

Biyani's Think Tank

Concept based material

Business Statistics

BBA –III Sem

Dr. Archana Dusad

Assistant Professor

Deptt. of Commerce & Management

Biyani Girls College, Jaipur



Biyani's

CUKPO
tional Portal in India
.COM

Published by :

Think Tanks

Biyani Group of Colleges

Concept & Copyright :

©**Biyani Shikshan Samiti**

Sector-3, Vidhyadhar Nagar,

Jaipur-302 023 (Rajasthan)

Ph : 0141-2338371, 2338591-95 • Fax : 0141-2338007

E-mail : acad@biyanicolleges.org

Website :www.gurukpo.com; www.biyanicolleges.org

First Edition : 2009

Second Edition: 2010

Price:

While every effort is taken to avoid errors or omissions in this Publication, any mistake or omission that may have crept in is not intentional. It may be taken note of that neither the publisher nor the author will be responsible for any damage or loss of any kind arising to anyone in any manner on account of such errors and omissions.

Leaser Type Setted by :

Biyani College Printing Department

Preface

I am glad to present this book, especially designed to serve the needs of the students. The book has been written keeping in mind the general weakness in understanding the fundamental concepts of the topics. The book is self-explanatory and adopts the “Teach Yourself” style. It is based on question-answer pattern. The language of book is quite easy and understandable based on scientific approach.

Any further improvement in the contents of the book by making corrections, omission and inclusion is keen to be achieved based on suggestions from the readers for which the author shall be obliged.

I acknowledge special thanks to Mr. Rajeev Biyani, *Chairman* & Dr. Sanjay Biyani, *Director (Acad.)* Biyani Group of Colleges, who are the backbones and main concept provider and also have been constant source of motivation throughout this Endeavour. They played an active role in coordinating the various stages of this Endeavour and spearheaded the publishing work.

I look forward to receiving valuable suggestions from professors of various educational institutions, other faculty members and students for improvement of the quality of the book. The reader may feel free to send in their comments and suggestions to the under mentioned address.

Author

GURUKR...
No. 1 Educational Web Portal in India

Syllabus

BBA III SEM.

BUSINESS STATISTICS

- Unit I. **Introduction:** Meaning and Definition of Statistics, Scope of Statistics in Economics, Management, Science and Industry. Concept of Population and sample with illustration, Methods of Sampling SRSWR, SRSWOR, Stratified, Systematic. Data condensation and graphical methods. Raw Data, Attributes and variables, classification frequency distribution, cumulative frequency distribution Graphs- Histogram, Frequency Polygon Diagrams - Multiple bar, Pie, Subdivided bar. .
- Unit II **Measuring of Central Tendency:** Criteria for good measures of central tendency Arithmetic Mean, Median, Mode for grouped and ungrouped data, combined mean.
- Unit III **Measures f Dispersion:** **Concept** of Dispersion, Absolute and Relative Measures of Dispersion, Range, Variance, Standard Deviation, Coefficient of variation, Quartile.
- Unit IV **Correlation and Regression(For ungrouped data)** : Concept of Correlation Positive and Negative correlation, Karl Pearson's Coefficient of Correlation Meaning of Regression, Two regression equations, regression coefficients and properties.
- Unit V **Index Numbers** : Meaning and Uses of Index Numbers, Simple and composite Index Numbers, Aggregative and average of price Relatives - simple and weighted index numbers. Construction of index numbers fixed and chain base, Laspear, Paschee, Kelly, and Fisher index number :Construction of consumer price index case of limit index

□ □ □

Content

S.No.	Name of Topic
1.	Meaning, Definition, and scope of Statistics
2.	Functions, Importance and Distrust of Statistics
3.	Statistical Investigation
4.	Collection Of Data
5.	Census and Sample Investigation
6.	Editing of collected Data and Statistical Error
7.	Classification and Tabulation of data
8.	Diagrammatic and Graphic Presentation
9.	Measure of central Tendency
10.	Measure of Dispersion
11.	Correlation
12.	Linear Regression
13.	Index Number

Chapter 1

Meaning, Definition and Scope of Statistics

Q.1 Define 'Statistics' and give characteristics of 'Statistics'.

Ans.: 'Statistics' means numerical presentation of facts. Its meaning is divided into two forms - in plural form and in singular form. In plural form, 'Statistics' means a collection of numerical facts or data example price statistics, agricultural statistics, production statistics, etc. In singular form, the word means the statistical methods with the help of which collection, analysis and interpretation of data are accomplished.

Characteristics of Statistics -

- a) Aggregate of facts/data
- b) Numerically expressed
- c) Affected by different factors
- d) Collected or estimated
- e) Reasonable standard of accuracy
- f) Predetermined purpose
- g) Comparable
- h) Systematic collection.

Therefore, the process of collecting, classifying, presenting, analyzing and interpreting the numerical facts, comparable for some predetermined purpose are collectively known as "Statistics".

Q.2 What is meant by 'Data'?

Ans.: Data refers to any group of measurements that happen to interest us. These measurements provide information the decision maker uses. Data are the

foundation of any statistical investigation and the job of collecting data is the same for a statistician as collecting stone, mortar, cement, bricks etc. is for a builder.

Q.3 Discuss the Scope of Statistics.

Ans.: The scope of statistics is much extensive. It can be divided into two parts –

- (i) **Statistical Methods** such as Collection, Classification, Tabulation, Presentation, Analysis, Interpretation and Forecasting.
- (ii) **Applied Statistics** – It is further divided into three parts:
 - a) **Descriptive Applied Statistics** : Purpose of this analysis is to provide descriptive information.
 - b) **Scientific Applied Statistics** : Data are collected with the purpose of some scientific research and with the help of these data some particular theory or principle is propounded.
 - c) **Business Applied Statistics** : Under this branch statistical methods are used for the study, analysis and solution of various problems in the field of business.

Q. 4 State the limitation of statistics?

Ans. Scope of statistics are very wide. In any area where problems can be expressed in qualitative form, statistical methods can be used. But statistics have some limitations

1. Statistics can study only numerical or quantitative aspects of a problem.
2. Statistics deals with aggregates not with individuals.
3. Statistical results are true only on an average.
4. Statistical laws are not exact.
5. Statistics does not reveal the entire story.
6. Statistical relations do not necessarily bring out the cause and effect relationship between phenomena.
7. Statistics is collected with a given purpose.
8. Statistics can be used only by experts.

Chapter-2

Functions, Importance and Distrust of Statistics

Q.1 Give reasons for distrust in Statistics.

Ans.: By distrust of statistics we mean lack of confidence in statistical statements and statistical methods. It is often commented by people

"Statistics can prove anything."

"There are three type of lies - lies, damned lies and statistics - wicked in the order of their naming."

The **main reasons** for such views are -

- a) Figures are convincing, and therefore people are easily led to believe them.
- b) Ignorance of limitation of statistics.
- c) Lack of test of accuracy.
- d) Contradiction of data from actual circumstances.
- e) Lack of specific ability to arrive at correct and appropriate results.
- f) Can easily be manipulated.

Q.2 Discuss the functions and importance/utility of Statistics.

Ans.: Statistical methods are used not only in the social, economic and political fields but in every field of science and knowledge. Statistical analysis has become more significant in global relations and in the age of fast developing information technology.

According to **Prof. Bowley**, *"The proper function of statistics is to enlarge individual experiences"*.

Following are some of the **important functions of Statistics** :

- a) To provide numerical facts.

- b) To simplify complex facts.
- c) To enlarge human knowledge and experience.
- d) Helps in formulation of policies.
- e) To provide comparison.
- f) To establish mutual relations.
- g) Helps in forecasting.
- h) Test the accuracy of scientific theories.
- i) To study extensively and intensively.

The use of statistics has become almost essential in order to clearly understand and solve a problem. Statistics proves to be much useful in **unfamiliar fields of application and complex situations** such as :-

- a) Planning
- b) Administration
- c) Economics
- d) Trade & Commerce
- e) Production management
- f) Quality control
- g) Helpful in inspection
- h) Insurance business
- i) Railways & transport Co
- a) Banking Institutions
- b) Speculation and Gambling
- c) Underwriters and Investors
- d) Politicians & social workers.

Chapter 3

Statistical investigation

Q.1 What do you mean by statistical investigation?

Ans. In statistical investigation the search for the knowledge is done by means of collection of data through statistical methods. For any statistical enquiry, whether it is business, economics, political or social science, the basic problem is to collect facts and figures relating to a particular phenomenon expressed in terms of quantity or numbers.

Q.2 Write different stages of statistical inquiry serially?

Ans. The following are the various stages of a statistical inquiry:

1. **Planning the Inquiry:** A clear plan for investigation is made in the light of the objects, scope, and nature, sources of information, units and method of collection of data.
2. **Collection of data:** The next step for investigation is the collection of raw data. Different types of questionnaires and schedules be prepared and dispatched to related respondents for collection of data.
3. **Editing of collected data:** After collection of data, data is edited in the light of standard of accuracy to be maintained.
4. **Classification and tabulation of data:** After editing, data is classified in different classes and present the same in the forms of tables so that analysis is convenient.
5. **Presentation of data:** after classification and tabulation the data are presented in the form of averages, diagrams or graphs etc.
6. **Analysis of Data:** The presented data are to be analyzed with the help of statistical methods. The mathematical measures like averages, dispersion, correlation etc. may be used for this purpose.
7. **Interpretation and writing of the report:** The last stage of investigation is the suitable and proper interpretation and presentation of a report of investigation by

the investigator. Report will be concise and touch all the points regarding inquiry to make it complete report.

Q.3 Discuss planning for a statistical investigation?

Ans. Quality of investigation results depends upon proper planning of investigation .The following matters need careful consideration before the commencement of investigation

1. **Definition of the problem:** The first step is to clearly state the problem indicating the purpose for which inquiry is to be conducted, the type of information needed and the use for which information obtained will be put.
2. **Object of investigation:** it is very essential for determining the nature of data to be collected and technique to be employed it will help in eliminating the collection of irrelevant data.
3. **Scope of investigation:** Scope of enquiry relates to the coverage with regards to the type or information, subject matter and geographical area.
4. **Time or Period of investigation :** The investigation should be carried out within a reasonable period of time, otherwise information collected may become outdated and have no meaning at all.
5. **Sources of information :** After determination of object and period of investigation, sources of information be decided. The sources may be Primary and Secondary.
6. **Types of investigation:** statistical investigation may be of the following types
 - **Census or sample**
 - **Direct or indirect**
 - **Initial or repetitive confidential and public**
 - **Regular and Ad hoc**
 - **Official , semi official, or non official**
 - **Postal or personal**
7. **Statistical unit :** it is a unit in terms of which the investigator counts or measures the variables or attributes selected for enumeration.
8. **Degree of accuracy desired:** degree of accuracy required depends upon the object of the enquiry.
9. **Organization of investigation:** Before the commencement of an enquiry the preliminaries be decided.
10. **Collection of data:** At last the investigator starts the work of collection data. The method of collection of data , the question to be included in a questionnaire, the size, manner and necessary considerations for the inquiry are predetermined.

