



NATIONAL OPEN UNIVERSITY OF NIGERIA

COURSE CODE: EDU 802

COURSE TITLE: ADVANCED EDUCATIONAL STATISTICS

**COURSE GUIDE EDU 802 ADVANCED EDUCATIONAL
STATISTICS**

Course Code: EDU 802

**Course Title: ADVANCED EDUCATIONAL
STATISTICS**

**Course Developer: Professor (Mrs.) C. N. Omoifo
Faculty of Education
University of Benin**

**Course Writer: Professor (Mrs.) C. N. Omoifo
Faculty of Education
University of Benin**

**Course Editor Prof. G A Badmus
School of Education
National Open University of Nigeria
Victoria Island-Lagos**

**Coures Cordnator Dr C A Okonkwo
School of Education
National Open University of Nigeria
Victoria Island-Lagos**

**Programme leader Dr C a Okonkwo
School of Education
National Open University of Nigeria
Victoria Island-Lagos**

COURSE GUIDE

Introduction

Welcome to EDU 802: Advance Educational Statistics which is a three credit unit course offered to doctoral students in Education. There are twenty-one Study Units in this course. The prerequisite for studying this course is EDU 702. It has been developed with appropriate examples in education suitable for education students.

This course guide is for distance learners enrolled in the Ph.D. programmes in Education of the National Open University of Nigeria (NOUN). This guide is one of the several resource tools available to help you successfully complete this course and ultimately your programme

In this guide, you will find very useful information about this course objectives, what the course is about, what course materials you will be using, available services to support your learning, and information on assignments and examination. It also offers you guidelines on how to plan your time for study, the amount of time you are likely to spend on each study unit and your tutor-marked assignments.

I strongly recommend that you go through this course guide and complete the feedback form at the end before you begin your study of the course. The feedback form must be submitted to your tutorial facilitator along with your first assignment. This guide also provides answers to several of your questions. However, do not hesitate to contact your study centre if you have further questions.

I wish you all the best in your learning experience and successful completion of this course.

Course Aims and Objectives

Course Aims

This course is aimed to review parametric statistics and then concentrate on the critical analysis of non-parametric statistical techniques.

Course Objectives

There are objectives to be achieved in each study units of this course. You should read them carefully before studying each unit. On completion of this course you should be able to:

- Explain rationale behind parametric and non-parametric tests presented
- Compare the parametric and non-parametric analogue when such analogue exists
- Select the “best” parametric and non-parametric test for a specific situation

- Perform parametric and non-parametric tests and apply the techniques to educational research.
- Conduct post hoc comparison on hypothesis rejected by parametric and non-parametric procedures in univariate analysis.
- Interpret outputs from computer analyses.

Course Summary

Module 1 introduces you to the basic principles of statistical measurements, distributions of measures, population parameters and sample statistics. Module 2 examines the concepts of probability and statistical inference. Module 3 deals with parametric techniques for testing hypothesis about the mean(s) of one or more populations. Module 4 deals with statistical techniques to explore relationships among variables. Module 5 explores non-parametric techniques to compare groups and Module 6 is on test construction statistics. There are twenty-one Study Units in this course. Each study unit consists of one week's work and should take you about three hours to complete. It includes specific objectives, guidance for study, reading materials and self assessment exercises. Together with Tutor-marked assignments, these exercises will assist you in achieving the stated learning objectives of the individual study units of the course.

Study Plan

The table below is a presentation of the course and how long it should take you to complete each study unit and the accompanying assignments.

Unit	Title of Study Unit	Weeks/Activity	Assignment
	Course Guide	1	Course Guide Form
Module 1: Introductory Concepts			
1	Principles of Statistical Measurements	2	Assignment
2	Distributions of Measures	3	Assignment
3	Population Parameters and Sample Statistics	4	TMA 1 to be submitted
Module 2: Probability and Tests of Statistical Significance			
1	Basic Concepts of Probability	5	Assignment
2	Probability Sampling Distribution	6	Assignment
3	Statistical Inference	7	TMA 2 to be submitted
Module 3: Parametric Techniques for Testing Hypothesis about the Mean(S) of One or More Populations			
1	t-tests	8	Assignment
2	Analysis of Variance (ANOVA)	9	Assignment
3	Analysis of Covariance (ANCOVA)	10	TMA 3 to be submitted

Module 4: Statistical Techniques to explore Relationships among Variables			
1	Correlation	11	Assignment
2	Regression	12	Assignment
3	Trend Analysis	13	TMA 4 to be submitted
Module 5: Non-Parametric Techniques to Compare Groups			
1	Chi-square Tests	14	Assignment
2	Mann-Witney Test	16	Assignment
3	Wilcoxon Signed Rank Test	17	Assignment
4	Kruskal-Wallis Test	18	Assignment
5	Friedman Test	19	TMA 5 to be submitted
Module 6 Test Construction Statistics			
1	Validity and Reliability of Scores	20	Assignment
2	Item Analysis	21	Assignment
3	Scaling Techniques	22	TMA 6 to be submitted

* Now use this overview to plan your personal timetable.

References/Further Readings

Your course material is the main text for this course. However, you are encouraged to consult other sources as provided for you in the list of references and further reading below:

References

- Adamu, S. O. (1997), *Statistics for Beginners Book 1*, Nigeria, Ibadan, SAAL Pub.
- Badmus, G.A. & Omoifo, C.N. (1998) "*The Essentials of Measurement and Evaluation in Education*" Benin: Osasu Publishers
- Cohen, J. (1988) *Statistical Power Analysis for the Behavioural Sciences*, Hillsdale, NJ: Erlbaum
- Hinton, P. R. (2003), *Statistics Explained A Guide for Social Science Students*, Hove and New York, Routledge
- Kerlinger, F. (1986), *Foundations of Behavioural Research*, New York: McGraw-Hill
- Leach, C. (1979), *Introduction to Statistics: A Non-parametric Approach for the Social Sciences* London: John Wiley and Sons
- Nworgu, B. G. (1991), *Educational Research: Basic Issues and Methodology*, Ibadan: Wisdom
- Owie, I. (2006), *Fundamentals of Statistics in Education and the Social Sciences*, Nigeria, Lagos, National Book Consortium
- Stevens, J. (1996), *Applied Multivariate Statistics for Social Sciences*, (3rd edition), Matiway, New Jersey: Lawrence Erlbaum Associates

Tabachnick, B. G. & Fidell, L. S. (1996), *Using Multivariate Statistics* (3rd edition), New York: HarperCollins

<http://www.gower.k12.il.us/Staff/ASSESS>

www.apsu.edu/oconnort/3760/3760lect03a.htm

faculty.vassar.edu/lowry/webtext.html

www.uwsp.edu/psych/stat/14/nonparm.htm

writing.colostate.edu/guides/research/stats/pop2a.cfm

www.shodor.org/interactivate/lessons/IntroStatistics

www.statbasics.co.uk

Further Reading

Conover, W. J. (1971), *Practical Non-parametric Statistic*, New York: John Wiley

Fleiss, J. L. (1973), *Statistical Methods for Rates and Proportions* New York: Wiley

How to Get the Most from this Course

In distance learning, the Study Units replace the university lecture. The advantage is that you can read and work through the course materials at your pace, and at a time and place that suits you best. Think of it as reading the lecture instead of listening to a lecturer. Just as a lecturer might give you in-class exercise, your Study Units provide exercises for you to do at appropriate times.

Each of the Study Units has common features which are designed to aid your learning. The first feature is an introduction to the subject matter of the unit and how a particular unit is integrated with other units and the course as a whole. Next is a set of learning objectives. These objectives let you know what you should be able to do by the time you have completed the unit. You should use these objectives to guide your study.

These exercises are designed to help you recall what you have studied and to evaluate your learning by yourself. You should do each Self Assessment Exercise as you come to it in the study unit. The summary at the end of each unit also helps you to recall all the main topics discussed in the main content of each unit. There are also tutor-marked questions at the end of each unit. Working on these questions will help you to achieve the objectives of the unit and prepare you for the assignments which you will submit and the final examination.

It should take you about three hours to complete a study unit, the exercises and assignments. When you have completed the first study unit take note of how long it took you and use this information to draw up a timetable to guide your study for the rest of your course. The wide margins on the left

and right side of the pages of your course book are meant for you to make notes of main ideas or key points which you can use when revising the course. If you make use of all these features, you will significantly increase your chances of passing the course.

Course Delivery

As an open and distance learner, you learn through several ways. You learn when you interact with the content in your course material in the same way as a student interacts with the teacher in a conventional institution. You also learn when you are guided through the course; however you are not taught the course. Instead, your course material is your teacher, and as such you will not be able to get answers to any questions which may arise from your study of the material. It is for this reason that, in addition to the course material which you have received, the delivery of this course is supported by tutorial, facilitation, and counselling support services. Although these services are not compulsory, you are encouraged to take maximum advantage of them.

Tutorial Sessions

The total number of tutorial hours for this course is 8 hours. Tutorial sessions form a part of your learning process as you have an opportunity to receive face-to-face contact with your tutorial facilitator and to receive answers to questions or clarifications which you may have. Also you may contact your tutorial facilitator by phone, email or mail.

On your part, you will be expected to prepare ahead of time by studying the relevant Study Units, write your questions so as to gain maximum benefit from tutorial sessions. Information about the location and time schedule for facilitation will be available at your study centre.

Tutorial sessions are a flexible arrangement between you and your tutorial facilitator. You will need to contact your study centre to arrange the time schedule for the sessions. You will also need to obtain your tutorial facilitator's phone number and email address.

Tutorial sessions are optional. However, the benefits of participating in them provide you a forum for interaction and peer group discussion which will minimise the isolation you may experience as a distance learner. You seriously need this interaction for the study of subject such as statistics.

Facilitation

Facilitation is learning that takes place both within and outside of tutorial sessions. Your tutorial facilitator guides your learning by doing the following:

- provide answers to your questions during tutorial sessions, or phone or by email;
- coordinate group discussions;
- provide feedback on your assignments;
- pose questions to confirm learning outcomes;
- coordinate, mark and record your assignment/examination score and;
- monitor your progress.

The language of instruction for this course is English. The course material is available in print or CD formats, and also on the university website.

On your part, you will be expected to prepare ahead of time by studying the relevant Study Units, write your questions so as to gain maximum benefit from facilitation.

Information about the location and time schedule for facilitation will be available at your study centre. Time of facilitation is a flexible arrangement between you and your tutorial facilitator. You should contact your tutorial facilitator if:

- you do not understand any part of the Study Units
- you have difficulty with the Self Assessment Exercises
- you have a question or a problem with an assignment, with your tutorial facilitator's comments on an assignment or with the grading of an assignment.

Counselling

Counselling forms a part of your learning because it is provided to make your learning experience easier. Counselling is available to you at two levels, academic and personal counselling. Student counsellors are available at the study centre to provide guidance for personal issues that may affect your studies. Your study centre manager and tutorial facilitators can assist you with questions on academic matters such as course materials, facilitation, grades and so on. Make sure that you have the phone numbers and email addresses of your study centre and the various individuals.

Assessment

There are three components of assessment for this course: Self Assessment Exercises and assignments at the end of each study unit; the Tutor-Marked Assignments; and a written examination. In doing these assignments, you are expected to use the information gathered during your study of the course. Below are detailed explanations on how to do each assignment.

Self Assessment Exercises (SAEs)

There are Self Assessment Exercises spread out through your course material. You should attempt each exercise immediately after reading the section that precedes it. Possible answers to the exercises are provided at the end of the course book; however, you should check the answers only after you must have attempted the exercises. The exercises are for you to evaluate your learning; they are not to be submitted. There are also questions spread through each study unit. You are required to attempt these questions after you have read a study unit. Again, the questions are to help you assess your knowledge of the contents of the unit. You are not required to submit the answers for SAEs.

Tutor-Marked Assignments (TMAs)

There are six Tutor-Marked Assignments for this course. The assignments are designed to cover all areas treated in the course. You will be given your assignments and the dates for submission at your study centre. You are required to attempt all six Tutor-Marked Assignments. You will be assessed on all six, but the best four performances will be used for your continuous assessment.

Each assignment carries 10% and together will count for 40% of your total score for the course. The assignments must be submitted to your tutorial facilitator for formal assessment on or before the stipulated dates for submission. The work that you submit to your tutorial facilitator for assessment will count for 40% of your total course score.

Guidelines for writing Tutor-Marked Assignments

1. On the cover page of your assignment, write the course code and title, assignment number (TMA 1, TMA 2), and date of submission, your name and matriculation number. It should look like this:

Course Code:

Course Title:

Tutor-Marked Assignment:

Date of Submission:
School and Programme:
Name:
Matriculation Number:

2. You should endeavour to be concise and to the point in your answers. You should give full details and working where so instructed. Your answer should be based on your course material, further readings and experience. However, do not copy from any of these materials. If you do, you will be penalised. Remember to give relevant examples and illustrations.
3. Use ruled foolscap sized paper for writing answers. Make and keep a copy of your assignments.
4. Your answers should be hand-written by you. Leave a margin of about 1.5 inches of the left side and about 5 lines before the answer to the next question for your tutorial facilitator's comments.
5. When you have completed each assignment, make sure that it reaches your tutorial facilitator on or before the deadline. If for any reason you cannot complete your work on time, contact your study centre manager and tutorial facilitator before the assignment is due to discuss the possibility of an extension. Extensions will not be granted after the due date unless under exceptional circumstances.

Final Examination and Grading

The final examination for EDU 802 will be of three hours duration, and will carry 60% of the total course grade. The examination will consist of questions which reflect the kinds of Self Assessment Exercises and questions in the Tutor-Marked Assignments which you have previously encountered. All areas of the course will be assessed. You should use the time between finishing the last unit and taking the examination to revise the entire course. You will find it useful to review your answers to Self Assessment Exercises and Tutor-Marked Assignments before the examination. For you to be eligible to sit for the final examinations, you must have done the following:

1. You should have submitted all the six Tutor-Marked Assignments for the course.
2. You should have registered to sit for the examination. The deadline for examination registration will be available at your study centre.

Failure to submit your assignments or to register for the examination (even if you sit for the examination) means that you will not have a score for the course.

Course Marking Scheme

The following table lays out the marks that constitute the total course score.

Assessment	Marks
Assignments 1-6 (six submitted, but the best four of all the assignments selected) out of 10%, totalling 40%	Four assignments, marked 60% of overall course score
Final examination score	60% of overall course score
Total	100% of course score

Conclusion

In conclusion, all the features of this course guide have been designed to facilitate your learning in order that you achieve the aims and objectives of the course. They include the aims and objectives, course summary, course overview, Self Assessment Exercises and study questions. You should ensure that you make maximum use of them in your study to achieve maximum results.

Summary

EDU 802 – Advanced Educational Statistics provides you with knowledge of statistical techniques available for research in Education. The course will equip you with the skills of choosing the right statistical test to use, how to use it and when to use it. During the review of the parametric tests, the assumptions underlying their uses and their “robustness” to violations of the assumptions will be discussed. Much of the discussions of non-parametric tests will emphasize the comparison of the non parametric procedures with their parametric analogies. This will help you in your ability to assess the pros and cons of using parametric and non-parametric statistics in given research situations. Interpretation of outputs from computer analysis of the various statistical tests is also well highlighted. This is because in this age of technology most analyses are no longer

manual but computer processed and you would be required to know how to interpret these outputs. Considering the importance of test in assessing students, the course also addressed test construction statistics for choosing good test items from teacher made tests.

I wish you success with the course and hope that you will find it both interesting and useful.

MAIN CONTENT

MODULE 1: INTRODUCTORY CONCEPTS

- Unit 1: Principles of Statistical Measurements
- Unit 2: Distributions of Measures
- Unit 3: Population Parameters and Sample Statistics

UNIT 1: PRINCIPLES OF STATISTICAL MEASUREMENTS

CONTENTS

- 1.0 Introduction
- 2.0 Objectives
- 3.0 Main Content
 - 3.1 Meaning of Statistics
 - 3.1.1 Statistics as Statistical Data
 - 3.1.2 Statistics as Statistical Methods
 - 3.1.3 Importance of Statistics
 - 3.2 Forms of Measurements
 - 3.2.1 Procedures for Measurement
 - 3.2.2 Properties of Measures
 - 3.2.3 Types of Scales
- 4.0 Conclusions
- 5.0 Summary
- 6.0 Tutor-Marked Assignment
- 7.0 References/Further Readings

1.0 INTRODUCTION

When you complete your course work and is to graduate, one condition you must fulfil is to present a report of a research study you undertook. This is known as ‘Thesis’ or ‘Dissertation’. In your research study, you are required to use statistics. This unit will present an explanation of this concept called statistics. In statistics you will make use of lots of measures or scores. It is therefore important that you begin the course with a review of the meaning of statistics and the principles of measurements. These may seem elementary to you since you have studied them previously but they are essential ingredients for understanding the course in statistics.

2.0 OBJECTIVES

At the end of this unit you should be able to:

- Explain the meaning of statistics
- Define statistics
- Give examples of the different types of statistical data
- Give examples of the different types of statistical methods
- Discuss the various scales of measurements
- List examples of different types of measures

3.0 MAIN CONTENT

3.1 Meaning of Statistics

What is statistics? In this section we will discuss the meanings of statistics. The term statistics is often used to mean either **statistical data** or **statistical methods**.

3.1.1 Statistics as Statistical Data

As statistical data, statistics refers to numerical descriptions of things. For example, you count or measure things such as counting the number of students in a class or school. You may also count the number of male and female students. These numerical descriptions you call statistics of students in that particular school or class. Other numerical descriptions may be the heights or weights of persons, students' scores in subject tests etc. Table 1.1 represents the statistical data or statistics of the distribution of students in an institution:

Table 1.1: Campus Distribution of Students

Programme	Number	Percentage
Undergraduate	700	70
Postgraduate	300	30
Total	1000	100

3.1.2 Statistics as Statistical Methods

You have been discussing one meaning of statistics – that of statistical data. The term can also mean statistical method. When you use the term statistics to mean statistical method you will refer to a body of methods that are used for collecting, organising and analysing numerical data. This is the more widely used meaning of statistics. There are different classifications of statistical methods. A statistical method can be classified as descriptive and inferential; parametric or non-parametric.

3.1.3 Importance of Statistics

Statistics tell us something about the population. We can use it to:

- **Estimate** the population parameter. After you have collected information from your sample, you can use it to estimate the population parameters. Let us illustrate this using teachers' salary. If the average teachers salary in your sample is **₦ 48,600** with a deviation of **₦ 400**, you might be able to say that the true mean salary is likely to be + or - **₦ 400** (i.e. **₦ 50,000 - ₦ 48,200**).
- **Determine change** in the population parameter:
These two sorts of applications results in two modes of statistical inference:

3.2 Forms of Measurements

Statistical methods are basically instruments for processing information. The information that they process is numerical in nature and derives from one or another of several forms of measurement. To understand the various forms of measurements we need to also understand the procedures for measurement and scales of measurement.

3.2.1 Procedures for Measurement

Let us consider some examples of measurements. First measure the length and height of your reading table by placing a meter rule across it and from up to down of the table respectively. What did you do to get the measures? One important thing you did to get the measures is to **count**. Counting is one procedure for measuring.

Let us assume that you have just received a telephone call informing you that you are one of the latest millionaires. Make a list of what you intend to do with your money. Your list may include buying the most expensive car in the world, trip to the United States to visit President Obama, buying a new home, changing your wardrobe etc. Supposing you are to rearrange your list to show the most important down to the least important, what you get is an **ordering** of your measures. Ordering is therefore another procedure of measurements. An alternative to ordering is **rating**. In rating, instead of rearranging, scores are awarded to each activity (say on a five-point scale) to indicate the rate of importance.

A third attempt will be made to understand procedures for measurements. This time let us assume that you are a Principal of a school. You have just conducted interview and admitted candidates into your school. How do you measure those admitted. One thing you could do is to separated the students into those admitted and those not admitted, or into males and females (if the school is a co-educational school). This is known as **sorting**. We say that the measures have been sorted out into different categories. Sorting is another procedure for measurement.

These three procedures of measurements produce different forms and types of measurement. It is therefore important that you understand the characteristics of the forms of measures from these three different procedures as they are determinants of the statistics that you will employ.

3.2.2 Properties of Measures

There are three important properties of measures. These are **magnitude, equal interval and absolute zero**. Magnitude is a property of ‘more ness.’ If it can be said that the attribute being measured is more in one case than the other the scale is said to have magnitude. A scale of measure has equal interval if the distance between any two points has the same meaning. For example, the positions on a meter rule has equal interval. The difference between 3cm and 5cm is the same as that between 11cm and 13 cm. This difference – 2cm means the same thing in both situations. Absolute zero exists when what is being measured is completely absent. For example when the presence of iodine is being measured in different leaves the absence of it in a particular leaf means absolute zero. However, when a student scores 0 in a test, it does not mean that there is no knowledge in the student. It simply means that the student did not answer correctly the points expected from him or her. In this case the 0 is not an absolute zero. The

absence or presence of these properties is used in classifying types of measures.

3.2.3 Types of Scale

There are four major types of scales – **nominal, ordinal, interval and ratio scales.**

Nominal Scale – This is a scale that does not have any of the properties of scales described above. In such measures numbers are assigned just for recognition. For example the team of 11 football players wears jersey numbered 1 to 11. It does not mean that the player wearing jersey number 11 is greater than or a better player than the one wearing jersey number 2 and so on. Similarly when you go out to collect data from males and females you may decide to assign the number 1 to males and 2 to females. It does not mean that males score lower or are inferior to females. Such measures are merely to identify, to name and to differentiate males from females, one player from another. Nominal scale is the simplest of all measurement scales.

Ordinal Scale – This scale has magnitude only. You will remember when you had to order your list of requirements in section above. In this scale there is implication that the steps in the rank order are equal. Let us consider students performance in a test. Student A scored 80%, B had 75%, C 63% and D 60%. Based on their performances, student A is first, B is second, C is third and D is fourth. However, the difference between first and second position is 5%, between second and third position you have 12% and between third and fourth position is 3%. In each situation there is no equal interval.

Interval Scale – This is a scale with magnitude and equal interval but no absolute zero. If you consider students performance in a test again, without ordering the measures you will agree that the measures on a scale of 100 has equal interval. Between 1% and 2% and between 2% and 3% we have equal interval as the difference. This is the most commonly used scale of measurement.

Ratio Scale – This is a scale that has the three properties discussed earlier. This scale is not commonly used in the field of education but for physical measurement in the pure sciences.

Based on the scale of measurement your data may be either **continuous** or **discrete**. Continuous data are such that the value of what is being measured may take unlimited number of intermediate values. For example, in measuring students' achievement or performance in mathematics, it is possible for a student to score 20%, another 20.6% yet another 21.1% etc. Between 20 and 21, there are intermediate values. Discrete data take on only a limited set of values. For example the number of children in the family, the number of students in the class is both examples of discrete data. You can have 1, 2, 3, 4, or 5 children in a family. You can have 30, 40 or 50 students in a class. It is not possible to have 2.5 or 30.3 children or students respectively in a family or class. Your data may also be described as **ordinal**. This is data that show order or ranking. For example a student who scored 90% may be the highest and thus ranked 1st. Another student with 75% is ranked 2nd. A third child with 70% ranks 3rd. 1st 2nd and 3rd are ordinal data.

Some times you will collect data from few persons and use them as they are. This is known as '**ungroup**' data. At other times the data will be so large that using them may become too cumbersome and confusing. In such situations the data are '**grouped**'.

4.0 CONCLUSIONS

Using Statistical methods we process measures. It is for this reason that we have began this study with principles of measurement. The statistical results you will get will, to a very large extent, depend on the type and quality of measures used. If you use the wrong type of measure, you will have a result which will be wrong and the interpretations and conclusions you will make will also be wrong. You must therefore make sure that you have understood the meaning of statistics and the different forms or types of measures available for use.

5.0 SUMMARY

In this unit you have learnt the two meanings of statistics and the principles of measurement. You also learnt the properties of measures and the different types of measures. The importance of an understanding of the meaning of statistics and type of measures for proper statistical analysis was also discussed. In the next unit you will be introduced to the concept of distribution of measures.

6.0 TUTOR-MARKED ASSIGNMENT

(1) The temperature in a classroom on a certain day is 30°C while the outside temperature is 26°C what type of scale is this measure? Give reason(s) to support your answer.

(2) The results of a student at the end of a semester programme in a certain institution are A, A⁺, C, D and E. What scale of measurement is this. Compare the type of measures in question 1 and 2.

7.0 REFERENCES/FURTHER READINGS

Adamu, S. O. (1997), *Statistics for Beginners Book 1*, Nigeria, Ibadan, SAAL Pub.

Badmus, G.A. & Omoifo, C.N. (1998) "*The Essentials of Measurement and Evaluation in Education*" Benin: Osasu Publishers

Hinton, P. R. (2003), *Statistics Explained A Guide for Social Science Students*, Hove and New York, Routledge

Owie, I. (2006), *Fundamentals of Statistics in Education and the Social Sciences*, Nigeria, Lagos, National Book Consortium

<http://www.gower.k12.il.us/Staff/ASSESS>

www.apsu.edu/oconnort/3760/3760lect03a.htm

faculty.vassar.edu/lowry/webtext.html

www.uwsp.edu/psych/stat/14/nonparm.htm

writing.colostate.edu/guides/research/stats/pop2a.cfm

www.shodor.org/interactivate/lessons/IntroStatistics

www.statbasics.co.uk

UNIT 2: DISTRIBUTIONS OF MEASURES

CONTENTS

- 1.0 Introduction
- 2.0 Objectives
- 3.0 Main Content
 - 3.1 Meaning of Distribution
 - 3.2 Frequency Distributions
 - 3.2.1 Ungrouped data
 - 3.2.2 Grouped Data
 - 3.3 Graphical Representations
 - 3.3.1 Bar Charts
 - 3.3.2 Histograms
 - 3.3.3 Frequency Polygon
 - 3.3.4 Pie Chart
 - 3.4 Parameters of Distributions
 - 3.4.1 Central Tendency
 - 3.4.2 Variability
 - 3.4.3 Skew, kurtosis and Modality
- 4.0 Conclusions
- 5.0 Summary
- 6.0 Tutor-Marked Assignment
- 7.0 References/Further Readings

1.0 INTRODUCTION

Measures of a variable are collected from several individuals, objects or cases. In most studies, huge amount of data is collected. The result is such that you may not find meaning in the result if left as collected. Often what is important is to present the data in forms or ways that will make the data meaningful. The method to use in describing or representing your data will depend on the type of data collected. You have studied this in your research method course but you will be required to briefly review it again. In a statistics course you will be dealing with list(s) of scores or measures. The list(s) simply is the distribution. To make meanings of the list of scores, we shall study in this unit the different types of distributions of measures. There are two major ways of presenting or describing your data. These are the *Frequency Distribution and Graphical Representation of Data*.

2.0 OBJECTIVES

At the end of this unit you should be able to:

- Draw Frequency distribution tables
- Draw Bar charts, histograms, line graph and pie charts
- List the parameters of distributions
- Calculate the mean, median and mode
- Calculate variance and standard deviation of set of measures
- Differentiate between positive and negative skewness
- Identify kurtosis and modality

3.0 MAIN CONTENT

3.1 Meaning of Distribution

The following is a list of students score in a test on integrated science:

78, 57, 50, 47, 76, 70, 74, 57, 71, 63, 60, 72, and 45

You can rearrange these measures from the lowest to the highest or from the highest to the lowest. You can also look at how related and how close the scores are. You may wish to see how clustered or how far apart these scores from one another. When you do these, you are looking at the distribution of scores. In its simplest form distributions is a list of scores or measures that are taken from a particular variable (in this case integrated science achievement).

3.2 Frequency Distributions

A frequency distribution is simply a table that shows the number of times each score in a set of data occurred. In constructing a frequency distribution table:

- The scores are arranged with the least score at the bottom of the table
- Tally (or count) is carried out on each score
- Finally the frequency of each score is determined

3.2.1 Ungrouped data

When the scores collected are few, you can distribute the list of scores based on the individual scores. This is known as ungroup data. You can construct a frequency distribution on the ungroup data as shown on Table 2.

Table 2: Frequency Distribution for Ungroup Data

Scores	Tally	Frequency
5	<i>ll</i>	2
4	<i>lll</i>	3
3	<i>lll</i>	3
2	<i>llll</i>	5
1	<i>llll</i>	5
0	<i>ll</i>	2

You can see from this distribution that some scores occurred more frequently than others.

3.2.2 Grouped Data

Some times the scores are too many that tabulating them as individual scores becomes cumbersome and uninteresting. In such situations the scores are grouped. Frequency distribution of group of scores is shown on table 3.

Table 3: Frequency Distribution for Group Data

Class	Tally	Frequency
81 – 85	<i>l</i>	1
76 – 80	<i>lll</i>	3
71 – 75	<i>ll</i>	2
66 – 70	<i>llll l</i>	6
61 – 65	<i>llll l</i>	6
56 – 60	<i>llll ll</i>	7
51 – 55	<i>llll llll</i>	10
46 – 50	<i>llll llll lll</i>	13
41 – 45	<i>ll</i>	2
36 – 40	<i>llll llll</i>	10
31 – 35	<i>llll</i>	4

You have constructed several of such tables in the past. You will need the skill in your study in this course.

3.3 Graphical Representations

Sometimes the data collected is presented in form of graphs. Graphs can be described as pictorial representation of raw data. There are different types of graphs.

3.3.1 Bar chart:

These are graphs with bars whose heights vary according to the frequencies of the scores.

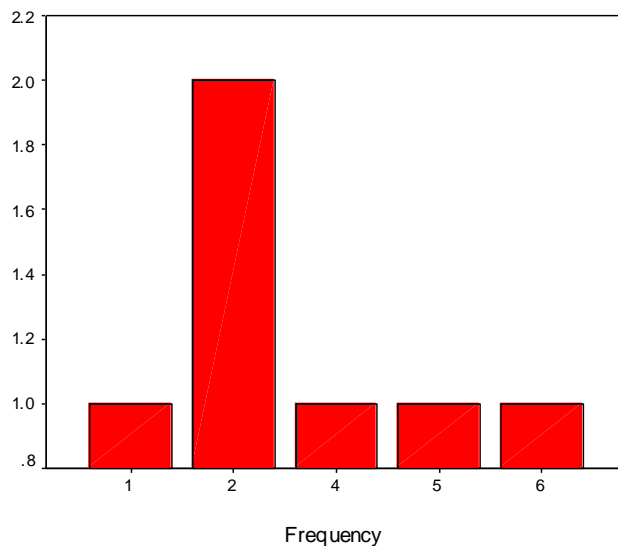


Figure 1: Sample Bar Chart

Usually the bars are separate with equal width. What is different is the height of the bars which represents the frequencies.

3.3.2 Histogram

These are graphs that are plotted using the mid-points of sets of data. The bars in a histogram are joined.

