UNIT 9: DATA PROCESSING

UNIT STRUCTURE

9.1 Learning Objectives

9.2 Introduction

9.3 Processing of Data

9.4 Editing of Data

9.4.1 Types of editing of Data

9.5 Coding of Data

9.6 Tabulation of Data

9.7 Let Us sum Up

9.8 Further Reading

9.9 Answers to Check your progress

9.10 Model Questions

9.1 LEARNING OBJECTIVES

After going through thus unit, you will be able to:

- learn the concept of editing
- discuss the meaning coding
- outline the process of tabulation
- describe the computer processing of data

9.2 INTRODUCTION

In this unit we are going to discuss the process and analysis of data. In the earlier units we discussed about the collection of data through primary and secondary methods, also we discussed on the concept and approaches of sampling and questionnaire design. Let us now discuss some of the basic terms and concepts related to research i.e. data analysis , data processing and data estimation. In our day to day life, we come across several information through print, audio and visual media, social gatherings and discussions, but we rarely think about that how these informations are being collected, processed and analyzed. As discussed in the previous

unit, the collection of data refers to a plan for gathering data, information from field situations. All these data collection effort will be become worthless, if the collected data are not properly processed and analyzed. **Data processing** is concerned with editing, coding, classifying, tabulating and charting and diagramming research data. Let us now discuss the approach in detail in the following sections and subsections.

9.3 PROCESSING OF DATA

The processing of data is an essential approach of managing and systemizing the facts and writing of a research report. Data processing is a series of actions or steps performed on data to verify, organize, transform, integrate, and extract data in an appropriate output form for subsequent use. Methods of processing must be rigorously documented to ensure the utility and integrity of the data. Processing of data is classified in the following five steps:

1. Editing of Data
2. Coding of Data
3. Classification of Data
4. Tabulation of Data
5. Diagrammatic representation of Data

Let us discuss these steps briefly:

1. **Editing of Data:** Editing is the first step in data processing. Editing is the process of examining the data collected in questionnaires/ schedules to detect errors and omissions and to see that they are corrected and the schedules are ready for tabulation. When the whole data collection is over a final and a thorough check up is made. (will be discussed in the following section 9.4)
2. **Coding of Data:** Coding is necessary for efficient analysis and through it the several replies may be reduced to a small number of classes which contain the critical information required for analysis. Coding decisions should usually be taken at the designing stage of the questionnaire. (will be discussed in the following section 9.5)

3. **Classification Data:** Classification or categorization is the process of grouping the statistical data under various understandable homogeneous groups for the purpose of convenient interpretation. A uniformity of attributes is the basic criterion for classification; and the grouping of data is made according to similarity. Classification becomes necessary when there is a diversity in the data collected for meaningless for meaningful presentation and analysis.
4. **Tabulation of Data:** Tabulation is the process of summarizing raw data and displaying it in compact form for further analysis. Therefore, preparing tables is a very important step. Tabulation may be by hand, mechanical, or electronic. The choice is made largely on the basis of the size and type of study, alternative costs, time pressures, and the availability of computers, and computer programmes.
5. **Presentation of data:** The presentation of data could be tabular, statistical and cartographic. In case of tabular form of presentation, data related to different variables should be classified and compared. Various statistical techniques are available to derive accurate and precise results. Since techniques have a large range coupled with the limitations of their own, selection of appropriate technique needs to be made for the purpose. The construction of graphs, charts, diagrams and maps are the various forms of cartographic presentations. The data is transformed into cartographic system which is used for visual presentation. A brief account of tabular, statistical as well as cartographic presentation of data is discussed below.

   i. Tabular Presentation: It is used for summarization of data in its micro form. It helps in the analysis of trends, relationship and other characteristics of a given data. Simple tabulation is used to answer question related to one characteristic of the data whereas complex tabulation is used to present several interrelated characteristics. Complex tabulation results in two way, three way tables which give information about two or three inter-related characteristic of data. The following points may be kept in mind while constructing a table.

ii. *Statistical Presentation of data:* The data collected through various sources needs to be processed statistically for precise explanations. Very often it becomes necessary to obtain a single representative value for the whole data set. The statistical measures that enable us to work out a single representative figure for the entire data distribution, is known as central tendency. Measures of central tendency help us to compare different distributions besides being representative for each distribution. These measures normally denote the central points of values, distance and occurrence in a distribution. The commonly used measures of central tendency are:
   a. Arithmetic mean or average
   b. Median
   c. Mode

iii. *Cartographic Presentation of Data:* The primary data collected through the field survey may be presented cartographically. The representation of data in visual form refers either to time, space or to both. The cartographic presentation refers to the display of data by constructing graphs, diagrams and maps. The set of data is transformed into some form of figure which is used for illustrations. These figures could be graphic, geometric or theme specific maps.

9.4 EDITING OF DATA

Editing of data is the first task of data processing. It is the process where the data are prepared for ensuing coding. When the researcher collects the data it is in a raw form and it needs to be edited, organized and analyzed. The raw data needs to be transformed into a comprehensible form of data. Editing of data is the process of examining errors and omissions in the collected data and making necessary correction in the same. Editing of data is very important when there is some inconsistency in the responses entered in the questionnaire or any half done or partial information in the questionnaire. The editing of data can be done at two stages: field and post-field editing. The field editing is a review of reporting by the researcher for completing what has been written in an abbreviated

form during interviewing the respondent. The post-field editing is carried out when field survey is completed and all the forms of schedule have been collected together.

A researcher should take care of the following pints while collecting the data:

1. Accuracy of the table
2. Consistent with other facts secured,
3. Unimormity of data
4. Completeness of the data,
5. Acceptable for tabulation and arranged to facilitate coding tabulation.

9.4.1 Types of Editing of Data

Let us now discuss the different types of editing of data:

1. **Editing for quality:** Quality is very important aspect while formulating and collecting data. To ensure the quality of the data the following should be taken care of:
   i. the data forms should be complete,
   ii. the data should be free from bias,
   iii. the recordings are free from errors,
   iv. the should be inconsistencies in responses within limits,
   v. there should be evidences to show dishonesty of enumerators or interviewers and
   vi. there should not be any willful manipulation of data.
2. **Editing for tabulation:** Editing for tabulation is done for certain accepted modification to data or even rejecting certain pieces of data in order to facilitate tabulation
3. **Field Editing:** The questionnaire filled up by the researcher or the respondent might have some abbreviated writings, illegible writings and the like. These are rectified by the researcher. This should be done soon after the researcher or interview before the loss of memory. The field editing should not extend to giving some guess data to fill up omissions.

4. **Central Editing:** It is done by the researcher after getting all schedules or questionnaires or forms from the enumerators or respondents. Obvious errors can be corrected. For missed data or information, the editor may substitute data or information by reviewing information provided by likely placed other respondents. A definite inappropriate answer is removed and "no answer" is entered when reasonable attempts to get the appropriate answer fail to produce results.



CHECK YOUR PROGRESS

Q1: What is Processing of Data

.....

.....

Q.2: What is editing of Data?

.....

.....

9.5 CODING OF DATA

Coding is the procedure of classifying the answers to questions into meaningful categories. Coding is necessary for efficient analysis and through coding many responses may be reduced to a small number of classes which contain the critical information required for analysis.

Coding is basically done at the designing phase of the questionnaire. This makes it possible to pre-code the questionnaire choices and which in turn is helpful for computer tabulation.Coding is the process by which data are organized into classes or categories and numerals or other symbols are given to each item according to the class in which it falls. So, coding involves two steps:

1. To specify the different categories or classes into which the responses are to be classified
2. To allocate respondents responses to different categories.

While allocating the coding to the categories, it should be taken care of that the categories must be 'all inclusive' and 'mutually exclusive'.

The all inclusive aspects can be taken care by adding one or more categories as “other”, ‘no information’ and ‘none’. The other aspect is that categories must be ‘mutually exclusive’ i.e. they must note be overlapping and ambigious.

To keep the response with in limited alternatives, the researcher need to assign some alphabetical or numerical symbols or both to the answers. For example in a question of educational qualifications alternative choices given are: Uneducated; Below Matriculation; Matriculation & above but below Graduate; Graduate & above; Technical Diploma; Technical Degree The alphabetical codes assigned to these alternatives could be A,B,C,D,E and F. Similarly, numerical codes to these alternatives could be 1,2,3,4.,and 5 respectively. It is necessary for the efficient analysis.

Let us now discuss the process of coding:

- Firstly, the study of the responses is very important. As discussed, coding begins at the preparation of interview schedules.
- Secondly, coding frame is developed by listing the possible answers to each question and assigning code numbers or symbols to each of them which are the indicators used for coding.
- Thirdly, after preparing the sample frame, the gradual process of fitting the answers to the questions must be begun.
- Lastly, transferring of the information from the schedules to a separate sheet called transcription sheet. Transcription sheet is a large summary sheet which contain the answer/codes of all the respondents. Transcription may not be necessary when only simple tables are required and the number of respondents are few.
- Coding is necessary to carry out the subsequent operations of tabulating and analyzing data.

9.6 TABULATION OF DATA

After the collection of data is completed, it is prepared for analysis. As the data is raw, it needs to be transformed in such a way, that it is appropriate for analysis. Tabulation comprises of sorting of data into different categories and counting the numbers of cases that belong to each category.

Tabulation is a process of summarizing data and presenting it in a compact form, by putting data into statistical table.

In other words, it is a systematic arrangement of data in columns and rows that represents data in concise and attractive way. One should follow the given guidelines for tabulation.

The simplest way to tabulate is to count the number of responses to one question, which is called univariate tabulation. Where two or more variables are involved in tabulation , it is called bivariate or multivariate tabulation.

Tabulation can be done by hand or by mechanical or computer aided. The choice depends on the size and type of study, cost considerations, time pressures and the availability of tabulating machines or computers. In relatively large inquiries, we may use mechanical or computer tabulation.

Let us now discuss the different types of tabulation:

There are generally two types of tables simple and complex. They are discussed following:

Simple Tabulation: Simple tabulation, gives information regarding one or more independent questions.Under it , the different attribute are stated in the left hand column and the frequency or extend of occurrence of each of these class are written in another column. In this three things are essential a) the classes made must be mutually exclusive, b) the tabulation must have internal logic and order, and c) the class intervals must carefully and reasonably selected. Following is an illustration of the same:

No of visit to joggers park per month	No. of Respondents	% of Respondents
Below 5 6-10	2 15	7 50
12-26	13	43
	30	100

Complex Tabulation or Cross Tabulation: Complex tabulation gives information regarding two mutually dependent questions. It may often be necessary to tabulate responses to two or more questions simultaneously. Such tabulations are known as bivariate or multivariate tabulations,

depending upon whether two or more than two variables are involved. In constructing cross tabulation tables, one has to first determine which data should be given primary emphasis and which should be given secondary emphasis. Data with primary emphasis are normally given in columns while those with secondary emphasis are shown in raws.

The following table is an example of cross tabulation.

Income Level	Shopping Centres		Total nos. of Households
	A	B	
Upto to Rs. 100	100	200	300
Rs. 1000+	120	80	200
Total	220	280	500

In the above table there are three variants i.e income level, shopping centres and households are being studied and tabulated.

Let us now discuss the steps are that are necessary in the **preparation of table:**

1. **Title of table:** The table should be first given a brief, simple and clear title which may express the basis of classification.
2. **Columns and rows:** Each table should be prepared in just adequate number of columns and rows.
3. **Captions and stubs:** The columns and rows should be given simple and clear captions and stubs.
4. **Ruling:** Columns and rows should be divided by means of thin or thick rulings.
5. **Arrangement of items;** Comparable figures should be arranged side by side.
6. **Deviations:** These should be arranged in the column near the original data so that their presence may easily be noted.
7. **Size of columns:** This should be according to the requirement.
8. **Arrangements of items:** This should be according to the problem.
9. **Special emphasis:** This can be done by writing important data in bold or special letters.
10. **Unit of measurement:** The unit should be noted below the lines.
11. **Approximation:** This should also be noted below the title.

12. Foot – notes: These may be given below the table.

13. Total: Totals of each column and grand total should be in one line.

14. **Source** : Source of data must be given. For primary data, write primary data.



CHECK YOUR PRIGRESS

Q3: What is Coding of Data?

.....

.....

Q4: What is Tabulation of Data?

.....

.....



9.7 LET US SUM UP

In this unit we discussed the following:

- Data processing is a series of actions or steps performed on data to verify, organize, transform, integrate, and extract data in an appropriate output form for subsequent use.
- Processing of data is classified in the following five steps:
  a. Editing of Data
  b. Coding of Data
  c. Classification of Data
  d. Tabulation of Data
  e. Diagrammatic representation of Data
- Editing of data is the first task of data processing. It is the process where the data are prepared for ensuing coding.
- The editing of data can be done at two stages: field and post-field editing.
- The different types of editing of data are :Editing for quality, Editing for tabulation, Field Editing and Central Editing

- Coding is the procedure of classifying the answers to questions into meaningful categories.
- Coding is the process by which data are organized into classes or categories and numerals or other symbols are given to each item according to the class in which it falls. So, coding involves two steps:
  a. To specify the different categories or classes into which the responses are to be classified
  b. To allocate respondents responses to different categories.
- Tabulation comprises of sorting of data into different categories and counting the numbers of cases that belong to each category. Tabulation is a process of summarizing data and presenting it in a compact form, by putting data into statistical table.



9.8 FURTHER READING

1. Kothari, C.R. and Garg, G.(2014), 'Research Methodology - Methods and Techniques', New Age International Publishers
2. Malhotra. N.K.(2011), Marketing Research – An Applied Orientation, Pearson



9.9 ANSWERS TO CHECK YOUR PROGRESS

Ans to Q1: Data processing is a series of actions or steps performed on data to verify, organize, transform, integrate, and extract data in an appropriate output form for subsequent use.

Ans to Q2: Editing of data is the first task of data processing. It is the process where the data are prepared for ensuing coding.

Ans to Q3: Coding is the procedure of classifying the answers to questions into meaningful categories.

Ans to Q4: Tabulation is a process of summarizing data and presenting it in a compact form, by putting data into statistical table.



9.10 MODEL QUESTIONS

Q.1: What is Data Processing?

Q.2: Write the classification of processing of data

Q.3: What is presentation of Data?

Q.4: Why editing of Data is important. Discuss.

Q.5: Discuss the different types of editing of data

Q.6: Explain Coding of Data

Q.7: What is tabulation of data? Write the types of tabulation data.

****

UNIT 10: UNIVARIATE AND BIVARIATE ANALYSIS OF DATA

UNIT STRUCTURE

10.1 Learning Objectives

10.2 Introduction

10.3 Descriptive vs Inferential Analysis

10.4 Descriptive Analysis of Univariate data

10.5 Measures of Central Tendency

10.6 Measures of Dispersion;

10.7 Graphical Analysis of Nominal scale data

10.8 Descriptive Analysis of Bivariate data

10.9 Analysis of Nominal scale data with multiple category responses

10.10 Let Us Sum Up

10.11 Further Reading

10.12 Answers to Check Your Progress

10.13 Model Questions

10.1 LEARNING OBJECTIVES

After going through this unit, you will be able to

- difference between descriptive and inferential Analysis
- analysis of univariate data
- different measures of central tendency
- different measures of dispersion
- graphical analysis of data
- analysis of bivariate data
- analysis of nominal scale data

10.2 INTRODUCTION

In the previous unit we disccussed the concepts of data and coding of data, in this unit we will discuss the analysis of data. Data analysis is the process of systematically applying statistical and/or logical techniques to describe and evaluate data. Data from various sources is gathered,

reviewed, and then analyzed to form some sort of finding or conclusion. Different data analysis techniques can be used based on the objective of the research or on the type of data gathered, as for example, univariate or bivariate or multivariate. Again, analysis of two kinds, viz. Descriptive and inferential analysis.Let us discuss these concepts in the following sections:

10.3 DESCRIPTIVE VS INFERENTIAL ANALYSIS

Descriptive analysis quantitatively described the features of data sample, while inferential analysis made inferences about the populations from which samples were drawn.

