ignou
THE PEOPLE'S
UNIVERSITY

Indira Gandhi National Open University
School of Social Sciences

BPCC108
STATISTICAL METHODS FOR PSYCHOLOGICAL RESEARCH II

"शिक्षा मानव को बन्धनों से मुक्त करती है और आज के युग में तो यह लोकतंत्र की भावना का आधार भी है। जन्म तथा अन्य कारणों से उत्पन्न जाति एवं वर्गगत विषमताओं को दूर करते हुए मनुष्य को इन सबसे ऊपर उठाती है।"

– इन्दिरा गाँधी

"Education is a liberating force, and in our age it is also a democratising force, cutting across the barriers of caste and class, smoothing out inequalities imposed by birth and other circumstances."

- Indira Gandhi

BPCC-108

STATISTICAL METHODS FOR PSYCHOLOGICAL RESEARCH-II

School of Social Sciences
Indira Gandhi National Open University

EXPERT COMMITTEE

Prof. Swaraj Basu, Former Director, School of Social Sciences, IGNOU, Delhi Prof. Vimala Veeraraghavan, Former Emeritus Professor, Discipline of Psychology, SOSS, IGNOU, New Delhi Prof. Preetam Khandelwal, Professor, Faculty of Management studies, Delhi	Dr. Swati Patra, Faculty, Discipline of Psychology, School of Social Sciences, IGNOU, New Delhi Dr. Monika Misra, Faculty, Discipline of Psychology, School of Social Sciences, IGNOU, New Delhi Dr. Smita Gupta, Faculty, Discipline of Psychology, School of Social Sciences, IGNOU, New Delhi	Prof. Suhas Shetgovekar (Convenor) Faculty, Discipline of Psychology, School of Social Sciences, IGNOU, New Delhi

COURSE COORDINATOR

Prof. Suhas Shetgovekar
Faculty, Discipline of Psychology,School of Social Sciences, IGNOU, New Delhi

GENERAL EDITOR

Prof. Suhas Shetgovekar
Faculty, Discipline of Psychology,School of Social Sciences, IGNOU, New Delhi**COURSE**

COURSE PREPARATION TEAM

Block/ Unit	Title	Unit Writer	Editor (Content, format Language)
Block 1	**Inferential Statistics: An Introduction**		
Unit 1	Introduction to Inferential Statistics	Prof. Suhas Shetgovekar	Dr. Mohsin Uddin
Unit 2	Introduction to Parametric and Non-Parametric Statistics	Prof. Suhas Shetgovekar	Dr. Mohsin Uddin
Block 2	**Parametric Statistics**		
Unit 3	Test of Significance of Difference between Two Means	Dr. Vivek Belhekar, adapted by Prof. Suhas Shetgovekar	Prof. Suhas Shetgovekar
Unit 4	Test of Significance of Difference between More than Two Means	Prof. Suhas Shetgovekar	Dr. Mohsin Uddin
Block 3	**Non-parametric Statistics**		
Unit 5	Chi-square	Dr. Vivek Belhekar, adapted by Prof. Suhas Shetgovekar	Prof. Suhas Shetgovekar
Unit 6	Computation of Chi-square	Dr. Vivek Belhekar, adapted by Prof. Suhas Shetgovekar	Prof. Suhas Shetgovekar
Unit 7	Mann Whitney U test	Prof. Suhas Shetgovekar	Dr. Mohsin Uddin
Unit 8	Median Test	Prof. Suhas Shetgovekar	Dr. Mohsin Uddin
Unit 9	Kruskal-Wallis Analysis of Variance	Ms. Sonia Puar adapted by Prof. Suhas Shetgovekar	Prof. Suhas Shetgovekar
Block 4	**Statistical Package for Social Sciences**		
Unit 10	Introduction to Statistical Package for Social Sciences (SPSS)	Prof. Suhas Shetgovekar	Dr. Mohsin Uddin
Unit 11	Computation of Descriptive Statistics using SPSS	Prof. Suhas Shetgovekar	Dr. Mohsin Uddin
Unit 12	Computation of Parametric and Non-parametric Statistics using SPSS	Prof. Suhas Shetgovekar	Dr. Mohsin Uddin

PRINT PRODUCTION

Rajiv Girdhar
Assistant Registrar,
IGNOU, New Delhi

Hemant Kumar Parida
Section Officer,
IGNOU, New Delhi

December, 2021

ISBN: 978-93-5568-071-6

Printed and published on behalf of the Indira Gandhi National Open University, New Delhi, by Registrar, MPDD, IGNOU, New Delhi

Cover Page Design: M/s. Selene Khosla

Laser Typeset by: Tessa Media & Computers, C-206, Shaheen Bagh, Jiamia Nagar, New Delhi-25

Printed at : Aravali Printers & Publishers Pvt. Ltd., W-30, Okhla Industrial Area, Phase-II, New Delhi-110020

ignou
THE PEOPLE'S
UNIVERSITY

BPCC-108 STATISTICAL METHODS FOR PSYCHOLOGICAL RESEARCH-II:HOW TO PROCEED IN THE COURSE?

The course on Statistical Methods for Psychological Research- II (BPCC-108) is a core course offered in IVth Semester of BA Psychology Honours Programme, under Bachelors Degree Programme of IGNOU. The course is of 06 Credits, comprising Theory (04 Credits) and Tutorials (02 Credits). The course is divided into four blocks. The main objectives of this course are to introduce the learners to inferential statistics andto help learners acquire knowledge about parametric and non parametric statistics.

Each of these blocks represents a specific theme which is discussed in terms of units. The units are arranged in a logical sequence so as to cover the main aspects of each theme. Before proceeding to read the units, you are advised to go through instructions about how to read the course material. Given below is the explanation of the organization and sequencing of the unit.

Organization and Sequencing of a Unit

The following is the structure of each unit:

1.0 Objectives

1.1 Introduction

1.2 Section (Theme of the section)

1.2.1 Subsection of 1 .....

Check Your Progress I

1.3 Section (Theme of the section)

1.3.1 Subsection of 2 .....

Check Your Progress II

Let Us Sum Up

References

Key Words

Answers to Check Your Progress

Unit End Questions

As the scheme suggests, each unit is divided into sections for easy reading and better comprehension. The numbering and length of each section and subsection may vary from one unit to the other, depending upon the depth of information in each unit. Each section is indicated by **BOLD CAPITALS** and each sub-section by a **relatively smaller but bold typeface**. Divisions within the sub-sections are in **relatively smaller bold typeface** so as to make it easy for you to understand.

Let us now discuss each section of a unit.

Objectives

We begin each unit with the section Objectives. It tells you briefly about the objectives of the unit and what you will learn after you study the unit.

Introduction

The section Introduction will mainly focus on introducing the theme of the unit.

Illustration

There are several illustrations in each unit in the form of figures and diagrams. The main purpose of these illustrations is to make the study comprehensive and interesting.

Check Your Progress

We have given self-check exercises under the caption Check Your Progress at the end of main sections. You can provide your answers in the space give below each question/ exercise.

You will be tempted to have a glance of the main text as soon as you come across an exercise. But we do hope that you will resist this temptation and turn to the main text only after completing the answers.

You should read each unit and note the important points in the margin provided in the course material. This will help in your study. It will also help you to answer the self- check exercises and the assignment questions, as well as help in revising your course before appearing for Term End Examination (TEE).

Let Us Sum Up

Let Us Sum Up summarises the whole unit for the purpose of ready reference and recapitulation.

References

We have given a list of references at the end of each unit. This is a list of books and articles used by the course writers to prepare the units. This reflects that your course material is based on a wide spectrum of literature available on a particular theme, related to your course. This also informs you of the wide literature available in the particular area of study. If interested in widening your knowledge, you may look for the mentioned references. Each reference mentions the name of the author, year of publication, title of the book/article, name of publisher and place of publication.

Further readings help you to increase your level of understanding of a particular theme in each unit.

Key words

The key words at the end of the unit explain the basic ideas, technical terms and difficult words.

Unit End Questions

Besides Check Your Progress, we have given Unit End Questions in each unit. Practicing these questions will help you in answering assignments and Term End Examination. Though the pattern and style of questions asked may not be similar.

Audio and Video Aids

Some units have been selected for the audio and video programmes to supplement the printed material. These will help you to understand the units with greater clarity.

Apart from this, you may also access IGNOU's FM radio channel, Gyanvani (105.6 FM), which is available across many cities in India, for regular programmes, related to themes on Psychology. You can listen to the live discussions by faculty and experts on the topic of the day and interact with them through telephone, email, and through chat mode.