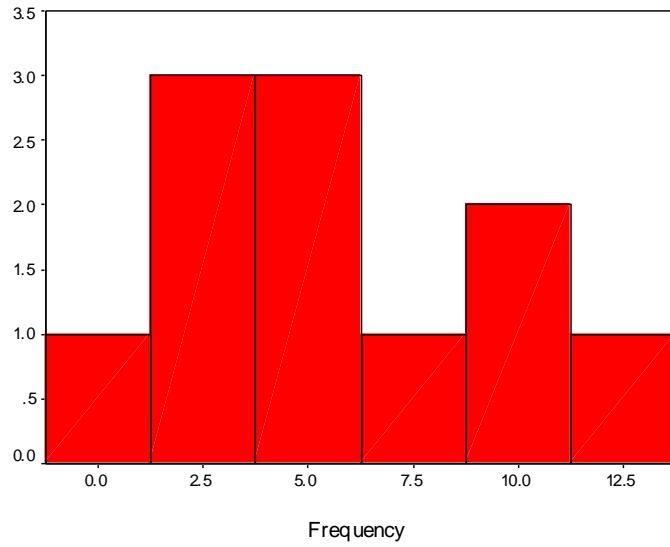


Figure 2: Sample Histogram

3.3.3 Frequency Polygon

This is a line graph. It is simply a line joining the mid-point of all the bars of a histogram.

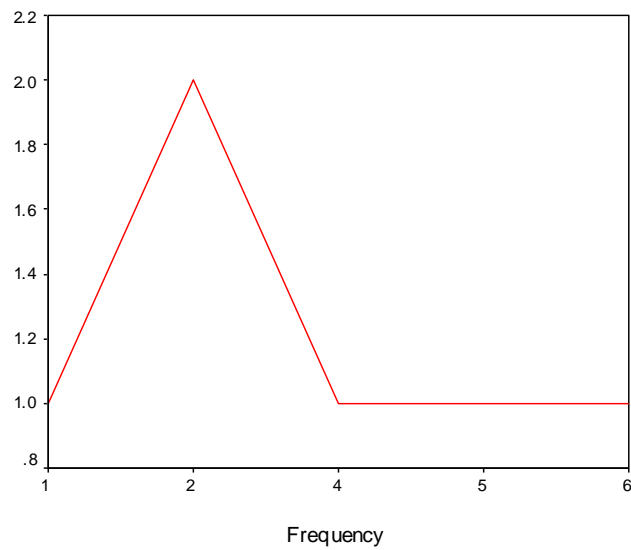


Figure 3: Sample Frequency Polygon

3.3.4 Pie-Chart

This is a circular graph.

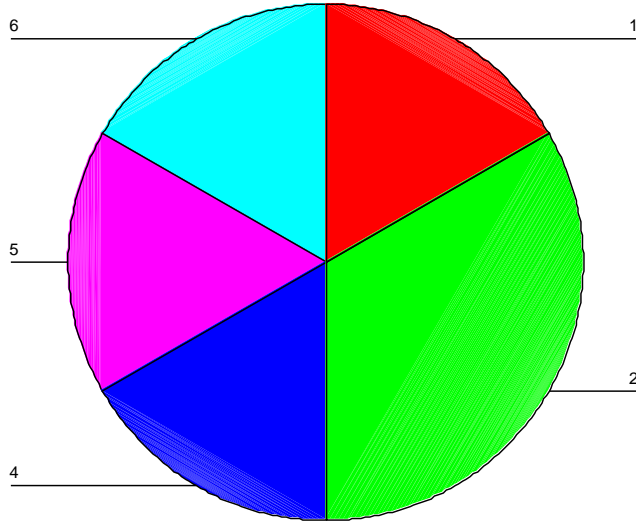


Figure 4: Sample Pie Graph

3.4 Parameters of Distributions

Generally, the main parameters of a distribution are measures of **central tendency** and measures of **variability**. Others are **skew**, **kurtosis** and **modality**. Several statistical methods will require a good understanding of central tendency and variability. You will be expected to study these two parameters in details. Skew, kurtosis and modality will only be treated briefly.

3.4.1 Central Tendency

This is also known as measures of location. The three common ways of measuring location or central tendency are **mean**, **median** and **mode**. You have learnt how to calculate these measures of location in your previous courses on statistics. In this section we will only review the formulae for computing them and work on some examples.

Arithmetic means: The symbol for arithmetic mean is \bar{x} read as 'x bar'. It is also known as average. It is a measure that describes sample of study. It is a single measure which can give a description of the attributes of study.

When the number of measures is few, you can calculate the mean without grouping the data. This is known as **mean for ungrouped data**. You will recall the statistics for mean:

For ungroup data:

$$\mathbf{Mean} = \bar{x} = \frac{1}{N} \sum \mathcal{X}_i = \frac{\Sigma X}{\Sigma f}$$

Where Σ = 'sum of'
 X = score
 N = Number of scores
 f = Frequency

For group data:

$$\mathbf{Mean} = \bar{x} = \frac{\Sigma fx}{\Sigma f} = \frac{\Sigma fx}{N}$$

Where x = mid-point of the class interval.

Median: This is the point in a distribution below which or above which half (50%) of the scores lie. It is also known as the midpoint of a distribution of scores when the scores are arranged in order of magnitude.

For example: 4, 7, 8, 9, 7, 3 and 6 are set of scores. When arranged in order of magnitude you will have 3, 4, 6, 7, 7, 8 and 9. The score in the middle or the midpoint of the distribution of the scores is 7. This is known as the median.

Mode: This is the score that occurs the greatest number of times in a distribution. It is the most frequent or typical value of the distribution. For example in the set of scores 4, 7, 8, 9, 7, 3 and 6, all the scores appear once except for 7. The score 7 appears twice and is the score that occurred the greatest number of time. The score 7 is therefore the mode of the set of scores.

3.4.2 VARIABILITY

In the previous section we considered one value that described all the scores in the set. There are also some statistics that inform you how far apart the measures of location are from each of the scores in a distribution.

These statistics are known as measures of variability or dispersion. Three commonly used measures of dispersion are;

The Range: This is the simplest measure of dispersion and very easy to calculate. The range is the difference between the largest and smallest scores. For example with the set of scores of 4, 7, 8, 9, 7, 3 and 6, the highest score is 9 and the lowest score is 3. The range is $9 - 3 = 6$. The more spread out the scores are the larger the range.

Variance: You would have noticed that the range makes use of only two scores in the set of data. There is a statistic that makes use of all the scores in the set of data and for this reason is more complex to calculate than the range. It is known as the variance. The variance is a measure of the distance of scores from the mean. To compute the variance you need to:

- First know the mean of the distribution.
- The mean is then subtracted from each of the scores. The result of some of the differences of the scores about the mean may be negative. This is the situation when the mean is greater than the score.
- To avoid negative numbers, the difference of the mean from the scores is squared.
- Finally the average of the deviation of the scores about their means is calculated and this is the variance.

Summarizing all the steps above, the formula for computing variance is:

$$\text{Variance} = \frac{1}{N} \sum (\chi_i - \bar{x})^2$$

The notation for variance is s^2 (for sample variance) or σ^2 (for population variance).

Standard Deviation: This is the most commonly used measure of dispersion. Because in variance the deviations are squared, the value of variance is also in square. This is not the same unit as the original scores. To achieve the unit of the original scores the square root of the variance is computed. This is known as standard deviation.

$$\text{SD or } S = \sqrt{\frac{1}{N} \sum (x_i - \bar{x})^2}$$

In summary, the variance is the square of the standard deviation, and the standard deviation is the square root of the variance.

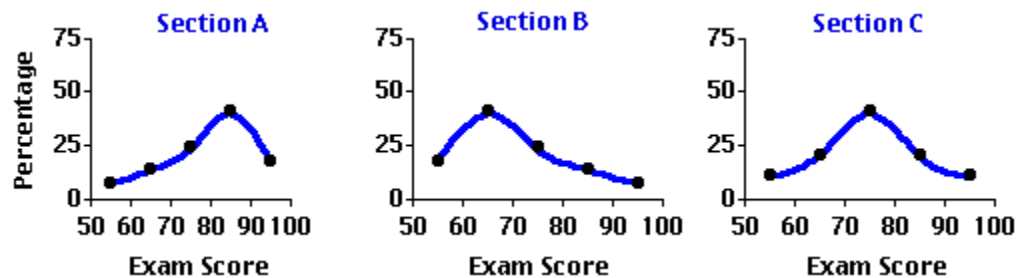
3.4.3 Skew, kurtosis and Modality

You have learnt that in both mean median and mode the interest is in the measure in the centre. Sometimes high or low scores occur more frequently than the scores in the middle or centre.

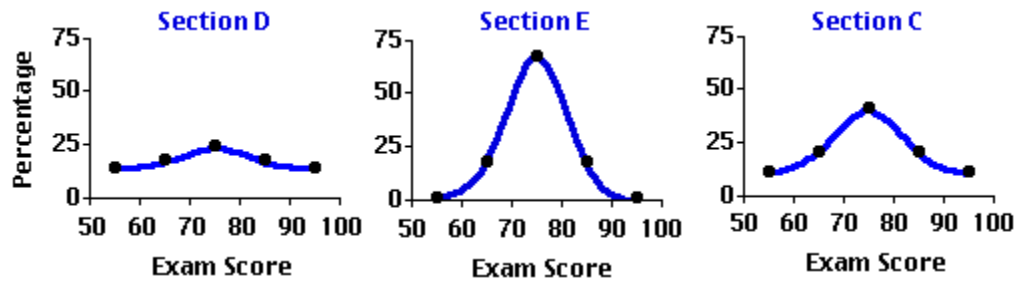
Skew: When a distribution is off centre it is said to be skewed. There are two types of skew – negative and positive skew. In a negative skew the scores cluster towards the higher range such that the tail of the distribution is longer at the lower end. In the positive skew, the scores cluster to the lower range such that the tail is longer at the higher range end.

SELF ASSESSMENT EXERCISE 1

- 1) Which of the three distributions below is (i) positively skewed (ii) negatively skewed?
- 2) How would you describe the third distribution?



Kurtosis: Sometimes distributions are short and flat or tall and slender. This is also a deviation from normality. **Kurtosis** describes the shape of a distribution.



The shapes above are examples of Kurtosis.

SELF ASSESSMENT TEST 2

- 1) Describe each of the shape above
- 2) If in a class majority of the students have equal mastery level what shape will their measure assume?

A distribution is said to be **platykurtic** (flat-curve) when the shape is short and flat (as with graph D). In a platykurtic distribution the individual measures are spread out fairly uniformly across their range. When tall and slender (graph E) it is described as **leptokurtic** (slender-curve). In a leptokurtic distribution the individual measures tend to cluster compactly at some particular point in the range. A shape in between these two extremes (graph C) is described as **mesokurtic** (medium-curve).

Modality

The number of distinct peaks or areas of cluster that appear within a distribution is known as the modality. Each of the peaks is called the mode. A distribution with only one distinct peak is described as **unimodal**; if it has two distinct peaks it is spoken of as **bimodal**; three peaks, **trimodal**; and so on. Most distributions that you are likely to encounter will be either unimodal or bimodal. Only rarely will you find distributions that have three or more distinct peaks.

4.0 CONCLUSIONS

You have learnt of some of the preliminary work that is usually carried out on data. This is basic in some investigations, while in others it may be the main statistical work to be done. What you have learnt is very important because many statisticians are interested in certain decision parameters which the population being studied has.

5.0 SUMMARY

In this unit you have tried to learn how to put data collected in numerical form into a meaningful form. You have summarized numerical data into tables, charts, graphs and have also made some summary calculations. You also learnt how to calculate measures of location and dispersion such as the mean, median, mode, standard deviation and variance.

6.0 TUTOR-MARKED ASSIGNMENT

The following is the score of 60 students in an integrated science examination

24	14	11	12	13	15	8	17	1	7
14	7	3	15	28	10	10	19	20	1
14	16	4	11	22	18	6	14	10	4
9	19	16	20	5	5	16	15	23	10
6	7	5	0	13	4	5	0	8	17
24	0	29	14	3	24	22	8	2	28

- Construct the frequency table using the ungrouped data.
- Group the data and using class interval/size of 3 construct a second frequency table.
- Draw the different types of graph for the data.
- Compute the mean, variance and standard deviation of the set of scores.

7.0 REFERENCES/FURTHER READINGS

Adamu, S. O. (1997), *Statistics for Beginners Book 1*, Nigeria, Ibadan, SAAL Pub.

Badmus, G.A. & Omoifo, C.N. (1998) "*The Essentials of Measurement and Evaluation in Education*" Osasu Publishers Ltd, Benin City

Hinton, P. R. (2003), *Statistics Explained A Guide for Social Science Students*, Hove and New York, Routledge

Owie, I. (2006), *Fundamentals of Statistics in Education and the Social Sciences*, Nigeria, Lagos, National Book Consortium

<http://www.gower.k12.il.us/Staff/ASSESS>

www.apsu.edu/oconnort/3760/3760lect03a.htm

faculty.vassar.edu/lowry/webtext.html

writing.colostate.edu/guides/research/stats/pop2a.cfm

www.shodor.org/interactivate/lessons/IntroStatistics

www.statbasics.co.uk

UNIT 3: POPULATION PARAMETERS AND SAMPLE STATISTICS

CONTENTS

- 1.0 Introduction
- 2.0 Objectives
- 3.0 Main Content
 - 3.1 Population versus Sample
 - 3.1.1 Definitions
 - 3.1.2 Differences between Population and Sample
 - 3.2 Sampling Techniques
 - 3.2.1 Ways of Generating Information about a population
 - 3.2.2 Types of Sampling Techniques
- 4.0 Conclusions
- 5.0 Summary
- 6.0 Tutor-Marked Assignment
- 7.0 References/Further Readings

1.0 INTRODUCTION

In your previous programmes you studied some statistics. In this unit will try to discuss some language of statistics. This is necessary because it helps to reduce the anxiety and confusion of dealing with many concepts in statistics. The word statistics is used to mean statistical data or statistical method. Statistical data are collected from units which may be referred to as population and samples. This unit the first of the three units of module 1 will consider these two concepts.

2.0 OBJECTIVES

At the end of this unit you should be able to:

- Define terms used in describing population and sample
- Identify concepts associated with population and samples
- Explain the different types of parameters, statistics and sampling
- Compare and contrast population and sample
- Make connections between parameters and statistics

3.0 MAIN CONTENT

3.1 POPULATION VERSUS SAMPLE

Population and sample are key concepts in statistics. Understanding these concepts is vital to understanding the principles of statistics. This section presents definitions, parameters and statistics as they relate to population and samples.

3.1.1 DEFINITIONS

This course is on Advanced Statistics. We will begin by defining the term 'Statistics'. Kerlinger, (1986) defined it as the theory and method of analyzing quantitative data obtained from samples of observations in order to study and compare sources of variation of phenomena, to help make decisions to accept or reject hypothesized relations between phenomena, and to aid in making reliable inferences from empirical observations. From this definition the concept of sample is very important in statistics. Related to sample is the term population. We will begin with the definition of population and sample.

There are different conceptions of the words population and sample. Let us now consider two such conceptions. The first is that in statistics you study groups of people, objects, or data measurements and produce summarizing mathematical information on the groups. The groups are called **samples**. The larger collection of people, objects or data measurements is called the **population**. In the second conception you study a clearly defined set of elements in which you are interested. This set of elements is called the **universe**. A set of values associated with these elements is called a **population** of values. Therefore the statistical term **population** has nothing to do with people or objects per se but is a collection of values. Let us illustrate these with a study of students at a particular school, all the students at that particular school make up the universe while students with a particular height or weight values make up the population. Both conceptions are acceptable and important. They are often used interchangeably. No matter the conception, it is important to note that population includes all objects of interest whereas the sample is only a portion of the population. However, it is often impossible to measure all of the values in a population. A collection of measured values is called a **sample**.

SELF ASSESSMENT EXERCISE 1

Which conception will be superior in a course in research method and statistics?

3.1.2 DIFFERENCES BETWEEN POPULATION AND SAMPLE

To understand the differences between a population and a sample, it is necessary to consider numbers which are used to describe the two. A number that describes some aspect of the population is known as ‘**Parameter**’, while the number which can be calculated from the sample data is described as ‘**Statistics**’. **Mean:** The arithmetic mean is one common population parameter and is symbolised as μ (Greek small letter mu) whereas the sample mean is written as \bar{x} .

Example: Teachers salary in Nigeria. There are different categories of teachers with different salaries. If it is possible to calculate one salary to represent teacher’s salary in Nigeria, the value or number will be known as the ‘mean’ or average teacher’s salary. Since it is calculated from the salaries of all teachers in Nigeria it describes the population and is known as parameter represented as μ (e.g. $\mu = \text{₦ } 50,000$)

In real applications, parameters are not generally known since it is often very difficult or impossible to reach all members of a population. You have to draw a random sample of all categories of teachers. If you calculate the average salary of teachers from this sample and assuming you get an amount of forty-eight thousand six hundred naira, this number is described as statistic and is represented as $\bar{x} = \text{₦ } 48,600$.

To distinguish between population parameters and sample statistics, different symbols are used. There is no universal agreement, but generally Greek letters refer to population values, and Roman letters to statistics. Let us consider some of the population parameters and sample statistics which you studied at the undergraduate or postgraduate diploma levels.

Variance: The population variance usually denoted by σ^2 (Greek small letter sigma squared), is the expected value of the squared difference of the values from the population mean. The sample variance is denoted by s^2 .

Unbiased Estimates of Population Parameters

Suppose we want to estimate the population mean from a sample of 50. We could use the sample mean or the sample median. Such an estimate is called a **point estimator**. In using the sample median, lots of different medians

corresponding to the different samples of 50 are computed. If you compute the expected value of the sample median and it equals the population mean it is said to be an **unbiased estimator** of the population mean, if not it is a **biased estimator** with bias equal to the difference between the expected value of the estimator and the value of the population parameter.

Sample Variability

Let us assume that 1000 teachers we sampled from the population to calculate the statistic $\bar{x} = \text{N } 48,600$. If you select another set of 1000 teachers (a second sample), it is most likely that the mean of this second sample will not be exactly $\bar{x} = \text{N } 48,600$. This is due to variations that exist in different samples. The variation of a sample statistic from one sample to the next is called *sampling variation* or *sampling variability*. When sampling variation is large, the sample contains little info about the population parameters, but when sampling variation is small, the sample statistic is informative about the parameter

SELF ASSESSMENT EXERCISE 2

What are the formulae for the following?

- (a) Population and Sample Means
- (b) Population and Sample Variance
- (c) Population and Sample Standard Deviation

Try calculating each of these parameters and statistics using examples

3.2 SAMPLING TECHNIQUES

In the last section you learnt that the sample is an estimate of a population and that information about a population is generated for the samples of the population. In this section you will learn the different ways by which samples are derived from a population.

3.2.1 Ways of Generating Information about a Population

There are two major ways of generating information about a population. These are through **Census** and **Sampling**. Taking a census involves collecting information from all elements in the population. For large population, this is almost impossible and when possible very expensive. Sampling is the less expensive and the most widely used method. Sampling is a process of selecting a subset of elements from a population for the purpose of estimating the population parameters.

3.2.2 Types of Sampling Techniques

There are two major types of sampling techniques – **random** and **non-random** sampling techniques.

1. Random sampling

In a random sampling technique, every element in the population has the same probability of being selected for inclusion in the sample. This type of sampling yields unbiased estimates of parameters and for this reason is mostly used. Four types of random sampling techniques are:

- Simple random sampling- Every person has an equal chance of being selected
- Systematic sampling- Every k^{th} person is selected
- Stratified random sampling- divide into groups and then sample
- Cluster (area) sampling- naturally occurring groups- stats on the groups

2. Non-random sampling

This is a technique in which every element in the population does not have the same probability of being selected for inclusion in the sample. This type of sample can yield biased and unreliable estimates of parameters.

SELF ASSESSMENT EXERCISE 2

Write short notes on each of the four types of random sampling techniques

4.0 CONCLUSIONS

Understanding sample statistic and population parameters is very important. Statistical analysis will depend on which parameter or statistic you wish to study. Besides a knowledge of population and sample will go a long way at assisting you to make right decisions on what statistical method to employ.

5.0 SUMMARY

In this unit you have learnt about population parameters and sample statistics. You also learnt of different ways of sampling so as to generate representative data that will aid making of conclusion about a population parameter from the sample statistics.

6.0 TUTOR-MARKED ASSIGNMENT

In the following data, two class frequencies are missing:

Class Interval	Frequency
100 – 110	4
110 – 120	7
120 – 130	15
130 – 140	---
140 – 150	40
150 – 160	---
160 – 170	16
170 – 180	10
180 – 190	6
190 – 200	3

However, it was possible to ascertain that the total number of frequencies was 150 and that the median has been correctly found out as 146.25. You are required to find out with the help of the information given:

- (i) The two missing frequencies
- (ii) Having found the missing frequencies, calculate the arithmetic mean and standard deviation.

7.0 REFERENCES/FURTHER READINGS

Badmus, G.A. & Omoifo, C.N. (1998) "*The Essentials of Measurement and Evaluation in Education*" Osasu Publishers Ltd, Benin City

Kerlinger, F. (1986), *Foundations of Behavioural Research*, New York: McGraw-Hill

Owie, I. (2006), *Fundamentals of Statistics in Education and the Social Sciences*, Nigeria, Lagos, National Book Consortium

<http://www.gower.k12.il.us/Staff/ASSESS>

www.apsu.edu/oconnort/3760/3760lect03a.htm

faculty.vassar.edu/lowry/webtext.html

writing.colostate.edu/guides/research/stats/pop2a.cfm

www.shodor.org/interactivate/lessons/IntroStatistics

www.statbasics.co.uk

MODULE 2 PROBABILITY AND TESTS OF STATISTICAL SIGNIFICANCE

- Unit 1: Basic Concepts of Probability
- Unit 2: Probability Sampling Distribution
- Unit 3: Statistical Inference

UNIT 1: BASIC CONCEPTS OF PROBABILITY

CONTENTS

- 1.0 Introduction
- 2.0 Objectives
- 3.0 Main Content
 - 3.1 Meaning of Probability
 - 3.1.1 Simple Probability
 - 3.1.2 Compound Probability
 - 3.2 Simple Laws of Probability
 - 3.2.1 The Addition Law of Probability for Mutually Exclusive Events
 - 3.2.2 Joint Probability
 - 3.2.3 The Multiplication Law of Probability for Independent Events
 - 3.2.4 Modification of Multiplication Law
- 4.0 Conclusions
- 5.0 Summary
- 6.0 Tutor-Marked Assignment
- 7.0 References/Further Readings

1.0 INTRODUCTION

Betting is an activity common among many Nigerians. We have pool houses, electronic betting and the rest. The activities of these often require one using intuition to predict what will happen. This we may call common sense. You and many others have had the opportunity to exhibit your intuition. Great Mathematicians have explained that probability has its roots in common sense. This goes to show that you already have some good, solid intuitions about the generalities of probability, even though you might not yet know anything at all about its technical details. In this unit you will consider the technical and mathematical considerations of our common sense translated into the concept of Probability.

2.0 OBJECTIVES

At the end of this unit you should be able to:

- Explain the meaning of probability
- Differentiate between simple and compound probability
- Differentiate between conjunction and disjunction probability
- Apply addition law or probability correctly
- Apply multiplication law of probability correctly

3.0 MAIN CONTENT

3.1 Meaning of Probability

Consider this simple scenario; assuming you have 4 balls and 3 of them are red balls while 1 is a blue ball, and you blindly want to choose one red ball. Assuming you placed a bet with your friends on whether the ball you will be drawing is a red or blue ball. This question of which of the balls you will draw is similar to the question ‘which of the two colours has the greater probability of being drawn?’ Your common sense should tell you that if you have 3 of the 4 balls as red, then you have 3 of the 4 chances of drawing a red ball, but only 1 out of the 4 chances of drawing a blue ball. Mathematically, the chance of drawing a red ball is $\frac{3}{4}$ which is equal to 0.75, while the chance of drawing a blue ball is $\frac{1}{4}$ which is equal to 0.25.

It follows that the probability of the occurrence of an event ‘X’ is given as:

$$P_{(X)} = \frac{\text{Number of possibilities favourable to the occurrence of } x}{\text{Total number of pertinent possibilities}}$$

This is known as **relative frequency concept of probability**. From the foregoing a simple definition of probability is **the relative frequency of an event within a reference class of similar events**.

3.1.1 Simple Probability

Consider a second scenario. For example a die has 6 faces 1 to 6. The probability of throwing a 5 in a single throw of an unbiased die is $\frac{1}{6}$. This is because of the 6 faces there is only 1 face labelled 5. Another example is a coin having 2 faces – a head and a tail. The probability of an unbiased coin falling heads is $\frac{1}{2}$. In these illustrations the probability of **individual**

events are considered. These are examples of simple concepts of probability.

Seeing probability as relative frequency means that in any calculation of probability the value is between 0 and 1. If an event is certain to occur its probability is 1. If an event is certain not to occur its probability is 0. It is possible to combine probabilities e.g. probability of 4 or 5 showing in a single throw of a die. To achieve this there are some simple laws of probability that will frequently speed up the calculation.

3.1.2 Compound Probability

In 3.1.1 you learnt of a simple probability event. It is possible to link two or more individual events and determine the probability of the events occurring. This is known as compound probability. There are two major ways of linking individual probabilities – **conjunction and disjunction**.

- **Conjunction Probability** – This is associated with the common sense meaning of the word **‘and’**. Conjunctive probability asks such questions like ‘What is the Probability of X **and** Y occurring?’
- **Disjunctive Probability** – This is associated with the common sense meaning of the word **‘or’**. It asks such questions as ‘What is the probability of X **or** Y occurring?’

3.2 Simple Laws of Probability

The two ways of linking individual probabilities have given rise to the laws of probability. In this section you will learn about the two major laws of probability and their amendments.

3.2.1 The Addition Law of Probability for Mutually Exclusive Events

This law states that if 2 events A and B are mutually exclusive, then the probability that either A or B will occur is equal to the sum of their separate probabilities. (events are said to be mutually exclusive if they cannot both occur simultaneously. For example, a coin has two faces head or tail. It is not possible to throw a coin and both faces will show at the same time. Only one at a time will show. Thus the head and the tail are said to be mutually exclusive. The probability of a head or a tail of a coin is:

$$\frac{1}{2} + \frac{1}{2} = \frac{1+1}{2} = \frac{2}{2} = 1$$

For a die with six faces, the probability of a 4 and a 5 occurring is:

$$\frac{1}{6} + \frac{1}{6} = \frac{2}{6} = \frac{1}{3}$$

The probability of a 2, a 3 and a 6 occurring is:

$$\frac{1}{6} + \frac{1}{6} + \frac{1}{6} = \frac{1+1+1}{6} = \frac{3}{6} = \frac{1}{2}$$

3.2.2 Joint Probability

Supposing the events are not mutually exclusive, the addition law is modified to calculate the probability of the events. Let us consider the die again. If you want to know the probability of events divisible by 2 or 3 occurring, you need to know the events. Faces 2, 4 and 6 are divisible by 2. Similarly faces 3 and 6 are divisible by 3. In such a situation 6 is occurring in both events, so we say the events are **not mutually exclusive**.

If you apply the addition law the probability of the events will be:

$$\text{For divisibility by 3} = \frac{2}{6}.$$

$$\text{For divisibility by 2} = \frac{3}{6}$$

$$\text{Addition law} = \frac{2}{6} + \frac{3}{6} = \frac{5}{6}$$

This is not true since the probability of 6 has been taken twice. In such a situation the Law is to subtract – the probability of that which make it not to be mutually exclusive from the added probability i.e.

$$\text{Probability of 2 or 3 divisible by} = \left(\frac{3}{6} + \frac{2}{6}\right) - \frac{1}{6}$$

$$= \frac{5}{6} - \frac{1}{6}$$

$$= \frac{4}{6} = \frac{1}{2}$$

This is more correct because if you consider the faces, there are 4 that meet the requirement - 2, 3, 4 and 6 therefore: $\frac{4}{6}$

The general formula for probability of A or B occurring is:

$$\text{Prob (A or B)} = \text{Prob (A)} + \text{Prob (B)} - \text{Prob (A and B)}$$

This formula is known as **JOINT PROBABILITY A and B**.

3.2.3 The Multiplication Law of Probability for Independent Events

Assuming you tossed two unbiased coins, the result from one coin does not affect the result in the second coin. Such events are said to be independent. Suppose you toss the two coins four results are possible – both coins displaying heads (HH), both coins displaying tail (TT), first coin displaying Head and the second coin tail (HT) and first coin displaying tail and the second Head (TH). Therefore the probability of getting each combination is $\frac{1}{4}$. The multiplication law gives a simpler way of arriving at the same result. This law states that “if two events are independent then the probability of both occurring is the product of the probabilities of each. For example:

$$\begin{aligned}\text{prob (HH)} &= \text{Prob (H with first coin)} \times \text{prob (H with 2}^{\text{nd}} \text{ con).} \\ &= \frac{1}{2} \times \frac{1}{2} \\ &= \frac{1}{4}\end{aligned}$$

When two events are mutually exclusive both the addition and the multiplication laws apply; for example finding the probability of getting only one head when two coins are tossed. The two events are HT and TH. If we take the simple explanation above,

$$\begin{aligned}\text{prob (HT) for the independent event} &= \frac{1}{4} \\ \text{prob (TH for the independent event)} &= \frac{1}{4} . \\ \therefore \text{For the 2 events} &= \frac{1}{4} + \frac{1}{4} = \frac{2}{4} = \frac{1}{2}\end{aligned}$$

Applying multiplication Law

$$\text{Prob (one head)} = \text{prob (HT or TH)} + \text{prob (H)} + \text{prob (TH)}$$

$$\text{Prob (HT)} = \text{prob (H)} \times \text{prob (T)} = \frac{1}{2} \times \frac{1}{2} = \frac{1}{4}$$

$$\text{Prob (TH)} = \text{prob (T)} \times \text{prob (H)} = \frac{1}{2} \times \frac{1}{2} = \frac{1}{4}$$

3.2.4 Modification of Multiplication Law

When events are not independent the multiplication law is modified. Consider a bunch of cards (for games) with many icons such as the club. Assuming you have a bunch containing 52 cards and that the number of club cards is 13, then the probability of drawing any club of the 13 clubs is

not independent. If you draw a first club, the probability of a second one will be dependent on the first. The probability of the first card being club is $13/52$. The probability of a second card being club is $12/52$. This is because a first club has been drawn ($13-1$) leaving one, 1 clubs left. In this situation, the multiplication law is given as: “The probability of two events x and Y both occurring is the probability of x occurring multiplied by the probability of Y occurring given that x has already occurred. Symbolically this is:

$$\text{Prob (X and Y) = prob (X) x prob (Y/X)}$$

Where Y/X means Y occurring given that X has already occurred.