Descriptive statistics are used to describe the basic features of the data in a study. They provide simple summaries about the sample and the measures. Together with the simple graphics analysis, they form the basis of virtually every quantitative analysis of data.

By inferential statistics we make judgements of the probability that an observed difference between groups (samples) is a dependable one or one that might have happened by chance in this study.

10.4 DESCRIPTIVE ANALYSIS OF UNIVARIATE DATA

When one measurement is made on each observation, univariate analysis is applied. Univariate is the simplest form of analyzing data. “Uni” means “one”, so in other words univariate data means data with one variable. The purpose of univariate data is to describe the pattern of the variable. For example, the age of 20 students, the height of 30 children, etc.

Some ways to describe the patterns found in univariate data include central tendency (mean, median and mode); range, variance, quartiles etc. Some other ways to describe the data through chart or graph are: bar charts, pie charts, histograms, frequency polygons, etc. There are three major characteristics of a single variable that we tend to look at:

(i) The distribution

(ii) The central tendency

(iii) The dispersion

Distribution: The distribution is a summary of the frequency of individual values or ranges of values for a variable. The simplest distribution would list every value of a variable and the number of persons who had each value. For instance, the distribution of college students year wise, or if we describe gender by listing the number or percent of males and females. In these cases, the variable has few enough values that we can list each one and summarize how many sample cases had the value. But if we have large number of possible values, we group the raw scores into categories according to ranges of values. This is known as frequency distribution.

One of the most common ways to describe a single variable is with a ***frequency distribution***. Depending on the particular variable, all of the data values may be represented, or may be grouped the values into categories first. Frequency distributions can be depicted in two ways, as a table or as a graph.

Distributions may also be displayed using percentages. For example, we could use percentages to describe the:

- percentage of people in different income levels
- percentage of people in different age ranges

10.5 MEASURES OF CENTRAL TENDENCY

Central Tendency: The central tendency of a distribution is an estimate of the "centre" of a distribution of values. There are three major types of estimates of central tendency:

- Mean
- Median
- Mode

Mean or average is probably the most commonly used method of describing central tendency. To compute the mean we need to add up all the values and divide by the number of values. For example, the mean or average quiz score is determined by summing all the scores and dividing by the number of students taking the exam. For example, consider the test score values:

15, 20, 21, 20, 36, 15, 25, 15

The sum of these 8 values is 167, so the mean is 167/8 = 20.875.

Median is the score found at the exact middle of the set of values. One way to compute the median is to list all scores in numerical order, and then locate the score in the centre of the sample. For example, if there are 500 scores in the list, score numbered 250 would be the median. If we order the 8 scores shown above, we would get:

15, 15,15,20,20,21,25,36

There are 8 scores and score number 4 and number 5 represent the halfway point. Since both of these scores are 20, the median is 20. If the two middle scores had different values, then the average between the two scores may be regarded as median.

Mode is the most frequently occurring value in the set of scores. To determine the mode, we might again order the scores as shown above, and then count each one. The most frequently occurring value is the mode. In our example, the value 15 occurs three times and is the model. In some distributions there is more than one modal value. For instance, in a bimodal distribution there are two values that occur most frequently.

Notice that for the same set of 8 scores we got three different values — 20.875, 20, and 15 — for the mean, median and mode respectively. If the distribution is truly normal (i.e., bell-shaped), the mean, median and mode are all equal to each other.



CHECK YOUR PROGRESS

Q1: Define Central Tendency

.....

.....

Q2: Fill in the blanks

a. To compute the mean we need to add up all the values and divide by the .....

b. Median is the score found at the exact ……. of the set of values.

10.6 MEASURES OF DISPERSION

Dispersion: Dispersion refers to the spread of the values around the central tendency. There are two common measures of dispersion, the range and the standard deviation. The **range** is simply the highest value minus the lowest value. In our example distribution, the high value is 36 and the low is 15, so the range is 36 - 15 = 21.

Standard Deviation is a more accurate and detailed estimate of dispersion because an outlier can greatly exaggerate the range (as was true in this example where the single outlier value of 36 stands apart from the rest of the values. The Standard Deviation shows the relation that set of scores has to the mean of the sample. It is denoted by ó and is given by the formula:

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

Again let us take the set of scores: **15, 15,15,20,20,21,25,36**

Calculation of Standard Deviation

Sl. No.	x	$(x - \bar{x})^2$
1	15	34.516
2	15	34.516
3	15	34.516
4	20	0.766
5	20	0.766
6	21	0.016
7	25	17.016
8	36	228.766
Total	167	350.875

$\bar{x}$ = 20.875

ó = 6.622

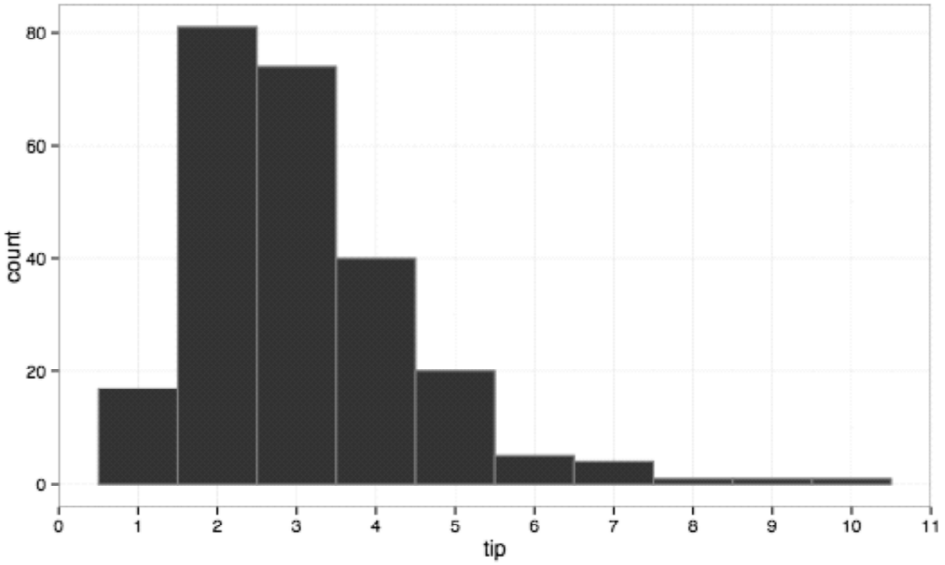
10.7 GRAPHICAL ANALYSIS OF NOMINAL SCALE DATA

Under graphical analysis following charts are used: bar charts, pie charts, histograms, frequency polygons:

Bar Charts: A bar chart or bar graph is a chart that presents grouped data with rectangular bars with lengths proportional to the values that they represent. The bars can be plotted vertically or horizontally. A vertical bar chart is sometimes called a Line graph if bars are replaced by lines.



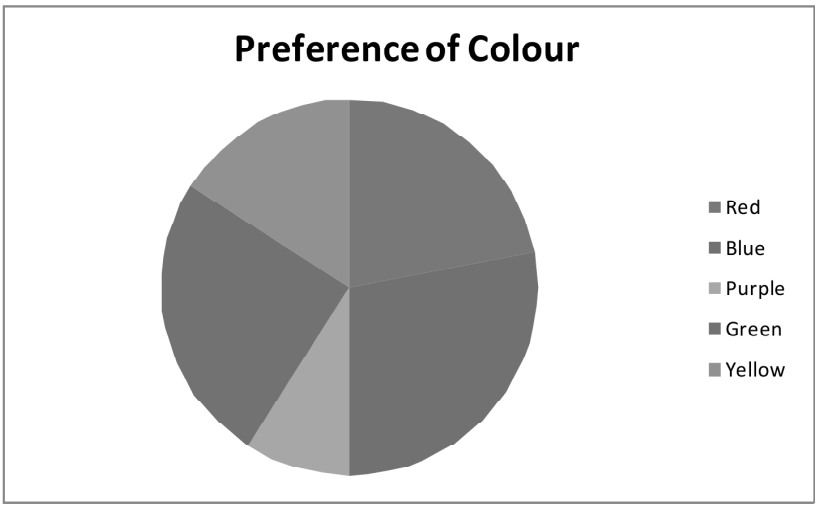
Histogram: Histogram is a graphical display of data using bars of different heights. It is drawn by placing class intervals on the x–axis and corresponding frequencies on y–axis. Thus we get a series of rectangles one attached to other. Histograms are commonly used in statistics to demonstrate how many of a certain type of variable occurs within a specific range. For example, a census focused on the demography of a country may use a histogram of how many people there are between the ages of 0 and 10, 11 and 20, 21 and 30, 31 and 40, 41 and 50 etc.



Pie Charts: Pie charts or circle graphs as they are sometimes known are different from the other types of graphs. This graph does not use a set of axes to plot points. Pie charts display percentages. A Pie Chart (or Pie Graph) is a circular chart divided into sectors; each sector shows the relative size of each value.

Example:

Preference of Favourite Colour		
Colour	**Frequency**	**Percentage**
Red	22	79.2
Blue	28	100.8
Purple	9	32.4
Green	25	90
Yellow	16	57.6





CHECK YOUR PROGRESS

Q3: Define Histogram.

.....

.....

Q4: Compute standard deviation for the following set of data:

25, 36, 46, 58, 39, 43, 60, 40, 55, 45

10.8 DESCRIPTIVE ANALYSIS OF BIVARIATE DATA

Bivariate data is that where exactly two measurements are made on each observation. The two measurements will be called X and Y. Since X and Y are obtained for each observation, the data for one observation is the air (X, Y). For example height and weight of 20 college students, price and demand of a good for 10 weeks, Stock market valuation (X) and quarterly corporate earnings (Y) are recorded for each company in a sample etc.

Bivariate analysis is one of the simplest forms of quantitative (statistical) analysis. It involves the analysis of two variables for the purpose of determining the empirical relationship between them. There is a difference between bivariate data and two sample data. In two sample data, the X and Y values are not paired, and there aren't necessarily the same number of X and Y values.

Two-sample data:

Sample 1: 3, 2, 5, 1, 3, 4, 2 ,3

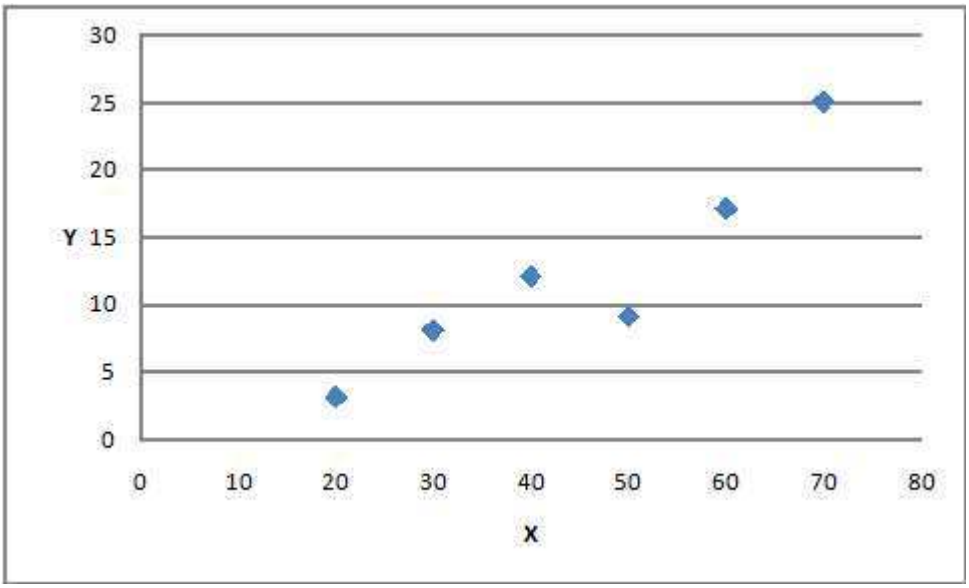
Sample 2: 4, 4, 3, 6, 5

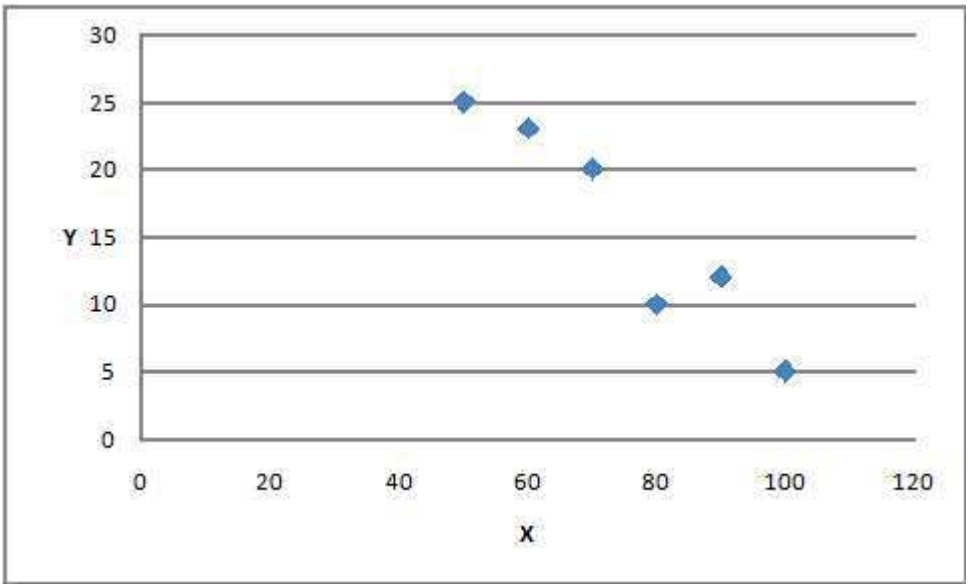
Bivariate data can be written $(X_1, Y_1), (X_2, Y_2), \ldots, (Xn, Yn)$. Each observation is a pair of values, for example (X_3, Y_3) is the third observation. In a bivariate data, the observations are independent of each other, but the two measurements within an observation may not be independent.

Types of Bivariate Analysis

Common types of bivariate analysis include:

1. **Scatter Plot:** The most important graphical summary of bivariate data is the scatter plot. This is simply a plot of the points (X_i, Yi) in the plane. A simple scatter plot gives an visual idea of the pattern of the variables of interest. It reflects the association between the variables.





2. Correlation Coefficient: A measure that determines the degree to which two variable's movements are associated. Correlation is a term that refers to the strength of a relationship between two variables. A strong, or high, correlation means that two or more variables have a strong relationship with each other, while a weak or low correlation means that the variables are hardly related. Correlation analysis is the process of studying the strength of that relationship with available statistical data.

The most widely used type of correlation coefficient is the Pearson r. Correlation coefficients can range from -1.00 to +1.00. The value of -1.00

represents a perfect negative correlation, which means that as the value of one variable increases, the other decreases. While a value of +1.00 represents a perfect positive relationship, meaning that as one variable increases in value, so does the other. Values like these—of + or - 1.00—signal a perfectly linear relationship between the two variables, so that if you plot the results on a graph it would make a straight line. A value of 0.00 means that there is no relationship between the variables being tested.

Example: Calculate and analyze the correlation coefficient between the number of study hours and the number of sleeping hours of different students.