You may also watch Gyandarshan TV channel (free to air educational channel), for programmes related to topics on Psychology. The schedule of Gyanvani and Gyandarshan is displayed on www.ignou.ac.in. The radio and TV channels may also be accessed on Gyandhara, webcast facility for Gyanvani and Gyandarshan, provided by the University.

Tutorial

The course on Statistical Methods for Psychological Research- II will include tutorial, that is compulsory. It will be in form of activities and evaluated by your academic counsellor. You should carefully read the course material and apply the information while answering the tutorial. These activities are meant to develop your ability to relate your knowledge to day-to day life experiences. The tutorial will be made available to you along with the assignments and they have to be submitted at your study centre.

Assignment

You will receive a set of assignments for the whole programme. These are Tutor Marked Assignments, which are to be submitted to the respective Study Centre after completion. These assignments will be evaluated by academic counsellor(s) from your Study Centre. Ensure that you complete all your assignments because the grades that you get in each of these assignments are included in the final evaluation of your degree. Before answering the assignments, read all the units and additional material (if available).

Guidelines for tutorial and assignments

While working on the tutorial and assignments, kindly ensure the following points:

1) Clearly write your Enrollment number

2) Answer them in your handwriting and in your own words (do not copy the sentences from the course material or any other source).

3) Write clearly and neatly so that it is easy to read your answers

4) Leave margins on one side of your answer-sheets so that evaluator may write his/ her comments on your performance.

5) Organise your answers well based on the question asked.

6) You will submit the tutorial/ assignments at your Study Centre on or before the date mentioned as per the admission cycle. Kindly check the dates from www.ignou.ac.in or your Regional Centre website.

Term End Examination (TEE)

Consider the following points while answering TEE.

1) Questions need to be answered in **one's own words** and they need to be focused based on the questions asked.

2) Answer the questions keeping in mind the word limit.

3) Organise answers well based on the questions asked and also keep in mind any bifurcation given in the marks.

4) Ensure that you mention correct question numbers for respective answers.

Preparation of Course Material

The syllabus of course material BPCC-108 is designed by an Expert Committee (see page 2 of this course) and prepared by Course Preparation Team which comprises the author(s) of units, content editor(s), language editor, and the course coordinator. The expert committee selected the themes and sub-themes of the blocks and units, keeping in view the prescribed syllabi of UGC (CBCS model).The authors of units have provided their expertise in elaborating them in the form of the main text of each unit. The content editor has carefully examined the course contents and has made an attempt to make the material clear and comprehendible.

For any query or feedback related to the course, you may kindly contact the course coordinator at:

Prof. Suhas Shetgovekar
Room No. 121, Block-F,
School of Social Sciences IGNOU, New Delhi
Email: sshetgovekar@ignou.ac.in

COURSE CONTENT

BLOCK 1 INFERENTIAL STATISTICS: AN INTRODUCTION

Unit 1 **Introduction to Inferential Statistics**

Concept and Meaning of Inferential Statistics

Inferential Procedures

Procedure for testing hypothesis

Unit 2 **Introduction to Parametric and Non-Parametric Statistics**

Concept of Parametric and Non-parametric Statistics

Assumptions of Parametric and Non-parametric Statistics

Application and Uses of Parametric and Non-parametric Statistics

Difference between Parametric and Non-parametric tests

BLOCK 2 PARAMETRIC STATISTICS

Unit 3 **Test of Significance of Difference between Two Means**

Need and Importance of the Significance of the Difference between Means.

Fundamental Concepts in Determining the Significance of the Difference between Means

Methods to Test the Significance of Difference between the Means of Two Independent Groups t-test

Significance of the Difference Between two Correlated Means

Unit 4 **Test of Significance of Difference between more than Two Means**

Concept of Variance

Concept of Analysis of Variance (ANOVA)

Computation of One Way Analysis of Variance (ANOVA)

Factorial Design

Computation of Two Way Analysis of Variance (ANOVA)

BLOCK 3 NON-PARAMETRIC STATISTICS

Unit 5 **Chi-square**

What is Ch- square?

Assumptions for the Application of χ2Test

X^2 Distribution

Properties of Chi-square Distribution

Applications of Chi-square Test

Precautions about Using the Chi-square Test

Unit 6 **Computation of Chi-square**

Computation of Chi- square test

Chi-square test when table entries are small (Yates Correction)

BPCC-108 COURSE INTRODUCTION

Dear Learner,

In the first semester you studied Statistical Methods for Psychological Research- I. In the fourth semester you will now study Statistical Methods for Psychological Research- II. The main objective of this course is to introduce the learners to inferential statistics and o help learners acquire knowledge about parametric and non parametric statistics. The present course will also be helpful in the sixth semester of this programme when you take up the course on project (BPCE144).

To provide an overview of this course, the course structure is divided in to four blocks.

Block 1: The first block of this course is titled 'Inferential statistics: An Introduction'. This block includes two units, unit 1 and 2. Unit 1 is titled Introduction to Inferential Statistics and will cover **c**oncept and meaning of inferential statistics, inferential procedures including estimation and hypothesis testing. The procedures for testing hypothesis will also be discussed in this unit. The second unit is titled Introduction to parametric and non-parametric statistics and will focus on the **c**oncept of parametric and non-parametric statistics. The assumptions of parametric and non-parametric statistics will also be explained in the unit. Further, the application and uses of parametric and non-parametric statistics will also be discussed. The unit will also focus on the difference between parametric and non-parametric tests

Block 2: The second block of this course is titled Parametric Statistics and includes two units, Units 3 and 4. Unit 3 is titled Test of Significance of Difference between Two Means and will cover the need and importance of the significance of the difference between means. The fundamental concepts in determining the significance of the difference betweenmeans will also be covered in the unit. The unit will then explain the methods to test the significance of difference between the means of two Independent groupst-test and significance of the difference between two correlated means. Unit 4 is titled Test of Significance of Difference Between more than Two Means and will cover concept of variance and concept of Analysis of Variance (ANOVA). It will also discuss computation of One Way Analysis of Variance (ANOVA), factorial design and computation of Two Way Analysis of Variance (ANOVA)

Block 3: Block 3 is titled Non-parametric Statistics and includes 5 units, Uni5, 6,7,8 and 9. Unit 5 is titled Chi-square and will include understanding what is chi-square by focusing on its assumptions and distribution. Properties and application of chi-square will also be explained. Further precautions about using the chi-square Test will also be discussed. Unit 6 will deal with the computation of chi-square that will be explained with the help of steps and examples. Unit 7 is titled Mann Whitney U Test and besides explaining what Mann Whitney U test is it will describe the computation of Mann Whitney U test from small sample, Mann Whitney U test from moderately large sample and Mann Whitney U test from large sample. Unit 8 is titled Median Test and will explain the concept of median test besides computation of one sample

median test, median test for two independent samples and median test for more than two independent samples. The last unit of this block is Unit 9 that is titled Kruskal- Wallis Analysis of Variance. This unit will introduce Kruskal-Wallis ANOVA Test and then focus on computation of Kruskal- Wallis ANOVA, considerations for large sample and comparison of ANOVA and Kruskal-Wallis ANOVA test.

Block 4: The last block of this course is titled Statistical Package for Social Sciences and includes 3 units, Unit 10, 11 and 12. Unit 10 is titled Introduction to Statistical Package for Social Sciences and includes the statistical analysis using Excel and introduction to Statistical Package for Social Sciences (SPSS) including how SPSS file is created, menu bar of SPSS, data view, variable view and output view of SPSS and data entry and coding. Unit 11 will cover computation of descriptive statistics using SPSS. The unit will include computation of frequency, descriptive statistics, crosstabs and graphs in SPSS. Unit 12 is titled Computation of Parametric and Non-parametric Statistics using SPSS and the unit will cover computation of correlation, independent t test, one way ANOVA, Chi- square, Mann-Whitney U test and Kruskal- Wallis ANOVA using SPSS.

It is very important that you read all the units very carefully and try to understand the subject matter. Knowledge of statistics is important in Psychology as the same can be applied to data analyse data while carrying out quantitative research.

Some suggestions and tips to enhance your study of this course are as follows:

1) Ensure that your basic concepts as given in this course, are clear. If you don't understand the terms, read again. The first unit is a foundation to rest of the units, so ensure that you read and learn this unit first.

2) Have a clear idea about the parametric and non-parametric statistics as that will help you in making decisions about when to use which statistical techniques.