From our example:

$$\begin{aligned}\text{Prob (X and Y)} &= \text{prob (X) x prob (Y/X)} \\ &= \frac{13}{52} \times \frac{12}{51} \\ &= \mathbf{0.059}\end{aligned}$$

4.0 CONCLUSIONS

The concept of probability is highly used in statistical analysis. The outcome of most statistical analysis is to determine how likely it is that a particular set of result occurred merely by chance. Probability therefore is the foundation of all statistical inference so make sure you don't take anything about it as trivial.

5.0 SUMMARY

In this unit you have learnt basic concepts of probability. You also learnt what simple and compound probabilities mean and two ways of linking individual probabilities to form compound probability. Based on the linkage, simple laws of probability were also discussed.

6.0 TUTOR-MARKED ASSIGNMENT

In an integrated science class of 76 students, 8 are Chemistry-education; 3, physics-education; 39, biology-education and 26, mathematics-education students. What is the probability of drawing a:

- i) Biology-education students from the pool?
- ii) Chemistry or Physics-education student from the pool?

iii) Mathematics or Physics-education student from the pool?

7.0 REFERENCES/FURTHER READINGS

Adamu, S. O. (1997), *Statistics for Beginners Book 1*, Nigeria, Ibadan, SAAL Pub.

Hinton, P. R. (2003), *Statistics Explained A Guide for Social Science Students*, Hove and New York, Routledge

Owie, I. (2006), *Fundamentals of Statistics in Education and the Social Sciences*, Nigeria, Lagos, National Book Consortium

<http://www.gower.k12.il.us/Staff/ASSESS>

www.apsu.edu/oconnort/3760/3760lect03a.htm

faculty.vassar.edu/lowry/webtext.html

www.uwsp.edu/psych/stat/14/nonparm.htm

writing.colostate.edu/guides/research/stats/pop2a.cfm

www.shodor.org/interactivate/lessons/IntroStatistics

www.statbasics.co.uk

UNIT 2: PROBABILITY SAMPLING DISTRIBUTION

CONTENTS

- 1.0 Introduction
- 2.0 Objectives
- 3.0 Main Content
 - 3.1 Types of Distributions
 - 3.1.1 Population Distribution
 - 3.1.2 Sample Distribution
 - 3.1.3 Relationship between Sample and Population Distributions
 - 3.2 Sampling Distribution
 - 3.2.1 Binomial Probability Sampling Distribution
 - 3.2.2 Parameters of Binomial Probability Sampling Distribution
 - 3.2.3 Importance of Sampling Distribution
- 4.0 Conclusions
- 5.0 Summary
- 6.0 Tutor-Marked Assignment
- 7.0 References/Further Readings

MAIN CONTENTS

1.0 INTRODUCTION

In module 1, units 2, you learnt various ways of gathering, describing and distributing data or measures. You also learnt of the various procedures of drawing samples from populations. In this unit we will expand the idea into different types of distributions and the relationship between them.

2.0 OBJECTIVES

At the end of this unit you should be able to:

- Define distribution formally
- List the different types of distribution
- State the relationships between the different types of distributions
- Give the importance of sampling distributions

3.0 MAIN CONTENT

3.1 Types of Distribution

You will recall that we considered variables as attributes that can assume two or more values. You have come across various variables such as sex, age, height, weight, achievement etc. Sex can either be male or female; age can be in form of ranges or by years such as 11 years, 15 years etc. You can represent height by tall or short or by measures such as 1.5 meters etc. In module 1, unit 2, we constructed frequency distribution tables and drew graphs to represent these distributions. In this section attempt will be made to expand on the meaning and type of distribution.

If you go back to module 1 unit 2 you will find distributions of ungroup and group data. From the distributions you simply observe how many cases or observations were seen in each class or category of your variable. With variable sex, the distribution will tell you how many males and how many females were seen. This then means that distribution can formally be defined as:

A statement of the frequency with which units of analysis (or cases) are assigned to the various classes or categories that makes up a variable.

There are three major types of distributions – Population distribution, Sample distribution and Sampling distribution.

3.1.1 Population Distribution

In module 1 unit 3 you learnt of population and sample. You learnt that population covers all the units of study or the universe. For example if in a study you are interested in second year middle basic education student in Nigeria, your population will be all the male and female students in junior secondary schools all over the 36 states and federal capital territory of Nigeria. If in the study you wish to determine those who approve or disapprove the introduction of the new curricula, then you have to observe the frequencies in each category. By implication, population distribution is defined as:

A statement of the frequency with which the units of analysis or cases that together make up a population are observed or are expected to be observed in the various classes or categories that make up a variable.

However, it is not always possible to carry out a census. Consequently it is not possible to study all members of the population. Often a sub-set of members of the population is selected for study. This sub-set you studied in the previous section is known as sample. The sample can also be distributed.

3.1.2 Sample Distribution

In module 1 unit 3 you also learnt that we can only study a population when it is small and the subjects few. You also learnt that to study the entire population requires carrying out a census, which often is not easy. Consequently representative sample(s) of the population is used. To have a representative sample means to randomly choose members of the sample. This is done through methods that give equally opportunity to all members of the population to be selected. A formal definition of sample is:

A sample distribution is a statement of the frequency with which the units of analysis or cases that together make up a sample are actually observed in the various classes or categories that make up a variable.

Considering the above discussion you will agree that population distribution represents the “total information” which you can get from measuring a variable, while the sample distribution represents an estimate of this information.

3.1.3 Relationship between Sample and Population Distributions

Let us use examples to illustrate some important features of sample distributions and their relationship to a population distribution. Let us assume we have a population which consists of only seven units of analysis - seven principals, each of whom had different numbers of conversations about class management with the classroom teachers. Table 2.1 is the data array along with the population parameters.

Table 2.1: Parameters of a Fictitious Population (N = 7)

Principal	X	X-M	(X -M) ²
1	4	-3	9
2	5	-2	4
3	6	-1	1
4	7	0	0
5	8	1	1
6	9	2	4
7	10	3	9
	$\sum X = 49$		$\sum (X -M)^2 = 28$
Population Mean (M) = $\sum X/N = 49/7 = 7$			
Population Variance (σ) = $\sum (X -M)^2/N = 28/7 = 4$			
Population Standard Deviation (σ^2) = $\sqrt{4} = 2$			

Let us consider three sample statistics from the above population and see how they estimate the population parameters. A summary of the statistics is presented in the Tables 2.2, 2.3 and 2.4.

Table 2.2: Descriptive Statistics of First Three-Samples (N = 4)

	X	X-M	(X -M)²
Sample 1	5	-2	4
	6	-1	1
	8	1	1
	9	2	4
	$\sum X = 28$		$\sum (X -M)^2 = 10$
Sample Mean (M) = $\sum X/N = 28/4 = 7$			
Sample Variance (σ) = $\sum (X -M)^2/N = 10/4 = 2.5$			
Sample Standard Deviation (σ^2) = $\sqrt{2.5} = 1.58$			

Table 2.3: Descriptive Statistics of Second Three-Samples (N = 4)

	X	X-M	(X -M)²
Sample 2	7	-1.5	2.25
	8	-0.5	0.25
	9	0.5	0.25
	10	1.5	2.25
	$\sum X = 34$		$\sum (X -M)^2 = 5$
Sample Mean (M) = $\sum X/N = 34/4 = 8.5$			
Sample Variance (σ) = $\sum (X -M)^2/N = 5/4 = 1.25$			
Sample Standard Deviation (σ^2) = $\sqrt{1.25} = 1.12$			

Table 2.4: Descriptive Statistics of Third Three-Samples (N = 4)

	X	X-M	(X -M)²
Sample 3	5	-2.75	7.56
	7	-0.75	0.56
	9	1.25	1.56
	10	2.25	5.06
	$\sum X = 31$		$\sum (X -M)^2 = 14.74$
Sample Mean (M) = $\sum X/N = 31/4 = 7.75$			
Sample Variance (σ) = $\sum (X -M)^2/N = 10/4 = 2.5$			
Sample Standard Deviation (σ^2) = $\sqrt{2.5} = 1.58$			

Tables 2.2, 2.3 and 2.4 show three of the many different sample distributions that can be obtained from the population of principals. You will observe that values of the three samples are not identical, the means, variances, and standard deviations are different among the samples, as well. You will also observe that the sample statistics are not identical to the population parameters. These sample statistics are said to be only estimates of the population parameters. We have distinguished between these estimated values and the actual descriptive values of the population by using the term parameter for population and statistic for sample. The meaning of the term statistic is parallel to the meaning of the term parameter: they both characterize distributions. Different symbols are used for population parameters and sample statistics and this is discussed in Module 1. It is important to note that a characteristic of statistics is that their values are always known. That is, if we draw a sample we will always be able to calculate statistics which describe the sample distribution. This however is not the same with parameters, which may or may not be known, depending on whether we have census information about the population.

At this point let us contrast the statistics computed from the three sample distributions with the parameters from the corresponding population distribution as presented in the Table 2.5.

Table 2.5: Descriptive Statistics from Three Samples of the same Population

	Mean	Variance	Standard Deviation
Population	7	4	2
Sample 1	7	8.5	7.75
Sample 2	2.50	1.25	2.50
Sample 3	1.58	1.12	1.58

You will observe from this table that the values for the mean, the variance, and the standard deviation in each of the samples are different. This difference between the population parameter value and the equivalent sample statistic indicates the error we make when we generalize from the information provided by a sample to the actual population values. This brings us to the third type of distribution known as **Sampling Distribution**.

3.2 Sampling Distribution

From our discussion so far, samples can also be distributed. Sampling distribution describes how sample results behave if you draw repeated samples of a particular size ‘n’ from a larger population. It is a statement of the frequency with which the units of analysis or cases that together make up a sample are actually observed in the various classes or categories that make up a variable. Sampling distribution is a kind of “second-order” distribution. Whereas the population distribution and the sample distribution are made up of data values, the sampling distribution is made up of values of statistics computed from a number of sample distributions. We can now define sampling distribution as:

A statement of the frequency with which values of statistics are observed or are expected to be observed when a number of random samples is drawn from a given population.

It is extremely important that you have a clear distinction between the concepts of sample distribution and of sampling distribution. A sample distribution refers to the set of scores or values that you obtain when you apply the operational definition to a subset of units chosen from the full population. Such a sample distribution can be characterized in terms of statistics such as the mean, variance, or any other statistic. A sampling

distribution emerges when you sample repeatedly and record the statistics that you observe. After a number of samples have been drawn, and the statistics associated with each computed, you can construct a sampling distribution of these statistics.

3.2.1 Binomial Probability Sampling Distribution

In unit 1 of this module you learned of the theory of probability. In this section we will further consider the idea of probability. Let us once again consider the example of coin tossing. If you toss 2 coins with the probability of head (H) occurring as 0.5, you represent such information as $p = 0.5$. Similarly the probability of head not occurring is $1 - p$. This is represented as q . So q is also 0.5. Table 2.6 presents the conjunctive outcome and the associated probabilities when the 2 coins are tossed.

Table 2.6: Conjunctive Outcome and the Associated Probabilities of Tossing Two Coins

Coin		Sub-Pathway Probability	Number of Heads	Main Pathway Probability
A	B			
-	-	$0.5 \times 0.5 = 0.25$	0	0.25 (25%)
-	H	$0.5 \times 0.5 = 0.25$	1	$0.25 + 0.25 = 0.50$ (50%)
H	-	$0.5 \times 0.5 = 0.25$	1	
H	H	$0.5 \times 0.5 = 0.25$	2	0.25 (25%)
TOTAL				1.0 (100%)

Information in Table 2.6 shows that:

- There are two identical trials
- Each trial results in one of two outcomes (Head or no Head or success or failure)
- The probability of Head or success on a single trial is equal to 'p' and remains the same for trial to trial
- The probability of no head or failure is 'q' which is $1-p$
- The trials are independent
- We are interested in the number of successes observed during the 'n' trials.

Such an experiment is known as **Binomial**. You can interpret the results in Table 2.6 as follows:

- The coin toss example has 2 coins ($n = 2$)

- The probability of head occurring is 0.5 ($p=0.5$) while the probability of head not occurring is 0.5 ($q=0.5$).

The sampling distribution therefore specifies that for any 2 randomly selected sample cons, there is 25% chance of including zero head, 50% chance of including 1 head and 50% chance of including 2 heads. It then follows that for such sample there is 75% (I e 50% + 25%) chance of including **at least 1 head**.

3.2.2 Parameters of Binomial Probability Sampling Distribution

Like all sampling distributions, binomial sampling distributions has parameters. These are:

- Measure of central tendency – this is about the mean of the distribution
- Measure of variability – which is the measure of the tendency of individual outcomes to be dispersed away from the average. Common among the measures are variance and standard deviation

Calculation of Parameters of Binomial Probability Sampling Distribution

Mean is represented by the symbol μ (lower case Greek letter ‘mu’) when the entire population is involved and σ (lower case Greek letter ‘sigma’) represents variance and standard deviation.

$$\text{Mean: } \mu = Np$$

$$\text{Variance: } \sigma^2 = Npq$$

$$\text{Standard deviation: } \sigma = \sqrt{Npq}$$

Considering our coin toss example:

$$\text{Mean: } \mu = 2 \times 0.5$$

$$= 1.0$$

$$\text{Variance: } \sigma^2 = 2 \times 0.5 \times 0.5$$

$$= 2 \times 0.25$$

$$= 0.50$$

$$\text{Standard Deviation: } \sigma = \sqrt{0.5}$$

$$= \pm 0.71$$

3.2.3 Importance of Sampling Distribution

Sampling distributions:

- **Give insight into sampling error**

If you look at the tables on sampling distributions above, you will observe that it is quite possible that there will be differences between sample characteristics and population characteristics. This means that when you draw random samples from a population there are no guarantees that the samples will indeed be exactly representative of the population. The discrepancy between the parameter of a population and the corresponding statistic computed for a sample drawn randomly from that population is known as sampling error.

- **Give insight into probability**

Perhaps the most important reason for constructing sampling distributions is that they can be interpreted in terms of the probability associated with the various statistics that are obtained when we sample randomly. The probability of any outcome is simply the ratio of that outcome's frequency to the total frequency. You will learn more of probability in our next unit.

- **Allow to test hypotheses.**

Using sampling distributions, you can assign probabilities to the various values of means obtained and probability statements are critical for hypothesis testing. For instance, if the mean of a population is 7.0 and if we were to draw a sample of 2 observations out of the population and if we then obtained a mean of 5.00 from this sample, we would consider this mean a relatively unusual event because according to the sampling distribution based on samples of this size, a mean of 5.00 had only a 4% chance of occurring. We would have expected to observe a value like 6.5, or 7, or 7.5, because these outcomes are much more likely to occur. If we obtain a mean of 5.00, we can say that it is relatively unlikely (but not impossible) that the population mean is really 7.0. Such statements led to hypothesis testing.

4.0 CONCLUSIONS

Sampling distribution is a device by which researchers rationally determine how confident they may be that their observed results reflect anything more than mere chance coincidence.

5.0 SUMMARY

In this unit we have learnt the different types of distribution – population, sample and sampling distribution. We have also tried to show the relationships between the different distribution types as well as the importance of each

6.0 TUTOR-MARKED ASSIGNMENT

A coin is thrown twice. The sample space is
 $S = (HH, HT, TH, TT)$
 $A = (HH, HT)$, $B = (HT, TT)$
Find the probability of A, B, AB and A + B

7.0 REFERENCES/FURTHER READINGS

Adamu, S. O. (1997), *Statistics for Beginners Book 1*, Nigeria, Ibadan, SAAL Pub.

Hinton, P. R. (2003), *Statistics Explained A Guide for Social Science Students*, Hove and New York, Routledge

Owie, I. (2006), *Fundamentals of Statistics in Education and the Social Sciences*, Nigeria, Lagos, National Book Consortium

<http://www.gower.k12.il.us/Staff/ASSESS>

www.apsu.edu/oconnort/3760/3760lect03a.htm

faculty.vassar.edu/lowry/webtext.html

writing.colostate.edu/guides/research/stats/pop2a.cfm

www.shodor.org/interactivate/lessons/IntroStatistics

www.statbasics.co.uk

UNIT 3 STATISTICAL INFERENCE

CONTENTS

- 1.0 Introduction
- 2.0 Objectives
- 3.0 Main Content
 - 3.1 Meaning of Statistical Inference
 - 3.1.1 Statistical Significance
 - 3.1.2 Significant Level
 - 3.2 Statistical Significance Testing
 - 3.2.1 Types of Hypothesis
 - 3.2.2 Two-tailed and One-tailed Tests
 - 3.2.3 Errors in Statistical Testing
- 4.0 Conclusions
- 5.0 Summary
- 6.0 Tutor-Marked Assignment
- 7.0 References/Further Readings

1.0 INTRODUCTION

In the previous units you learnt that using sample data one can calculate sample statistics and that because it is not always easy to study the entire population, samples are used to estimate the population parameters. You also learnt that one importance of statistics is for hypothesis testing. In this unit you will learn more about hypothesis testing. Types of hypothesis and errors due to statistical testing will also be considered.

2.0 OBJECTIVES

At the end of this unit you should be able to:

- Formulate null hypothesis
- Formulate alternative hypothesis
- Differentiate directional from non-directional hypothesis
- Define hypothesis testing
- List steps in hypothesis testing
- Identify errors due to statistical testing
- Differentiate between type 1 and type 2 errors

3.0 MAIN CONTENT

3.1 Meaning of Statistical Inference

Let us start with a scenario in which your friend reaches a cartoon on a table and then hands over to you a wrap of vicks blue sweet. What would be you thinking? Many ideas will lay before you. You may think that your friend brought the sweet from the box. You may also think that the sweet was hidden in the palm before reaching out for the box. Several other thoughts will come to play. Assuming your friend again dips the hand into the cartoon and brings out more blue sweets you will become better convinced that the vicks blue sweets come from the cartoon and that the content of the cartoon is vicks blue. But suppose on another draw what came out is lemon yellow plus sweet, then your first thought may be modified. You may find at this point that your confidence of getting vicks blue is not sure.

In our previous units we considered the process of calculating the probability of getting vicks blue or lemon vicks. This is by probability sampling techniques ensuring that the samples are drawn randomly. From the various samples we then infer or generalize about the population. **Statistical inference is any procedure by which one generalizes from sample data to the population.**

3.1.1 Statistical Significance

"Significant" in normal English means 'important' however, in statistics it means '**probably true**' (not due to chance). A research finding may be true without being important. When statisticians say a result is "highly significant" they mean it is very probably true. They do not (necessarily) mean it is highly important. Significance is a statistical term that tells how sure you are that a difference or relationship exists.

3.1.2 Significant Level

The amount of evidence required to accept that an event is unlikely to have arisen by chance is known as the **significance level** or **critical p-value**. Significance levels show you how likely a result is due to chance. In the field of Education, the most common level, used to mean something is good enough to be believed, is .95. This means that the finding has a 95% chance of being true. However, this value is also used in a misleading way. No statistical package will show you "95%" or ".95" to indicate this level. Instead it will show you ".05," meaning that the finding has a five percent

(.05) chance of not being true, which is the converse of a 95% chance of being true. To find the significance level, subtract the number shown from one. For example, a value of ".01" means that there is a 99% ($1-.01=.99$) chance of it being true.

3.2 Statistical Hypothesis Testing

In the last few sections you have learnt of statistical inference and significant levels. The philosophy of making a decision about statistical significance has resulted in the practice of **hypothesis testing**. Hypothesis testing has become so important in research that in most departments research students are required to list the hypotheses to be tested in their projects.

The hypotheses are often statements about population parameters like expected value and variance. A hypothesis might also be a statement about the distributional form of a characteristic of interest. They are simply informed or intelligent guess about the solution to a problem. To understand more about hypothesis it will be necessary to consider the different types.

3.2.1 Types of Hypothesis

Hypothesis can be classified as either **research hypothesis** or **statistical hypothesis**. When postulations are about the relationships between two or more variables it is known as **research hypothesis**. Such a hypothesis is not directly testable statistically because it is not expressed in measurable terms. For an example 'Students poor performance in science is due to use of inadequate methods of teaching the subject. However, it is possible to express the relationship between two or more variables in statistical and measurable terms. This is known as **statistical hypothesis**. Statistical hypothesis is formulated in two forms – **null and alternative hypothesis**. A null hypothesis is a hypothesis of no difference, no relationship or no effect while the alternative hypothesis specifies any of the possible conditions not anticipated in the null Hypothesis (Nworgu, 1991). The null hypothesis is represented as H_0 while the alternative is H_1 or H_a .

Examples:

H_0 = There is no significant difference in the mean scores in science of students taught using inquiry method and those taught using no inquiry method

H₁ or H_a = There is a significant difference in the mean score in science of students taught using inquiry and non-inquiry methods.

Alternative hypothesis can further be described as **Non-directional and Directional Alternatives**. When an alternative hypothesis gives the direction of the difference or effect it is known as a **Directional Alternative**, but when direction is not given it is non-directional.

Example:

Directional Alternative: Boys taught using inquiry method has higher mean score in science than girls taught using the same method of teaching is Directional Alternative.

Non-Directional Alternative: The mean score in science for boys taught

Steps in Hypothesis Testing

Hypothesis testing is the use of statistics to determine the probability that a given hypothesis is true. The usual process of hypothesis testing consists of four steps :

1. Formulate the **null hypothesis** and the **alternative hypothesis**.
2. Identify a **test statistic** that can be used to assess the truth of the **null hypothesis**.
3. Compute the **P-value**, which is the probability that a test statistic at least as significant as the one observed would be obtained assuming that the null hypothesis were true (the smaller the *P*-value, the stronger the evidence against the null hypothesis).
4. Compare the *p*-value to an acceptable significance value α (sometimes called an **alpha value**). If $p \leq \alpha$, that the observed effect is statistically significant, the null hypothesis is ruled out, and the alternative hypothesis is valid.

3.2.2. Two-tailed and One-tailed Tests

One important concept in significance testing is whether you use a one-tailed or two-tailed test of significance. The answer is that it depends on your hypothesis. When your hypothesis states the direction of the difference or relationship, you use a one-tailed probability. For example, a one-tailed test would be used to test these null hypotheses: Females will not score significantly higher than males on an IQ test. Blue collar workers are will not buy significantly more product than white collar workers. Superman is not significantly stronger than the average person. In each

case, the null hypothesis (indirectly) predicts the direction of the difference. A two-tailed test would be used to test these null hypotheses: There will be no significant difference in IQ scores between males and females. There will be no significant difference in the amount of product purchased between blue collar and white collar workers. There is no significant difference in strength between Superman and the average person. The one-tailed probability is exactly half the value of the two-tailed probability. There is a raging controversy (for about the last hundred years) on whether or not it is ever appropriate to use a one-tailed test. The rationale is that if you already know the direction of the difference, why bother doing any statistical tests. While it is generally safest to use two-tailed tests, there are situations where a one-tailed test seems more appropriate. The bottom line is that it is the choice of the researcher whether to use one-tailed or two-tailed research questions.

3.2.3 Errors in Statistical Testing

We define a *type I error* as the event of **rejecting the null hypothesis when the null hypothesis was true** (i.e. H_0 is wrongly rejected). For example, in a trial of a new instructional method the null hypothesis might be that the new method is no better, on average, than the current method; that is H_0 : there is no difference between the two methods of teaching. A type I error would occur if we concluded that the two methods produced different effects when in fact there was no difference between them. The probability of a type I error (α) is called the significance level. A **type II error** (with probability β) is defined as the event of **failing to reject the null hypothesis when the null hypothesis was false**. A type II error is frequently due to sample sizes being too small. Table 2.7 gives a summary of possible results of any hypothesis test:

Table 2.7: Summary of Statistical Decisions and Types of Errors

		Decision	
		Reject H_0	Don't reject H_0
Truth	H_0	Type I Error	Right Decision
	H_1	Right Decision	Type II Error

4.0 CONCLUSIONS

The purpose of collecting and analysing data is to find a reasonable and objective criterion for deciding on a proper line of action. In most statistical investigation we either accept or reject the hypothesis. What you have

studied in this unit will help you to the right decision to make in a given situation.

5.0 SUMMARY

In this unit we have discussed different types of hypothesis and how to formulate these types of hypothesis. You also learnt conditions for making different decisions and the types or errors when wrong decisions are made.

6.0 TUTOR-MARKED ASSIGNMENT

A panel is set up to try students involved in examination malpractice. The panel decides to set up a hypothesis to aid the decision concerning cases brought before it. How best can this be done?

7.0 REFERENCES/FURTHER READINGS

Adamu, S. O. (1997), *Statistics for Beginners Book 1*, Ibadan: SAAL Pub

Hinton, P. R. (2003), *Statistics Explained A Guide for Social Science Students*, Hove and New York: Routledge

Kerlinger, F. (1986), *Foundations of Behavioural Research*, New York: McGraw-Hill

Owie, I. (2006), *Fundamentals of Statistics in Education and the Social Sciences*, Lagos: National Book Consortium

Nworgu, B. G. (1991), *Educational Research: Basic Issues and Methodology*, Ibadan: Wisdom
<http://www.gower.k12.il.us/Staff/ASSESS>

www.apsu.edu/oconnort/3760/3760lect03a.htm

faculty.vassar.edu/lowry/webtext.html

www.uwsp.edu/psych/stat/14/nonparm.htm

writing.colostate.edu/guides/research/stats/pop2a.cfm

www.shodor.org/interactivate/lessons/IntroStatistics

www.statbasics.co.uk

MODULE 3: PARAMETRIC TECHNIQUES FOR TESTING HYPOTHESIS ABOUT THE MEAN(S) OF ONE OR MORE POPULATIONS

Unit 1: t-tests

Unit 2: Analysis of Variance (ANOVA)

Unit 3: Analysis of Covariance (ANCOVA)

UNIT 1: t-Tests

CONTENTS

- 1.0 Introduction
- 2.0 Objectives
- 3.0 Main Content
 - 3.1 T-Test
 - 3.1.1 Origin of t-variable
 - 3.1.2 Assumptions of t-test
 - 3.2 Types of t-test
 - 3.2.1 One Sample t-test
 - 3.2.2 Independent Samples T-test
 - 3.2.3 Paired Samples t-tests
 - 3.3 Interpretation of Computer Output of a t-test analysis
- 4.0 Conclusions
- 5.0 Summary
- 6.0 Tutor-Marked Assignment
- 7.0 References/Further Readings

1.0 INTRODUCTION

In unit 1 you learnt about population parameters and sample statistics. You have also learnt from probability that a random sample will tend to reflect the properties of the population from which it is drawn. This means that the mean of any particular sample can be taken as an unbiased estimate of the population mean. When you estimate a population parameter by a single number, it is called **point estimate** of that parameter. Consequently you can use sample mean (\bar{x}) as a point estimate for the population mean (μ). You can also use sample standard deviation (s) as a point estimate for the population standard deviation (σ). In this module you will learn about those techniques that are used to explore differences between means. Specifically you shall learn of parametric techniques used for testing hypotheses about the means of one or more populations.

2.0 OBJECTIVES

At the end of this unit you will be able to:

- Calculate one sample t-tests
- Calculate paired sample t-tests
- Calculate independent sample t-test
- Calculate effect size, and
- Interpret computer outputs on t-tests analyses

3.0 Main Content

3.1 t-Test

When sample size is 30 or more, we can approximate the population parameters with the sample statistics. Unfortunately, in many practical and important situations large samples are not simple available. Does it mean that in such situations we can not approximate population parameters with the sample statistics? Before we answer this question it is important to note that with small samples errors are likely to occur. So estimating population parameters with the sample statistics may introduce these errors. To avoid the errors involved a new variable called the student's t-variable has been introduced. The test statistic is known as student's t-test.

3.1.1 Origin of t-variable

W. S. Gosset discovered the distribution of samples drawn from a normally distributed population. His pen name 'Student' was used to publish the work in 1908 while he was still a staff of Guinness and was working on stout. He referred to the quantity under study as 't' and it has ever since been known as 'student's t'.

The t variable is defined by the following formula:

$$t = \frac{\bar{x} - \mu}{\frac{s}{\sqrt{n}}}$$

Where \bar{x} is the sample mean

μ is the population mean

To use the t-test, you must have one dependent variable and one independent variable. The independent variable must be of two levels only.

3.1.2 Assumptions of t-test

The following are the assumptions that must be satisfied before the t-test can be used:

- The Level of Measurement for the dependent variable must be Continuous. That is, the dependent variable is measured at the interval or ratio level rather than discrete categories.
- The sampling must be random. That is, the scores are obtained using a random sample from the population.

When these assumptions are violated t-test cannot be applied.

3.2 Types of t-test

We will discuss three types of t-tests. These are:

- One sample t-test
- Independent samples t-test, and
- Paired sample t-test

We will now consider each type in some detail.

3.2.1 One Sample t-test

This is typically used to find out whether the mean of a sample drawn from a normal population deviates significantly from a stated value (the hypothetical value) of the population mean.

The formula for one sample t-test is given as:

$$t = \frac{(\bar{x} - \mu)\sqrt{N}}{S}$$

Where \bar{x} = Sample mean
 μ = Population
 S = Standard deviation of sample
 N = Number of items in sample

Let us consider an example. Ten students are chosen at random from a population and their scores in English are found to be 63, 63, 66, 67, 68, 69,

70, 70, 71 and 71. Considering the data discuss the suggestion that the mean score in English in the population is 66.

Before you begin your test, it is important to follow the following steps:

Step 1: State the Hypothesis.

In this example the hypothesis can be stated as:

“The English mean score in the population is 66”.

Step 2: Calculate the sample mean (\bar{x}) and standard deviation (s) as illustrated in Table 3.1.

Table 3.1: Steps in Calculating Sample Mean and Standard Deviation

Students	Scores in English (X)	Deviation from the Mean (67.8) (x)	Deviation from the Mean Squared (x ²)
1	63	-4.8	23.04
2	63	-4.8	23.04
3	66	-1.8	3.24
4	67	-0.8	.64
5	68	0.2	.04
6	69	1.2	1.44
7	70	2.2	4.84
8	70	2.2	4.84
9	71	3.2	10.24
10	71	3.2	10.24
N = 10	$\sum X = 678$		$\sum x^2 = 81.6$

$$\bar{X} = \frac{678}{10} = 67.8$$

$$S = \sqrt{\frac{81.6}{10-1}} = 3.01$$

Step 3: Substitute the values of \bar{X} , S and N into the formula to calculate ‘t’

$$t = \frac{(\bar{x} - \mu)\sqrt{N}}{S} = \frac{(67.8 - 66)\sqrt{10}}{3.01} = 1.9$$

Step 4: Interpretation of result

To interpret this result you need to compare the calculated t with a critical t-value also known as table t-value. To determine the table t-value, you need you read off the value of the degree of freedom against the level of significance from a table of t-distribution which can be found in any good statistical table.