Number of Study hours	**2**	**4**	**6**	**8**	**10**
Number of Sleeping hours	**10**	**9**	**8**	**7**	**6**

Solution: Pearson's correlation coefficient is given by

$$r = \frac{\sum (x - \bar{x})(y - \bar{y})}{\sqrt{\sum (x - \bar{x})^2}\sqrt{\sum (y - \bar{y})^2}}$$

Calculation:

x	y	$(x-\bar{x})$	$(y-\bar{y})$	$(x-\bar{x})^2$	$(y-\bar{y})^2$	$(x-\bar{x})(y-\bar{y})$
2	10	-4	2	16	4	-8
4	9	-2	1	4	1	-2
6	8	0	0	0	0	0
8	7	2	-1	4	1	-2
10	6	4	-2	16	4	-8
$\sum x$ =30	$\sum y$ =40			$\sum (x-x)^2$ =40	$\sum (y-\bar{y})^2$ −10	$\sum (x-\bar{x})(y-\bar{y})$ =-20

$$\therefore r = \frac{-20}{\sqrt{10X40}} = -1$$

3. **Regression Analysis:** Scatter plot is giving the pattern of relationship between bivariate data whereas regression analysis is a statistical process for estimating the relationships among variables. Its main focus is on the relationship between a dependent variable and one or more independent variables (or 'predictors').

How to Find the Regression Equation?

In the table below, the x_i column shows scores on the aptitude test. Similarly, the y_i column shows statistics grades. The last two rows show sums and mean scores that we will use to conduct the regression analysis.

	Student	x_i	y_i	$(x_i - x)$	$(y_i - y)$	$(x_i - x)^2$	$(y_i - y)^2$	$(x_i - x)(y_i - y)$
	1	95	85	17	8	289	64	136
	2	85	95	7	18	49	324	126
	3	80	70	2	-7	4	49	-14
	4	70	65	-8	-12	64	144	96
	5	60	70	-18	-7	324	49	126
Sum		390	385			730	630	470
Mean		78	77					

Let us consider the regression equation be a linear equation of the form:

$$\hat{y} = b_0 + b_1 x$$

To conduct a regression analysis, we need to solve for b_0 and b_1. Computations are shown below.

$b_1 = \Sigma [(x_i - x)(y_i - y)] / \Sigma [(x_i - x)^2]$ $b_1 = 470/730 = 0.644$		$b_0 = y - b_1 * x$ $b_0 = 77 - (0.644)(78) =$ 26.768

Therefore, the regression equation is: $\hat{y}$ = 26.768 + 0.644x.



CHECK YOUR PROGRESS

Q5: What are the different ways of analysing bivariate data?

.....

.....

Q6: Plot a scatter diagram using following data:

X: 29	35	10	37	45	39	57	68	75
Y: 44	53	18	42	55	60	82	85	91

10.9 ANALYSIS OF NOMINAL SCALE DATA WITH MULTIPLE CATEGORY RESPONSES

Multiple response data routinely arise in the study of consumer behaviour. The use of specific products, brands, and distribution channels (e.g., the Internet); magazine readership; product consumption contexts; activities engaged in outside the marketplace (e.g., hobbies); and

perceptions about brand attributes are examples of behavioural constructs that are commonly studied by means of multiple response questions. Respondents are directed to consider the response options within a common universe (e.g., brands within a product class, contexts of use for a specific product) that focuses cognitive effort to a particular domain. The use of a nominal scale to capture discrete behaviour (e.g., use, do not use) and perceptions (e.g., provides benefit, does not provide benefit) further reduces the response effort, there by resulting in a question format that is easy for respondents to use. Surveys containing multiple response questions provide a useful format for collecting behavioural and perceptual data across many response options.

Multiple response data are often collected as part of a descriptive research study in which the goal is to move inductively from data to theory, not from theory to data. When a well-developed theory is available to guide a research effort, richer response scales are often developed, and the domain of the inquiry is focused on specific hypotheses. In contrast, descriptive research typically employs less developed scales and attempts to be more encompassing in its domain. A goal of analyzing multiple response data in descriptive research is to uncover structure in the binary responses that can ultimately suggest new theory. Many techniques are available to marketing researchers for the exploratory analysis of data, particularly for continuous data from elliptically symmetric distributions (e.g., the multivariate normal). Principal component, canonical correlation, and cluster analysis are examples of multivariate data–reduction techniques that can provide insights into the structure underlying high-dimensional data.



10.10 LET US SUM UP

After going through the unit we have come through the following:

- Descriptive analysis quantitatively described the features of data sample, while inferential analysis made inferences about the populations from which samples were drawn.

- Data may be of the type univariate, bivariate or multivariate.
- Some ways to describe the patterns found in univariate data include central tendency (mean, median and mode) and dispersion (range, variance, quartiles etc.) Some other ways to describe the data through chart or graph are: bar charts, pie charts, histograms, frequency polygons, etc.
- Bivariate data is that where exactly two measurements are made on each observation.
- Scatter plot, Correlation and regression analysis are some common form of analysis of bivariate data.
- Multiple response data are often collected as part of a descriptive research study in which the goal is to move inductively from data to theory, not from theory to data.



10.11 FURTHER READING

1. Malhotra. N.K.(2011), *Marketing Research -An Applied Orientation*, Pearson Education



10.12 ANSWERS TO CHECK YOUR PROGRESS

Ans to Q.1: **Central Tendency:** The central tendency of a distribution is an estimate of the “centre” of a distribution of values.

Ans to Q.2: **(a)** To compute the mean we need to add up all the values and divide by the number of values.

(b) Median is the score found at the exact middle of the set of values.

Ans to Q.3: Histogram is a graphical display of data using bars of different heights. It is drawn by placing class intervals on the x–axis and corresponding frequencies on y–axis. Thus we get a series of rectangles one attached to other.

Ans to Q.4: Calculation of Standard Deviation

Sl. No.	x	$(x-\bar{x})^2$
1	25	388.09
2	46	1.69
3	39	32.49
4	60	234.09
5	55	106.09
6	36	75.69
7	58	176.89
8	43	2.89
9	40	22.09
10	45	0.09
Total	447	1040.1

$\bar{x} = 44.7$

$\sigma = \sqrt{(\frac{1}{10} X 1040.1)}$ = 10.20

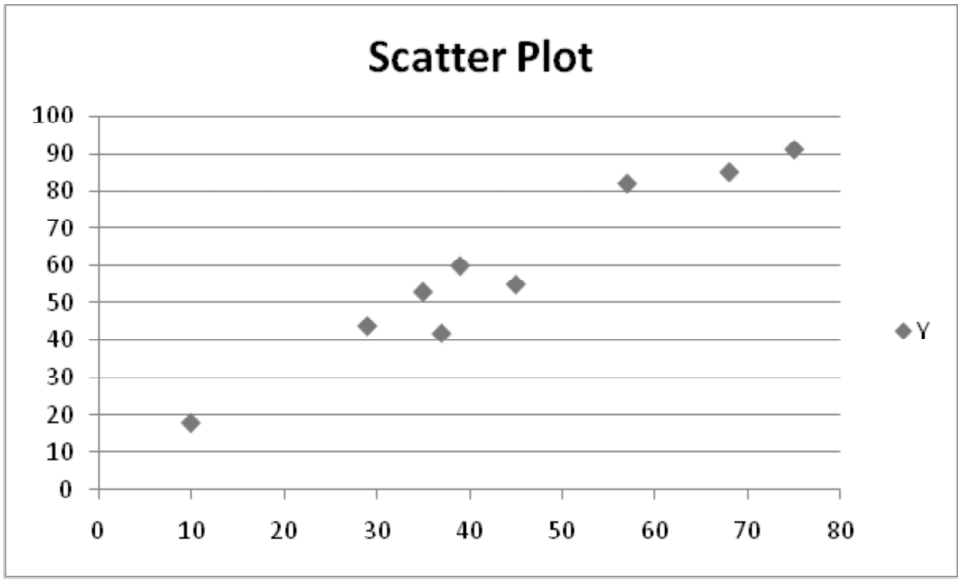
Ans to Q.5: Bivariate analysis is one of the simplest forms of quantitative (statistical) analysis. It involves the analysis of two variables for the purpose of determining the empirical relationship between them. There are different methods for analysing bivariate data, viz.

(a) Scatter Plot

(b) Correlation and

(c) Regression Analysis

Ans to Q.6:





10.13 MODEL QUESTIONS

Q.1: What are the measures of central tendency?

Q.2: What is the difference between descriptive analysis and inferential analysis?

Q.3: Explain the requirement of a measure of dispersion?

Q.4: What is regression and what are its importance in the analysis of bivariate data?

Q.5: Write short notes on correlation.

*****

UNIT 11: TESTING OF HYPOTHESES

UNIT STRUCTURE

11.1 Learning Objectives

11.2 Introduction

11.3 Concepts in Testing of Hypothesis

11.4 Steps in testing of hypothesis

11.5 Applications of Hypothesis Testing

11.6 Test of Single Mean (Large Sample)

11.7 Test of Single Mean (Small Sample)

11.8 Test of Difference between Means (Large Sample)

11.9 Test of Difference between Means (Small Sample)

11.10 Test of Single Population (Large Sample)

11.11 Tests for Difference between Two Population Proportions

11.12 Let Us Sum Up

11.13 Answers to Check Your Progress

11.14 Further Reading

11.15 Model Questions

11.1 LEARNING OBJECTIVES

After going through this unit, you will be able to

- learn the concept of hypothesis
- discuss the basic concepts concerning testing of hypothesis
- discuss the test statistic and critical region
- outline the procedure for hypothesis testing
- describe the different test statistic

11.2 INTRODUCTION

Hypothesis is usually considered as the principal instrument in research. Its main function is to suggest new experiments and observations. In fact, many experiments are carried out with the deliberate object of testing hypotheses. Decision-makers often face situations wherein they are

interested in testing hypotheses on the basis of available information and then take decisions on the basis of such testing.

11.3 CONCEPTS IN TESTING OF HYPOTHESIS

Most people including researchers make rough confidence statements on many aspects of their areas of interest, be it business research, social science or industrial engineering, etc. These statements or hypotheses are often about particular facets of their work areas. These hypotheses need verification for possible acceptance or rejection. To this end, researcher gather data and allow them to support or cast doubt on their hypotheses. This process is termed hypothesis testing and is one of two main areas of statistical inference.

Hypothesis testing is a statistical technique that is used in a variety of situations. Though the technical details differ from situation to situation, all hypothesis tests use the same core set of terms and concepts.

Null Hypothesis: The null hypothesis is a clear statement about the relationship between two (or more) statistical objects. These objects may be measurements, distributions, or categories. Typically, the null hypothesis, as the name implies, generally expresses the idea of no difference and usually given the symbol H_0.

Alternative Hypothesis: The alternative hypothesis, which we denote by H1 , expresses the idea of some difference. Alternative hypotheses may be one-sided or two-sided. Usually the setting of the problem determines the alternative even before the data has been collected.

For example suppose we are interested in testing whether there is any statistical evidence, from a random sample of potential customers, to support the hypothesis that more than 10% of the potential customers will purchase a new product. In this case the null and alternative hypotheses are

H_0 : P = 0.01 (Null Hypothesis)

H_1 : P ‘“ 0.01 (Alternate Hypothesis)

Note: Alternative hypothesis determines the tail of a test.

For example: If we consider the null hypothesis as

H_0 : ì = 100 then the alternative hypothesis may be

H_1 : ì “ 100 (Two tail test) or

H_1 : ì > 100 (Right tail test) or

H_1 : ì < 100 (Left tail test)

Errors in Testing of Hypothesis: All hypothesis testing is liable to errors. There are two basic kinds of error:

- Type I error: Reject H_0 when it is, in fact, true; the probability of committing a type I error is denoted by α.
- Type II error: Accept H_0 when it is, in fact, false; the probability of committing a type II error is denoted by β.

The objective in all hypothesis testing is to set the Type I error level, also known as the significance level, at a low enough value, and then to use a test statistic which minimises the Type II error level for a given sample size. As we fix the Type I error level, it is best to devise the test in such a way that the Type I error is most serious, in terms of cost.

Table: 11.1 possible Solutions in Hypothesis Testing

Actual	Decision	
	Accept H_0	Reject H_0
H_0 is true	Correct Decision	Wrong Decision (Type I Error)
H_0 is false	Wrong Decision (Type II Error)	Correct Decision

Level of Significance: The level of significance is a measure of the statistical strength of the hypothesis test. It is often characterized as the probability of incorrectly concluding that the null hypothesis is false. The significance level is something that you should specify up front. In applications, the significance level is typically one of three values: 10%, 5%, or 1%. A 1% significance level represents the strongest test of the three. For this reason, 1% is a higher significance level than 10%.

Power: Related to significance, the power of a test measures the probability of correctly concluding that the null hypothesis is true.

Test Statistic: After setting up the null hypothesis and alternative hypothesis we compute the test statistic. The test statistic is a statistic based on appropriate probability distribution. It is used to test whether the null hypothesis set up should be accepted or rejected. It is the main yardstick which helps us to decide whether to accept or reject the null hypothesis.

$$\text{TestStatistic} = \frac{\text{Sample Statistic - Hypothesised Parameter}}{\text{Standard Error of Statistic}}$$

Different probability distribution values are used in appropriate cases while testing the null hypothesis. The following are the most commonly used test statistic.

Z-Distribution: We use Z-Distribution under normal curve for large sample (i.e. if the sample size n $\geq$ 30). The Z-statistic is defined as

$$Z = \frac{t - E(T)}{S.E.(t)} \sim N(0,1) \text{ as } n \to \infty$$

t being the statistic and T the parameter.

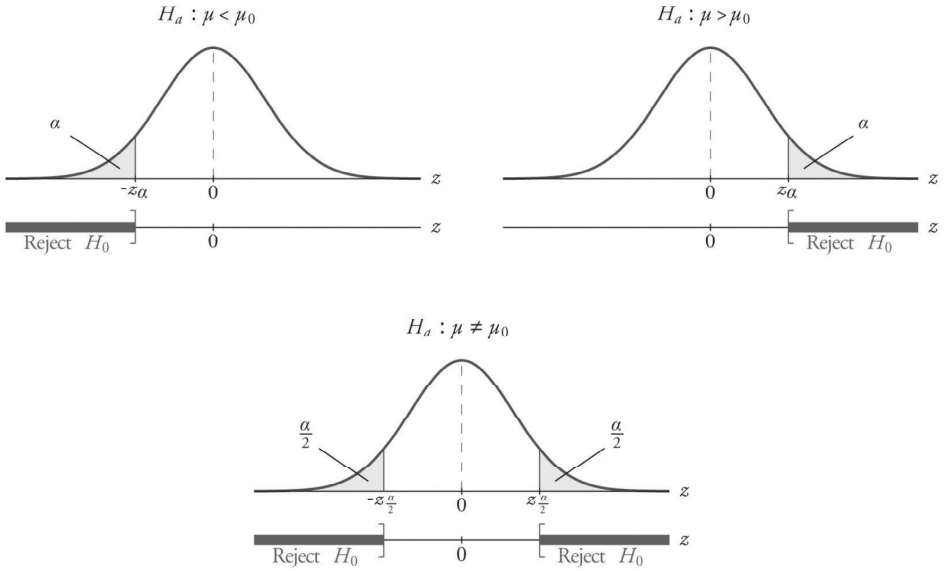
t - Test: t-test is used for small sample i.e., if the sample size n<30.

Critical Region

In testing of hypothesis each test statistic follow certain sampling distribution. As these are also probability distribution so each distribution covers some specified region. Based on the level of significance the region covered by each probability distribution is divided into two mutually exclusive regions, viz. acceptance region and critical region. The critical region of the sampling distribution of a statistic is also known as the α region. It is the area, or areas, of the sampling distribution of a statistic that will lead to the rejection of the hypothesis tested when that hypothesis is true.

