तेजस्वि नावधीतमस्तु
ISO 9001:2008 & 14001:2004

FAIRFIELD
Institute of Management & Technology
Managed by 'The Fairfield Foundation'
(Affiliated to GGSIP University, New Delhi)

GURU GOBIND SINGH INDRAPRASTHA UNIVERSITY, DELHI
BACHELOR OF COMMERCE (Hons)

BCOM 110- Business Statistics **L-5 T/P-0 Credits-5**

Objectives: The objective of this course is to familiarize students with the basic statistical tools used to summarize and analyze quantitative information for decision making.

COURSE CONTENTS

Unit I **Lectures: 20**

Statistical Data and Descriptive Statistics: Measures of Central Tendency: Mathematical averages
including arithmetic mean, geometric mean and harmonic mean, properties and applications, positional averages, mode, median (and other partition values including quartiles, deciles, and percentile;

Unit II **Lectures: 15**

Measures of variation: absolute and relative, range, quartile deviation, mean deviation, standard deviation, and their co-efficients, properties of standard deviation/variance; Moments: calculation and significance; Skewness, Kurtosis and Moments.

Unit III **Lectures: 15**

Simple Correlation and Regression Analysis: Correlation Analysis, meaning of correlation simple, multiple and partial; linear and non-linear, Causation and correlation, Scatter diagram, Pearson co-efficient of correlation; calculation and properties, probable and standard errors, rank correlation; Simple Regression Analysis: Regression equations and estimation.

Unit IV **Lectures: 20**

Index Numbers: Meaning and uses of index numbers, construction of index numbers, univariate and composite, aggregative and average of relatives – simple and weighted, tests of adequacy of index numbers, Base shifting, problems in the construction of index numbers.

Text Books:
1. Levin, Richard and David S. Rubin. (2011), Statistics for Management. 7th Edition. PHI.
2. Gupta, S.P., and Gupta, Archana, (2009), Statistical Methods. Sultan Chand and Sons, New Delhi.

Reference Books:
1. Berenson and Levine, (2008), Basic Business Statistics: Concepts and Applications. Prentice Hall.

BUSINESS STATISTICS:PAPER CODE 110

Dr. Shweta Gupta

Asst. professor (Management)

FIMT

Unit-1

STATISTICAL DATA AND DESCRIPTIVE STATISTICS

LESSON STRUCTURE

1.1 Introduction

1.2 Definitions of Measures of Central Tendency

1.3 Mathematical averages

1.4 Arithmetic Mean, Definition Properties and Applications

1.5 Geometric Mean, Definition Properties and Applications.

1.6 Harmonic Mean, Definition Properties and Applications.

1.7 positional averages

1.8 Mode, Definition Properties and Applications.

1.9 Median, Definition Properties and Applications.

2.0 Partition Values, Definition Properties and Applications.

1.1 INTRODUCTION

In a general sense we can say statistics means numerical fact and figures like GDP, Per Capita Income, Sales, profit, Etc. as a subject of study statistics refer to the body of principles and procedure developed for the collection ,classification, summarization and interpretation of numerical data for some special purpose.

In plural Sense stat is a set of Data or Numerical statement of Fact and in singular sense statistics is a process collection classification tabulation analysis and interpretation of data. As the name the Descriptive statistics merely describe the data and consists of the methods and techniques used in the collection, organization, presentation and analysis of data in order to describe the various features and characteristics of such data. these methods either be graphical or computational.

1.2 DEFINITIONS OF MEASURES OF CENTRAL TENDENCY

"an average is a single figure that represents the whole group-clark.

1.3MATHEMATICAL AVERAGES

a- Arithmetic Mean
b- -Geometric Mean
c- Harmonic Mean

1.4ARITHMETIC MEAN, DEFINITION AND PROPERTIES

Arithmetic Mean is defined by some of items Divided by number of items.

Properties-

(1)The sum of the deviation of the items from arithmetic mean is always Zero.

(2)The sum of the squared deviation of the items from arithmetic mean is minimum.

(3)If each item of a series is increased, decreased, multiplied or is divided by the same constant the A.M. is also affected accordingly.

(4)if each item of the original series is replaced by the actual mean then the sum of these substitutions will be equal to the sum of the individual items.

1.5 GEOMETRIC MEAN, DEFINITIO N AND APPLICATIONS

Geometric mean is defined at the *nth* root of the product of *n* observations of a distribution.

Applications

The geometric mean is most suitable in the following three cases:
1. Averaging rates of change.
2. The compound interest formula.
3. Discounting, capitalization

1.6 HARMONIC MEAN, DEFINITION PROPERTIES AND APPLICATIONS

The harmonic mean is defined as the reciprocal of the arithmetic mean of the reciprocals of individual observations.

Applications

It is useful in finding average involving speed, time, price and Ratios.

1.7 POSITIONAL AVERAGES

Positional average Describe Position of the Average value in Whole of distribution. There are two positional average Median and Mode.

1.8 MODE, DEFINITION AND APPLICATIONS

Mode is the value at the point around which the items are most heavily concentrated.

Application

when companies try to find out that which size are variety should be focused for the purpose of production in such condition Mode is the average that help them

1.9 MEDIAN, DEFINITION PROPERTIES AND APPLICATIONS

" The Median is that value of the variables which divides the group into two equal parts,one part comprising all values greater and the other less than the median."

Properties

An important property of Median is that the sum of the absolute deviations of the items from the median is less than the sum from any other value or average..

Application

we have need to catteries the variables than Median will we helpful average.

2.0 PARTITION VALUES, DEFINITION PROPERTIES AND APPLICATIONS

Just as Median there are other useful measures which divide the series into 4,8,10 and 100 equal parts, they are called quartiles deciles and percentiles.

UNIT-2

MEASURES OF VARIATION

LESSON STRUCTURE

1.1 INTRODUCTION

The dispersion or variability provides us one more step in increasing our understanding of the pattern of the data that how variables are scattered. Understanding of variability of data is very necessary to reach on any decision.

DEFINITIONS OF DISPERSION AND PROPERTIES OF GOOD DISPERSION

1. "Dispersion is the measure of the variation of the items." -A.L. Bowley
2. "The degree to which numerical data tend to spread about an average value is called the variation of dispersion of the data." -Spiegel

A good measure of dispersion should possess the following properties

1. It should be simple to understand.
2. It should be easy to compute.
3. It should be rigidly defined.
4. It should be based on each and every item of the distribution.
5. It should be amenable to further algebraic treatment.
6. It should have sampling stability.
7. Extreme items should not unduly affect it

1.2 MEANING OF ABSOLUTE AND RELATIVE MEASUREMENT

Absolute measure is not able to comparative study if the measuring unit are different .so for the purpose of comparison we find out respective measurement that's generally called coefficient of that particular measurement.