3) Understand the formulae and also the steps in computing various statistical techniques. These have been discussed with the help of examples. Practicing these steps along with the formulae will help you learn them in a better way.

4) Answering the exercises given in Check Your Progress (given in each unit) will help you practice and learn the subject matter and the method for computing various statistical techniques.

BLOCK 1

INFERENTIAL STATISTICS: AN INTRODUCTION

ignou
THE PEOPLE'S
UNIVERSITY

UNIT 1 INTRODUCTION TO INFERENTIAL STATISTICS*

Structure

1.0 OBJECTIVES

After reading this unit, you will be able to:

- discuss the concept and meaning of inferential statistics;
- describe inferential procedures; and
- explain the procedure for testing hypothesis.

1.1 INTRODUCTION

To refresh your memory with regard to what you learned in BPCC104: Statistical Methods for Psychological Research- I and to set a background for further discussion and explanation for the present course, that is, BPCC108: Statistical Methods for Psychological Research- II, let us focus on the following points.

Statistics can be described as a branch or sub-field of mathematics that mainly deals with the organisation as well as analysis and interpretation of a group of numbers (Aron, Aron and Coups, 2009).

* Prof. Suhas Shetgovekar, Faculty, Discipline of Psychology, SOSS, IGNOU, Delhi

1. Statistics as a subject area has a vast scope and application. It finds its application in fields like policy planning, management, education, marketing, agriculture, medicine and so on, though, one of its major application is in research.

2. Statistics can be categorised in to descriptive and inferential statistics. In BPCC 104, we discussed in detail about descriptive statistics (and also briefly touched upon inferential statistics) and its techniques.

3. In the present unit, we will mainly focus on inferential statistics and in subsequent units we will discuss various statistical techniques under inferential statistics.

BOX 1.1: Revisiting Descriptive Statistics

Descriptive statistics mainly comprises of description and organisation of the data. It can be termed as a technique that helps in summarisation of prominent characteristics of a distribution. Based on the properties of the sample, the descriptive statistics can be categorised in to the following (Mohanty and Misra, 2016, page 7):

- **Statistics of location:** Covers techniques like measures of central tendency including mean, median and mode, frequency distribution, percentiles and so on.
- **Statistics of dispersion:** Covers techniques related to measures of dispersion including quartile deviation, standard deviation, range, average deviation and variance.
- **Statistics of correlation:** Includes coefficients of correlation like Pearson's product moment correlation, Spearman's rank order correlation and Kendall's rank correlation. Correlation mainly helps us understand the relationship between variables.

Refer to Box 1.1, that gives a brief description about descriptive statistics.

As you may recall, in BPCC104: Statistics for Psychological Research- I, we discussed about the techniques mentioned in the box 1.1. We learned about measures of central tendency including mean, median and mode, frequency distribution, percentiles. We also focused on measures of dispersion including quartile deviation, standard deviation, range, average deviation and variance. Correlation was also discussed in the course.

These techniques are relevant mainly in univariate analysis of data, that is, when there is one variable. But when we want to carry out bivariate analysis (where there are two variables) or when we have multiple independent variables and dependent variables, where we want to study cause and effect relationship and so on, we could use inferential statistics.

In BPCC104: Statistics for Psychological Research- I, we briefly explained inferential statistics. Let us once again discuss the same in this unit as this unit forms the foundation to the units that we will subsequently discuss in this course.

1.2 CONCEPT AND MEANING OF INFERENTIAL STATISTICS

Let us start our discussion with some examples.

Example 1: A Researcher wanted to study whether significant difference exists in the organisational citizenship behavior of junior and senior manager in a Multinational Company (MNC). For this purpose the researcher selected a sample of 100 each (that is, 100 junior managers and 100 senior managers) from an MNC. The selected sample was then administered the Organisational Citizenship Scale and data was obtained.

Example 2: A clinical psychologist wanted to study the effectiveness of a new psychotherapy (Therapy A) on patients diagnosed with depression. The clinical psychologist therefore administered the Becks' Depression Inventory on a group (N= 50) of patients diagnosed with depression, which was followed by six months intervention based on Therapy A and then after six months the same patients were again administered Becks' Depression Inventory to study the effect of the Therapy A. In this research the Pretest-Posttest design was used.

Example 3: Yet another researcher carried out a study on emotional maturity of early, middle and late adolescents. The researcher collected data from the three groups (200 from each group) with the help of Emotional Maturity Scale.

For the above examples, different statistical techniques, that necessarily fall under Inferential statistics can be used. For instance, with regard to the first example, the researcher could use independent t test, as there are two groups and significant difference between the two groups with regard to organisational citizenship behaviour needs to be studied. Though if the assumptions of parametric tests are not fulfilled then Mann- Whitney U test can also be used.

With regard to the example 2, the paired t test can be used if the assumptions of parametric statistics are met. This is so because the significant difference between a same group, pretest and posttest, is studied.

With regard to the last example (that is, example 3), to carry out data analysis One Way Analysis of Variance (this is a parametric test) or Kruskal- Wallis One Way ANOVA (this is non-parametric test) can be used.

Above we mentioned about parametric and non-parametric tests. the same will be discussed in detail by us in unit 4 of this course.

Let us now focus on what is inferential statistics.

In inferential statistics, inferences are drawn about the population based on a representative sample. After the data is collected, it is organised and summarised and once this is done, inferential statistics can be carried out in order to analyse the data and draw conclusions and make inferences. Thus, with the help of inferential statistics, inferences can be drawn about the population of the study based on the characteristics of the sample on whom the study was carried out (Salkind, 2014).

As stated by Veeraraghavan and Shetgovekar (2016, page 5) "Inferential statistics refers to the mathematical methods based on probability theory and helps in reasoning and inferring the characteristic features of the sample drawn from the larger population". Inferential statistics can also be effectively used to make estimations and predictions.

As stated by Aron, Aron and Coups (2013, page 2), inferential statistics is employed by psychologists in order to make inferences and draw conclusions based on certain data. Further, inferential statistics though computed based on descriptive statistics of a given sample, it goes beyond the sample and it helps in generalisation of inferences to the whole population (Mohanty and Mishra, 2016, page 8). King and Minium (2008, page, 4) described that "the purpose of inferential statistics is to draw conclusions about the conditions that exist in a population from study of a sample". The process of inferential statistics can be explained with the help of the following flowchart (refer to figure 1.1).

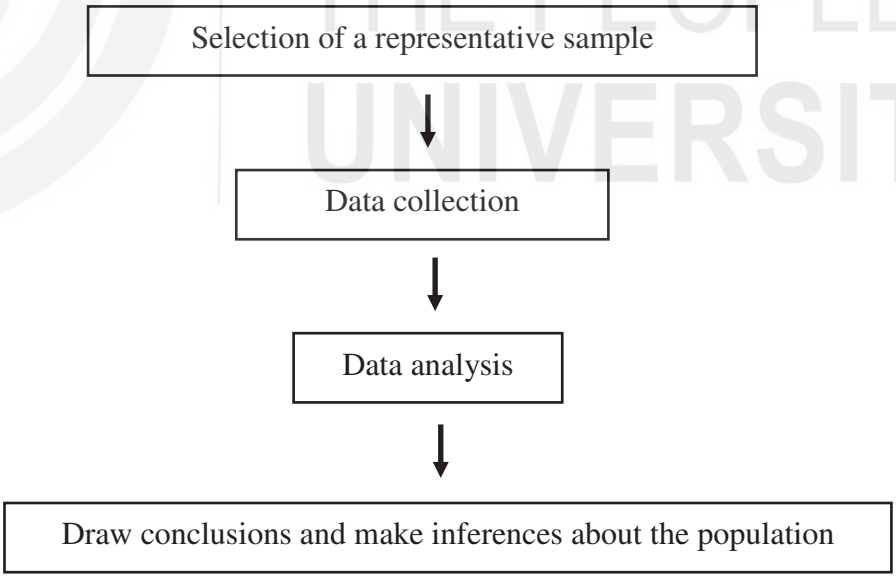

Fig 1.1: Process of Inferential Statistics

Thus, for example, if we want to study the environmental attitude of adolescents in Mumbai, we will take a representative sample (N =500) from the population of adolescents in Mumbai. We will then administer a standardised psychological test that measures environmental attitude to the selected adolescents. Once the data is collected, it is organised and then

analysed using inferential statistics and then conclusions are drawn and inferences are made for the population of adolescents in Mumbai.

Check Your Progress I

1) What is inferential statistics?