Example: When we are testing the null hypothesis $H_0 : \mu = 60$ against the two-sided alternative hypothesis $H_1 : \mu$ "60, the critical region consists of both tails of the sampling distribution of the test statistic. Such a test is a *two-tailed test.*

On the other hand, if we are testing the null hypothesis $H_0 : \mu = 60$ against one-sided alternative $H_1 : \mu > 60$ *or* $H_1 : \mu < 60$, the critical regions are the *right tail* or *left tail* of the sampling distribution of the test statistic respectively.



11.4 STEPS IN TESTING OF HYPOTHESIS

Steps in Hypothesis Testing: The steps follows in testing of hypothesis are as follows:

1. The first and foremost step in hypothesis testing is tospecify the null hypothesis (H_0) and the alternative hypothesis (H_1).
2. The second step is to specify the level of significance α.
3. The third step is the calculation of test statistic which is given by

$$\text{Test Statistic} = \frac{\text{Sample Statistic - Hypothesised Parameter}}{\text{Standard Error of Statistic}}$$

4. The next step is to locate a critical region depending on the level of significance.
5. The final step is to describe the result and the statistical conclusion in a concrete way.



CHECK YOUR PROGRESS

Q1: Distinguish between the following:

(i) Null hypothesis and alternative hypothesis

(ii) One-tailed test and two-tailed test

(iii) Type I error and Type II error

Q2: The procedure of testing hypothesis requires a researcher to adopt several steps. Describe in brief all such steps.

.....

.....

11.5 APPLICATIONS OF HYPOTHESIS TESTING

In this section, some of the standard tests that are most widely used in applications are presented. These are:

1. Test of Single Mean (Large Sample)
2. Test of Single Mean (Small Sample)
3. Test of Difference between Means (Large Sample)
4. Test of Difference between Means (Small Sample)
5. Test of Single Population (Large Sample)
6. Tests for Difference between Two Population Proportions

11.6 TEST OF SINGLE MEAN (LARGE SAMPLE)

Suppose we want to test the null hypothesis $H_0 : \mu = \mu_0$ against one of the alternatives

$H_0 : \mu \neq \mu_0$ or $H_0 : \mu > \mu_0$ or $H_0 : \mu < \mu_0$ on the basis of a large sample n ($n \geq 30$) from the normal population with known variance σ^2. The appropriate test statistic for this test is

$$z = \frac{\bar{x} - \mu}{\sigma / \sqrt{n}}$$

Example : Suppose that it is known from experience that the standard deviation of the weight of 8-ounce packages of cookies made by a certain bakery is 0.16 ounce. To check whether its production is under control on a given day, they select a random sample of 25 packages and find that their mean weight is 8.112 ounce. Is the production process under control? Use 5% level of significance.

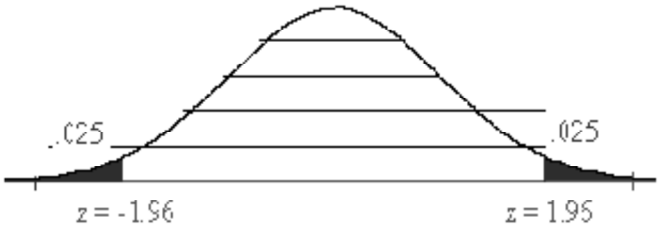
Solution:

1. Null hypothesis : $H_0 : \mu = 8$
2. Alternative hypothesis : $H_1 : \mu \neq 8$
3. Value of Test Statistic,

$$z = \frac{\bar{x} - \mu}{\sigma / \sqrt{n}} = \frac{8.112 - 8}{0.16 / \sqrt{25}} = 3.5$$

4. Significance level : 5 %

5. Critical regions :



6. Conclusion : Since calculated $z = 3.50$ exceeds the critical value 1.96, we reject the null hypothesis at the 5% significance level and conclude that the process is not under control.

11.7 TEST OF SINGLE MEAN (SMALL SAMPLE)

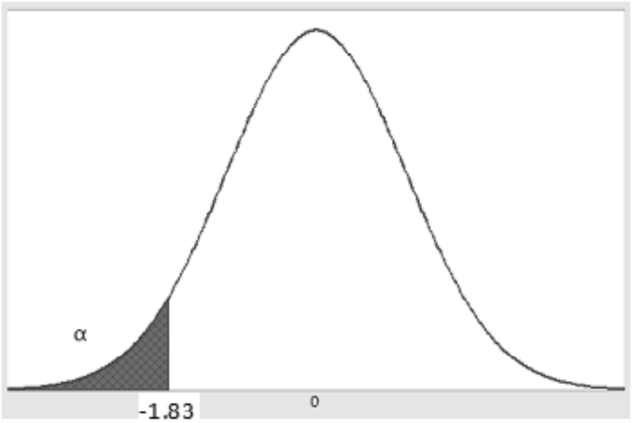
Example: The height of 10 soldiers selected at random had a mean height of 158 cm and variance of 39.0625 cms. Assuming a significance level of 5%, test the hypothesis that the soldiers of the population are on the average lesser than 162.5 cms tall.

Solution

1. Null hypothesis : $H_0 : \mu = 162.5$
2. Alternative hypothesis : $H_1 : \mu < 162.5$
3. Value of Test Statistic,

$$t = \frac{\bar{x} - \mu}{\sqrt{s^2/(n-1)}} = \frac{158 - 162.5}{\sqrt{39.0625/(10-1)}} = -2.163$$

4. Significance level: 5 %
5. Critical regions: Critical value of t at 5% level of significance with n-1 = 10-1 = 9 degrees of freedom is -1.83.



6. Conclusion: Since calculated t = -2.163 falls in the region of rejection, we reject the null hypothesis at the 5% significance level and conclude that the average height of the soldiers is less than 162.5cm.



CHECK YOUR PROGRESS

Q3: The average score in an aptitude test administered at the national level is 80.To evaluate a state's education system, the average score of 100 of the state's students selected on random basis was 75 with a standard deviation. The state wants to know if there is a significant difference between the local scores and national scores. Conduct the testing of hypothesis for the above situation.

Q.4: The height of ten children selected at random from a given locality had a mean 63.2 cms and variance 6.25 cms. Test at 5% level of significance the hypothesis that the children of given locality are on the average less than 65

11.8 TEST OF DIFFERENCE BETWEEN MEANS (LARGE SAMPLE)

This test is done to detect whether or not the means of two samples drawn from two different sources differ significantly. It is also done to check whether the difference is due to a chance and whether the samples belong to the same population.

The steps in the process of testing the significance of the difference between the two means are as follows:

1. The null hypothesis to be tested is written as $H_0: \mu_1 = \mu_2$ against the alternative hypothesis $H_1: \mu_1 \neq \mu_2$
2. Specify the level of significance α.
3. Computing the test statistic

$$z = \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{\left[\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}\right]}}$$

where $\bar{x}_1$ and $\bar{x}_2$ are sample means, σ_1 and σ_2 are standard deviations of population 1 and population 2 respectively and n_1, n_2 are the sample sizes.

4. Determination of critical value and conclusion.

Example: Given that the standard deviations of each of the two populations is 2.5, and the means of two large samples of sizes 1000 and 2000 are 67.5 and 68 respectively. Test the equality of means.

Solution:

1. Null Hypothesis H_0 : $\mu_1 = \mu_2$ (H_0 : There is no difference between means)
2. Alternative Hypothesis H_1: $\mu_1 \neq \mu_2$ (H_1 : There is difference between means)
3. α = 5% or 0.05
4. Test Statistic, z

$$z = \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{\left[\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}\right]}} = \frac{67.5 - 68}{\sqrt{\left[\frac{6.25}{1000} + \frac{6.25}{2500}\right]}} = -5.16$$

5. From the Standard Normal Distribution Table, the critical value of z, at 5% level of significance is ± 1.96.
6. As the computed value of z is greater than critical value of z, it falls in the region of rejection and hence the null hypothesis is rejected.
   It can be concluded that the means are significantly different from each other.

11.9 TEST OF DIFFERENCE BETWEEN MEANS (SMALL SAMPLE)

The steps in the process of testing the significance of the difference between the two means in case of small samples are as follows:

1. The null hypothesis to be tested is written as H_0 : $\mu_1 = \mu_2$ against the alternative hypothesis H_1: $\mu_1 \neq \mu_2$
2. Specify the level of significance α .
3. Computing the test statistic

$$t = \frac{\bar{x}_1 - \bar{x}_2}{S\sqrt{\left[\frac{1}{n_1} + \frac{1}{n_2}\right]}}$$

$$\text{where } S^2 = \frac{n_1 s_1^2 + n_2 s_2^2}{n_1 + n_2 - 2}$$

and s_1, s_2 are the sample standard deviations and n_1 and n_2 are the sample sizes.

The test statistic follows Student's t-distribution with $n_1 + n_2 - 2$ d.f. (Degrees of Freedom).

4. Determination of critical value and conclusion.

Example: Samples of two types of electric bulbs were tested for length of life and following data were obtained.

	Type I	Type II
Sample No.	8	7
Sample Mean	1234 hours	1036 hours
Sample SD	36 hours	40 hours

Is the difference in the means sufficient to warrant that Type I is superior to Type II regarding the length of life?

Solution:

1. Null Hypothesis H_0 : $\mu_1 = \mu_2$ (H_0 : There is no difference between mean life of bulbs)
2. Alternative Hypothesis H_1: $\mu_1 > \mu_2$ (H_1 : Type I is superior to Type II)
3. α = 5% or 0.05
4. Test Statistic is given by

$$S^2 = \frac{n_1 s_1^2 + n_2 s_2^2}{n_1 + n_2 - 2} = \frac{1}{13}\left[8X(36)^2 + 7X(40)^2\right] = 1659.08$$

$$\therefore t = \frac{1234 - 1036}{\sqrt{1659.08\left[\frac{1}{8} + \frac{1}{7}\right]}} = 9.39$$

5. The critical value of t, at 5% level of significance with 13 d.f. is 1.77.
6. As the computed value of t is greater than critical value of t, it falls in the region of rejection and hence the null hypothesis is rejected.

   It can be concluded that the Type I is superior to Type II.

11.10 TEST OF SINGLE PROPORTION (LARGE SAMPLE)

If the observations on various items are categorized into two classes C_1 and C_2, we often want to test the hypothesis, whether the proportion of items in a particular class, say C_1, is P_0 or not. In such cases the null hypothesis becomes

$H_0 : P = P_0$

and $H_1 : P \neq P_0$ or $H_1 : P > P_0$ or $H_1 : P < P_0$

The test statistic for the single proportion is

$$z = \frac{\hat{P} - P_0}{\sqrt{P_0 Q_0 / n}}$$

where $Q_0 = 1 - P_0$ and $\hat{P}$ is the estimated proportion in class C_1.

z follows the standard normal distribution.

Example: To test the conjecture of the management that 60 percent employees favour a new bonus scheme, a sample of 150 employees was drawn and their opinion was taken whether they favour it or not. Only 55 employees out of 150 favoured the bonus scheme. Test the validity of the conjecture of the management at 1% level of significance.

Solution:

1. Null Hypothesis $H_0 : P = 0.60$ (H_0 : 60 percent employees favoured the bonus scheme)
2. Alternative Hypothesis $H_1 : P \neq 0.60$
3. α = 1% or 0.01
4. Test Statistic, z

$$z = \frac{\hat{P} - P_0}{\sqrt{P_0 Q_0 / n}} = \frac{0.367 - 0.60}{\sqrt{(0.60 X 0.40)/150}} = -5.825, \text{ since } \hat{P} = \frac{55}{150} = 0.367$$

5. From the Standard Normal Distribution Table, the critical value of z, at 1% level of significance is ± 2.58.
6. As the computed value of z is greater than critical value of z, it falls in the region of rejection and hence the null hypothesis is rejected.

   It can be concluded that 60 percent employees do not favour the new bonus scheme.

11.11 TESTS FOR DIFFERENCE BETWEEN TWO POPULATION PROPORTIONS

If we have two populations and each item of a population belonged to either of the two classes C_1 and C_2. One may be interested to know whether the proportion of items in C_1 in the both the populations is the same or not. In such cases we apply tests for difference between two population proportions.

The null hypothesis is

$H_0 : P_1 = P_2$

and $H_1 : P_1 \neq P_2$ or $H_1 : P_1 > P_2$ or $H_1 : P_1 < P_2$

The test statistic for tesing the difference between two population proportions is

$$z = \frac{P_1 - P_2}{\sqrt{\hat{P}\hat{Q}\left[\frac{1}{n_1} + \frac{1}{n_2}\right]}}$$

where $\hat{Q} = 1 - \hat{P}$ and $\hat{P}$ is the pooled estimate of proportion given by

$$\hat{P} = \frac{n_1 P_1 + n_2 P_2}{n_1 + n_2}$$

z follows the standard normal distribution.

Example: A sample of 400 families in an old city is selected randomly, and a sample of 500 families is randomly selected from several new colonies of the same city. A survey is conducted for the number of houses possessing washing machine. The number of washing machine holders in old city is 48 out of 400 selected families wheras it was 120 in new colonies out of 500. Test whether there is any difference between proportion of washing machine holders in two colonies.

Solution:

1. Null Hypothesis $H_0 : P_1 =$
2. Alternative Hypothesis $H_1 : P_1 \neq P_2$
3. α = 5% or 0.05
4. Calculation: $P_1 = 0.12$, $P_2 = 0.24$

$$\hat{P} = \frac{n_1 P_1 + n_2 P_2}{n_1 + n_2} = 0.187 \text{ and } \hat{Q} = 1 - \hat{P} = 0.813$$

$$z = \frac{P_1 - P_2}{\sqrt{\hat{P}\hat{Q}\left[\frac{1}{n_1} + \frac{1}{n_2}\right]}} = \frac{0.12 - 0.24}{\sqrt{(0.187)(0.813)(0.0025 + 0.002)}} = -4.59$$

5. From the Standard Normal Distribution Table, the critical value of z, at 1% level of significance is ± 2.58.
6. As the computed value of z is greater than critical value of z, it falls in the region of rejection and hence the null hypothesis is rejected.
   It can be concluded that the proportion of washing machine holders in old city and in new colonies is not the same at 1% level of significance.



11.12 LET US SUM UP

After going through the unit we have come through the following:

- Hypothesis testing is a statistical technique that is used in a variety of situations including decision making.
- The null hypothesis, as the name implies, generally expresses the idea of no difference and usually given the symbol H_0.
- The alternative hypothesis, which we denote by H1, expresses the idea of some difference. Alternative hypothesis determines the tail of a test.
- There are two types of errors in testing of hypothesis: Type I error and Type II error.
- Level of significance plays an important role in testing of hypothesis.
- Test statistic is used to test whether the null hypothesis set up should be accepted or rejected. It is the main yardstick which helps us to decide whether to accept or reject the null hypothesis.
- $$\text{TestStatistic} = \frac{\text{Sample Statistic - Hypothesised Parameter}}{\text{Standard Error of Statistic}}$$

- Some of the standard tests that are most widely used in applications are presented.

  These are:

  - Test of Single Mean (Large Sample)
  - Test of Single Mean (Small Sample)
  - Test of Difference between Means (Large Sample)
  - Test of Difference between Means (Small Sample)
  - Test of Single Population (Large Sample)
  - Tests for Difference between Two Population Proportions



11.13 FURTHER READING

1. Malhotra, N.K. and Dash, S. (2011). Marketing research: An Applied Orientation (6th edition), New Delhi: Pearson Education.