1.3 DIFFERENT MEASURE OF VARIATION

1.3(A) RANGE

The simplest measure of dispersion is the range, which is the difference between the maximum value and the minimum value of data.

LIMITATIONS OF RANGE

There are some limitations of range, which are as follows:
1. It is based only on two items and does not cover all the items in a distribution.
2. It is subject to wide fluctuations from sample to sample based on the same Population.
3. It fails to give any idea about the pattern of distribution. This was evident from the data given in Examples 1 and 3.
4. Finally, in the case of open-ended distributions, it is not possible to compute The range.

1.3(B) INTERQUARTILE RANGE AND QUARTILE DEVIATION

Interquartile range denotes the difference between the third quartile and the first Quartile. When interquartile range divided by two its called semi interquartile range or quartile deviation.

MERITS OF QUARTILE DEVIATION

The following merits are entertained by quartile deviation:
1. As compared to range, it is considered a superior measure of dispersion.
2. In the case of open-ended distribution, it is quite suitable.
3. Since it is not influenced by the extreme values in a distribution, it is particularly suitable in highly skewed or erratic distributions.

LIMITATIONS OF QUARTILE DEVIATION
1. Like the range, it fails to cover all the items in a distribution.
2. It is not amenable to mathematical manipulation.
3. It varies widely from sample to sample based on the same population.
4. Since it is a positional average, it is not considered as a measure of dispersion. It merely shows a distance on scale and not a scatter around an average.
In view of the above-mentioned limitations, the interquartile range or the quartile deviation has a limited practical utility.

1.3(C) MEAN DEVIATIO

The mean deviation is also known as the average deviation. As the name implies, it is the average of absolute amounts by which the individual items deviate from the mean.Since the positive deviations from the mean are equal to the negative deviations,while computing the mean deviation, we ignore positive and negative signs.

MERITS OF MEAN DEVIATION
1. A major advantage of mean deviation is that it is simple to understand and easy to calculate.
2. It takes into consideration each and every item in the distribution. As a result, a change in the value of any item will have its effect on the magnitude of mean deviation.
3. The values of extreme items have less effect on the value of the mean deviation.
4. As deviations are taken from a central value, it is possible to have meaningful comparisons of the formation of different distributions.

LIMITATIONS OF MEAN DEVIATION
1. It is not capable of further algebraic treatment.
2. At times it may fail to give accurate results. The mean deviation gives best results when deviations are taken from the median instead of from the mean. But in a series, which has wide variations in the items, median is not a satisfactory measure.
3. Strictly on mathematical considerations, the method is wrong as it ignores the algebraic signs when the deviations are taken from the mean.
In view of these limitations, it is seldom used in business studies. A better measure known as the standard deviation is more frequently used.

1.3(D) STANDARD DEVIATION, , PROPERTIES OF STANDARD

STANDARD DEVIATION

The standard deviation is similar to the mean deviation in that here too the deviations are measured from the mean. At the same time, the standard deviation is preferred to the mean deviation or the quartile deviation or the range because it has desirable mathematical properties.

PROPERTIES OF STANDARD STANDARD DEVIATION

1. Standard deviation is only used to measure spread or dispersion around the mean of a data set.
2. Standard deviation is never negative.
3. Standard deviation is sensitive to outliers. A single outlier can raise the standard deviation and in turn, distort the picture of spread.
4. For data with approximately the same mean, the greater the spread, the greater the standard deviation.
5. If all values of a data set are the same, the standard deviation is zero (because each value is equal to the mean).

COEFFICIENT OF VARIATION

The coefficient of variation (CV) is the ratio of the standard deviation to the mean. The higher the coefficient of variation, the greater the level of dispersion around the mean. It is generally expressed as a percentage.
More value of CV present less consistency, and less value of CV present more consistency.

1.3(D) SKEWNESS: MEANING AND DEFINITIONS

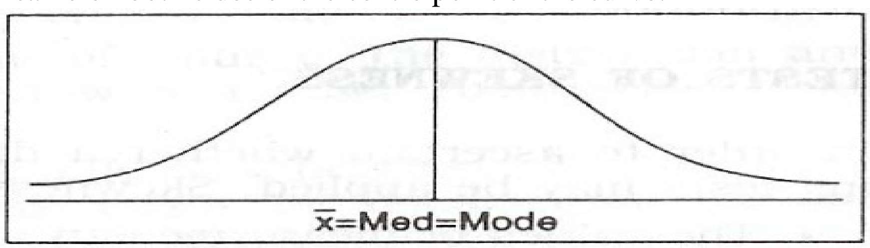
skewness help us to understand a distribution. Two distributions may have the same mean and standard deviation but may differ widely in their overall appearance as can be seen from the following. In both these distributions the value of mean and standard deviation is the same (X = 15, σ = 5). But it does not imply that the distributions are alike in nature.
The distribution on the left-hand side is a symmetrical one whereas the distribution on the right-hand side is symmetrical or skewed. Measures of skewness help us to distinguish between different types of distributions.

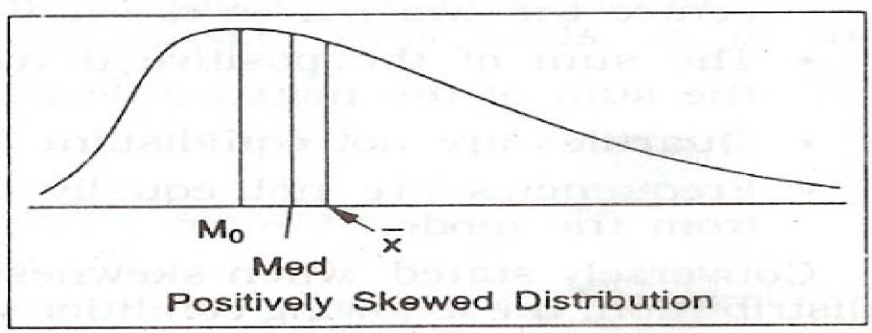
Some important definitions of skewness are as follows:
1. "When a series is not symmetrical it is said to be asymmetrical or skewed."
-Croxton & Cowden.
2. "Skewness refers to the asymmetry or lack of symmetry in the shape of a

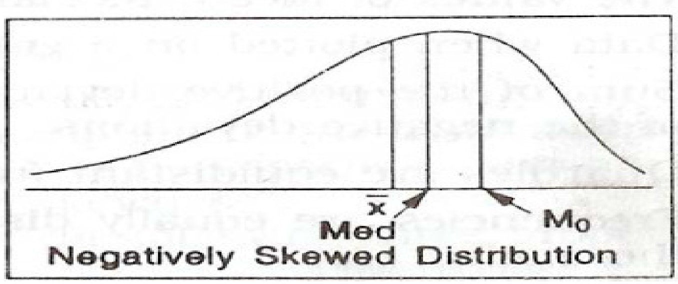
Frequency distribution." -Morris Hamburg.

