such as running of machines in industries, offices, households, trains, and so on.

(e) Tastes and Preferences of the Consumer

Where the customer is particular about his taste and preferences, the product is said to be inelastic. For the customers who are particular or loyal to certain brands such as Colgate, Tata Tea, Annapurna Atta, and so on, price increases do not matter. They tend to buy that brand inspite of the price changes.

(f) Availability of Close Substitutes

Where there are a good number of close substitutes, the demand is said to be elastic and vice versa. For gold, there is no close and literal substitute and hence the demand for gold is inelastic. If coffee and tea are equally good for me, if there is an increase in price of coffee, I may tend to switch over to tea. But this may not hold good when I am particular about coffee only. I may be prepared to pay higher price for coffee.

(g) In case of Complementaries or Joint Goods

In case of complementaries or goods having joint demand, the elasticity is comparatively low.

(h) Level of Prices

If the price is very expensive (such as diamonds) or very cheap (such as salt), then the product is likely to have an inelastic demand. If the price is too high, a fall in it will

not increase the demand much. Similarly, if the price is too low, a further fall in its price is not likely to result in more demand. The demand of the relatively poor people is more sensitive to price changes. In order to derive maximum satisfaction from their limited income, they try to plan their purchases in response to changes in prices. The rich may not bother about price changes.

(i) Availability of Subsidies

Subsidy refers to money paid by a government or other public authority in order to help a company financially or to make something cheaper for the public. There is need for subsidies in case of goods with inelastic demand such as LPG, sugar, wheat and so on.

(j) Expectation of Prices

Where people expect a fall in the price, the demand for the product is likely to be inelastic.

(k) Durability of the Product

Where the product is durable in case of consumer durables such as TV, the demand is elastic. In the case of perishable goods such as milk, the demand is inelastic.

(l) Government Policy

Where the government policy is liberal, the product is likely to have elastic demand and vice versa. Government, in the interest of the lower income group consumers, closely monitors the prices of certain products (such as, ration goods as sold in fair price shops are likely to have inelastic demand). Also, another example could be taxes. Government can raise tax collections with a little reduction in the tax rates.

3.4 Introduction to Supply - Methods of Determining Supply

Q18. Define Supply? Explain the deter- minants of Supply.

Ans :

Supply of a commodity is the amount of its which the sellers are able and willing to offer for sale at a price during a certain period of time. Supply is a relative term - related to price and time. Market supply means the total quantity of a commodity that all the firms are willing to sell at a given during a given time period. It is found by adding the supply of the firms selling the commodity.

Supply analysis can be used to determine the impact of changes in product and factor prices, in technology, and in access on factor demands (including labor), production, marketed output, aggregate supply, and incomes. Generally, it can be used to analyze the impact on production of the removal of barriers to access or other changes in markets. Supply analysis, in the employment context, deals with key staffing questions related to current staffing levels in an organization.

Definitions

(i) According to Thomas Supply of goods is the quantity offered for sale in a given market at a given time at various prices.

(ii) According to Samuelson Supply refers to the amounts of a good that producer in a given market in a given market desired to sell, during a given time period at various prices, Ceteris Paribus.

Determinants of Supply

Supply of a commodity by a firm is generally not a fixed quantity. It keeps on changing. The factors that effect the supply of a commodity are :

1. Price of the Commodity
2. Other Factors
   a) Prices of other Commodities
   b) Prices of Factors of Production
   c) Objective of the Producer
   d) Production Technology

1. Price of the Commodity

While studying the effects of changes in the price of a commodity on its supply, we assume that no other factor is affecting the supply of the commodity. When other things are constant, a rise in the price of a commodity provides an incentive to its seller to sell more of it as it fetches more profits and a fall in the price is a disincentive to the seller as his profits.

2. Other Factors

While studying the effects of changes in the price of a commodity on it supply we assume that other factors that can affect supply remain constant. We assume that the price of a commodity does not change.

(a) Prices of other Commodities

If the prices of other commodities increase, it becomes more profitable for a firm to produce these other commodities. So it will shift its resources from the production of the commodity whose price has not changed to the production of the commodities whose price have increases.

(b) Prices of Factors of Production

If the cost of production of a commodity rises due to a rise in the price of any one or more of its factors of production, and price of the commodity remaining the same then the margin of profit of its producers will fall. They will reduce the supply of the commodity through its price has not changed. Similarly a fall in the price of factors of production will increase the supply of a commodity.

(c) Objective of the Producer

The objective of a producer of a commodity is to earn maximum profits. So he producer that much of a commodity which will fetch him maximum profits. It is possible that a producer may be interested in maximizing his sales rather than his profits. The producer goes on increasing the production and sales so long as his target of profit is not adversely affected. Therefore, the objective of a producer also affects the supply of the commodity.

(d) Production Technology

Technological advancement means introduction of new machines and better methods of production. This reduces the cost of production and increases profit. So the producer is able to supply more of the commodity at the same price. Thus improvement in technology increases the supply of a commodity. Price remaining the same.

Q19. Define Supply Function.

Ans :

The mathematical function explaining the quantity supplied in terms of its various determinants, including price; thus the algebraic representation of the supply curve.

Mathematical or functional relationship between supply and its determinants is called supply function. It is a given as :

$$Q_x = f(P_x, C_x, T_x)$$

Where

Q_x = Supply of Commodity X

P_X = Price of Commodity X

C_x = Cost of Production of Commodity X

T_x = Technology of its Production.

Q20. Define law of supply explain assumptions of law of supply.

Ans :

The direct relationship between the price of a commodity and its supply is stated in the form of a law called Law of Supply. The law of supply states that "Other things remaining unchanged, the supply of a commodity expands with a rise in price and contracts with a fall in price." The market supply curve is a diagrammatic representation of the law of supply. Law of supply states the relationship between the price and the supply of a commodity. Price of the commodity affects its supply.

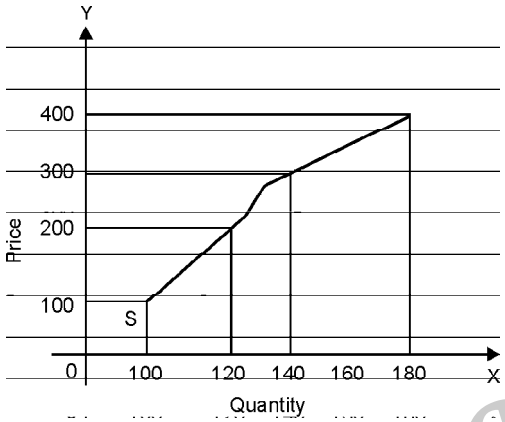


Figure: Market Supply Curve

Assumptions underlying Law of Supply

- Cost of production is unchanged
- No change in technique of production
- Fixed scale of production
- Government policies are unchanged
- No change in transport cost
- No speculation
- Prices of other goods held constant.

Expectations to Law of Supply

(a) Supply of Labour

If we take the supply of labour at very high wage, we may find that the supply of labour had decreased instead of increasing.

(b) Agricultural Products

Since the production of agricultural products cannot be increased beyond a certain limit, the supply cannot be increased beyond this limit even on an increase in their prices.

(c) Artistic Goods

Supply of artistic goods cannot be increased or decreased easily.

(d) Goods of Auction

Supply of goods of auction is limited as such cannot neither be increased not decreased.

(e) Hope of Change in the Prices of Commodities in Near Future

If the price of commodity is on rising pace, then the supply of such commodity decreases as producers and sellers will like to store this commodity and vice-versa.

3.4.1 Price Elasticity of Supply

Q21. What do you understand by price elasticity of supply ? Explain different types of price elasticity of supply.

Ans : **(July-21)**

The degree to which the supply of the product increases or decreases due to the change in the product price is termed as the elasticity of supply. Elasticity of supply is considered as the responsiveness of the sellers to a change in the price of the product.

The following formula can be used for measuring the elasticity of supply,

$$E_s = \frac{\text{Proportionate change in quality supplied}}{\text{Proportionate change in price}}$$

$$E_2 = \frac{\Delta Q_S}{\Delta P} \times \frac{P}{Q_S}$$

Where,

E_s = Elasticity of supply

ΔQ_s = The change in quantity supplied

ΔP = The change in price

Q_s = The original quantity supplied

P = The original price.

The elasticity of supply is depicted in the figure given below,

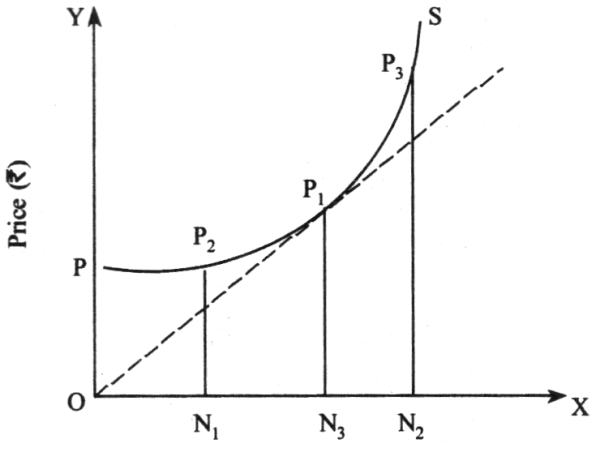


Fig. Quantity Supplied (Units)

Figure represents the following,

(a) P and P_2 = perfectly elastic supply

(b) P_3 and S = perfectly inelastic supply

(c) P_2 and P_1 = Elastic supply

(d) P_1 and P_3 = Inelastic supply.

The time element basically has significant impact on the elasticity of supply. There is more elasticity of supply in the long-period of time, while the supply curve in a short period of time is perfectly inelastic.

Example

ABC firm supplied 3,000 umbrellas at ` 75 per unit. There is a rise in the price of umbrella to ` 100 per umbrella. Lotus firm also raised the supply of umbrellas to 5,000. Work-out on the elasticity of supply for the umbrellas.

Given,

The original price (P) = ` 75

Change in price (ΔP) = ` 100 – ` 75

= ` 25

Original quantity supplied (Q) = 3,000

Change in quantity (ΔQ) = 5,000 – 3,000

= 2,000 Umbrellas

Elasticity of supply (E_s) is given by,

$$E_s = \frac{\Delta Q}{\Delta P} \cdot \frac{P}{Q}$$

$$= \frac{2,000}{25} \times \frac{75}{3,000}$$

$$= \frac{1,50,000}{75,000}$$

= ` 2

Types of Elasticity of Supply

The elasticity of supply is broadly classified into five types, which are as follows,

1. Perfectly elastic supply
2. Perfectly inelastic supply
3. Relatively elastic supply
4. Relatively inelastic supply
5. Unitary elastic supply.

1. Perfectly Elastic Supply

The situation in which there is a change in the supply or sales of a product, inspite of no change in the price level is known as perfectly elastic supply. In such case, the elasticity is infinite.

$(E = \infty)$

2. Perfectly Inelastic Supply

The situation in which there is no change in the supply or sales of a product inspite of a drastic change in the price level is termed as perfectly inelastic supply. Here, the elasticity is equal to zero.

$(E = 0)$

3. Relatively Elastic Supply

Under relatively elastic supply, the rate of change of supply or sales is greater than the rate of change of price. Here, the elasticity of supply is greater than one.

$(E > 1)$

4. Relatively Inelastic Supply

The situation in which the rate of change of supply or sales is less than the rate of change of price is termed is relatively inelastic supply. The elasticity of supply here is less than one.

$(E < 1)$

5. Unitary Elastic Supply

In unitary elastic supply, the rate of change of supply or sales is perfectly equal to the rate of change of price. Here, the elasticity is equal to one.

$(E = 1)$

3.5 LEONTIEF'S METHODS OF DETERMINING DEMAND CURVE FROM TIME SERIES DATA AND THEIR LIMITATIONS

Q22. Explain Leontief's Methods of Deter- mining Demand Curve from Time Series Data and their Limitations ?

Ans : **(Nov.-20, Dec.-18)**

Leontief's Method

Leontief's method of determining demand curve from time series data is based on the following assumptions,

1. Every market transaction denotes the intersection of rapid supply and demand curves, which tend to shift from one position to another at times. This means that the range at which the curves tend to shift from time to time should also be studied, along with the determination of demand and supply elasticities.
2. The variations occurring in the shifts of demand and supply curves are independent of each other and the shape (elasticities) of the curve stays unaffected by them, which means that a variation in the shifting of the demand curve to the right is more likely to be related with the variation of the shifting of supply curve to the left as to the right.
3. Every demand and supply curve consists of constant elasticity, which implies that the supply and demand curves should be straight lines when plotted on a double logarithmic scale.

Let X_t and Y_t represent the consumption and price logarithms of a product respectively, at time $t(t=1,2,3,...,n)$, .

We get,

Demand Curve = $X_t = \eta_1 Y_t + P_t$ ... (1)

Supply Curve = $X_t = \eta_2 Y_t + Q_t$ ... (2)

Where η_1 and η_2 denote demand elasticity and supply elasticity respectively and P_t and Q_t are two random variables distributed independently with,

$$E(P_t) = 0 = E(Q_t) \quad ...(*)$$

It can be observed in (1) and (2), that X_t is taken for supply as well as consumption, as market equilibrium has,

$$d = S \Rightarrow \log d = \log S = X_t \text{ (Say)}$$

(1) and (2) can also be written as,

$$X_t - \eta_1 Y_t = P_t \quad .....(3)$$

and $$X_t - \eta_2 Y_t = Q_1 \quad .....(4)$$

Multiply (3) and (4), we get,

$$X_t^2 + \eta_1\eta_2 Y_t^2 - (\eta_1 + \eta_2) Y_t X_t = P_t Q_t \quad .....(5)$$

As P_t and Q_t are independently distributed with,

$E(P_t) = 0$ and $E(Q_t) = 0$,

$Cov(P_t, Q_t) = 0 \Rightarrow E(P_t, Q_t) - E(P_t).E(Q_t) = 0$

By using eq. (*), we get,

$E(P_t, Q_t) = 0$ ... (**)

By adding time t and using eq (**), we get,

$$\Sigma X_t^2 + \eta_1\eta_2\Sigma Y_t^2 - (\eta_1 + \eta_2)\Sigma Y_t X_t = 0 \quad ...(6)$$

Equation (6) cannot be solved with two unknown quantities of η_1 and η_2 that's why the time range t: [1, n] is divided into two equal halves,

$$t_1 : \left[1, \frac{n}{2}\right] \text{ and } t_2 : \left[\frac{n}{2}+1, n\right]$$

Adding (5) w.r.t. time for each half separately and using eq. (**), we get,

$$\left.\begin{aligned} &\sum_{t=1}^{n/2} X_t^2 + \eta_1\eta_2 \sum_{t=1}^{n/2} Y_t^2 - (\eta_1 + \eta_2)\sum_{t=1}^{n/2} Y_t X_t = 0 \\ \text{and } &\sum_{t=\frac{n}{2}+1}^{n} X_t^2 + \eta_1\eta_2 \sum_{t=\frac{n}{2}+1}^{n} Y_t^2 - (\eta_1 + \eta_2)\sum_{t=\frac{n}{2}+1}^{n} Y_t X_t = 0 \end{aligned}\right\} \quad .....(7)$$

Equation (7) can be solved for η_1 and η_2 and $\Sigma X_t^2, \Sigma Y_t^2$ and $\Sigma X_t Y_t$ can be calculated from the time series data.

Limitations

The following are the limitations of Leontief's method from statistical and economic point of view,

1. The assumption that the variation in shifting of the demand and supply curves are independent of each other and that the Coumot-Marshall demand and supply cutVe may shift in any possible direction, violates the principle of general theory of equilibrium's fundamental principle, which states that "the demand of any one commodity is a function not only of its price but also of all other prices".
2. The assumption of Leontief's method that the demand and supply curves are simultaneous is not a valid assumption in case of agricultural commodity. For example, on an average, low (high) price for a commodity in a specific year is related with low (high) production of the commodity in the following year.
3. The assumption which states that demand elasticity and supply elasticity are constant might not stand true.

Criticism

The Leontief's method received criticism for the method which seems statistically defective to calculate $q_1 + q_2$- He splits up his series into the ellipse of the two curves formed from the mathematical solution are not similar and the respective axes are not parallel to each other. In addition to this, there is a noticeable difference between the first and the second half period, which shows that the data is not homogeneous. It is in such a way that every period must be studied separately.

3.6 PIGOU'S METHODS OF DETERMINING DEMAND CURVE FROM TIME SERIES DATA AND THEIR LIMITATIONS

Q23. Explain Pigou's Methods of Determining Demand Curve from Time Series Data and their Limitations.

Ans : **(July-19)**

Pigou's Method

Pigou's method of determining demand curve from time series data is based on the following assumptions,

1. The first assumption of Pigou is that, at each time interval the demand curve has a smooth appearance, which implies that at each time interval, the demand curve has constant elasticity.

$$d = cp^{-\alpha} \Rightarrow \log d = \log c - \alpha \log p$$

i.e., log d = log c + a log p; (a = $-\alpha$)

$\therefore$ X = aY + b

Where X are Y are the logarithms of consumption and price respectively.

2. The shifting of the demand curve over various time periods is steady and the extent to which the curves have shifted is equal for every two successive time intervals.