11.14 ANSWERS TO CHECK YOUR PROGRESS

Ans to Q.1. (i) Null Hypothesis: The statistical hypothesis that is set up for testing a hypothesis is known as null hypothesis and it is denoted by H_0. **Alternative Hypothesis:** The negation of null hypothesis is called the alternative hypothesis and it is denoted by H_1.

(ii) The alternative hypothesis determines the tail of a test. For example, let us consider the null hypothesis to be tested be

$H_0: \mu_1 = \mu_2$ against the alternative hypothesis

$H_1: \mu_1 \neq \mu_2$

Then it is called two tailed test.

If the alternative hypothesis is

$H_1: \mu_1 > \mu_2$ (Right Tail Test)

Or $H_1: \mu_1 < \mu_2$ (Left Tail Test)

Then it is known as one tail test.

(iii) Type I Error: Rejection of H_0 when it is, in fact, true.

Type II Error: Acceptance of H_0 when it is false.

Ans to Q2: Steps in Hypothesis Testing: The steps follows in testing of hypothesis are as follows:

1. The first and foremost step in hypothesis testing is tospecify the null hypothesis (H_0) and the alternative hypothesis (H_1).
2. The second step is to specify the level of significance á.
3. The third step is the calculation of test statistic which is given by

$$\text{Test Statistic} = \frac{\text{Sample Statistic - Hypothesised Parameter}}{\text{Standard Error of Statistic}}$$

4. The next step is to locate a critical region depending on the level of significance.
5. The final step is to describe the result and the statistical conclusion in a concrete way.

Ans to Q3:

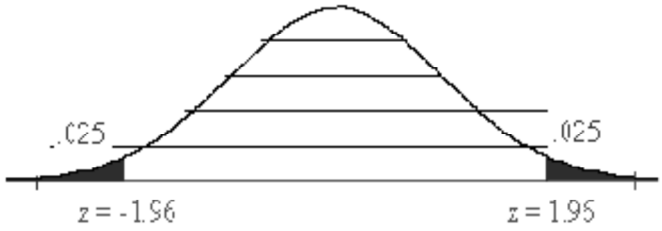
1. Null hypothesis : $H_0 : \mu = 80$
2. Alternative hypothesis : $H_1 : \mu \neq 80$
3. Value of Test Statistic,

$$z = \frac{\bar{x} - \mu}{\sigma/\sqrt{n}} = \frac{75 - 80}{14/\sqrt{100}} = -3.57$$

, since it is a two tail test

$$\therefore |z| = |-3.57| = 3.57$$

4. Significance level : 5 %
5. Critical region :



6. Conclusion: Since calculated *z* = 3.57 exceeds the critical value 1.96, we reject the null hypothesis at the 5% significance level and conclude that there is a significant difference between the local scores and the national scores.\

Ans to Q4:

1. Null hypothesis : $H_0 : \mu = 65$
2. Alternative hypothesis : $H_1 : \mu < 65$
3. Value of Test Statistic,

$$t = \frac{\bar{x} - \mu}{\sqrt{s^2/(n-1)}} = \frac{63.2 - 65}{\sqrt{6.25/(10-1)}} = -2.158$$

4. Significance level: 5 %
5. Critical regions: Critical value of t at 5% level of significance with (n – 1) = (10 – 1) = 9 degrees of freedom is -1.83.
6. Since $|t| = 2.158 > t_{0.05,9} = 1.83$, therefore the null hypothesis is rejected and can be concluded that the mean height of the children is less than 65 cms.



11.15 MODEL QUESTIONS

Q.1. In a random sample of 400 persons from a large population, 120 is females. Can it be said that males and females are in the ratio 5:3 in the population? Use 10% level of significance.

Q.2. The mean I.Q. of a sample of 1600 children was 99. Is it likely that this was a random sample from a population with mean I.Q. 100 and standard deviation 15?

Q.3. The mean weekly sales of the chocolate bar in candy stores was 140.3 bars per store. After an advertisement campaign, the mean weekly sales in 22 stores for a typical week increased to 15307 and showed a standard deviation of 17.2. Was the advertising campaign successful?

Q.4. On a certain day, 74 trains were arriving on time at Delhi station during the rush hours and 83 were late. At New Delhi there were 65 on time and 107 late. Is there any difference in the proportions arriving on time at the two stations?

Q.5. The mean yield of wheat from a district A was 210 kgs with standard deviation 10 kgs. per acre from a sample of 100 plots. In another

district B, the mean yield was 220 kgs with standard deviation 12 kg from a sample of 150 plots. Assuming that the standard deviation of the yield in the entire state was 11 kg, test whether there is any significant difference between the mean yield of crops in the two districts.

*****

UNIT 12: CHI-SQUARE ANALYSIS

UNIT STRUCTURE

12.1 Learning Objectives

12.2 Introduction

12.3 Chi square test

12.4 Chi square test for the Goodness of Fit

12.5 Chi square test for the independence of variables

12.6 Chi square test for the equality of more than two population proportions

12.7 Let Us Sum Up

12.8 Further Reading

12.9 Answers to Check Your Progress

12.10 Model Questions

12.1 LEARNING OBJECTIVES

After going through this unit, you will be able to

- discuss the non-parametric test
- discuss the Independence of attributes
- describe the degrees of freedom
- dicuss chi- Square Distribution
- discuss the Goodness of Fit

12.2 INTRODUCTION

In this Unit we are going to discuss the various concepts of Chi-Square Analaysis.Chi-square is a statistical test commonly used to compare observed data with data we would expect to obtain according to a specific hypothesis. This unit will give us a fair idea about the Chi square test for the Goodness of Fit, Chi square test for the independence of variables and Chi square test for the equality of more than two population proportions.

Let us dicuss this concepts in the following sections.

12.3 CHI SQUARE (χ^2) TEST

Chi square test is applies to those problems in which we study whether the frequency with which a given event has occurred, is significantly different from the one as expected theoretically. The measure of Chi square enables us to find out the degree of discrepancy between observed frequencies and theoretical frequencies and thus to determine whether the discrepancy so obtained between observed frequencies and theoretical frequencies is due to error of sampling or due to chance.

The Chi-square is computed on the basis of frequencies in a sample and thus the value of Chi-square so obtained is a statistic. Chi-square is not a parameter as its value is not derived from the observations in a population. Hence Chi-square test is a non-parametric test.The Chi Square (χ^2) Test is defined as

$$\chi^2 = \sum_{i=1}^{n} \frac{(O_i - E_i)^2}{E_i}, \text{ where}$$

$O_i = \text{observed frequency of the } i^{th} \text{ event}$

$E_i = \text{expected frequency of the } i^{th} \text{ event}$

Degrees of Freedom (df)

The number of data that are given in the form of a series of variables in a row or column or the number of frequencies that are put in cells in a contingency table, which can be calculated independently is called the degrees of freedom and is denoted by ν.

Case I If the data is given in the form of a series of variables in a row or column, then the degrees of freedom = (number of items in the series) - 1, i.e., $\nu = n - 1$.

Case II When the number of frequencies are put in cells in a contingency table, the degrees of freedom will be the product of (number of rows less one) and the (number of columns less one) i.e., ν = (R - 1)(C-1), where R is the number of rows and C is the number of columns.

Uses of Chi Square (χ^2) Test

The χ^2 test is a very powerful test for testing the hypothesis of a number of statistical problems. The important uses of ÷² test are:

12.4 CHI SQUARE TEST FOR GOODNESS OF FIT

χ^2 test is a measure of probabilities of association between the attributes. It gives us an idea about the divergence between the observed and expected frequencies. Thus the test is also described as the test of "Goodness of Fit". If the curves of these two distributions, when superimposed do not coincide or appear to diverge much we say the fit is poor. On the other hand if they do not diverge much, then the fit is less poor.

Example: A survey with 320 families with 5 children each revealed the

No. of Boys:	5	4	5	2	1	0
No. of Girls:	0	1	2	3	4	5
No. of Families:	14	56	110	88	40	12

Is the result consistent with the hypothesis that the male and female births are equally probable?

Solution:

Null Hypothesis H_0 : Male and female births are equaly probable

Alternative Hypothesis H_1 : Male and female births are not equaly probable

Calculation of ÷²: Based on the null hypothesis the expected frequencies are given by

$$320\left[\frac{1}{2}+\frac{1}{2}\right]^5 = 320\left[\frac{1}{32},\frac{5}{32},\frac{10}{32},\frac{10}{32},\frac{5}{32},\frac{1}{32}\right] = 10,50,100,100,50,10$$

Table for χ^2

O	E	O-E	$(O-E)^2$	$(O-E)^2/E$
14	10	4	16	1.60
56	50	6	36	1.72
110	100	10	100	1.00
88	100	-12	144	1.44
40	50	10	100	2.00
12	10	2	4	0.40
				$\chi^2 = 7.16$

Level of Significance: α = 0.05

Critical Value: The table value of χ^2 at 5% level of significance with 6 - 1=5 df is 11.07.

Decision: Since the calculated value of χ^2 is less than table value of χ^2 so null hypothesis is accepted and can be concluded that the male and female births are equally probable.

12.5 CHI SQUARE TEST FOR THE INDEPENDENCE OF VARIABLES

χ^2 test can be applied to test, the independence between various attributes when the sample data is presented in the form of a contingency table with any number of rows 'r' and columns 'c' the null hypothesis and alternative hypothesis are set as follows:

Null Hypothesis H_0 : The attributes are independent

Alternative Hypothesis H_1 : The attributes are not independent

Example: Samples of 400 students of undergraduate and 400 postgraduate classes were taken to know their opinion about autonomous college. 290 undergraduate and 310 postgraduate students favoured the autonomous status. Test whether the opinion about the autonomous status of students independent of their education level or not.

Solution: The contingency table of observed frequencies is

Level of Education	Favour	Not in Favour	Total
Undergraduate	290	110	400
Post graduate	310	90	400
Total	600	200	800

Table of expected frequencies will be

Level of Education	Favour	Not in Favour
Undergraduate	(600 x 400)/800 = 300	(600 x 400)/800 = 300
Post graduate	(200 x 400)/800 = 100	(200 x 400)/800 = 100

Null Hypothesis H_0: Opinion regarding an autonomous status are independent of level of education.

$$\chi^2 = \sum \frac{(O-E)^2}{E} = \frac{(290-300)^2}{300} + \frac{(310-100)^2}{100} + \frac{(110-300)^2}{300} + \frac{(90-100)^2}{300} = 556$$

df = (2-1) (2-1) = 1

Critical Value: Table value of χ^2 at 5% level of significance with 1 df is 3.84.

Decision: Since calculated χ^2 is greater than table value of χ^2, so H_0 is rejected and can be concluded that opinion about autonomous is not independent of level of education.



CHECK YOUR PROGRESS

Q.1: Write short notes about Chi-square test. Why it is called a non parametric test?

.....

.....

Q.2. Define degrees of freedom.

.....

.....

Q.3. Give the test statistic of chi-square test.

.....

.....

Q.4. What are the uses of chi-square test?

.....

.....

12.6 CHI SQUARE TEST FOR THE EQUALITY OF MORE THAN TWO POPULATION PROPORTIONS

Here, the null hypothesis is that the proportion in each the response variable categories is the same for all treatment groups. The alternate hypothesis is that the set of proportions differs somewhere among the treatment groups.

It is possible to write the hypotheses in terms of population parameters. Let π_{ij} be the proportion of treatment population i that is found in response category j. There are G treatment groups and K response categories. The null hypothesis is:

$H : \pi_{11} = \pi_{21} = \pi_{31} = \ldots = \pi_{G1}$

The basic summary statistic in a chi-square analysis is the contingency table which summarizes the number of observations for each combination of the explanatory and response variable. Sample percentages are often computed (denoted as p_{ij} to distinguish them from the π_{ij} in the population).

Example: Suppose assembly elections are announced in three major states. A major political party wishes to test if the proportions of its supporters in the three states are the same or not. The party conducts a sample survey of 1000 people in each state and finds that there are 300, 350 and 425 supporters in the sample in the surveyed state. Using these data test if the three proportions are the same or not.

Solution: Here, we wish to test

H_0: $\Pi_1 = \Pi_2 = \Pi_3$ (All the proportions are the same)

H_1: Not all $\Pi_1 = \Pi_2 = \Pi_3$ are equal

Note that, we only test if all the proportions are the same or not. In case we conclude that all the proportions are not the same, we cannot find which one is different.

We first make the 2X3 contingency table of the given data as below:

Supporter	State 1	State 2	State 3	Total
Yes	300	350	425	1075
No	700	650	575	1925
Total	1000	1000	1000	3000

The expected cell frequencies will be as follows:

Supporter	State 1	State 2	State 3
Yes	358.33	358.33	358.33
No	641.67	641.67	641.67

Test Statistic

$$\chi^2 = \sum \frac{(O-E)^2}{E}$$
$$= \frac{(300-358.33)^2}{358.33} + \frac{(350-358.33)^2}{358.33} + \frac{(425-358.33)^2}{358.33}$$
$$+ \frac{(700-641.67)^2}{641.67} + \frac{(650-641.67)^2}{641.67} + \frac{(575-641.67)^2}{641.67}$$
$$= 9.495 + 0.194 + 12.404 + 5.302 + 0.108 + 6.927$$
$$= 34.43$$

The test statistic follows chi-square distribution with (2-1)(3-1)=2 d.f. Therefore critical value at 5% level of significance is 5.991. Since computed test statistic is larger than the critical value, we reject the null hypothesis at 5% level of significance and conclude that all proportions of supporters of the political party in three states are not the same.



12.7 LET US SUM UP

After going through the unit we have come through the following:

- Chi square test is computed on the basis of frequencies in a sample and thus the value of chi square so obtained is a statistic.
- Chi-square is not a parameter as its value is not derived from the observations in a population. Hence chi-square test is a non parametric test.
- It is not concerned with any population distribution and its observation.
- The number of data that are given in the form of a series of variables in a row or column or the number of frequencies that are put in cells in a contingency table, which can be calculated independently is called the degrees of freedom and is denoted by ν.
- Chi square test can be applied for
  - Test of Goodness of Fit
  - Test for the independence of variables
  - Test for the equality of more than two population proportions



12.8 FURTHER READING

1. Kothari, C.R. and Garg, G.(2014), 'Research Methodology - Methods and Techniques', New Age International Publishers
2. Arora, P.N. and Arora, S.(2009), ' Statistics for Management', S. Chand



12.9 ANSWERS TO CHECK YOUR PROGRESS

Ans to Q.1: The Chi-square is computed on the basis of frequencies in a sample and thus the value of Chi-square so obtained is a statistic. Chi-square is not a parameter as its value is not derived from the observations in a population. Hence Chi-square test is a non-parametric test.

Ans to Q.2: The number of data that are given in the form of a series of variables in a row or column or the number of frequencies that are put in cells in a contingency table, which can be calculated independently is called the degrees of freedom and is denoted by ν.

Ans to Q.3: The χ^2 test is a very powerful test for testing the hypothesis of a number of statistical problems. The important uses of χ^2 test are:

- Test of Goodness of Fit
- Test for the independence of variables
- Test for the equality of more than two population proportions



12.10 MODEL QUESTIONS

Q.1: A sample of 300 students of undergraduate and 300 students of Postgraduate classes of a University were asked to give their opinion toward the autonomous colleges. 190 of the undergraduate and 210 of the Postgraduate students favoured the autonomous status.

Test at 5% level of significance whether the opinion of students towards the status are independent of the classes or not.

Q.2: In 60 throws of a dice, face one turned up 6 times, face two or three 18 times, face four or five 24 times and face six 12 times. Test at 10% significance level whether the dice is unbiased?