The above definitions show that the term 'skewness' refers to lack of symmetry" i.e.,
When a distribution is not symmetrical (or is asymmetrical) it is called a skewed Distribution. The concept of skewness will be clear from the following three diagrams showing a Symmetrical distribution. a positively skewed distribution and a negatively skewed Distribution.

1. **Symmetrical Distribution.** It is clear from the diagram (a) that in a symmetrical distribution the values of mean, median and mode coincide. The spread of the frequencies is the same on both sides of the centre point of the curve.







2. Asymmetrical Distribution. A distribution, which is not symmetrical, is called a skewed distribution and such a distribution could either be positively skewed or negatively skewed as would be clear from the diagrams (b) and (c).

3. Positively Skewed Distribution.
In the positively skewed distribution the value of the mean is maximum and that of mode least- the median lies in between the two as is clear from the diagram (b)

4. Negatively Skewed Distribution.

The following is the shape of negatively skewed distribution. In a negatively skewed distribution the value of mode is maximum and that of mean least-the median lies in between the two. In the positively skewed distribution the frequencies are spread out over a greater range of values on the high-value end of the curve (the right-hand side) than they are on the low-value end. In the negatively skewed distribution the position is reversed, i.e. the excess tail is on the left-hand side. It should be noted that in moderately symmetrical distributions the interval between the mean and the median is approximately one-third of the interval between the mean and the mode. It is this relationship, which provides a means of measuring the degree of skewness.

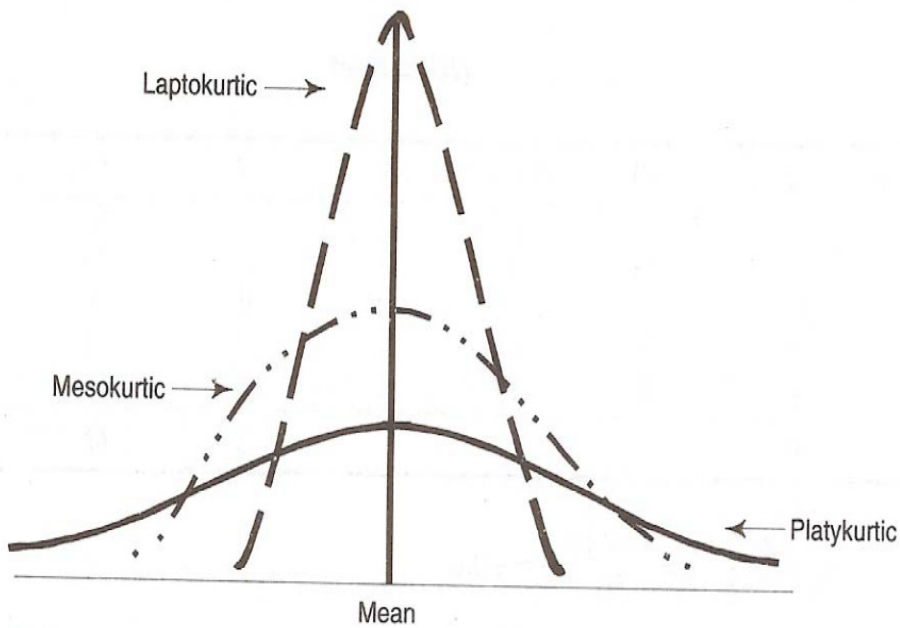
MEASURES OF SKEWNESS

There are four measures of skewness, each divided into absolute and relative measures. The relative measure is known as the coefficient of skewness and is more frequently used than the absolute measure of skewness. Further, when a comparison between two or more distributions is involved, it is the relative measure of skewness, which is used. The measures of skewness are: (i) Karl Pearson's measure, (ii) Bowley's measure, (iii) Kelly's measure, and (iv) Moment's measure.

1.3(E) KURTOSIS

Kurtosis is another measure of the shape of a frequency curve. It is a Greek word, which means bulginess. While skewness signifies the extent of asymmetry, kurtosis measures the degree of peakedness of a frequency distribution. Karl Pearson classified curves into three types on the basis of the shape of their peaks. These are mesokurtic, leptokurtic and platykurtic. These three types of curves are shown in figure below:

It will be seen from Fig.



/pes of Curves

that mesokurtic curve isneither too much flattened nor too much peaked. In fact, this is the frequency curve of a normal distribution. Leptokurtic curve is a more peaked than the normal curve. In contrast, platykurtic is a relatively flat curve. In case of a normal distribution, that is, mesokurtic curve, the value of β_2=3. If β_2 turn out to be > 3, the curve is called a leptokurtic curve and is more peaked than the normal curve.Again, when $\beta_2 < 3$, the curve is called a platykurtic curve and is less peaked than the normal curve. The measure of kurtosis is very helpful in the selection of an appropriate average. For example, for normal distribution, mean is most appropriate; for a leptokurtic distribution, median is most appropriate; and for platykurtic distribution, the quartile range is most appropriate.

1.3(F) MOMENTS: CALCULATION AND SIGNIFICANCE

In mechanics, the term *moment* is used to denote the rotating effect of a force. In Statistics, it is used to indicate peculiarities of a frequency distribution. The utility of moments lie,s in the sense that they indicate different aspects of a given distribution. Thus, by using moments, we can measure the central tendency of a series, dispersion or variability, skewness and the peakedness of the curve. The moments about the actual arithmetic mean are denoted by μ. The first four moments about mean or *central moments* are as follows: It may be noted that the first central moment is zero, that is, $\mu = 0$.The second central moment is $\mu_2 = \sigma$, indicating the variance.
The third central moment μ_3 is used to measure skewness. The fourth central moment gives an idea about the Kurtosis.

UNIT -3

SIMPLE CORRELATION AND REGRESSION ANALYSIS

LESSON STRUCTURE

. 1.1 INTRODUCTION

The statistical methods of ***Correlation*** and ***Regression*** are helpful in knowing the relationship between two or more variables which may be related in same way, *like* interest rate of bonds and prime interest rate; advertising expenditure and sales; income and consumption; crop-yield and fertilizer used; height and weights and so on.

In all these cases involving two or more variables, we may be interested in seeing:

1. if there is any association between the variables;
2. if there is an association, is it strong enough to be useful;
3. if so, what form the relationship between the two variables takes;
4. how we can make use of that relationship for predictive purposes, that is, forecasting;
5. how good such predictions will be.

Since these issues are inter related, correlation and regression analysis, as two sides of a single process, consists of methods of examining the relationship between two or more variables. If two (or more) variables are correlated.

1.2 MEANING OF CORRELATION

"The correlation between variables is a measure of the nature and degree of association between the variables".

1.3 TYPE OF CORRELATION

Correlation can be classified in several ways. The important ways of classifying correlation are:

(i) Positive and negative,

(ii) Linear and non-linear (curvilinear) and

(iii) Simple, partial and multiple.