It can also be said that the extent of the shifting of the demand curve is such that the distance between the i^{th} position of $(i + 1)^{th}$ position (on a logarithmic scale) is equal to the distance between $(i + 1)^{th}$ position and $(i + 2)^{th}$ position.

From the above assumptions, at different time points,

t = 1,2,..., we have,

$$\left.\begin{aligned} X_1 &= aY_1 + b \\ X_2 &= aY_2 + b + r \\ &\vdots \quad \vdots \\ X_i &= aY_i + b + (i-1)r \\ X_{i+2} &= aY_{i+1} + b + ir \\ X_{i+2} &= aX_{i+2} + b + (i+1)r \\ &\vdots \quad \vdots \end{aligned}\right\}$$

According to the above equation, we have,

$$(X_{i+1} - X_i) - a(Y_{i+1} - Y_t) = (X_{i+2} - X_{i+1}) - a(Y_{i+2} - Y_{i+1})$$

$$\therefore \quad a = \frac{X_{i+2} - 2X_{i+1} + X_i}{Y_{i+2} - 2Y_{i+1} + Y_i} = a_i \text{ (say) } (i = 1, 2,...) \qquad \text{..... (1)}$$

The steps involved in Pigou's method are as follows,

Step-1: A table of logarithms consisting of time-series values of price (Y) and consumption i.e., log p and log d.

Step-2: Calculate a. (i = 1, 2,...) from eq. (1). As demand is a diminishing price's functions, if the value of 'a_i' at any interval is said to be positive, then it cannot be considered as a measure of demand elasticity for appropriate set of times. But, if it is said to be negative, then it can be considered as a measure of the elasticity.

Step-3: Every negative a_i can be considered as an observation on the unknown elasticity belonging to the demand curve, if the negative values of a_i. s are exceeding the positive values and if they are closely grouped into a given value and also if the data is not suspected otherwise.

Limitations

The following are the limitations of Pigou's method,

1. The assumption that the demand curve for a commodity is x = f(p, t)

   Where x denotes the quality of the demand commodity, p denotes the commodity's price and t denotes time. It implies that the prices of commodities which are closely related to that specific commodity have zero or no impact on the commodity or all the effecting factors are considered are ineffective while studying the variations occurring in x due to the variations occurring in y. But this assumption does not stand true as it is not possible to put a stop on all the other factors without consideration.

2. The method proves to be a complete failure if there are three collinear price-quantity points in three successive sets of observations. This means that, Pigou's method is not applicable to collinear successive observations but, it is applicable to functions which are non-linear and change in directions.

3. This method is not useful in some cases where the underlying theoretical demand curves tends to change its motion's direction.

3.7 PARETO LAW OF INCOME

Q24. What do you understand by Pareto Law of Income.

Ans :

During 19th and 20th century Vilfredo Pareto proposed the law of distribution of income. He is the first analyst who have conducted the study from the statistical perspective for solving the problem of income distribution among a group of citizens. This interpretation of his law of income distribution was based on the empirical study of the income data of different countries at different times.

The distribution of income in a stable economy is derived approximately by the following formula,

$$P = A (x - a)^{-v}$$

Where,

P = Number of people having income x or greater

a = Lowest income at which the curve starts

A and V = parameters

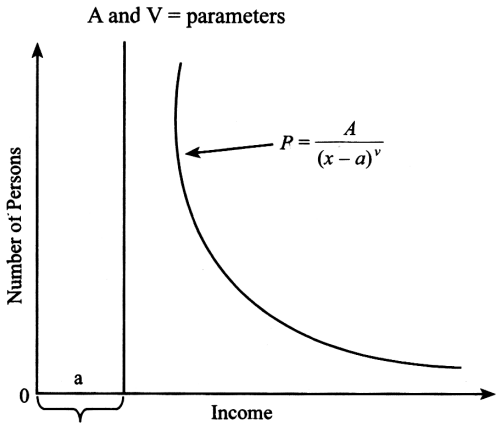


Fig. Pareto Curve

The curve which is derived with the help of above formula is extremely asymmetrical, similar to a hyperbola.

The Pareto curve becomes,

$P = A x^{-v}$ when origin is shifted to $x = a$

Logarithm of both sides are taken as,

$\log P = \log A - v \log x$

Therefore, the graph will be a straight line with its slope equal to -v when Pareto curve is graphically represented on a double logarithmic scale.

Some of the important consideration of Pareto's law of income distribution are as follows,

1. Generally, simplified form of Pareto curve is used due to lack of information regarding number of persons with less incomes.
2. Pareto curve is best fitted for higher incomes as compared to lower incomes.
3. According to Pareto value of 'v' varies from 1.2 to 1.9 in several countries. Thus, the value of 'v' can be taken as 1.5 on an average.
4. If $\log P = \log A - v. \log x$ is differentiated.

   We get,

$$\frac{d \log P}{d \log x} = -v$$

   Therefore, V is interpreted as the income elasticity of P.
5. If $P = Ax^{-v}$ is differentiated, with reference to x, we get,

$$dP = -Avx^{-v-1}dx$$

   Hence, corresponding to an increment dx in income of class with lower limit x, the relative increment in P would be derived by,

$$\frac{dP}{P} = -\frac{v}{x}dx$$

6. Pareto's law received various criticism by scientific investigators for its rigid and uncompromising form.

Formulation of Problem of Income Distribution

The problem of distribution of income can be formulated as,

(i) The form of frequency function $\phi(x)$ is determined for total income distribution of society members from poorest to the richest.

(ii) Checking whether $\phi(x)$ is governed by the type of society from which income is derived or is its form is unavoidable.

(iii) Checkin any deductive reason be assigned for the form of $\phi(x)$.

3.7.1 Distribution Curves of Concentration

Q25. Explain in detail about Distribution curves of concentration?

Ans :

If from of the distribution i.e., $\phi(x)$ is unknown, the formulation of the problem of distribution of income can be achieved by 'curves of concentration'. The formula through which formulation can be done are as follows,

$$Sb_x = \frac{\text{No of Persons with Income } \leq x}{F}$$

$$T_x = \frac{\text{Total Income of Persons with Income } \leq x}{I}$$

Where,

F = Total Frequency of the population

I = Total Income

If the variables S ant T has a functional relationship, then the equations will be,

$S = S_x$ and $T = T_x$

If income distribution law is well known then for eliminating x in above equation i.e., $S = S_x$ and $T = T_x$, we get the equation T = f(s).

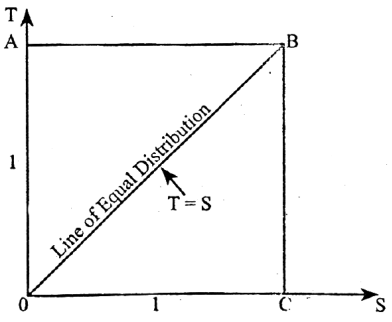


Figure : Line of Equal Distribution

The straight line T = S is referred as line of equal distribution

Another formulation is useful to income data that can be given to equation T = f(s)

$$S_x = \frac{F - F_x}{F} \text{ and } T_x = \frac{I - I_x}{I}$$

Where,

F_x = No. of persons with income x or more

I_x = Total income possessed by F_x

M.O. Lorenz, Gini and others suggested curves for representation of income distribution.

It can be explained with help of an illustration, assuming Pareto's form for F_x, which as follows,

$$F_x = Ax^{-v}$$

$$\log F_x = \log A^{-v} \log x$$

$$= A_1^{-v} \log x \quad [\because A_1 = \log A]$$

The frequency function of distribution of income is derived by,

$$\phi(x) = -\frac{d}{dx} Fx = Avx^{-(v+1)}$$

Subsequently,

I_x = Income possessed by $F_x = \int_x^{x_1} x\phi(x)dx$

Where,

x_1 = upper limit of distribution

Assuming $x_1 = \infty$ we get,

$$I_x = Av\int_x^{\infty} x^{-v}dx = Av\left|\frac{x^{-(v-1)}}{-(v-1)}\right|_x^{\infty} = \frac{AV}{V-1}.x^{-(v-1)}$$

$$= A\,\delta\, x^{-(v-1)}, \delta = \frac{v}{v-1}$$

$$\Rightarrow \log I_x = B - (v-1)\log x,$$

Where,

B = $\log(A\delta)$, which is a constant

$$\frac{\log A - \log F_x}{v} = \frac{\log(A\delta) - \log I_x}{(v-1)}$$

$$\frac{A}{F_x} = \left(\frac{A\delta}{I_x}\right)^{\delta}$$

F = Total population of people earning income

= No. of persons with $x \geq 1$

= $(N_x)_{x=1}$ = A

Similarly, I = $(I_x)_{x=1}$ = Aδ

Substituting in $\frac{A}{F_x} = \left(\frac{A\delta}{I_x}\right)^{\delta}$, we get,

$$\frac{F}{F_x} = \left(\frac{I}{I_x}\right)^{\delta}, \left(\delta = \frac{v}{(v-1)}\right)$$

The expression in above equation is derived under the assumption that $(F_x)_{x=1}$ = A is applied over the complete range of income.

From $S_x = \frac{F - F_x}{F}, T_x = \frac{I - I_x}{I}$ and

$$\frac{F}{F_x} = \left(\frac{I}{I_x}\right)^{5}, \left(\delta = \frac{v}{v-1}\right), \text{we get,}$$

$$I - S_x = \frac{F_x}{F} = \left(\frac{I_x}{I}\right)^{\delta} = (1 - T_x)^{\delta} \Rightarrow T = 1 - (1 - S_x)^{1/\delta}$$

$\therefore$ $T = -(1 - P)^{1/s}$ gives the function T = f(s). The graph of this function is known as Lorenz curve, which is the design of income concentration and is compared in figure of Lorenz curve with line of equal distribution.

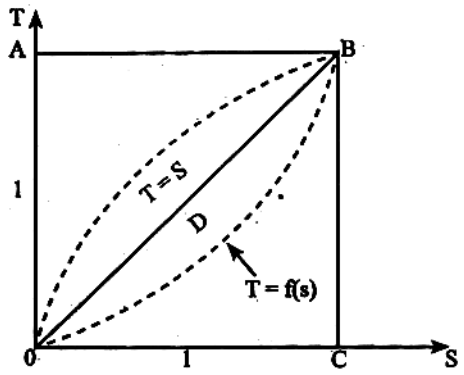


Fig : Lorenz Curve

Gini profounded a concentration ratio (ρ) as a measure of difference between distributions, which as follows,

$$\rho = \frac{\text{Area BDOB}}{\Delta BOC} = \frac{\Delta BOC - \text{Area BDOCB}}{\Delta BOC}$$

$$= 1 - \frac{\text{Area BDOCB}}{\Delta BOC}$$

However,

$$\Delta BOC = \frac{1}{2} \times OC \times BC = \frac{1}{2} \quad (\because OC = BC = 1)$$

P = 1 – 2 Area BDOCB

$$= 1 - 2\int_0^1 Td_s = 1 - 2\int_0^1 \left[1 - (1-S)^{1/\delta}\right]^{ds}$$

$$= 1 - 2\left| S + \frac{(1-S)^{\frac{1}{\delta}+1}}{(\frac{1}{\delta}+1)} \right|_0^1$$

$$= 1 - 2\left[1 - \frac{\delta}{1+\delta}\right] = \frac{\delta - 1}{\delta + 1}, \left(\delta = \frac{V}{V-1}\right)$$

$$= \frac{1}{2v-1}$$

d varies from 1 to ∞, ρ varies from 0 to 1, when Pareto's value v = 1.5 is assumed, we get,

$\rho = \frac{1}{2} = 0.5$

3.8 INDEX NUMBERS - CONCEPT

Q26. What do you understand by index numbers ?

Ans :

Some prominent definitions of index numbers are given below:

Definitions

(i) **According to Croxton & Cowden** "Index numbers are devices for measuring differences in the magnitude of a group of related variables."

(ii) **According to Spiegel** "An index number is a statistical measure designed to show changes in a variable or a group of related variables with respect to time, geographic loca-tion or other characteristics such as income, profession, etc."

(iii) **According to Patterson** "In its simplest form an index number is the ratio of two index numbers expressed as a per cent. An index number is a statistical measure—a measure designed to show changes in one variable or in a group of related variables over time, or with respect to geographic location, or other characteristic."

(iv) According to Morris Hamburg "In its simplest form, an index number is nothing more than a relative number, or a 'relative' which expresses the relationship between two figures, where one of the figures is used as a base."

(v) According to Berenson & Levine "Generally speaking, index numbers measure the Size or magnitude of some object at a particular point in time as a percentage of some base or reference object in the past."

For a proper understanding of the term index number, the following points are worth considering :

- **Index numbers are specialized averages :**

  As explained in the chapter on measures of central value, an average is a single figure representing a group of figures. However, to obtain an average the items must be comparable: for example, the average weight of men, women and children of a certain locality has no meaning at all. Furthermore, the unit of measurement must be the same for all the items. Thus an average of the weight expressed in kg., lb., etc., has no meaning. However, this is not so with index numbers. Index numbers are used for purposes of comparison in situations where two or more series are expressed in different units or the series are composed of different types of items. For example, while constructing a consumer price index the various items are divided into broad heads, namely (i) Food, (ii) Clothing, (iii) Fuel and Lighting, (iv) House Rent, and (v) Miscellaneous. These items are expressed in different units : thus, under the head 'food' wheat and rice may be quoted per quintal, ghee per kg., etc. Similarly, cloth may be measured in terms of metres. An average of all these items expressed in different units is obtained by using the technique of index numbers.

- Index numbers measure the net change in a group of related variables.Since index numbers are essentially averages they describe in one single figure the increase or decrease in a group of related variables under study. The group of variables may be the prices of a specified set of commodities, the volume of production in different sectors, etc. Thus, if the consumer price index of working class for Delhi has gone up to 113 in February 2000 compared to February 1999 it means that there is a net increase of 13 per cent in the prices of commodities included in the index.

- Index numbers measure the effect of changes over a period of time. Index numbers are most widely used for measuring changes over a period of time. Thus we can find out the net change in agricultural prices from the beginning of First Plan period to the end of the Eighth Plan period, i.e., from 1951 to 1996. Similarly, we can compare the agricultural production, industrial production, imports, exports, wages, etc., at two different times. However, it should be noted that index numbers not only measure changes over a period of time but also compare economic conditions of different locations, different industries, different cities or different countries.

Q27. Explain the uses of index number

Ans :

Index numbers are indispensable tools of economic and business analysis. Their significance can be best appreciated by the following points:

(i) They help in framing suitable policies.

Many of the economic and business policies are guided by index numbers. For example, while deciding the increase in dearness allowance of the employees, the employers have to depend primarily upon the cost of living index. If wages and salaries are not adjusted in accordance with the cost of living, very often it leads to strikes and lock - outs

which in turn cause considerable waste of resources. The index numbers provide some guideposts that one can use in making decisions.

Though index numbers are most widely used in the evaluation of business and economic conditions, there is a large number of other fields also where index numbers are useful. For example, sociologists may speak of population indices; psychologists measure intelligence quotients which are essentially index numbers comparing a person's intelligence score with that of an average for his or her age; health authorities prepare indices to display changes in the adequacy of hospital facilities and educational research organizations have devised formulae to measure changes in the effectiveness of school systems.

(ii) They reveal trends and tendencies.

Since index numbers are most widely used for measuring changes over a period of time, the time series so formed enable us to study the general trend of the phenomenon under study. For example, by examining index number of imports for India for the last 8-10 years we can say that our imports are showing an upward tendency, i.e., they are rising year after year. Similarly, by examining the index numbers of industrial production, business activity, etc., for the last few years we can conclude about the trend of production and business activity. By examining the trend of the phenomenon under study we can draw very important conclusions as to how much change is taking place due to the effect of seasonality, cyclical forces, irregular forces, etc. Thus index numbers are highly useful in studying the general business conditions.

(iii) They are important in forecasting future economic activity.

Index numbers are useful not only in studying the past and present workings of our economy, but they are also important in forecasting future economic activity. Index numbers then are often used in time series analysis, the historical study of long-term trend, seasonal variations and business cycle development, so that business leader may keep pace with changing economic and business conditions and have better information available for decision-making purposes.

(iv) Index numbers are very useful in deflating.

Index numbers are highly useful in deflating, i.e., they are used to adjust the original data for price changes, or to adjust wages for cost of living changes and thus transform nominal wages into real wages. Moreover, nominal income can be transformed into real income and nominal sales into real sales through appropriate index numbers.

3.8.1 Construction of Index Number

Q28. Explain the problems in the construction of index number.

Ans :

1. The Purpose of the Index

At the very outset the purpose of constructing the index must be very clearly decided - what the index is to measure and why? There is no all-purpose index. Every index is of limited and particular use. Thus, a price index that is intended to measure consumers' prices must not include wholesale prices. And if such an index is intended to measure the cost of living of poor families, great care should be taken not to include goods ordinarily used by middle class and upper-income groups. Failure to decide clearly the purpose of the index would lead to confusion and wastage of time with no fruitful results. Other problems such as the base year, the number of commodities to be included, the prices of the 'commodities, etc., are decided in the light of the purpose for which the index is being constructed.

2. Selection of a Base Period

Whenever index numbers are constructed a reference is made to some base period. The base period of an index number (also called the reference period) is the period against

which comparisons are made. It may be a year, a month or a day. The index for base period is always taken as 100. Though the selection of the base period would primarily depend upon the object of the index, the following , points need careful consideration of base period.