If you look at the values in this table, the t-value for our example above which has $df = 9$ at 5% (.05) significant level, is 2.262. The calculated t (1.9) is less than the table t (2.262), hence our hypothesis that the population mean is 66 is retained.

Example 2: Let us consider a situation where you wish to compare the reading achievement of boys and girls. For example, a group of 17 SSI students are assigned to a new style of science teaching. Their performance before entering the new classroom was average, though as a result of the new style of instruction we believe they should score higher than average on tests of science knowledge if the new curriculum is working. After 6 weeks in the classroom they are given a test assessing their knowledge of basic science concepts to see if the teaching has been effective. The result is $\bar{X} = 84$, $SD = 16$, $N = 17$. Previous research has shown that SSI students in general score an average of 78 on this exam ($N = 78$). Is the difference we observed after 6 weeks of instruction consistent with what is likely under conditions of chance above or does it reflect a difference?

- **Research Hypothesis:**
The new teaching style resulted in better test performance than average SSI students.
- **Statistical Hypothesis:**
SSI students do not differ in knowledge of science concepts from the average performance $H_0: \mu_1 = \mu_2$

$$t = \frac{(\bar{x} - \mu)\sqrt{N}}{S} = \frac{(84 - 78)\sqrt{17}}{16} = 1.5$$

The table 't' value is 2.12. The calculated t-value (1.5) is lower, so we fail to reject H_0 i.e. we accept H_0 .

Interpretation of Computer Output of One-Sample t-test

The computer output for a one sample t-test is given below:

One Sample t-test

One-Sample Statistics

	N	Mean	Std. Deviation	Std. Error Mean
Post-test Integrated Science Achievement Score	48	36.96	6.813	.983

One-Sample Test

	Test Value = 100					
	t	df	Sig. (2-tailed)	Mean Difference	95% Confidence Interval of the Difference	
					Lower	Upper
Post-test Integrated Science Achievement Score	-64.103	47	.000	-63.04	-65.02	-61.06

There are two table titled One-Sample Statistics and One-Sample Test. In the one sample t-test the mean of the variable is compared with a hypothesised (or test value). From the table titled **One-Sample Test** you observe the column marked **Sig. (2-tailed)**. This is the probability value. You conclude that there is a significant difference between the two scores if the probability value is equal or less than .05. If the value is larger than .05 you conclude that there is no significant difference between the two scores.

3.2.2 Independent Samples T-test

This is used when you have **two different** groups (known as independent groups) and you want to compare the groups' scores. In this case you only collect data on one occasion from the two different groups and the scores must be continuous data.

What Does T-test for Independent Samples tell you?

The test will tell you whether:

- There is a statistically significant difference in the mean scores for the 2 groups (e.g. do males differ significantly from females in the measured variable?)

- Statistically you are testing the probability that the two sets of scores came from the same population.

The formula for Independent Sample t-test is given as:

$$t = \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}}$$

Where:

$$S\Sigma = \sqrt{\frac{S_1^2}{N_1} + \frac{S_2^2}{N_2}} \quad \text{is known as sum of squares.}$$

Further Assumptions of Independent t-test

In addition to the general assumptions of the t-test, the following are assumptions of the independent t-test:

- Independence of observations - the observations that make up your data must be independent of one another. This means that your observations or measurements should not be collected in a group setting. (If you suspect that this assumption is violated. Sterens 1996 p. 241 recommends setting stringent alpha value of $p < 0.01$).
- Normal Distribution – the population from which the samples are taken is assumed to be normally distributed. To avoid violation of this assumption the sample size must be large (e.g. 30+).
- Homogeneity of Variance – samples are obtained from populations of equal variances. This means that the variability of scores for each of the groups is similar.

Calculating ‘t’ for Independent Samples

The formula for calculating t for two independent sample means is:

$$t = \frac{\bar{x}_1 - \bar{x}_2}{S \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}} \quad \text{or} \quad \frac{\bar{x}_1 - \bar{x}_2}{S} \sqrt{\frac{n_1 n_2}{n_1 + n_2}}$$

Where:

\bar{x}_1 and \bar{x}_2 = means of the first and second samples respectively

n_1 and n_2 are the number of items in the first and second samples respectively

S = stand deviation between the difference between mean of two samples

$$S = \sqrt{\frac{\sum (x_1 - \bar{x}_1)^2 + \sum (x_2 - \bar{x}_2)^2}{N_1 + N_2 - 2}}$$

Test of difference between the means of two samples

$$t = \frac{\bar{x}_1 - \bar{x}_2}{S \sqrt{\frac{1}{N_1} + \frac{1}{N_2}}}$$

$$= \frac{\bar{x}_1 - \bar{x}_2}{S} \sqrt{\frac{N_1 N_2}{N_1 + N_2}}$$

S = Standard deviation of the difference between the two samples

$$S = \sqrt{\frac{\sum (x_1 - \bar{x}_1)^2 + \sum (x_2 - \bar{x}_2)^2}{N_1 + N_2 - 2}}$$

$$= \sqrt{\frac{\sum x_1^2 - \frac{(\sum x_1)^2}{N_1} + \sum x_2^2 - \frac{(\sum x_2)^2}{N_2}}{N_1 + N_2 - 2}}$$

Where x_1 and x_2 are squares of deviations of the 2 samples

Example

A new method of teaching science acquisition skills is to be introduced into the schools. Two groups of nine teachers were trained for a period of three weeks one group on the use of the new method and the other following the standard method of teaching. The degree of acquisition of science skills were recorded at the end of the three weeks for the two groups of teachers and are summarized in Table 3.1 below.

New Procedure	Standard Procedure
32	35
37	31
35	29
28	25
41	34
44	40
35	27
31	32
34	31

Do the data present sufficient evidence to indicate that the degree of acquisition of science skills at the end of the three weeks was high for the new method of teaching?

Solution

- Let μ_1 and μ_2 equal the degree of acquisition of science skills for the new and standard teaching methods respectively.
- The sample means and sum of squares of deviations are:
 $\bar{X}_1 = 35.22$ SS of deviation = 195.56
 $\bar{X}_2 = 31.56$ SS of deviation = 160.22

$$S = \sqrt{\frac{\sum (x_1 - \bar{x}_1)^2 + \sum (x_2 - \bar{x}_2)^2}{N_1 + N_2 - 2}}$$

$$S = \sqrt{\frac{195.56 + 160.22}{9 + 9 - 2}}$$

$$S = \sqrt{22.24}$$

$$t = \frac{\bar{x}_1 - \bar{x}_2}{S \sqrt{\frac{1}{N_1} + \frac{1}{N_2}}}$$

$$t = \frac{35.22 - 31.56}{4.72 \sqrt{\frac{1}{9} + \frac{1}{9}}}$$

$$= 1.64$$

Interpretation of Computer Output for Independent-Sample t-test

Several computer programmes are used for data analysis. Common among them is the Statistical Package for Social Sciences (SPSS) and any other spread sheet programme. However in we will consider the use of SPSS in this unit. An example of the output generated by the computer for independent-sample t-test using SPSS is presented below:

t-Test

Group Statistics

	Sex	N	Mean	Std. Deviation	Std. Error Mean
Integrated Science Achievement Score	Males	24	30.83	4.631	.945
	Females	24	30.33	4.556	.930

Independent Samples Test

		Levene's Test for Equality of Variances		t-test for Equality of Means						
		F	Sig.	t	df	Sig. (2-tailed)	Mean Difference	Std. Error Difference	95% Confidence Interval of the Difference	
									Lower	Upper
Integrated Science Achievement Score	Equal variances assumed	.000	.987	.377	46	.708	.50	1.326	-2.169	3.169
	Equal variances not assumed			.377	45.988	.708	.50	1.326	-2.169	3.169

Specifically, two tables are generated; one named **Group Statistics** and the other **Independent Sample test**. In the group statistics box, the number of subjects in each group, the mean, standard deviation and standard error of mean are given. In the independent sample test box, there are two sections. The first is the Levene's test for equality of Variance while the second part is the t-test results.

Interpretations of Output for Independent Sample t-test

We will discuss this in steps:

Step 1: Checking the information about the groups

You are expected to check the values from this box to ascertain that they are correct with respect to your sample size. Such information will enable you to determine when data are missing. If there are missing data you are expected to check the data that you fed into the computer. Perhaps you have entered something wrongly. The box also gives you information on the descriptive statistics on your sample. For example in the output presented above, there are 24 males and 24 females. The mean score for the males is 30.83 while that for females is 30.33. The standard deviation and standard error of mean are 4.631 and 0.945 respectively for males and 4.556 and 0.930 respectively for females.

Step 2: Checking Assumptions

You need to confirm that the assumptions for t-test for independent sample are not violated. To do this you check the first part of the independent sample test box marked Levene's Test. This tests whether the variance of scores for the groups is the same. The outcome of this test determines which of the t-values that SPSS provided is the correct one for you to use.

- If sig. value (second column of this test) is larger than 0.05, you use the first line in the table, which refers to **Equal variance assumed**.
- If sig. value is equal or less than 0.05, it means that the variance of the two groups are not the same. They are different. This means that your data violates the assumption of equal variance. In such a situation you are not expected to use the t-test. However, SPSS is able to provide you with an alternative t-value which compensates that the variances are not equal. So you use the information in the second line which refers to **Equal variances not assumed**.

Step 3: Assessing Differences between the Groups

To find out if there is a significant difference between the groups, you check from the second part of the box labelled **t-test for equality of means**. You refer to the column **Sig. (2 tailed)**. Two values are given. Considering the result from the Levene test, you choose which is appropriate.

If the Sig. (2-tailed) value is larger than .05 you concluded that there is no significant difference in the two mean scores. In the example output above, the sig. value from the Levene test is .987. This is larger than .05. It means that there is no significant difference in the variances of the two groups. That is to say, they have equal variance. So we use the information in the first line. Looking at the column Sig. (2-tailed), the value .708 is

above the required cut off of .05. You conclude that there is no significant difference between the two groups in the mean score on integrated science. If the Sig. (2-tailed) value is equal to or less than .05, you concluded that there is a significant difference in the two mean scores.

It is important to determine the magnitude of the differences between your groups. There are a number of different effect size statistics. However, the most commonly used is the '**eta squared**'. Eta squared can range from 0 to 1 and represents the proportion of variance in the

dependent variable that is explained by the independent variable. SPSS does not provide this value so you have to calculate it using the formula:

$$\text{Eta squared} = \frac{t^2}{t^2 + (N1 + N2 - 2)}$$

Although in the result presented in the output above, there is no significant difference, let us assume that there is so we can practice the calculation of eta squared.

Replacing with the appropriate values from the table above:

$$\begin{aligned}\text{Eta squared} &= \frac{.377^2}{.377^2 + (24 + 24 - 2)} \\ &= \frac{.142}{.142 + 46} \\ &= .003\end{aligned}$$

According to Cohen (1988), the guidelines for interpreting this result are:

- .01 = small effect
- .06 = moderate effect
- .14 = large effect

For our example .003 is very small. Expressed as percentage (i.e. multiply by 100) it is only .3%. Thus only .3% of the variance in integrated science score is explained by sex.

3.2.3 Paired Samples t-tests

This is also known as repeated measures. In this technique you have only one group of people and you collect data from them at two different occasions or under 2 different conditions (e.g. pre-test/post-test). This technique requires that you have one categorical independent variable measure at two different levels and the dependent variable which must be continuous also measured at two different occasions or under 2 different conditions.

$$t = \frac{\bar{x}}{s/\sqrt{n}} = \frac{\bar{x}\sqrt{n}}{s}$$

Where:

$$\bar{x} = \frac{1}{n} \sum x$$

$$s^2 = \frac{1}{n-1} \sum (xi - \bar{x})^2$$

3.3 Interpretation of Computer Output for Paired Sample t-test

In the previous section you learnt of the steps of interpreting t-test for independent sample. In this section you will also learn of steps to interpret the output for paired-sample t-test. Below is an example of a computer output for paired sample t-test.

Paired Sample Statistics Output

Paired Samples Statistics

		Mean	N	Std. Deviation	Std. Error Mean
Pair 1	Pre-test Integrated Science Achievement Score	30.58	48	4.552	.657
	Post-test Integrated Science Achievement Score	36.96	48	6.813	.983

Paired Samples Correlations

		N	Correlation	Sig.
Pair 1	Pre-test Integrated Science Achievement Score & Post-test Integrated Science Achievement Score	48	.034	.816

		Paired Differences				t	df	Sig. (2-tailed)	
		Mean	Std. Deviation	Std. Error Mean	95% Confidence Interval of the Difference				
					Lower				Upper
Pair 1	Pre-test Integrated Science Achievement Score - Post-test Integrated Science Achievement Score	-6.38	8.063	1.164	-8.72	-4.03	-5.478	47	.000

The output consists of three tables:

- Paired Sample Statistics – This table presents the descriptive statistics of the variables. These are the number of subjects for study, the mean, the standard deviation and the standard error of mean for each of the variable measured.
- Paired Sample Correlation – This provides the correlation between the two variables of study.
- Paired Sample test – This table provides the t-test statistics of the paired variables.

There are two steps involved in interpreting the results of this analysis.

Step 1: Determining Overall Significance

In the table labelled **Paired Sample Test** you look in the final column labelled **Sig. (2-tailed)**. This is the probability value. You conclude that there is a significant difference between the two scores if the probability value is equal or less than .05. If the value is larger than .05 you conclude that there is no significant difference between the two scores.

In our example above, the probability value is .000. This means that the actual probability value was less than .0005. This is smaller than .05. You can conclude that there is a significant difference between the two scores (pre and post test scores) in Integrated Science. It is important that you note the t-value and the degree of freedom (df) as you will need to quote the values when you make your report.

Step 2: Comparing Mean Values

Having established that there is significant difference from step 1, the next step is to find which of the set of scores is higher. You get this information from the table labelled **Paired Sample Statistics**. In the example here, the

post test is higher (6.813) than the pre-test score (4.552). You can then conclude that there is a significant increase in the post test score.

Using the formula and the procedure as discussed in the previous section you determine the effect size.

4.0 CONCLUSIONS

The introduction of t-distribution is very important. It should enhance your ability to make statistical inferences.

5.0 SUMMARY

In this unit we have discussed examples to illustrate the method of using t distribution for testing the significance of various results obtained from small samples. We have also discussed interpretation of computer output for the various types of t-tests.

6.0 TUTOR-MARKED ASSIGNMENT

(1) Nine students were chosen at random from a population and their weights were found to be (in pounds) 110, 115, 118, 120, 122, 125, 128, 130, and 139. In the light of these data, discuss the suggestion that the mean weight of the population is 120 pounds.

(2) Two laboratories measured the fat content of ice cream produced by a company. The results are:

Lab. A	7	8	7	3	8	6	9	4
	7	8						

Lab. B	9	8	8	4	7	7	9	6
	6	6						

Is there a significant difference between the mean fat content obtained by the two laboratories?

(3) The following data show ten students performance before and after an educational intervention:

Student	1	2	3	4	5	6	7	8
	9	10						

Before	24 23	26 23	20	21	23	30	32	25
After	26 24	26 25	22	22	24	30	32	26

Test whether there is any significant change in performance before and after the intervention.

7.0 REFERENCES/FURTHER READINGS

- Adamu, S. O. (1997), *Statistics for Beginners Book 1*, Nigeria, Ibadan, SAAL Pub.
- Cohen, J. (1988) *Statistical Power Analysis for the Behavioural Sciences*, Hillsdale, NJ: Erlbaum
- Hinton, P. R. (2003), *Statistics Explained A Guide for Social Science Students*, Hove and New York, Routledge
- Owie, I. (2006), *Fundamentals of Statistics in Education and the Social Sciences*, Nigeria, Lagos, National Book Consortium

<http://www.let.rug.nl/nerbonne/teach/Statistiek-I/>

<http://www.gower.k12.il.us/Staff/ASSESS>

www.apsu.edu/oconnort/3760/3760lect03a.htm

faculty.vassar.edu/lowry/webtext.html

www.uwsp.edu/psych/stat/14/nonparm.htm

writing.colostate.edu/guides/research/stats/pop2a.cfm

www.shodor.org/interactivate/lessons/IntroStatistics

www.statbasics.co.uk

UNIT 2: ANALYSIS OF VARIANCE (ANOVA)

CONTENTS

- 1.0 Introduction
- 2.0 Objectives
- 3.0 Main Content
 - 3.1 Conceptual Introduction to Analysis of Variance
 - 3.1.1 Assumptions of ANOVA
 - 3.1.2 Calculation of Several Measures of Variability
 - 3.1.3 The f-ratio
 - 3.2 Types of Analysis of Variance
 - 3.2.1. One-way ANOVA (between groups)
 - 3.2.2 One-way ANOVA (repeated measure)
 - 3.2.3 Two-way ANOVA (between groups)
 - 3.3 Interpretation of Computer Output
- 4.0 Conclusions
- 5.0 Summary
- 6.0 Tutor-Marked Assignment
- 7.0 References/Further Readings

1.0 INTRODUCTIONS

In the previous unit you used t-tests to compare the mean scores of two different groups or conditions. Consider a situation where the performances of two schools were compared in a certain subject. Because the comparison is for two group you learnt that you can use the t-test to test the hypothesis. Assuming we have three groups in place of two, what we have learnt so far does not accommodate such testing. An alternative you may argue is to use a t-test to compare the schools two at a time. The implication of doing so is that the 5% chance thereIn many situations you are likely to have more than two groups or conditions. In such a situation you use Analysis of variance (ANOVA). ANOVA is a statistical technique for testing hypotheses about many population means. In this unit you will learn of the different types, the assumptions and applications of Analysis of Variance.

2.0 OBJECTIVE

At the end of this unit, you will be able to:

- Calculate 'f' for different samples
- State assumptions governing use of ANOVA

- Differentiate between the different types of ANOVA
- Give conditions for use of the different types of ANOVA

3.0 MAIN CONTENT

3.1 Conceptual Introduction to Analysis of Variance

In unit 1 we discussed the independent-samples **t**-test with the aimed of determining if the independent variable with two levels or groups have different effects on the dependent variable. Let us assume that we have an independent variable with three levels or groups. If the dependent variable has effect on the groups, it will be reflected in significant differences among the means of the three groups. At first glance, you might suppose you could determine whether the three group means significantly differ from one another by performing a separate independent-samples **t**-test for each possible pair of means: that is, a **t**-test for

- Mean of group 1 versus Mean of group 2
- Mean of group 1 versus Mean of group 3 and
- Mean of group 2 versus Mean of group 3

This is a form of pair-wise **t**-test comparisons.

However, it is important that you note that this simple strategy is not possible. Let us now consider the reasons for the simple strategy not being possible. Remember what it means for a particular result to be significant. If an observed result is found to be significant at the basic .05 level, what this means is that there is only a 5% chance of its having occurred through mere chance. It then follows that for any of the pair-wise comparison, there would be a 5% probability by mere chance even when the null hypothesis is true. If this is true, you will end up with three times 5% which will amount to 15%. This is an indication that the probability that one or another of the comparisons might end up significant by mere chance is substantially greater than 0.05. Considering this from the law of probability as discussed earlier, the disjunctive probability would be $0.05 + 0.05 + 0.05 = .15$.

To overcome this complication, the Analysis of Variance (ANOVA) was developed. ANOVA is essentially an extension of the logic of **t**-tests to those situations where we wish to compare the means of three or more samples concurrently. As its name suggests, the analysis of variance focuses on variability. It involves the calculation of several measures of variability.

3.1.1 Assumptions of ANOVA

One-Way Analysis of Variance (ANOVA) makes the following assumptions about the data that are being fed into it:

1. that the scale on which the dependent variable is measured has the properties of an equal interval scale;
2. that the **k** samples are independently and randomly drawn from the source population(s);
3. that the source population(s) can be reasonably supposed to have a normal distribution; and
4. that the **k** samples have approximately equal variances

3.1.2 Calculation of Several Measures of Variability

You will remember that you learnt how to obtain the raw measure of variability in module 1, unit 2. Let us review the process again. For any set of **N** values of X_i that derive from an equal-interval scale of measurement, a deviate is the difference between an individual value of X_i and the mean of the set:

$$\text{deviate} = X_i - M_x$$

a **squared deviate** is the square of deviate:

$$\text{squared deviate} = (X_i - M_x)^2$$

and the **sum of squared deviates** is the sum of all the squared deviates in the set:

$$SS = \sum (X_i - M_x)^2$$

The algebraic equivalent formula is:

$$SS = \sum X_i^2 - \frac{(\sum X_i)^2}{N}$$

It is more convenient to use this formula during computation

Aggregate Differences among Sample Mean

At the beginning you learnt that ANOVA focuses on variability. The basic concept is that, whenever you have three or more numerical values, the measure of their variability is equivalent to the measure of their aggregate differences. That, indeed, is precisely what "variability" means: aggregate differences.

Calculating the aggregate differences requires first the calculation of each sample mean and grand mean of the samples. Then the difference between each sample mean and the grand mean is derived. To avoid negative scores the difference is then squared. Each of the squared difference is then weighted. This is done by multiplying each with the number in the sample (N). The sum of these resultant values gives what is known as **Sum of squared deviates between groups (SS_B)**. This is an aggregate measure of the degree to which the three sample means differ from one another

Let us illustrate these stages with a hypothetical sample. Assuming that three sample means are 11, 8 and 5 and that the sample size is 4, as presented in Table 3.2, let us practice the calculation of ANOVA.

Table 3.2: Sum of Squares Deviate between Groups

Sample A	Sample B	Sample C	Mean of Mean
$N_A = 4$	$N_B = 4$	$N_C = 4$	$11 + 8 + 5/3$
$M_A = 11$	$M_B = 8$	$M_C = 5$	$= 8$
$4(11 - 8)^2$ $= 36$	$4(8 - 8)^2$ $= 0$	$4(5 - 8)^2$ $= 36$	

The sum of these three values ($36 + 0 + 36 = 72$) gives what is known as the sum of squares deviate between groups. That is:

$$SS_{Bg} = 72$$

Once we have this measure, all that remains is to figure a way to determine whether it differs significantly from the zero that would be specified by the null hypothesis.

Once the SS between groups have been found, the next stage is to calculate the sum of squares within-group (SS_{wg}). The sum of squares of the samples

when summed together constitute a quantity known as sum of squared deviates **within-groups**, symbolized as SS_{wg}

In ANOVA therefore, three quantities of sum of squares are required. These are Sum of Squares total (SS_T), Sum of Squares between groups (SS_{Bg}) and Sum of Squares within group (SS_{wg}). The relationship between the three is:

- $SS_T = SS_{wg} + SS_{bg}$
- $SS_{bg} = SS_T - SS_{wg}$
- $SS_{wg} = SS_T - SS_{bg}$

These procedures can be summarized into the following stages.

3.1.3 The f-ratio

We will begin our discussion of this section with the explanation of two concepts – ‘**degree of freedom**’ (df) and ‘**mean square**’ (MS).

Degree of Freedom (df)

The basic concept of degrees of freedom in this context is "N—1," where "N" refers to the number of items on which the measure of sum of squared deviates is based. Let us assume that the value of the between-groups SS is based on the means of three groups, so the number of degrees of freedom associated with SS_{bg} is $df_{bg}=3-1=2$. Similarly if the between-groups SS is based on the mean of four groups, the number of degree of freedom is $4 - 1 = 3$.

When you consider within-groups SS , it is the sum of the separate SS measures for each of the samples. If we consider our example above with three samples we will have SS_a , SS_b , and SS_c . Each of these separate within-groups measures of SS is associated with a certain number of degrees of freedom: N_a-1 , N_b-1 , and N_c-1 , respectively. Let us assume that sample size for each group is 5, so the number of degrees of freedom associated with the composite within-groups measure, SS_{wg} , is $(N_a-1)+(N_b-1)+(N_c-1)$. Therefore:

$$\begin{aligned} df_{wg} &= (N_a-1)+(N_b-1)+(N_c-1) \\ &= (5 - 1) + (5 - 1) + (5 - 1) \\ &= 4 + 4 + 4=12 \end{aligned}$$

Mean Square (MS)

Within the context of the analysis of variance, an estimate of a source population variance is spoken of as a **mean square** (shorthand for "mean of the squared deviates") and conventionally symbolized as **MS**. The variance estimate is $\mathbf{MS}_{bg} = \mathbf{SS}_{bg} / \mathbf{df}_{bg}$

Consider the example in section 3.1.2 in which the between group SS is found to be 72 and calculated from three samples, the variance estimate Mean Square between-groups is:

$$\begin{aligned}\mathbf{MS}_{bg} &= \mathbf{SS}_{bg} / \mathbf{df}_{bg} \\ \mathbf{MS}_{bg} &= 72 / 3 \\ &= 24\end{aligned}$$

For within sum of squares, the variance estimate is $\mathbf{MS}_{wg} = \mathbf{SS}_{wg} / \mathbf{df}_{wg}$

The relationship between the two values of Mean square (**MS**) is conventionally described by a ratio known as **F**, which is defined for the general case as:

$$\mathbf{F} = \frac{\mathbf{MS}_{\text{effect}}}{\mathbf{MS}_{\text{error}}}$$

Where $\mathbf{MS}_{\text{effect}}$ is a variance estimate pertaining to the particular fact whose significance you wish to assess (e.g., the differences among the means of several independent samples), and $\mathbf{MS}_{\text{error}}$ is a variance estimate reflecting the amount of sheer, cussed random variability that is present in the situation. For the present example, $\mathbf{MS}_{\text{effect}}$ would be the same as \mathbf{MS}_{bg} and $\mathbf{MS}_{\text{error}}$ would be the same as \mathbf{MS}_{wg} . When the null hypothesis is **true**, \mathbf{MS}_{bg} will tend to be equal to or less than \mathbf{MS}_{wg} ; and when the null hypothesis is **not true**, \mathbf{MS}_{bg} will tend to be greater than \mathbf{MS}_{wg} . Consequently F ratio comes out as:

$$\mathbf{F} = = \frac{\mathbf{MS}_{bg}}{\mathbf{MS}_{wg}}$$

3.2 Types of Analysis of Variance

There are different types of ANOVA. In this section we will consider three types – one-way between groups ANOVA, one-way repeated measure ANOVA and two-way between groups ANOVA. It is called ANOVA because the statistics compares the variance (i.e. variability of scores) between or within the different groups.

3.2.1. One way ANOVA (Between Groups)

This is called one-way because it involves only one Independent Variable (referred to as a factor) which must have a number of different levels. These levels correspond to the different groups or conditions. For example, consider the effect of a teaching style on students' performance in Mathematics. The teaching style is the independent variable while student mathematics score from an achievement test is the dependent variable. Assuming the independent variable (teaching style) has 3 levels (whole class lecture, small group activities and self-paced computer assisted activities) you are expected to use ANOVA

Example 1

A	B	C	
14	8	8	
10	14	6	
11	3	5	
9	7	1	
<hr/>	<hr/>	<hr/>	
11	8	5	= Group \bar{x}
$\frac{11+8+5}{3} = 8$			= Grand \bar{x}

Observations

- (1) There seems to be an overall \bar{x} difference

Questions

- (1) Is this difference statistically significant?
- (2) If so, is the size of the difference significant?

Step-by-Step Computational Procedure: One-Way Analysis of Variance for Independent Samples

Step 1: Combining all **k** groups together, calculate Total Sum of Squares

$$\begin{aligned}SS_T &= \sum X_i^2 - \frac{(\sum X_i)^2}{N_T} \\&= \sum (\text{Each observation} - \text{Grand mean})^2 \\&= (14-8)^2 + (10-8)^2 + (11-8)^2 + (9-8)^2 + (8-8)^2 \\&\quad + (14-8)^2 + (3-8)^2 + (7-8)^2 + (8-8)^2 + (6-8)^2 + \\&\quad (5-8)^2 + (1-8)^2 = 174\end{aligned}$$

Step 2 For each of the **k** groups separately, calculate the sum of squared deviates within the group ("g") as:

$$\begin{aligned}SS_g &= \sum X_{gi}^2 - \frac{(\sum X_{gi})^2}{N_g} \\&= (\text{Each Observation} - \text{Treatment mean})^2 \\&= (14-11)^2 + (10-11)^2 + (11-11)^2 + (9-11)^2 + (8-8)^2 + (14-8)^2 + (3-8)^2 \\&\quad + (7-8)^2 + (8-5)^2 + (6-5)^2 + (5-5)^2 + (1-5)^2 = 102\end{aligned}$$

Step 3: Between Treatment Sum of Squares

Take the sum of the SS_g values across all **k** groups to get

$$\begin{aligned}SS_{wg} &= SS_a + SS_b + SS_c + SS_d \\&= (\text{Average treatment outcome} - \text{Grand mean})^2 \times \text{Number of observations in treatment} \\&= (11-8)^2 \times 4 + (8-8)^2 \times 4 + (5-8)^2 \times 4 \\&= 36 + 0 + 36 \\&= 72\end{aligned}$$

Between treatment SS is the explained variance

Step 4 Calculate the relevant degrees of freedom as

$$df_T = N_T - 1$$

$$df_{bg} = k - 1$$

$$df_{wg} = N_T - k$$

Or

$$Df = c - 1, n - c$$

$$= 3 - 1, 12 - 3$$

$$= 2, 9$$

Step 5 Calculate the relevant mean-square values as

$$MS_{bg} = \frac{SS_{bg}}{df_{bg}}$$

and

$$MS_{wg} = \frac{SS_{wg}}{df_{wg}}$$

Step 7 Calculate F as

$$F = \frac{MS_{bg}}{MS_{wg}}$$

Step 8. Refer the calculated value of **F** to the table of critical values of **F**, with the appropriate pair of numerator/denominator degrees of freedom,

For example if table or critical value of $F = 4.26$ ($\alpha = 0.05$) f must equal or exceed 4.26 for the means to be significantly different.

Example 1:

Consider the following scores and means for three students from three classes:

	Class 1	Class 2	Class 3
	4	16	10
	14	18	19
	3	14	7
\bar{x}	7	16	12

There seems to be an overall \bar{x} difference. The question is:

- Is the difference statistically significant?
- If so, is the size of the difference significant?

Let us practice the step by step ANOVA calculation using this example.