Positive and Negative Correlation

If both the variables move in the same direction, we say that there is a positive correlation, *i. e.,* if one variable increases, the other variable also increases on an average or if one variable decreases, the other variable also decreases on an average.
On the other hand, if the variables are varying in opposite direction, we say that it is a case of negative correlation; *e. g.,* movements of demand and supply.

Linear and Non-linear (Curvilinear) Correlation

If the change in one variable is accompanied by change in another variable in a constant ratio, it is a case of linear correlation. Observe the following data:
X : 10 20 30 40 50
Y : 25 50 75 100 125
The ratio of change in the above example is the same. It is, thus, a case of linear correlation. If we plot these variables on graph paper, all the points will fall on the same straight line.
On the other hand, if the amount of change in one variable does not follow a constant ratio With the change in another variable, it is a case of non-linear or curvilinear correlation. If a Couple of figures in either series X or series Y are changed, it would give a non-linear Correlation.

Simple, Partial and Multiple Correlations

The distinction amongst these three types of correlation depends upon the number of variables involved in a study. If only two variables are involved in a study, then the correlation is said to be simple correlation. When three or more variables are involved in a study, then it is a problem of either partial or multiple correlation. In multiple correlation, three or more variables are studied simultaneously. But in partial correlation we consider only two variables influencing each other while the effect of other variable(s) is held constant.

1.4 CORRELATION AND CAUSATION

Correlation helps us to determine the degree of correlation between two or more variables but does not focus the causes and effect relationship. In some cases there is no relation between variables still correlation exists but Existence of causation always implies correlation**.** The high degree of correlation may be due to following reasons

1-Mutual Dependence-sometime there is a high degree of correlation between two variables but it may be difficult to point out the causes and effect variables as exam. The demand of the product due to growth of population or due season, income, competition etc.

2-Influence of Third variable-

Sometime the degree of correlation between two variables may be due to Effect of a third variable. They have acted upon both the variable causing them to respond together. However, neither of the two is the cause of the other.

3-Pure Chance

Sometime there is no functional relation between variables still high degree correlation found between them. It's may be a chance or due to biasness of investigator.

1.5 SCATTER DIAGRAM

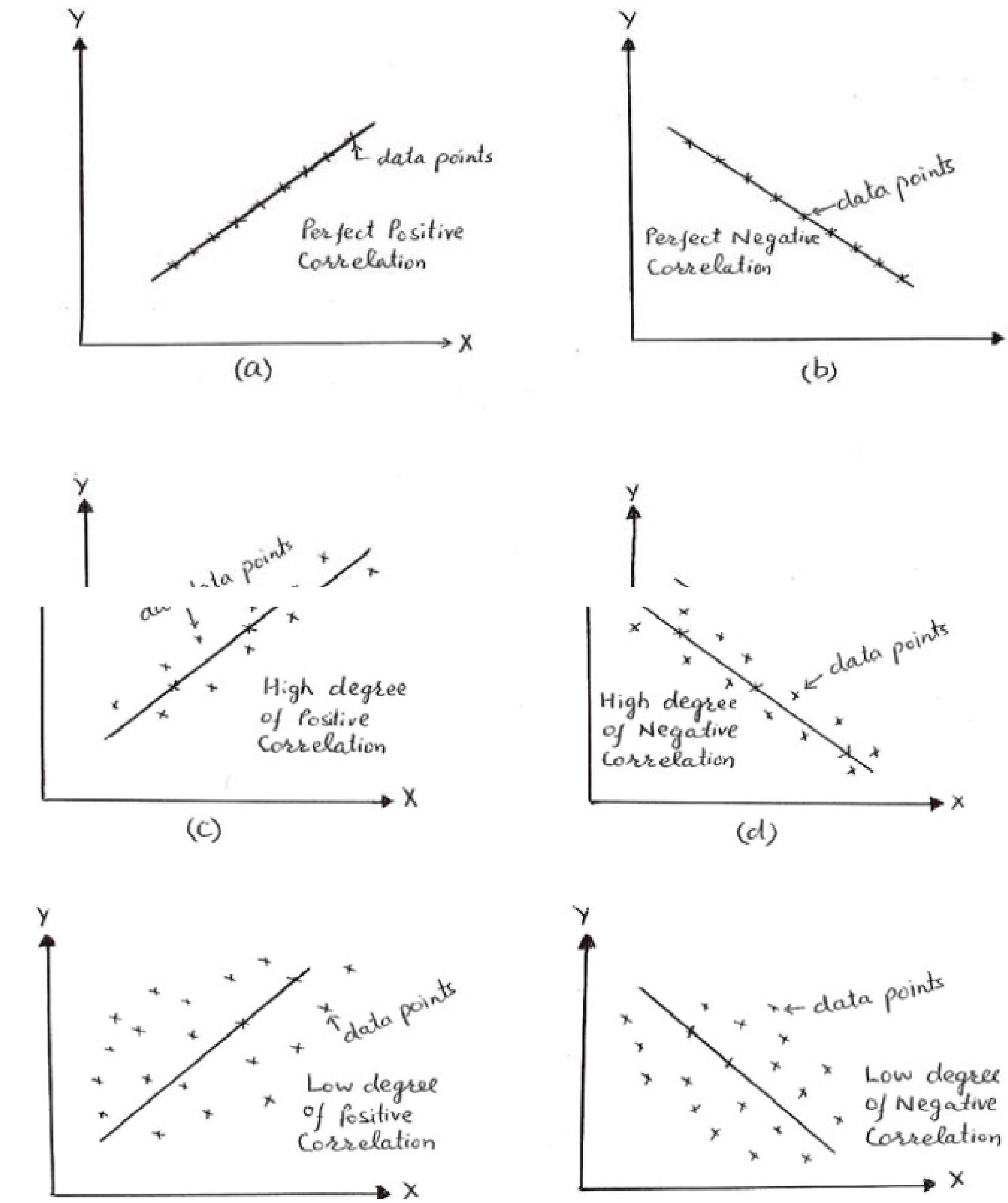
SCATTER DIAGRAM

This method is also known as Dotogram or Dot diagram. Scatter diagram is one of the implest methods of diagrammatic representation of a bivariate distribution. Under this method, both the variables are plotted on the graph paper by putting dots. The diagram so obtained is called "Scatter Diagram". By studying diagram, we can have rough idea about the nature and degree of relationship between two variables. The term scatter refers to the spreading of dots on the graph. We should keep the following points in mind while interpreting correlation:

1. if the plotted points are very close to each other, it indicates high degree of correlation. If the plotted points are away from each other, it indicates low degree of correlation. if the points on the diagram reveal any trend (either upward or downward), the variables are said to be correlated and if no trend is revealed, the variables are uncorrelated.
2. if there is an upward trend rising from lower left hand corner and going upward to the upper right hand corner, the correlation is positive since this reveals that the values of the two variables move in the same direction. If, on the other hand, the points depict a downward trend from the upper left hand corner to the lower right hand corner, the correlation is negative since in this case the values of the two variables move in the opposite directions.
3. in particular, if all the points lie on a straight line starting from the left bottom and going up towards the right top, the correlation is perfect and positive, and if all the points like on a straight line starting from left top and coming down to right bottom, the correlation is perfect and negative. if the points on the diagram reveal any trend (either upward or downward), the variables are said to be correlated and if no trend is revealed, the variables are uncorrelated.
4. if there is an upward trend rising from lower left hand corner and going upward to the upper right hand corner, the correlation is positive since this reveals that the values of the two variables move in the same direction. If, on the other hand, the points depict a downward trend

from the upper left hand corner to the lower right hand corner, the correlation is negative since in this case the values of the two variables move in the opposite directions.