Chapter 4

Collection of Data

Q.1 What do you mean by Collection of Data? Differentiate between Primary and Secondary Data.

Ans.: Collection of data is the basic activity of statistical science. It means collection of facts and figures relating to particular phenomenon under the study of any problem whether it is in business economics, social or natural sciences.

Such material can be obtained directly from the individual units, called primary sources or from the material published earlier elsewhere known as the secondary sources.

Difference between Primary & Secondary Data

	Primary Data	Secondary Data
Basis nature	Primary data are original and are collected for the first time.	Data which are collected earlier by someone else, and which are now in published or unpublished state.
Collecting Agency	These data are collected by the investigator himself	Secondary data were collected earlier by some other person.
Post collection alterations	These data do not need alteration as they are according to the requirement of the investigation	These have to be analyzed and necessary changes have to be made to make them useful as per the requirements of investi-

		gation.
Time & Money	More time, energy and money has to be spent in collection of these data.	Comparatively less time and money is to be spent.

Q.2 What do you mean by Questionnaire? Give merits of a good Questionnaire.

Ans.: Questionnaire is a document containing questions related to the specific requirement of a statistical investigation for collection of information which is filled by the informants personally.

Requirements of a good **questionnaire** :-

- Questions should be simple, clear and short.
- Simple alternative or multiple choice questions.
- Unambiguous and precise.
- Questions should be in sequence.
- Directly relative questions.
- Test of accuracy.
- No restricted questions affecting personal whims.
- Assurance of secrecy to the informants.
- Probability of a perfect answer.

Chapter 5

Census and sample investigation

Q.1 What is Law of 'Statistical Regularity' and the Law of 'Inertia of Large Numbers'?

Ans.: Based on the mathematical theory of probability, Law of Statistical Regularity states that if a sample is taken at random from a population, it is likely to possess the characteristics as that of the population. A sample selected in this manner would be representative of the population. If this condition is satisfied, it is possible for one to depict fairly accurately the characteristics of the population by studying only a part of it.

The Law of 'Inertia of Large Numbers' is a corollary of the law of statistical regularity. It states that, other things being equal, larger the size of sample, more accurate the results are likely to be. This is because when large numbers are considered, the variations in the component parts tend to balance each other and, therefore, the variation in the aggregate is insignificant.

Q.2 What is Random Sampling?

OR

Define Random Sampling.

Ans.: Random Sampling is one in which selection of items is done in such a way that every item of the universe has an equal chance of being selected.

Random Sampling is based on probability and it is free from bias.

The different **methods of Random Sampling** are:-

- a) Lottery method.
- b) Rotating the drum.
- c) By systematic arrangement.

- d) Routed wheel method
- e) By random numbers

Q.3 Define Stratified Random Sampling?

Ans.3. This sampling method is used when the population is comprised of natural subdivision of units. This method consists in classifying the population units into a certain number of groups called strata.

Then random samples are selected independently from each group.

Gurukpo.com
No. 1 Educational Web Portal in India

Chapter 6

Editing of Collected Data and Statistical Error

Q.1 What do you mean by Statistical Error? Give the difference between Mistake and Error?

Ans.: The difference between the actual value of the figure and its approximated value is called statistical error. For example if number of students of a college is 1,255 but in round figures it is written as 1,300, then the difference is called 'Statistical error'. In the words of Prof. Connor, *"In statistical sense, 'error' means the difference between the approximate value and the true or ideal value, accurate determination of which is not possible"*.

Difference between Mistake and Error

	Basis	Mistake	Error
a)	Nature	Committed deliberately	Not deliberate
b)	Source	Due to use of wrong method	Difference between actual and approximate value.
c)	Estimation	Difficult to estimate	Can be estimated
d)	Prevention	Can be avoided by	Cannot be avoided

		carefulness	easily.
e)	State Occurrence	of Can be committed at any stage	Creeps in only at the state of collecting, analysis and interpretation.

Q.2 Write a note on the Editing of Primary Data and Secondary Data for the purpose of Analysis and Interpretation.

Ans.: Before analysis and interpretation, it is necessary to edit the data to detect possible errors and inaccuracies, so that accurate and impartial results may be obtained. Thus editing means the process of checking for the errors and omissions and making corrections, if necessary.

The task of editing is a highly specialized one and requires high level of skill and carefulness to attain the proper degree of accuracy.

Editing of Primary Data : While editing primary data, the following points should be considered :-

- Editing for consistency
- Editing for completeness
- Editing for accuracy
- Editing for uniformity or homogeneity

Editing of Secondary Data : Since secondary data have already been obtained, it is highly desirable that a proper scrutiny of such data is made before they are used by the investigator. **Bowley** rightly points out that "*secondary data should not be accepted at their face value*".

Hence before using secondary data, the investigators should consider the following aspects :-

- **Whether the Data are Suitable for the Purpose of Investigation in View :**
Quite often secondary data do not satisfy immediate needs because they have

been compiled for other purpose. The variation can be in units of measurement, variation in the date/period to which data is related etc.

- **Whether the Data is Adequate for the Investigation :** Adequacy of data is to be judged in the light of the requirements of the survey and the geographical area covered by the available data. For example if our object is to study ;the wage rates of the workers in sugar industry in India and the available data cover only the state of Rajasthan, it would not serve the purpose.
- **Whether the Data are Reliable :** The following points should be checked to find out the reliability of secondary data -
 - (i) The collecting agency was unbiased.
 - (ii) The enumerators are properly trained.
 - (iii) A proper check on the accuracy of the field work.
 - (iv) Was the editing, tabulating and analysis carefully and conscientiously done

□ □ □

Chapter 7

Classification and Tabulation of Data

Q.1 What is the meaning of Classification? Give objectives of Classification and essentials of an ideal classification.

Ans.: Classification is the process of arranging data into various groups, classes and sub-classes according to some common characteristics of separating them into different but related parts.

Main objectives of Classification :-

- (i) To make the data easy and precise
- (ii) To facilitate comparison
- (iii) Classified facts expose the cause-effect relationship.
- (iv) To arrange the data in proper and systematic way
- (v) The data can be presented in a proper tabular form only.

Essentials of an Ideal Classification :-

- (i) Classification should be so exhaustive and complete that every individual unit is included in one or the other class.
- (ii) Classification should be suitable according to the objectives of investigation.
- (iii) There should be stability in the basis of classification so that comparison can be made.
- (iv) The facts should be arranged in proper and systematic way.
- (v) Data should be classified according to homogeneity.
- (vi) It should be arithmetically accurate.

Q.2 What is Manifold Classification?

Ans.: When the data are classified into more than two classes according to more than one attribute, it is called manifold classification.

Q.3 What are Class Limits?

Ans.: The two values which determine a class are known as class limits. First or the smaller one is known as lower limit (L_1) and the greater one is known as upper limit (L_2)

Q.4 How many types of Series are there on the basis of Quantitative Classification? Give the difference between Exclusive and Inclusive Series.

Ans.: There are three types of frequency distributions -

(i) **Individual Series :** In individual series, the frequency of each item or value is only one for example ;marks scored by 10 students of a class are written individually.

(ii) **Discrete Series :** A discrete series is that in which the individual values are different from each other by a different amount.

For example:

Daily wages	5	10	15	20
No. of workers	6	9	8	5

(iii) **Continuous Series :** When the number of items are placed within the limits of the class, the series obtained by classification of such data is known as continuous series.

For example

Marks obtained	0-10	10-20	20-30	30-40
No. of students	10	18	22	25

Difference between Exclusive and Inclusive Series

	Exclusive Series	Inclusive Series
Limits	Upper limit of one class is equal to the lower limit of next class.	The two limits are not equal.
Inclusion	The value equal to the upper limit is included in the next class.	Both upper & lower limits are included in the same class.
Conversion	It does not require any conversion.	Inclusive series is converted into exclusive series for calculation purpose.

Suitability	It is suitable in all situations.	It is suitable only when the values are in integers.
-------------	-----------------------------------	--

Q.5 What is Bivariate Frequency Distribution?

Ans.: A frequency distribution obtained by the simultaneous classification of data according to two characteristics is known as a bivariate frequency distribution.

Q.6 Give Sturges Formula for determining Magnitude of Classes?

Ans.: According to Prof. A. H. Sturges, class interval can be found using the following formula.

$$I = \frac{L - S}{1 + 3.322 \log N}$$

Where -

I = class interval N = No. of observations
L = Largest value S = Smallest value

Q.7 Define Tabulation. State the objectives of Tabulation and kinds of Tables.

Ans.: According to Blair, "*Tabulation in its broad sense is an orderly arrangement of data in columns and rows.*"

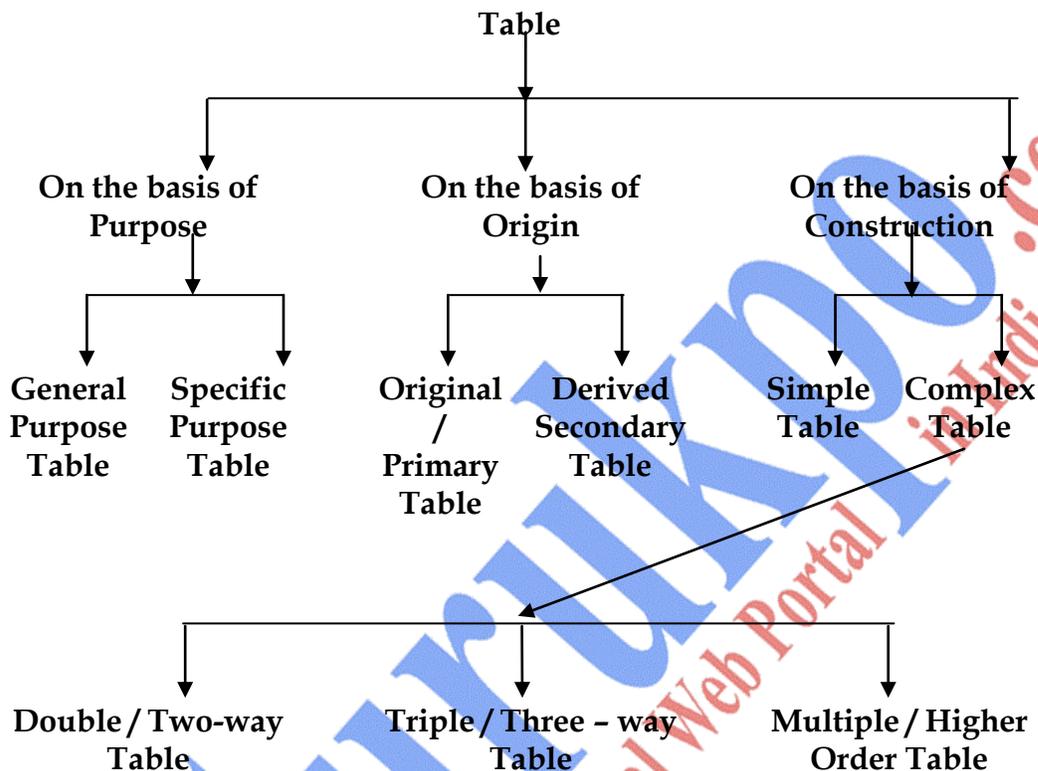
Tabulation is a process of presenting the collected and classified data in proper order and systematic way in columns and rows so that it can be easily compared and its characteristics can be elucidated.