Step 1: To begin, you need to calculate the grand mean for the three classes:

$$\text{Grand } \bar{x} = \frac{7 + 16 + 12}{3} = 11.7$$

Step 2: You calculate the total variance sum of squares:

$$\begin{aligned} \text{Total Variance sum of Squares} &= \sum (\text{Each observation} - \text{Grand } \bar{x})^2 \\ &= (4 - 11.7)^2 + (14 - 11.7)^2 + (3 - 11.7)^2 + (16 - 11.7)^2 + (18 - 11.7)^2 + (14 - 11.7)^2 + \\ &\quad (10 - 11.7)^2 + (19 - 11.7)^2 + (7 - 11.7)^2 \\ &= 59.29 + 5.29 + 75.69 + 18.49 + 39.69 + \\ &\quad 5.29 + 2.89 + 53.29 + 22.09 \end{aligned}$$

Step 3: You calculate between Treatment Sum of Squares

$$\begin{aligned} &= \sum (\text{Group } \bar{x} - \text{Grand } \bar{x})^2 N \\ &= (7 - 11.7)^2_3 + (16 - 11.7)^2_3 + (12 - 11.7)^2_3 \\ &= 22.09 \times 3 + 18.49 \times 3 + 0.09 \times 3 \\ &= 66.27 + 55.47 + 0.27 \\ &= 122.01 \end{aligned}$$

Step 4: Then Within-Treatment SS (Unexplained variance)

$$\begin{aligned} &\sum (\text{Each observation} - \text{its group } \bar{x})^2 \\ &= (4 - 7)^2 + (14 - 7)^2 + (3 - 7)^2 + (16 - 16)^2 + (18 - 16)^2 + (14 - 16)^2 + (10 - 12)^2 + (19 - 12)^2 + (7 - 12)^2 \end{aligned}$$

$$= 9 + 49 + 16 + 0 + 4 + 4 + 4 + 49 + 25 = 160$$

When reporting the results of an analysis of variance, it is good practice to present a summary table as shown in Table 3.3. Clearly identifying each component of **SS**, **df**, and **MS**, allows you to take in the main details of the analysis at a single glance.

Table 3.3: ANOVA Summary Table

Source of Variation	SS	Df	MS	F-cal	F-table
Between group	122.01	2	61.00	2.29	5.14
Within group	160.00	6	26.67		
Total	282.01	8			

Interpretation of Computer Output of One-Way ANOVA

The output generated from oneway ANOVA is presented below.

Oneway ANOVA

Descriptives

Post-test Integrated Science Achievement Score

	N	Mean	Std. Deviation	Std. Error	95% Confidence Interval for Mean		Minimum	Maximum
					Lower Bound	Upper Bound		
Discovery Group	16	39.69	6.172	1.543	36.40	42.98	34	55
Lecture Group	16	40.06	6.115	1.529	36.80	43.32	31	52
Control Group	16	31.13	3.931	.983	29.03	33.22	27	39
Total	48	36.96	6.813	.983	34.98	38.94	27	55

Test of Homogeneity of Variances

Post-test Integrated Science Achievement Score

Levene Statistic	df 1	df 2	Sig.
2.418	2	45	.101

ANOVA

Post-test Integrated Science Achievement Score

	Sum of Squares	df	Mean Square	F	Sig.
Between Groups	817.792	2	408.896	13.489	.000
Within Groups	1364.125	45	30.314		
Total	2181.917	47			

Post Hoc Tests

Multiple Comparisons

Dependent Variable: Post-test Integrated Science Achievement Score

Tukey HSD

(I) Experimental Groups	(J) Experimental Groups	Mean Difference (I-J)	Std. Error	Sig.	95% Confidence Interval	
					Lower Bound	Upper Bound
Discovery Group	Lecture Group	-.38	1.947	.980	-5.09	4.34
	Control Group	8.56*	1.947	.000	3.84	13.28
Lecture Group	Discovery Group	.38	1.947	.980	-4.34	5.09
	Control Group	8.94*	1.947	.000	4.22	13.66
Control Group	Discovery Group	-8.56*	1.947	.000	-13.28	-3.84
	Lecture Group	-8.94*	1.947	.000	-13.66	-4.22

*. The mean difference is significant at the .05 level.

Homogeneous Subsets

Post-test Integrated Science Achievement Score

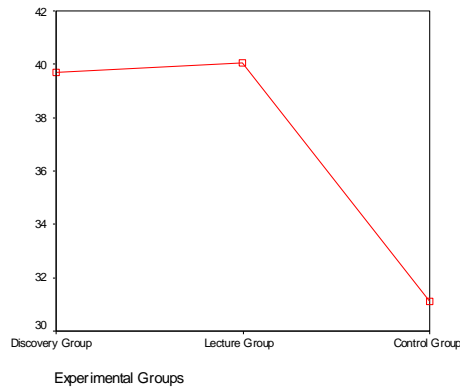
Tukey HSD^a

Experimental Groups	N	Subset for alpha = .05	
		1	2
Control Group	16	31.13	
Discovery Group	16		39.69
Lecture Group	16		40.06
Sig.		1.000	.980

Means for groups in homogeneous subsets are displayed.

a. Uses Harmonic Mean Sample Size = 16.000.

Means Plots



There are five separated pieces of information. These are:

- Table titled **Descriptive** – gives information about the different groups. You should always check this first to confirm that the number of subjects is correct, and to consider the descriptive statistics of the groups.
- The second table is titled **Test of Homogeneity of Variance** – You look at this table to determine if the assumption on homogeneity of variance is violated or not. To do so you look at the column marked **Sig.** If the value is larger than .05 it means you have not violated the assumption of homogeneity of variance. If the value is equal or less than .05 you have violated the assumption. In the example used for the analysis, the value is .101. This is larger than .05 so the assumption of homogeneity of variance is not violated.
- The third table is labelled **ANOVA** – This table gives the various sums of squares in ANOVA computation. However, you are interested in the column marked **sig.** If the value is equal or less than .05, then there is a significant difference among the means of the different groups. The value in the above output is .000. This is less than .0005 which is by far smaller than .05. You conclude that there is a significant difference among the means of the three groups. The ANOVA table only tells you that a difference exists without indicating where the difference lies. It does not show the groups that are different.
- To determine which group is different from which other group, the table titled **Multiple Comparison** is consulted. This gives the results of the post-hoc tests. On this table you look down the column labelled **Mean Difference.** You look for any asterisks (*) next to the values listed. The asterisks indicate that the two groups being compared are significantly different from one another.

- The final piece of information is the **Mean Plots** – This gives a visual comparison of the means of the groups.

Calculating Effect Size

You have learnt how to compute this for t-tests in the previous unit. For ANOVA, the formula is:

$$\text{Eta squared} = \frac{\text{SumofSquaresbetween} - \text{groups}}{\text{TotalSumofSquares}}$$

Self Assessment

Question 1: Are the means significant different?

Question 2

Schools

A	B	C	D
8	7	5	10
7	5	3	5
4	5	4	6
5	4	4	4
5	3	3	8
5	4	5	7
6	6	4	8
6	4	4	4

3.2.2 One-way ANOVA (repeated measure)

In unit 1 section 3.2.3 you learnt of pair sample t-test also known as repeated measure or correlated-sample t-test. You can think of this version of the analysis of variance as its extension. The most conspicuous similarity between the two is in the way the data are arrayed. In the correlated-samples t-test we typically have a certain number of subjects; each measured under two conditions, A and B. Or alternatively, we have a certain number of matched pairs of subjects, with one member of the pair measured under condition A and the other measure under condition B. It is the same structure with the correlated-samples ANOVA, except that now the number of conditions is three or more. When the analysis involves each subject being measured under each of the **k** conditions, it is sometimes spoken of as a **repeated measures** or **within subjects** design. When it

involves subjects matched in sets of three for $k=3$, four for $k=4$, and so on, with the subjects in each matched set randomly assigned to one or another of the k conditions, it is described as a **randomized blocks** design. (In this latter case, each set of k matched subjects constitutes a "block.")

Step-by-Step Computational Procedure: One-way ANOVA (repeated measure)

The logic and procedure of the One-way repeated or correlated ANOVA are the same as for the independent-samples version, described in section 3.2.1 above. In the independent-samples ANOVA, SS_T is analyzed into two complementary components, SS_{bg} and SS_{wg} . Each divided by its respective value of df yields a value of MS ; and these, in turn, yield the F -ratio. The numerator of the ratio, MS_{bg} , reflects the aggregate differences among the means of the k groups of measures; and the denominator, MS_{wg} , reflects the random variability, commonly described as "error," that exists inside the k groups.

In most real-life situations, a certain amount of the variability that exists inside the k groups will reflect pre-existing individual differences among the subjects. The pre-existing individual differences are sources of variability and could be entirely extraneous to the question the investigators were aiming to answer. To avoid the extraneous clutter all subjects in a study must be equivalent. Unfortunately this is not easy to achieve. Repeated measure ANOVA is a design that identifies these pre-existing differences and removes them. The portion of SS_{wg} that is attributable to pre-existing individual differences is designated as $SS_{subjects}$. In repeated measure ANOVA it is dropped from the analysis; and the portion that remains, SS_{error} , is then used as the measure of random variability. The formula for calculating $SS_{subjects}$, is:

$$SS_{subj} = \frac{\sum (\sum X_{subj*})^2}{k} - \frac{(\sum X_{Ti})^2}{N_T}$$

When the SS_{subj} is removed from the SS_{wg} what remains is known as SS_{error} . The formula for calculating SS_{error} is:

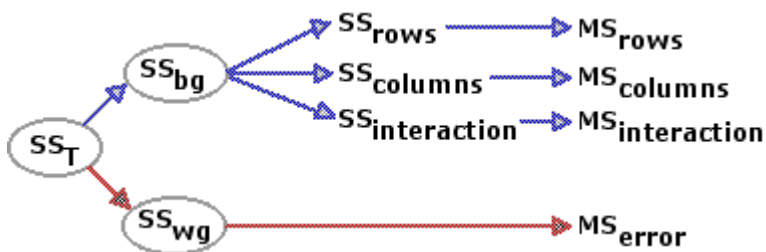
$$SS_{\text{error}} = SS_{\text{wg}} - SS_{\text{subj}}$$

3.2.3 Two-way ANOVA (between groups)

This section will explore two-way between-groups ANOVA. Two-way means that there are two independent variables and between-groups means that different people are in each group. In our example in the last section, we considered the effect of teaching methods on integrated science achievement. Results indicated that there was a significant difference of the effect of the different methods with post hoc indicating that the major difference was between discovery method and control. We may further ask: Is this the case for both males and females? With this question, a second independent variable sex is introduced. One-way ANOVA can not answer these questions; rather we use what is known as two-way ANOVA. Two-way ANOVA is a procedure that examines the effects of two independent variables concurrently. The advantage of using a two-way design is that you can test the *'main effect'* for each of the independent variables and also explore the possibility of an *'interaction effect'*. An interaction effect occurs when the effect of one independent variable on a dependent variable depends on the level of a second independent variable.

Two-Way ANOVA Computational Procedure

The two-way ANOVA for independent samples proceeds exactly the same way like the corresponding one-way ANOVA. Like with the one-way you are required to calculate the sum of squares between-group (SS_{bg}) as you did in section on one-way ANOVA. This is the measure of the aggregate differences among the several groups. With two independent variables and the data arrayed in the form of a rows-by-columns matrix, SS_{bg} is further divided into three parts. As illustrated in the diagram below.



One part measures the differences among the means of the two or more rows (SS_{rows}), another measure the mean differences among the two or more columns (SS_{columns}), and the third is a measure of the degree to which

the two independent variables interact ($SS_{\text{interaction}}$). Each of these three components is then converted into a corresponding value of MS , with the result that three separate F -ratios are calculated and three separate tests of significance are performed (one for the row variable, one for the column variable, and one for the interaction between the two variables).

You are also required to calculate the sum of squares within-groups (SS_{wg}) which is the measure of random variability inside the groups. The SS_{wg} is treated the same way as in the one-way analysis to generate the value for MS_{error} that appears in the denominator of the F -ratio. These are illustrated as step by step as below:

Step-by-Step Two-Way ANOVA Computational Procedure

The discussion above can be summarized in a step by step procedure. To illustrate this let us consider an example with 2 rows and 3 columns (hence $rc=6$). We will also assume that you have already done the basic calculation to get $\sum X_i$ and $\sum X_i^2$ for each of the groups separately and for all groups combined.

Step 1: Combining all rc groups together, calculate

$$SS_T = \sum X_i^2 - \frac{(\sum X_i)^2}{N_T}$$

Step 2. For each of the rc groups separately, calculate the sum of squared deviates within the group ("g") as

$$SS_g = \sum X_{gi}^2 - \frac{(\sum X_{gi})^2}{N_g}$$

Step 3. Take the sum of the SS_g values across all rc groups to get

$$SS_{\text{wg}} = SS_a + SS_b + SS_c + SS_d + SS_e + SS_f$$

Step 4. Calculate SS_{bg} as

$$SS_{bg} = SS_T - SS_{wg}$$

Step 5. Calculate SS_{rows} as:

$$SS_{rows} = \frac{(\sum X_{r1})^2}{N_{r1}} + \frac{(\sum X_{r2})^2}{N_{r2}} - \frac{(\sum X_T)^2}{N_T}$$

Step 6. Calculate SS_{cols} as:

$$SS_{cols} = \frac{(\sum X_{c1})^2}{N_{c1}} + \frac{(\sum X_{c2})^2}{N_{c2}} + \frac{(\sum X_{c3})^2}{N_{c3}} - \frac{(\sum X_T)^2}{N_T}$$

Step 7. Calculate SS_{rxc} as:

$$SS_{rxc} = SS_{bg} - SS_{rows} - SS_{cols}$$

Step 8. Calculate the relevant degrees of freedom as:

$$df_T = N_T - 1$$

$$df_{wg} = N_T - rc$$

$$df_{bg} = rc - 1$$

$$df_{rows} = r - 1$$

$$df_{cols} = c - 1$$

$$df_{rxc} = (r - 1)(c - 1)$$

Step 9. Calculate the relevant mean-square values as:

$$MS_{rows} = \frac{SS_{rows}}{df_{rows}} \quad MS_{cols} = \frac{SS_{cols}}{df_{cols}} \quad MS_{rxc} = \frac{SS_{rxc}}{df_{rxc}} \quad MS_{error} = \frac{SS_{wg}}{df_{wg}}$$

Step 10. Calculate **F** as:

$$F_{\text{rows}} = \frac{MS_{\text{rows}}}{MS_{\text{error}}} \quad F_{\text{cols}} = \frac{MS_{\text{cols}}}{MS_{\text{error}}} \quad F_{\text{rxc}} = \frac{MS_{\text{rxc}}}{MS_{\text{error}}}$$

Example 1 – Class

Location	Class 1	Class 2	Class 3
R1	4	16	16
R2	14	18	19
R3	3	14	7

Example 2

Types of School	Class I	Class II	Class III	Total \bar{x}
All girls	4	16	10	10
All boys	14	18	19	17
Mixed	3	14	7	8

Group \bar{x} Classes = 11.7

Group \bar{x} for type 2 of school

$$\begin{aligned}
 &= (10 - 11.7)_3^2 + (17 - 11.7)_3^2 + (8 - 11.7)_3^2 \\
 &= 8.67 + 84.27 + 41.07 \\
 &= 134
 \end{aligned}$$

Two-Way ANOVA Calculation

Example class	1	2	3	
Types of School				Total
All girls	4	16	10	10
All boys	14	18	19	17
Mixed	3	14	7	8
Total	7	16	12	

3.3 Interpretation of Computer Output from 2-Way ANOVA

Four tables and a graph are produced when computer is used for the analysis of a 2-way ANOVA:

Descriptive Statistics

Dependent Variable: Post-test Integrated Science Achievement Score

Sex	Experimental Groups	Mean	Std. Deviation	N
Males	Discovery Group	40.00	7.540	8
	Lecture Group	40.00	5.855	8
	Control Group	31.50	4.408	8
	Total	37.17	7.100	24
Females	Discovery Group	39.38	4.955	8
	Lecture Group	40.13	6.770	8
	Control Group	30.75	3.655	8
	Total	36.75	6.661	24
Total	Discovery Group	39.69	6.172	16
	Lecture Group	40.06	6.115	16
	Control Group	31.13	3.931	16
	Total	36.96	6.813	48

Levene's Test of Equality of Error Variances

Dependent Variable: Post-test Integrated Science Achievement Score

F	df 1	df 2	Sig.
1.350	5	42	.263

Tests the null hypothesis that the error variance of the dependent variable is equal across groups.

- a. Design: Intercept+SEX+EXPGROUP+SEX * EXPGROUP

Tests of Between-Subjects Effects

Dependent Variable: Post-test Integrated Science Achievement Score

Source	Type III Sum of Squares	df	Mean Square	F	Sig.	Partial Eta Squared	Noncent. Parameter	Observed Power ^a
Corrected Model	821.667 ^b	5	164.333	5.074	.001	.377	25.370	.971
Intercept	65564.083	1	65564.083	2024.401	.000	.980	2024.401	1.000
SEX	2.083	1	2.083	.064	.801	.002	.064	.057
EXPGROUP	817.792	2	408.896	12.625	.000	.375	25.251	.994
SEX * EXPGROUP	1.792	2	.896	.028	.973	.001	.055	.054
Error	1360.250	42	32.387					
Total	67746.000	48						
Corrected Total	2181.917	47						

a. Computed using alpha = .05

b. R Squared = .377 (Adjusted R Squared = .302)

Post Hoc Tests

Experimental Groups

Multiple Comparisons

Dependent Variable: Post-test Integrated Science Achievement Score

Tukey HSD

(I) Experimental Groups	(J) Experimental Groups	Mean Difference (I-J)	Std. Error	Sig.	95% Confidence Interval	
					Lower Bound	Upper Bound
Discovery Group	Lecture Group	-.38	2.012	.981	-5.26	4.51
	Control Group	8.56*	2.012	.000	3.67	13.45
Lecture Group	Discovery Group	.38	2.012	.981	-4.51	5.26
	Control Group	8.94*	2.012	.000	4.05	13.83
Control Group	Discovery Group	-8.56*	2.012	.000	-13.45	-3.67
	Lecture Group	-8.94*	2.012	.000	-13.83	-4.05

Based on observed means.

*. The mean difference is significant at the .05 level.

Homogeneous Subsets

Post-test Integrated Science Achievement Score

Tukey HSD^{a,b}

Experimental Groups	N	Subset	
		1	2
Control Group	16	31.13	
Discovery Group	16		39.69
Lecture Group	16		40.06
Sig.		1.000	.981

Means for groups in homogeneous subsets are displayed.

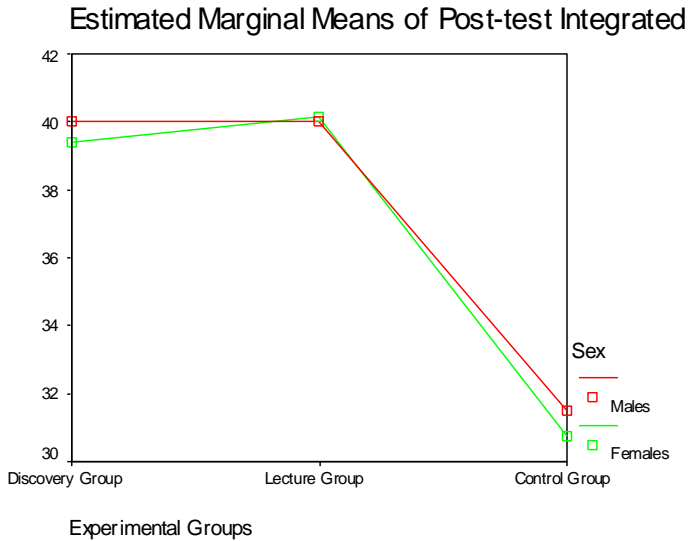
Based on Type III Sum of Squares

The error term is Mean Square(Error) = 32.387.

a. Uses Harmonic Mean Sample Size = 16.000.

b. Alpha = .05.

Profile Plots



1. **Descriptive Statistics:** This table provides the mean scores (\bar{x}), standard deviations (SD) and the number (N) of item or candidates for each group check that these values are correct.

2. **Levene’s Test of Equality of Error Variances**
 This is a test of one of the assumptions underlying ANOVA. From this table, you should be interested on the column marked “sig”. If the value is greater than 0.05, the test is not significant. If the value is equal to or less than 0.05 the test is significant. A significant result suggests that the variance of your Dependent Variable across the groups is not equal. They are different. This situation can be studied by setting a more stringent significance level (e.g. 0.01). If there is no significant difference, it means that the variance of your Dependent Variables across the groups is equal and that you have not violated the homogeneity of variance assumption.

Test of Between – Subjects Effects Table

This is the main output from two-way ANOVA. This table gives information on:

- **Main Effects:** In the left-hand column the Independent Variables are listed. To determine if there is a main effect for each Independent Variable, the column marked **Sig** is read. If the value is equal to or less than 0.05, then the main effect for that

Independent Variable is said to be significant. If greater than 0.05 the main effect is not significant.

- **Effect Size:** This is provided in the column labeled **Eta Squared**, unlike in t-test and one-way ANOVA where you have to manually compute it. You will recall the interpretation of eta squared. It is same as previously discussed. You may at this point read Cohen's (1988) criterion as discussed under one-way ANOVA.
- **Interaction Effects:** This is the row with the 2 Independent Variables separated by an asterisk (SEX*EXPGROUP). SPSS tells you whether there is an interaction between the two independent variables in their effect on the dependent variable. To check if the interaction effect of the two dependent variables is significant, you look at the column marked **Sig** for that line for interaction. If the value is equal to or less than .05, the effect is significant. If the value is larger than .05 the interaction effect is not significant. From the output above, there is no significant interaction effect. It means that there is significant difference in the effect of method of teaching on integrated science achievement for males and females.
- **Post-hoc Tests:** The 'f' test can only identify the existence of differences. However, it is only possible to know where the difference lies through post-hoc tests. This is a test that compares the levels of Independent Variables in pairs. The result of the post-hoc is presented on the table marked **multiple comparisons**. The column marked **sig** gives an indication of significance or not.
- **Plots:** Plots allow you **visually** inspect the relationship among your variables

4.0 Conclusions

In many research situations we are interested in comparing the mean scores of more than two groups. One way analysis of variance is important in this respect. It is so called because it compares the variances (variability of scores) between the different groups.

5.0 Summary

In this unit we have discussed the basic requirements and assumptions of one-way analysis of variance. We also learnt of the steps in calculating f-ratio and interpretation of computer output for a one-way ANOVA

6.0 Tutor-Marked Assignment

- 1) A track coach developed three different methods of coaching 'track'. A sample of 21 subjects believed to be of equal ability was randomly assigned to each of the three teaching methods. The data below show the time it took each subject to run a particular distance following the coaching programme.

Method A	Method B	Method C
6	3	5
8	4	6
9	4	7
8	3	6
9	4	6
7	3	5
7	3	5

- a) Calculate the appropriate analysis of variance and draw relevant conclusions
 - b) Which of the coaching method is superior? (Owie, 2006)
- 2) Explain the terms:
 - a) Sum of squared deviates within the group
 - b) Sum of squares between groups
 - c) F-ratio
 - d) Interaction effect
 - e) Degree of freedom

7.0 References/Further Readings

Adamu, S. O. (1997), *Statistics for Beginners Book 1*, Nigeria, Ibadan, SAAL Pub.

Hinton, P. R. (2003), *Statistics Explained A Guide for Social Science Students*, Hove and New York, Routledge

Owie, I. (2006), *Fundamentals of Statistics in Education and the Social Sciences*, Nigeria, Lagos, National Book Consortium

Kerlinger, F. (1986), *Foundations of Behavioural Research*, New York:
McGraw-Hill

<http://www.gower.k12.il.us/Staff/ASSESS>

www.apsu.edu/oconnort/3760/3760lect03a.htm

faculty.vassar.edu/lowry/webtext.html

www.uwsp.edu/psych/stat/14/nonparm.htm

writing.colostate.edu/guides/research/stats/pop2a.cfm

www.shodor.org/interactivate/lessons/IntroStatistics

www.statbasics.co.uk

UNIT 3: ANALYSIS OF COVARIANCE (ANCOVA)

CONTENTS

- 1.0 Introduction
- 2.0 Objectives
- 3.0 Main Content
 - 3.1 Understanding Analysis of Covariance
 - 3.1.1 When to Use Analysis of Covariance
 - 3.1.2 Assumptions of Analysis of Covariance
 - 3.2 Types of Analysis of Covariance
 - 3.2.1 One-way ANCOVA
 - 3.2.2 Two-way ANCOVA
 - 3.2.3 Multivariate ANCOVA
- 4.0 Conclusions
- 5.0 Summary
- 6.0 Tutor-Marked Assignment
- 7.0 References/Further Readings

1.0 INTRODUCTION

In the previous sections and units we discussed designs for analyzing variances. You will recall that both in t-tests and analysis of variance, we discussed the use of repeated measure design. In such a design, a subject is subjected repeatedly to the different conditions under study. Unfortunately, there are certain situations where a repeated-measures design might not be feasible; and there are others where, even if it is feasible, it might not be desirable. In this unit we will consider one other design that has the advantage of removing pre-existing individual differences without resorting to repeated measures. This is known as Analysis of Variance (ANCOVA).

2.0 OBJECTIVES

By the end of this unit you should be able to:

- Explain the uses of Analysis of Variance
- List the Assumptions of ANCOVA
- List the types of ANCOVA
- Differentiate among the different types of ANCOVA
- Interpret computer output of ANCOVA

3.0 MAIN CONTENT

3.1 Understanding Analysis of Covariance

The purpose of all research study is to explain the effect of independent variable(s) on the dependent variable(s). We use a research design to provide structure for the research. Using a research design a researcher controls Independent variables that can help to explain the observed variation in the dependent variable, which in turn reduces error variance (unexplained variation). This method of control is called **experimental control** since the research design is structured before the research begins. Sometimes experimental control is difficult or impossible. An alternative is to use what is known as **statistical control**. An analysis procedure employed in statistical control is the **Analysis of Variance (ANOVA)**. Analysis of Covariance is a design that is capable of removing the obscuring effects of pre-existing individual differences among subjects without resorting to the repeated-measures strategy. It is a very powerful and robust package. ANCOVA is a merger of ANOVA and regression for continuous variables. ANCOVA tests whether certain factors have an effect after removing the variance for which quantitative predictors (covariates) account. It allows you to explore differences between groups while statistically controlling for an additional (continuous) variable. This additional variable is one that you suspect may be influencing scores on the dependent variable and is called COVARIATE.

3.1.1 When to Use Analysis of Covariance

ANCOVA can be used when you have:

- Two-group pre-test/post-test design. For example you are interested in finding the effect of a new educational intervention. What you do is to test the subjects before the intervention (pre-test) and after the intervention (post-test). The scores on pre-test are treated as a covariate to 'control' for pre-existing differences between the groups.
- ANCOVA is also useful when you have been unable to randomly assign your subjects to the different groups but instead has had to use existing grouping. This is peculiar to school situations and research in education. Many schools may not allow you to disorganize the school arrangement during your data collection. In such situation, you use what is known as intact class. As the students may differ in a number of attributes (and not necessarily on the one

you are interested in), ANCOVA can be used in an attempt to reduce some of the differences.

3.1.2 Assumptions of Analysis of Covariance

You will recall that we discussed the assumptions underlying the analysis of variance in unit 2. These assumptions also affect analysis of covariance. You may take time now to re-visit the unit on ANOVA to remind your self of these assumptions.

In addition to the assumptions for ANOVA, there are key issues and assumptions associated with ANCOVA. These additional assumptions and key issues are:

- *Influence of Treatment on Covariate Measurement*
When you plan to use ANCOVA, you must ensure that the covariate is measured prior to the treatment or experiment. This is to avoid the influence of treatment on the covariate.
- *Reliability of Covariates*
ANCOVA assumes that covariates are measured without error. Unfortunately in education research there are variables that can be measured without avoiding errors. Most measures that rely on scale may not meet this assumption. To improve on the reliability of your measurement tools you must ensure that you use good and validated scales or questionnaire. If it is an observation study, make sure you train your observers.
- *Correlations Among Covariates*
Your covariates should correlate substantially with the Dependent Variable and not with one another. If covariates correlate strongly ($r=80$ and above) you should consider removing some of the covariates.
- *Linear Relationship between Dependent Variable and Covariate*
ANCOVA assumes that the relationship between the dependent variable and each of your covariates is linear (straight line). It also assumes a linear relationship between pairs of your covariates if you are using more than one. Violation of this assumption reduces the power (sensitivity) of your test.

- *Homogeneity of Regression Slopes*
ANOVA assumes that the relationship between the covariate and dependent variable is the same for each of the groups. There should be no interaction between the covariate and the dependent variable. If interaction exists, the result of the ANCOVA will be misleading.

3.2 Types of Analysis of Covariance

Analysis of covariance can be used as part of one-way, two-way and multivariate ANOVA techniques. In this section we will discuss the different types of ANCOVA. However, computer output for only one-way and two-way ANCOVA will be interpreted.

3.2.1 One-way ANCOVA

You will recall that the one-way ANOVA discussed in unit 2 has only one independent variable and one dependent variable. The situation is similar with the one-way ANCOVA. This involves one independent variable which is usually categorical variable (with two or more levels or conditions), one dependent variable which is a continuous variable and one or more continuous covariates. The main difference between the two design is that one-way ANOVA has no covariates. One way ANCOVA will tell you if the mean dependent variable scores at time 2 (post test) for the Independent Variable groups are significantly different after initial pre-test (covariate) scores are controlled for.

Calculating F-ratio for One-way ANCOVA

Four sets of calculations are necessary in one-way ANCOVA.

- The first is similar to what you learnt in ANOVA. These calculations are on the dependent variable of your interest. The set of calculations require you to calculate the sum of squares for total (SS_T), sum of squares within-group (SS_{wg}) and sum of squares between-groups (SS_{bg}). You may wish to refer to these calculations in the section on ANOVA.
- The second and third sets are aimed at the covariance aspect. The focus of the second and third sets of calculations is on the covariate that you wish to statistically control. For these sets you are required to calculate the sum of squares total and within-group for the covariate, and the sum of covariates.
- The final set ties the two aspects together.

Because of the complexity in the calculations, specific examples are not considered here. You are however, advised to read any good statistical group for detailed calculations. We will only concentrate on interpretation of computer output since most persons employ the computer to analyse their data when complicated designs are used.

Interpretation of Output from One-Way ANCOVA

There are at least 4 tables generated in the one-way computer analysis:

1. Table Labelled Descriptive Statistics

Descriptive Statistics

Dependent Variable: Post-test Integrated Science Achievement Score

Experimental Groups	Mean	Std. Deviation	N
Discovery Group	39.69	6.172	16
Lecture Group	40.06	6.115	16
Control Group	31.13	3.931	16
Total	36.96	6.813	48

This table gives you the group mean scores, standard deviations and the number of subjects in each group.

2. Table Labelled Levene's Test of Equality of Error Variances

This allows you check that you have not violated the assumptions of equality of variance. if value is greater than 0.05 there is no significant difference. This means that there is equality of variance and the assumption is not violated. However, if the value is equal to or less than .05 (as in the output below), it means that there is a significant difference in the variance and that there is no equality of means. Therefore the assumption is violated.