5. in particular, if all the points lie on a straight line starting from the left bottom and going up towards the right top, the correlation is perfect and positive, and if all the points like on a straight line starting from left top and coming down to right bottom, the correlation is perfect and negative.



PROPERTIES OF PEARSON CO-EFFICIENT OF CORRELATION,

A mathematical method for measuring the intensity or the magnitude of *linear relationship* between two variables was suggested by Karl Pearson (1867-1936), a great British Biometrician and Statistician and, it is by far the most widely used method in practice.

Karl Pearson's measure, known as Pearsonian correlation coefficient between two variables X and Y, usually denoted by $r(X, Y)$ or r_{xy} or simply r is a numerical measure of linear relationship between them and is defined as the ratio of the covariance between X and Y, to the product of the standard deviations of X and Y.

Properties of Pearsonian Correlation Coefficient

1. *Pearsonian correlation coefficient cannot exceed 1 numerically*. In other words it lies between –1 and +1.

Value of *r* Degree of correlation

±1 perfect correlation
±0.90 or more very high degree of correlation
±0.75 to ±0.90 sufficiently high degree of correlation
±0.60 to ±0.75 moderate degree of correlation
±0.30 to ±0.60 only the possibility of a correlation
less than ±0.30 *possibly no correlation*
0 absence of correlation

2. *Pearsonian Correlation coefficient is independent of the change of origin and scale*.
3. *Two independent variables are uncorrelated but the converse is not true*

If X and Y are independent variables then

$r_{xy} = 0$

However, the converse of the theorem is not true *i.e.,* uncorrelated variables need not necessarily be independent. As an illustration consider the following bivariate distribution.

X : 1 2 3 -3 -2 -1
Y : 1 4 9 9 4 1

For this distribution, value of r will be 0.

Hence in the above example the variable X and Y are uncorrelated. But if we examine the data carefully we find that X and Y are not independent but are connected by the relation $Y = X^2$. The above example illustrates that uncorrelated variables need not be independent.

4. Pearsonian coefficient of correlation is the geometric mean of the two regression coefficients,

5. *The square of Pearsonian correlation coefficient is known as the coefficient of determination.*

1.6 PROBABLE AND STANDARD ERRORS

Probable Error of Correlation Coefficient

The correlation coefficient establishes the relationship of the two variables. After ascertaining this level of relationship, we may be interested to find the extent upto which this coefficient is dependable. Probable error of the correlation coefficient is such a measure of testing the reliability of the observed value of the correlation coefficient, when we consider it as satisfying the conditions of the random sampling.

There are two main functions of probable error:

1. ***Determination of limits***: The limits of population correlation coefficient are $r \pm PE(r)$, implying that if we take another random sample of the size N from the same population, then the observed value of the correlation coefficient in the second sample can be expected to lie within the limits given above, with 0.5 probability.

When sample size N is small, the concept or value of PE may lead to wrong conclusions. Hence to use the concept of PE effectively, sample size N it should be fairly large.

2. ***Interpretation of 'r':*** The interpretation of $'r'$ based on PE is as under:

• If $r < PE(r)$, there is no evidence of correlation, $i.e.$ a case of insignificant correlation.

• If $r > 6\ PE(r)$, correlation is significant. $If\ r < 6\ PE(r)$, it is insignificant.

• If the probable error is small, correlation exist where $r > 0.5$

Standard Errors

Standard error disclose that how much a sample mean deviate from population mean.

1.7 RANK CORRELATION

SPEARMAN'S RANK CORRELATION

Sometimes we come across statistical series in which the variables under consideration are not capable of quantitative measurement but can be arranged in serial order. This happens when we are dealing with qualitative characteristics (attributes) such as honesty, beauty, character, morality, *etc.*, which cannot be measured quantitatively but can be arranged serially. In such situations Karl Pearson's coefficient of correlation cannot be used as such. Charles Edward Spearman, a British Psychologist, developed a formula in 1904, which consists in obtaining the correlation coefficient between the ranks of N individuals in the two attributes under study.

1.8 SIMPLE REGRESSION ANALYSIS: REGRESSION EQUATIONS AND ESTIMATION

In general sense, regression analysis means the estimation or prediction of the unknown value of one variable from the known value(s) of the other variable(s). **simple regression** – linear regression involving only two variables: a dependent variable and an independent variable. Regression analysis for studying more than two variables at a time is known as **multiple regressions.**

INDEPENDENT AND DEPENDENT VARIABLES

Simple regression involves only two variables; one variable is predicted by another variable.

The variable to be predicted is called the **dependent variable**. *The predictor* is called the**independent variable,** or *explanatory variable*

UNIT -4

INDEX NUMBERS

LESSON STRUCTURE

1.1 Introduction

1.2 Meaning and uses of index numbers

1.3 construction of index numbers

1.4 univariate and composite,
1.5 simple aggregative and average of relative

1.6 weighted aggregative and average of relative

1.7 tests of adequacy of
index numbers
1.8 Base shifting

1.9 problems in the construction of index numbers.

1.1 INTRODUCTION

"Index numbers are statistical devices designed to measure the relative changes in the level of a certain phenomenon in two or more situations". The phenomenon under consideration may be any field of quantitative measurements. It may refer to a single variable or a group of distinct but related variables. In Business and Economics, the phenomenon under consideration may be:

- the prices of a particular commodity like steel, gold, leather, *etc.* or a group of commodities like consumer goods, cereals, milk and milk products, cosmetics, *etc.*
- volume of trade, factory production, industrial or agricultural production, imports or exports, stocks and shares, sales and profits of a business house and so on.
- the national income of a country, wage structure of workers in various sectors, bank deposits, foreign exchange reserves, cost of living of persons of a particular community, class or profession and so on.