Objects of Tabulation :

- Orderly and systematic presentation of data.
- Making data precise and stable.
- To facilitate comparison.
- To make the problem clear and self evident.

- To facilitate analysis & interpretation of data.

Kinds of Table : The different kinds of tables are shown in the following chart -



Q.8 What are the main parts of a good Table?

Ans.: The number of parts depends mostly on the nature of the data. However, a table should have the following parts.

- (i) **Table No. :** Each table should be numbered so that the table may be referred with that number.
- (ii) **Title :** Every table must be given a suitable title which should be short, clear and complete.
- (iii) **Captions :** Caption refers to the column heading which explains what the column represents.
- (iv) **Stubs :** Stubs are the designations of the rows or row headings.
- (v) **Body :** It is the heart of the table. The body of the table contains the numerical information.

- (vi) **Ruling and Spacing** : Ruling and leaving the space depends on the needs of the topic and makes the table attractive and beautiful.
- (vii) **Footnotes** : In order to explain the figures shown in the table, explanatory notes may be given at the end of the table.
- (viii) **Source** : At the end of the table, the source or origin of given data is mentioned.

□ □ □

Gurukpo.com
No. 1 Educational Web Portal in India

Chapter 8

Diagrammatic and Graphic Presentation

Q.1 what is Diagrammatic Representation? State the importance of Diagrams.

Ans.: Depicting of statistical data in the form of attractive shapes such as bars, circles, and rectangles is called diagrammatic presentation.

A diagram is a visual form of presentation of statistical data, highlighting their basic facts and relationship. There are geometrical figures like lines, bars, squares, rectangles, circles, curves, etc. Diagrams are used with great effectiveness in the presentation of all types of data.

When properly constructed, they readily show information that might otherwise be lost amid the details of numerical tabulation.

Importance of Diagrams :

A properly constructed diagram appeals to the eye as well as the mind since it is practical, clear and easily understandable even by those who are unacquainted with the methods of presentation. Utility or importance of diagrams will become clearer from the following points -

- (i) **Attractive and Effective Means of Presentation:** Beautiful lines; full of various colours and signs attract human sight, and do not strain the mind of the observer. A common man who does not wish to indulge in figures, get message from a well prepared diagram.
- (ii) **Make Data Simple and Understandable :** The mass of complex data, when prepared through diagram, can be understood easily. According to Shri Morane, *"Diagrams help us to understand the complete meaning of a complex numerical situation at one sight only"*.

- (iii) **Facilitate Comparison** : Diagrams make comparison possible between two sets of data of different periods, regions or other facts by putting side by side through diagrammatic presentation.
- (iv) **Save Time and Energy** : The data which will take hours to understand, becomes clear by just having a look at total facts represented through diagrams.
- (v) **Universal Utility** : Because of its merits, the diagrams are used for presentation of statistical data in different areas. It is widely used technique in economic, business, administration, social and other areas.
- (vi) **Helpful in Information Communication** : A diagram depicts more information than the data shown in a table. Information concerning data to general public becomes more easy through diagrams and gets into the mind of a person with ordinary knowledge.

Q.2 Explain in brief the various types of Diagrams?

Ans.: The different types of diagrams can be divided into following heads -

- (1) One dimensional diagrams
- (2) Two dimensional diagrams
- (3) Three dimensional diagrams
- (4) Pictograms
- (5) Cartograms

(1) **One Dimensional Diagrams or Bar Diagrams :**

Bar diagrams are the most common types of diagram. A bar is a thick line whose width is shown merely for attention. They are called one dimensional because it is only the length of the bar that matters and not the width.

Kinds of Bar Diagrams :

- (i) **Line Diagrams** : When the number of items is large, but the proportion between the maximum and minimum is low, lines may be drawn to economise space. Only individual or time series are represented by these diagrams

- (ii) **Simple Bar Diagrams** : A simple bar diagram is used to represent only one variable. For example the figures of sales, production etc. of various years may be shown by means of simple bar diagram. The bars are of equal width only the length varies.

These diagrams are appropriate in case of individual series, discrete series and time series.

- (iii) **Multiple Bar Diagrams** : In a multiple bar diagram, two or more sets of interrelated data are represented. Different shades, colors or dots are used to distinguish between the bars.

These are used to compare two or more related variables based on time and place.

- (iv) **Sub-divided Bar Diagrams** : If a bar is divided into more than one parts, it will be called sub-divided bar diagram. Each component occupies a part of the bar proportional to its share in the total. For example total expenditure incurred by a family on various items such as food, clothing, education, house rent etc can be represented by means of sub-divided bar diagram.

- (v) **Percentage Sub-Divided Bar Diagrams** : Percentage sub-divided bars are particularly useful to measure relative changes of data. When such diagrams are prepared, the length of the bar is kept equal to 100 and segments are cut in these bars to represent the components of an aggregate.

- (vi) **Profit - Loss Diagrams** : If relative change of cost & sales or profit or loss are to be represented with the help of bars, then profit - loss diagram are constructed. These diagrams are similar to percentage sub-divided bars and are prepared in the same way.

- (vii) **Duo-Directional Bar Diagrams** : In duo-directional bar diagram comparative study of two major parts of data is represented in a single bar. Such duo directional diagrams are represented on both sides of the horizontal axis, i.e. above and below the base line.

- (viii) **Paired Bars** : If two different informations which are in different units are to be presented then paired bar diagram are used. These bars are

not vertical but horizontal and the first scale is in the first half and second scale is in the second half.

- (ix) **Deviation Bar Diagram** : Deviation bars are popularly used for representing net quantities i.e. net profit, net loss, net exports or net imports etc. Such bars can have both positive or negative values. Positive values are shown above the base line and negative values below it.
- (x) **Progress Chart or Gant Chart** : These charts are mainly used in factories for comparing the actual production with targeted production. By looking at it, it can be known how much production has been achieved and how much they are lacking behind the capacity.
- (xi) **Pyramid Bar Diagram** : These diagrams are constructed to show population distribution. The distribution of population according to sex, age, education etc are represented by this diagram. In this diagram, the base line is in the middle and its shape is like a pyramid.
- (xii) **Sliding Bar Diagrams** : These bars are like deo-directional bars but instead of absolute figures percentages of two variables are shown. One of them is shown on the right side of the base and the other on its left.

(2) **Two Dimensional Diagrams :**

In two dimensional diagrams, the height as well as the width of the bars will be considered. The area of the bars represents the magnitude of data. Such diagrams are also known as area diagrams or surface diagrams.

The important types are -

- (i) Rectangle diagram
- (ii) Square diagram
- (iii) Circle or Pie diagram
- (i) **Rectangle Diagram:** Rectangles are often used to represent the relative magnitude of two or more values. The area of rectangles is kept in proportion to the values.

They are placed side by side like bars and uniform space is left between different rectangles.

The rectangles may be of different types -

- a) Simple Rectangles
- b) Sub-divided Rectangles
- c) Percentage sub-divided Rectangles

- (ii) **Square Diagrams** : When there is a large difference between the extreme values (example the smallest value is 4 and the biggest value is 800) is such a case square diagram is more appropriate.

First of all square roots of the given values are calculated and the sides are taken in the proportion of square roots. The squares are drawn on the common base line, serially either in increasing or decreasing heights to have beautiful and attractive appearance.

For calculating the scale, the area is calculated by squaring its side, on the basis of that value of 1 sq.cm. is calculated.

- (iii) **Circle or Pie Diagram** : These diagrams are more attractive, therefore, pie diagrams are preferred to square diagrams. These diagrams are used to represent data of population, foreign trade, production etc.

The square roots of the given values are calculated and then it is divided by some common factor so as to attain the radii for the circles. The area of the circle is calculated by the formula - (πr^2) . The sides for squares are taken as the radii for different circles.

A circle can also be sub-divided on the basis of angles to be calculated for each component. There is 360 degree at the centre of the circle and proportionate sectors are cut taking the whole data equal to 360 degrees. Such a circle is known as sub-divided circle or Angular diagram.

- (3) **Three Dimensional Diagrams** :

In three-dimensional diagrams length, width and height (depth) are taken into consideration. If the difference between the minimum and maximum value is so wide as it is difficult to represent them by square or circle diagram, then three dimensional diagram is used. For this cubic roots of the given numbers is calculated. Three dimensional diagrams include cubes, blocks, spheres and cylinders etc.

- (i) **Pictograms** : Pictograms are used by government and non-government organizations for the purpose of advertisement and publicity through appropriate pictures. It is a popular technique particularly when statistical facts are to be presented for a layman having no background of mathematics or statistics.

For representing data relating to social, business and economic phenomena for general masses in fairs and exhibitions, this method is used.

- (ii) **Cartograms** : Cartograms or statistical maps are also used to represent data. Cartograms are simple and elementary form of visual presentation and are very easy to understand.

While highlighting the regional or geographical comparisons, mapographs or cartograms are generally used.

Q.3 What is Graphic Presentation?

Ans.: According to M.M. Blair - "*The simplest to understand, the easiest to make, the most variable and the most widely used type of chart is graph.*"

Graphic presentation is a visual form of presenting statistical data. Graph acts as a tool of analysis and makes complex data simple and intelligible.

Graphs are more appropriate in the following cases -

- a) If tendency instead of real measurement is important.
- b) When comparative study of many data series is required on one graph.
- c) If estimation and interpolation are to be presented by graph.
- d) If frequency distribution is presented by two or more curves.

Q.4 What is False Base Line?

Ans.: While making the graph the vertical scale (y-axis) must start from zero origin. But in some cases adjustment of scale is not possible on (y-axis) starting with zero since the values are big. If the vertical scale begins with zero, the curve will be concentrated mostly on the top of the graph paper.

In such cases that portion of the scale from zero to the minimum value of items to be depicted is omitted. This fact is indicated in the graph paper itself either by drawing a double saw tooth lines in the empty space or by a cut in y-axis. This is known as using false base line. It is drawn in this way to show that Y axis is lost in between.