Levene's Test of Equality of Error Variances

Dependent Variable: Post-test Integrated Science Achievement Score

F	df1	df2	Sig.
3.464	2	45	.040

Tests the null hypothesis that the error variance of the dependent variable is equal across groups.

a. Design: Intercept+TOTPRESC+EXPGROUP

3. Table Labelled Test of Between-Subjects Effects

Tests of Between-Subjects Effects

Dependent Variable: Post-test Integrated Science Achievement Score

Source	Type III Sum of Squares	df	Mean Square	F	Sig.	Partial Eta Squared	Noncent. Parameter	Observed Power ^a
Corrected Model	832.132 ^b	3	277.377	9.042	.000	.381	27.126	.993
Intercept	1118.909	1	1118.909	36.474	.000	.453	36.474	1.000
TOTPRESC	14.340	1	14.340	.467	.498	.011	.467	.103
EXPGROUP	829.547	2	414.774	13.521	.000	.381	27.041	.997
Error	1349.785	44	30.677					
Total	67746.000	48						
Corrected Total	2181.917	47						

a. Computed using alpha = .05

b. R Squared = .381 (Adjusted R Squared = .339)

This is the main ANCOVA results. Very important pieces of information are derived from this table. They are:

- Column marked **Sig**: The value on this column tells you if your (Independent Variable) groups are significantly different in terms of their scores on the Dependent Variable. It also indicates the influence of your covariate. To read the values correctly, you find the lines in the table which correspond to the Independent Variable and covariate. If **sig** is less than 0.05, the difference is significant. If greater the difference is not significant.
- Column marked **Eta squared**: The value on this column helps you consider the effect size of your independent variable and covariate when the value is converted to percentage by multiplying by 100, the value indicates how much of the variance in the dependent variable is explained by the independent variable or the covariate.

4. Table Marked Estimated Marginal Means

Experimental Groups

Dependent Variable: Post-test Integrated Science Achievement Score

Experimental Groups	Mean	Std. Error	95% Confidence Interval	
			Lower Bound	Upper Bound
Discovery Group	39.705 ^a	1.385	36.914	42.496
Lecture Group	40.103 ^a	1.386	37.310	42.896
Control Group	31.067 ^a	1.387	28.271	33.863

a. Evaluated at covariates appeared in the model: Pre-test Integrated Science Achievement Score = 30.58.

This provides you with the adjusted means on the dependent variable for each of your groups. “Adjusted” tells you that the effect of the covariate has been statistically removed.

3.2.2 Two-way ANCOVA

Two-way ANCOVA is so called because it involves two independent variables which are categorical and consists of two or more levels or conditions. There is only one dependent variable which is continuous variable and one or more continuous covariates. Let us consider an example of the effect of teaching style on students’ achievement in mathematics. Let us also assume that there are 3 levels of teaching style and that the achievement test was given before the treatment (pre-test) and after the treatment (post-test). In this example, the teaching style becomes the independent variable, score on Mathematics test before treatment (pre-test) is the covariate and score on mathematics test after treatment (post-test) is the dependent variable. Assuming that the situation involves sex (boys and girls), the sex becomes a second independent variable. In this example involving two independent variables, one dependent variable and one covariate, the best statistics is two-way ANCOVA.

Interpretation of Computer Output from Two-Way ANCOVA

Six tables and a graph are generated for the two-way ANCOVA. Tables generated are interpreted to give pieces of information as follows:

1. The first table is Labelled **Descriptive statistics**. This gives the value of the mean, standard deviation and number of subjects for each group for the individual variables. In the output below, the mean and number of males and females in the different experimental groups and sex are given. This will also help you check for errors especially during data entry.

Descriptive Statistics

Dependent Variable: Post-test Integrated Science Achievement Score

Experimental Groups	Sex	Mean	Std. Deviation	N
Discovery Group	Males	40.00	7.540	8
	Females	39.38	4.955	8
	Total	39.69	6.172	16
Lecture Group	Males	40.00	5.855	8
	Females	40.13	6.770	8
	Total	40.06	6.115	16
Control Group	Males	31.50	4.408	8
	Females	30.75	3.655	8
	Total	31.13	3.931	16
Total	Males	37.17	7.100	24
	Females	36.75	6.661	24
	Total	36.96	6.813	48

2. Table Labelled **Levene's Test of Equality of Error Variances**
 This tests the non violation of the assumption of equality of variance. The interpretation is same as in previous units.

Levene's Test of Equality of Error Variances^a

Dependent Variable: Post-test Integrated Science Achievement Score

F	df1	df2	Sig.
1.730	5	42	.149

Tests the null hypothesis that the error variance of the dependent variable is equal across groups.

a. Design:

Intercept+TOTPRESC+EXPGROUP+SEX+EXPGROUP
 * SEX

3. Table labelled **Test of Between-Subjects Effects**
 This is the main two-way ANCOVA results. From this table you should consider the following important pieces of information:

- If there is a significant main effect of the two independent variables on the dependent variable
- If the interaction between the two independent variables is significant
- The influence of your covariates(s)

Tests of Between-Subjects Effects

Dependent Variable: Post-test Integrated Science Achievement Score

Source	Type III Sum of Squares	df	Mean Square	F	Sig.	Partial Eta Squared	Noncent. Parameter	Observed Power ^a
Corrected Model	836.064 ^b	6	139.344	4.245	.002	.383	25.470	.960
Intercept	1109.746	1	1109.746	33.807	.000	.452	33.807	1.000
TOTPRESC	14.397	1	14.397	.439	.512	.011	.439	.099
EXPGROUP	829.605	2	414.803	12.637	.000	.381	25.273	.994
SEX	1.512	1	1.512	.046	.831	.001	.046	.055
EXPGROUP * SEX	2.408	2	1.204	.037	.964	.002	.073	.055
Error	1345.853	41	32.826					
Total	67746.000	48						
Corrected Total	2181.917	47						

a. Computed using alpha = .05

b. R Squared = .383 (Adjusted R Squared = .293)

To do this you look at the column marked **sig** for each of the independent variables and interactions. If the value is equal or less than 0.05 then there is a significant difference, if value is greater than 0.05 then the differences are not significant.

- The effect size is derived from the corresponding eta squared value. You can determine how much of the variance in the dependent variable is explained by the independent variables.

4. Tables Labeled Estimated Marginal Means

These tables give you the adjusted means on the dependent variables for the groups after the effects of the covariate(s) has been statistically removed.

Estimated Marginal Means

1. Experimental Groups

Dependent Variable: Post-test Integrated Science Achievement Score

Experimental Groups	Mean	Std. Error	95% Confidence Interval	
			Lower Bound	Upper Bound
Discovery Group	39.705 ^a	1.433	36.812	42.599
Lecture Group	40.103 ^a	1.434	37.208	42.999
Control Group	31.066 ^a	1.435	28.168	33.965

a. Evaluated at covariates appeared in the model: Pre-test Integrated Science Achievement Score = 30.58.

2. Sex

Dependent Variable: Post-test Integrated Science Achievement Score

Sex	Mean	Std. Error	95% Confidence Interval	
			Lower Bound	Upper Bound
Males	37.136 ^a	1.170	34.772	39.500
Females	36.781 ^a	1.170	34.417	39.144

a. Evaluated at covariates appeared in the model: Pre-test Integrated Science Achievement Score = 30.58.

3. Experimental Groups * Sex

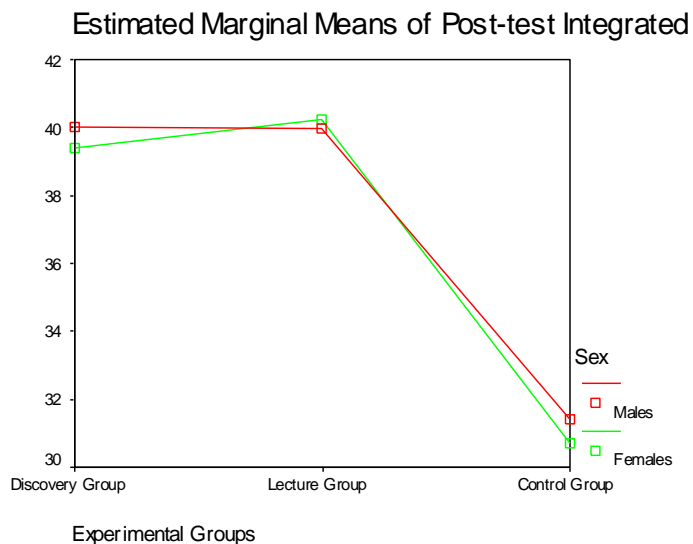
Dependent Variable: Post-test Integrated Science Achievement Score

Experimental Groups	Sex	Mean	Std. Error	95% Confidence Interval	
				Lower Bound	Upper Bound
Discovery Group	Males	40.025 ^a	2.026	35.934	44.117
	Females	39.385 ^a	2.026	35.294	43.476
Lecture Group	Males	39.964 ^a	2.026	35.872	44.057
	Females	40.242 ^a	2.033	36.136	44.349
Control Group	Males	31.418 ^a	2.029	27.320	35.517
	Females	30.714 ^a	2.026	26.622	34.807

a. Evaluated at covariates appeared in the model: Pre-test Integrated Science Achievement Score = 30.58.

Sometimes you may request for a plot (graph) of the adjusted means. The plot is presented below.

Profile Plots



3.2.3 Multivariate ANCOVA

Multivariate ANOVA or MANOVA

This is called multivariate because it has more than one (many) dependent variables. For example you may wish to find out the effect of different methods of teaching on students performance and interest in science. Here you have just one independent variable – methods of teaching which must have more than 2 levels. (e.g. lecture, inquiring and demonstration) and two dependent variables – performance in science and interest in science (Multi-dependent variable) In this example with just one independent variable it is known as one-way MANOVA. If there are **two** independent variables, it is known as **two-way** MANOVA, more than two independent variables it is **higher factorial design**.

What do you need?

- one, two or more independent (categorical variables)
- Two or more continuous dependent variable.

4.0 Conclusions

Analysis of covariance is an extension of analysis of variance. It allows you to explore differences between groups while controlling for an additional variable.

5.0 Summary

In this unit we have learnt about Analysis of Covariance. We discussed the assumptions of ANCOVA and types of ANCOVA. We also learnt how to interpret computer output for ANCOVA.

6.0 Tutor-Marked Assignment

- 1) List and explain two major differences between an Analysis of Variance and Analysis of Covariance
- 2) What are key issues and assumptions associated with ANCOVA
- 3) In a study to reduce fear of statistics two groups of students made up of male and female students were treated with two programmes mathematics skills and confidence building. An instrument to measure the students' fear of statistics before the intervention was used. Data on fear of statistics was collected before and after the intervention. Part of the output for the analysis is:

Descriptive Statistics

Intervention	Sex	Mean	Std. Deviation	N
Maths Skills	Male	37.25	5.50	8
	Female	38.14	3.44	7
	Total	37.67	4.51	15
Confidence Building	Male	40.57	5.56	7
	Female	34.50	4.78	8
	Total	37.33	5.88	15
Total	Male	38.80	5.60	15
	Female	36.20	4.48	15
	Total	37.50	5.15	30

Part of Test of Between Subject Effects

Source	Type III Sum of Squares	df	Mean Square	f	Sig.
Fear of Statistics 1					
Group	545.299	1	545.299	164.698	.000
Sex	4.739	1	4.739	1.431	.243
Group* Sex	4.202	1	4.202	1.269	.271
Error	104.966	1	104.966	31.703	.000
	82.772	25	3.311		

All assumptions satisfied, interpret the results in the output above.

7.0 References/Further Readings

Adamu, S. O. (1997), *Statistics for Beginners Book 1*, Nigeria, Ibadan, SAAL Pub.

Hinton, P. R. (2003), *Statistics Explained A Guide for Social Science Students*, Hove and New York, Routledge

Owie, I. (2006), *Fundamentals of Statistics in Education and the Social Sciences*, Nigeria, Lagos, National Book Consortium

Kerlinger, F. (1986), *Foundations of Behavioural Research*, New York: McGraw-Hill

<http://www.gower.k12.il.us/Staff/ASSESS>

www.apsu.edu/oconnort/3760/3760lect03a.htm

faculty.vassar.edu/lowry/webtext.html

www.uwsp.edu/psych/stat/14/nonparm.htm

writing.colostate.edu/guides/research/stats/pop2a.cfm

www.shodor.org/interactivate/lessons/IntroStatistics

www.statbasics.co.uk

MODULE 4 STATISTICAL TECHNIQUES TO EXPLORE RELATIONSHIPS AMONG VARIABLES

- Unit 1: Correlation
- Unit 2: Regression
- Unit 3: Trend Analysis

UNIT 1 Correlation

CONTENTS

- 1.0 Introduction
- 2.0 Objectives
- 3.0 Main Content
 - 3.1 Understanding Correlation
 - 3.1.1 Types of Relationship
 - 3.1.2 Graphical Representations of Relationships
 - 3.1.3 Assumptions of Correlation
 - 3.2 The Measurement of Linear Correlation
 - 3.2.1 Pearson Product- Moment Correlation Coefficient
 - 3.2.2 Coefficient of Determination
 - 3.2.3 Interpretation of Computer Output for Pearson Product-Moment Correlation
 - 3.3 The Measurement of Partial Correlation
 - 3.3.1 Partial Correlation
 - 3.3.2 Interpretation of Computer output on Partial Correlation
- 4.0 Conclusions
- 5.0 Summary
- 6.0 Tutor-Marked Assignment
- 7.0 References/Further Readings

1.0 INTRODUCTION

In modules 3 we looked at techniques for testing statistical significance of differences between and among means. There are situations when the interest is on exploring relationship among variables. In this section we will focus on dictating and describing relationships among variables. These techniques are mostly used by researchers engaged in non-experimental research designs. Unlike experimental designs, variables are not

deliberately manipulated or controlled; rather they are described as they exist naturally. In this unit we consider the technique known as Correlation.

2.0 OBJECTIVES

At the end of this unit you should be able to:

- Describe the different types of relationships
- List the different forms of relationship
- Differentiate between correlation and causality
- Calculate Pearson product-moment correlation coefficient
- Calculate partial correlation coefficient
- Calculate coefficient of determination
- Interpret correlation

3.0 MAIN CONTENT

3.1 Understanding Correlation

Correlation is a description of the strength and direction of a linear relationship. In correlation you are interested in exploring the association between pairs of variables. Let us begin by considering two variables, X and Y , in the case where each particular value of X_i is paired with one particular value of Y_i . For example: the measures of height for individual human subjects, paired with their corresponding measures of weight; the number of hours that individual students in a statistics course spend studying prior to an exam, paired with their corresponding measures of performance on the exam; the amount of class time that individual students in a statistics course spend snoozing and daydreaming prior to an exam, paired with their corresponding measures of performance on the exam; and so on. These are examples of what is known as relationship. We will consider the different types and forms of these relationships.

3.1.1 Types of Relationship

Relationship can be of different types and forms. There are two major types of relationship – linear and curvilinear:

1. Linear Relationship

A relationship that can be described by a straight line is spoken of as *linear* (short for 'rectilinear'). The various forms that linear correlation is capable of taking are:

Positive correlation: The plot for a positive correlation, on the other hand, will reflect the tendency for high values of X_i to be associated with high values of Y_i , and vice versa; hence, the data points will tend to line up along an upward slanting diagonal.

Negative correlation - The plot for negative correlation will reflect the opposite tendency for high values of X_i to be associated with low values of Y_i , and vice versa; hence, the data points will tend to line up along a downward slanting diagonal.

Zero correlation - In the case of zero correlation, the coordinate plot will look something like the rather pattern-less jumble shown in Figure 3.2a, reflecting the fact that there is no systematic tendency for X and Y to be associated, either the one way or the other.

Perfect correlation

When the data points line up along the diagonal like beads on a taut string it is typically spoken of as *perfect* correlation.

2. Curvilinear Relationship

This is a relationship that can be described by a curved line.

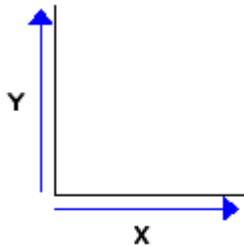
3.1.2 Graphical Representation of Relationships

You can plot the relationship between two variables graphically. This coordinate plot is known **scatterplot** or **scattergram**. Scatterplot is a standard method for graphically representing the relationship that exists between two variables, X and Y, in the case where each particular value of X_i is paired with one particular value of Y_i . This is a **bivariate list**.

Steps in Constructing a Bivariate Coordinate Plot

When you have a bivariate list, you have to follow the following steps to construct a coordinate plot:

- **Lay out two axes at right angles to each other.** By convention, the horizontal axis is assigned to the X variable and the vertical axis to the Y variable, with values of X increasing from left to right and values of Y increasing from bottom to top.

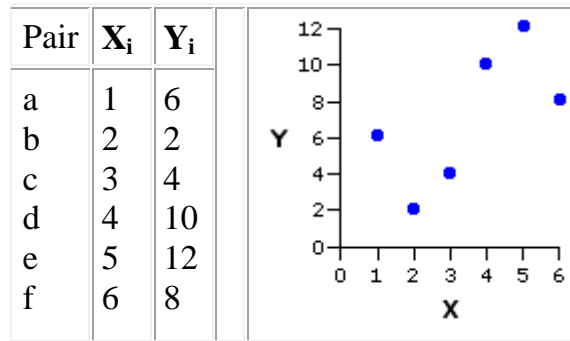


In examining the relationship between two causally related variables, the **independent variable** is the one that is capable of *influencing* the other, and the **dependent variable** is the one that is capable of being *influenced* by the other. For example, the amount of time you spend studying before an exam can affect your subsequent performance on the exam, whereas your performance on the exam cannot retroactively affect the prior amount of time you spent studying for it. Hence, amount of study is the independent variable and performance on the exam is the dependent variable. By convention X axis is left for the independent variable and the Y axis for the dependent variable. For cases where the distinction between "independent" and "dependent" does not apply, it makes no difference which variable is called X and which is called Y.

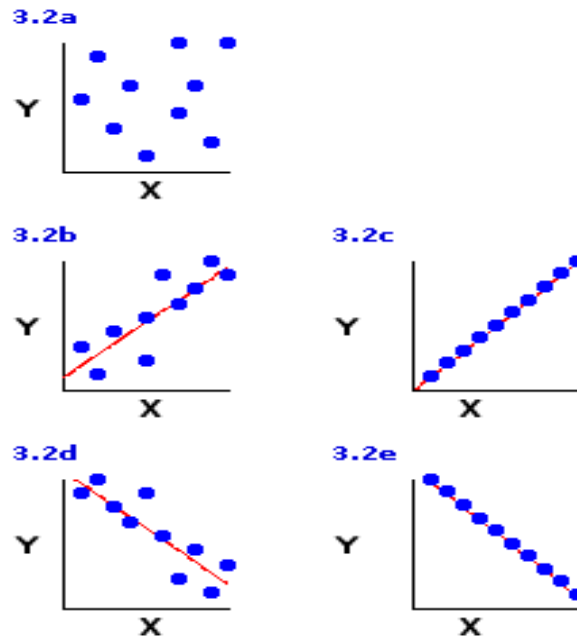
In designing a coordinate plot of this type, it is not generally necessary to begin either the X or the Y axis at zero. The X axis can begin at or slightly below the lowest observed value of X_i , and the Y axis can begin at or slightly below the lowest observed value of Y_i .

- Plot the various values of X against Y

Figure below illustrates a scatter plot.



The different forms of relationships can also be represented graphically as shown in the figures below



3.2a is a case of zero relationship. The coordinate plot looks rather patternless, reflecting the fact that there is no systematic tendency for X and Y to be associated, either the one way or the other. 3.2b is a case of a positive relationship. This reflects the tendency for high values of X_i to be associated with high values of Y_i , and vice versa; hence, the data points will tend to line up along an upward slanting diagonal. 3.2d is a case of negative relationship. This reflects the opposite tendency for high values of X_i to be

associated with low values of Y_i , and vice versa; hence, the data points will tend to line up along a downward slanting diagonal. In 3.2c and 3.2e data points line up along the diagonal like beads on a string. This arrangement is typically spoken of as *perfect* relationship. These would represent the maximum degree of relationship that could possibly exist between two variables. In the real world you will normally find perfect relationship only in the realm of basic physical principles; for example, the relationship between voltage and current in an electrical circuit with constant resistance. Among the less tidy phenomena of the behavioral and biological sciences, positive and negative relationships are much more likely to occur.

3.1.3 Assumptions of Correlation

There are a number of assumptions for correlation. These are:

- Level of Measurement – The scale of measurement should be interval or ratio scale (except when you have one dichotomous variable such as sex)
- Related Pairs – each subject must provide a score on both variable X and variable Y
- Independence of Observation – the observations that make up your data must be independent of one another.
- Scores on each variable should be normally distributed
- The relationship must be linear

3.2 The Measurement of Linear Correlation

3.2.1 Pearson Product-Moment Correlation Coefficient

The primary measure of linear correlation is the **Pearson product-moment correlation coefficient**, symbolized by the lower-case Roman letter r , which ranges in value from $r=+1.0$ for a perfect positive correlation to $r=-1.0$ for a perfect negative correlation. The midpoint of its range, $r=0.0$, corresponds to a complete absence of correlation. Values falling between $r=0.0$ and $r=+1.0$ represent varying degrees of positive correlation, while those falling between $r=0.0$ and $r=-1.0$ represent varying degrees of negative correlation.

The raw measure of the tendency of two variables, X and Y, to co-vary is a quantity known as the **covariance**. Covariance is a measure of the degree to which two variables, X and Y, co-vary. Pearson product-moment correlation coefficient is a simple ratio between:

- the amount of covariation between X and Y that is actually observed, and
- the amount of covariation that *would* exist if X and Y had a perfect (100%) positive correlation.

Thus

$$r = \frac{\textit{observed covariance}}{\textit{maximum possible positive covariance}}$$

The quantity listed above as "maximum possible positive covariance" is precisely determined by the two separate variances of X and Y. Specifically, the maximum possible positive covariance that can exist between two variables is equal to the *geometric mean* of the two separate variances. So the structure of the relationship now comes down to:

$$r = \frac{\text{observed covariance}}{\text{sqrt}[(\text{variance}_X) \times (\text{variance}_Y)]}$$

(where sqrt. = square root)

Although in principle this relationship involves two variances and a covariance, in practice, through the magic of algebraic manipulation, it boils down to something that is computationally much simpler.

The third item, SC_{XY} , denotes a quantity that we will speak of as the **sum of co-deviates**; and as you can no doubt surmise from the name, it is something very closely akin to a sum of squared deviates. SS_X is the raw measure of the variability among the values of X_i ; SS_Y is the raw measure of the variability among the values of Y_i ; and SC_{XY} is the raw measure of the **co**-variability of X and Y together.

$$r = \frac{SC_{XY}}{\text{sqrt}[SS_X \times SS_Y]}$$

From our previous Modules and units you will recall that For any

particular item in a set of measures of the variable X, $\text{deviate}_X = X_i - \mathbf{M}_X$; Similarly, for any particular item in a set of measures of the variable Y, $\text{deviate}_Y = Y_i - \mathbf{M}_Y$

The relationship between codeviates and deviates is given by:

$$\text{co-deviate}_{XY} = (\text{deviate}_X) \times (\text{deviate}_Y)$$

And finally, the analogy between a co-deviate and a squared deviate:

For a value of X_i , the squared deviate is $(\text{deviate}_X) \times (\text{deviate}_X)$

For a value of Y_i it is $(\text{deviate}_Y) \times (\text{deviate}_Y)$

And for a pair of X_i and Y_i values, the co-deviate is $(\text{deviate}_X) \times (\text{deviate}_Y)$

This should give you a sense of the underlying concepts. Just keep in mind, no matter what particular computational sequence you follow when you calculate the correlation coefficient, that what you are fundamentally calculating is the ratio

$$r = \frac{\text{observed covariance}}{\text{maximum possible positive covariance}}$$

which, for computational purposes, comes down to

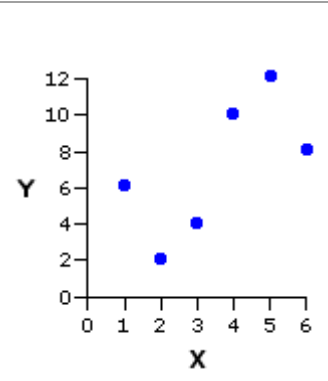
$$r = \frac{SC_{XY}}{\text{sqrt}[SS_X \times SS_Y]}$$

Example

Consider the pairs of X and Y as shown below, the steps for calculating r are:

Step 1: Calculating the square of each value of X_i and Y_i , along with the cross-product of each X_iY_i pair. These are the items that will be required for the calculation of the three summary quantities in the above formula: SS_X , SS_Y , and SS_{XY} .

Pair	X_i	Y_i	X_i^2	Y_i^2	X_iY_i
a	1	6	1	36	6
b	2	2	4	4	4
c	3	4	9	16	12
d	4	10	16	100	40
e	5	12	25	144	60
f	6	8	36	64	48
sums	21	42	91	364	170



Step 2 Calculate the sum of squared deviates for X_i and Y_i values using the formulae

$$SS_X = \sum X_i^2 - \frac{(\sum X_i)^2}{N}$$

$$\begin{aligned}
 N &= 6 \text{ [because there are 6 values of } X_i\text{]} \\
 \sum X_i^2 &= 91 \\
 \sum X_i &= 21 \\
 (\sum X_i)^2 &= (21)^2 = 441
 \end{aligned}$$

Thus:

$$SS_X = 91 - (441/6) = 17.5$$

Similarly, the sum of squared deviates for a set of Y_i values can be calculated according to the formula

$$SS_Y = \sum Y_i^2 - \frac{(\sum Y_i)^2}{N}$$

$$\begin{aligned} N &= 6 \text{ [because there are 6 values of } Y_i\text{]} \\ \sum Y_i^2 &= 364 \\ \sum Y_i &= 42 \\ (\sum Y_i)^2 &= (42)^2 = 1764 \end{aligned}$$

Thus:

$$SS_Y = 364 - (1764/6) = 70.0$$

Step 3: Calculate (SC_{XY}) sum of co-deviates for paired values of X_i and Y_i using the formula:

$$SC_{XY} = \sum (X_i Y_i) - \frac{(\sum X_i)(\sum Y_i)}{N}$$

$$\begin{aligned} N &= 6 \text{ [because there are 6 } X_i Y_i \text{ pairs]} \\ \sum X_i &= 21 \\ \sum Y_i &= 42 \\ (\sum X_i)(\sum Y_i) &= 21 \times 42 = 882 \\ \sum (X_i Y_i) &= 170 \end{aligned}$$

Thus:

$$SC_{XY} = 170 - (882/6) = 23.0$$

Step 4: Calculate the Correlation Coefficient using the formula:

$$r = \frac{SC_{XY}}{\sqrt{SS_X \times SS_Y}}$$

$$= \frac{23.0}{\sqrt{17.5 \times 70.0}} = +0.66$$

3.2.2 Coefficient of Determination

A closely related companion measure of linear correlation is the **coefficient of determination**, symbolized as r^2 , which is simply the square of the correlation coefficient. The coefficient of determination can have only positive values ranging from $r^2=+1.0$ for a perfect correlation (positive or negative) down to $r^2=0.0$ for a complete absence of correlation. The advantage of the correlation coefficient, r , is that it can have either a positive or a negative sign and thus provide an indication of the positive or negative *direction* of the correlation. The advantage of the coefficient of determination, r^2 , is that it provides an equal interval and ratio scale measure of the *strength* of the correlation. In effect, the correlation coefficient, r , gives you the true direction of the correlation (+ or —) but only the square root of the strength of the correlation; while the coefficient of determination, r^2 , gives you the true strength of the correlation but without an indication its direction. Both of them together give you the whole works

From the example in the previous section 3.2.1:

$$r^2 = (+0.66)^2 = 0.44$$

3.2.3 Interpretation of Computer Output for Pearson Product-Moment Correlation

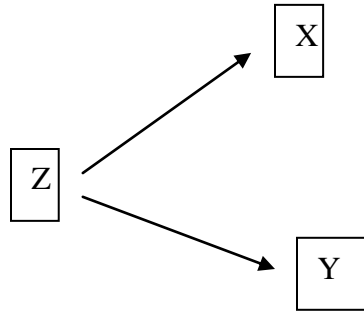
Correlations

		Score on Reading	Score in Exam
Score on Reading	Pearson Correlation	1	.010
	Sig. (2-tailed)	.	.949
	N	48	48
Score in Exam	Pearson Correlation	.010	1
	Sig. (2-tailed)	.949	.
	N	48	48

- Step 1 Check and make sure the N (number of cases) is correct.
- Step 2 Determine the direction of the relationship. What sign is in front of r? Is it positive or negative? In the table above there is a positive relationship. Meaning that increase reading score increase exam score.
- Step 3 Determine the strength of the relationship. This has to do with the size of the value of r. Cohen (1988) suggest:
 $r = .10$ to $.29$ or $r = -.10$ to $-.29$ is small relationship
 $r = .30$ to $.49$ or $r = -.30$ to $-.49$ is medium relationship
 $r = .50$ to 1.0 or $r = -.50$ to -1.0 is large relationship
 From table above the relationship is extremely small (.010)
- Step 4 Calculate the Coefficient of Determination. This is simply the square of r. You can convert this to percentage by multiplying by 100. From our example it is:
 $r^2 = .010 \times .010 = .0001$
 $= 100 \times .0001 = .01\%$
- Step 5 Assess the significant level – This is from the row marked **(Sig. 2 tailed)**. Report this but ignore it since it is greatly influenced by sample size.

3.3 The Measurement of Partial Correlation

Partial correlation is similar to Pearson product-moment correlation, except that it allows you to control for an additional variable. The additional variable is usually a variable that you suspect might be influencing your two variables of interest. Let us consider two variables X and Y. The relationship between the two variables is influence by a third one Z. Graphically the relationship is:



Partial correlation statistically helps to remove the influence of the third variable which if left to exist will inflate the correlation between the two variables of interest.

For partial correlation you need three continuous variables. Two of the variables are those you wish to explore their relationship while the third is that you will control.

3.3.1 The Measurement of Partial Correlation

Suppose you measured each of N subjects on each of your three variables, X , Y , and Z , and found the following correlations:

$$X \text{ versus } Y: r_{XY} = +.50 \quad r^2_{XY} = .25$$

$$X \text{ versus } Z: r_{XZ} = +.50 \quad r^2_{XZ} = .25$$

$$Y \text{ versus } Z: r_{YZ} = +.50 \quad r^2_{YZ} = .25$$

Considering the value of r^2 , which in each case is equal to .25, it means that for each pair of variables— XY , XZ , and YZ —the covariance, or variance overlap, is 25%. Partial correlation is a procedure that allows us to measure the region of three-way overlap precisely, and then to remove it from the picture in order to determine what the correlation between any two of the variables would be (hypothetically) **if** they were not each correlated with the third variable. Alternatively, you can say that partial correlation allows us to determine what the correlation between any two of the variables would be (hypothetically) **if** the third variable were held constant.