……

……

……

……

……

1.3 INFERENTIAL PROCEDURES

There are two types of procedures under inferential statistics, namely estimation and hypothesis testing. These two are discussed in detail as follows:

1.3.1 Estimation

Estimating probability of a phenomenon is referred to as estimation (Veeraraghavan and Shetgovekar, 2016). As we know from the explanation of inferential statistics, that inferences are drawn from sample that is representative of a population and these inferences can then be generalised to the whole population. In these inferences, the researcher will make an estimation that needs to be close to the actual or true population value.

There are two types of estimation: point estimation and interval estimation.

Point estimation: This is a type of estimation in which the value is a single point. For example, the estimation for sample mean is made as 46.8 that is expected to be equal to the population mean. Point estimate comprises of sample mean and sample proportion. The population mean is 'μ', the sample mean will be '$\bar{x}$'. In similar manner, if the population proportion is '*P'* then sample proportion will be '*p'*.

Interval estimation: An interval estimate is an interval or two numbers within which the population parameter could lie. Thus, for population mean 'μ', the interval estimate will be a<x< b. The interval estimate is greater than 'a' but lesser than 'b'. For example, an interval estimate could be 45- 47 within which it is expected that the population mean will lie. As the researcher has an interval, he/ she is thus able to trust that the estimate is close to the population value with 95% or 99% level of confidence. Interval estimate comprises of confidence interval for mean and confidence interval for proportions.

While estimations are made there could be fluctuations and these could be due to varied reasons including chance factors and sampling error.

The inferences drawn by the researcher needs to be free of any chance factors. For example, a researcher is studying if there exists significant difference in job satisfaction of government and private bank employees. After carrying out data collection and data analysis, he/she obtains results that such a difference does exist, then such results should not be an outcome of chance factors. If the difference falls within the range ± 1.96, then the significant difference can be said to be real and not due to chance factors.

Fluctuations can also be as a result of sampling error that occur when the sample selected by the researcher is not representative of the population being studied. A sample that is not representative of the population will not possess the same characteristics as the population and thus the results obtained from such a sample cannot be used to draw inferences for the population. Sampling errors can be avoided by being careful while selecting a sample and also by having a larger sample.

1.3.2 Hypothesis Testing

One of the significant aspects of research is hypothesis testing. Hypothesis is a tentative statement that is investigated during the process of research. Hypothesis pertains to certain phenomenon and is based on a theory. Data is then collected by the researcher in order to validate the hypothesis. Thus, the hypothesis is rejected or accepted by a researcher based on the results obtained. Hypothesis testing is referred to as "a process involving statistical decision making with regard to population value that is based on the sample value" (Veeraraghavan and Shetgovekar, 2016, page 9). Though, while drawing inferences one needs to ensure that hypothesis is not incorrectly accepted or rejected due to results that are influenced by chance factor or confounding factors.

1.3.2.1 Types of Hypothesis

There are two types of hypothesis:

Null hypothesis: Once a researcher finalises his/her research topic and variables after reviewing literature, he/ she will have to formulate hypothesis based on the problem statement. In a research attempts are made by the researcher to reject the null hypothesis(es). Null hypothesis is also referred to as hypothesis of 'no difference' as null hypothesis implies that there is no difference between two groups. It is denoted by H_o. For example, if a researcher wants to study if there is significant difference in job satisfaction of government and private bank employees, he/ she will formulate a null hypothesis as 'No significant difference exists in job satisfaction of public and private bank employees'. And then based on the results obtained the hypothesis will be rejected or accepted.

Alternative hypothesis (also called as research hypothesis)***:*** Alternative hypothesis can be termed as a counter proposition to the null hypothesis (Veeraraghavan and Shetgovekar, 2016). It is formulated by a researcher

based on a theory. The hypothesis states that there is a significant difference between two groups. It is denoted by H_a. For example, the researcher can formulate the alternative hypothesis as 'There will be a significant difference in job satisfaction of public and private bank employees'. This is also a neutral alternative hypothesis that does not provide any direction and thus is termed as nondirectional hypothesis. Alternative hypothesis can also be directional. For example, 'Public bank employee have higher (or lower) job satisfaction when compared with private bank employees'. Such hypotheses are formulated based on the review of literature.

Difference between null and alternative hypothesis: The difference between null and alternative hypothesis (Salkind, 2014, page 139) are discussed as follows:

1) The very first difference is that the alternative hypothesis is about inequality, where it is stated that a relationship exists between the given two variables. Whereas, null hypothesis is about equality where it is stated that there is no relationship between the given two variables.

2) Further, the alternative hypothesis is regarding the sample and the null hypothesis is regarding the population.

3) The third difference is that the null hypothesis is to be tested indirectly and the alternative hypothesis is to be tested directly. This is because we make inferences about the population based on the sample.

4) The fourth difference is quite interesting, as the alternative hypothesis are usually written using Roman symbols, whereas null hypothesis are written using Greek symbols.

   For instance

   The alternative hypothesis is $H_1: \bar{x}_1 < \bar{x}_2$

   The null hypothesis is $H_O: \mu_1 = \mu_2$

   If you recall, what we have learned in BPCC104, μ('mu') is a symbol for parameter mean, whereas, $\bar{x}$ is a symbol for sample mean.

5) The alternative hypothesis is an explicit hypothesis, whereas, the null hypothesis is an implied hypothesis, mainly because it cannot be directly tested.

Characteristics of a good hypothesis:Let us now discuss the characteristics of a good hypothesis

1. The statement of hypothesis is not stated as a question but is in a declarative form.
2. It states a relationship that is expected between the given variables.
3. The theory or literature on the basis of which the hypothesis is formulated is reflected in that hypothesis.

4. The hypothesis needs to be clear, to the point as well as brief.
5. It needs to be possible to test the hypothesis.

1.3.2.2 Level of Significance

The term significance in the context of statistics implies 'probably true' that indicates that results are free from chance factor at the specified level of significance. This level of significance could be 0.05 level or 0.01 level. For example, if there exists a significant difference in job satisfaction of government and private bank employees then the researcher could either be 95% (0.05 level of significance, $P<0.05$) confident or 99% (0.01 level of significance, $P<0.01$) confident about the results obtained. This is because when carrying out research with human participants, 100% accuracy cannot be achieved. Thus, there could be 5% or 1% chance that the results are due to chance or confounding factors. Whether the null hypothesis is accepted or rejected will depend on whether the statistical value obtained after data analysis is more or less than the table value (tables for various statistical techniques are provided at the end of any book on Statistics) specified at 0.05 or 0.01 level of significance. If the obtained value is higher than the table value, then the null hypothesis is rejected, and if the obtained value is less than the table value than the null hypothesis is accepted.

1.3.2.3 One-tailed and Two- tailed Tests

Further, any hypothesis can also be one tailed or two tailed. It is termed as one- tailed when certain direction is given to the hypothesis or the hypothesis is directional. For example, if the researcher wants to study if gender difference exists with regard to emotional intelligence, a one-tailed hypothesis would be "Females have higher emotional intelligence than males" or "Males have higher emotional intelligence than females". Refer to figure no. 1.2 that provides the figure for both 0.05 and 0.01 levels of significance. In one-tailed test, in order to reject a null hypothesis, the score needs to fall in the upper tail, that is in the top 5% of the distribution. A one-tailed test can be tested in either of the direction.

In case of two- tailed hypothesis, the hypothesis is nondirectional and will be stated as "Gender difference will exist with regard to emotional intelligence". In case of two-tailed test, a null hypothesis can be rejected when the score falls in the either of the top 2.5% of the distribution. Thus, the level of significance is maintained at 0.05 level (refer to figure 1.2). As the level of significance is taken as 0.05 level, it can be taken at 0.01 level as well, in which case the prediction will be based on the lower tail.