Q.5 What is Historigram? Give the difference between Absolute Historigram and Index Historigram.

Ans. The curve drawn on graph paper by presenting variables of time series is called 'Historigram', because it reveals the past history of data. A time series is an arrangement of statistical data in a chronological order with reference to occurrence of time. Time period may be a year, quarter, month, week, days, hours etc. Such graphs may be drawn on natural scale or ratio scale.

When actual values of a variable are plotted on a graph paper, it is called absolute historigram. Absolute historigram can be of (i) one variable (ii) two or more variables.

Index historigrams are obtained by plotting index numbers of actual values on a graph paper. If index numbers are not given, the same may be calculated of the original values. It shows relative changes.

Q.6 What are the various types of graphs of frequency distribution?

Ans.: Frequency distribution can also be presented by means of graphs. Such graphs facilitate comparative study of two or more frequency distributions as regards their shape and pattern. The most commonly used graphs are as follows -

- (i) Line frequency diagram
- (ii) Histogram
- (iii) Frequency Polygon
- (iv) Frequency curves
- (v) cumulative frequency curves or Ogine curves

Line Frequency Diagram : This diagram is mostly used to depict discrete series on a graph. The values are shown on the X-axis and the frequencies on the Y axis. The lines are drawn vertically on X-axis against the relevant values taking the height equal to respective frequencies.

Histogram : It is generally used for presenting continuous series. Class intervals are shown on X-axis and the frequencies on Y-axis. The data are plotted as a series of rectangles one over the other. The height of rectangle represents the frequency of that group. Each rectangle is joined with the other so as to give a continuous picture.

Histogram is a graphic method of locating mode in continuous series. The rectangle of the highest frequency is treated as the rectangle in which mode lies. The top corner of this rectangle and the adjacent rectangles on both sides are joined diagonally. The point where two lines intersect each other a perpendicular line is drawn on OX-axis. The point where the perpendicular line meets OX-axis is the value of mode.

Frequency Polygon : Frequency polygon is a graphical presentation of both discrete and continuous series.

For a discrete frequency distribution, frequency polygon is obtained by plotting frequencies on Y-axis against the corresponding size of the variables on X-axis and then joining all the points by a straight line.

In continuous series the mid-points of the top of each rectangle of histogram is joined by a straight line. To make the area of the frequency polygon equal to histogram, the line so drawn is stretched to meet the base line (X-axis) on both sides.

Frequency Curve : The curve derived by making smooth frequency polygon is called frequency curve. It is constructed by making smooth the lines of frequency polygon.

This curve is drawn with a free hand so that its angularity disappears and the area of frequency curve remains equal to that of frequency polygon.

Cumulative Frequency Curve or Ogive Curve : This curve is a graphic presentation of the cumulative frequency distribution of continuous series. It can be of two types -

(a) Less than Ogive and

(b) More than Ogive.

Less than Ogive : This curve is obtained by plotting less than cumulative frequencies against the upper class limits of the respective classes. The points so obtained are joined by a straight line. It is an increasing curve sloping upward from left to right.

More than Ogive : It is obtained by plotting 'more than' cumulative frequencies against the lower class limits of the respective classes. The points so obtained are joined by a straight line to give 'more than ogive'. It is a decreasing curve which slopes downwards from left to right.

Chapter 9

Measures of Central Tendency

Q.1 What do you mean by Measures of Central Tendency? Define Arithmetic Mean, Median and Mode.

Ans. : The central tendency of a variable means a typical value around which other values tend to concentrate; hence this value representing the central tendency of the series is called measures of central tendency or average.

According to Clark, "*Average is an attempt to find one single figure to describe whole of figures.*"

Arithmetic Mean (\bar{X}) : The most popular and widely used measure of representing the entire data by one value is known as arithmetic mean. Its value is obtained by adding together all the items and by dividing this total by the number of items.

Arithmetic mean may be of **two types** :

- Simple Arithmetic mean
- Weighted arithmetic mean

Calculation of Arithmetic Mean :

Individual Series	Discrete & continuous series
Direct Method : $\bar{X} = \frac{\sum X}{N}$ Where - X → Values N → No. of Items	Direct Method : $\bar{X} = \frac{\sum fX}{N}$ Where - fX → Values X frequencies N → Total of frequencies
Shortcut Method : $\bar{X} = A + \frac{\sum dx}{N}$ Where - A → Assumed Mean	Shortcut Method : $\bar{X} = A + \frac{\sum f dx}{N}$

$dx \rightarrow X - A$	
Step-Deviation Method : Not applicable	Step-Deviation Method : $\bar{X} = A + \frac{\sum f dx' \times i}{N}$ Where - $dx' \rightarrow \frac{X - A}{i}$ $i \rightarrow \text{Class interval}$

Median (M) : Median is that value of the variable which divides the group into two equal parts, one part comprising all values greater than, and the other all values less than the median

Calculation of Median :

a) Individual Series :

- Arrange the variables in ascending or descending order.
- $M = \text{size of } \left(\frac{N+1}{2} \right)^{\text{th}} \text{ item}$

Where N \rightarrow No. of items

b) Discrete Series :

- Arrange the variables in ascending or descending order.
- Calculate cumulative frequencies.
- Apply formula $M = \left(\frac{N+1}{2} \right)^{\text{th}} \text{ item}$; N = Total of frequency.
- The value (X) corresponding to this in the cumulative frequency will be the median.

c) Continuous Series :

- Arrange the variables in ascending or descending order.
- Calculate cumulative frequencies.
- Determine median class by using $(N/2)$.
- Apply formula -

$$M = l_1 + \frac{i}{f}(m - c)$$

$l_1 \rightarrow$ lower limit of median group

$i \rightarrow$ class interval

$m \rightarrow N/2$

$c \rightarrow$ cumulative frequency preceding the median group.

$f \rightarrow$ frequency of median group

Mode (Z) : Mode is the value that appears most frequently in a series i.e. it is the value of the item around which frequencies are most densely concentrated.

Calculation of Mode :

a) Individual Series :

- (i) By inspection - value repeated most.
- (ii) By converting individual series into discrete series.
- (iii) By empirical relationship between the averages -

$$Z = 3M - 2X$$

b) Discrete Series :

- (i) By inspection - value having highest frequency.
- (ii) By grouping.
- (iii) By empirical relationship.

c) Continuous Series :

- (i) First calculate modal class by inspection or by grouping.
- (ii) Then apply the following formula -

$$l_1 + \frac{\Delta_1}{\Delta_1 + \Delta_2}$$

where $l_1 \rightarrow$ lower limit of modal class

$\Delta_1 \rightarrow f_1 - f_0$

$\Delta_2 \rightarrow f_1 - f_2$

$f_1 \rightarrow$ frequency of modal class

$f_0 \rightarrow$ frequency of preceding class.

$f_2 \rightarrow$ frequency of succeeding class

$i \rightarrow$ class interval

Q.2 What are the essentials of an Ideal Average.

- Ans.:** (i) Should be easy to understand.
 (ii) Clearly and rigidly defined.
 (iii) Based on all the observations.
 (iv) Simple to compute.
 (v) Least affected by fluctuations.
 (vi) Capable of further Algebraic treatment.
 (vii) Sampling stability.

Q.3 What is Geometric Man? Give Algebraic Characteristics of Geometric Mean and state when Geometric Mean is useful.

Ans.: Geometric mean is the n^{th} root of the product of N items or values.

Calculation of Geometric Mean (G) :

Individual Series	Discrete & Continuous Series
$G = \text{Antilog} \left(\frac{\sum \log X}{N} \right)$	$G = \text{Antilog} \left(\frac{\sum f \log X}{N} \right)$

Algebraic Characteristics of Geometric Mean :

- (ii) The product of the items remains unchanged if each item is replaced by geometric mean.
 (iii) Geometric mean cannot be found if the value of some item in the series is negative or zero.
 (iv) The product of corresponding ratios on either side of the geometric mean is always equal.
 (v) Not affected by changing the sequence of items.

Geometric mean is **appropriate or useful :-**

- When ratios or percentages are to be found.

- In determining rates of increase or decrease.
- When the different values are at vast difference.

Q.4 What is Harmonic Mean? In which circumstances Harmonic Mean is used.

Ans.: Harmonic mean of a series is the reciprocal of the arithmetic mean of the reciprocal of the values of its items.

Calculation of Harmonic Mean (H.M.) :

Individual Series	Discrete & Continuous Series
H.M. = Reciprocal $\left(\frac{\sum \text{Reci.}X}{N} \right)$	H.M. = Reciprocal $\left(\frac{\sum (\text{Reci.}X.f)}{N} \right)$

Harmonic Mean is **used in the following cases :-**

- For determining average speed or velocity.
- To find out average price.
- If the item given in the question which is variable is to be kept as constant in the answer, or vice versa, then harmonic mean will be calculated.

Q.5 What is Partition Value. Give formula for calculating different Partition Values?

Ans.: Values of the items that divide the series into many parts are known as partition values. A variable may be divided into four, five, eight, ten and hundred equal parts known as Quartiles, Quintiles, Octiles, Deciles and Percentiles. The aforesaid partition values gives an idea of the formation of the series which are used in the calculation of dispersion and skewness.

Formula used for Calculating Partition Values :

Measures	Individual & Discrete Series	Continuous Series
Quartiles :		
Q_1	Size of $\left(\frac{N+1}{4}\right)^{\text{th}}$ item	$q_1 = (N/4)^{\text{th}}$ item & $Q_1 = l_1 + \frac{i}{f}(q_1 - c)$
Q_3	Size of $3\left(\frac{N+1}{4}\right)^{\text{th}}$ item	$q_3 = 3(N/4)^{\text{th}}$ item & $Q_3 = l_1 + \frac{i}{f}(q_3 - c)$
Quintiles : Q_{n_4}	Size of $4\left(\frac{N+1}{5}\right)^{\text{th}}$ item	$q_{n_4} = 4(N/5)^{\text{th}}$ item & $Q_3 = l_1 + \frac{i}{f}(q_{n_4} - c)$
Octiles : O_2	Size of $2\left(\frac{N+1}{8}\right)^{\text{th}}$ item	$o_2 = 2(N/8)^{\text{th}}$ item & $O_2 = l_1 + \frac{i}{f}(o_2 - c)$
Deciles : D_7	Size of $7\left(\frac{N+1}{10}\right)^{\text{th}}$ item	$d_3 = 7(N/10)^{\text{th}}$ item & $D_3 = l_1 + \frac{i}{f}(d_7 - c)$
Percentiles: P_{75}	Size of $75\left(\frac{N+1}{100}\right)^{\text{th}}$ item	$7_{75} = 75(N/100)^{\text{th}}$ item & $P_{75} = l_1 + \frac{i}{f}(p_{75} - c)$

Q.6 How are Arithmetic Mean, Geometric Mean and Harmonic Mean related to each other? Why is Arithmetic Mean greater than the Geometric Mean and Geometric Mean is greater than the Harmonic Mean.