3. Selection of Number of Items

The items included in an index should be determined by the purpose for which the index is constructed. Every item cannot be included while constructing an index number and hence one has to select a sample. For example, while constructing a price index it is impossible to include each and every commodity. Hence it is necessary to decide what commodities to include. The commodities should be selected in such a manner that they are representative of-the tastes, habits and customs of the people for whom the index is meant.

4. Price Quotations

After the commodities have been selected, the next problem is to obtain price quotations for these commodities. It is a well known fact that prices of many commodities vary from place to place and even from shop to shop in the same market. It is impracticable to obtain price quotations from all the places where a commodity is dealt in. A selection must be made of representative places and persons.

5. Choice of an Average

Since index numbers are specialized averages a decision has to be made as to which particular average (i.e. arithmetic mean, median, mode, geometric mean or harmonic mean) should be used for constructing the index.

6. Selection of Appropriate Weights

The problem of selecting suitable weights is quite important and at the same time quite difficult to decide. The term 'weight' refers to the relative importance of the different items in the construction of the index. All items are not of equal importance and hence it is necessary to devise some suitable method whereby the varying importance of the different items is taken into account. This is done by allocating weights. Thus, we have broadly two types of indices unweighted indices and weighted indices.

7. Selection of an Appropriate Formula

A large number of formulae have been devised for constructing the index. The problem very often is that of selecting the most appropriate formula. The choice of the formula would depend not only on the purpose of the index but also on the data available. Prof. Irving Fisher has suggested that an appropriate index is that which satisfies time reversal test and factor reversal test. Theoretically. Fisher's method is considered as "ideal" for constructing index number. However, from a practical point of view there are certain limitations of this index which which shall be discussed later. As such, no one particular formula can be regarded as the best under all circumstances. On the basis of this knowledge of the characteristics of different formulae, a discriminating investigator will choose technical methods adapted to his data and appropriate to his purposes. None of the above problems is simple to solve in practice and the final index is usually the product of compromise between theoretical standards and the standards attainable with the given data.

3.8.2 Uses and Limitations of Index Number

Q29. Explain the uses and limitations of index numbers

Ans :

Uses

Index numbers are extensively used for a variety of purposes in Economics, Business Management, etc., for quantitative data relating to production, consumption, profits, personnel and financial matters as well as for comparing changes in the level of phenomenon for periods, places, etc. In fact, the importance of Index numbers stems from the fact that there is hardly any field of quantitative measurement where index numbers are not used.

The main uses of Index numbers can be summarized as below:

1. Economic Barometers

Index Numbers can be constructed for any phenomenon for which quantitative information is available. They capture the various changes taking place in the general economy and business activities. They provide a fair view of the general trade, the economic development and business activity of the country. Thus, they are aptly termed as 'Economic barometers'.

2. Study of Trend

Index Numbers study the relative changes in the level of a phenomenon over different periods of time. They are especially useful for study of general trend for a group phenomenon in a time series data.

3. Policy Formulation

Index Numbers are indispensable for any organization in efficient planning and formulation of executive decisions. For example, the Dearness Allowance payable to employees is determined on the basis of Cost of living index numbers. Similarly, Psychiatrists use Intelligence Quotients to assess a child's intelligence in relation to his age, which further can be used in framing the education policy.

4. Deflation

Index Numbers can be deflated to find out the real picture pertaining to the phenomenon. For example, we can know if the real income of employees has been increasing by deflating the nominal wages with the help of index numbers.

5. Forecasting

Index Numbers provide valuable information that aids forecasting. For example an Index of sales, along with related indices such as cost of living index, helps in forecasting future demand and future sales of business.

6. Measurement of Purchasing Power

The cost of Living Index helps us in finding out the intrinsic worth of money. It is one of the key indicators touching the life of the common man.

7. Simplicity

Index Numbers eliminate the clutter of large numbers and complicated calculations, providing the underlying information in a manner that is simple and easy to understand. For example, many people may not understand the dynamics of stock- markets, but they can follow the movements of the BSE Sensex with ease.

Limitations

Index Numbers are of great use in studying trends, events, pertaining to a defined phenomenon. However, they suffer from certain limitations. There are

1. Approximate Representation

Index Numbers are based on sample data. Hence, they are only approximate indicators. They may not fully reflect the changes in the relative level of a phenomenon.

2. Likelihood of Error

There is a possibility of error in the choice of the sample, selection of base period, data collection, and assignment of weights or the appropriate formula to be used.

3. Not Responsive

There are rapid changes in Technology, tastes, fashions, customs and consequently, the consumption pattern of various commodities. Index numbers may not be able to keep pace with such changes. Hence, they may not be able to reflect the changes in the phenomenon being studied.

4. No Universally Acceptable formula

There is no universally acceptable standard formula that can capture the changes in a phenomenon with perfection. None of the

formulae seems to satisfy all the tests of consistency. Hence, some amount of formula error is present in any calculation of Index numbers.

5. Scope for manipulation

On account of wide variety of formulae index numbers can be manipulated by appropriate choice of base year, sample set, price and quantity quotations.

6. Other Limitations

Apart from the above, Index numbers cannot capture the aspects of quality, reliability service etc. There is no index that is universally applicable. An Index constructed for one purpose cannot be used for another. Index numbers being averages, they are subject to all limitations of an average. Lastly, the utility of Index Numbers is limited by the availability of adequate and accurate data.

3.9 METHODS OF CONSTRUCTING INDEX NUMBERS

3.9.1 Simple and Weighted Index Numbers.

Q30. Explain the various methods of constructing index numbers.

Ans :

The various methods of constructing index, numbers are shown in the following figure:

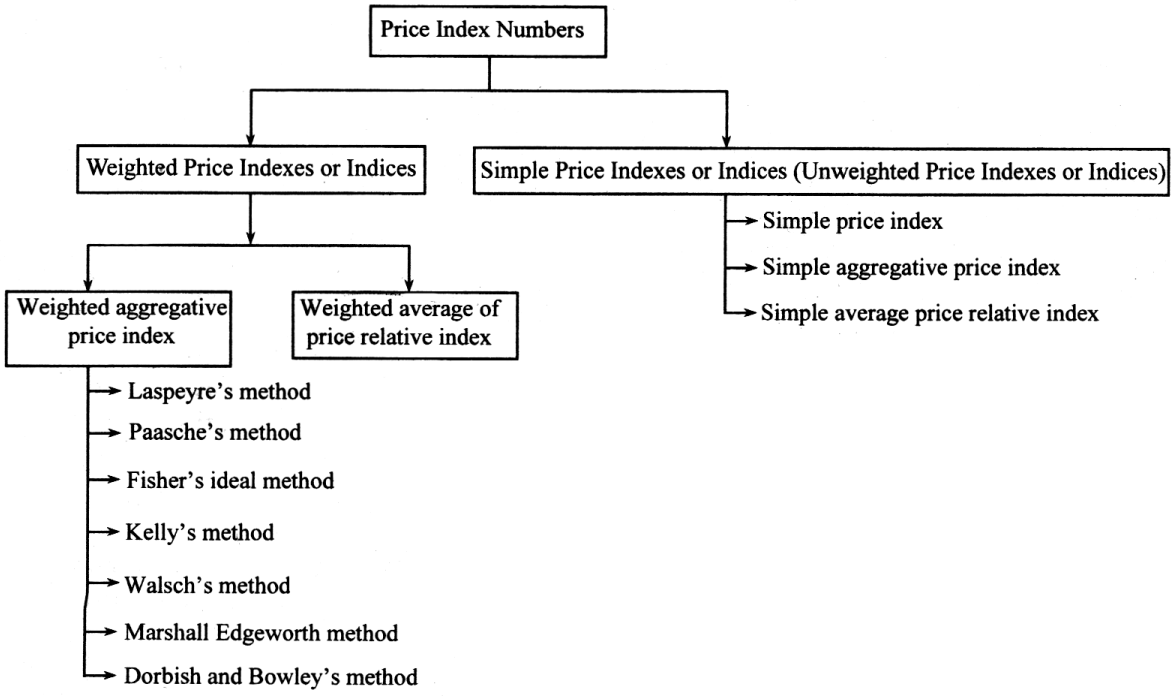


Fig : Types and Methods of Price Index Numbers

1. Weighted Price Indexes

At the time of constructing the weighted price indexes or indices, the rational weights are allocated in an explicit manner. These rational weights show the relative significance of items or commodities which are related with the computation of an index. Quantity weights and value weights are used in

this weighted indexes or indices. Weighted price indexes,or indices are further divided into two types as follows,

(a) Weighted aggregate price index

(b) Weighted average of price relative index.

(a) Weighted Aggregate Price Index

In a weighted aggregate price index, certain weight is assigned to each and every commodity or item of group in accordance with its significance. This helps in gathering more information and improving accuracy of the estimates. The following methods are used in weighted aggregate price index:

(i) Laspeyre's method

(ii) Pasche's method

(iii) Fisher's ideal method

(iv) Kelly's method

(v) Walsch's method

(vi) Marshal Edgeworth's method

(vii) Dorbish and Bowley's method

(b) Weighted Average of Price Relative Index

In weighted average of price relative index, value of each commodity or item related with the calculation of composite index is ascertained by multiplying the price of each item with its quantity consumed. Quantityconsumed is considered for computing the weighted average of price relative. The formula for weighted average of price relative index is as follows,

$$P_{01} = \frac{\Sigma\left(\left(p_1 \div p_0\right) \times 100\right)\left(p_0 q_0\right)}{\Sigma p_0 q_0} = \frac{\Sigma PV}{\Sigma V}$$

$$= \frac{\Sigma p_1 q_0}{\Sigma p_0 q_0} \times 100$$

Where,

$V(= p_0 q_0)$ = Base prices and quantities determining values

$$P\left(= \frac{p_1}{p_0} \times 100\right) = \text{Price relative}$$

This formula is equivalent to the formula of Laspeyre's index formula,

If 'V' is taken as p_0, q_1, then the formula would,

$$P_{01} = \frac{\Sigma p_1 q_1}{\Sigma p_0 q_1} \times 100$$

Then it would be equal to Paasche's index method.

2. Unweighted Price Indexes or Indices

The unweighted indexes or indices include the following methods,

(a) Simple price index

(b) Aggregate price index

(c) Average price relative index

(a) Simple/Single Price Index

Single price index is computed by dividing the current year price of the commodity with its base year price. It is a percentage ratio which represents the comparison of a particular commodity price. The general formula used for single price index is as follows,

$$\text{Single price index in period 'n'} = \frac{P_n}{P_0} \times 100$$

Where,

P_n = Price of the commodity in the n^{th} year

P_0 = Price of the commodity in the base year

(b) Simple Aggregate Price Index

In aggregate price index, the sum of current year prices of various commodities is divided with the sum of base year prices of that various commodities. The formula is given as follows,

$$\text{Aggregate price index, } P_{01} = \frac{\Sigma p_1}{\Sigma p_0} \times 100$$

Where,

P_1 = Unit price of a current year prices of all commodities

P_0 = Sum of base year prices of all commodities

(c) Simple Average Price Relative Index

The method is an improvement over the aggregate price method is,

$$P_{01} = \sum\left(\frac{P_1}{P_0} \times 100\right)$$

Where,

n = Number of commodities included in the computation of the index.

3.9.2 Laspeyre's Method

Q31. Explain Laspeyre's Method of index numbers.

Ans :

Laspeyre's Method:

This method takes the quantities of the commodities in the base period as the weight of that commodity for the purpose of calculating the index numbers. The following steps may be following.

Step I: Multiply the current year price (represented by p_0) with the quantities of the base year (q_0) for each commodity.

Step 2 : Add the numbers obtained in step 1. The resultant sum is represented as $\Sigma p_1 q_0$

Step 3: Multiply the prices of base year (represented by p_0) with the quantities of the base year for each commodity.

Step 4 : Add the numbers obtained in step 3. The resultant sum is represented as $\Sigma p_1 q_0$

Step 5 : The index number as per Laspeyre's method

$$= P_{0,1} = \frac{\Sigma p_1 q_0}{\Sigma p_0 q_0} \times 100$$

3.9.3 Paasche's Method

Q32. Explain Paasche's Method of index number.

Ans :

Paasche's Method:

Paasche's Method is similar to Laspeyre's method. The only difference is in assignment of weights. As per this method quantities consumed of the commodities in the current year is taken as basis. The following steps need to be followed.

Step 1 : Multiply current year's prices (p_1) with current year's quantities (q_1)

Step 2 : Add the numbers obtained in step (1). The resultant sum is $\Sigma p_1 q_1$

Step 3 : Multiply base year's Prices (p_0) with current year's quantities (q_1)

Step 4 : Add the numbers obtained in step (2). The resultant sum is $\Sigma p_0 q_1$

Step 5 : Index Number as per Paasche's method = $P_{0,1} = \frac{\Sigma p_1 q_1}{\Sigma p_0 q_1} \times 100$

3.9.4 Fisher's Ideal Index

Q33. Explain Fisher's Ideal Index

Ans : **(July-19)**

Fisher's Ideal Index :

This is the most popular amongst all weighted aggregative index numbers. It is obtained by calculating the Geometric Mean (G.M) of Laspeyre's and Paasche's index numbers. The formula for calculating fisher's ideal index is an under.

$$P_{0.1} = \left[\sqrt{\frac{\Sigma p_1 q_0 \times \Sigma p_1 q_1}{\Sigma p_0 q_0 \times \Sigma p_0 q_1}} \right] \times 100$$

Reasons for Fisher's Index being called an Ideal Index:

We noticed that Fisher's index is called an Ideal index. The reasons for the same are as follows:

1. It gives weightage to both current consumption and base year consumption.
2. It is free from upward or downward bias.
3. It satisfies both time reversal and factor reversal tests (to be discussed later)
4. It is a Geometric mean of Laspeyre's index and Paasche's index.

PROBLEMS ON INDEX NUMBER

1. From the following data, construct an Index number for 1994 taking 1999 as base as per simple Aggregative method

Commodities	Price in 1993 (Rs)	Price in 1994 (Rs)
A	40	60
B	60	90
C	85	125
D	25	35
E	30	40

Sol :

Construction of Price Index

Commodities	Price in 1993 in Rs. (P_0)	Price in 1994 in Rs. (P_1)
A	40	60
B	60	90
C	85	125
D	25	35
E	30	40
	$\Sigma P_0 = 240$	$\Sigma P_1 = 350$

$$P_{0,1} = \left(\frac{\Sigma p_1}{\Sigma p_0}\right) \times 100 = \frac{350}{240} \times 100 = 145.83$$

2. Following are the prices of commodities in 1970 and 1975. Calculate a price index based on price relatives using the arithmetic mean as well as geometric mean.

Year	Commodity					
	A	B	C	D	E	F
1970	45	60	20	50	85	120
1975	55	70	30	75	90	130

Sol :

(a) Arithmetic Mean :

Commodity	P_0 (Price in 1970)	P_1 (Price in1971)	Price Relative P	Log P
A	45	55	55/45 × 100 = 122.22	2.0871
B	60	70	116.67	2.0668
C	20	30	150.00	2.1761
D	50	75	150.00	2.1761
E	85	90	105.88	2.0245
F	120	130	108.33	2.0346
			ΣP = 753.10	**12.565.2**
				= ΣLogp

$$\text{Average of Price Relatives} = \frac{753.10}{6} = 125.52$$

(b) Geometric Mean

$$P_{0,1} = \text{Antilog}\left[\frac{\Sigma \text{Logp}}{N}\right] = \text{Antilog}\left(\frac{12.5652}{6}\right) = \text{Antilog}\ (2.0942) = 124.3$$

3. Construct index numbers of price from the following data by applying :

1. Laspeyres method

2. Paasche method

3. Fisher's ideal method

Commodity	1999		2000	
	Price	Quantity	Price	Quantity
A	2	8	4	6
B	5	10	6	5
C	4	14	5	10
D	2	19	2	13

Sol :

Calculation of Various Indices

Commodity	1999		2000		p_1q_0	p_0q_0	p_1q_1	p_0q_1
	Price p_0	Qty q_0	Price p_1	Qty q_1				
A	2	8	4	6	32	16	24	12
B	5	10	6	5	60	50	30	25
C	4	14	5	10	70	56	50	40
D	2	19	2	13	38	38	26	26
					200	160	130	103

1. Laspeyres Method : $P_{0,1} = \frac{\Sigma p_1 q_0}{\Sigma p_0 q_0} \times 100$; where $\Sigma p_1 q_0 = 200$, $\Sigma p_0 q_0 = 160$

$$P_{0,1} = \frac{200}{160} \times 100 = 125$$

2. Paasche's Method : $P_{0,1} = \frac{\Sigma p_1 q_1}{\Sigma p_0 q_1} \times 100$; where $\Sigma p_1 q_1 = 130$, $\Sigma p_0 q_1 = 103$

$$P_{0,1} = \frac{130}{103} \times 100 = 126.21$$

3. Fisher's Ideal Method: $P_{0,1} = \sqrt{\frac{\Sigma p_1 q_0}{\Sigma p_0 q_0} \times \frac{\Sigma p_1 q_1}{\Sigma p_0 q_1}} \times 100 = \sqrt{\frac{200}{160} \times \frac{130}{103}} \times 100$

$= \sqrt{1.578} \times 100 = 1.256 \times 100 = 125.6$

4. From the following data construct a price index number of the group of four commodities by using the appropriate formula :

Commodity	Base Year		Current Year	
	Prcie per unit	Expenditure (Rs.)	Price per unti	Expenditure (Rs.)
A	2	40	5	75
B	4	16	8	40
C	1	10	2	24
D	5	25	10	60

Sol :

Since we are given the base year and current year prices and expenditure, Fisher's ideal. Formula shall be most appropriate

Construction of Price Index

Commodity	p_0	q_0	p_1	q_1	p_1q_0	p_0q_0	p_1q_1	p_0q_1
A	2	20	5	15	100	40	75	30
B	4	4	8	5	32	16	40	20
C	1	10	2	12	20	10	24	12
D	5	5	10	6	50	25	60	30
					202	91	199	92

$$P_{0,1} = \sqrt{\frac{\Sigma p_1q_0}{\Sigma p_0q_0} \times \frac{\Sigma p_1q_1}{\Sigma p_0q_1}} \times 100$$

$$= \sqrt{\frac{202}{91} \times \frac{199}{92}} \times 100 = 2.1912 \times 100 = 219.12$$

5. Compute Index Numbers from the following data for 2010 with 2005 as base by

(i) Lasperyes Method (ii) Pasche's Method and (iii) Fisher's Ideal formula.