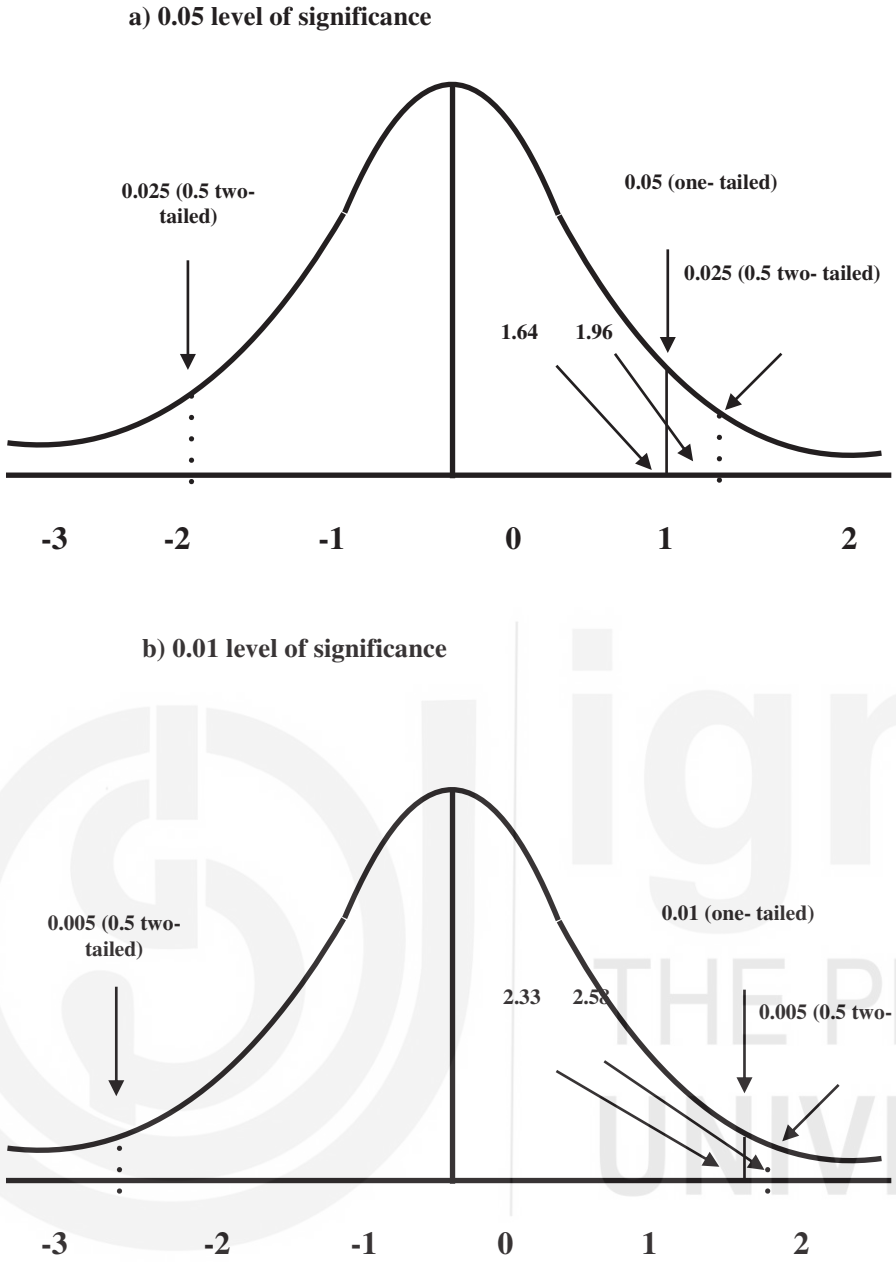

Fig. 1.2: One Tailed and two Tailed Tests

The z scores for one tailed test would be -1.64 or 1.64 at 0.01 level and -2.33 or 2.33 at 0.05 level. The z scores for two tailed test would be -1.96 or 1.96 at 0.01 level and -2.58 or 2.58 at 0.01 level (as you can see in figure 1.2).

1.3.2.3 Errors in Hypothesis Testing

If we look at the process of hypothesis testing, it is based on inferences that we draw for the population based on a sample. Thus, we can say that we draw conclusions based on limited information and as a result there is a scope of making errors. This can despite of ensuring that the sample is representative of the population. Thus, in this context we can discuss about

Type I and Type II errors in hypothesis testing. Refer to the table 1.1 for clarity in this regard.

Table 1.1: Type I and Type II Errors

	Null Hypothesis is true	**Null Hypothesis is false**
Null Hypothesis is Rejected	Type I error	Decision is correct
Null Hypothesis is Accepted	Decision is correct	Type II error

As can be seen in table 1.1 a researcher will be making a correct decision when a false null hypothesis is rejected and when a null hypothesis that is true is accepted. However, it may so happen that a null hypothesis is rejected even when it is true and this is termed as type I error. On the other hand when a false null hypothesis is accepted then it is termed as type II error.

Let us now discuss the correct decisions as well as the types of errors.

Correct decision: A researcher will be making a correct decision when a false null hypothesis is rejected and when a null hypothesis that is true is accepted.

Type I error: In this regard a decision is taken by the researcher to reject a null hypothesis that is actually true. The probability with regard to whether type I error is committed is related with the level of significance, the alpha level (α). The chances of a true null hypothesis getting rejected is high when the alpha level is high. To put this in other words, if the level of significance that we use to reject the null hypothesis is high, more the likelihood that type I error is being committed.

Type II error: In this regard a decision is taken by the researcher to accept a null hypothesis that is actually false. Type II error is represented in terms of beta (β) and this error is committed when on the basis of test of significance, null hypothesis is accepted even when it is false.

In a way, whether the Type I error is committed or not will depend on the level of significance that is taken by the researcher, which can be specified before the research process starts. Therefore, the likelihood of making type I error would be lower if a stringent alpha level is used. Though, when, the alpha level is stringent, the likelihood of making Type II error could increase (Mohanty and Misra, 2016). Thus, it can be said that the relationship between making type I and type II error is inverse in nature. Though, if we want to ensure that both the errors are avoided, then the sample size can be increased.

1.3.2.4 Power of a Test

Let us now come to the next significant term, that is power of the test. Power of a test can be related with type II error. Power of a test can be explained as a measure of how sensitive the experiment or research is in detecting the real effect that the independent variable has on the dependent variable (Mohanty and Misra, 2016). Thus, it mainly involves rejecting the null hypothesis, as when null hypothesis is rejected it can be said that the independent variable is having an effect on the dependent variable. Thus, the power of a test can also be described in terms of probability related to rejection of null hypothesis when it is false. Thus,

Power = 1- probability of type II error

Putting this differently,

Power = 1- β or β = 1- Power

Power of a test can also be described as probability of rejecting a null hypothesis that is actually false. Thus,

Power = P (rejection of null hypothesis)= 1- β

Where, P stands for probability.

The values of probability ranges from 0.00 to 1.00 (Mohanty and Misra, 2016). The more is this value the more the experiment or research will be sensitive in detecting the real effect and the less is the value, the less is the experiment or research will be sensitive in detecting the real effect. Further, as we increase the sample size, the Power of a test will also increase.

Check Your Progress II

1) What is power of a test?

……

……

……

……

……

1.4 PROCEDURE FOR TESTING HYPOTHESIS

We understood some of the important aspects of hypothesis testing. Let us now look at the procedure for testing the hypothesis in previous section.

Step 1: Null hypothesis(es)/ alternative hypothesis(es) are specified.

In the earlier sub section, we discussed about the null hypothesis (es) and alternative hypotheses. In the very first step of hypothesis testing procedure,

null hypothesis(es)/ alternative hypothesis(es) are specified. This will mainly depend on the objective,(s) and nature of the study as well as the existing review of literature.

Step 2: A level of significance is selected.

We discussed about the level of significance in the previous section. In step two, the level of significance is selected. As we have mentioned under the type I error, the level of significance or alpha level is to be stated before you start with the research process. Thus, in the second step we will state the level of significance for our research.

Step 3: Based on the parameter specified in the null hypothesis (es), a suitable statistic is calculated.

A sample is taken by the researcher and data is collected. Statistic that is thus obtained from the sample (representative) is then used in order to make an estimation about the population parameters.

Step 4: **Decision is made whether to accept or reject the null hypothesis (es).**

In this regard, based on the P value or probability level selected under step 2, decision is taken by the researcher whether to accept or reject the null hypothesis.

Check Your Progress III

1) List the steps in hypothesis testing.

……

……

……

……

……

1.5 LET US SUM UP

To summarise, in inferential statistics, inferences are drawn about the population based on a representative sample. After the data is collected, it is organised and summarised and once this is done, inferential statistics can be carried out in order to analyse the data and draw conclusions and make inferences. The types of procedures under inferential statistics, namely estimation and hypothesis testing were also discussed in the unit. Point estimation is a type of estimation in which the value is a single point. Whereas an interval estimate is an interval or two numbers within which the population parameter could lie. Hypothesis testing is referred to as a process involving statistical decision making with regard to population value that is based on the sample value. The unit also covered the types of hypotheses,

level of significance, one tail and two tail test, rrors in hypothesis testing and power of a test. Lastly the procedure for hypothesis testing was also highlighted.