Ans.: Relation between arithmetic mean, geometric mean and harmonic mean can be given by -

$$(i) \quad \bar{X} > g > h \quad \text{and} \quad (ii) \quad g^2 = \bar{X}.h.$$

While calculating arithmetic mean, the bigger values are given more weightage than the small values, whereas in harmonic mean the smaller values are given much more weightage than to the larger values. Therefore, arithmetic mean is greater than geometric mean and geometric mean is greater than harmonic mean.

Chapter 10

Measures of Dispersion

Q.1 What do you mean by Dispersion. Give the meaning of Absolute Measure and Relative Measure with example.

Ans. : Dispersion is a measure of the extent to which the individual item vary from a central value Dispersion is used in two senses, (i) difference between the extreme items of the series and (ii) average of deviation of items from the mean.

Absolute Measure : The figure showing the limit or magnitude of dispersion is known as absolute measure and it is shown in the same unit as those of the original data, example measures of dispersion in the age of students, their height, weight etc.

Relative Measure : For comparative study the concerning absolute measure is divided by the corresponding mean or some other characteristic value to obtain a ratio or percentage, which is known as the relative measure.

Q.2 Explain the various methods for measuring Dispersion. Also give their merits and demerits?

Ans.: Following are the important methods of studying dispersion -

(1) **Numerical Methods:**

a) Methods of limits

i) Range

ii) Inter-quartile range

iii) Percentile Range

b) Methods of average deviation:

i) Quartile Deviation

ii) Mean Deviation

iii) Standard Deviation

(2) **Graphic Method - Lorenz Curve:**

i) **Range:** The difference between the value of the smallest item and the value of the largest item of the series is called range.

Range = Largest item - Smallest item

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

Merits of Range :

- Simple and easy to be computed.
- It takes minimum time to calculate.
- Not necessary to know all the values, only smallest and largest value is required.
- Helpful in quality control of products.

Demerits of Range :

- Not based on all the items.
- Subject to fluctuation/uncertain measure.
- Cannot be computed in case of open-end distributions.
- As it is not based on all the values it is not considered as a good or appropriate measure.

ii) **Inter-Quartile Range :** Inter-quartile range represents the difference between the third-quartile and the first quartile. It is also known as the range of middle 50% values.

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

Merits:

- It is easy to calculate.
- Can be measured in open end distributions.
- It is least affected by the uncertainty of the extreme values.

Demerits:

- It does not represent all the values.
- It is an uncertain measure.
- It is very much affected by sampling fluctuations.

iii) **Percentile Range:** It is the difference between the values of the 90th and 10th percentile. It is based on the middle 80% items of the series.

$$\text{Percentile Range} = P_{90} - P_{10}$$

Merits and Demerits : Its use is limited. Percentile range has almost the same merits and demerits as those of inter-quartile range.

iv) **Quartile Deviation / Semi - Interquartile Range :**

Quartile Deviation gives the average amount by which the two quartiles differ from the median. Quartile deviation is an absolute measure of dispersion.

It is calculated by using the formula -

$$\text{Q.D.} = Q_3 - Q_1, \text{ coefficient of Q.D.} = \frac{Q_3 - Q_1}{Q_3 + Q_1}$$

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

$$Q_3 + Q_1$$

Merits:

- It is easy to calculate and understand.
- It has a special utility in measuring variation in open end distributions.
- QD is not affected by the presence of extreme values.

Demerits :

- It is very much affected by sampling fluctuations.
- It does not give an idea of the formulation of the series.
- It is not capable of further algebraic treatment.

- v) **Mean Deviation (δ)** : Mean Deviation is also known as average deviation or first measure of dispersion. It is the average difference between the items in a distribution and the median mean or mode of that series.

Computation of Mean Deviation:

Individual Series	Discrete & continuous series
a) Deviation from Mean : $\delta\bar{X} = \frac{\sum d\bar{x} }{N}, d\bar{x} = X - \bar{X} $ Coefficient of $\delta\bar{X} = \frac{\delta\bar{X}}{\bar{X}}$	a) Deviation from Mean : $\delta\bar{X} = \frac{\sum f d\bar{x} }{N}$
b) Mean Deviation from Median : $\delta M = \frac{\sum dM }{N}, dM = X - M $ Coefficient of $\delta M = \frac{\delta M}{M}$	b) Mean Deviation from Median : $\delta M = \frac{\sum f dM }{N}$

<p>c) Mean Deviation from Mode : $\delta Z = \frac{\sum dZ }{N}, dZ = X - Z$ Coefficient of $\delta Z = \frac{\delta Z}{M}$</p>	<p>b) Mean Deviation from Mode : $\delta M = \frac{\sum f dZ }{N}$</p>
--	--

Merits :

- It is simple to understand and easy to compute.
- Based on each and every item of the data.
- Least affected by the extreme values.
- It can be computed from any average, mean, median or mode.

Demerits :

- Signs are ignored therefore mathematically it is incorrect and not a significant measure.
- Cannot be compared if mean deviations of different series are based on different averages.
- Does not give accurate results.

vi) **Standard Deviation (σ) :**

Standard Deviation was introduced by Karl Pearson in 1823. It is the most important and widely used measure of studying dispersion, as it is free from those defects from which the earlier methods suffer and satisfies most of the properties of a good measure of dispersion.

Standard Deviation is the square root of the average of the square deviations from the arithmetic mean of a distribution.

Co-efficient of Standard Deviation :

Standard deviation is an absolute measure. When comparison of variability in two or more series is required, relative measure of standard deviation is computed which is called coefficient of standard deviation.

Individual Series	Discrete & Continuous Series
<p>Direct Method :</p> $\sigma = \sqrt{\frac{\sum d^2}{N}}$ <p>Where - $d^2 \rightarrow (X - \bar{X})^2$ $N \rightarrow$ No. of Items</p>	<p>Direct Method :</p> $\sigma = \sqrt{\frac{\sum fd^2}{N}}$
<p>Shortcut Method :</p> $\sigma = \sqrt{\frac{\sum dx^2}{N} - \left(\frac{\sum dx}{N}\right)^2}$ <p>Where - $dx^2 \rightarrow (X - A)^2$ $A \rightarrow$ Assumed Mean</p>	<p>Shortcut Method :</p> $\sigma = \sqrt{\frac{\sum fdx^2}{N} - \left(\frac{\sum fdx}{N}\right)^2}$

Computation of Standard Deviation:

Coefficient of S.D. = σ / \bar{X}

Variance = σ^2

Coefficient of variation: Coefficient of variation is used for the comparative study of stability or homogeneity in more than two or more series.

$$C.V = \sigma \times \frac{100}{\bar{X}}$$

Merits:

- Based on all the items.
- Well-defined and definite measure of dispersion.
- Least affected by sample fluctuations.
- Suitable for algebraic treatment.

Demerits :

- Standard Deviation is comparatively difficult to calculate.
- Much importance is given to the extreme values.

vii) **Graphical Method - Lorenz Curve :**

This curve was devised by Dr. Max O'Lorenz. He used this technique to show inequality of wealth and income of a group of people. It is simple, attractive and effective measure of dispersion, yet it is not scientific since it does not provide a figure to measure dispersion.

Merits :

- Easy to understand from the graph.
- Comparison can be done in two or more series.
- Attractive & effective.
- Concentration or density of frequency can be known from the curve.

Demerits :

- It is not a numerical measure; therefore it is not definite as a mathematical measure.
- Drawing of curve requires more time and labour.

Q.3 What is the best method of measuring Dispersion. Write the formula for calculating combined S.D.

Ans.: Standard Deviation is the best method of measuring dispersion as deviations taken from mean and algebraic signs are not ignored and it is algebraically correct.

Formula for calculating combined Standard Deviation :

$$\sigma_{12} = \sqrt{\frac{N_1 (\sigma_1^2 + D_1^2) + N_2 (\sigma_2^2 + D_2^2) + \dots}{N_1 + N_2}}$$

Where -

N_1 & $N_2 \rightarrow$ No. of items of two series

σ_1 & $\sigma_2 \rightarrow$ S.D. of the two series

$D_1 \rightarrow \bar{X}_1 - \bar{X}_{12}$

$D_2 \rightarrow \bar{X}_2 - \bar{X}_{12}$

$\bar{X}_{12} \rightarrow$ Combined arithmetic mean

Chapter 11

Correlation

Q.1 What is Correlation. State the different types and degrees of Correlation.

Ans.: If two series vary in such a way, that fluctuations in one are accompanied by the fluctuations in the other, these variables are said to be correlated. Like rise in price of a commodity, reduces its demand and vica-versa.

Some relationship exists between age of husband and wife, rainfall and production. Two variables are said to be correlated if the change in one variable results in a corresponding change in the other variable.

According to A. M. Tuttle, "*Analysis of co-variation of two or more variables is usually called correlation*".

Types of Correlation : Correlation can be of following types -

- (i) **Positive and Negative Correlation :** If changes in two connected series is in the same direction, i.e. increase in one variable is associated with increase in other variable, the correlation is said to be positive. For example increase in father's age, increase in son's age.

If the two related series change in opposite direction i.e. increase in one variable is associated with the corresponding decrease in other variable, the correlation is said to be negative.

- (ii) **Linear an Non-Linear :** If the amount of change in one variable tends to bear constant ratio of change in the other variable, the correlation is said to be linear. We get a straight line if the variables of these series are marked on graph paper.

Correlation would be called non-linear or curvilinear if the amount of change in one variable does not bear a constant ratio to the amount of change in the other variable. For example if we double the amount of rainfall the production would not necessarily be doubled.

- (iii) **Simple, Partial and Multiple Correlation** : When only two variables are studied, it is called simple correlation.

If the common effect of two or more independent variables on one dependent data series is studied, it is called multiple correlations. For example if the study of rain, soil, temperature on potato production per acre is studied then it is multiple correlation.

On the other hand, in partial correlation we recognize more than two variables, but consider only two variables to be influencing each other, the effect of other influencing variables being kept constant.

Degree of Correlation : The interpretation of co-efficient of correlation is based on the degree of correlation. The coefficient may be in the following degrees -

- (i) **Perfect Correlation** :

- a) Perfect positive correlation ($r = +1$)
 b) Perfect negative correlation ($r = -1$)

- (ii) **Absence of Correlation or No Correlation** :

$$r = 0$$

- (iii) **Limited Degree of Correlation** :

- | | |
|--|----------------------|
| (a) High degree positive or negative | ± 0.75 to 1.00 |
| (b) Moderate degree positive or negative | ± 0.25 to 0.75 |
| (c) Low degree positive or negative | 0 to ± 0.25 |

Q.3 Explain the Mathematical Methods of finding out Correlation.