Q.3: One thousand girls in a college were graded according to their I.Q. and the economic conditions of their homes. Use chi-square test to find out whether there is any association between economic conditions at home and I.Q. of girls.

Q.4: From the table given below, test whether the colour of son's eye is associated with that of father's eye.

EYE COLOUR OF SONS

Eye Colour of Father	Light	Dark
Light	471	151
Dark	148	230

*****

UNIT 13: ANALYSIS OF VARIANCE

UNIT STRUCTURE

13.1 Learning Objectives

13.2 Introduction

13.3 Introduction of Analysis of Variance

13.4 One-way ANOVA

13.5 Two-way ANOVA

13.6 Design of Experiments

13.7 Principles of Design of Experiments

13.8 Completely Randomized Design (CRD)

13.9 Randomized Block Design (RBD)

13.10 Factorial Design

13.11 Let Us Sum Up

13.12 Further Reading

13.13 Answers to Check Your Progress

13.14 Model Questions

13.1 LEARNING OBJECTIVES

After going through this unit, you will be able to

- learn the meaning of ANOVA
- discuss One-way ANOVA and Two-way ANOVA
- discuss Principles of Design of Experiments
- learn Completely Randomised Design(CRD)
- learn Randomised Block Design(RBD)
- describe Factorial Design

13.2 INTRODUCTION

In this unit we are going to discuss about ANOVA and its various concepts. ANOVA is a statistical method that stands for analysis of variance. This unit will give us a fair idea about the One-way ANOVA and Two-way ANOVA, the principles of Design of Experiments.

At the end of this unit we will discuss the Completely Randomised Design(CRD) and Randomised Block Design(RBD).

13.3 INTRODUCTION OF ANALYSIS OF VARIANCE

ANOVA was developed by Ronald Fisher in 1918 and is the extension of the t and the z test. Before the use of ANOVA, the t-test and z-test were commonly used. But the problem with the T-test is that it cannot be applied for more than two groups. In 1918, Ronald Fisher developed a test called the analysis of variance. This test is also called the Fisher analysis of variance, which is used to do the analysis of variance between and within the groups whenever the groups are more than two.

To use the ANOVA we made the following assumptions:

(i) Each group sample is drawn from a normally distributed population.

(ii) All populations have a common variance,

(iii) All samples are drawn independently of each other.

ANOVA is a method of splitting the total variation of a data into integral parts which measures different sources of variations. The total variations split up into two components:

(a) Variations within the subgroups of samples

(b) Variation between the subgroups of samples

These two variations are tested for their significance by F test which is given by

$$F = \frac{\text{Variance between the samples}}{\text{Variance within the samples}}$$

General Purpose of ANOVA: Researchers are using ANOVA in many ways. The use of ANOVA depends on the research design. Commonly, researchers are using ANOVA in three ways: one-way ANOVA, two-way ANOVA, and N-way Multivariate ANOVA.

Let us discuss these concepts in the following sections:

13.4 ONE-WAY ANOVA

When we compare more than two groups, based on one factor (independent variable), this is called one way ANOVA. For example, it is used if a manufacturing company wants to compare the productivity of three or more employees based on working hours. This is called one way ANOVA.

The one-way analysis of variance (ANOVA) is used to determine whether there are any statistically significant differences between the means of two or more independent (unrelated) groups.

The one-way ANOVA compares the means between the groups you are interested in and determines whether any of those means are statistically significantly different from each other. Specifically, it tests the null hypothesis:

$$H_0 : \mu_1 = \mu_2 = \mu_3 = ... = \mu_k$$

where μ = group mean and k = number of groups.

Under one-way ANOVA, we consider only one factor and then observe that the reason for said factor to be is that several possible types of samples can occur within that factor. We then determine if there are differences within that factor. The technique involves the following steps:

(i) Obtain the mean of each sample i.e., obtain

$\bar{x}_1, \bar{x}_2, \bar{x}_3, ...\bar{x}_k$

when there are k samples.

(ii) Determine the mean of the sample means as follows:

$$\bar{\bar{x}} = \frac{\bar{x}_1 + \bar{x}_2 + \bar{x}_3 + ... + \bar{x}_k}{\text{No. of samples (k)}}$$

(iii) Find the Sum of Squares for variance between the samples (or SS between). Symbolically this can be written as:

$$\text{SS between} = n_1(\bar{x}_1 - \bar{\bar{x}})^2 + n_2(\bar{x}_2 - \bar{\bar{x}})^2 + ... + n_k(\bar{x}_k - \bar{\bar{x}})^2$$

(iv) Obtain mean square between samples. This is given by

$$\text{MS between} = \frac{\text{SS between}}{(k-1)}$$

where (k-1) represents the degrees of freedom between samples.

(v) Obtain sum of squares for variance within samples which is given by

$$\text{SS within} = \sum (x_{1i} - \bar{x}_1)^2 + \sum (\bar{x}_{2i} - \bar{x}_2)^2 + ... + \sum (\bar{x}_{ki} - \bar{x}_k)^2$$

i=1,2,3,...

(vi) Determine mean square within samples which is given by

$$\text{MS within} = \frac{\text{SS within}}{(n - k)}$$

where (n-k) represents the degrees of freedom within samples.

(vii) Finally F-ratio may be worked out as under:

$$\text{F - ratio} = \frac{\text{MS between}}{\text{MS within}}$$

This F-ratio worked as the test statistic and follows F- distribution with (k-1), (n-k) degrees of freedom.

Table 13.1 Analysis of Variance Table for One-way ANOVA

(There are k samples having in all n items)

Source of Variation	Sum of squares (SS)	Degrees of freedom (df)	Mean Square (MS)	F- ratio
Between samples	SS between	(k-1)	MS between = $\frac{\text{SS between}}{(k-1)}$	$\frac{\text{MS between}}{\text{MS within}}$
Within Samples	SS within	(n-k)	MS within = $\frac{\text{SS within}}{(n-k)}$	Source of Variation
Total	SS Total	(n-1)		

13.5 TWO-WAY ANOVA

Two way ANOVA technique is used when the data are classified on the basis of two factors. For example, based on the working hours and working conditions, if a company wants to compare employee productivity, it can do that through two way ANOVA. Two-way ANOVA can be used to see the effect of one of the factors after controlling for the other, or it can be used to see the interaction between the two factors.

Analysis of two-way classified data with one observation per cell

We can plan an experiment in such a way to study the effects of two factors in the same experiment. Let the factors be A and B and the levels $A_1, A_2, ..., A_p$ and $B_1, B_2, ...B_q$. Let y_{ij} be the observation under the i^{th} level of A and the j^{th} level of B.

Here, the mathematical model may be represented as

$$y_{ij} = \mu + \alpha_i + \beta_j + \gamma_{ij},$$

where μ is a constant general effect, present in all observatio ns, α_i is an effect due to the i^{th} level of the factor A, β_j is an effect due to the j^{th} level of the factor B and γ_{ij} is called the interactio n between i^{th} level of the factor A and the j^{th} level of the factor B.

The null hypotheses to be tested here are:

$$H_A : \alpha_1 = \alpha_2 = ... = \alpha_p = 0$$
$$H_B : \beta_1 = \beta_2 = ... = \beta_q = 0$$

for the equality of the effects of the different levels of A and of the different levels of B, respectively. To derive appropriate test for the hypotheses we prepare the following ANOVA Table:

Table 13.2 Analysis of Variance Table for Two-way Classified Data with One Observation per cell

Source of Variation	Sum of squares (SS)	Degrees of freedom (df)	Mean Square (MS)	F-ratio
Between the leveles of A	SSA	(p-1)	$MSA = \frac{SSA}{(p-1)}$	$F = \frac{MSA}{MSE}$
Between the leveles of B	SSB	(q-1)	$MSB = \frac{SSB}{(q-1)}$	$F = \frac{MSB}{MSE}$
Error	SSE	(p-1)(q-1)	$MSE = \frac{SSE}{(p-1)(q-1)}$	
Total	TSS	(pq-1)		

If

$$H_A : \alpha_1 = \alpha_2 = ... = \alpha_p = 0$$

is true, E(MSA) = E(MSE), and hence F =MSA/MSE will give the test of H_A. So a test for the hypothesis of equality of the effects of the different levels of A is provided by this F, which follows the F-distribution with df = (p-1), (p-1)(q-1). The null hypothesis will be rejected at the á level of significance if(and only if)

$$F = \frac{MSA}{MSE} \rangle F_{\alpha;(p-1),(p-1)(q-1)}$$

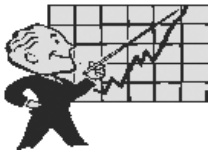
Similarly,

$$H_B : \beta_1 = \beta_2 = ... = \beta_q = 0$$

for the equality of the effects of the different levels of B, is rejected at the α

level of significance if

$$F = \frac{MSB}{MSE} \rangle F_{\alpha;(q-1),(p-1)(q-1)}$$



CHECK YOUR PROGRESS

Q.1: What is ANOVA?

.....

.....

Q.2: What are the assumptions of ANOVA?

.....

.....

Q.3. Which test statistic is used in ANOVA? Write down the test statistic.

.....

.....

13.6 DESIGN OF EXPERIMENTS

Experimental Design is a blueprint of the procedure that enables the researcher to test his hypothesis by reaching valid conclusions about relationships between independent and dependent variables. It refers to the conceptual framework within which the experiment is conducted. One of the main objectives of designing an experiment is how to verify the hypothesis in an efficient and economical way. In the contest of the null hypothesis of equality of several means of normal populations having same variances, the analysis of variance technique can be used. Note that such techniques are based on certain statistical assumptions. If these assumptions are violated, the outcome of the test of hypothesis then may also be faulty and the analysis of data may be meaningless. So the main question is how to obtain the data such that the assumptions are met and the data is readily available for the application of tools like analysis of variance. The designing of such mechanism to obtain such data is achieved by the design of experiment.

13.7 PRINCIPLES OF EXPERIMENTAL DESIGN

There are three basic principles of design which were developed by Sir R.A. Fisher.

(i) Randomization

(ii) Replication

(iii) Local control

(i) Randomization

The principle of randomization involves the allocation of treatment to experimental units at random to avoid any bias in the experiment resulting from the influence of some extraneous unknown factor that may affect the experiment. In the development of analysis of variance, we assume that the errors are random and independent. In turn, the observations also become random. The principle of randomization ensures this.

(ii) Replication:

In the replication principle, any treatment is repeated a number of times to obtain a valid and more reliable estimate than which is possible with one observation only. Replication provides an efficient way of increasing the precision of an experiment. The precision increases with the increase in the number of observations.

(ii) Local control (error control)

The replication is used with local control to reduce the experimental error. For example, if the experimental units are divided into different groups such that they are homogeneous within the blocks, then the variation among the blocks is eliminated and ideally the error component will contain the variation due to the treatments only. This will in turn increase the efficiency.

13.8 COMPLETELY RANDOMISED DESIGN (CRD)

The CRD is the simplest design. Suppose there are *v* treatments to be compared.

- All experimental units are considered the same and no division or grouping among them exist.

- In CRD, the *v* treatments are allocated randomly to the whole set of experimental units, without making any effort to group the experimental units in any way for more homogeneity.
- Design is entirely flexible in the sense that any number of treatments or replications may be used.

  Example: Suppose there are 4 treatments and 20 experimental units, then

    - the treatment 1 is replicated, say 3 times and is given to 3 experimental units,
    - the treatment 2 is replicated, say 5 times and is given to 5 experimental units,
    - the treatment 3 is replicated, say 6 times and is given to 6 experimental units

  and

    - finally, the treatment 4 is replicated [20-(6+5+3)=]6 times and is given to the remaining 6 experimental units
- All the variability among the experimental units goes into experimented error.
- CRD is used when the experimental material is homogeneous.
- CRD is more useful when the experiments are conducted inside the lab.
- CRD is well suited for the small number of treatments and for the homogeneous experimental material.

Layout of CRD

Following steps are needed to design a CRD:

- Divide the entire experimental material or area into a number of experimental units, say *n*.
- Fix the number of replications for different treatments in advance (for given total number of available experimental units).
- No local control measure is provided as such except that the error variance can be reduced by choosing a homogeneous set of experimental units.

Analysis

There is only one factor which is affecting the outcome – treatment effect. So the set-up of one-way analysis of variance is to be used.

13.9 RANDOMISED BLOCK DESIGN (RBD)

The completely randomized design (CRD) will seldom be used if the experimental units are not alike. For in that case the variation among the units will vitiate the test of significance of the treatment effects. The simplest design which enables us to take care of the variability among the units is the RBD. This is also the simplest design using all the three principles enunciated by Fisher.

Suppose we want to compare the effects of p treatments, each treatment being replicated an equal number o times, say q times. Then we used n = pq experimental units, and these units are not homogeneous. The RBD consists of two steps. The first step is to divide the units into p more or less homogeneous groups. In each group or block, we take as many units as there are treatments. Thus the number of blocks is the same as the common replication number (p). The same technique should be applied to the units of a block.

The second step is to assign the treatments at random to units of a block. This randomisation has to be done afresh for each block. This is the difference between of an RBD from a CRD. In an RBD randomisation is restricted within a homogeneous block.

Layout and analysis of RBD follows the analysis of variance for a two-way classified data.

13.10 FACTORIAL DESIGN

Experiments where the effects of more than one factor, say for example variety, manure, etc., each at two or more levels, are considered together are called factorial experiments, while experiments with one factor at varying levels, say only variety or manure, may be called the simple experiments. A Factorial Design is an experimental setup that consists of

multiple factors and their separate and conjoined influence on the subject of interest in the experiment. A factor is an independent variable in the experiment and a level is a subdivision of a factor. Factors and levels are different conditions that the experimental subjects are exposed to. A study or experiment is used to see if any of the conditions influence the subject and in what ways they are influential. The benefit of a factorial design is that it allows the researchers to look at multiple levels at a time and how they influence the subjects in the study.



13.11 LETS US SUM UP

After going through the unit we have come through the following:

- ANOVA is a method of splitting the total variation of a data into integral parts which measures different sources of variations.
- It is used to do the analysis of variance between and within the groups whenever the groups are more than two.
- In the contest of the null hypothesis of equality of several means of normal populations having same variances, the analysis of variance technique can be used.
- The use of ANOVA depends on the research design. Commonly, researchers are using ANOVA in three ways: one-way ANOVA, two-way ANOVA, and N-way Multivariate ANOVA.
- Experimental Design is a blueprint of the procedure that enables the researcher to test his hypothesis by reaching valid conclusions about relationships between independent and dependent variables.
- There are three basic principles of design which were developed by Sir R.A. Fisher.
  (i) Randomization
  (ii) Replication
  (iii) Local control
- Based on the application of the principles of design, many design are being developed, viz.

(a) Completely Randomized Design (CRD)

(b) Randomised Block Design (RBD)

(c) Latin Square Design (LSD)

(d) Graco Latin Square Design

(e) Factorial Design, etc.



13.12 FURTHER READING

1. Kothari, C.R. and Garg, G.(2014), '*Research Methodology - Methods and Techniques'*, New Age International Publishers
2. Malhotra. N.K.(2011), *Marketing Research – An Applied Orientation*, Pearson



13.13 ANSWERS TO CHECK YOUR PROGRESS

Ans to Q.1. ANOVA is a statistical method that stands for analysis of variance. ANOVA was developed by Ronald Fisher in 1918 and is the extension of the t and the z test. Before the use of ANOVA, the t-test and z-test were commonly used. But the problem with the T-test is that it cannot be applied for more than two groups. In 1918, Ronald Fisher developed a test called the analysis of variance. This test is also called the Fisher analysis of variance, which is used to do the analysis of variance between and within the groups whenever the groups are more than two.