1.6 REFERENCES

Aaron, A., Aaron, E. N., & Coups, E. J. (2013). Statistics for psychology. New Delhi: Dorling Kindersley (India) Pvt. Ltd.

Best, J. W and Kahn, J. V. (1999). Research in Education. New Delhi: Prentice Hall of India Pvt. Ltd. for information on research designs.

Kerlinger, Fred, N. (1995). Foundations of Behavioural Research. Bangalore: Prism Books Pvt. Ltd. for information on research, research designs, types of research and methods of data collection.

King, M; Minium, W. (2008). Statistical Reasoning in the Behavioural Sciences. Delhi: John Wiley and Sons, Ltd.

Mangal, S. K. (2002). Statistics in psychology and Education. new Delhi: Phi Learning Private Limited.

Minium, E. W., King, B. M., & Bear, G. (2001). Statistical reasoning in psychology and education. Singapore: John-Wiley.

Mohanty, B and Misra, S. (2016). Statistics for Behavioural and Social Sciences. Delhi: Sage.

Salkind, N. J. (2014). Statistics for People who (think that) hate statistics. New Delhi: Sage.

Veeraraghavan, V and Shetgovekar, S. (2016). Textbook of Parametric and Non-parametric Statistics. Delhi: Sage.

1.7 KEY WORDS

Alternative hypothesis*:* Alternative hypothesis can be termed as a counter proposition to the null hypothesis.

Inferential statistics: In inferential statistics, inferences are drawn about the population based on a representative sample. With the help of inferential statistics, inferences can be drawn about the population of the study based on the characteristics of the sample on whom the study was carried out (Salkind, 2014).

Interval estimation: An interval estimate is an interval or two numbers within which the population parameter could lie.

Null hypothesis: Null hypothesis is also referred to as hypothesis of 'no difference' as null hypothesis implies that there is no difference between two groups.

Point estimation: This is a type of estimation in which the value is a single point.

1.8 ANSWERS TO CHECK YOUR PROGRESS

Check Your Progress I

1) What is inferential statistics?

   As stated by Veeraraghavan and Shetgovekar (2016, page 5) "Inferential statistics refers to the mathematical methods based on probability theory and helps in reasoning and inferring the characteristic features of the sample drawn from the larger population".

Check Your Progress II

1) What is power of a test?

   Power of a test can be explained as a measure of how sensitive the experiment on research is in detecting the test effect that the independent variable has on the dependent variable.

Check Your Progress III

1) List the steps hypothesis testing.

   The steps in testing the hypothesis are as follows:

   Step 1: Null hypothesis(es)/ alternative hypothesis(es) are specified.

   Step 2: A level of significance is selected.

   Step 3: Based on the parameter specified in the null hypothesis (es), a suitable statistic is calculated.

   Step 4: Decision is made whether to accept or reject the null hypothesis (es).

1.9 UNIT END QUESTIONS

1) Discuss the concept and meaning of inferential statistics.
2) Explain the types of hypothesis and level of significance.
3) Discuss one-tailed and two-tailed test with the help of a diagram.
4) Explain the types of errors in hypothesis testing with suitable examples.
5) Discuss the procedure for hypothesis testing.

UNIT 2 INTRODUCTION TO PARAMETRIC AND NON-PARAMETRIC STATISTICS*

Structure

2.0 OBJECTIVES

After reading this unit, you will be able to,

- explain the concept of parametric and non-parametric statistics;
- describe the assumptions of parametric and non-parametric statistics;
- discuss the application and uses of parametric and non-parametric statistics; and
- differentiate between parametric and non-parametric tests.

2.1 INTRODUCTION

Let us us go to a classroom scene where a teacher is interacting with her learners to learn about parametric and Non-parametric statistics.

Prof. Anuradha, an academic counsellor at one of the IGNOU study centres, was discussing about psychological statistics with her students. The learners seemed to be confused about what was parametric and non-parametric statistics. Let us look at the doubts raised by the student and the explanations provided by Prof. Anuradha.

* Prof. Suhas Shetgovekar, Faculty, Discipline of Psychology, SOSS, IGNOU, Delhi

Supriya (Learner): *Maam what would be a basic difference between parametric and non-parametric statistics?*

Prof. Anuradha: *Well, Supriya, basically, in parametric statistics, there are certain assumptions that need to be met. Whereas, there is no such condition in Non-parametric statistics.*

Rashid (Learner): *But Maam, what are these assumptions?*

Prof. Anuradha: *Yes, that is a relevant questions. There are various assumptions, like the assumed distribution is normal in parametric statistics.*

Divya (Learner): *So Maam, in non-parametric statistics, there is no such assumptions.*

Prof. Anuradha: *Absolutely right Divya. In non-parametric statistics, the assumed distribution may not be normal. It can be any distribution.*

Rashid (Learner): *So Maam, that means the distribution can be skewed or have kurtosis, as we had studied in BPCC104.*

Prof. Anuradha: *Absolutely right Rashid.*

Meanwhile Sara raised her hand and was eager to say something. So Prof. Anuradha asked Sara to speak.

Sara (Learner): *Maam, another assumption is that in parametric statistics, the variance is homogeneous.Whereas, the variance could be heterogeneous or no assumption is made with regard to the variance.*

Prof. Anuradha was happy to hear this and praised Sara.

Prof. Anuradha: *Now who will tell me about scales of measurement that we studies in BPCC104?*

Salil (Learner): *But Maam, What do we have to do about scales of measurement here?*

Prof. Anuradha: *Well Salil, when we discuss about parametric and non-parametric statistics, we need to revisit scales of measurement. So who will tell me the scales of measurement. And please give an example as well so that I know that you have not forgotten what we learned under BPCC104.*

Sahil (Learner) raised his hand and said: *Maam, nominal scale can be used to measure variables that are qualitative as well as exclusive in nature. For example, the Jersey numbers that are assigned to football players.*

Shirin (Learner): *Maam, ordinal scale involves ranks, that is, the data can be assigned ranks based on whether they are less or more, low or high, bad or good and so on. For example, ranks obtained by us in the surprise class test that you gave us last week.*

Amarjeet (Learner): *Maam interval scale are similar to the ordinal scale as the categories can be ranked and are exclusive as well, but the degree of difference between two participants is same. Interval scales are most commonly used to measure psychological variables.*

John (Learner): *And Maam example of an interval scale would be scores obtained by individuals on an emotional intelligence scale.*

Sheetal (Learner): *Maam, ratio scale has all the properties of all the scales, nominal, ordinal and interval. It also has an absolute zero, that indicates presence or absence of certain property or characteristics. For example, errors by Individuals on a memory test.*

Prof. Anuradha looked pleased and said: *Thats very good, looks like you have studied well. So let us now go to other assumptions. The next assumption is that the observations need to be independent of each other. Whereas, in non-parametric statistics, there is no assumption with regard to the independence of the observations.*

To understand more about parametric statistics, mean is the measure of central tendency that is used here.

Ventakatesh (Learner): *So Maam, in non-parametric statistics we don't use mean as a measure of central tendency?*

Prof. Anuradha: *In non-parametric statistics, median is the measure of central tendency that is used. Further, outliers can have an influence in parametric statistics, Whereas in non-parametric statistics, the influence of outliers is comparatively less.*

Sheetal (Learner): *Also Maam, I feel, non-parametric statistical techniques are easier to calculate.*

Prof. Anuradha said with a smile: *That's right Sheetal. So I hope we have now some clear idea about what is parametric and Non-parametric statistics.*

The learners nodded their head in unison.

With the above conversation, I hope you all also must have developed an idea about what the parametric and non-parametric statistics are.

In BPCC104, we had discussed that statistics can be categorised in different ways. It can be categorised into two main branches, descriptive statistics and inferential statistics and we discussed in detail the concept of descriptive statistics and its techniques in BPCC104. Further, statistics can also be categorised as parametric and Non-parametric statistics. Though we briefly touched upon this in BPCC104, in the present unit we will further focus on what are parametric and non parametric statistics.

In BPCC104, we studied about the Pearson's product moment correlation and Spearman's rank order correlation. Here, Pearson's product moment

correlation was parametric statistical technique and Spearman's rank order correlation was non-parametric statistical technique.

You must have heard about certain statistical techniques like t test, Analysis of Variance (ANOVA), Mann Whitney U test, Kruskal- Wallis test and so on. t test and Mann- Whitney, for example, are used to study significant difference between two sample means. t test is parametric statistical technique and Mann Whitney U is non-parametric statistical technique. Similarly, ANOVA is parametric statistical technique and Kruskal-Wallis test is non-parametric statistical technique.