The various situations requiring comparison may refer to either

- the changes occurring over a time, or
- the difference(s) between two or more places, or
- the variations between similar categories of objects/subjects, such as persons, groups of persons, organisations *etc.* or other characteristics such as income, profession, *etc.*

1.2 MEANING AND USES OF INDEX NUMBERS

1. Index Numbers as Economic Barometers

Index numbers are indispensable tools for the management personnel of any government organisation or individual business concern and in business planning and formulation of executive decisions. The indices of prices (wholesale & retail), output (volume of trade, import and export, industrial and agricultural production) and bank deposits, foreign exchange and reserves *etc.,* throw light on the nature of, and variation in the general economic and business activity of the country. They are the indicators of business environment. A careful study of these indices gives us a fairly good appraisal of the general trade, economic development and business activity of the country. In the world of G simpson and F Kafka:

"Index numbers are today one of the most widely used statistical devices. They are used to take the pulse of the economy and they have come to be used as indicators of inflationary or deflationary tendencies."

Like barometers, which are used in Physics and Chemistry to measure atmospheric pressure, index numbers are rightly termed as "economic barometers", which measure the pressure of economic and business behaviour.

2. Index Numbers Help in Studying Trends and Tendencies

Since the index numbers study the relative change in the level of a phenomenon at different periods of time, they are especially useful for the study of the general trend for a group phenomenon in time series data. The indices of output (industrial and agricultural production), volume of trade, import and export, *etc.,* are extremely useful for studying the changes in the level of phenomenon due to the various components of a time series, *viz.* secular trend, seasonal and cyclical variations and irregular components and reflect upon the general trend of production and business activity. A a measure of average change in extensive group, the index numbers can be used to forecast future events. For instance, if a businessman is interested in establishing a new undertaking, the study of the trend of changes in the prices, wages and incomes in different industries is extremely helpful to him to frame a general idea of the comparative courses, which the future holds for different undertakings.

3. Index Numbers Help in Formulating Decisions and Policies

Index numbers of the data relating to various business and economic variables serve an important guide to the formulation of appropriate policy. *For example,* the cost of living index numbers are used by the government and, the industrial and business concerns for the regulation of dearness allowance (D.A.) or grant of bonus to the workers so as to enable them to meet the increased cost of living from time to time.The excise duty on the production or sales of a commodity is regulated according to the index numbers of the consumption of the commodity from time to time. Similarly, the indices of consumption of various commodities help in the planning of their future production. Although index numbers are now widely used to study the general economic and business conditions of the society, they are also applied with advantage by sociologists (population indices), psychologists (IQs'), health and educational authorities *etc.,* for formulating and revising their policies from time to time.

4. Price Indices Measure the Purchasing Power of Money

A traditional use of index numbers is in measuring the purchasing power of money. Since the changes in prices and purchasing power of money are inversely related, an increase in the general price index indicates that the purchasing power of money has gone down.

5. Index Numbers are Used for Deflation

Consumer price indices or cost of living index numbers are used for deflation of net national product, income value series in national accounts. The technique of obtaining real wages from the given nominal wages (as explained in use 4 above) can be used to find real income from inflated money income, real sales from nominal sales and so on by taking into account appropriate index numbers.

1.3 Methods of constructing index numbers:

A large number of formulae have been derived for constructing index numbers. They can be

1) Unweighted indices
   a) Simple aggregative method
   b) Simple average of relatives.
2) Weighted indices
   a) Weighted aggregative method
      i) Lasperey's method
      ii) Paasche's method
      iii) Fisher's ideal method
      iv) Dorbey's and Bowley's method
      v) Marshal-Edgeworth method
      vi) Kelly's method
   b) Weighted average of relatives

1.4 Univariate index: An index which is calculated from a single variable is called *univariate index.*

Composite index: An index which is calculated from group of variables is called *Composite index*

1.5 Simple indices:

i) Simple aggregative method:

This is the simplest method of constructing index numbers. When this method is used to construct a price index number the total of current year prices for the various commodities in question is divided by the total of the base year prices and the quotient is multiplied by 100.

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

Where P_0 are the base year prices

P_1 are the current year prices

P_{01} is the price index number for the current year with reference to the base year.

Problem:

Calculate the index number for 1995 taking 1991 as the base for the following data

Commodity	Unit	Prices 1991 (P_0)	Prices 1995 (P_1)
A	Kilogram	2.50	4.00
B	Dozen	5.40	7.20
C	Meter	6.00	7.00
D	Quintal	150.00	200.00
E	Liter	2.50	3.00
Total		166.40	221.20

Price index number = $P_{01} = \frac{\sum P_1}{\sum P_0} \times 100 = \frac{221.20}{166.40} \times 100 = 132.93$

∴ There is a net increase of 32.93% in 1995 as compared to 1991.

Limitations:

There are two main limitations of this method

1. The units used in the prices or quantity quotations have a great influence on the value of index.
2. No considerations are given to the relative importance of the commodities.

ii) Simple average of relatives

When this method is used to construct a price index number, first of all price relatives are obtained for the various items included in the index and then the average of these relatives is obtained using any one of the averages i.e. mean or median etc.

When A.M is used for averaging the relatives the formula for computing the index is

$$P_{01} = \frac{1}{n} \sum \left(\frac{P_1}{P_0} \times 100 \right)$$

When G.M is used for averaging the relatives the formula for computing the index is

$$P_{01} = Anti\log\left[\frac{1}{n}\sum\log\left(\frac{P_1}{P_0}\times 100\right)\right]$$

Where n is the number of commodities

and price relative = $\frac{P_1}{P_0}\times 100$

Problem:

Calculate the index number for 1995 taking 1991 as the base for the following data

Commodity	**Unit**	**Prices 1991 (P_0)**	**Prices 1995 (P_1)**	$\frac{P_1}{P_0}\times 100$
A	Kilogram	50	70	$\frac{70}{50}\times 100 = 140$
B	Dozen	40	60	150
C	Meter	80	90	112.5
D	Quintal	110	120	109.5
E	Liter	20	20	100
Total				

Price index number = $P_{01} = \frac{1}{n}\sum\left(\frac{P_1}{P_0}\times 100\right) = \frac{1}{5}\sum 612 = 122.4$

∴ There is a net increase of 22.4% in 1995 as compared to 1991.

Merits:

1. It is not affected by the units in which prices are quoted
2. It gives equal importance to all the items and extreme items don't affect the index number.
3. The index number calculated by this method satisfies the unit test.

Demerits:

1. Since it is an unweighted average the importance of all items are assumed to be the same.
2. The index constructed by this method doesn't satisfy all the criteria of an ideal index number.
3. In this method one can face difficulties to choose the average to be used.

1.6 Weighted indices:

i) Weighted aggregative method:

These indices are same as simple aggregative method. The only difference is in this method, weights are assigned to the various items included in the index.