The partial correlation of X and Y , with the effects of Z removed (or held constant), would be given by the formula

$$r_{XY \cdot Z} = \frac{r_{XY} - (r_{XZ})(r_{YZ})}{\sqrt{1 - r_{XZ}^2} \times \sqrt{1 - r_{YZ}^2}}$$

Considering our example it would work out as:

$$r_{XY \cdot Z} = \frac{.50 - (.50)(.50)}{\sqrt{1 - .25} \times \sqrt{1 - .25}}$$

$$= +.33$$

$$\text{Hence } r^2_{XY \cdot Z} = .11$$

3.3.2 Interpretation of Computer output on Partial Correlation

There are two outputs as shown below:

--- PARTIAL CORRELATION COEFFICIENTS ---
-

Zero Order Partial

	PRESCOA	TOTPOSCO	PRESCOB
PRESCOA	1.0000	.0096	.4308
	(0)	(46)	(46)
	P= .	P= .949	P= .002
TOTPOSCO	.0096	1.0000	.0584
	(46)	(0)	(46)
	P= .949	P= .	P= .693
PRESCOB	.4308	.0584	1.0000
	(46)	(46)	(0)

P= .002 P= .693 P= .

(Coefficient / (D.F.) / 2-tailed Significance)

". " is printed if a coefficient cannot be computed

-

- - - P A R T I A L C O R R E L A T I O N C O E F F I C I E N T S - -

-

Controlling for.. PRESCOB

PRESCOA TOTPOSCO

PRESCOA 1.0000 -.0173
(0) (45)
P= . P= .908

TOTPOSCO -.0173 1.0000
(45) (0)
P= .908 P= .

(Coefficient / (D.F.) / 2-tailed Significance)

". " is printed if a coefficient cannot be computed

The first is the normal Person product-moment correlation between the two variables of interest. In this case it is .0096 or .01. In the second matrix the third variable has been controlled for and the new partial correlation given as -.0173. This has changed the direction of the relationship.

4.0 CONCLUSIONS

There are a range of techniques available to explore relationships. These vary according to the type of research question that needs to be addressed and the type of data available. Correlation is used when you wish to describe the strength and direction of the relationship between two variables (usually continuous). It can also be used when one of the variables is dichotomous (that is it has only two values).

5.0 SUMMARY

In this unit we learn of different types and forms of relationships. We also learnt how to calculate correlation coefficient. We practiced the calculation

of Pearson product-moment coefficient and partial correlation coefficient. Interpretation of computer out-put for correlation was also considered.

6.0 TUTOR-MARKED ASSIGNMENT

- (1) The following data are marks obtained by 10 students out of a maximum of 10 marks for each subject:

Botany	3	6	4	6	4	7	5	5
	4	7						
Zoology	4	6	5	7	4	7	6	6
	5	8						

Draw the scatter diagram and calculate the correlation co-efficient

- (2) The following data are scores made on teaching and administrative abilities of 10 headmasters in some secondary schools in Lagos:

Teaching Ability (X)	7	2	6	3	6	6
	6	1	1	3		
Admin. Ability (Y)	5	1	3	4	8	3
	6	0	2	4		

Show that there is a relationship between teaching and administrative abilities by:

- Drawing the scatter diagram of the data
- Calculating the coefficient of correlation

7.0 REFERENCES/FURTHER READINGS

Adamu, S. O. (1997), *Statistics for Beginners Book 1*, Nigeria, Ibadan, SAAL Pub.

Hinton, P. R. (2003), *Statistics Explained A Guide for Social Science Students*, Hove and New York, Routledge

Owie, I. (2006), *Fundamentals of Statistics in Education and the Social Sciences*, Nigeria, Lagos, National Book Consortium

Kerlinger, F. (1986), *Foundations of Behavioural Research*, New York: McGraw-Hill

<http://www.gower.k12.il.us/Staff/ASSESS>

www.apsu.edu/oconnort/3760/3760lect03a.htm

faculty.vassar.edu/lowry/webtext.html

writing.colostate.edu/guides/research/stats/pop2a.cfm

UNIT 2: REGRESSION

CONTENTS

- 1.0 Introduction
- 2.0 Objectives
- 3.0 Main Content
 - 3.1 Understanding Regression
 - 3.1.1 Types of Regression Models
 - 3.1.2 Major Types of Multiple Regressions
 - 3.1.3 Assumptions of Multiple Regressions
 - 3.2 Interpretations of Computer Output of different Multiple Regression
 - 3.2.1 Standard Multiple Regression
 - 3.2.2 Hierarchical Multiple Regression
 - 3.2.3. Stepwise Multiple Regression
- 4.0 Conclusions
- 5.0 Summary
- 6.0 Tutor-Marked Assignment
- 7.0 References/Further Readings

1.0 INTRODUCTION

Regression analysis refers to a class of statistical techniques that study relationship between a criterion (dependent variable) ‘Y’ and one or more predictor (independent variables) ‘X’. In the last unit you learnt about correlation. You learnt that the theory of interdependence lead to the theory of correlation. In this unit you will learn of regression which relates to the theory of dependence. The unit will not go into details of calculation; rather the emphasis will be computer analysis which is mostly used these days in

research. If you are interested in detailed manual computation, you are advised to consult any good statistical textbook.

2.0 OBJECTIVES

At the end of this unit you should be able to:

- Differentiate between correlation and regression
- Write regression equations
- Explain the different regression models
- List the assumptions of multiple regression
- Interpret computer outputs for the different types of multiple regression

3.0 MAIN CONTENT

3.1 Understanding Regression

The means by which a regression study is effected is known as regression equation. The general form of regression equation is:

$$\hat{y} = f(X_1, X_2, \dots, X_p)$$

where the circumflex on the y denotes that what is being represented by the function of the X's is a predicted or modeled y-value rather than an actually observed one.

The function $f ()$ may be a simple linear function of a single predictor or complicated weighted sum of several predictors raised to various powers and may also include products of two or more predictors

3.1.1 Types of Regression Models

There are different types of regression models:

- Simple Linear Model – This takes the form of a single-predictor linear equation

$$\hat{y} = b_0 + b_1X$$

This is the simplest regression model.

- Multiple Linear Regression – When you add a second predictor variable to a simple linear model you get what is known as multiple linear regression. This takes the form of

$$\hat{y} = b_0 + b_1X_1 + b_2 + X_2$$

- Polynomial Regression – It is also possible to make the linear regression model more complex by adding successively a term in X^2 , X^3 etc while holding the number of actual predictor variable to one. Such regression equation is known as polynomial regression.
- Multiple Non-linear Regression – This is a model that combines the complexities of multiple linear and polynomial regression models

3.1.2 Major Types of Multiple Regressions

There are a number of multiple regression types that you can use for your analysis depending on the nature of the questions you wish to address:

- Standard Multiple Regression – This is when all the independent (predictor) variables are entered into a regression equation simultaneously. This is the most widely used type.
- Hierarchical or sequential Multiple Regression – In this type the independent variables are entered in an order specified by a researcher. For example if you are interest in knowing how well ones reading predicts performance in an exam after the effect of age is controlled for, you will enter age as block 1 and reading as block 2
- Stepwise Multiple Regression – In stepwise model, the statistical programme is allowed to select which variable will enter the regression equation first and in which order all the variables will go. This is unlike the hierarchical where the researcher determines which goes first.

3.1.3 Assumptions of Multiple Regressions

The major assumptions of multiple regression are:

- **Sample size** – The issue here is that of generalization. Many authors argue on the sample size for multiple regressions. Stevens (1996, p. 72) recommends a sample size of 15 subjects per predictor for a

reliable equation. Tabachnick and Fidell (1996, p. 132) gave a formula for calculating sample size as $N = 50 + 8m$ (where m = number of independent variables) If you have say 4 variables you will need 82 cases. For stepwise it is suggested that 40 cases are needed per every independent variables.

- **Multicollinearity and Singularity** – Multiple regression does not like multicollinearity and singularity. Multicollinearity exists when the independent variables are highly correlated. ($r = .9$ and above). Singularity is when one independent variable is actually a combination of other independent variables.
- **Outliers** – Multiple regression is very sensitive to outliers (very high and very low scores). They should be removed in the data for regression.
- **Normality, linearity, homoscedasticity, independence of residuals** – These all refer to various aspects of distribution of scores. The scores should be normally distributed, have a straight line relationship and the predicted dependent variable scores should be the same for all predicted scores.

3.2 Interpretations of Computer Output of different Multiple Regression

When assumptions are satisfied, the computer programme will generate output for the regression analysis. In the following sections we will interpret the outputs for the different types of regression.

3.2.1 Standard Multiple Regression

Correlations

		Post-test Score for Section A	Post-test Score for Section B	Post-test Score for Section C
Pearson Correlation	Post-test Score for Section A	1.000	.706	.496
	Post-test Score for Section B	.706	1.000	.565
	Post-test Score for Section C	.496	.565	1.000
Sig. (1-tailed)	Post-test Score for Section A	.	.000	.000
	Post-test Score for Section B	.000	.	.000
	Post-test Score for Section C	.000	.000	.
N	Post-test Score for Section A	48	48	48
	Post-test Score for Section B	48	48	48
	Post-test Score for Section C	48	48	48

Model Summary^b

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	.716 ^a	.513	.491	2.197

a. Predictors: (Constant), Post-test Score for Section C, Post-test Score for Section B

b. Dependent Variable: Post-test Score for Section A

ANOVA^b

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	228.451	2	114.225	23.664	.000 ^a
	Residual	217.216	45	4.827		
	Total	445.667	47			

a. Predictors: (Constant), Post-test Score for Section C, Post-test Score for Section B

b. Dependent Variable: Post-test Score for Section A

Coefficients^a

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.	Collinearity Statistics	
		B	Std. Error	Beta			Tolerance	VIF
1	(Constant)	.538	1.841		.292	.771		
	Post-test Score for Section B	.790	.159	.625	4.958	.000	.681	1.468
	Post-test Score for Section C	.179	.157	.143	1.136	.262	.681	1.468

a. Dependent Variable: Post-test Score for Section A

Collinearity Diagnostics^a

Model	Dimension	Eigenvalue	Condition Index	Variance Proportions		
				(Constant)	Post-test Score for Section B	Post-test Score for Section C
1	1	2.964	1.000	.00	.00	.00
	2	1.987E-02	12.214	.99	.17	.26
	3	1.636E-02	13.459	.00	.82	.74

a. Dependent Variable: Post-test Score for Section A

Residuals Statistics^a

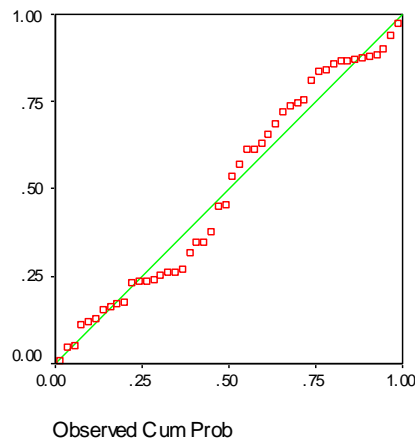
	Minimum	Maximum	Mean	Std. Deviation	N
Predicted Value	9.26	17.19	12.42	2.205	48
Std. Predicted Value	-1.432	2.166	.000	1.000	48
Standard Error of Predicted Value	.320	.844	.537	.118	48
Adjusted Predicted Value	9.20	16.96	12.42	2.204	48
Residual	-5.06	4.23	.00	2.150	48
Std. Residual	-2.304	1.927	.000	.978	48
Stud. Residual	-2.487	1.987	-.001	1.016	48
Deleted Residual	-5.90	4.50	.00	2.320	48
Stud. Deleted Residual	-2.648	2.058	-.003	1.032	48
Mahal. Distance	.016	5.963	1.958	1.313	48
Cook's Distance	.000	.341	.027	.052	48
Centered Leverage Value	.000	.127	.042	.028	48

a. Dependent Variable: Post-test Score for Section A

Charts

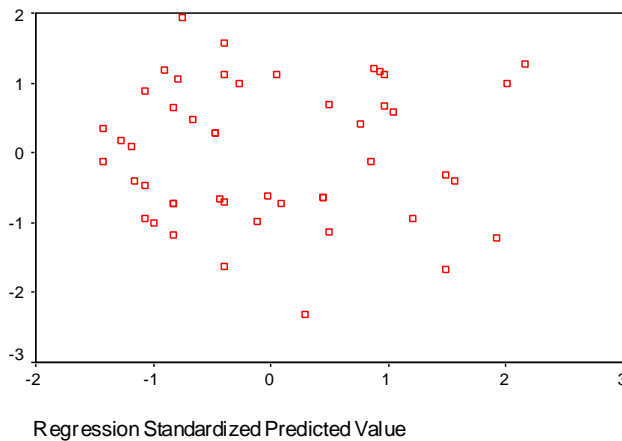
Normal P-P Plot of Regression Standard

Dependent Variable: Post-test Score for



Scatterplot

Dependent Variable: Post-test Score for Section



Step 1 Checking the Assumptions

From the table labelled Correlation confirm that your independent variable show some relationship with the dependent variable (at least 0.3). Also check that the correlation between your independent variables is not too high (7 and above). In this example you have $r = .706$, $.496$ and $.565$ respectively. You can also check for assumptions from table labelled **Coefficient**. Look for the values given under the column headed **Tolerance**. If this value is very low (near 0) then this indicates that the multiple correlation with other variables is high. In this example it is quite respectable ($.681$), so we do not appear to have violated the assumption of multicollinearity.

Examine the scatter plot and the normal probability plot. In the normal probability plot you expect that your points will lie in a reasonable straight diagonal line from bottom left to top right. This will suggest no major deviation from normality.

Step 2 Evaluating the Model

Look at the model summary and check the value given under the heading **R Square**. This tells you how much of the variance in the dependent variable is by the model. In the example it is .513 or 51.3%. This is quite respectable result.

Step 3 Evaluating each of the Independent Variables

This enables you to know which of the variables included in the model contributed to the prediction of the dependent variable. This you will find from the table labelled **Coefficient**. You look in the column labelled **Beta** under **Standard Coefficients**. Look for the variable that made the largest contribution. In our example it is Post test B (.625). This variable also made a significant unique contribution because the column marked **Sig.** is (.000).

3.2.2 Hierarchical Multiple Regression

Step 1 Evaluating the Model

Check **R Square** from the table labelled Model Summary. In this case we have .499 (49.9%) after block1 variable has been entered and .513 (51.3%) after block 2. After the third variable was controlled you have .014

Model Summary

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate	Change Statistics				
					R Square Change	F Change	df 1	df 2	Sig. F Change
1	.706 ^a	.499	.488	2.204	.499	45.750	1	46	.000
2	.716 ^b	.513	.491	2.197	.014	1.290	1	45	.262

a. Predictors: (Constant), Post-test Score for Section B

b. Predictors: (Constant), Post-test Score for Section B, Post-test Score for Section C

ANOVA^c

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	222.226	1	222.226	45.750	.000 ^a
	Residual	223.441	46	4.857		
	Total	445.667	47			
2	Regression	228.451	2	114.225	23.664	.000 ^b
	Residual	217.216	45	4.827		
	Total	445.667	47			

a. Predictors: (Constant), Post-test Score for Section B

b. Predictors: (Constant), Post-test Score for Section B, Post-test Score for Section C

c. Dependent Variable: Post-test Score for Section A

Step 2 Evaluating each of the Independent Variables

This will help you find out how well each of the variables contributed to the equations. You look at the table marked **Coefficients**. Look in the model 2 which summarizes all the results with all the variables entered. In this example only post test B made a significant contribution.

Coefficients^a

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.
		B	Std. Error	Beta		
1	(Constant)	1.484	1.647		.901	.372
	Post-test Score for Section B	.892	.132	.706	6.764	.000
2	(Constant)	.538	1.841		.292	.771
	Post-test Score for Section B	.790	.159	.625	4.958	.000
	Post-test Score for Section C	.179	.157	.143	1.136	.262

a. Dependent Variable: Post-test Score for Section A

Excluded Variables^b

Model		Beta In	t	Sig.	Partial Correlation	Collinearity Statistics
						Tolerance
1	Post-test Score for Section C	.143 ^a	1.136	.262	.167	.681

a. Predictors in the Model: (Constant), Post-test Score for Section B

b. Dependent Variable: Post-test Score for Section A

3.2.3. Stepwise Multiple Regression

Descriptive Statistics

	Mean	Std. Deviation	N
Post-test Score for Section A	12.42	3.079	48
Post-test Score for Section B	12.25	2.436	48
Post-test Score for Section C	12.29	2.466	48

Correlations

		Post-test Score for Section A	Post-test Score for Section B	Post-test Score for Section C
Pearson Correlation	Post-test Score for Section A	1.000	.706	.496
	Post-test Score for Section B	.706	1.000	.565
	Post-test Score for Section C	.496	.565	1.000
Sig. (1-tailed)	Post-test Score for Section A	.	.000	.000
	Post-test Score for Section B	.000	.	.000
	Post-test Score for Section C	.000	.000	.
N	Post-test Score for Section A	48	48	48
	Post-test Score for Section B	48	48	48
	Post-test Score for Section C	48	48	48

Variables Entered/Removed^a

Model	Variables Entered	Variables Removed	Method
1	Post-test Score for Section B	.	Stepwise (Criteria: Probability-of-F-to-enter <= .050, Probability-of-F-to-remove >= .100).

a. Dependent Variable: Post-test Score for Section A

Model Summary

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	.706 ^a	.499	.488	2.204

a. Predictors: (Constant), Post-test Score f or Section B

ANOVA^b

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	222.226	1	222.226	45.750	.000 ^a
	Residual	223.441	46	4.857		
	Total	445.667	47			

a. Predictors: (Constant), Post-test Score for Section B

b. Dependent Variable: Post-test Score f or Section A

Coefficients^a

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.
		B	Std. Error	Beta		
1	(Constant)	1.484	1.647		.901	.372
	Post-test Score f or Section B	.892	.132	.706	6.764	.000

a. Dependent Variable: Post-test Score f or Section A

Excluded Variables^b

Model		Beta In	t	Sig.	Partial Correlation	Collinearity Statistics Tolerance
1	Post-test Score f or Section C	.143 ^a	1.136	.262	.167	.681

a. Predictors in the Model: (Constant), Post-test Score f or Section B

b. Dependent Variable: Post-test Score f or Section A

The interpretation of the tables are similar to what we have done in the other regression tables.

4.0 CONCLUSIONS

Regression is based on correlation discussed in the last unit. However, unlike correlation, it allows a more sophisticated exploration of the interrelationship among a set of variables. This makes ideal for the investigation of more complex real-life, rather than laboratory based research questions.

5.0 SUMMARY

In this unit we have discussed differences between correlation and regression. We also discussed different regression models and learn how to write regression equation. We discussed the assumptions of regression and also learnt how to interpret computer output for a regression analysis.

6.0 TUTOR-MARKED ASSIGNMENT

The following are marks obtained out of a maximum of ten marks in each subject by 7 students:

Mathematics (X)	4	8	6	5	3	2	5
Physics (Y)	3	9	8	7	2	1	6

- Draw the scatter diagram of the data
- Find the slope of a freehand line that passes through the data
- Fit a regression line by method of least squares and compare the slope of this line with that obtained in (b).

7.0 REFERENCES/FURTHER READINGS

Cohen, J. W. (1988) Statistical Power Analysis for the Behavioural Sciences (2nd edition), Hillsdale, NJ: Lawrence Erlbaum Associates

Stevens, J. (1996), Applied Multivariate Statistics for Social Sciences, (3rd edition), Matiway, New Jersey: Lawrence Erlbaum Associates

Tabachnick, B. G. & Fidell, L. S. (1996), Using Multivariate Statistics (3rd edition), New York: HarpeCollins

<http://www.gower.k12.il.us/Staff/ASSESS>

www.apsu.edu/oconnort/3760/3760lect03a.htm

faculty.vassar.edu/lowry/webtext.html

www.uwsp.edu/psych/stat/14/nonparm.htm

writing.colostate.edu/guides/research/stats/pop2a.cfm

www.shodor.org/interactivate/lessons/IntroStatistics

[**www.statbasics.co.uk**](http://www.statbasics.co.uk)

UNIT 3: TREND ANALYSIS

CONTENTS

- 1.0 Introduction
- 2.0 Objectives
- 3.0 Main Content
 - 3.1 Understanding Trend Analysis
 - 3.1.1 Definition of Trend
 - 3.1.2 Methods of finding the Trend
 - 3.1.3 Computation of Example of Trend Analysis
- 4.0 Conclusions
- 5.0 Summary
- 6.0 Tutor-Marked Assignment
- 7.0 References/Further Readings

1.0 INTRODUCTION

Sometimes we need to collect data over a time period. This is known as time series. Time series is important as it is used in projecting for the future. There are statistical techniques developed to achieve this. Most time series are broken into different components one of which is **the trend**. In this unit we will study trend analysis.

2.0 OBJECTIVES

At the end of this unit you should be able to:

- Explain trend analysis
- List procedures for trend analysis
- Calculate an example of trend analysis

3.0 MAIN CONTENT

3.1 Understanding Trend Analysis

3.1.1 Defining Trend Analysis

The trend is the general path which the data have followed over a long period (Adamu, 1997). It is a technique for studying functional relationship between variables. For instance a plot of the annual number of students in secondary schools in Nigeria for the past 20 years will show that the trend

has been an upward one even though there are variations from one year to another.

3.1.2 Methods of finding the Trend

The two methods usually used in trend analysis are the calculation of regression line and the moving averages.

1 Method of Regression line

In the previous unit we discussed regression equation. We will here go through how to find the equation of the regression line for a set of observations. When a scatter plot is drawn it is not always possible to have all the points on a single straight line. What is expected of you is to try and get the best line through some of the points. This is achieved by drawing the line such that the sum of the squared deviations of the original values from the line is as small as possible. This line is known as the regression line and its slope is the regression coefficient. Let us now consider two methods of getting regression line:

- **Graph** – From the graph the intercept (c) and the slope (m) are read.
- **Calculation** – In this case we use the equation of the line that we wish to find. This is:

$$y = mx + c$$

2 Moving Averages

Moving averages of a series are the successive averages for a given size along the series.

3.1.3 Computation of Example of Trend Analysis

Let us consider examples of trend analysis.

Example 1: Find the trend for the data on the number of students offered scholarship in a particular state for the period shown.

Year	1999	2000	2001	2002	2003	2004	2005	2006	2007	2008
Scholarship	31	33	34	36	39	37	38	40	41	43

To find the trend we require the equation: $s = mY + c$

Where the variable 'Y' is the year and 's' is number of scholarships.

Using moving averages – the steps involve averages for series less than that which the trend is to be found. If in the example above we take three years we will sum the total of the scholarships for those three years.

Step 1: Find the 3-year moving totals
For 1999, 2000 and 2001
= 31 + 33 + 34
= 98

Similarly we sum for the next three years 2000, 2001 and 2002 until we get to the last one which will be 2006, 2007 and 2008.

Step 2: Find the 3-year moving average
The average of the 3-year moving totals is obtained by dividing the totals by 3. This gives the trend.

4.0 CONCLUSIONS

The method of moving averages in finding the trend is the main elementary method. It is used mainly because it is simple.

5.0 SUMMARY

In this unit you have learnt the meaning of trend analysis and its relationship with regression model. You have also learnt of two methods of computing trend.

6.0 TUTOR-MARKED ASSIGNMENT

- 1) Explain the terms trend and moving average.
- 2) Using the method of moving averages, calculate the trend of this series:

Year	1980	1981	1982	1983	1984	1985	1986	1987	1988
% Pass	40	55	53	56	60	50	48	62	65

7.0 REFERENCES/FURTHER READINGS

Cohen, J. W. (1988) Statistical Power Analysis for the Behavioural Sciences (2nd edition), Hillsdale, NJ: Lawrence Erlbaum Associates

Stevens, J. (1996), Applied Multivariate Statistics for Social Sciences, (3rd edition), Matiway, New Jersey: Lawrence Erlbaum Associates

Tabachnick, B. G. & Fidell, L. S. (1996), Using Multivariate Statistics (3rd edition), New York: HarpeCollins

<http://www.gower.k12.il.us/Staff/ASSESS>

www.apsu.edu/oconnort/3760/3760lect03a.htm

faculty.vassar.edu/lowry/webtext.html

www.uwsp.edu/psych/stat/14/nonparm.htm

writing.colostate.edu/guides/research/stats/pop2a.cfm

www.shodor.org/interactivate/lessons/IntroStatistics

www.statbasics.co.uk

MODULE 5: NON-PARAMETRIC TECHNIQUES TO COMPARE GROUPS

- Unit 1: Chi-square Tests
- Unit 2: Mann-Witney Test
- Unit 3: Wilcoxon Signed Rank Test
- Unit 4: Kruskal-Wallis Test
- Unit 5: Friedman Test

UNIT 1 CHI-SQUARE TESTS

CONTENTS

- 1.0 Introduction
- 2.0 Objectives
- 3.0 Main Content
 - 3.1: The Chi-Square Test for Independence
 - 3.1.1 Test Statistic – χ^2
 - 3.1.2 Determination of Critical (or table) χ^2 Value
 - 3.1.3 Interpretation of Computer Output for Chi-square for Independence
 - 3.2 One-sample χ^2 Test or Goodness of Fit
 - 3.2.1 Degree of Freedom (df)
- 4.0 Conclusions
- 5.0 Summary
- 6.0 Tutor-Marked Assignment
- 7.0 References/Further Readings

1.0 INTRODUCTION

These are techniques that do not make assumptions about the underlying population distributions. For this reason they are also known as distribution-free tests. They do not have stringent requirements as parametric techniques and so tend to be less sensitive and powerful. Consequently these techniques may fail to detect differences between groups that actually do exist.

Circumstances to Use Non-Parametric Techniques

- When data is measured on nominal (categorical) and ordinal (ranked) scales
- When you have very small samples
- When your data do not meet the stringent assumptions of the parametric techniques.

Assumptions for non-parametric Techniques

There are some general assumptions that you need check for non-parametric techniques. These are:

- Random Samples
- Independent observations: each person or case can only be counted once. They cannot appear in more than one category or group and the data from one subject cannot influence the data from another except for repeated measures techniques.

2.0 OBJECTIVES

3.0 MAIN CONTENT

3.1: The Chi-Square Test for Independence

This is used to determine if two categorical variables are related. That is if the two variables are independent or associated. The test compares the frequency of cases found in the various categories of one variable across the different categories of another variable. For example, consider a situation to find relationship between smoking and sex. Smoking has two categories (yes/No) while sex also has two categories (male/female). Both variables are categorical.

The null hypothesis tested by the chi-square test for independence is that the proportion of people in the response variable is the same for the explanatory variable. In the example above, sex (male/female) is the explanatory variable while act of smoking (yes or no) is the response variable. The alternative hypothesis is that the proportion differ.

3.1.1 Test Statistic – χ^2

The variables for a chi-square test are first presented in a table known as contingency table. Let us represent the column total by 'c_i' and the row total by r_i. The observed differences are given in the cells. Consider a case

of 436 students, 184 males and 252 females. These students were asked if they smoke. 33 male students responded yes and 151 responses were No. 52 female students responded yes while 200 responses were 'No'. The contingency table for this observation is presented in the table below:

Sex		Smoking		
		Yes	No	Row Total (r_i)
Males		33	151	184
Females		52	200	252
Column	Total	85	351	436 = π = Grand Total
(c_j)				

This table is called the observed frequency table under the null hypothesis we expect the proportion of students who smoke to be the same for males and females. If this is so the proportion of students who smoke (yes) will be $85/436$ or 0.195 since 85 of the 436 students indicated "yes" (that they smoke). Similarly the proportion of students who do not smoke (No) will be $351/436$ or 0.815 since 351 of the 436 students indicated 'No' (that they do not smoke). We know that 184 students are males. If there is no difference between the sexes, we expect 0.195×184 or 35 of those who smoke to be males, 0.195×252 or 49 of those who smoke to be females.

Similarly, we expect 0.805×184 or 148 as male students who do not smoke and 0.805×252 or 203 as female students who do not smoke. These expected frequencies are shown on table below.

Sex		Smoking		Row Total (r_i)
		Yes	No	
Male		36	148	184
Female		49	203	252
Column	Total	85	351	436
(c_j)				

By convention the two tables (observed and expected frequencies) are put as one table. The expected frequencies are put in parenthesis to differentiate them from the observed ones as shown below:

Sex	Smoking		
	Yes	No	Row Total (π)
Male	33 (36)	151 (148)	184
Female	52 (49)	200 (203)	252
Column Total (c_j)	85	351	436

If the null hypothesis is false you will observe differences between the observed and the expected frequencies. To get an idea of the degree of the $(O - E)^2 / E$ for each cell as presented on the table below.

Sex	Smoking		
	Yes	No	Row Total
Male	0.250	0.061	0.311
Female	0.184	0.044	0.228
Total	0.434	0.105	0.539 = χ^2

$$\text{Male/yes cell} = \frac{(33 - 36)^2}{36} = \frac{9}{36} = 0.250$$

$$\text{Female/yes cell} = \frac{(52 - 49)^2}{49} = \frac{9}{49} = 0.184$$

$$\text{Male/No cell} = \frac{(151 - 148)^2}{148} = \frac{9}{148} = 0.061$$

$$\text{Female/No cell} = \frac{(200 - 203)^2}{203} = \frac{9}{203} = 0.044$$

The χ^2 (Greek letter chi) statistic is the sum of these differences and is given by the formula:

$$\chi^2 = \sum \frac{(O - E)^2}{E}$$

For our example $\chi^2 = 0.539$ and is called calculated χ^2

Small values of χ^2 favour the Null hypothesis (hypothesis of no difference) while large values favour the alternative hypothesis (hypothesis of difference) for decision making a critical χ^2 value is compared with the calculated χ^2 value. There is a standard χ^2 distribution table in almost all statistic books.

3.1.2 Determination of Critical (or table) χ^2 Value

To use the standard χ^2 distribution table, we need to first know the degrees of freedom (df) associated with our value of χ^2 . The general rule for a contingency table is that:

$$df = (r - 1) \times (c - 1)$$

where r = number of rows

c = number of columns

In our example, we have 2 rows and two columns so $df = (2-1) \times (2-1) = 1 \times 1 = \underline{1}$

On the standard χ^2 distribution table, the df are found on the left column. This is marked with the χ^2 values to get the critical χ^2 value at a given significance level (α).

From the standard critical χ^2 values df = 1 at 0.05 is 3.841 and at 0.01 is 6.635. Since small values of χ^2 favour null hypothesis, and the χ^2 obtained from our example is 0.5..... which is smaller than the critical value of 3.841 at 0.05 and 6.635 at 0.01 α level, we do not reject the null hypothesis. We conclude that there is no significant difference in the proportion of males and females that smoke.