In the next few pages of this unit, we will discuss the above points and learn more about parametric and non-parametric statistics.

2.2 CONCEPT OF PARAMETRIC AND NON-PARAMETRIC STATISTICS

Let us now try to understand the concepts of parametric and non-parametric statistics

2.2.1 Parametric Statistics

To start with parametric statistics, statistical techniques like Pearson's product moment correlation, t test and ANOVA can be termed as parametric statistical techniques, As such these mainly involve data that is in terms of absolute number or in terms of values. They are as such not in terms of ranks. In BPCC104, we discussed about the term parameter. A parameter can be termed as a value that provides information about the population that is investigated in the research. It can be described as "a measure of the population and refers to the indices of a central value, dispersion, correlation and so on of all the individuals of the population" (Mohanty and Misra, 2016, page 3). Thus it can be said that the parametric statistics mainly depends on the characteristics of a population (Mohanty and Misra, 2016).

Salkind (2014, page 468) described parametric statistics as "statistics used for the inference from a sample to a population that assume the variances of each group are similar and that the sample is large enough to represent the population". The definitions highlights the homogeneity of the variance as well as also focuses on the sample being representative of the population.

Mohanty and Misra (2016, page 675) defined parametric statistical tests as "a test whose model specifies certain conditions about the parameters of the population from which the research sample was drawn".

2.2.2 Non-parametric Statistics

With reference to non-parametric statistics, statistical techniques like Spearman's rank order correlation, chi- square, Mann Whitney U test can be termed as non-parametric statistical techniques. Non-parametric statistical

techniques are distribution free test. In other words, these tests are not based on data that is normally distributed or any such assumption (as is the case with parametric statistical techniques). Non-parametric statistics can also be described as tests that do not involve testing of hypotheses related to population parameters (King and Minium, 2013).

Salkind (2014, page 46) described non-parametric statistics as "distribution-free statistics that do not require the same assumptions as do parametric statistics". The model of the non-parametric statistical tests do not state any conditions about the population parameters from which the sample for research has been taken (Mohanty and Misra, 2016). Thus, in the case of non-parametric statistics, the methods are distribution free, the data is in terms of ranks of the observations and there is as such no assumptions that are provided with reference to the structure of the model.

Let us us now see what are some of the parametric statistical techniques and their Non-parametric counter parts with the help of table 2.1.

Table 2.1: Parametric and Non-parametric Statistics

Category	**Parametric Statistical Techniques**	**Non-parametric Statistical Techniques**
Correlation	Pearson's Product Moment Coefficient of Correlation (r)	Spearman's Rank Coefficient Correlation (Rho), Kendall's Tau
Two groups, independent measures	Independent t-test	Mann-Whitney 'U' test
More than two groups, independent measures,	One-way ANOVA	Kruskal-Wallis one-way ANOVA
Two groups, repeated measures	Paired t-test	Wilcoxon matched pair signed rank test
More than two groups, repeated measures,	One-way repeated measures ANOVA	Friedman's test

Check Your Progress I

1) Define parametric statistics.

……

……

……

2.3 ASSUMPTIONS OF PARAMETRIC AND NON-PARAMETRIC STATISTICS

As we have discussed the concepts of parametric and non-parametric statistics, Let us us try to understand the assumptions of parametric and non-parametric statistics.

2.3.1 Assumptions of Parametric Statistics

The following are the assumptions of parametric statistics:

1. **The population from which the research sample is taken is to be normally distributed:** This is also referred to as assumption of normality (Mangal, 2010 page 276). In BPCC104, we discussed about normal distribution. A normal distribution is a continuous probability distribution for a variable. It is also known as Gaussian/ Gauss or LAPlace – Gauss distribution. The normal distribution is determined by two parameters, mean and variance. When the normal distribution is represented in form of a graph, it is known as normal probability distribution curve or simply normal curve. A normal curve is a bell shaped curve, bilaterally symmetrical and is continuous frequency distribution curve. Such a curve is formed as a result of plotting frequencies of scores of a continuous variable in a large sample. The curve is known as normal probability distribution curve because its y ordinates provides relative frequencies or the probabilities instead of the observed frequencies. A continuous random variable can be said to be normally distributed if the histogram of its relative frequency has shape of a normal curve.

2. **The variables are measured in terms of interval or ratio scale:** As you have studied in BPCC104, there are four scales of measurement nominal, ordinal, interval and ratio. In case of parametric statistics, the variables are measured in terms of the interval and ratio scale.

   Interval scale are similar to the ordinal scale as the categories can be ranked and are exclusive as well, but the degree of difference between two participants is same. Ratio scale has all the properties of all the scales, nominal, ordinal and interval. Further, it also has an absolute zero, that indicates presence or absence of certain property or characteristics.

3. **The observations need to be independent of each other:** If any case is included or deleted from the sample, this should not have an effect on the results of the research. Thus, is a certain case is selected from the population and the same is included in the sample should not have an influence the chance of inclusion of any other case and a score that is given to certain case should not be biased by a score assigned to another case (Mohanty and Misra, 2016). This point mainly focuses on the

random selection of sample from the population and an unbiased assignment of scores to the observations.

4. **The population from which the research sample is taken needs to display a variance that is homogeneous:** It is important that population from which the research sample is taken displays variance that is homogeneous or display variance that is same or in certain cases have a ratio of variance that is known. To understand the term variance, it is a measure of the dispersion of a set of data points around their mean value. It is a mathematical expectation of the average squared deviations from themean. The variance (s^2) or mean square (MS) is the arithmetic mean of the squared deviations of individual scores from their means. In other words, it is the mean of the squared deviation of scores.Variance is expressed as $V = SD^2$.

This assumption is significant when we compute independent t test and ANOVA as they use t and F statistics and these are quite strong even if some of the assumptions are not met, but only when the size of the groups are equal. Equal group size can be explained in terms of a ratio, that needs to be less than 1:5 between the smallest and the largest group. Thus if the size of the groups are not equal, then the assumption of homogeneity is not met.

When the group sizes that are small are associated with large sample variances, there will be a bias and an underestimation of level of significance which will in turn lead to incorrect rejection of null hypothesis. Whereas, when the group sizes that are large are associated with large variances, the level of significance is overestimated, and as a result the Power of a test can decrease.

In this context, we can also discuss about homoscedasticity of distribution. Homoscedasticity denotes the variance between columns and rows display equality. This mainly denotes that the different samples that are taken from different populations have variance that is same. This is mainly used in the context of linear regression where the assumption is made about the data that it is homoscedastic.

It is important to focus on homoscedasticity because its absence can make the coefficient estimates inaccurate or decrease its precision. Thus, the coefficient estimates may not be close to the correct value of population.

The main difference between homogeneity of variance and homoscedasticity of variance is that, homogeneity of variance is mainly used in the context of ANOVA and homoscedasticity of variance is used in the context of regression.

Thus, if a scatterplot is used to understand the homogeneity of variance, a data that is homogeneous will have same scatter and a data that is heterogeneous or has heteroscedasticity will display a scattered pattern on the scatter plot.

In parametric statistics, it is also important that there are few outliers in the data. Outliers can be described as extreme values in data or distribution.

2.3.2 Assumptions of Non-parametric Statistics

It is often possible that, the assumptions and conditions as stated under parametric statistics are not met. In such a situation, non-parametric statistics can be used.

The following are the assumptions of non-parametric statistics:

1. **There is no assumptions of normality of the data:** In the earlier section of this unit, we discussed about the concept of non-parametric statistics and there we explained that non-parametric statistics is distribution free as there is no assumption made about the distribution of the data. Further, in this case the sample size can also be small, for example, N could be 4 or 6 (Mangal, 2010).

2. **The variables are measured in terms of nominal or ordinal scales:** We discussed under parametric statistics that the variables are measured in terms of interval and ratio scales. In non-parametric statistics, the variables are measured in terms of nominal and ordinal scales. Nominal scale can be used to measure variables that are qualitative as well as exclusive in nature. Whereas, ordinal scale involves ranks, that is, the data can be assigned ranks based on whether they are less or more, low or high, bad or good and so on and the data is ranked in terms of its magnitude.

3. **The observations need not be independent of each other:** Under non-parametric statistics, as such there is no assumption regarding the independence of relationship between the data.

4. **The population from which the research sample is taken from need not display homogeneous variance:** The sample can be heterogeneous and the variance could be heterogeneous or no assumption is made with regard to the variance.

Further, in non-parametric statistics, median is the measure of central tendency that is used here and it is less likely to be influenced by the outliers in the data.