Ans.: Correlation coefficient can be determined by the following **methods** -

- (i) **Karl Pearson's Coefficient of Correlation (r)** : Karl Pearson's Coefficient of Correlation is widely used in practice. It is an assumption of Karl Pearson's coefficient of correlation that linear relations exist in both the series.

This method is considered as the best measure because it provides the knowledge of directions of changes in data i.e. positive or negative, and also shows the degree of correlation which should always lie between $+1$ and -1 .

Computation of Karl Pearson's Coefficient of Correlation :

Direct Method	Short cut Method
<p>Formula -</p> <p>(1) $r = \frac{\sum xy}{\sqrt{(\sum x^2 \times \sum y^2)}}$</p> <p>(2) $r = \frac{\sum xy}{N \cdot \sigma_x \cdot \sigma_y}$</p> <p>Here $x = X - \bar{X}$ & $Y = Y - \bar{Y}$ $N \rightarrow$ No. of observations $\sigma_x \rightarrow$ Standard deviation of X-series $\sigma_y \rightarrow$ Standard deviation of Y-series</p>	<p>Formula -</p> <p>(1) $r = \frac{\sum dx dy - N(\bar{X} - A)(\bar{Y} - A)}{N \cdot \sigma_x \cdot \sigma_y}$</p> <p>(2) $r = \frac{N \cdot \sum dx dy - (\sum dx \times \sum dy)}{\sqrt{[N \cdot \sum d^2 x - (\sum dx)^2][N \cdot \sum d^2 y - (\sum dy)^2]}}$</p> <p>Here $dx = X - A_x$ & $dy = Y - A_y$ $A \rightarrow$ Assumed mean</p>

- (ii) **Spearman's Rank Correlation:** When it is not possible to measure the perfect quantities due to absence of numerical facts, ranking figures are used. These ranks are determined according to the size of the data. Using ranks rather than actual observations gives the coefficient of rank correlation.

This method was developed by the British Psychologist Charles Edward Spearman in 1904. This measure is especially useful to measure honesty, beauty, skill, wisdom etc.

Computation of Rank Correlation :

- First assign the ranks to the data of both the series. Rank can be assigned by taking either the highest value as I or the lowest value as 1. But whether we start with the lowest value or the highest value we must follow the same method in case of both the variables.
- Take the difference of two ranks ($R_1 - R_2$) i.e. (D)
- Square the rank difference i.e. (D^2)
- Apply the **formula -**

$$r_R = \frac{6 \sum D^2}{N(N^2 - 1)}$$

Where -

$N \rightarrow$ No. of pairs of observations.

$$D \rightarrow (R_1 - R_2)$$

In case of Tied Ranks or Equal Ranks: If some values are equal in a distribution than average rank is given to those items. The formula for calculation of correlation will be :

$$R_p = 1 - \frac{6 \left[\sum \frac{d^2 + 1}{12} (m^3 - m) + \frac{1}{12} (m^3 - m) \right]}{N(N^2 - 1)}$$

M \rightarrow No. of items where ranks are common.

- (iii) **Concurrent Deviation Method:** This method is simplest of all methods. It is useful to know correlation of short-term changes in time-series. Sometimes it is not necessary to know the actual degree of correlation but it is important to know the direction of correlation i.e. positive or negative. For this purpose concurrent Deviation method is appropriate.

Computation of Correlation:

- Compare each item of both the series from its previous item. If the second item is bigger put a (+) sign; if it is smaller put a (-); and if it is equal, put a sign of (=).
- Multiply the signs of both the series.
- Total the positive signs = (C) No. of concurrent deviations.
- Apply the **formula:**

$$r_c = \pm \frac{\pm (2C - N)}{N}$$

Where -

C \rightarrow No. of concurrent deviation

N \rightarrow No. of pairs - 1

Q. 4 What is Probable Error and Standard Error? What are the rules of testing the significance of coefficient of Correlation.

Ans. Probable Error is an expected error in Karl Pearson's coefficient of correlation, which facilitates the upper limit and lower limit of possible correlation. With the help of probable error it is possible to determine the reliability of the value of the coefficient in so far as it depends on the conditions of random sampling. Formula to calculate probable error :

$$\text{P.E.} = \frac{0.6745 \times (1 - r^2)}{\sqrt{N}}$$

Where -

$r \rightarrow$ Coefficient of correlation.

$N \rightarrow$ No. of items

Standard Error: Now a day's Standard Error is preferred to probable error. Degree of standard error is standard deviation of sampling distribution in any statistics of random sampling.

Other things remaining the same, it is essential that standard error should be small as far as possible. Smaller the standard error, more uniformity will be there in sampling distribution.

$$\text{S.E. of } r = \frac{(1 - r^2)}{\sqrt{N}}$$

Rules for Testing the Significance of Coefficient of Correlation:

- (i) If coefficient of correlation is more than 6 times of probable error ($r > 6 \text{ P.E.}$), it is significant.
- (iii) If r is less than 0.3 P.E ($r < 0.3 \text{ P.E.}$) it is insignificant.

□□□

Chapter 12

Linear Regression

Q.1 Define Regression. Why are there two Regression Lines? Under what conditions can there be only one Regression Line?

Ans.: Regression literally means “return” or “go back”. In the 19th century, Francis Galton at first used regression in his paper “Regression towards Mediocrity in Hereditary Stature” for the study of hereditary characteristics.

Use of regression in modern times is not limited to hereditary characteristics only but it is widely used for the study of expected dependence of one variable on the other.

Therefore, the method by which best probable values of unknown data of a variable are calculated for the known values of the other variable is called regression.

Regression helps in forecasting, decision making and in studying two or more variables in economic field. It also shows the direction, quality and degree of correlation.

Regression Lines :

Regression line is that line which gives the best estimate of dependent variable for any given value of independent variable. If we take the case of two variables X and Y, we shall have two regression lines as the regression of X on Y and the regression of Y on X.

Regression Line X and Y : In this formation, Y is independent and X is dependent variable, and best expected value of X is calculated corresponding to the given value of Y.

Regression Line Y on X : Here Y is dependent and X is independent variable, best expected value of Y is estimated equivalent to the given value of X.

An important reason of having two regression lines is that they are drawn on least square assumption which stipulates that the sum of squares of the deviations from different points to that line is minimum. The deviations from the points from the line of best fit can be measured in two ways - vertical, i.e. parallel to Y - axis, and horizontal i.e. parallel to X axis.

For minimizing the total of the squares separately, it is essential to have two regression lines.

Single line of Regression : When there is perfect positive or perfect negative correlation between the two variables ($r = \pm 1$) the regression lines will coincide or overlap and will form a single regression line in that case.

Q.2 What is the difference between Regression and Correlation?

Ans.: Difference between Correlation & Regression :

Basis	Correlation	Regression
Relationship	Correlation tells the degree of average relationship between two or more variables.	Using relationship between known variables and unknown variables, the unknown variable is estimated
Cause & Effect	Unable to tell which series is the cause and which is the effect, despite high degree of correlation.	The given independent variable is the 'cause' and dependent variable is the effect.
Change in scale	Unaffected by the change of origin and scale.	Regression coefficient is affected by change in scale.
Application	Limited Application.	Wider Application.

Q.3 How are Regression Equations derived? Explain.

Ans.: Computation of Regression Equations : Algebraic expression of regression lines is called regression equations. Like lines equations are also two :

Regression Equation X on Y	Regression Equation Y on X
Original Form : $X = a + by$	Original Form : $Y = a + bx$
Formula : $\bar{X} - X = b_{xy} (Y - \bar{Y})$	Formula : $Y - \bar{Y} = b_{yx} (X - \bar{X})$

Here - $\bar{X} \rightarrow$ Mean of x series. $\bar{Y} \rightarrow$ Mean of y series. $b_{xy} \rightarrow$ Regression coefficient X on Y	Here - $\bar{X} \rightarrow$ Mean of x series. $\bar{Y} \rightarrow$ Mean of y series. $b_{yx} \rightarrow$ Regression coefficient Y on X
Computation of Regression Coefficients :	
1. Direct Method : $b_{xy} = \frac{\sum xy}{\sum y^2}$ or $b_{xy} = r \frac{\sigma_x}{\sigma_y}$	1. Direct Method : $b_{yx} = \frac{\sum xy}{\sum x^2}$ or $b_{yx} = r \frac{\sigma_y}{\sigma_x}$
2. Short Cut Method : $b_{xy} = \frac{N \cdot \sum dx dy - (\sum dx \cdot \sum dy)}{N \cdot \sum d^2 y - (\sum dy)^2}$ Here - $dx = X - A$ $dy = Y - A$	2. Short Cut Method : $b_{yx} = \frac{N \cdot \sum dx dy - (\sum dx \cdot \sum dy)}{N \cdot \sum d^2 x - (\sum dx)^2}$ Here - $dx = X - A$ $dy = Y - A$

Q.4 Give the interpretation of Regression Coefficients. How Correlation is calculated from Regression Coefficients?

Ans.: Interpretation of Regression Coefficients :

- (i) If both the coefficients are positive correlation coefficient will be positive, and if both the coefficients are negative, the coefficient of correlation will also be negative.
- (ii) Both the regression coefficients will have the same sign.
- (iii) The product of both the regression coefficients cannot be more than 1.

Calculation of Correlation Coefficient from Regression Coefficient :

$$r = \sqrt{(b_{xy}) \times (b_{yx})}$$

$b_{xy} \rightarrow$ Regression coefficient x on y

$b_{yx} \rightarrow$ Regression coefficient y on x.

Chapter 13

Index Numbers

Q.1 What is Index Numbers? Give the importance or utility of Index Numbers?

Ans.: Index Numbers are specialized averages designed to measure the change in a group of related variables over a period of time. Index numbers have today become one of the most widely used statistical devices and there is hardly any field where they are not used.

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 location or other characteristics such as income, profession etc.”*

Importance of Index Numbers :

- (i) **Help in Framing Suitable Policies :** Index numbers provide guidance in administrative policy determination. For example in determining dearness allowance of the employees cost of living index are used.
- (ii) **They Reveal Trends :** Index numbers are most widely used for determining the commercial and industrial trends. By examining the index numbers of industrial production of last few years, we can conclude about the trend of production.
- (iii) **Helps in Comparative Study :** Data which cannot be compared with the help of simple averages, index numbers can be used because they are in relative form.
- (iv) **Important in Forecasting :** Index numbers are often used in time series analysis, the historical study of long-term trend, seasonal variations etc.
- (v) **Universal Application :** Index numbers are useful in economic, commercial, social and in every field such as agriculture etc.

Q.2 What is Base Year. Distinguish between Fixed Base Method and Chain Base Method.