	2005		2010	
Commodity	**Price**	**Qty**	**Price**	**Qty**
A	**20**	**8**	**40**	**6**
B	**50**	**10**	**60**	**5**
C	**40**	**15**	**50**	**15**
D	**20**	**20**	**20**	**25**

Sol :

Comm's	p_0	q_0	p_1	q_1	p_1q_0	p_0q_0	p_1q_1	p_0q_1
A	20	8	40	6	320	160	240	120
B	50	10	60	5	600	500	300	250
C	40	15	50	15	750	600	750	600
D	20	20	20	25	400	400	500	500
					2070	**1600**	**1790**	**1470**

i) Laspayer's Method

$$P_{0,1} = \frac{\Sigma p_1q_0}{\Sigma p_0q_0} \times 100 = \frac{2070}{1660} \times 100 = 124.69$$

ii) Paasches Method

$$P_{0,1} = \frac{\Sigma p_1 q_1}{\Sigma p_0 q_1} \times 100 = \frac{1790}{1470} \times 100 = 121.77$$

iii) Fisher's ideal formula

$$P_{0,1} = \sqrt{\frac{\Sigma p_1 q_0}{\Sigma p_0 q_0} \times \frac{\Sigma p_1 q_1}{\Sigma p_0 q_1}} \times 100$$

$$= \sqrt{\frac{2070}{1660} \times \frac{1790}{1470}} \times 100$$

$$= \sqrt{1.2469 \times 1.2176} \times 100$$

$$= \sqrt{1.5182} \times 100$$

$$= 1.2321 \times 100$$

$$= 123.21.$$

6. From the following data, calculate Price Index numbers by (i) Laspeyre's (ii) Paasche's and (iii) Fisher's ideal method.

Commodity	2005		2009	
	Price `	Qty.	Price `	Qty.
A	10	4	20	3
B	25	5	30	3
C	20	8	25	8
D	10	10	10	13

Sol :

Calculate price index

Commodity	2005		2009					
	p_0	q_0	p_1	q_1	P_1q_0	$P_0 q_0$	P_1q_1	$P_0 q_1$
A	10	4	20	3	80	40	60	30
B	25	5	30	3	150	125	90	75
C	20	8	25	8	200	160	200	450
D	10	10	10	13	100	100	130	100
					530	425	480	655

i) Laspeyres ($P_{0,1}$) = $\frac{\sum P_1 q_0}{\sum P_0 q_0} \times 100$

= $\frac{530}{425} \times 100$

= 124.70

ii) Paasches ($P_{0,1}$) = $\frac{\sum P_1 q_1}{\sum P_0 q_1} \times 100$

= $\frac{480}{655} \times 100$

= 73.28

iii) Fishers ideal ($P_{0,1}$) = $\sqrt{\frac{\sum P_1 q_0}{\sum P_0 q_0} \times \frac{\sum p_1 q_1}{\sum p_0 q_1}} \times 100$

= $\sqrt{\frac{530}{425} \times \frac{480}{655}} \times 100$

= $\sqrt{1.247 \times 0.732} \times 100$

= 0.9554 × 100

= 95.54

3.9.5 Criterion of Good Index Numbers

Q34. Explain the criterion for testing the consistency of good index numbers.

Ans : **(July-19)**

1. Test

This test states that the formula of index number should be independent of the units in which the prices or quantities of various commodities (or items) are quoted. All the formulae, except the index number based on simple aggregate of prices (quantities) satisfy this test.

2. Time Reversal Test

This test was proposed by Prof. Irwin Fisher. According to fisher 'the formula for calculating the index number should be such that it gives the same ratio between one point of comparison and the other, no matter which of the two is taken as the base, or putting it another way, the index number reckoned forward should be reciprocal of the one reckoned backward.

In simple terms, given two time periods I and II, if an index number is calculated for period II taking period I as base, its value should be the reciprocal value of the index number for period I taking period II as base. The index numbers for the purpose of the test, should be in decimal form and not

in percentage form. In other words, $P_{0,1}$ and $P_{1,0}$ should not be multiplied with 100. Symbolically.

$$\boxed{P_{0,1} \times P_{1,0} = 1}$$

Where $P_{0,1}$ = Fishers index for period II taking period I as base and

$P_{0,1}$ = Fishers index for period I taking Period II as base

Time Reversal test is satisfied by Marshall-Edge worth, Fisher, Walsh. Kelly' s index numbers and also by simple Aggregative Index, simple geometric mean of price relatives and weighted average of price relatives. Laspeyre's and Paasche's index numbers do not satisfy the time reversal test.

3. Factor Reversal Test : This test was also proposed by proof Fisher. This test requires that the product of two index numbers, one measuring price taking quantities as base, and the other measuring quantities taking price as base, should be equal to the net increase in total value from one period to another. Let us illustrate the same with the help of an example.

If $P_{0,1}$is price index number, $Q_{0,1}$ is quantity index number, the product of the two index numbers should be equal to the value index number $V_{0,1}$

$$\boxed{P_{0,1} \times Q_{0,1} = V_{0,1}}$$

Fishers index number satisfies the factor reversal test. No other method satisfies the factor reversal test.

4. Circular Test: The circular test was proposed by Weztergaard. It is an extension of the time-reversal test. It more than two time periods are considered, price index is calculated for each period with the previous year as base period. Lastly, the price index for the first year is calculated taking the last period as the base. The product of all the price index numbers should be equal to 1. Symbolically, if three time peril are considered,

$$\boxed{P_{0,1} \times P_{1,2} \times P_{2,1} = 1}$$

Only simple geometric mean of price relatives method and Kelly's method Satisfy the circular test.

PROBLEMS

7. Compute Fisher's Index Number and apply both factor reversal test and time reversal test from the following data :

Commodity	Base year		Current year	
	Price Rs.	Expenditure Rs.	Price Rs.	Expenditure Rs.
A	5	25	10	60
B	1	10	2	24
C	4	16	8	40
D	2	40	5	75

Sol :

Comm's	P_0	q_0	P_1	q_1	P_1q_0	P_0q_0	P_1q_1	P_0q_1
A	5	5	10	6	50	25	60	30
B	1	10	2	12	20	10	24	12
C	4	4	8	5	32	16	40	20
D	2	20	5	15	100	40	75	30
					202	91	199	92

Fishers Ideal Index :

$$P_{0,1} = \sqrt{\frac{\sum p_1q_0}{\sum p_0q_0} \times \frac{\sum p_1q_1}{\sum p_0q_1}} \times 100$$

$$= \sqrt{\frac{202}{91} \times \frac{199}{92}} \times 100$$

$$= \sqrt{\frac{40,198}{8372}} \times 100 = \sqrt{4.8015} \times 100$$

$$= 2.191 \text{ x } 100 = 219.1$$

Time Reversal Test :

$$P_{0,1} = \sqrt{\frac{\sum p_1q_0}{\sum p_0q_0} \times \frac{\sum p_1q_1}{\sum p_0q_1}}$$

$$P_{1,0} = \sqrt{\frac{\sum p_0q_1}{\sum p_1q_1} \times \frac{\sum p_0q_0}{\sum p_1q_0}}$$

$$P_{0,1} \text{ x } P_{1,0} = \sqrt{\frac{\sum p_1q_0}{\sum p_0q_0} \times \frac{\sum p_1q_1}{\sum p_0q_1} \times \frac{\sum p_0q_1}{\sum p_1q_1} \times \frac{\sum p_0q_0}{\sum p_1q_0}}$$

$$= \sqrt{\frac{202}{91} \times \frac{199}{92} \times \frac{92}{199} \times \frac{91}{202}} = \sqrt{1} = 1$$

$$\therefore P_{0,1} \times P_{1,0} = 1$$

Factor Reversal Test

$$P_{0,1} \text{ x } q_{0,1} = \sqrt{\frac{\sum p_1q_0}{\sum p_0q_0} \times \frac{\sum p_1q_1}{\sum p_0q_1} \times \frac{\sum p_0q_1}{\sum p_0q_0} \times \frac{\sum p_1q_1}{\sum p_1q_0}}$$

$$= \sqrt{\frac{202}{91} \times \frac{199}{92} \times \frac{92}{91} \times \frac{199}{202}} = \sqrt{\frac{199}{91} \times \frac{199}{91}} = \frac{199}{91}$$

$$V_{0,1} = \frac{\Sigma p_1 q_0}{\Sigma p_0 q_0} = \frac{199}{91}$$

$$\therefore P_{0,1} \times Q_{0,1} = V_{0,1}$$

8. Calculate Fisher's price index from the following data and check whether Time Reversal Test is satisfied :

Commodity	Base Year		Current Year	
	Price (Rs.)	Qty. (Kg.)	Price (Rs.)	Qty (Kg.)
A	32	50	30	50
B	30	35	25	40
C	16	55	18	50

Sol :

Calculation of Fisher's Price Index

Commodity	p_o	q_o	p_1	q_1	$p_o q_o$	$p_1 q_1$	$p_o q_1$	$p_1 q_o$
A	32	50	30	50	1,600	1,500	1,600	1,500
B	30	35	25	40	1,050	1,000	1,200	875
C	16	55	18	50	880	900	800	990
					3,530	3,400	3,600	3,365

Fisher's Ideal Index $= P_{0,1} = \sqrt{\frac{\Sigma p_1 q_o}{\Sigma p_o q_o} \times \frac{\Sigma p_1 q_1}{\Sigma p_o q_1}} \times 100$

$$= \sqrt{\frac{3,365}{3,530} \times \frac{3,400}{3,600}} \times 100$$

$$= \sqrt{0.95 \times 0.094} \times 100$$

$$= \sqrt{0.0893} \times 100 = 0.298 \times 100$$

$$= 29.8$$

Time Reversal Test $= P_{0,1} \times P_{1,0} = 1$

$$= \sqrt{\frac{\Sigma p_1 q_o}{\Sigma p_o q_o} \times \frac{\Sigma p_1 q_1}{\Sigma p_o q_1}} \times \sqrt{\frac{\Sigma p_o q_1}{\Sigma p_1 q_1} \times \frac{\Sigma p_o q_o}{\Sigma p_1 q_o}}$$

$$= \sqrt{\frac{3,365}{3,530} \times \frac{3,400}{3,600}} \times \sqrt{\frac{3,600}{3,400} \times \frac{3,530}{3,365}}$$

$= \sqrt{1} = 1$

Hence Time Reversal Test is Satisfied.

9. Calculate Fisher's ideal index from the following data and prove that it satisfies both the time reversal and factor reversal tests :

Commodity	2000		2001	
	Price	Expenditure	Price	Expenditure
A	8	80	10	120
B	10	120	12	96
C	5	40	5	50
D	4	56	3	60
E	20	100	25	150

Sol :

Calculation of Fisher's Ideal Index

Commodity	2000		2001					
	p_0	q_0	p_1	q_1	p_1q_0	p_0q_0	p_1q_1	p_0q_1
A	8	10	10	12	100	80	120	96
B	10	12	12	8	144	120	96	80
C	5	8	5	10	40	40	50	50
D	4	14	3	20	42	56	60	80
E	20	5	25	6	125	100	150	120
					451	396	476	426

$$P_{0,1} = \sqrt{\frac{\Sigma p_1q_0}{\Sigma p_0q_0} \times \frac{\Sigma p_1q_1}{\Sigma p_0q_1}} \times 100 = \sqrt{\frac{451}{396} \times \frac{476}{426}} \times 100$$

$$\sqrt{1.2726} \times 100 = 1.128 \times 100 = 112.8$$

Time Reversal Test : Time reversal test is satisfied when $P_{0,1} \times P_{1,0} = 1$

$$P_{1,0} = \sqrt{\frac{\Sigma p_0q_1}{\Sigma p_1q_1} \times \frac{\Sigma p_0q_0}{\Sigma p_1q_0}} = \sqrt{\frac{426}{476} \times \frac{396}{451}}$$

$$P_{0,1} \times P_{1,0} = \sqrt{\frac{451}{396} \times \frac{476}{426} \times \frac{426}{476} \times \frac{396}{451}} = \sqrt{1} = 1$$

Hence, time reversal test is satisfied.

Factor Reversal Test : Factor reversal test is satisfied when :

$$P_{0,1} \times Q_{0,1} = \frac{\Sigma p_1 q_1}{\Sigma p_0 q_0}$$

$$Q_{0,1} = \sqrt{\frac{\Sigma q_1 p_0}{\Sigma q_0 p_0} \times \frac{\Sigma q_1 p_1}{\Sigma q_1 p_0}}$$

$$P_{0,1} \times Q_{0,1} = \sqrt{\frac{451}{396} \times \frac{476}{426} \times \frac{426}{396} \times \frac{476}{451}} = \frac{476}{396}$$

This is also the value of $\frac{\Sigma p_1 q_1}{\Sigma p_0 q_0}$. Hence, the above data also satisfies the Factor Reversal Test.

10 Construct a Fisher's ideal index from the following data and show that it satisfies time reversal and factor reversal tests :

Items	1995		1996	
	p_0	q_0	p_1	q_1
A	10	40	12	45
B	11	50	11	52
C	14	30	17	30
D	8	28	10	29
E	12	15	13	20

Sol :

Construction of Fisher's Ideal Index

Items	p_0	q_0	p_1	q_1	p_1q_0	$p_0 q_0$	$p_1 q_1$	p_0q_1
A	10	40	12	45	480	400	540	450
B	11	50	11	52	550	550	572	572
C	14	30	17	30	510	420	510	420
D	8	28	10	29	280	224	290	232
E	12	15	13	20	195	180	260	240
					2015	1774	2172	1914

Fisher's Ideal Index : $P_{0,1} = \sqrt{\frac{\Sigma p_1 q_0}{\Sigma p_0 q_0} \times \frac{\Sigma p_1 q_1}{\Sigma p_0 q_1}}$

$$= \sqrt{\frac{2015}{1774} \times \frac{2172}{1914}} \times 100 = 1.135 \times 100 = 113.5$$

Time Reversal Test : Time reversal test is satisfied when :

$$P_{0,1} \times P_{1,0} = 1$$

$$P_{1,0} = \sqrt{\frac{\Sigma p_0 q_1}{\Sigma p_1 q_1} \times \frac{\Sigma p_0 q_0}{\Sigma p_1 q_0}} = \sqrt{\frac{1914}{2172} \times \frac{1774}{2015}}$$

$$P_{0,1} \times P_{1,0} = \sqrt{\frac{2015}{1774} \times \frac{2172}{1914} \times \frac{1914}{2172} \times \frac{1774}{2015}}$$

Hence, time reversal test is satisfied by the given data.

Factor Reversal Test : Factor reversal test is satisfied when :

$$P_{0,1} \times q_{0,1} = \frac{\Sigma p_1 q_1}{\Sigma p_0 q_0}$$

$$q_{0,1} = \sqrt{\frac{\Sigma q_1 p_0}{\Sigma q_0 p_0} \times \frac{\Sigma q_1 p_1}{\Sigma q_0 p_1}} = \sqrt{\frac{1914}{1774} \times \frac{2172}{2015}}$$

$$P_{0,1} \times q_{0,1} = \sqrt{\frac{2015}{1774} \times \frac{2172}{1914} \times \frac{1914}{1774} \times \frac{2172}{2015}} = \frac{2172}{1774}$$

$\frac{\Sigma p_1 q_1}{\Sigma p_0 q_0}$ is also equal to $\frac{2172}{1774}$. Hence factor reversal test is satisfied by the given data.

3.9.6 Problems Involved in Construction of Index Numbers

Q35. What are the problems Involved in Construction of Index Numbers

Ans : **(Dec.-18(MGU)**

Construction of index numbers involves resolving a few problems, which are described below :

1. Purpose of an Index

Every index is of limited and specific use. There is no all purpose Index available. Thus, great care needs to be taken to select or build an appropriate index to measure a defined phenomenon. Failure to decide clearly the purpose of the index would lead to confusion and wastage of time with no fruitful results.