There are various methods of assigning weights and consequently a large number of formulae for constructing weighted index number have been designed.
Some important methods are

i. **Lasperey's method:** This method is devised by Lasperey in year 1871.It is the most important of all the types of index numbers. In this method the base year quantities are taken weights. The formula for constructing Lasperey's price index number is

$$P_{01}^{La} = \frac{\sum p_1 q_0}{\sum p_0 q_0} \times 100$$

ii. **Paasche's method:** In this method the current year quantities are taken as weights and the formula is given by

$$P_{01}^{Pa} = \frac{\sum p_1 q_1}{\sum p_0 q_1} \times 100$$

iii. **Fisher's ideal method:** Fishers price index number is given by the G.M of the Lasperey's and Paasche's index numbers.

Symbolically

$$P_{01}^{F} = \sqrt{P_{01}^{La} P_{01}^{Pa}}$$

$$= \sqrt{\frac{\sum p_1 q_0}{\sum p_0 q_0} \times 100 \frac{\sum p_1 q_1}{\sum p_0 q_1} \times 100}$$

$$= \sqrt{\frac{\sum p_1 q_0}{\sum p_0 q_0} \frac{\sum p_1 q_1}{\sum p_0 q_1}} \times 100$$

iv. **Dorbey's and Bowley's method**

Dorbey's and Bowley's price index number is given by the A.M of the Lasperey's and Paasche's index numbers.

Symbolically

$$P_{01}^{DB} = \frac{P_{01}^{La} + P_{01}^{Pa}}{2}$$

Quantity index numbers:

i. Lasperey's quantity index number: Base year prices are taken as weights

$$Q_{01}^{La} = \frac{\sum q_1 p_0}{\sum q_0 p_0} \times 100$$

ii. Paasche's quantity index number : Current year prices are taken as weights

$$Q_{01}^{Pa} = \frac{\sum q_1 p_1}{\sum q_0 p_1} \times 100$$

iii. Fisher's ideal method: $Q_{01}^{F} = \sqrt{Q_{01}^{La} Q_{01}^{Pa}} = \sqrt{\frac{\sum q_1 p_0}{\sum q_0 p_0} \times \frac{\sum q_1 p_1}{\sum q_0 p_1}} \times 100$

Fisher's index number is called ideal index number. Why?

The Fisher's index number is called ideal index number due to the following characteristics.

1) It is based on the G.M which is theoretically considered as the best average of constructing index numbers.
2) It takes into account both current and base year prices as quantities.
3) It satisfies both time reversal and factor reversal test which are suggested by Fisher.
4) The upward bias of Lasperey's index number and downward bias of Paasche's index number are balanced to a great extent.

Example: Compute price index numbers for the following data by

(i) Laspeyre's method,
(ii) Paasche's method,
(iii) Fisher's ideal method,
(iv) Dorbish-Bowley's method,
(v) Marshall-Edgeworth's method.

Year	Commodity A		Commodity B		Commodity C	
	Price	Quantity	Price	Quantity	Price	Quantity
1980	4	50	3	10	2	5
1985	10	45	6	8	3	4

Base year : 1980

Price and quantity given in arbitrary units.

Calculation of Indices

Commodities	1980		1985					
	Price p_0	Quantity q_0	Price p_1	Quantity q_1	p_1q_0	p_0q_0	p_1q_1	p_0q_1
A	4	50	10	45	500	200	450	180
B	3	10	6	8	60	30	48	24
C	2	5	3	4	15	10	12	8
Total	—	—	—	—	575	240	510	212

(*i*) Laspeyre's method :

$$L_{01} = \frac{\Sigma p_1 q_0}{\Sigma p_0 q_0} \times 100 = \frac{575}{240} \times 100 = 239.58$$

(*ii*) Paasche's method :

$$P_{01} = \frac{\Sigma p_1 q_1}{\Sigma p_0 q_1} \times 100 = \frac{510}{212} \times 100 = 240.57$$

(*iii*) Fisher's ideal method :

$$F_{01} = \sqrt{L_{01} \times P_{01}} = \sqrt{239.58 \times 240.57} = 240.07.$$

(*iv*) Dorbish-Bowley's method :

$$DB_{01} = \frac{L_{01} + P_{01}}{2}$$

$$= \frac{239.58 + 240.57}{2} \doteq 239.82.$$

Comparison of Lasperey's and Paasche's index numbers:-

In Lasperey's index number base year quantities are taken as the weights and in Paasche's index the current year quantities are taken as weights.

From the practical point of view Lasperey's index is often proffered to Paasche's for the simple reason that Lasperey's index weights are the base year quantities and do not change from the year to the next. On the other hand Paasche's index weights are the current year quantities, and in most cases these weights are difficult to obtain and expensive.

Lasperey's index number is said to be have upward bias because it tends to over estimate the price rise, where as the Paasche's index number is said to have downward bias, because it tends to under estimate the price rise.

When the prices increase, there is usually a reduction in the consumption of those items whose prices have increased. Hence using base year weights in the Lasperey's index, we will be giving too much weight to the prices that have increased the most and the numerator will be too large. Due to similar considerations, Paasche's index number using given year weights under estimates the rise in price and hence has down ward bias.

If changes in prices and quantities between the reference period and the base period are moderate, both Lasperey's and Paasche's indices give nearly the same values.

Demerit of Paasche's index number:

Paasche's index number, because of its dependence on given year's weight, has distinct disadvantage that the weights are required to be revised and computed for each period, adding extra cost towards the collection of data.

What are the desiderata of good index numbers?

Irving Fisher has considered two important properties which an index number should satisfy. These are tests of reversibility.

1. Time reversal test
2. Factor reversal test

If an index number satisfies these two tests it is said to be an ideal index number.

Weighted average of relatives:

Weighted average of relatives can be calculated by taking values of the base year (p_0q_0) as the weights. The formula is given by

$$\text{When A.M is used } P_{01} = \frac{\sum PV}{\sum V}$$

$$\text{When G.M is used } P_{01} = Anti\log\frac{\sum V\log P}{\sum V}$$

Where $P = \frac{p_1}{p_0}\times 100$ and $V = p_0q_0$ i.e. base year value

Illustration 8. From the following data compute price index by supplying weighted average of price method using :

(*a*) arithmetic mean, and
(*b*) geometric mean.

Commodity	p_0 *(Rs.)*	q_0	p_1 *(Rs.)*
Sugar	3.0	20 kg.	4.0
Flour	1.5	40 kg.	1.6
Milk	1.0	10 lt.	1.5

Solution. (*a*) INDEX NUMBER USING WEIGHTED ARITHMETIC MEAN OF PRICE RELATIVES

Commodity	p_0	q_0	p_1	p_0q_0 *V*	$\frac{p_1}{p_0}\times 100$ *p*	*PV*
Sugar	Rs. 3.0	20 kg.	Rs. 4.0	60	$\frac{4}{3}\times 100$	8,000
Flour	Rs. 1.5	40 kg.	Rs. 1.6	60	$\frac{1.6}{1.5}\times 100$	6,400
Milk	Re. 1.0	10 lt.	Rs. 1.5	10	$\frac{1.5}{1.0}\times 100$	1,500
				$\sum V = 130$		$\sum PV = 15{,}900$

$$p_{01} = \frac{\sum PV}{\sum V} = \frac{15{,}900}{130} = 122.31$$

This means that there has been a 22.3 per cent increase in prices over the base level.