Ans.: Base year is such a year with which the changes in the current year are compared. Base year should be normal year from every aspect and it should not be affected by war, famines, inflation etc.

Base year should not be too old. The index for base period is always taken as 100.

In fixed base method, the year selected for construction of index numbers remains constant for all times and the base shall remain fixed.

While in chain base method the base year is changed every year, generally preceding year and not fixed year.

Q.3 Give the types of Weights. Why are Weights used in construction of Index Numbers?

Ans.: The term 'Weight' refers to the relative importance of the different items in the construction of the index. For example in daily use the importance of wheat and rice is more than jute and iron.

Weights can be classified as follows -

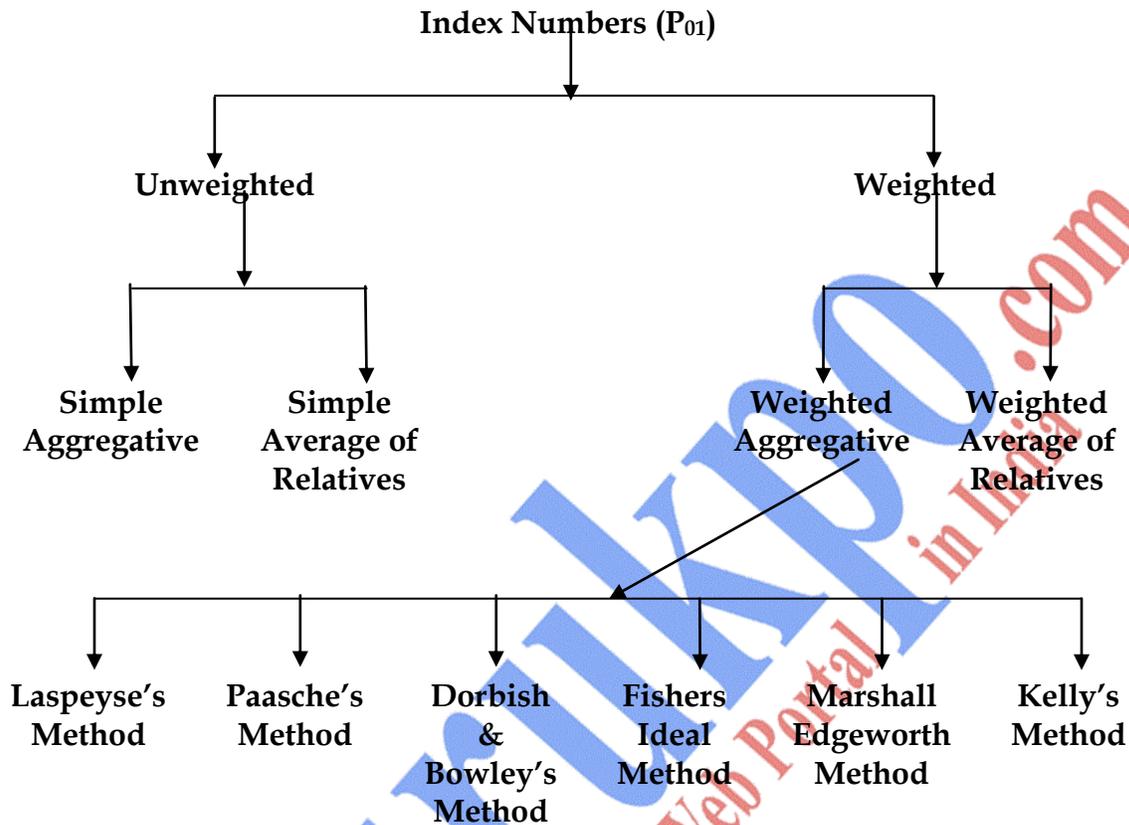
- (i) **Explicit and Implicit weights** : Explicit weights are called direct weight. They may be in the form of quantity or value. In implicit weighing, a commodity or its variety is included in the index a numbers of times.
- (ii) **Fixed and Fluctuating Weight** : If same weights are used from year to year, they are called fixed weights. If weights are changed from time to time, they are called fluctuating weights.
- (iii) **Arbitrary and Real Weights** : If real quantities or units are used, then those weights are called real weights.

If some unrealistic weights, according to the own will and assumptions of the investigator are used, they are called arbitrary weights.

Weights are used in construction of Index numbers because all items are not of equal importance and hence it is necessary to device some suitable method where by the varying importance of the different items is taken into account. This is done by allocating weights.

Q.4 Explain the various methods of constructing Index Numbers?

Ans.: Various methods of calculating index numbers can be shown by the following chart -



(I) **Unweighted Index Numbers :**

(a) **Simple Aggregative Method :**

$$P_{01} = \frac{\sum P_1}{\sum P_0} \times 100$$

Where -

$\sum P_1 \rightarrow$ Total of current year prices of various commodities.

$\sum P_0 \rightarrow$ Total of base year prices of various commodities.

(b) **Simple Average of Relative Method :**

$$P_{01} = \frac{\sum R}{N} \quad \text{and} \quad R = \frac{P_1}{P_0} \times 100$$

Where -

$P_1 \rightarrow$ Prices of current year.

$P_0 \rightarrow$ Prices of base year.

Base year can be price of one year or multi-year or average price.
Geometric mean can also be used for averaging the price relatives.

(II) **Weighted Index Numbers:**

(a) **Weighted Aggregative Method:**

(i) **Lapsler's Method:**

$$P_{01} = \frac{\sum P_1 q_0}{\sum P_0 q_0} \times 100$$

Where -

P_1 → Price of current year.

P_0 → Price of base year.

q_0 → Quantity of base year

(ii) **Paasche's Method:**

$$P_{01} = \frac{\sum P_1 q_1}{\sum P_0 q_1} \times 100$$

Where -

q_1 → Quantity of current year.

(iii) **Dorbish & Bowley's Method:**

$$P_{01} = \frac{\text{Laspeyre's Index} + \text{Paasche's Index}}{2}$$

$$\text{or } P_{01} = \left(\frac{\frac{\sum P_1 q_0}{\sum P_0 q_0} + \frac{\sum P_1 q_1}{\sum P_0 q_1}}{2} \right) \times 100$$

(iv) **Marshall - Edge worth's Method:**

$$P_{01} = \left(\frac{\frac{\sum P_1 q_0 + \sum P_1 q_1}{\sum P_0 q_0 + \sum P_0 q_1}}{2} \right) \times 100$$

(v) **Fisher's Ideal Method:**

$$P_{01} = \sqrt{\left(\frac{\sum P_1 q_0}{\sum P_0 q_0} \times \frac{\sum P_1 q_1}{\sum P_0 q_1} \right)} \times 100 \quad \text{or } \sqrt{L \times P}$$

(iv) **Kelly's Method:**

$$P_{01} = \frac{\sum P_1 q}{\sum P_0 q} \times 100$$

$$\text{Where - } q = \frac{q_0 + q_1}{2}$$

- (b) **Weighted Average of Relatives:** In this method, the price relatives for each commodity are obtained and these price relatives are multiplied by the value weights for each item and the product is divided by the total of weights.

$$\text{Consumer price index} = \frac{\sum RW}{\sum W}$$

$$\text{Where } R = \frac{P_1}{P_0} \times 100 \text{ and } W = P_0 q_0$$

Q.5 Fisher's Formula is called the Ideal Formula. Why?

Ans.: Fisher's formula is called the ideal because of the following reasons:

- (i) It is based on geometric mean which is considered best for constructing index numbers.
- (ii) It fulfills both the time reversal and factor reversal tests.
- (iii) It takes into account both current year as well as base year prices and quantities.
- (iv) It is free from bias.

Q.6 Explain the various tests of adequacy of Index Number.

Ans.: Following tests are applied for choosing an appropriate formula -

- (i) **Time Reversal Test :** Time reversal test is a test to determine whether a given method will work both ways in time, forward and back ward. This test means that if the index number of the current year is computed on the basis of base year (P_{01}) and again the index number of the base year is calculated on the basis of current year (P_{10}), both the index numbers should be reciprocal of each other i.e. product of both of them should be one.

Symbolically, the following relation should be satisfied -

$$P_{01} \times P_{10} = 1$$

- (ii) **Factor Reversal Test :** According to this test, an index of price, if multiplied by an index of quantity, with the same base, given years and commodities, should be equal to the true value ratio, i.e. equal to the ratio of current year price to total price of the base year.

$$P_{01} \times Q_{01} = \frac{\sum P_1 q_1}{\sum P_0 q_0}$$

If the product is not equal to the value ratio, there is, with reference to this test, an error in one or both of the index numbers.

- (iii) **Circular Test:** This test is just an extension of the time reversal test. The test requires that if an index is constructed for the year 'a' on base year 'b' and for the year 'b' on base year 'c', we ought to get the same result as if we calculated directly an index for 'a' on a base year 'c' without going through 'b' as an intermediary. So, if there are three indices P_{01} , P_{12} and P_{20} , the circular test will be satisfied if :

$$P_{01} \times P_{12} \times P_{20} = 1$$

Q.7 What are the methods of constructing Consumer Price Index or Cost of Living Index Numbers?

Ans.: Consumer price index or cost of living index numbers are designed to measure the effect of changes in the price of group of commodities and services on the purchasing power of a particular class of society, during any given period with reference to some fixed base. Its object is to find out how much the consumers of a particular class have to pay more for a certain basket of goods and services in a given period as compared to the base period.

Following two methods are used to construct consumer price index numbers -

- (i) Family Budget method or weighted average of relatives method.
- (ii) Aggregative Expenditure Method.

Family Budget Method : In this method, the price relatives for each commodity are obtained and these price relatives are multiplied by the value weights for each item and the product is divided by the total of weights.

$$\text{Consumer price index} = \frac{\sum RW}{\sum W}$$

$$\text{Where } R = \frac{P_1}{P_0} \times 100 \text{ and } W = P_0 q_0$$

This method is same as weighted average of price relatives method.

Aggregative Expenditure Method : This method is most popular method for constructing consumer price index numbers. The prices of commodities of current year as well as base year are multiplied by the quantities consumed in the base year. The aggregate expenditure of current year is divided by the aggregate expenditure of the base year and the quotient is multiplied by 100.

$$\text{Consumer price Index} = \frac{\Sigma P_1 q_0}{\Sigma P_0 q_0} \times 100$$

This method is same as Laspeyre's method.

Q.8 What do you understand by base Shifting? Give the formula for converting the Fixed Base Index Number to Chain Base Index Number.

Ans.: Sometimes it becomes necessary to change the base year used for calculating index number of a series from one period to another without returning to the original date. This change of reference base period is usually referred to as 'shifting the base'.

Formula for converting Fixed Base Index to Chain Base Index -

$$\text{Chain Base Index No.} = \frac{\text{Fixed base Index of Current Year}}{\text{Fixed base Index of Previous Year}} \times 100$$

Q.9 What is Splicing?