2. Selection of Base Period

The base period refers to that time period with which comparisons of relative changes are made. The period that is selected as base should be a 'normal' year. In other words, it should be free from abnormalities like wars, earthquakes, famines, booms, depressions, etc. Sometimes, it is really difficult to select a year which is normal in all respects. Secondly, the base period should not be too distant in the past as decisions cannot be taken based on the situation that was prevalent very long ago. Lastly, the user must have clarity as to whether the base would be fixed or will he be following a Chain Base Index (will be explained later).

3. Constituents of an Index

The next problem is with respect to what will constitute the Index. The first issue is of sample size. In other words, how many items should be included in the Index? The sample size should neither be too small (to avoid one or two items having a huge impact on the index) nor should it be too large (to avoid the impact getting diluted). The second issue is in deciding which items would constitute the index. Ideally, the items that are most representative of the phenomenon being studied should constitute the index. Lastly, the grade or quality of the item included should be kept constant. In other words, you cannot take the price of Sona Masoori rice in one year and Basmati rice in another year and say that "rice" is a constituent of the index.

4. Price Quotations

Prices of various items included in the index may vary from one market to another. It is not practicable to obtain price quotations from all the markets. Thus, prices should be taken from markets which are large, liquid markets, facilitating a high volume of trade in that particular item. It is preferable to identify an official source or a reliable agency to provide the prices at agreed frequency.

5. Method of Calculating the Index

Once the index is constituted, the next problem is to decide the most appropriate formula for calculating the Index. Which average should be used to calculate the index? While Median, Mode and Mean are almost never used, the Geometric Mean is ideally the best but Arithmetic Mean is most popular due to its simplicity and ease of calculation. Another decision to be taken at the time of constructing the index is the usage of weights. Should simple averages be used or should weighted averages be used? If weighted averages are to be calculated, assignment of weights will involve considerable amount of skill. Lastly, a large number of methods have been presented by eminent statisticians and it is often very difficult to select the formula that is most appropriate. The Fisher's index is considered as an Ideal index but it has its own limitations.

Thus, construction of an Index is a complex process. It must be driven by the objective for which the index is being constructed.

3.10 FIXED AND CHAIN BASED INDEX NUMBERS

Q36. Define Chain Index Numbers. Explain the steps in constructing a chain index numbers.

Ans :

In the fixed base method, discussed so far the base remains the same Throughout the series of the index. This method, though convenient, has certain limitations. As time elapses conditions which were once important become less significant and it becomes more difficult to compare accurately present conditions with those of a remote period. New items may have to be included and old ones may have to be deleted in order to make the index more representative. In such cases it may be desirable to use the chain base method.

Steps in Constructing a Chain Index

In constructing a chain index following steps are desirable:

(i) Express the figures for each year as percentages of the preceding year.

The results so obtained are called link relatives.

(ii) Chain together these percentages by successive multiplication to form a chain index. Chain index of any year is the average link relative of that year multiplied by chain index of previous year divided by 100. In the form of formula

Chain Index for current year

$$= \frac{\text{Average link relative of current year} \times \text{Chain Index of previous year}}{100}$$

The link relatives obtained in step (i) facilitate comparison from one year to another, i.e., between closely situated periods in which the *q's* are a process of chaining binary comparisons facilitate long-term comparisons.

Chain relatives differ from fixed-base relatives in computation. Chain relatives are computed from link relatives whereas fixed base relatives are computed directly from the original data. The results obtained by the two different methods should be the same, but they may differ from each other slightly due to rounding off of decimal places. Since the process of computing chain relatives is quite complicated and the results are same as the fixed base relatives obtained from the original data, chain relatives should be used when the original data are not available but the link relatives are.

Q37. Explain the merits and demerits of chain based methods.

Ans :

Merits of the Chain Base Method

The merits of this method are numerated here:

1. The chain base method has a great significance in practice because in economic and business data, we are more often concerned with making comparisons with the previous period and not with any distant past. The link relatives obtained by chain base method serve this purpose.

2. Chain base method permits the introduction of new commodities and the detection of old ones without necessitating either the recalculation of entire series or other drastic changes. Because of this flexibility, chain index is used in many types of indices such as the consumer price index and the wholesale price index.

3. Weights can be adjusted as frequently as possible. This flexibility is of great significance in many types of index numbers.

4. Index numbers calculated by the chain base method are free to a greater extent from seasonal variations than these obtained by the other method.

Limitations of the Chain Index

The main limitation of the chain index is that while the percentages of previous year figures give accurate comparisons of year-to-year changes, the long-range comparisons of chained percentages are not strictly valid. However, when the index number user wishes to make year-to-year comparisons, as is so often done by the businessman, the percentages of the preceding year provide a flexible and useful too.

Q38. Explain the conversion of chain index to fixed index.

Ans :

At times it may convert the chain base index numbers into fixed base In such a case the following procedure is followed :

1. For the first year the fixed base index will be taken the same as the chain base index. However, if the index numbers are to be constructed by taking first year as the base in that case the index for the first year is taken as 100.

2. For calculating the indices for other years the following formula is used :

$$\text{Current year s F.B.I.} = \frac{\text{Current year's C.B.I.} \times \text{Previous year's F.B.I}}{100}$$

F. B. I = Fixed base index no. ; C.B.I. = Chain base index no.

PROBLEMS ON CHAIN INDEX NUMBERS

11. From the following data of the wholesale prices of wheat for ten years construct index numbers taking (a) 1989 as base, and (b) by chain base method :

Year	Price of Wheat (Rs. per 40 Kg.)	Year	Price of Wheat (Rs. per 40 Kg.)
1989	50	1994	78
1990	60	1995	82
1991	62	1996	84
1992	65	1997	88
1993	70	1998	90

Sol :

(a) Construction of Index Numbers Taking 1989 as Base

Year	Price of Wheat	Index Number (1989 = 100)
1989	50	100
1990	60	$\frac{60}{50} \times 100 = 120$
1991	62	$\frac{62}{50} \times 100 = 124$
1992	65	$\frac{65}{50} \times 100 = 130$
1993	70	$\frac{70}{50} \times 100 = 140$
1994	78	$\frac{78}{50} \times 100 = 156$
1995	82	$\frac{82}{50} \times 100 = 164$
1996	84	$\frac{84}{50} \times 100 = 168$
1997	88	$\frac{88}{50} \times 100 = 176$
1998	90	$\frac{90}{50} \times 100 = 180$

This means that from 1989 to 1990 there is a 20 per cent increase; from 1990 to 1991 there is a 24 per cent increase, from 1991 to 1992 there is a 30 per cent increase. If we are interested in finding out increase from 1989 to 1990, from 1990 to 1991, from 1991 to 1992, we shall have to compute the chain indices.

(b) Construction of Chain Indices

Year	Price of Wheat	Link Relatives	Chain Indices (1989 = 100)
1989	50	100.00	100.00
1990	60	$\frac{60}{50} \times 100 = 120.00$	$\frac{120 \times 100}{100} = 120$
1991	62	$\frac{62}{60} \times 100 = 103.33$	$\frac{103.33 \times 120}{100} = 124$
1992	65	$\frac{65}{62} \times 100 = 104.84$	$\frac{104.84 \times 124}{100} = 130$
1993	70	$\frac{70}{65} \times 100 = 107.69$	$\frac{107.69 \times 130}{100} = 140$
1994	78	$\frac{78}{70} \times 100 = 111.43$	$\frac{111.43 \times 140}{100} = 156$
1995	82	$\frac{82}{78} \times 100 = 105.13$	$\frac{105.13 \times 156}{100} = 164$
1996	84	$\frac{84}{82} \times 100 = 102.44$	$\frac{102.44 \times 164}{100} = 168$
1997	88	$\frac{88}{84} \times 100 = 104.76$	$\frac{104.76 \times 168}{100} = 176$
1998	90	$\frac{90}{88} \times 100 = 102.27$	$\frac{102.27 \times 176}{100} = 180$

Note :

The chain indices obtained in (b) above with 1989 = 100 are the same as the fixed base indices obtained in (a) above. In fact chain index figure will always be equal to fixed base index figure if there is only one series.

12. Compute the chain index number with 1995 prices as base from the following table giving the average wholesale prices of the commodities A, B and C for the year 1996.

Commodity	Average wholesale price (in Rs.)				
	1995	1996	1997	1998	1999
A	20	16	28	35	21
B	25	30	24	36	45
C	20	25	30	24	30

Sol :

Computation of Chain Indices

Commodity	Relatives based on precending year				
	1995	**1996**	**1997**	**1998**	**1999**
A	100	$\frac{16}{20}\times 100 = 80$	$\frac{28}{16}\times 100 = 175$	$\frac{35}{28}\times 100 = 125$	$\frac{21}{35}\times 100 = 60$
B	100	$\frac{30}{25}\times 100 = 120$	$\frac{24}{30}\times 100 = 80$	$\frac{36}{24}\times 100 = 150$	$\frac{45}{36}\times 100 = 125$
C	100	$\frac{25}{20}\times 100 = 125$	$\frac{30}{25}\times 100 = 120$	$\frac{24}{30}\times 100 = 80$	$\frac{30}{24}\times 100 = 125$
Total of link Relatives	300	325	375	355	310
Average of link Relatives	100	108.33	125	118.33	103.33
Chain Index (1995 = 100)	100	$\frac{108.33\times 100}{100}$ = 108.33	$\frac{125\times 108.33}{100}$ = 135.41	$\frac{118.33\times 135.41}{100}$ = 160.23	$\frac{103.33\times 160.23}{100}$ = 165.57

13. From the chain base index numbers given below prepare fixed base index numbers:

1991	**1992**	**1993**	**1994**	**1995**
80	**110**	**120**	**90**	**140**

Sol :

Computation of Fixed Base Index Numbers

Year	Chair base index number	Fixed base index numbers
1991	80	80
1992	110	$\frac{110\times 80}{100} = 88.00$
1993	120	$\frac{120\times 88}{100} = 105.60$
1994	90	$\frac{90\times 105.6}{100} = 95.04$
1995	140	$\frac{140\times 95.04}{100} = 133.06$

3.11 Cost of Living Index Numbers

Q39. What do you understand by consumer price index ?

Ans : **(July-21, Dec.-18(MGU)**

Consumer Price Index (CPI)

The Consumer Price Index (CPI) is a measure of the average change over time in the prices paid by urban consumers for a market basket of consumer goods and services.

The CPI is widely used as an indicator of the change in the general level of consumer prices or the rate of inflation. Since the purchasing power of money is affected by changes in prices, the CPI is useful to virtually all Canadians. Consumers can compare movements in the CPI to changes in their personal income to monitor and evaluate changes in their financial situation.

The CPI affects nearly all Americans because of the many ways it is used. Three major uses are :

1. As an economic indicator

The CPI is the most widely used measure of inflation and is sometimes viewed as an indicator of the effectiveness of government economic policy. It provides information about price changes in the nation's economy to government, business, labour, and other private citizens, and is used by them as a guide to making economic decisions. In addition, the President, Congress, and the Federal Reserve Board use trends in the CPI to aid in formulating fiscal and monetary policies.

2. As a deflator of other economic series

The CPI and its components are used to adjust other economic series for price changes and to translate these series into inflation-free dollars. Examples of series adjusted by the CPI include retail sales, hourly and weekly earnings, and components of the national income and product accounts. An interesting example of this is the use of the CPI as a deflator of the value of the consumer's dollar to find its purchasing power. The purchasing power of the consumer's dollar measures the change in the value to the consumer of goods and services that a dollar will buy at different dates. In other words, as prices increase, the purchasing power of the consumer's dollar declines.

3. As a means of adjusting dollar values

The CPI is often used to adjust consumers' income payments, (for example, Social Security); to adjust income eligibility levels for government assistance; and to automatically provide cost-of-living wage adjustments to millions of American workers. The CPI affects the income of about 80 million persons as a result of statutory action: 48.4 million Social Security beneficiaries, about 19.8 million food stamp recipients, and about 4.2 million military and federal Civil Service retirees and survivors. Changes in the CPI also affect the cost of lunches for 26.5 million children who eat lunch at school, while collective bargaining agreements that tie wages to the CPI cover over 2 million workers. Another example of how dollar values may be adjusted is the use of the CPI to adjust the federal income tax structure. These adjustments prevent inflation-induced increases in tax rates, an effect called "bracket creep".

Q40. Explain the methods of Consumer Price Index Numbers.

Ans :

There are two methods for the compute of consumer price index numbers.

(a) Aggregate Expenditure method

(b) Family Budget Method

Cost of Living Index Numbers are weighted index numbers. The commodities constitute the index are given weights according to their importance. Normally, the weights are in the ratio of amounts spent on each item. There are two methods of constructing the cost of Living Index Numbers.

(a) Aggregate Expenditure Method or Weighted Aggregative :

Thus method is similar to Laspeyre's method. The quantities consumed in the base year are taken as weights. The formula is :

$$\text{Consumer Price Index = Cost of Living Index } = \frac{\Sigma p1q0}{\Sigma p0q1} \times 100$$

Since P represents price ans q represents quantity. pq is the amount spent of given commodity. Thus $\sum pq$ represents total amount spent on all items. In other words it represents total expenditure. $\sum p_0q_0$ is total expenditure incurred in the base period. $\sum p_1q_0$ is total expenditure in the current year at base period. Thus

$$\text{Consumer Price Index } = \frac{\text{Total Expenditure Current Year at Base Year Price}}{\text{Total Expenditure in Base Year}} \times 100$$

(b) Family Budget Method or Method of Weighted Relatives :

The cost of living index is obtained by taking a weighted average of price relatives. The quantities consumed in the base year are taken as weights.

The formula is :

$$\text{Cost of Living Index } = \frac{\sum PV}{\sum V} \text{ Where}$$

P = $(p_1/p_0) \times 100$ for each item and V = Values Weight = 100

$$\text{Thus, Cost of Living Index } = \sum \frac{\left(\frac{p1}{p0} \times 100\right) \times p0q0}{\sum p0q0} = \frac{\sum p1q0}{\sum p0q0} \times 100$$

Thus, the cost of living index figure is one and the same, irrespective of method of construction.

PROBLEMS ON CONSUMER PRICE INDEX

14. Calculate the index number using both the Aggregate Expenditure method and Family Budget method for the year 1973 with 1960 as base year from the following data.

Commodity	Quantity in units in 1960	Price per units in 1960 (Rs)	Price per units in 1973 (Rs)
A	100	8.00	12.00
B	25	6.00	7.50
C	10	5.00	5.25
D	20	48.00	52.00
E	25	15.00	16.50
F	30	9.00	27.00

Sol :

Calculation of Consumer Price Index

Commodity	Quantity (q_0)	Price (p_0)	p_0q_0 / p_1	p_0q_0 = V	p_1q_0 × 100	$P=(P_1/P_0)$	PV
A	100	8.00	12.00	800	1200	150	120000
B	25	6.00	7.50	150	187.50	125	18750
C	10	5.00	5.25	50	52.50	105	52.50
D	20	48.00	52.00	960	1040.00	108.33	1040000
E	25	15.00	16.50	375	412.50	110	41250
F	30	9.00	27.00	270	810.00	300	81000
				2605	3702.50		370205

(1) Aggregate Expenditure Method :

$$CP_1 = \frac{\sum p1q0}{\sum p0q0} \times 100 = \frac{3702.05}{2605} \times 100 = 142.13$$

(2) Family budget Method :

$$CP_1 = \frac{\sum PV}{\sum V} = \frac{370250}{2605} = 142.13$$

15. In the construction of a certain cost of living number, the following group index numbers are found. Calculate the Cost of Living Index Number by using (i) Weighted Arithmetic Mean and (ii) Weighted Geometric Mean.

Group	Index Numbers	Weights
Food	350	5
Fuel and Lighting	200	1
Clothing	240	1
House Rent	160	1
Miscellaneous	250	2

Sol :

Computation of Consumer Price Index

Group	Index No. (I)	Weights (W)	Weighted (WI)	Log I	W.log I
Food	350	5	1750	2.5441	12.7205
Fuel Lighting	200	1	200	2.3010	2.3010
Clothing	240	1	240	2.3802	2.3802
House Rent	160	1	160	2.2041	2.2041
Miscellaneous	250	2	500	2.3979	4.7958
		10	2850		24.4016

Consumer Price Index :

(i) Using Arthimethic Mean $= \frac{\sum IW}{\sum W} = \frac{2850}{10} = 285$

(ii) Using Geometric Mean = Antilog $\left[\frac{\sum W \log I}{\sum W}\right] = \text{anti log}\left[\frac{24.4016}{10}\right]$

= Antilog (2.44016) = 275.55

3.12 WHOLESALE PRICE INDEX NUMBERS

Q41. Define Wholesale Price Index Number. State various series of Wholesale Price Index Number.

Ans : **(Nov.-20)**

Wholesale Price Index Number

The general change occurred in the wholesale prices of the commodities which are represented in index numbers is termed as wholesale price index number. Wholesale price indices are used to study the changes that take place in the general price level of a country.

The Indian Ministry of Commerce and Industry constructed the first wholesale price index number in January 1947 with 1939 as base year.