Check Your Progress II

1) List the assumptions of non-parametric statistics.

……

……

……

……

2.4 APPLICATION AND USES OF PARAMETRIC AND NON-PARAMETRIC STATISTICS

When we discuss about the application and uses of parametric and non-parametric statistics, we need to remember that they are mainly used based on whether the above assumptions have been met or not.

While deciding about where to employ parametric or non-parametric statistics, we need to remember that parametric statistics need to be used if the assumptions met and if not then one can go for non-parametric statistics. Thus the assumptions discussed in the previous section are to be kept in mind while using parametric statistics.

When the assumptions and conditions under parametric statistics are met and we want to measure the difference between two groups or samples that are independent, we can use statistical technique Independent t test. When we want to measure two groups having repeated measures, we can use the parametric statistical technique paired t test. When we want to measure the difference between more than two groups or samples that are independent, we can use ANOVA. And when we want to study relationship between two variables, statistical technique like Pearson product moment correlations can be used.

The main advantage of parametric statistics is that all the necessary information about the population is available and as such parametric statistics is more powerful especially when its assumptions are met, though it could be difficult to meet these assumptions. Further, parametric statistics also involve complex calculations and computations.

In non-parametric statistics there are very few assumptions under Non-parametric statistics which make them easy to use. Non-parametric statistics can be used when the data is not meeting the assumptions as we have discussed under parametric statistics. For instance is the data is skewed, we can use non parametric statistics. This is relevant as when parametric statistics is used with a data that is skewed, the it will become less powerful. When sample is small, non-parametric statistics can be effectively used. Non-parametric statistics are often useful when we need to draw inferences from a sample size. Non-parametric statistics can be effectively used with categorical data and for sample that is drawn from a heterogeneous population. They are also simple to compute. Though they are less sensitive compared to parametric statistics and as such less powerful.

Given that the assumptions as stated under parametric statistics are not met, the non-parametric statistics can be used to measure the difference between groups or samples that are independent. (using Mann-Whitney U test).

It can also be employed while measuring the difference between groups or samples that are dependent or related. Techniques like Wilcoxon matched

pair test and sign test are some of the statistical techniques that can be used for the purpose.

It can also be used to measure the correlation between the variable. Some of the non-parametric statistical techniques that can be used for this purpose are Spearman's rank order correlation (that was discussed by us in BPCC104) and Kendall's tau.

Check Your Progress III

1) State any one use of non-parametric statistics.

……

……

……

……

……

2.5 DIFFERENCE BETWEEN PARAMETRIC AND NON-PARAMETRIC STATISTICS

To distinguish between parametric and non-parametric statistics, the following table (table no. 2.2) can be referred to:

Table 2.2: Difference between Parametric and Non-parametric Statistics

Parametric Statistics	Non-parametric Statistics
The assumed distribution is normal.	The assumed distribution may not be normal. It can be any distribution.
The variance is homogeneous.	The variance could be heterogeneous or no assumption is made with regard to the variance.
The scales of measurement used are interval or ratio.	The scales of measurement used are nominal or ordinal.
The observations need to independent of each other	There is no assumption with regard to the independence of the observations
Mean is the measure of central tendency that is used here.	Median is the measure of central tendency that is used here.
It is more complex to compute when compared to the non parametric techniques.	It is simple to calculate.

Can get affected by outliers	Is comparatively less affected by outliers

Check Your Progress IV

1) Fill in the blanks

   a) In parametric statistics, the assumed distribution is _____.

   b) In non parametric statistics, the scales of measurement used are _____.

   c) In parametric statistics, the observations need to _____ _____of each other.

2.6 LET US SUM UP

To summarise, in the present unit we discussed about parametric and non parametric statistics. Salkind (2014, page 468) described parametric statistics as "statistics used for the inference from a sample to a population that assume the variances of each group are similar and that the sample is large enough to represent the population" Statistical techniques like Pearson's Product Moment Correlation, t test and ANOVA can be termed as parametric statistical techniques. On the other hand, non-parametric statistics can also be described as tests that do not involve testing of hypotheses related to population parameters (King and Minium, 2013). Salkind (2014, page 46) described Non-parametric statistics as "distribution- free statistics that do not require the same assumptions as do parametric statistics". Statistical techniques like Spearman's rank order correlation, chi- square, Mann Whitney U test can be termed as non-parametric statistical techniques. Various assumptions of parametric and non-parametric statistics were also discussed. The unit also focused on the application and uses of parametric and non-parametric statistics and difference between parametric and non-parametric statistics was also discussed.

2.7 REFERENCES

Aron and Aron (2009). Statistics for Psychology (5th ed). New Delhi: Pearson 2. Howell, D. (2009).Statistical Methods for Psychology (7th ed.). Wadsworth.

Aaron, A., Aaron, E. N., & Coups, E. J. (2013). Statistics for psychology. New Delhi: Dorling Kindersley (India) Pvt. Ltd.

Cohen, B. H. (2008). *Explaining Psychological Statistics*. Hoboken, NJ: John Wiley & Sons, Inc.

Best, J. W and Kahn, J. V. (1999). Research in Education. New Delhi: Prentice Hall of India Pvt. Ltd. for information on research designs.

Elmes, D. G., Kantrowitz, B. H. and Roedier, H. L. III. (2003). *Research Methods in Psychology*. Belmont, CA: Thomson/Wandsworth Learning.

Kerlinger, Fred, N. (1995). Foundations of Behavioural Research. Bangalore: Prism Books Pvt. Ltd. for information on research, research designs, types of research and methods of data collection.

King, Bruce. M and Minium, Edward. W. (2008). Statistical Reasoning in the Behavioural Sciences. Delhi: John Wiley and Sons, Ltd.

Mangal, S. K. (2002). Statistics in psychology and Education. new Delhi: Phi Learning Private Limited.Minium, E. W., King, B. M., & Bear, G. (2001). Statistical reasoning in psychology and education. Singapore: John-Wiley.

Mohanty, B and Misra, S. (2016). Statistics for Behavioural and Social Sciences. Delhi: Sage.

Salkind, N. J. (2014). Statistics for People who (think that) hate statistics. New Delhi: SAGE.

Stephanie Glen. "Homogeneity, Homogeneous Data & Homogeneous Sampling" From StatisticsHowTo.com: Elementary Statistics for the rest of us! https://www.statisticshowto.com/homogeneity-homogeneous/

Tate, M. W.(1955). Statistics in Education. New York: Macmillion Co. Stephens, L. (2009). Statistics in psychology. New York: McGraw-Hill.

Veeraraghavan, V and Shetgovekar, S. (2016). Textbook of Parametric and Non-parametric Statistics. Delhi: Sage.

2.8 KEY WORDS

Homoscedasticity: Homoscedasticity denotes the variance between columns and rows display equality.

Parametric statistics: Salkind (2014, page 468) described parametric statistics as “statistics used for the inference from a sample to a population that assume the variances of each group are similar and that the sample is large enough to represent the population”

Non-parametric statistics: Non-parametric statistics can be described as tests that do not involve testing of hypotheses related to population parameters (King and Minium, 2013). Salkind (2014, page 46) described non-parametric statistics as “distribution- free statistics that do not require the same assumptions as do parametric statistics”

2.9 ANSWERS TO CHECK YOUR PROGRESS

Check Your Progress I

1) Define parametric statistics.

Salkind (2014, page 468) described parametric statistics as "statistics used for the inference from a sample to a population that assume the variances of each group are similar and that the sample is large enough to represent the population"

Check Your Progress II

1) List the assumptions of non-parametric statistics.

   The following are the assumptions of non-parametric statistics.

   - There is no assumptions of normality of the data
   - The variables are measured in terms of nominal or ordinal scales:
   - The observations need not be independent of each other
   - The population which the research sample is taken need not display variance needs to be homogeneous.

Check Your Progress III

1) State any one use of non-parametric statistics.

   Non-parametric statistics can be effectively used with categorical data and for sample that is drawn from a heterogeneous population.

Check Your progress IV

1) Fill in the blanks

a) In parametric statistics, the assumed distribution is normal.

b) In non parametric statistics, the scales of measurement used are nominal and ordinal.

c) In parametric statistics, the observations need to independent of each other.

2.10 UNIT END QUESTIONS

1) Discuss the concepts of parametric and non-parametric statistics.
1) Explain the assumptions of parametric statistics.
2) What are the assumptions of non-parametric statistics?
3) Differentiate between parametric and non-parametric statistics
4) Discuss the application of parametric and non-parametric statistics.