Ans.: When a series of index numbers is prepared by taking a particular year as base and then it is stopped, and the last year of that series is taken as the base year and another series is prepared. At times, it may be necessary to convert these two series into a continuous series. This procedure is called splicing.

If old series is to be related with new series, this is called forward splicing and to convert old series on the basis of new series is known as backward splicing.

a) **Formula for Forward Splicing :**

$$\frac{\text{Old Index No. of new base year} \times \text{Index No. to be adjusted}}{100}$$

b) **Formula for Backward Splicing :**

$$\frac{\text{Index No. to be adjusted} \times 100}{\text{Old Index No. of new base year}}$$

Q.10 What do you mean by Deflating of Index Numbers?

Ans.: Deflating means making allowances for the effect of changing price levels. A rise in price level means a reduction in the purchasing power of money.

The process of adjusting a series of salary or wages or income according to current price changes to find out the level of real salary, wages or income is called deflating of index numbers. It is necessary when price level is increasing and cost of living is also increasing.

$$\text{Index No. of real wage} = \frac{\text{Real wage of current year}}{\text{Real wage of base year}} \times 100$$

Where

$$\text{Real wage} = \frac{\text{Money wage or income}}{\text{Price Index or Cost of living index}} \times 100$$

□ □ □

Gurukpo.com
No. 1 Educational Web Portal in India

Practical Questions

Q.1 Calculate Mean, Mode and Median for the following data :

Wages in `	Below 13	13-15	15-17	Below 19	Below 21	21-23	Below 25
No. of Labourers	30	50	60	210	260	40	330

Solution.

Calculation of Mean, Mode and Median

Wages in `	Frequency f	M.V.	X-A A = 18dx	Step dx'	fdx'	c.f.
11-13	30	12	-6	-3	-90	30
15-15	50	14	-4	-2	-100	80
15-17	60	16	-2	-1	-60	140
17-19	70 (210-140)	18	0	0	0	210
19-21	50 (260-210)	20	+2	+1	50	260
21-23	40	22	+4	+2	80	300
23-25	30(330-300)	24	+6	+3	90	330
Total	N=330				$-30\sum fdx'$	

$$\text{Mean : } \bar{X} = A + \frac{\sum fdx'}{N} \times i = 18 + \frac{(-30 \times 2)}{330} = 18 - \frac{6}{33} \text{ or } 18 - 0.18 = \text{ ` } 17.82$$

Mode (z) : By inspection method, the modal group is 17-19

$$Z = l_1 + \frac{\Delta}{\Delta_1 + \Delta_2} \times i \text{ where } \Delta_1 = 70 - 60 = 10, \Delta_2 = 70 - 50 = 20$$

$$\text{Thus, } Z = 17 + \frac{10}{10+20} \times 2 \text{ or } 17 + 0.67 = \text{ ` } 17.67$$

Median $m = \frac{N^{th}}{2}$ item $\frac{330}{2} = 165^{th}$ item. It lies in the median group = (17-19)

$$M = l_1 + \frac{i}{f} (m-c) \text{ or } 17 + \frac{2}{70} (165-140) \text{ or } 17 = \frac{2 \times 25}{70} \text{ or } 17 = 0.71 = \text{ ` } 17.71$$

Q.2 Complete mean and standard deviation of the marks obtained by 100 students in an examination.

Marks	0.10	10.20	20.30	30.40	40.50	Total
Frequency	12	21	23	34	10	100

Computation of Mean and Std. Deviation

Marks	No. of Students (f)	Mid Value (X)	dx A = 25, i = 10	fdx'	fd^2x' (4x5)
0-10	12	5	-2	-24	48
10-20	21	15	-1	-21	21
20-30	23	25	0	0	0
30-40	34	35	+1	34	34
40-50	10	45	+2	20	40
Total	100 (N)	-		$\Sigma fdx' = 54 - 45 = 9$	143

$$\bar{X} = A \frac{\Sigma fdx'}{N} \quad \text{or } \bar{X} = 25 + \frac{9}{100} \times 10 \text{ or } 25.9 \text{ marks}$$

$$\text{or } \sigma = \frac{i}{N} \sqrt{\Sigma fd^2x' \cdot N - (\Sigma fdx')^2}$$

$$\text{or } \sigma = \frac{10}{100} \sqrt{143 \times 100 - (9)^2}$$

$$\sigma = \frac{1}{10} \sqrt{14300 - 81} = 11.92 \text{ Marks}$$

Thus $\bar{X} = 25.9$ Marks and $\sigma = 11.92$ marks.

Q.3 The results of B.Com. Examination are given below: From the following data, calculate coefficient of correlation between age and success in the examination.

Age in Years	18	19	20	21	22	23	24	25	26	27
% of failure	38	40	35	32	34	37	42	46	52	56

Sol. In the problem the percentage of failures is given, but coefficient of correlation is required to ascertain in age and success in examination. As such [100- percentage of failure] will be the % of success in the examination. Thus series shall be presented as under:

Age in years (X)	dx A = 23	d^2x	% of success 100-% of failure (Y)	dx A = 59	d^2y	$dx dy$
18	-5	25	62	3	9	-15
19	-4	16	60	1	1	-4
20	-3	9	65	6	36	-18
21	-2	4	68	9	81	-18
22	-1	1	66	7	49	-7
23	0	0	63	4	16	0

24	1	1	58	-1	1	-1
25	2	4	54	-5	25	-10
26	3	9	48	-11	121	-33
27	4	16	44	-15	225	-60
(N=10)	Σdx = -5	Σd^2x = 85		Σdy = -2	Σd^2y = 564	$\Sigma dxdy$ = -166

$$r = \frac{\Sigma dxdy \times N - (\Sigma dx)(\Sigma dy)}{\sqrt{\Sigma d^2x \times N - (\Sigma dx)^2} \times \sqrt{\Sigma d^2y \times N - (\Sigma dy)^2}}$$

Substituting the values formula, we get:

$$r = \frac{-166 \times 10 - (-5)(-2)}{\sqrt{85 \times 10 - (-5)^2} \times \sqrt{564 \times 10 - (-2)^2}}$$

$$= \frac{-1660 - 10}{\sqrt{850 - 25} \times \sqrt{5640 - 4}}$$

$$= \frac{-1670}{2156.3} = -0.774$$

$$= -\text{A.L.} [\log \text{ of } 1670 - \frac{1}{2} (\log 825 + \log 5636)]$$

$$= -\text{A.L.} [\log 3.2227 - \frac{1}{2} (2.9165 + 3.7513)]$$

$$= -\text{A.L.} [3.2227 - 3.3339] = -\text{A.L.} \bar{1}.8888 \text{ or } -0.7741$$

It shows high degree of negative correlation.

Q.4 Calculate the regression equations from the following data:

X	1	2	3	4	5	6	7	8	9
Y	9	8	10	12	11	13	14	16	15

X	dx A = 5	d ² x	Y	dy A = 12	d ² y	dxdy
1	-4	16	9	-3	9	12
2	-3	9	8	-4	16	12
3	-2	4	10	-2	4	4
4	-1	1	12	0	0	0
5	0	0	11	-1	1	0
6	1	1	13	+1	1	1
7	2	4	14	+2	4	4
8	3	9	16	+4	16	12

Sol.

9	4	16	15	+3	9	12
45	0	60	108	0	60	57

Thus $N = 9, \Sigma X = 45,$
 $\Sigma Y = 108, \Sigma d^2 x = 60,$
 $\Sigma d^2 y = 60, \Sigma dx dy = 57,$

$$\bar{X} = \frac{\Sigma X}{N} = \frac{45}{9} \text{ or } 5$$

Regression Coefficients (b) :

x on y

$$b_{xy} = \frac{\Sigma dx dy}{N \sigma_y^2} \text{ or } \frac{\Sigma dx dy}{\Sigma d_y^2}$$

$$= \frac{57}{60} = 0.95$$

Regression Equation x on y

$$(X - \bar{X}) = b_{xy}(Y - \bar{Y})$$

$$\text{Or } (X - 5) = .95(Y - 12)$$

$$\text{Or } X = .95y - 11.4 + 5$$

$$\text{Or } X = 0.95y - 6.4$$

$$\bar{Y} = \frac{\Sigma Y}{N} = \frac{108}{9} \text{ or } 12$$

Regression Coefficients (b)

y on x

$$b_{yx} = \frac{\Sigma dx dy}{N \sigma_x^2} \text{ or } \frac{\Sigma dx dy}{\Sigma d_x^2}$$

$$= \frac{57}{60} = 0.95$$

Regression Equation y on x

$$(Y - \bar{Y}) = b_{yx}(X - \bar{X})$$

$$\text{Or } (Y - 12) = .95(X - 5)$$

$$\text{Or } y = 0.95x - 4.74 + 12$$

$$\text{Or } y = .95x + 7.25$$

Q.5 An enquiry into the budgets of middle class families of Jodhpur gave the following information:

Expenses on		Price 2011	Price 2012	Standard Price (C)
Food	35%	150	145	100
Rent	15%	30	30	25
Clothing	20%	75	65	50

Fuel	10%	25	25	20
Miscellaneous	20%	40	45	40

What changes in the cost of living figures of 2011 and 2012 are seen as compared to standard price? Also find out the percentage increase in prices in 2012 as compared to 2011?

Sol. This question is of Family Budget Method. Hence, first price relative (R_1 & R_2) will be computed on the basis of standard prices by using following formula:

$$R_1 = \frac{P_1}{P_0} \times 100, R_2 = \frac{P_2}{P_0} \times 100$$

Construction of Cost of Living Index Nos.

Groups	Std. Price P_0	Price 2011 P_1	Price 2011 P_2	R_1	P_2	W (%)	$P_1 W$	$P_w W$
Food	100	150	145	150	145	35	5250	5075
Rent	25	30	30	120	120	15	1800	1800
Clothing	50	75	65	150	130	20	3000	2600
Fuel	20	25	25	125	125	10	1250	1250
Miscellaneous	40	40	45	100	112.5	20	2000	2250
Total	P_0	P_1	P_2	-	-	100 (ΣW)	13,300 ($\Sigma R_1 W$)	12975 ($\Sigma R_2 W$)

$$\text{Cost of Living Index for 2011} = \frac{\Sigma R_1 W}{\Sigma W} \text{ or } \frac{13300}{100} = 133.00$$

$$\text{Cost of Living Index for 2012} = \frac{\Sigma R_2 W}{\Sigma W} \text{ or } \frac{12975}{100} = 129.75$$

$$\text{Prices in 2012 have fallen } (133 - 129.75) = 3.25 \text{ Percentage decline} = \frac{3.25}{133} \times 100 = 2.44\%$$