Various Wholesale Price Index Numbers

On the recommendation of Wholesale Industrial Price Review Committee, the series of construction of index number was started. They are discussed as below,

(i) Revised index number of wholesale prices

(ii) New series of index numbers of wholesale prices

(iii) Series of index numbers of wholesale prices

(iv) Wholesale price index number series

(v) Current series of wholesale price indices

(i) Revised Index Number of Wholesale Prices

The revised index number of wholesale prices was constructed in 1956 with (1952-53 = 100) as the base year. According to the Standard International Trade, the revised index number of wholesale price is based on 112 commodities which are classified into 6 groups. It comprises of 555 individual quotations. The groups for the construction of revised index number of wholesale prices are shown as below,

(a) Food

(b) Liquor and Tobacco

(c) Fuel, Power, Light and Lubricants

(d) Industrial Raw Materials

(e) Manufactured Articles:

- Intermediate Products
- Finished Products.

(ii) New Series of Index Numbers of Wholesale Prices

On the recommendations of 'wholesale price index revision committee' a new series on index numbers of wholesale prices in India is constructed with (1961-62 = 100) as base year. This new series of index numbers is also computed as weighted arithmetic mean of the price relatives. It consists of 139 commodities which comprises of 774 quotations. The groups for the construction of new series of index numbers of wholesale prices are shown as below,

(a) Food Articles

(b) Liquor and Tobacco

(c) Fuel, Power, Light and Lubricants

(d) Industrial Raw Materials

(e) Chemicals

(f) Machinery and Transport

(g) Manufactured Articles,

- Intermediate Products
- Finished Products.

(iii) Series of Index Numbers of Wholesale Prices

The series of Index numbers of wholesale prices was constructed since the first week of January 1977. It comprises of enlarged coverage and improved weighted system. According to this new classification, the commodities are grouped into three major groups as shown below,

I. Primary Articles,

(a) Food Articles

(b) Non-food Articles

(c) Minerals.

II. Fuel, Power, Light and Lubricants

III. Manufactured Products.

The Whole Price Index (WPI) (1970 - 71 = 100) is calculated on the principle of weighted arithmetic mean with accordance to the Laspeyre's formula in its modified form,

(iv) Wholesale Price Index Number Series

The wholesale price index number is constructed in July 1989 with (1981-82 = 100) as a base year. The commodities under this series are classified into 3 groups and 16 sub-groups. It covers around a total of 447 commodities as against 360 commodities in the previous WPI series with base 1970-71. The commodities are grouped into three major groups as shown below,

I. Primary Articles,

(a) Food Articles

(b) Non-food Articles

(c) Minerals

II. Fuel, Power, Light and Lubricants

III. Manufactured Products.

(v) Current Series of Wholesale Price Indices

The current series of wholesale price indices is constructed with (1993-94 = 100) as base year. WPI series (1981-82 = 100) was revised again because it was getting quite distant which adequately reflects the economic growth in various sectors of the economy. So, the current WPI consists of various commodities which are divided into 3 major groups and 16 sub-groups as shown below,

I. Primary Articles,

1. Food Articles
2. Non-food Articles
3. Minerals.

II. Fuel, Power, Light and Lubricants

III. Manufactured Products,

1. Food Products
2. Beverages, Tobacco and Tobacco Products
3. Textiles
4. Wood and Wood Products
5. Leather and Leather Products
6. Paper and Paper Products
7. Chemical and Chemical Products

8. Rubber and Plastic Products
9. Non-metallic Mineral Products
10. Basic Metals, Alloys and Metal Products
11. Machinery and Machine tools including electrical machinery
12. Transport equipment and parts.

3.13 BASE SHIFTING

Q42. What do you understand by base shifting of index numbers ?

Ans : **(Dec.-18)**

For a variety of reasons, it frequently becomes necessary to change the reference base of an index number series from one time period to another without returning to the original raw data and recomputing the entire series. This change of reference base period is usually referred to as "shifting the base". There are two important reasons for shifting the base:

1. The previous base has become too old and is almost useless for purposes of comparison. By shifting the base it is possible to state the series in terms of a more recent time period.
2. It may be desired to compare several index number series which have been computed on different base periods; particularly if the several series are to be shown on the same graph, it may be desirable for them to have the same base period. This may necessitate a shift in the base period.

When base period is to be changed, one possibility is to recompute all index numbers using the new base period. A simpler approximate method is to divide all index numbers for the various years corresponding to the old base period by the index number corresponding to the new base period, expressing the results as percentages. These results represent the new index numbers, the index number for the new base period being 100 per cent. Mathematically speaking, this method is strictly applicable only if the index numbers satisfy the circular test. However, for many types of index numbers the method, fortunately, yields results which in practice are close enough to those which would be obtained theoretically.

PROBLEMS ON BASE SHIFTING

16. The following are the index numbers of prices (1990 = 100) :

Year	Index	Year	Index
1990	100	1995	410
1991	110	1996	400
1992	120	1997	380
1993	200	1998	370
1994	400	1999	340

Shift the base from 1990 to 1996 and recast the index numbers.

Sol :

Index Numbers with 1990 As Base (1990 = 100)

Year	Index Numbers (1990 = 100)	Index Number (1996 = 100)
1990	100	$\frac{100}{400} \times 100 = 25.0$
1991	110	$\frac{110}{400} \times 100 = 27.5$
1992	120	$\frac{120}{400} \times 100 = 30.0$
1993	200	$\frac{200}{400} \times 100 = 50.0$
1994	400	$\frac{400}{400} \times 100 = 100.0$
1995	410	$\frac{410}{400} \times 100 = 102.5$
1996	400	$\frac{400}{400} \times 100 = 100.0$
1997	380	$\frac{380}{400} \times 100 = 95.0$
1998	370	$\frac{370}{400} \times 100 = 92.5$
1999	340	$\frac{340}{400} \times 100 = 85.0$

The new series with 1996 as base is obtained very easily by dividing each entry of the first column by 400, i.e., the values of the index for 1996 and multiplying the ratio by 100. Thus,

$$\text{Index number for 1990} = \frac{\text{Index number for 1996}}{\text{Index number for 1990}} \times 100 = \frac{100}{400} \times 100 = 25$$

$$\text{Index number for 1991} = \frac{\text{Index number for 1991}}{\text{Index number for 1996}} \times 110 = \frac{100}{400} \times 110 = 27.5$$

In a similar manner other indices can also be obtained.

17. The following are the index numbers of wholesale prices of a commodity based on 1990 :

Year	Index Number	Year	Index Number
1993	100	1997	210
1994	108	1998	225
1995	120	1999	240
1996	150		

Prepare New Index Numbers taking 1995 as base.

Sol :

Shifting the Base from 1993 to 1995

Year	Index Numbers (1993 = 100)	Index Number (1995 = 100)
1993	100	83.33
1994	108	90.00
1995	120	100
1996	150	125
1997	210	175.5
1998	225	187.5
1999	240	200.0

$$\text{Index number for 1993} = \frac{\text{Index number for 1993}}{\text{Index number for 1995}} \times 100 = \frac{100}{120} \times 100 = 83.33$$

$$\text{Index number for 1994} = \frac{108}{120} \times 100 = 90 \text{ etc.}$$

3.14 SPLING OF INDEX NUMBER

Q43. Define Splicing.

Ans : **(Dec.-18)**

Sometimes an index number series is available for a period of time. and then undergoes substantial revision including a shift in the reference period.

For example, the weights of an index number may become out of date and we may construct another index with new weight. Thus two indices would result. At times, it may be necessary to convert these two indices into a continuous series. The procedure employed for this conversion is called splicing

The process of splicing is very simple and is akin to that used in shifting the base. It is expressed in the form of a formula as follows :

$$\text{Spliced Index No.} = \frac{\text{Index No. of current year} \times \text{Old Index of New Base Year}}{100}$$

The following example would illustrate the procedure.

Q44. What are the different types of splicing ?

Ans :

Types of Splicing

Generally, splicing is categorized into two types and they are as follows,

1. Backward splicing
2. Forward splicing.

1. Backward Splicing

Backward splicing is used for splicing a new series of indices to continue with old series of indices.

Formula

$$\text{Back ward Splicing} = \frac{\text{Index A of Current Year}}{\text{Index A of Common Year}} \times 100$$

2. Forward Splicing

Forward splicing is used for splicing old series of indices to continue with new series of indices.

Formula

$$\text{Forward Splicing} = \frac{\text{Index B of Current Year} \times \text{Index A of Common Year}}{100}$$

PROBLEM ON SPLICING

18. The following table gives two series of index numbers with 2005 and 2010 as base. Obtain a continuous series of index number by considering 2010 as base year.

Years	2005	2006	2007	2008	2009	2010	2011	2012	2013	2014
Index No. (2005 = 200)	200	240	250	280	300	320				
Index No. (2010 = 200)						200	220	240	260	280

Sol :

Years	Index Number (2005 = 200)	Index Number (2010 = 200)	Backward Splicing
2005	200		$\frac{200}{320} \times 100 = 62.5$
2006	240		$\frac{240}{320} \times 100 = 75$
2007	250		$\frac{250}{320} \times 100 = 78.12$
2008	280		$\frac{280}{320} \times 100 = 87.50$
2009	300		$\frac{300}{320} \times 100 = 93.75$
2010	320	200	$\frac{320}{320} \times 100 = 100$
2011		220	220
2012		240	240
2013		260	260
2014		280	280

$$\text{Back Splicing} = \frac{\text{Index A of Current Year}}{\text{Index A of Common Year}} \times 100$$

3.15 Deflation of Index Number

Q45. What do you mean by Deflating with an example ?

Ans : **(Dec.-18)**

By deflating we mean making allowances for the effect of changing price levels. A rise in price level means a reduction in the purchasing power of money. To take the case of a single commodity, suppose the price of wheat rises from Rs. 300 per quintal in 1986 to Rs. 600 per quintal in 1996 ; it means that in 1996 one can buy only half of wheat if he spends the same amount which he was spending on wheat in 1986. Thus the value (or purchasing power) of a rupee is simply the reciprocal of an appropriate price index written as a proportion. If prices increase by 60 per cent the price index is 1.60 and what a rupee will buy is only 1/1.60 or 5/8 of what it used to buy. In other words, the purchasing power of rupee is 5/8 of what it was or approximately 63 paise. Similarly, if prices increase by 25 per cent, the price index is 1.25 (125 per cent), and the purchasing power of the rupee is 1/1.25 = 0.80 or 80 paise.

This is the same as saying that the purchasing power of money is the reciprocal of the price index. The general expression may be given thus :

$$\text{Purchasing Power of Money} = \frac{1}{\text{Price Index}}$$

It shall be clear from above that since the value of money goes down with rising prices the workers or the salaried people are interested not so much in money wages as in real wages, *Ie.,* not how much they earn but how much their income or wage will buy.

For calculating real wages we can multiply money wages by a quantity measuring the purchasing power of the rupee, or better we divide the cash wages by an appropriate price index. This process is referred to as deflating. In principle it appears to be very simple but in practice the main difficulty consists in finding appropriate index to deflate a given set of values or appropriate deflators. The process of deflating can be expressed in the form of a formula as follows :

$$\text{Real wage} = \frac{\text{Money wage}}{\text{Price index}} \times 100$$

$$\text{Real wage or income Index No.} = \frac{\text{Index of Money Wages}}{\text{Consumer Price Index}}$$

PROBLEM ON DEFLATING

19. The following table gives the annual income of a worker and the general Index Numbers of price during 1988 - 1996. Prepare Index Number to show the changes in the real income of the teacher and comment on price increase.

Year	1988	1989	1990	1991	1992	1993	1994	1995	1996
income	3600	4200	5000	5500	6000	6400	6800	7200	7500
Price Index No.	100	120	145	160	250	320	450	530	600

Sol :

Index Number Showing Changes in The Real Income of the Worker

Year	Income	Price Index No.	Real Income	Real income Index No.
1988	3600	100	$\frac{3600}{100} \times 100 = 3600.00$	100.00
1989	4200	120	$\frac{4200}{120} \times 100 = 3500.00$	97.00
1990	5000	145	$\frac{5000}{145} \times 100 = 3448.27$	95.78
1991	5500	160	$\frac{5500}{160} \times 100 = 3437.50$	95.48

1992	6000	250	$\frac{6000}{250} \times 100 = 2400.00$	66.60
1993	6400	320	$\frac{6400}{320} \times 100 = 2000.00$	55.55
1994	6800	450	$\frac{6800}{450} \times 100 = 1511.11$	41.97
1995	7200	530	$\frac{7200}{530} \times 100 = 1358.49$	37.73
1996	7500	600	$\frac{7500}{600} \times 100 = 1250.00$	34.72

The method discussed above is frequently used to deflate individual values, value series of value indices. Its special use is in problems dealing with such diversified things as rupee sales, rupee inventories of manufacturer's, wholesaler's and retailer's income, wages and the like.

Exercises Problems

1. Compute (i) Laspeyres index and (ii) Paasches index from the data given below and test whether they satisfy (i) Time Reversal Test and (ii) Factor Reversal Tests

Commodity	p_0	q_0	p_1	q_1
A	5	10	4	12
B	8	6	7	7
C	6	3	5	4

[Ans: $P_{0,1}$ = 83.62, $P_{0,1}$ = 119.66, $Q_{0,1}$= 120.69, $V_{0,1}$ = 100.86

$P_{0,1}$ =83.57, $P_{1,0}$ = 119.59, $Q_{0,1}$ =120.62, $V_{0,1}$ =100.86]

2. Convert CBI to FBI (chained to 1996) :

Year :	1996	1997	1998	1999	2000
CBI	100	120	90	100	125

[Ans: 100, 120, 108, 108, 135]

3. From the Chain base index numbers given below, find fixed base index numbers.

Year :	1975	1976	1977	1978	1979
Chain Base Index :	80	110	120	90	140

[Ans: 80, 88, 105.60, 95.04, 133.06]

4. The following are the index numbers of wholesale prices of a commodity based on 1990.

Year	Index Numbers
1990	100
1991	108
1992	120
1993	150
1994	210
1995	225
1996	240

Prepare new index numbers taking 1992 as base.

[Ans: 83.33, 90, 100, 125, 175, 187.5, 200]

5. Construct the Cost of Living Index Number for the following data.

Item	Weights	Price Relatives
A	55	140
B	30	120
C	12	130
D	3	110

[Ans: 131.90]

Short Question and Answers

1. Define Demand?

Ans :

It is necessary to estimate the demand for the goods or services before they are produced and provided. The producers, for this purpose, heavily depend upon the data relating to the pattern of consumption of these goods and services. The demand analysis provides them the basis to take decisions relating to volume of production (How many products required to produce), capital to be invested (How much amount to be invested) and so on.

Demand

Demand for a commodity refers to the quantity of the commodity which an individual consumer is willing to purchase at a particular time at a particular price.

A product or service is said to have demand when three conditions are satisfied.

(a) Desire to acquire - Desire of the consumer to buy the + Product

(b) Willingness to pay - His willingness to buy the product and

(c) Ability to pay - Ability to pay the specified price for it.

Definitions

a) Demand refers to the quantities of commodity that the consumers are able to buy at each possible price during a given period of time, other things being equal.

- Ferguson

b) Demand is the ability and the willingness to buy a specific quantity of a good at alternatives prices in a given time period

-B.R.Schiller.

2. Features of Demand.

Ans :

The various features of demand are:

a) Difference between Desire and Demand: Demand is the amount of commodity for which a consumer has the willingness and the ability to buy .These is difference between need and demand. Demand is not only the need, it also implies that the consumer has the money to purchase it.

b) Relationship between Demand and price: Demand is always at a price .unless price is stated, the amount demanded has no meaning. the consumer must know both the price and the commodity and he will tell his amount demanded.

c) Demanded at a point of time: The amount demanded must refer to some period of time such as 10 quintals of wheat per year or six shirts per year of five kilos of sugar per month .not only this, the amount demanded and the price must refer to a particular data.

3. State Demand function ?

Ans :

Demand function is a function which describes a relationship between one variable and its determinants, u describes how much quantity of goods is bought at alternative prices of good and related goods, alternative income levels, and alternative values of other variables affecting demand. Thus, the demand function for a good relates the quantity of a good which consumers demand during a given period to the factors which influence the demand. The above factors can be built up into a demand function. Mathematically, the demand function for a product A can be expressed as follows:

$$Qd = f(P, I, T, P_R, E_P, E_I, S_P, D_C, A, O)$$

Where

Qd - refers to quantity of demand and it is a function of the following variables

P - refers to price of the product

I - refers to Income level of the consumer

T - refers to tastes and preferences of the consumer

P_R - refers to prices of related goods (substitutes/complementary)

E_P - refers to expectations about the prices in future

E_I - refers to expectations about the incomes in future

S_P - refers to size of population

D_c - refers to distribution of consumers over different regions

A - refers to advertising efforts and

O - refers to any other factors capable of affecting the demand .

4. Define Law of Demand ?

Ans :

The law of demand states that other factors being constant (cetris peribus), price and quantity demand of any good and service are inversely related to each other. When the price of a product increases, the demand for the same product will fall.

Description

Law of demand explains consumer choice behavior when the price changes. In the market, assuming other factors affecting demand being constant, when the price of a good rises, it leads to a fall in the demand of that good. This is the natural consumer choice behavior. This happens because a consumer hesitates to spend more for the good with the fear of going out of cash.