(*b*) INDEX NUMBER USING GEOMETRIC MEAN OF PRICE RELATIVES

Commodity	p_0	q_0	p_1	V	p	Log p	V Log p
Sugar	Rs. 3.0	**20 kg.**	Rs. 4.0	60	133.3	2.1249	127.494
Flour	Rs. 1.5	40 kg.	Rs. 1.6	60	106.7	2.0282	121.692
Milk	Re. 1.0	10 lt.	Rs. 1.5	10	150.0	2.1761	21.761
				$\Sigma V = 130$			$\Sigma V \log p$ **= 270.947**

$$p_{01} = \text{Antilog}\left[\frac{\Sigma V \log p}{\Sigma V}\right] = \text{Antilog}\left[\frac{270.947}{130}\right] = \text{Antilog } 2.084 = 120.9$$

1.7 TESTS OF ADEQUACY OF INDEX NUMBERS

We have discussed various formulae for the construction of index numbers. None of the formulae measures the price changes or quantity changes with perfection and has some bias. The problem is to choose the most appropriate formula in a given situation. As a measure of the formula error a number of mathematical tests, known as the *tests of consistency* or *tests of adequacy* of index number formulae have been suggested. In this section we will discuss these tests, which are also sometimes termed as the criteria for a good index number.

*1.***Time Reversal Test:** The time reversal test, proposed by Prof Irving Fisher requires the index number formula to possess time consistency by working both forward and backward *w. r. t.* time. In his (Fisher's) words:

"The formula for calculating an 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."

I**Factor Reversal Test:** This is the second of the two important tests of consistency proposed by Prof Irving Fisher. According to him:

"Just as our formula should permit the interchange of two times without giving inconsistent results, so it ought to permit interchanging the price and quantities without giving inconsistent results - i.e., the two results multiplied together should give the true value ratio, except for a constant of proportionality."

Circular Test: Circular test, first suggested by Westergaard, is an extension of time reversal test for more than two periods and is based on the shift ability of the base period. This requires the index to work in a circular manner and this property enables us to find the index numbers from period to period without referring back to the original base each time.

1.9, BASE SHIFTING

The need for shifting the base may arise either
(i) when the base period of a given index number series is to be made more recent, or
(ii) when two index number series with different base periods are to be compared, or
(iii) when there is need for splicing two overlapping index number series.
Whatever be the reason, the technique of shifting the base is simple:

New Base Index Number = Old Index Number of Current Year *x100/*Old Index Number of New BaseYear

Conversion of Fixed-base Index into Chain-base

To convert fixed-base index numbers into chain-base index numbers, the following procedure is adopted:

- The first year's index number is taken equal to 100
- For subsequent years, the index number is obtained by following formula:

Current Year's CBI = Current Year's FBI *x100/* PreviousYear's CBI

2.0 PROBLEMS IN THE CONSTRUCTION OF INDEX NUMBERS.

The above discussion enables us to identify some of the important problems, which may be faced in the construction of index numbers:

1. ***Choice of the Base Period:*** Choice of the base period is a critical decision because of its importance in the construction of index numbers. A base period is the reference period for describing and comparing the changes in prices or quantities in a given period. The selection of a base year or period does not pose difficult theoretical questions. To a large extent, the choice of the base year depends on the objective of the index. A major consideration should be to ensure that the base year is not an abnormal year. *For example,* a base period with very low price/quantity will unduly inflate, while the one with a very high figure will unduly depress, the entire index number series. An index number series constructed with any such period as the base may give very misleading results. It is, therefore, necessary that the base period be selected carefully.

2. ***Selection of Weights to be Used:*** It should be amply clear from the various indices discussed in the lesson that the choice of the system of weights, which may be used, is fairly large. Since any system of weights has its own merits and is capable of giving

results amenable to precise interpretations, the weights used should be decided keeping in view the purpose for which an index is constructed.
It is also worthwhile to bear in, mind that the use of any system of weights should represent the relative importance of individual commodities that enter into the construction of an index. The interpretations that are intended to be made from an index number are also important in deciding the weights. The use of a system of weights that involves heavy computational work deserves to be avoided.

3. ***Type of Average to be Used:*** What type of average should be used is a problem specific to simple average indices. Theoretically, one can use any of the several averages that we have, such as mean, median, mode, harmonic mean, and geometric mean. Besides being locational averages, median and mode are not the appropriate averages to use especially where the number of years for which an index is to be computed, is not large.
While the use of harmonic mean and geometric mean has some definite merits over mean, particularly when the data to be averaged refer to ratios, mean is generally more frequently used for convenience in computations.

4. ***Choice of Index:*** The problem of selection of an appropriate index arises because of availability of different types of indices giving different results when applied to the same data. Out of the various indices discussed, the choice should be in favour of one which is capable of giving more accurate and precise results, and which provides answer to specific questions for which an index is constructed.
While the Fisher's index may be considered ideal for its ability to satisfy the tests of adequacy, this too suffers from two important drawbacks. First, it involves too lengthy computations, and second, it is not amenable to easy interpretations as are the Laspeyre's and Paasche's indices. The use of the term ideal does not, however, mean that it is the best to use under all types of situations. Other indices are more appropriate under situations where specific answers are needed.

5. ***Selection of Commodities:*** Commodities to be included in the construction of an index should be carefully selected. Only those commodities deserve to be included in the construction of an index as would make it more representative. This, in fact, is a problem of sampling, for being related to the selection of commodities to be included in the sample.
In this context, it is important to note that the selection of commodities must not be based on random sampling. The reason being that in random sampling every commodity, including those that are not important and relevant, have equal chance of being selected, and consequently, the index may not be representative. The choice of commodities has, therefore, to be deliberate and in keeping with the relevance and importance of each individual commodity to the purpose for which the index is constructed.

6. ***Data Collection:*** Collection of data through a sample is the most important issue in the construction of index numbers. The data collected are the raw material of an

index. Data quality is the basic factor that determines the usefulness of an index. The data have to be as accurate, reliable, comparable, representative, and adequate, as possible.

The practical utility of an index also depends on how readily it can be constructed. Therefore, data should be collected from where these can be easily available. While the purpose of an index number will indicate what type of data are to be collected, it also determines the source from where the data can be available.

REFERENCES

1-Business Statistics, Thukral Dr. J K, Scholar Tech. Press

2- **Business Statistics,Gupta S. P.** S. Chand and Company Ltd.

3- Business Statistics, Jain T. R. Vk Indian Enterprises

4-www.statsinfoindia.weebly.com/uploads/7/3/9/1/.../**index_numbers**.