Ans to Q.2. To use the ANOVA we made the following assumptions:

(i) Each group sample is drawn from a normally distributed population.

(ii) All populations have a common variance,

(iii) All samples are drawn independently of each other.

Ans to Q.3. The test statistic used in ANOVA technique is F-test. It is given by

$$F = \frac{\text{Variance between the samples}}{\text{Variance within the samples}}$$



13.14 MODEL QUESTIONS

Q.1: What are the assumptions of ANOVA technique?

Q.2: What is Design of Experiment?

Q.3: What are the principles of Design of Experiment laid by R.A. Fisher?

*****

UNIT 14: RESEARCH REPORT WRITING

UNIT STRUCTURE

14.1 Learning Objectives

14.2 Introduction

14.3 Meaning and Significance of Report Writing

14.4 Types of Research Reports

14.4.1 Brief Research Report

14.4.2 Detailed Research Report

14.5 Structure of the Research Report

14.6 Report Writing: Formulation rules for writing report

14.7 Guidelines for presenting tabular data

14.8 Guidelines for visual representations

14.9 Let Us Sum Up

14.10 Further Reading

14.11 Answers to Check Your Progress

14.12 Model Questions

14.1 LEARNING OBJECTIVES

After going through this unit, you will be able to

- discuss the concept and the importance of research report
- learn to write a Brief report
- learn to write Detailed report
- outline the structure of the research report
- discuss the formulation rules for writing the report
- describe the guidelines for presenting tabular data,
- describe the guidelines for visual Representations

14.2 INTRODUCTION

Report writing is very important in a reserach process.The purpose of writing a report is to convey the results or progress of research done.Writing of report is the last step in a research study and requires a set

of skills. We will discuss the types of research reports, structure of the research report, formulation rules for writing the report.

We will also discuss the guidelines for presenting tabular data and Guidelines for visual Representations.

14.3 MEANING AND SIGNIFICANCE OF REPORT WRITING

Research report is considered a major component of the research study for the research task remains incomplete till the report has been presented and/or written. As a matter of fact even the most brilliant hypothesis, highly well designed and conducted research study, and the most striking generalizations and findings are of little value unless they are effectively communicated to others. The purpose of research is not well served unless the findings are made known to others. Research results must invariably enter the general store of knowledge. All this explains the significance of writing research report. There are people who do not consider writing of report as an integral part of the research process. But the general opinion is in favour of treating the presentation of research results or the writing of report as part and parcel of the research project. Writing of report is the last step in a research study and requires a set of skills somewhat different from those called for in respect of the earlier stages of research. This task should be accomplished by the researcher with utmost care; he may seek the assistance and guidance of experts for the purpose.

14.4 TYPES OF RESEARCH REPORTS

Reports can be broadly classified into two categories: brief report and detailed report.

14.4.1 Brief Research Reports

Brief reports are most often used by newspapers and other periodicals to share research or information about a currently trending topic or as an enhancement to a longer story. These reports are similar in structure to detailed reports, but without any concrete

opinions or the drawing of conclusions by the author. Brief reports are essentially meant to inform the reader and provide the essential information about a particular topic. Different components of a brief report are - Summary, Background, Purpose, Result and Conclusions.

Summary: The first part of a brief report is typically a short summary that names the main points of the research or topic being covering, followed by the names of author(s) and where they did their research. The summary serves as the introduction and usually includes no more than one to three sentences. Examples of appropriate summary sentences would be, “Those with a family history of mental illness and addiction are more likely to suffer from substance abuse than others. A recent study conducted by the National Institute of Mental Health found that those who had a family history of mental illness and addiction were 10 times more likely to become addicts themselves.”

Background: The next part of the brief report provides background to the statements made in the summary (introduction). Background information should include more details about the subject, why the study is important and the current state of the research. Example: “Researchers found that of the 200 students surveyed, 20 percent reported being exposed to drugs or alcohol before the age of 12 by a parent or other caregiver who had diagnosed mental illness and suffered from addiction.”

Purpose: The purpose of the study is to provide readers with reasons they should care about the information and to lend credibility to the research. Research must have a purpose in order to carry out the research. For example the summarisation can be written for the above study as: “The purpose of the study is to increase awareness of the effects of mental illness on society and to encourage those suffering to seek treatment.”

Results and Conclusion: Brief reports do not require the drawing of detailed conclusions as in the case of detailed reports.

However, it requires to make a few writings as to why the research or study is important to the society. Using the results of these reports various measures can be taken to check the problem. Organizations can use these results to help increase community awareness about the dangers of substance abuse and mental illness, as well as to provide assistance to those who are struggling.

14.4.2 Detailed Research Report

Research reports are the product of slow, painstaking, accurate inductive work. The research report is a means of communication that can be understood, believed, trusted by everyone who are likely to be affected by the research, and acted upon by the decision maker. Effective report writing is an art. Some basic points to note in writing a report includes:

Readers: The report should take into account the level of readers' technical sophistication, their interest in the project, ability to understand as well as the circumstances under which they will read the report and how they will use it.

Adherence to study objectives: A research report must show that the research objectives have been accomplished.

Easy to follow: The most basic characteristics of a good report is that it is easy to follow, well organised, logically structured and clearly and lucidly written.

Different Steps in Writing Report

(i) **Logical analysis of the subject-matter:** It is the first step which is primarily concerned with the development of a subject. There are two ways in which to develop a subject logically and chronologically. The logical development is made on the basis of mental connections and associations between the one thing and another by means of analysis. Logical treatment often consists in developing the material from the simple possible to the most complex structures. Chronological development is based on a connection or sequence in time or occurrence. The

directions for doing or making something usually follow the chronological order.

(ii) **Preparation of the final outline:** It is the next step in writing the research report. Outlines are the framework upon which long written works are constructed. They are an aid to the logical organization of the material and a reminder of the points to be stressed in the report.

(iii) **Preparation of the rough draft:** This follows the logical analysis of the subject and the preparation of the final outline. Such a step is of utmost importance for the researcher now sits to write down what he has done in the context of his research study. He will write down the procedure adopted by him in collecting the material for his study along with various limitations faced by him, the technique of analysis adopted by him, the broad findings and generalizations and the various suggestions he wants to offer regarding the problem concerned.

(iv) **Rewriting and polishing:** This step happens to be most difficult part of all formal writing. Usually this step requires more time than the writing of the rough draft. The careful revision makes the difference between a mediocre and a good piece of writing. While rewriting and polishing, one should check the report for weaknesses in logical development or presentation. The researcher should also "see whether or not the material, as it is presented, has unity and cohesion; does the report stand upright and firm and exhibit a definite pattern, like a marble arch? Or does it resemble an old wall of moldering cement and loose brick." In addition the researcher should give due attention to the fact that in his rough draft he has been consistent or not. He should check the mechanics of writing—grammar, spelling and usage.

(v) **Preparation of the final bibliography:** Next in order comes the task of the preparation of the final bibliography. The bibliography, which is generally appended to the research report,

is a list of books in some way pertinent to the research which has been done. It should contain all those works which the researcher has consulted. The bibliography should be arranged alphabetically and may be divided into two parts; the first part may contain the names of books and pamphlets, and the second part may contain the names of magazine and newspaper articles. Generally, this pattern of bibliography is considered convenient and satisfactory from the point of view of reader, though it is not the only way of presenting bibliography. The entries in bibliography should be made adopting the following order:

For books and pamphlets the order may be as under:

(a) Name of author, last name first.
(b) Title, underlined to indicate italics.
(c) Place, publisher, and date of publication.
(d) Number of volumes.

Example:

Kothari, C.R., Quantitative Techniques, New Delhi, Vikas Publishing House Pvt. Ltd., 1978.

For magazines and newspapers the order may be as under:

(a) Name of the author, last name first.
(b) Title of article, in quotation marks.
(c) Name of periodical, underlined to indicate italics.
(d) The volume number and issue number.
(e) The date of the issue.
(f) The pagination.

Example:

Robert V. Roosa, "Coping with Short-term International Money Flows", *The Banker,* London, September, 1971, p.995

(vi) Writing the Final Draft: The final draft should be written in a concise and objective style and in simple language, avoiding vague expressions such as "It seems", "There may be", etc. While writing the final draft, the researcher must avoid abstract

terminology and technical jargons. Finally, a report should not be dull, but must enthuse people and maintain interest and must show originality.

14.5 STRUCTURE OF THE RESEARCH REPORT

A comprehensive layout of the research report should comprise (a) Preliminary section, (b) the main report, and (c) the end matter.

(a) Preliminary section: The preliminary section comprise of *Title and Date*, followed by acknowledgements in the form of *Preface* or *Foreword.* Then there should be a *Table of Contents* followed by *List of Tables and Illustrations* so that anybody can locate the required information in the report.

(b) Main Report: The main text provides the complete outline of the research report along with all details. Title of the research study is repeated at top of the first page of the main text and then follows the other details on pages numbered consecutively, beginning with the second page. Each main section of the report should begin on a new page. The main text of the report should have the following sections:

(i) Introduction: The purpose of introduction is to introduce the research project to the readers. It shoild contain a clear statement of the objectives of research. a brief summary of the other relevant may also be stated so that the present study can be seen in that context. the hypotheses of study,if any, and the definitions of the major concepts employed in the study should be explicitlystated in the introduction of the report.

(ii) Research Methodology: In this section research methodology adopted in conducting the study must be fully explained. The statistical analysis applied must also be clearly stated. In addition to all these the scope of the study and limitations under which the project was completed must also be narrated.

(iii) Statement of Findings and Recommendations: In this section findings and recommendations in non-technical language must be included so that it can be understood by all concerned.

(iv) Results: A detailed presentation of the findings of the study, with supporting data in the form of tables and charts together with a validation of results, is the next step in writing the main text of the report. This generally comprises the main body of the report, extending over several chapters. The result section of the report should contain statistical summaries and reductions of the data rather than the raw data. All the results should be presented in logical sequence and splitted into readily identifiable sections. All relevant results must find a place in the report.

(v) Implications of the Results: Towards the end of the main text, the researcher should again put down the results of his research clearly and precisely. He should, state the implications that flow from the results of the study, for the general reader.

(v) Summary: It has become customary to conclude the research report with a very brief summary, resting in brief the research problem, the methodology, the major findings and the main conclusions drawn from the research results.



CHECK YOUR PROGRESS

Q.1: What are the components of main report?

.....

.....

Q.2. Give an example of preparing bibliography.

.....

.....

Q.3. Illustrate the steps of writing a research report.

.....

.....

14.6 REPORT WRITING: FORMULATION RULES FOR WRITING REPORT

There are very definite and set rules which should be followed in the actual preparation of the research report.

Size and physical design: The manuscript should be written/typed on unruled paper. A margin of at least one inch should be allowed at both hands, top and bottom. The paper should be neat and legible. If the manuscript is to be typed, then all typing should be double-spaced on one side of the page only except for the insertion of the long quotations.

Procedure: Various steps in writing the report should be strictly adhered.

Layout: Keeping in view the objective and nature of the problem, the layout of the report should be thought of and decided and accordingly adopted.

Treatment of quotations: Quotations should be placed in quotation marks and double spaced, forming an immediate part of the text. But if a quotation is of a considerable length (more than four or five type written lines) then it should be single-spaced and indented at least half an inch to the right of the normal text margin.

Footnotes: Regarding footnotes one should keep in view the followings:
The footnotes serve two purposes viz., the identification of materials used in quotations in the report and the notice of materials not immediately necessary to the body of the research text but still of supplemental value. In other words, footnotes are meant for cross references, citation of authorities and sources, acknowledgement and elucidation or explanation of a point of view.

Documentation style: Regarding documentation, the first footnote reference to any given work should be complete in its documentation, giving all the essential facts about the edition used.

Use of statistics, charts and graphs: A judicious use of statistics in research reports is often considered a virtue for it contributes a great deal towards the clarification and simplification of the material and research results. Statistics are usually presented in the form of tables, charts, bars and line-graphs and pictograms. Such presentation should be self explanatory and complete in itself.

The final draft: Revising and rewriting the rough draft of the report should be done with great care before writing the final draft.

Bibliography: Bibliography should be prepared and appended to the research report as discussed earlier.

Preparation of the index: At the end of the report, an index should invariably be given, the value of which lies in the fact that it acts as a good guide, to the reader. Index may be prepared both as subject index and as author index. The former gives the names of the subject-topics or concepts along with the number of pages on which they have appeared or discussed in the report, whereas the latter gives the similar information regarding the names of authors. The index should always be arranged alphabetically.

14.7 GUIDELINES FOR PRESENTING TABULAR DATA

Tables, charts, and graphs are often used to report data and statistics. Each of these three graphic devices conveys information by showing

1) values
2) labels and
3) context.

Values are data; they may be numbers, symbols, or text. Values by themselves are meaningless: 120, +, or “mild,” for example. Thus, they must be accompanied by labels. *Labels* identify the values and often add additional characteristics: in the above examples, a mean systolic pressure of 120 mm Hg; a positive diagnostic test result; and “mild” symptoms of Alzheimer’s disease. Even values with labels, however, are meaningless in the absence of a larger context, such as a specific research project.

Context is the background against which values and labels should be interpreted. To continue the above example, a mean systolic pressure of 120 mm Hg might be the therapeutic endpoint for patients with hypertension; a positive diagnostic test result might define a subgroup of patients with comorbidities (the presence of one or more additional disorders); and “mild” symptoms of Alzheimer’s disease might be an inclusion criterion in a drug study. In tables, charts, and graphs, the context should be identified in the titles or captions, as well as be apparent from the larger document.

14.8 GUIDELINES FOR VISUAL REPRESENTATIONS

At times oral or visual presentation of the results of the study is effective, particularly in cases where policy recommendations are indicated by project results. The merit of this approach lies in the fact that it provides an opportunity for give-and-take decisions which generally lead to a better understanding of the findings and their implications.

Visual presentation is effective when supplemented by various visual devices. Use of slides, wall charts and boards is quite helpful in contributing to clarity and in reducing the boredom, if any. Distributing a board outline, with a few important tables and charts concerning the research results, makes the listeners attentive who have a ready outline on which to focus their thinking. This very often happens in academic institutions where the researcher discusses his research findings and policy implications with either in a seminar or a group discussion.



14.9 LET US SUM UP

After going through the unit we have come through the following:

- Research report is considered a major component of the research study for the research task remains incomplete till the report has been presented and/or written.
- The purpose of research is not well served unless the findings are made known to others. Research results must invariably enter the general store of knowledge.
- Reports can be broadly classified into two categories: brief report and detailed report.
- Different steps in writing report involve:
  - (i) Logical analysis of the subject-matter
  - (ii) Preparation of the final outline
  - (iii) Preparation of the rough draft
  - (iv) Rewriting and polishing
  - (v) Preparation of the final bibliography

(vi) Writing the Final Draft

- A comprehensive layout of the research report should comprise
  (a) Preliminary section,
  (b) the main report, and
  (c) the end matter.



14.10 FURTHER READING

1. Kothari, C.R. and Garg, G.(2014), 'Research Methodology - Methods and Techniques', New Age International Publishers
2. Malhotra. N.K.(2011), Marketing Research – An Applied Orientation, Pearson



14.11 ANSWERS TO CHECK YOUR PROGRESS

Ans to Q1. The main text provides the complete outline of the research report along with all details. The main text of the report should have the following sections:

(a) Introduction
(b) Research Methodology
(c) Statement of Findings and Recommendations
(d) Results
(e) Implications of the Results
(f) Summary

Ans to Q.2. The entries in bibliography should be made adopting the following order:

For books and pamphlets the order may be as under:

(a) Name of author, last name first.
(b) Title, underlined to indicate italics.
(c) Place, publisher, and date of publication.
(d) Number of volumes.