5. Define Elasticity of Demand?

Ans :

In the words of Dr.Marshall, "Elasticity of Demand may be defined as the percentage change in the quantity demanded divided by the percentage change in the price."

According to Boulding, "Price elasticity of demand measures the responsiveness of the quantity demanded to the change in price."

In the words of Dooley, "The price elasticity of demand measures the responsiveness of the quantity demanded to a change in its price."

According to Antol Murad, "Elasticity of demand is the ratio of relative change in quantity to relative change in price."

Thus, price elasticity of demand is a device to measure the rate of change in the quantity of a product demanded in response to a small change in its price.

6. Income Elasticity Demand.

Ans :

Income elasticity of demand refers to the quantity demanded of a commodity in response to a given change in income of the consumer.

Income elasticity is normally positive, which indicates that the consumer tends to buy more and more with every increase in income.

It is measured as follows:

$$\text{Income elasticity of demand} = \frac{\text{Proportionate change in quantity demanded for product X}}{\text{Proportionate change in income}}$$

The same is expressed as

$$\text{Edi} = \frac{(Q_2 - Q_1)/Q_1}{(I_2 - I_1)/I_1}$$

Where Q_1 is the quantity demanded before change, Q_2 is quantity demanded after change I_1 is income before change and I_2 is the income after change.

A positive income elasticity indicates that the demand for the product rises more quickly than the rise in disposable income. In other words, the demand is more responsive to a change in income.

7. Cross Elasticity of Demand.

Ans :

Cross elasticity of demand refers to the quantity demanded of a commodity in response to a change in the price of a related good, which may be substitute or complement.

It is measured as follows :

$$\text{Cross elasticity of demand} = \frac{\text{Proportionate change in quantity demanded for product X}}{\text{Proportionate change in price of product Y}}$$

The same is expressed as

$$Edc = \frac{(Q_2 - Q_1)/Q_1}{(P_2y - P_1y)/P_1y}$$

Where Q_1 is the quantity demanded before change, Q_2 is quantity demanded after change, P_1y is the price before change and P_2 y is the price after change in the case of product Y.

Cross elasticityis always positive for substitutes (which means that the demand for tea goes up if there is an increase in the price of coffee) and negative for complements (which menas that if there is an increase in the price of sugar, the demand for coffee tends to fall).

8. Define Supply?

Ans :

Supply of a commodity is the amount of its which the sellers are able and willing to offer for sale at a price during a certain period of time. Supply is a relative term - related to price and time. Market supply means the total quantity of a commodity that all the firms are willing to sell at a given during a given time period. It is found by adding the supply of the firms selling the commodity.

Supply analysis can be used to determine the impact of changes in product and factor prices, in technology, and in access on factor demands (including labor), production, marketed output, aggregate supply, and incomes. Generally, it can be used to analyze the impact on production of the removal of barriers to access or other changes in markets. Supply analysis, in the employment context, deals with key staffing questions related to current staffing levels in an organization.

Definitions of Supply

a) Supply of goods is the quantity offered for sale in a given market at a given time at various prices.

– Thomas

b) Supply refers to the amounts of a good that producer in a given market in a given market desired to sell, during a given time period at various prices, Ceteris Paribus.

– **Samuelson**

9. Define Supply Function.

Ans :

The mathematical function explaining the quantity supplied in terms of its various determinants, including price; thus the algebraic representation of the supply curve.

Mathematical or functional relationship between supply and its determinants is called supply function. It is a given as :

$$Q_x = f(P_x, C_x, T_x)$$

Where

Q_x = Supply of Commodity X

P_X = Price of Commodity X

C_x = Cost of Production of Commodity X

T_x = Technology of its Production.

10. Index Numbers.

Ans :

Some prominent definitions of index numbers are given below:

1. "Index numbers are devices for measuring differences in the magnitude of a group of related variables."

— Croxton & Cowden

2. "An index number is a statistical measure designed to show changes in a variable or a group of related variables with respect to time, geographic loca-tion or other characteristics such as income, profession, etc."

— Spiegel

3. "In its simplest form an index number is the ratio of two index numbers expressed as a per cent. An index number is a statistical measure—a measure designed to show changes in one variable or in a group of related variables over time, or with respect to geographic location, or other characteristic."

— Patterson

4. "In its simplest form, an index number is nothing more than a relative number, or a 'relative' which expresses the relationship between two figures, where one of the figures is used as a base."

— Morris Hamburg

5. "Generally speaking, index numbers measure the Size or magnitude of some object at a particular point in time as a percentage of some base or reference object in the past."

— Berenson & Levine

11. Limitations of Index Numbers.

Ans :

Index Numbers are of great use in studying trends, events, pertaining to a defined phenomenon. However, they suffer from certain limitations. There are

1. Approximate Representation

Index Numbers are based on sample data. Hence, they are only approximate indicators. They may not fully reflect the changes in the relative level of a phenomenon.

2. Likelihood of Error

There is a possibility of error in the choice of the sample, selection of base period, data collection, and assignment of weights or the appropriate formula to be used.

3. Not Responsive

There are rapid changes in Technology, tastes, fashions, customs and consequently, the consumption pattern of various commodities. Index numbers may not be able to keep pace with such changes. Hence, they may not be able to reflect the changes in the phenomenon being studied.

4. No Universally Acceptable formula

There is no universally acceptable standard formula that can capture the changes in a phenomenon with perfection. None of the formulae seems to satisfy all the tests of consistency. Hence, some amount of formula error is present in any calculation of Index numbers.

5. Scope for manipulation

On account of wide variety of formulae index numbers can be manipulated by appropriate choice of base year, sample set, price and quantity quotations.

6. Other Limitations

Apart from the above, Index numbers cannot capture the aspects of quality, reliability service etc. There is no index that is universally applicable. An Index constructed for one purpose cannot be used for another. Index numbers being averages, they are subject to all limitations of an average. Lastly, the utility of Index Numbers is limited by the availability of adequate and accurate data.

---

12. Paasche's Method of index number.

Ans :

Pascoe's Method is similar to Laspeyre's method. The only difference is in assignment of weights. As per this method quantities consumed of the commodities in the current year is taken as basis. The following steps need to be followed.

Step 1 : Multiply current year's prices (p_1) with current year's quantities (q_1)

Step 2 : Add the numbers obtained in step (1). The resultant sum is $\Sigma p_1 q_1$

Step 3 : Multiply base year's Prices (p_0) with current year's quantities (q_1)

Step 4 : Add the numbers obtained in step (2). The resultant sum is $\Sigma p_0 q_1$

Step 5 : Index Number as per Paasche's method = $P_{0,1} = \frac{\Sigma p_1 q_1}{\Sigma p_0 q_1} \times 100$

---

13. Time Reversal Test.

Ans :

This test was proposed by Prof. Irwin Fisher. According to fisher 'the formula for calculating the index number should be such that it gives the same ratio between one point of comparison and the other, no matter which of the two is taken as the base, or putting it another way, the index number reckoned forward should be reciprocal of the one reckoned backward.

In simple terms, given two time periods I and II, if an index number is calculated for period II taking period I as base, its value should be the reciprocal value of the index number for period I taking period II as base. The index numbers for the purpose of the test, should be in decimal form and not in percentage form. In other words, $P_{0,1}$ and $P_{1,0}$ should not be multiplied with 100. Symbolically.

$$\boxed{P_{0,1} \times P_{1,0} = 1}$$

Where $P_{0,1}$ = Fishers index for period II taking period I as base and

$P_{0,1}$ = Fishers index for period I taking Period II as base

Time Reversal test is satisfied by Marshall-Edge worth, Fisher, Walsh. Kelly' s index numbers and also by simple Aggregative Index, simple geometric mean of price relatives and weighted average of price relatives. Laspeyre's and Paasche's index numbers do not satisfy the time reversal test.

14. Factor Reversal Test.

Ans :

This test was also proposed by proof Fisher. This test requires that the product of two index numbers, one measuring price taking quantities as base, and the other measuring quantities taking price as base, should be equal to the net increase in total value from one period to another. Let us illustrate the same with the help of an example.

If $P_{0,1}$ is price index number, $Q_{0,1}$ is quantity index number, the product of the two index numbers should be equal to the value index number $V_{0,1}$

$$\boxed{P_{0,1} \times Q_{0,1} = V_{0,1}}$$

Fishers index number satisfies the factor reversal test. No other method satisfies the factor reversal test.

15. Define Chain Index Numbers.

Ans :

In the fixed base method, discussed so far the base remains the same Throughout the series of the index. This method, though convenient, has certain limitations. As time elapses conditions which were once important become less significant and it becomes more difficult to compare accurately present conditions with those of a remote period. New items may have to be included and old ones may have to be deleted in order to make the index more representative. In such cases it may be desirable to use the chain base method.

Steps in Constructing a Chain Index

In constructing a chain index following steps are desirable:

(i) Express the figures for each year as percentages of the preceding year.

The results so obtained are called link relatives.

(ii) Chain together these percentages by successive multiplication to form a chain index. Chain index of any year is the average link relative of that year multiplied by chain index of previous year divided by 100. In the form of formula

Chain Index for current year

$$= \frac{\text{Average link relative of current year} \times \text{Chain Index of previous year}}{100}$$

The link relatives obtained in step (i) facilitate comparison from one year to another, i.e., between closely situated periods in which the *q's* are a process of chaining binary comparisons facilitate long-term comparisons.

16. Consumer Price Index.

Ans :

The Consumer Price Index (CPI) is a measure of the average change over time in the prices paid by urban consumers for a market basket of consumer goods and services.

The CPI is widely used as an indicator of the change in the general level of consumer prices or the rate of inflation. Since the purchasing power of money is affected by changes in prices, the CPI is useful to virtually all Canadians. Consumers can compare movements in the CPI to changes in their personal income to monitor and evaluate changes in their financial situation.

17. Define Wholesale Price Index Number.

Ans :

The general change occurred in the wholesale prices of the commodities which are represented in index numbers is termed as wholesale price index number. Wholesale price indices are used to study the changes that take place in the general price level of a country.

The Indian Ministry of Commerce and Industry constructed the first wholesale price index number in January 1947 with 1939 as base year.

18. Base Shifting of Index Numbers ?

Ans :

For a variety of reasons, it frequently becomes necessary to change the reference base of an index number series from one time period to another without returning to the original raw data and recomputing the entire series. This change of reference base period is usually referred to as "shifting the base". There are two important reasons for shifting the base:

1. The previous base has become too old and is almost useless for purposes of comparison. By shifting the base it is possible to state the series in terms of a more recent time period.

2. It may be desired to compare several index number series which have been computed on different base periods; particularly if the several series are to be shown on the same graph, it may be desirable for them to have the same base period. This may necessitate a shift in the base period.

19. Define Splicing.

Ans :

Sometimes an index number series is available for a period of time. and then undergoes substantial revision including a shift in the reference period.

For example, the weights of an index number may become out of date and we may construct another index with new weight. Thus two indices would result. At times, it may be necessary to convert these two indices into a continuous series. The procedure employed for this conversion is called splicing

The process of splicing is very simple and is akin to that used in shifting the base. It is expressed in the form of a formula as follows :

$$\text{Spliced Index No.} = \frac{\text{Index No. of current year} \times \text{Old Index of New Base Year}}{100}$$

The following example would illustrate the procedure.

20. Backward Splicing.

Ans :

Backward splicing is used for splicing a new series of indices to continue with old series of indices.

Formula

$$\text{Back ward Splicing} = \frac{\text{Index A of Current Year}}{\text{Index A of Common Year}} \times 100$$

2. Forward Splicing

Forward splicing is used for splicing old series of indices to continue with new series of indices.

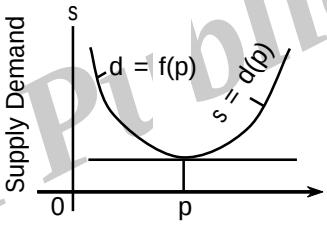
Formula

$$\text{Forward Splicing} = \frac{\text{Index B of Current Year} \times \text{Index A of Common Year}}{100}$$

21. What is Equilibrium price?

Ans :

The price determined by the point of intersection of demand and supply curves is known as equilibrium price i.e., equilibrium price is the solution of d = s Graphically it is represented as follows.



p is the equilibrium price

22. Distinguish between complementary and competitive commodities.

Ans :

The commodities A_1 and A_2 are said to be complementory if $\frac{\partial x_1}{\partial p_2} < 0$ and $\frac{\partial x_2}{\partial p_1} < 0$.

They are competitive if $\frac{\partial x_1}{\partial p_2} > 0$ and $\frac{\partial x_2}{\partial p_1} > 0$.

23. What is Giffen's paradox?

Ans :

There are some exceptional demand curves which instead of steping downwards rise upwards. Robert Giffen discovered one situation in the realways of workers. Even after a sharp rise in the price of bread its consumption was increase since bread was still the cheapest food compared with other foods. Since meat, egg such phenomenon is known as giffens paradox and it holds that the demand for Giffen's goods.

24. Price elasticity of supply

Ans :

The degree to which the supply of the product increases or decreases due to the change in the product price is termed as the elasticity of supply. Elasticity of supply is considered as the responsiveness of the sellers to a change in the price of the product.

The following formula can be used for measuring the elasticity of supply,

$$E_s = \frac{\text{Proportionate change in quality supplied}}{\text{Proportionate change in price}}$$

$$E_2 = \frac{\Delta Q_S}{\Delta P} \times \frac{P}{Q_S}$$

Where,

E_s = Elasticity of supply

ΔQ_s = The change in quantity supplied

ΔP = The change in price

Q_s = The original quantity supplied

P = The original price.

Choose the Correct Answers

1. Quantity in the base year is denoted by _____. [a]

   (a) q_0 (b) p_0

   (c) q_1 (d) p_1

2. _____ is helpful in business fore casting. [c]

   (a) Cost of Living (b) Splicing

   (c) Wholesale price (d) Base shifting

3. _____ denotes the price of the j^{th} commodity in the i^{th} year. [b]

   (a) q_{ij} (b) p_{ij}

   (c) p_i (d) p_j

4. _____ = $p_{ij} \times q_{ij}$. [d]

   (a) Q_{ij} (b) P_{ij}

   (c) E_{ij} (d) V_{ij}

5. Laspeyer's gives a _____ aggregate index. [b]

   (a) Price (b) Weighted

   (c) Quantity (d) None

6. $P_{0,1} \times P_{1,2} \times P_{2,0} =$ _____ [c]

   (a) 0 (b) 2

   (c) 1 (d) All

7. In chain base index numbers p_{01} is called _____ link. [a]

   (a) First (b) Second

   (c) Third (d) None

8. Deflation means _____ [d]

   (a) Correcting (b) Adjusting

   (c) Changing (d) a & b

9. _____ for any commodity means the requirement of an tem. [a]

(a) Demand (b) Supply

(c) a & b (d) None

10. $x = cp^{-\alpha}$ is called the _____ function with constant price elasticity. [b]

(a) Supply (b) Demand

(c) a & b (d) None

Fill in the Blanks

1. _____ for any commodity means the requirement of an item.
2. _____ Means the amount of commodity available at a given price.
3. The demand analysis deals with no. of important aspects of _____.
4. The price elasticity of demand is defined as the value of the ratio of relative change in the _____ to the relative change in the price.
5. The negative sign in demand and price moves in an _____ direction.
6. Demand function with constant price elasticity is _____.
7. Index numbers are _____ barometers.
8. _____ helps in formulating decisions and policies.
9. _____ is the test that the change in the price multiply by change in the quality should be equal to total change in the value.
10. _____ means joining in order.

ANSWERS

1. Demand
2. Supply
3. Economic statistics
4. Demand
5. Opposite
6. $x = cp^{-\alpha}$
7. Economic
8. Index Numbers
9. Factor Reversal test
10. Splicing

One Mark Answers

1. Define Demand.

Ans :

The demand for any commodity means the requirement of an item.

2. Define supply.

Ans :

Supply means the amount of commodity available at a given price.

3. Define index numbers.

Ans :

Index numbers can be calculated by different methods as such it is not possible to give a precise definition of index numbers.

4. State any to uses of index number.

Ans :

(i) Index numbers help in studying trends and tendencies.

(ii) Index numbers help in studying trends and tendencies.

5. Write the formula for simple aggregate method.

Ans :

$$P_{oi} = \frac{\Sigma P_{ij}}{\Sigma P_{oj}} \times 100 \cdot$$

Practicals

1. Simple random sampling with and without replacement comparison between SRSWR and SRSWOR.

Ans :

Refer to Unit-I, Problem No. 4, 5

2. Stratified random sampling with proportional and optimum allocations comparison between propotional and optimum allocation with SRSWOR.

Ans :

Refer to Unit-II, Problem No. 1

3. Systematic sampling with N = nk. comparison of systematic sampling with stratified and SRSWOR.

Ans :

Refer to Unit-II, Q.No. 8

4. Measurement of trend by method of least squares and moving averages.

Ans :

Refer to Unit-II, Problem No. 8, 14

5. Determination of seasonal indices by the method of ratio to moving average.

Ans :

Refer to Unit-II, Problem No. 9

6. Determination of seasonal indices by the method of ratio to trend.

Ans :

Refer to Unit-II, Problem No. 9

7. Determination of seasonal indices by the method of link relatives.

Ans :

Refer to Unit-II, Problem No. 18

FACULTY OF SCIENCE

B.Sc. III Year V-Semester(CBCS) Examination

July - 2021

STATISTICS

PAPER - V (DSC) : Sampling Theory, Time Series, Index Numbers and Demand Analysis

Time : 2 Hrs] **[Max. Marks : 60**

PART - A (4 × 5 = 20 Marks)
[Short Answer Type]

ANSWERS

1. Define parameter, statistic and standard error. **(Unit-I, Q.No.4,5)**
2. Explain what are called non-sampling errors and state their sources. **(Unit-I, Q.No.12)**
3. Describe stratified random sampling procedure with an example. **(Unit-II, Q.No.1)**
4. Define Time Series. Explain about additive and multiplicative models of time series. **(Unit-II, SQA-11,25,26)**
5. Explain about cyclic variations. **(Unit-II, SQA-28)**
6. Explain about price elasticity of supply. **(Unit-III, SQA-24)**
7. Explain time reversal and factor reversal tests. **(Unit-III, SQA-13,14)**
8. Define chain base index numbers. **(Unit-III, SQA-15)**

PART - B (2 × 20 = 40 Marks)
[Essay Answer Type]

Note: Answer any **Two** questions.

9. Explain about principal steps of sample surveys. **(Unit-I, Q.No.6)**
10. In SRSWOR show that variance of sample mean is given by,

$$V(\bar{y}_n) = \frac{s^2}{n} \cdot \frac{N-n}{N}$$ **(Unit-I, Theorem.3)**

11. Define systematic sampling procedure. If $\bar{y}_{sys}$ is the mean of systematic sampling then show that

$$V\left(\bar{y}_{sys}\right) = \frac{N-1}{N} S^2 - \frac{(n-1)k}{N} S^2_{wsy}$$

Where,

$$S^2_{wsy} = \frac{1}{k(n-1)} \sum_{i=1}^{k} \sum_{j=1}^{n} \left(y_{ij} - \bar{y}_i\right)^2$$ **(Unit-II, Q.No.10)**

12. Explain link relatives method to find seasonal indices along with merits dand demerits. **(Unit-II, Q.No.30)**

13. Derive curve of concentration in demand analysis. **(Unit-III, Q.No.7)**

14. Define cost of living index number. Explain its construction methods and uses. **(Unit-III, Q.No.39,40)**

FACULTY OF SCIENCE

B.Sc. III Year V-Semester(CBCS) Examination

OCtober / November - 2020

STATISTICS

Paper - V (Sampling Theory, Time Series, Index Numbers and Demand Analysis)

Time : 2 Hrs] **[Max. Marks : 60**

PART - A (4 × 5 = 20 Marks)

Note: Answer any **Four** questions.

ANSWERS

1. Define sample. What are its limitations? **(Unit-I, Q.No.4,14)**
2. What are mixed sampling methods? **(Unit-I, Q.No.18)**
3. What is a time series? Name its components. **(Unit-II, SQA.11, Q.No.17)**
4. Define systematic sampling procedure. **(Unit-II, Q.No.8)**
5. What is whole sale price index? Give an example. **(Unit-III, SQA-17)**
6. Explain about demand and supply. **(Unit-III, SQA-1,8)**
7. Explain factor reversal test. **(Unit-III, SQA-14)**
8. Write about mixed models of time series. **(Unit-II, SQA-26)**

PART - B (2 × 20 = 40 Marks)

Note: Answer any **Two** questions.

9. Give advantages of sampling over census. **(Unit-I, Q.No.13)**
10. In SRSWR show that sample mean is unbiased estimator of population mean and derive its variance. **(Unit-I, Theorem.1)**
11. Explain stratified random sampling procedure. Prove that $V(\bar{y}st)$ is minimum for fixed total sample size 'n' if $n_i \alpha N_i S_i$. **(Unit-II, Q.No.6)**
12. Explain modified exponential curve and its fitting by partial sums method. **(Unit-II, Q.No.24)**
13. Explain Leontiefs method for estimating demand curve, stating assumptions. **(Unit-III, Q.No.22)**
14. Explain various methods of weighted price indices. **(Unit-III, Q.No.41)**

FACULTY OF SCIENCE

B.Sc. III Year V-Semester (CBCS) Examination

June / July - 2019

STATISTICS

PAPER - V (DSC) : Sampling Theory, Time Series, Index Numbers and Demand Analysis

Time : 3 Hrs] **[Max. Marks : 60**

PART - A (5 × 3 = 15 Marks)

[Short Answer Type]

Note : Answer any **Five** of the following questions

ANSWERS

1. What is subjective sampling? **(Unit-I, Q.No.16)**
2. In SRSWOR show that sample mean is an unbiased estimator for population mean. **(Unit-I, Theorem.1)**
3. Define cost function. **(Unit-II, SQA-4)**
4. Explain the Decomposition of Time series data by additive model. **(Unit-II, SQA-24)**
5. Explain price elasticity of supply and Demand. **(Unit-III, Q.No.21,11)**
6. What is Equilibrium price? **(Unit-III, SQA-21)**
7. Explain the concept of Base shifting in Index Numbers. **(Unit-III, SQA-18)**
8. Prove that Fisher's Ideal Index Number lies between Laspeyre's and paasche's index numbers. **(Unit-III, Q.No.33)**

PART - B (3 × 15 = 45 Marks)

[Essay Answer Type]

Note : Answer all from the following questions.

9. (a) What is a sample survey? Explain the main steps involved in the planning and execution of a sample survey. **(Unit-I, Q.No.2,7)**

OR

(b) Define SRS (i) with replacement (ii) without replacement, Explain the various methods of drawing a random sample. **(Unit-I, Q.No.22,23)**

10. (a) Explain any one method for fitting the following growth curves

(i) Modified Exponential curve **(Unit-II, Q.No.24)**

(ii) Logistic curve. **(Unit-II, Q.No.26)**

OR

(b) What are the various methods of allocating a sample in stratified random sampling? Compare the efficiencies of Neyman and proportional allocations with that of simple random sample of the same size. **(Unit-II, Q.No.1,3,5)**

11. (a) Explain Pigou's method of deriving demand curve from Time series data stating the limits. **(Unit-III, Q.No.23)**

OR

(b) Outline the Mathematical tests for an ideal Index Number. Illustrate these with reference to Fisher's Ideal Index number. **(Unit-III, Q.No.34, Prob.7)**

FACULTY OF SCIENCE

B.Sc. III Year V-Semester (CBCS) Examination

November / December - 2018

STATISTICS

PAPER - V (DSC) : Sampling Theory, Time Series, Index Numbers and Demand Analysis

Time : 3 Hrs] **[Max. Marks : 60**

PART - A (3 × 5 = 15 Marks)

[Short Answer Type]

Note : Answer any **Five** of the following questions

ANSWERS

1. Define sampling unit and sampling frame. **(Unit-I, SQA-12,4)**
2. Explain probability sampling. **(Unit-I, SQA-11)**
3. Explain about proportional allocation. **(Unit-II, Q.No.3)**
4. Explain about random fluctuations in time series data. **(Unit-II, SQA-28)**
5. Distinguish between complementary and competitive commodities. **(Unit-III, SQA-22)**
6. What is Giffen's paradox? **(Unit-III, SQA-23)**
7. Explain chain base index numbers. **(Unit-III, SQA-15)**
8. Explain the multiplicative and mixed model of a time series data. **(Unit-II, SQA-25,26)**

PART - B (3 × 15 = 45 Marks)

[Essay Answer Type]

Note : Answer **ALL** questions

9. (a) Distinguish between sampling and non sampling errors. Give the soures of Non sampling errors. **(Unit-I, Q.No.12)**

Ans :

BASIS FOR COMPARISON	SAMPLING ERROR	NON-SAMPLING ERROR
Meaning	Sampling error is a type of error, occurs due to the sample selected does not perfectly represents the population of interest.	An error occurs due to sources other than sampling, while conducting survey activities is known as non sampling error.
Cause	Deviation between sample mean and population mean	Deficiency and analysis of data
Type	Random	Random or Non-random
Occurs	Only when sample is selected.	Both in sample and census.
Sample size	Possibility of error reduced with the increase in sample size.	It has nothing to do with the sample size.

OR

(b) Define SRSWOR and SRSWR. Show that is SRSWOR the probability of selecting a specified unit of the population at any given draw is equal to the probability of selecting it at the first draw.

Ans :

$P(E_r)$ = P[Sepcified unit is not selected in any one of the previous (r – 1) draw cut then selected at the rth draw]

$$\therefore \quad P(E_r) = \prod_{i=1}^{r-1}\left[1-\frac{1}{N-(i-1)}\right] \times \frac{1}{N-(r-1)} = \prod_{i=1}^{r-1}\frac{N-i}{N-i+1} \times \frac{1}{N-r+1}$$

$$= \frac{N-1}{N} \times \frac{N-2}{N-1} \times \frac{N-3}{N-2} \ldots\ldots \frac{N-r+1}{N-r+2} \times \frac{1}{N-r+1} = \frac{1}{N}$$

$$\therefore \quad P(E_r) = \frac{1}{N} = P[E_i]$$

The probability of selecting a specified unit of the population at any given draw in equal to the probability of its being selected at the first draw.

10. (a) What are the seasonal variations? Explain ratio to trend method of calculating seasonal variations. Also give its merits and demerits. **(Unit-II, Q.No.27,29)**

OR

(b) Define cost function. With a cost function $C = a + \Sigma_h c_h n_h$ prove that the variance of the estiamted mean $\bar{y}_{st}$ is minimum when nh is proportional to $N_h S_h / \sqrt{C_h}$. **(Unit-II, Q.No.4,6)**

11. (a) Describe Leontief's method of estimating price elasticity of demand for time series data and its limitations. **(Unit-III, Q.No.22)**

OR

(b) What is meant by (i) Base shifiting (ii) Deflating (iii) Splicing of Index Numbers? Explain and illustrate. **(Unit-III, Q.No.42,45,43)**

MAHATMA GANDHI UNIVERSITY
FACULTY OF SCIENCE
B.Sc. III Year V-Semester(CBCS) Examination
November / December - 2018
STATISTICS
PAPER - V : Applied Statistics - 1

Time : 2½ Hrs] **[Max. Marks : 60**

PART - A (3 × 5 = 15 Marks)

[Short Answer Type]

Note: Answer **ALL** of the following questions.

ANSWERS

1. Explain Simple Random Sampling. **(Unit-I, QNo.19)**
2. Explain Time Series analysis uses. **(Unit-II, SQA-12)**
3. Show that Fisher's Index Number satisfies the Time and Factor Reversal Test. **(Unit-III, SQA-13,14)**

PART - B (3 × 15 = 45 Marks)

[Essay Answer Type]

Note: Answer **ALL** of the following questions.

4. (a) In SRSWOR, show that the sample mean square is an unbiased estimate of population mean square. **(Unit-I, Theorem.1)**

OR

(b) If X_i, Y_i are the pair of variates define for every unit (i = 1, 2, ..., N) of the population and $\bar{x}$ and $\bar{y}$ are the corresponding sample means of a simple random sample of size 'n' taken without replacement then. Prove that cov

$$(\bar{X}, \bar{Y}) = \frac{N-n}{nN} \cdot \frac{1}{N-1} \sum_{i=1}^{N} (X_i - \bar{X})(Y_i - \bar{Y}).$$ **(Unit-I, Theorem.4)**

5. (a) Explain the components of Time Series Analysis. **(Unit-II, QNo.17)**

OR

(b) Explain the link relative method for measuring the seasonal variations merits and demerits. **(Unit-II, QNo.30)**

6. (a) Explain the problems in construction of Index Numbers. **(Unit-III, QNo.35)**

OR

(b) Define Cost of living Index Number. Describe the methods of construction and mention its uses. **(Unit-III, QNo.39,40)**

FACULTY OF SCIENCE

B.Sc. III Year V-Semester(CBCS) Examination

Model Paper - I

STATISTICS

PAPER - V (DSC) : Sampling Theory, Time Series, Index Numbers and Demand Analysis

Time : 3 Hrs] **[Max. Marks : 60**

PART - A (5 × 3 = 15 Marks)

[Short Answer Type]

Note : Answer any **Five** of the following questions

ANSWERS

1. Sample survey. **(Unit-I, SQA-2)**
2. Advantages of sampling. **(Unit-I, SQA-9)**
3. Define Probability Sampling. **(Unit-I, SQA-11)**
4. Stratified Random Sampling. **(Unit-II, SQA-1)**
5. Population Proportion. **(Unit-II, SQA-3)**
6. Advantages of Stratified Random Sampling. **(Unit-II, SQA-6)**
7. Features of Demand. **(Unit-III, SQA-2)**
8. Factor Reversal Test. **(Unit-III, SQA-14)**

PART - B (3 × 15 = 45 Marks)

[Essay Answer Type]

Note : Answer all from the following questions.

9. (a) In SRSWOR the variance of the sample mean is given by

$$\text{Var}(\bar{y}) = \frac{S^2}{n} \cdot \frac{N-n}{N}$$ **(Unit-I, Theorem.3)**

OR

(b) Explain Simple Random Sampling with Replacement(SRSWR). **(Unit-I, Q.No.23)**

10. (a) In stratified random sampling with given cost function of the form

$C = a + \sum_{i=1}^{k} C_i n_i$, $V(\bar{y}_{st})$ is minimum of $n_i \propto \frac{N_i S_i}{\sqrt{C_i}}$. **(Unit-II, Q.No.6)**

OR

(b) Explain the various models are used in Time Series Analysis. **(Unit-II, Q.No.18)**

11. (a) Explain the various methods of deter- ming demand. **(Unit-III, Q.No.7)**

OR

(b) Explain Pigou's Methods of Determining Demand Curve from Time Series Data and their Limitations. **(Unit-III, Q.No.23)**

FACULTY OF SCIENCE

B.Sc. III Year V-Semester(CBCS) Examination

Model Paper - II

STATISTICS

PAPER - V (DSC) : Sampling Theory, Time Series, Index Numbers and Demand Analysis

Time : 3 Hrs] **[Max. Marks : 60**

PART - A (5 × 3 = 15 Marks)

[Short Answer Type]

Note : Answer any **Five** of the following questions

ANSWERS

1. Sampling **(Unit-I, SQA-1)**
2. Sample Frame **(Unit-I, SQA-4)**
3. Census Method **(Unit-I, SQA-8)**
4. Define cost function. **(Unit-II, SQA-4)**
5. Advantages of Simple Random Sampling **(Unit-II, SQA-9)**
6. Utility of Time Series Analysis. **(Unit-II, SQA-12)**
7. Define Chain Index Numbers. **(Unit-III, SQA-15)**
8. What is Giffen's paradox? **(Unit-III, Q.No.23)**

PART - B (3 × 15 = 45 Marks)

[Essay Answer Type]

Note : Answer all from the following questions.

9. (a) Explain the notation and terminology used in simple random sampling. **(Unit-I, Q.No.19)**

OR

(b) Explain Simple Random Sampling without Replacement (SRSWOR). **(Unit-I, Q.No.22)**

10. (a) Distinguish between relative efficiency of systematic sampling over stratified random sampling. **(Unit-II, Q.No.10)**

OR

(b) What is a growth curve? Write in detail about modified exponential curve with example. **(Unit-II, Q.No.24)**

11. (a) What do you understand by price elasticity of supply ? Explain different types of price elasticity of supply. **(Unit-III, Q.No.21)**

OR

(b) Explain the criterion for testing the consistency of good index numbers. **(Unit-III, Q.No.34)**

FACULTY OF SCIENCE

B.Sc. III Year V-Semester(CBCS) Examination

Model Paper - III

STATISTICS

PAPER - V (DSC) : Sampling Theory, Time Series, Index Numbers and Demand Analysis

Time : 3 Hrs] **[Max. Marks : 60**

PART - A (5 × 3 = 15 Marks)

[Short Answer Type]

Note : Answer any **Five** of the following questions

ANSWERS

1. Standard Error. **(Unit-I, SQA-5)**
2. Limitations of sampling. **(Unit-I, SQA-10)**
3. Need for sampling. **(Unit-I, SQA-7)**
4. Define Time Series. **(Unit-II, SQA-11)**
5. Additive Model **(Unit-II, SQA-24)**
6. Optimum allocation. **(Unit-II, SQA-5)**
7. Time Reversal Test. **(Unit-III, SQA-13)**
8. What is Equilibrium price? **(Unit-III, SQA-21)**

PART - B (3 × 15 = 45 Marks)

[Essay Answer Type]

Note : Answer all from the following questions.

9. (a) Explain the notation and terminology used in simple random sampling. **(Unit-I, Q.No.19)**

OR

(b) In a simple random sampling without replacement the sample mean is an unbiased estimate of the population mean. **(Unit-I, Theorem.1)**

10. (a) What are the Components of Time Series ? **(Unit-II, Q.No.17)**

OR

(b) What is logistic curve ? How it is prepared ? **(Unit-II, Q.No.26)**

11. (a) Explain Leontief's Methods of Deter- mining Demand Curve from Time Series Data and their Limitations ? **(Unit-III, Q.No.22)**

OR

(b) Define Wholesale Price Index Number. State various series of Wholesale Price Index Number. **(Unit-III, Q.No.41)**