Example:

Kothari, C.R., Quantitative Techniques, New Delhi, Vikas Publishing House Pvt. Ltd., 1978.

For magazines and newspapers the order may be as under:

(a) Name of the author, last name first.

(b) Title of article, in quotation marks.

(c) Name of periodical, underlined to indicate italics.

(d) The volume number and issue number.

(e) The date of the issue.

(f) The pagination.

Example:

Robert V. Roosa, "Coping with Short-term International Money Flows", *The Banker,* London, September, 1971, p.995

Ans to Q3: Different steps in writing report involve:

(i) Logical analysis of the subject-matter

(ii) Preparation of the final outline

(iii) Preparation of the rough draft

(iv) Rewriting and polishing

(v) Preparation of the final bibliography

(vi) Writing the Final Draft



14.12 MODEL QUESTIONS

Q.1: "It is only through interpretation the researcher can expose the relations and processes that underline his findings". Explain, giving examples.

Q.2: Explain the significance of a research report and narrate the various steps involved in writing such a report.

Q.3: Describe, in brief, the layout of a research report, covering all relevant points.

Q.4: What is a visual presentation?

Q.5: What are the different forms in which a research work may be reported. Describe.

*****

UNIT 15: ETHICS IN RESEARCH

UNIT STRUCTURE

15.1 Learning Objectives

15.2 Introduction

15.3 Meaning of Research Ethics

15.4 Clients Ethical code

15.5 Researchers Ethical code

15.6 Ethical Codes related to respondents

15.7 Responsibility of ethics in research

15.8 Let Us Sum Up

15.9 Further Reading

15.10 Answers to Check Your Progress

15.11 Model Questions

15.1 LEARNING OBJECTIVES

After going through this unit, you will be able to

- learn the concept of research ethics
- discuss the Clients Ethical code
- discuss the Researchers Ethical code
- describe the Ethical Codes related to respondents
- outline the responsibility of ethics in research

15.2 INTRODUCTION

Ethics or moral philosophy is a branch of philosophy that involves systematizing, defending, and recommending concepts of right and wrong conduct. The term ethics derives from the Greek word ‘ethos’, which means habit and custom. Research ethics involves the application of fundamental ethical principles to be applied in planning, conducting and reporting of research.Let us discuss the differnt aspects of ethics in reserach in the following sections.

15.3 MEANING OF RESEARCH ETHICS

Research that involves human subjects or participants raises unique and complex ethical, legal, social and political issues. Research ethics is specifically interested in the analysis of ethical issues that are raised when people are involved as participants in research. There are three objectives in research ethics. The first and the broadest objective is to protect human participants. The second objective is to ensure that research is conducted in a way that serves the interests of individuals, groups and/or society as a whole. Finally the third objective is to examine the specific research activities and projects for their ethical soundness, looking at issues such as the management of risk, protection of confidentiality and the process of informed consent.

For the most part, research ethics has traditionally focused on issues in biomedical research. The application of research ethics to examine and evaluate biomedical research has been well developed over the last century and has influenced much of the existing guidelines for the ethical issues arise. However, in humanities and social science research, different kinds of ethical issues arise. New and emerging methods of conducting research, such as autoethnography and participatory action research raise important but markedly different ethical issues and obligations for research.

Research involving vulnerable persons, which may include children, persons with developmental or cognitive disabilities, persons who are institutionalized, the homeless or those without legal status, also raises unique issues in any research context.

Research ethicists everywhere today are challenged by issues that reflect global concerns in other domains, such as the conduct of in developing countries, the limits of research involving genetic material and protection of privacy in light of advances in technology and internet capabilities.

15.4 CLIENTS' ETHICAL CODE

Client means any or organisation, department or division (including any belonging to the same organisation as the researcher) that requests,

commissions or subscribes to all or any part of a market, social and organisational research project.

Relationships between a Survey Research Organization and Clients for whom the surveys are conducted should be of such a nature that they foster confidence and mutual respect. They must be characterized by honesty and confidentiality.

The following specific approaches describe in more detail the responsibilities of Research Organizations in this relationship:

1. A Survey Research Organization must assist its clients in the design of effective and efficient studies that are to be carried out by the Research Organization. If the Survey Research Organization questions whether a study design will provide the information necessary to serve the client's purposes, it must make its reservations known.
2. A Research Organization must conduct the study in the manner agreed upon. However, if it becomes apparent in the course of the study that changes in the plans should be made, the Research Organization must make its views known to the client promptly.
3. A Research Organization has an obligation to allow its clients to verify that work performed meets all contracted specifications and to examine all operations of the Research Organization that are relevant to the proper execution of the project in the manner set forth. While clients are encouraged to examine questionnaires or other records to maintain open access to the research process, the Survey Research Organization must continue to protect the confidentiality and privacy of survey Respondents.
4. When more than one client contributes to the cost of a project specially commissioned with the Research Organization, each client concerned shall be informed that there are other Participants (but not necessarily their identity).
5. Research Organizations will hold confidential all information that they obtain about a client's general business operations, and about matters connected with research projects that they conduct for a client.

6. For research findings obtained by the agency that are the property of the client, the Research Organization may make no public release or revelation of findings without expressed, prior approval from the client.

15.5 RESEARCHERS' ETHICAL CODE

Researcher means any individual or organisation (or that part of an organisation) that carries out, or act as a consultant or subcontractor for market, social and organisational research, or offers their services or the services of others to do so, including those working in client organisations.

General rules for Researcher

Few general rules researcher must follow are:

1. Researchers must conform to all relevant state, national and international laws.
2. Researchers must behave ethically and must not do anything which might damage the reputation of market, social and organisational research.
3. Researchers must take special care when carrying out research among children and young people and other vulnerable groups in the community.
4. Participants' cooperation is voluntary and must be based on adequate, and not misleading, information about the general purpose and nature of the project when their agreement to participate is being obtained and all such statements must be honoured.
5. The rights of participants as private individuals must be respected by researchers and they must not be harmed or adversely affected as a result of cooperating in a research project.
6. Participants' identifiable research information must not, without their consent, be revealed to anyone not directly involved in the research project and not be used for any non-research activity directed to individual participants.

7. Researchers must ensure that projects and activities are designed, carried out, reported and documented accurately, transparently and objectively.
8. Researchers must conform to accepted principles of fair competition.

Moreover, the researcher has a duty to safeguard the issues pertaining to: Participants' anonymity must be strictly preserved. If the participant on request from the researcher has given consent for their personal information or identifiable research information to be disclosed:

(a) The participant must first have been told:
   i) to whom the information would be supplied and
   ii) the purposes for which it will be used.

(b) The researcher must take reasonable steps to ensure that:
   i) the personal information will not be used for any non-research activity, and
   ii) only that part of the information considered necessary for the purpose is disclosed.

(c) Researchers must ensure that research projects are designed, carried out, reported and documented accurately, transparently and objectively.

(d) Researchers must only collect sensitive information (whether from a participant or from a third party) where the participant has consented, and the information is reasonably necessary for a research purpose.



CHECK YOUR PROGRESS

Q.1. What do you mean by research ethics?

.....

.....

Q.2. Write about clients' ethical code.

.....

.....

Q.3. Fill in the blanks

a. Participants' ..... must be strictly preserved.

b. Researchers must take special care when carrying out research among children and young people and other ..... groups in the community.

15.6 ETHICAL CODES RELATED TO RESPONDENTS

Researchers have professional and legal responsibilities to their respondents that are embodied in the procedures of a research study. Underlying these specific responsibilities are four fundamental ethical principles:

Respondents should be:

(a) willing participants in survey research;

(b) appropriately informed about the survey's intentions and how their personal information and survey responses will be used and protected;

(c) sufficiently satisfied with their survey experience;

(d) willing to participate again in survey research.

Confidentiality of Respondents:

1. Since individuals who are interviewed are the lifeblood of the Survey Research Industry, it is essential that Survey Research Organizations be responsible for protecting from disclosure to third parties—including Clients and members of the Public—the identity of individual Respondents as well as Respondent identifiable information, unless the Respondent expressly requests or permits such disclosure.
2. A minimal amount of Respondent-identifiable information will be disclosed to the client to permit the client to validate interviews and to determine an additional fact of analytical importance to the study.
3. Before disclosing Respondent-identifiable information to a client for purposes of interview validation or re-inquiry, the Survey Research Organization must take whatever steps are needed to ensure that the client will conduct the validation or recontact in a fully professional manner.

4. Assurance to respondents will be given that the Client will respect such limitations and maintain Respondent confidentiality.
5. To assure client compliance, the Survey Research Organization must obtain written confirmation from the client before releasing any data.
6. The research firm has obtained a written agreement from the client assuring that no other use will be made of Respondent-identifiable information.

15.7 RESPONSIBILITY OF ETHICS IN RESEARCH

There are several reasons why it is important to adhere to ethical norms in research.

First norm promote the aim of research, such as knowledge, truth and avoidance of error. For example, prohibiting against fabricating and falsifying or misrepresenting research data promote the truth.

Second, research often involves great deal of cooperation and coordination among many different people in different disciplines and institutions, ethical standards promote the values that are essential to collaborative work, such as trust, accountability, mutual respect and fairness. For example, many ethical norms in research, such as guidelines for authorship, copyright and patenting policies, data sharing policies and confidentiality rules in peer review, are designed to protect intellectual property interests while encouraging collaboration. Most researchers want to receive credit for their contributions and do not want to have their ideas stolen or disclosed prematurely.

Third, many of the ethical norms help to ensure that researchers can be held accountable to the public. For instance, conflicts of interest, the human subjects' protections and animal care and use are necessary to make sure that researchers who are funded by public money can be held accountable to the public.

Fourth, ethical norms in research also help to build public support for research. People are more likely to fund a research project if they can trust the quality and integrity of research.

Finally, many of the norms of research promote a variety of other important moral and social values, such as social responsibility, human rights, animal welfare, compliance with the law, and public health and safety.



15.8 LET US SUM UP

After going through the unit we have come through the following:

- Ethics or moral philosophy is a branch of philosophy that involves systematizing, defending, and recommending concepts of right and wrong conduct.
- Research that involves human subjects or participants raises unique and complex ethical, legal, social and political issues. Research ethics is specifically interested in the analysis of ethical issues that are raised when people are involved as participants in research.
- Relationships between a Survey Research Organization and Clients for whom the surveys are conducted should be of such a nature that they foster confidence and mutual respect. They must be characterized by honesty and confidentiality.
- Few general rules researcher must follow are:
  1. Researchers must conform to all relevant state, national and international laws.
  2. Researchers must behave ethically and must not do anything which might damage the reputation of market, social and organisational research.
  3. Researchers must take special care when carrying out research among children and young people and other vulnerable groups in the community.
- Researchers have professional and legal responsibilities to their respondents that are embodied in the procedures of a research study. Underlying these specific responsibilities are four fundamental ethical principles:

  (a) willing participants in survey research;

(b) appropriately informed about the survey's intentions and how their personal information and survey responses will be used and protected;

(c) sufficiently satisfied with their survey experience;

(d) willing to participate again in survey research.



15.9 FURTHER READING

1. Malhotra, N.K. and Dash, S. (2011). Marketing research: An Applied Orientation (6th edition), New Delhi: Pearson Education.



15.10 ANSWERS TO CHECK YOUR PROGRESS

Ans to Q.1. Research that involves human subjects or participants raises unique and complex ethical, legal, social and political issues. Research ethics is specifically interested in the analysis of ethical issues that are raised when people are involved as participants in research. There are three objectives in research ethics. The first and the broadest objective is to protect human participants. The second objective is to ensure that research is conducted in a way that serves the interests of individuals, groups and/or society as a whole. Finally the third objective is to examine the specific research activities and projects for their ethical soundness, looking at issues such as the management of risk, protection of confidentiality and the process of informed consent.

Ans to Q2. Client means any or organisation, department or division (including any belonging to the same organisation as the researcher) that requests, commissions or subscribes to all or any part of a market, social and organisational research project.

Relationships between a Survey Research Organization and Clients for whom the surveys are conducted should be of such a nature that they foster confidence and mutual respect. They must be characterized by honesty and confidentiality.

Ans to Q.3.

(a) Participants' **anonymity** must be strictly preserved.

(b) Researchers must take special care when carrying out research among children and young people and other **vulnerable** groups in the community.



15.11 MODEL QUESTIONS

Q.1: What is ethics and why it is important in business research?

Q.2: Write about clients' ethical code.

Q.3: Write about researchers' ethical code.

Q.4: Describe ethical codes related to respondents.

Q.5: Illustrate the responsibility of ethics in research.

******



REFERENCES (For All Units of the Course)

1. Malhotra, N.K. and Dash, S. (2011). Marketing research: An Applied Orientation (6th edition), New Delhi: Pearson Education.
2. Shields, Patricia; Rangarjan, Nandhini (2013). A Playbook for Research Methods: Integrating Conceptual Frameworks and Project Management. Stillwater, Oklahoma: New Forums Press. ISBN 1-58107-247-3.
3. Kakati R. P. and Choudhury S. (2013),Measuring Customer-Based Brand Equity Through Brand Building Blocks for Durables, The IUP Journal of Brand Management, Vol. X, No. 2, pp.24-42
4. Kothari, C.R. and Garg, G.(2014), 'Research Methodology - Methods and Techniques', New Age International Publishers
5. Malhotra. N.K.(2011), Marketing Research – An Applied Orientation, Pearson Education
6. Beri, G C (2009), Marketing Research, Fourth Edition, Tata McGrew Hill Publishing Company Limited, New Delhi
7. Bajpai N(2011), Business Research Methods, Pearson Education, India
8. Sachdeva J K (2011), Business Research Methodology, 2nd Edition , Himalaya Publishing House Pvt. Ltd.

Name of the Paper:
Course Code:

Centre for Internal Quality Assurance (CIQA)
Krishna Kanta Handiqui State Open University
City Office, Housefed Complex, Guwahati- 06

Learner's Feedback on Course

Dear Learner,

Regarding the course as mentioned above, we would like to know your opinions and comments so as to improve the quality of self learning materials in future. Please respond to the following statements by ticking the number you feel most reflect your opinion. After completion of the additional comments, please detach the page and send/mail the same to us at the address given below:

The Director, Centre for Internal Quality Assurance, KKHSOU, Housefed Complex, Dispur, Guwahati- 781006
(E-mail id: ciqa@kkhsou.in)

1. Approximately how many hours did you spend for studying the units in the course?
2. Please give your opinions (by ✓ mark) to the following items based on your reading of the block.

Sl. No.	Statements	Strongly Agree	Agree	Neither Agree nor Disagree	Disagree	Strongly Disagree
I.	The SLMs of the course fulfil the learning objectives set out in the beginning of each unit					
II.	The units of the SLM could satisfy my academic needs and expectations					
III.	The Unit writers have excellent knowledge about the course contents					
IV.	Language and contents of the units were presented at a level which I could readily understand					
V.	Ample opportunity for participation in the activities provided in the units					
VI.	Used enough illustrations (Diagrams, tables etc.) for conceptual clarity					
VII.	Quality of content is engaging, relevant, and up-to-date					
VIII.	The self-check questions are very helpful					
IX.	The Possible model questions and the answers to check my progress have benefited me a lot					

Additional Comments: *(Please feel free to provide your open comments)*

1. Which aspects of the SLM, according to you, worked well?

   ……

   ……

2. What sort of changes/improvements do you feel KKHSOU could implement to improve the overall quality of the SLM?

   ……

   ……

Thank you for taking the time to complete this form.