However, for a 2 x 2 contingency table, there is an intimate relation between a normal distribution and the chi-square distribution with 1 df. The values of χ^2 distribution with 1 df is simply the square of the normal distribution.

For example 1.96 is the critical value for a two-tailed test using the normal distribution at 0.05 while the chi-square distribution with 1df is 3.841 which is $(1.96)^2$.

3.1.3 Interpretation of Computer Output for Chi-square for Independence

The output generated is shown below.

Sex * Ever had blood transfusion? Crosstabulation

			Ever had blood transfusion?		Total
			Yes	No	
Sex	Male	Count	92	365	457
		Expected Count	105.8	351.2	457.0
		% within Sex	20.1%	79.9%	100.0%
		% within Ever had blood transfusion?	43.8%	52.4%	50.4%
		% of Total	10.1%	40.2%	50.4%
	Female	Count	118	332	450
		Expected Count	104.2	345.8	450.0
		% within Sex	26.2%	73.8%	100.0%
		% within Ever had blood transfusion?	56.2%	47.6%	49.6%
		% of Total	13.0%	36.6%	49.6%
Total	Count	210	697	907	
	Expected Count	210.0	697.0	907.0	
	% within Sex	23.2%	76.8%	100.0%	
	% within Ever had blood transfusion?	100.0%	100.0%	100.0%	
	% of Total	23.2%	76.8%	100.0%	

Chi-Square Tests

	Value	df	Asymp. Sig. (2-sided)	Exact Sig. (2-sided)	Exact Sig. (1-sided)
Pearson Chi-Square	4.728 ^b	1	.030		
Continuity Correction ^a	4.392	1	.036		
Likelihood Ratio	4.736	1	.030		
Fisher's Exact Test				.033	.018
Linear-by-Linear Association	4.723	1	.030		
N of Valid Cases	907				

a. Computed only for a 2x2 table

b. 0 cells (.0%) have expected count less than 5. The minimum expected count is 104.19.

Step 1 Assumptions

The first thing you do is to ensure you have not violated the assumption that minimum expected frequency should be 5 (or that at least 80% of the cells have expected frequencies of 5 or more). To confirm this you look at the

footnote below the table labelled **Chi-Square Tests**. In this example 0 cells have expected count less than 5. This means that we have not violated the assumption.

Step 2 Chi-Square Tests

The main value of interest is the Pearson chi-square value. This you will find in the second table labelled Chi-Square Tests. For a 2 x 2 table as our example above, you use the value in the second row (Continuity Correction). This value estimates the over estimation of the chi-square. You also read the column marked '**Asymp. Sig.**' to determine if your value is significant. In our example it is 4.392 significant at .036. This means that the proportion of males that have ever had blood transfusion is significantly different to the proportion of females that had been transfused

Step 3 Summary Information

You need to find out the percentage of each sex that had ever been transfused. This information is in the table labelled Crosstabulation.

3.2 One-sample χ^2 Test or Goodness of Fit

In the previous example two groups (males and females) were used for the chi-square test. In this section we will consider the use of one sample. One-sample χ^2 is used to test whether data obtained are consistent with a specific prediction. It is also known as χ^2 for Goodness of fit.

For example: A school offers a total of seven subjects. 35 students of the school were independently asked to name their favourite subject. The choices were: Nigerian language, 0; Mathematics 1; Biology, 11; English 8; Chemistry, 7; Physics, 5; and Music, 3. Are these data consistent with students whose preferences are evenly distributed over the subjects offered in the school?

Solution

From the Null hypothesis of even distribution we expect each subject to be chosen by 5 students, making 5 the expected frequency

$$\chi^2 = \sum \frac{(O - E)^2}{E} = \frac{(0-5)^2}{5} + \frac{(1-5)^2}{5} + \frac{(11-5)^2}{5} + \frac{(8-5)^2}{5} + \frac{(7-5)^2}{5} + \frac{(5-5)^2}{5} + \frac{(3-5)^2}{5} = 18.8$$

3.2.1 Degree of Freedom (df)

For a one-sample χ^2 , df is calculated by the formula n-1, where n = number of categories. In the example above there are seven subjects

$$\therefore df = 7 - 1 = 6$$

Critical χ^2 with df = 6 is 12.592

Since the calculated χ^2 is greater than 12.592 alternative hypothesis is favoured so we reject H_0 . In conclusion, students did not prefer all subjects evenly or equally.

4.0 Conclusions

The Chi-square tests are used to determine if two categorical variables are related. That is if the two variables are independent or associated.

5.0 Summary

In this unit you have learnt the use of chi-square statistics. You have also learnt of the two main types of chi-square. Finally you learnt how to interpret computer output of a chi-square analysis.

6.0 Tutor-Marked Assignment

A Study of the relationship between height and IQ provided the data contained in the following contingency table.

	Tall	Short
IQ > \bar{X}	55	48
IQ < \bar{X}	52	42

Showing all necessary step, draw appropriate conclusion based on the proper level of significance (Owie, 2006).

7.0 References/Further Readings

Cohen, J. W. (1988) *Statistical Power Analysis for the Behavioural Sciences* (2nd edition), Hillsdale, NJ: Lawrence Erlbaum Associates

Owie, I. (2006), *Fundamentals of Statistics in Education and the Social Sciences*, Nigeria, Lagos, National Book Consortium

Stevens, J. (1996), *Applied Multivariate Statistics for Social Sciences*, (3rd edition), Matway, New Jersey: Lawrence Erlbaum Associates

Tabachnick, B. G. & Fidell, L. S. (1996), *Using Multivariate Statistics* (3rd edition), New York: HarperCollins

<http://www.gower.k12.il.us/Staff/ASSESS>

www.apsu.edu/oconnort/3760/3760lect03a.htm

faculty.vassar.edu/lowry/webtext.html

writing.colostate.edu/guides/research/stats/pop2a.cfm

www.shodor.org/interactivate/lessons/IntroStatistics

www.statbasics.co.uk

www.uwsp.edu/psych/stat/14/nonparm.htm

UNIT 2: MANN-WHITNEY TEST

CONTENTS

- 1.0 Introduction
- 2.0 Objectives
- 3.0 Main Content
 - 3.1. Getting to Understand Mann-Whitney test
 - 3.1.1 Assumptions of Mann-Whitney
 - 3.2 Procedure for Carrying Out Mann-Whitney Test
 - 3.3 Interpretation of Computer Output from Mann-Whitney U Test
- 4.0 Conclusions
- 5.0 Summary
- 6.0 Tutor-Marked Assignment
- 7.0 References/Further Readings

1.0 INTRODUCTION

In module 2 unit 3 you learnt the use of t-test to compare the means of two independent-samples. In this unit you will learn the non-parametric equivalent called Mann-Whitney u test. In this test, instead of comparing means of the two groups as in the t-test, medians are compared. In Mann-Whitney u test continuous scores are first converted into ranks across the two groups. Then test evaluates the two groups to indicate significant difference. It is used to test the null hypothesis that the two samples come from the same population (i.e. have the same median).

2.0 OBJECTIVES

At the end of this unit you should be able to:

- Explain the Mann-Whitney test
- List the assumptions of Mann-Whitney test
- Compute the Mann-Whitney U using different methods
- Interpret Computer Output of Mann-Whitney Analysis

3.0 MAIN CONTENT

- 3.1. Getting to Understand Mann-Whitney test
 - 3.1.1 Assumptions of Mann-Whitney

The only assumptions of the Mann-Whitney test are

1. that the two samples are randomly and independently drawn
2. that the dependent variable is continuous and
3. that the measures within the two samples have the properties of at least an ordinal scale of measurement, so that it is meaningful to speak of "greater than," "less than," and "equal to."

3.2 Procedure for Carrying Out Mann-Whitney Test

- i. Arrange all the observations in order of magnitude (from the smallest to the largest).
- ii. Write an indication to show the sample each observation come from.
- iii. Calculate the statistic u

There are two ways of calculating Mann-Whitney u :

1. Direct method – This is used when the samples are small.

Considering one of the samples, write down the number of measures from the other sample which is to the left of it (i.e. smaller than it). This is known as u .

- Mann-Whitney u is the summation of all the U s i.e. $U = u_1 + u_2 + u_3 + u_4 + \dots + u_n$.

U is the smaller of the sum of U s for the 2 samples i.e. $U = \min(u_1, u_2)$. To find the probability of observing a value of u or lower. For a one-sided test u is the p-value; for two-sided test, double the probability to obtain the p-value.

Example

The following are Mathematics scores from eight students from two samples A and B.

Sample A	Sample B
25	29

26	28
27	35
31	32

Calculate the Mann-Whitney u test statistic. Is the null hypothesis rejected or retained?

Solution

Step 1: Arrange all the observations in order of magnitude

25, 26, 27, 28, 29, 31, 35

Step 2: Write an indication to show the sample each observation come from:

25	26	27	28	29	31	32	35
A	A	A	B	B	A	B	B

Step 3: Calculate u – considering each sample write down the number of means from the other sample which is to the left or smaller than it.

$$\begin{aligned}
 U_A &= u_1 = u_2 + u_3 + u_4 \\
 &= 0 + 0 + 0 + 2 \\
 &= 2
 \end{aligned}$$

$$\begin{aligned}
 U_B &= u_1 = u_2 + u_3 + u_4 \\
 &= 3 + 3 + 4 + 4 \\
 &= 14
 \end{aligned}$$

Step 4: Mann-Whitney $U = \min (U_A, U_B) = 2$

Step 5: Get p-value from table

To be filled later i.e. use of table

Practice

The order of performance of some students in two schools (T and H) are:

School T	School H
58	42
33	40
43	60
50	34
56	37
48	35

What is the U value?

Is there a significant difference in the two observations?

2 - Indirect Methods or Use of Formula to Calculate U

Mann-Whitney u can also be determined through the formula:

$$U = n_1 n_2 + \frac{n_1(n_1 + 1)}{2} - R_1$$

where n_1 = number of subjects in sample 1

n_2 = number of subjects in sample 2

R_1 = Rank sum for sample 1

Steps:

- (i) Order all observations according to magnitude. Let the smallest observation have rank 1. When there are ties, average the ranks, assign the value to all tied observations and continue with the next possible rank e.g. if 3 observations which would have occupied ranks 3,4,5 are tied, assign the average which is $\frac{3+4+5}{3} = \frac{12}{3} = 4$ to all the three observations. The next observation will receive the rank of 6.

Step (ii): Sum of ranks

Step (iii): Substitute into the formulae.

Example:

Two Educational Training Programmes A and B were conducted by NERSC. Four groups of 50 prospective teachers were trained within each training programme. Two months after the completion of training, the number of teachers remaining of the staff NERDC was observed. The result is as follows:

Programme A	Programme B
28	33
31	29
27	35
25	30

Do these data present sufficient evidence to indicate a difference in the population distribution for A and B?

Step 1: Ranking the observations. The same table is used and the ranks placed in parenthesis.

Programme A	Programme B
Rank Observation	Rank Observation
(3) 28	33 (7)
(6) 31	29 (4)
(2) 27	35 (8)
(1) 25	30 (5)

Step 2: Sum of Ranks (12)

Step 3: Substituted into the equations

$$\begin{aligned}
 \text{(i)} \quad U_A &= n_a n_b + \frac{n_a(n_a + 1)}{2} - R_A \\
 &= 4 \times 4 + \frac{4(4 - 1)}{2} - 12 \\
 &= 16 + \frac{4 \times 3}{2} - 12 \\
 &= 16 + \frac{12}{2} - 12 \\
 &= 16 + 6 - 12 \\
 &= 26 - 12 \\
 U_A &= 14
 \end{aligned}$$

U_B can be determined in two ways either by the formula:

$$\begin{aligned}U_B &= n_A n_B + \frac{n_B(n_B + 1)}{2} - R_B \\&= 4 \times 4 + \frac{4(4+1)}{2} - 24 \\&= 16 + \frac{4 \times 5}{2} - 24 \\&= 16 + \frac{20}{2} - 24 \\&= 16 + 10 - 24 \\&= 26 - 24\end{aligned}$$

$$U_B = \underline{2}$$

Or

By the formula

$$\begin{aligned}U_B &= n_A n_B - U_A \\&= 4 \times 4 - 14 \\&= 16 - 14 \\U_B &= \underline{2}\end{aligned}$$

From the table:

For $n_1 = 4$ and $n_2 = 4$

$P(u \leq 2) = 0.0571$

Practice

Test the hypothesis of no difference in the distribution below:

Sample A	Sample B
125	89
116	101

133	97
115	95
123	94
120	102
132	98
128	106
121	98

3.3 Interpretation of Computer Output from Mann-Whitney U Test

The output from Mann-Whitney U-test is a simple one as shown on the figure below:

Mann-Whitney Test

Ranks

	Sex	N	Mean Rank	Sum of Ranks
Post-test Integrated Science Achievement Score	Males	24	24.94	598.50
	Females	24	24.06	577.50
	Total	48		

Test Statistics^a

	Post-test Integrated Science Achievement Score
Mann-Whitney U	277.500
Wilcoxon W	577.500
Z	-.217
Asymp. Sig. (2-tailed)	.828

a. Grouping Variable: Sex

The first table titled 'Ranks' summarizes the ranks of the scores. The second table titled 'Test Statistics' is the most important for interpretation of results. Two values are important in interpreting the u statistic. You need to look for the Z value and the significance level which is given as 'Asymp. Sig. (2-tailed)'. If your sample size is larger than 30 the value of Z approximation test is given. If $p \leq 0.5$ there is a significant difference but If $p > 0.05$ there is no significant difference.

UNIT 3: WILCOXON SIGNED RANK TEST

CONTENTS

- 1.0 Introduction
- 2.0 Objectives
- 3.0 Main Content
 - 3.1 Understanding Wilcoxon Rank Test
 - 3.1.1 When to Use Wilcoxon Rank Test
 - 3.1.2 Assumptions of Wilcoxon Rank Test
 - 3.2 Interpretation of Wilcoxon Statistic 'W'
 - 3.2.1 Step by Step Calculation of W
 - 3.2.2 Calculation of Z-ratio
 - 3.2.3 Use of Critical Table of W
 - 3.2.4 Interpretation of Computer Output from Wilcoxon Signed Rank Test
- 4.0 Conclusions
- 5.0 Summary
- 6.0 Tutor-Marked Assignment
- 7.0 References/Further Readings

1.0 INTRODUCTION

In module 2 unit 1 you learnt tests for comparing differences between two means when you have only are group and data collected from them on two different occasions. This test we called '**Paired-Sample t-test**'. In this unit you will learn an alternative test you can use if the assumptions and conditions for using paired-sample t-test are not met. This is the non-parametric alternative and is known as '**Wilcoxon Signed Rank Test**' also referred to as the '**Wilcoxon Matched Pair Signed Rank Test**'.

2.0 OBJECTIVES

At the end of this unit you should be able to:

- Explain the assumptions of Wilcoxon Rank test
- List the step by step procedure for calculating Wilcoxon Statistic
- Calculate the Wilcoxon test statistic
- Use critical table of W to interpret Wilcoxon test statistics
- Calculate Z-ratio
- Interpret computer output for Wilcoxon statistic

3.0 MAIN CONTENT

3.1 Understanding Wilcoxon Signed Rank Test

Like the paired-sample t-test, Wilcoxon signed rank test is designed for use with repeated measures. That is when data or measures are collected from a group data or measures are collected from a group on two different occasions or under two different conditions. The main difference between paired-sample t-test and Wilcoxon signed rank test is that instead of comparing means (as with t-test) ranks for the two set of measures are compared.

For an example, let us consider 16 students ratings of two questions on probability in an introductory statistics course. The students framed each answer in terms of a zero to 100 percent rating scale, with 0% corresponding to $P=0.0$, 27% corresponding to $P=.27$, and so forth. Let us designate the student's responses as question A and question B. The following table shows the probability ratings of the 16 subjects for each of the two questions.

Subject	Question A (X_A)	Question B (X_B)
1	78	78
2	24	24
3	64	62
4	45	48
5	64	68
6	52	56
7	30	25
8	50	44
9	64	56
10	50	40
11	78	68
12	22	36
13	84	68
14	40	20
15	90	58
16	72	32

Assuming you are interested in a hypothesis that the probability ratings do on average end up higher for question A than for question B, you will agree

from the table that the observed results are consistent with the hypothesis. Wilcoxon statistics will enable us to determine whether the degree of the observed difference (differences in students rating of question A and B) reflects anything more than some lucky guessing.

3.1.1 When to Use Wilcoxon Signed Rank Test

To use Wilcoxon Sign Test you need:

- **One group** of subjects measured on the same **continuous scale** on two different occasions (X_A and X_B).
- Assumption for Wilcoxon test must not be violated.

3.1.2 Assumptions of Wilcoxon Signed Rank Test

The assumptions of the Wilcoxon test are that the :

- paired values of X_A and X_B are randomly and independently drawn (i.e., each pair is drawn independently of all other pairs);
- dependent variable (e.g., a subject's probability estimate) is intrinsically continuous, capable in principle, if not in practice, of producing measures carried out to the n^{th} decimal place; and
- measures of X_A and X_B have the properties of at least an ordinal scale of measurement, so that it is meaningful to speak of "greater than," "less than," and "equal to."

3.2 Interpretation of Wilcoxon Statistic ‘W’

In this section we will consider how to manually calculate and interpret the **W** statistic. We will also interpret computer output of Wilcoxon analysis.

3.2.1 Step by Step Calculation of W

Step 1: Determine the difference between $X_A - X_B$.

Subj.	X_A	X_B	$X_A - X_B$
1	78	78	0
2	24	24	0
3	64	62	+2
4	45	48	-3
5	64	68	-4
6	52	56	-4
7	30	25	+5

8	50	44	+6
9	64	56	+8
10	50	40	+10
11	78	68	+10
12	22	36	—14
13	84	68	+16
14	40	20	+20
15	90	58	+32
16	72	32	+40
mean difference = +7.75			

Table above shows the observed differences and indicated the sign of the results.

Step 2: Absolute Value Transformation

The Wilcoxon test begins by transforming each instance of $X_A - X_B$ into its absolute value. This is accomplished simply by removing all the positive and negative signs from the observed differences between question A and B. Thus the entries in column 4 of the table below become those of column 5 (without signs).

Step 3: Ranking of Absolute Differences

Remove cases in which there is zero difference between X_A and X_B (at this point since they provide no useful information) and rank the remaining absolute differences from lowest to highest, with tied ranks included where appropriate. The result of this step is shown in column 6.

Step 4: Create Signed Ranks

Attach to each rank the positive or negative sign that was removed from the $X_A - X_B$ difference. This is shown in column 7 and gives clue to why the Wilcoxon procedure is known as the signed-rank test.

1	2	3	4	5	6	7
Subj.	X_A	X_B	original $X_A - X_B$	absolute $X_A - X_B$	rank of absolute $X_A - X_B$	signed rank
1	78	78	0	0	---	---
2	24	24	0	0	---	---
3	64	62	+2	2	1	+1
4	45	48	-3	3	2	-2
5	64	68	-4	4	3.5	-3.5
6	52	56	-4	4	3.5	-3.5
7	30	25	+5	5	5	+5
8	50	44	+6	6	6	+6
9	64	56	+8	8	7	+7
10	50	40	+10	10	8.5	+8.5
11	78	68	+10	10	8.5	+8.5
12	22	36	-14	14	10	-10
13	84	68	+16	16	11	+11
14	40	20	+20	20	12	+12
15	90	58	+32	32	13	+13
16	72	32	+40	40	14	+14
W			=		67.0	
N = 14						

Step 5: Calculation of W Statistic

The sum of the signed ranks (column 7) is a quantity symbolized as **W**. This is equal to 67 in our example above. The sample size (N) for our observed value of **W** is 14. This is because two of the original 16 subjects were removed from consideration because of the zero difference they produced.

3.2.2 Calculation of Z-ratio

The sum of the N unsigned ranks (column 6) is equal to:

$$\text{sum} = N(N+1)$$

$$\frac{14(14+1)}{2} = 105$$

Thus the maximum possible positive value of \mathbf{W} (in the case where all signs are positive) is $\mathbf{W}=+105$, and the maximum possible negative value (in the case where all signs are negative) is $\mathbf{W}=-105$. For the present example, a preponderance of positive signs among the signed ranks would suggest that subjects tend to rate the probability higher for question A than for question B. A preponderance of negative signs would suggest the opposite. The null hypothesis is that there is no tendency in either direction, hence that the numbers of positive and negative signs will be approximately equal. In that event, we would expect the value of \mathbf{W} to approximate zero, within the limits of random variability.

To interpret \mathbf{W} you need to understand the sampling distribution of \mathbf{W} . As the size of N increases, the sampling distribution of \mathbf{W} comes closer and closer to the outlines of the normal distribution. With a sample of size $N=10$ or greater, the approximation is close enough to allow for the calculation of a z -ratio, which can then be referred to the unit normal distribution. However, when N is smaller than 10, the observed value of \mathbf{W} must be referred to an exact sampling distribution as shown in a table of critical values of \mathbf{W} for small sample sizes which are available in any good statistics text book

The null hypothesis would expect the value of \mathbf{W} to approximate zero, within the limits of random variability. This means that any particular observed value of \mathbf{W} belongs to a sampling distribution whose mean is equal to zero. Hence

$$\mu_{\mathbf{W}} = 0$$

For any particular value of N , it can be shown that the standard deviation of the sampling distribution of \mathbf{W} is equal to:

$$\sigma_{\mathbf{W}} = \text{sqrt} \left[\frac{N(N+1)(2N+1)}{6} \right]$$

For our example, with $N=14$, this works out as:

$$\sigma_{\mathbf{W}} = \text{sqrt} [14(14+1)(28+1)] = \pm 31.86$$

Thus, the structure of the z -ratio for the Wilcoxon test is:

$$z = \frac{(W - \mu_W) \pm .5}{\sigma_W}$$

Where a " $\pm .5$ " is a correction for continuity

The correction for continuity is " $-.5$ " when W is greater than μ_W and " $+.5$ " when W is less than μ_W . Since μ_W is in all instances equal to zero, the simpler computational formula is:

$$z = \frac{W - .5}{\sigma_W}$$

For the present example, with $N=14$, $W=67$, and $\sigma_W=\pm 31.86$, the result is

$$z = \frac{67 - .5}{31.86} = +2.09$$

3.2.3 Use of Critical Table of W

From the following table of critical values of z , you can see that the observed value of $z=+2.09$ is significant just a shade beyond the .025 level for a directional test, which is the form of test called for by our investigator's directional hypothesis. For a two-tailed non-directional test, it would be significant just beyond the .05 level. When N is smaller than 10, the observed value of W must be referred to an exact sampling distribution.

3.2.4 Interpretation of Computer Output from Wilcoxon Signed Rank Test

The output generated for Wilcoxon signed rank test is just a simple table captioned test statistics.

Test Statistics^b

	Test at Time 2 - Test at Time 1
Z	-4.596 ^a
Asymp. Sig. (2-tailed)	.000

a. Based on negative ranks.

b. Wilcoxon Signed Ranks Test

You should be interested in just two things in the output: the z value and the associated significant levels. As usual significant level = or less than 0.05 indicates significant difference and a conclusion that the set of scores or data are significantly different.

4.0 CONCLUSIONS

Wilcoxon test helps to correct some of the problems of sign test. You must therefore ensure that the assumptions are all met so that the sign test may not be rendered more useful under certain conditions.

5.0 SUMMARY

In this unit you have learnt of the assumptions of Wilcoxon tes. You also learnt how to compute the Wilcoxon 'W' test statistic and Z-ratio. Finally, you learnt how to interpret computer output on Wilcoxon statistics.

6.0 TUTOR-MARKED ASSIGNMENT

Using Wilcoxon statistics draw the appropriate conclusions on the following set of data:

Subjects	Assessment 1	Assessment 2
1	60	75
2	56	73
3	50	66
4	42	49
5	40	38
6	36	34
7	32	33
8	30	30
9	30	27
10	25	27

7.0 REFERENCES/FURTHER READINGS

Cohen, J. W. (1988) *Statistical Power Analysis for the Behavioural Sciences* (2nd edition), Hillsdale, NJ: Lawrence Erlbaum Associates

Stevens, J. (1996), *Applied Multivariate Statistics for Social Sciences*, (3rd edition), Matiway, New Jersey: Lawrence Erlbaum Associates

Tabachnick, B. G. & Fidell, L. S. (1996), *Using Multivariate Statistics* (3rd edition), New York: HarpeCollins

<http://www.gower.k12.il.us/Staff/ASSESS>

www.apsu.edu/oconnort/3760/3760lect03a.htm

faculty.vassar.edu/lowry/webtext.html

www.uwsp.edu/psych/stat/14/nonparm.htm

writing.colostate.edu/guides/research/stats/pop2a.cfm

www.shodor.org/interactivate/lessons/IntroStatistics

www.statbasics.co.uk

UNIT 4: KRUSKAL-WALLIS TEST

CONTENTS

- 1.0 Introduction
- 2.0 Objectives
- 3.0 Main Content
 - 3.1 Understanding Kruskal-Wallis
 - 3.1.1 Assumptions of Kruskal-Wallis
 - 3.2 Interpretation of Kruskal-Wallis 'K' Statistic
 - 3.2.1 Steps in Calculating 'k'
 - 3.2.2 General Formula
 - 3.3 Use of Table of the Null Distribution 'K'
 - 3.3.1 Chi-Square Approximation to K
 - 3.3.2 The Treatment of Ties
 - 3.3.3 Multiple Comparisons When K is Significant
 - 3.3.4 Interpretation of Computer Output from Kruskal-Wallis Test
- 4.0 Conclusions
- 5.0 Summary
- 6.0 Tutor-Marked Assignment
- 7.0 References/Further Readings

1.0 INTRODUCTION

In module 2 you learnt that when you wish to compare means of more than 2 groups e.g. 3 or more groups the correct statistics is ANOVA. When you have one independent variable (with more than 2 levels) and one dependent variable the type of ANOVA is one-way ANOVA. In this unit you will learn the non-parametric alternative of one-way between groups ANOVA. This is known as Kruskal-Wallis Test.

2.0 OBJECTIVES

At the end of this unit you should be able to:

- Explain the principles of Kruskal-Wallis test
- List the basic assumptions for the test
- Calculate the Kruskal-Wallis 'K' statistic
- Interpret computer output for Kruskal-Wallis analysis

3.0 MAIN CONTENT

3.1 Understanding Kruskal-Wallis

Kruskal-Wallis is used to test whether the sample scores come from the same population or from several populations that differ in location. Consider 3 psychological treatments to reduce anxiety of students. Suppose four students were treated, student by treatment 1, two by treatment 2 and 1 by treatment 3 and the percentage reduction on anxiety for the four students collected as shown on table below:

Treatments

1	2	3
18.2	15.8	25.5
	14.8	

You may wish to determine which of the three treatments is more effective.

A look at the data on the table indicates that treatment 3 gave the highest percentage reduction that treatment 2 and 1. You may wish to conclude that treatment 3 is more effective in reducing anxiety than treatment and treatment 1 more effective than treatment 2. Assuming that one null hypothesis is that there is no significant difference among the three treatments, you can use Kruskal-Wallis test to establish this. The test statistics is “k”.

3.1.1 Assumptions of Kruskal-Wallis

One assumption of Kruskal-Wallis is that the data are independent and that the scores on the response variable are continuous ordinal data.

3.2 Interpretation of Kruskal-Wallis ‘K’ Statistic

3.2.1 Steps in Calculating ‘k’

Step 1: Rank the data from the lowest to the highest

Treatment		
1	2	3
3	2	4
	1	

Step 2: Compute the Rank-sum for the three treatments

	Treatment		
	1	2	3

Ranks (R)	3	2 1	4
Rank-sum (R _i)	3	3	4

Step 3: Calculation of Mean Rank

Since there are different numbers of students for each treatment calculate the mean Rank for each. The mean rank is obtained by dividing the rank-sum R_i for a given group by the number of students (n_i) in that group. This is shown in the table below:

	Treatment		
	1	2	3
Ranks (R)	3	2 1	4
Rank-sum R _i	3	3	4
Average rank R _i n _i	$\frac{3}{1} = 3$	$\frac{3}{2} = 1.5$	$\frac{4}{1} = 4$

If the null hypothesis is true, the three averages should be relatively close together. If the averages measure of dispersion for the average ranks by:

(i) Calculate the grand mean of all few ranks i.e. Rank Grand mean

$$= \frac{1 + 2 + 3 + 4}{4} = 2.5$$

(ii) Deviations from Grand mean = Average rank – Rank Group mean

Considering our example

$$= 3 - 2.5 = 0.5$$

$$= 1.5 - 2.5 = -1.0$$

$$= 4 - 2.5 = 1.5$$

Step 5: Calculation of Average Rank

Calculate single number to reflect the dispersion of the average ranks by summing the squares of the deviations; with each squared deviations weighted by sum of squares of deviations =

$$1x(0.5)^2 + 2x(-1.0)^2 + 1x(1.5)^2$$

$$= 0.25 + 2.00 + 2.25$$

$$= 4.50$$

The larger this number is the greater is the differences among the samples

3.2.2 General Formula

- (i) - mean of 'n' ranks = $(n + 1)/2$. In our example $n = 4$ students
 \therefore mean of ranks = $(4+1)/2 = \frac{5}{2} = 2.5$

This is much easier than what we did before

- (ii) Formula for measure of dispersion of the average ranks is then

$$\sum t_i \left(\frac{R_i}{t_i} - \frac{n+1}{2} \right)^2$$

where t_i = number of sample in each treatment

R_i = Rank-sum

N = total subjects/sample

In our example:

$$\begin{aligned} \sum t_i \left(\frac{R_i}{t_i} - \frac{n+1}{2} \right)^2 &= 1x \left(\frac{3}{1} - \frac{4+1}{2} \right)^2 + 2x \left(\frac{3}{2} - \frac{4+1}{2} \right)^2 + 1x \left(\frac{4}{1} - \frac{4+1}{2} \right)^2 \\ &= 1x (3 - 2.5)^2 + 2x (1.5 - 2.5)^2 + 1x (4 - 2.5)^2 \\ &= 1x (0.5)^2 + 2x (1.0)^2 + 1x (1.5)^2 \\ &= 0.25 + 2.00 + 2.25 \\ &= 4.50 \end{aligned}$$

- (iii) **Kruskal-Wallis 'K' Statistics Formula**

This is gotten by multiplying the sum of deviations by a constant

$$\begin{aligned} \therefore K &= \frac{12}{n(n+1)} x \sum t_i \left(\frac{R_i}{t_i} - \frac{n+1}{2} \right)^2 \\ &= \frac{12}{4x5} x 4.50 \\ &= \underline{\underline{2.7}} \end{aligned}$$

Note

- (i) K must always be positive

- (ii) If the H_0 is true, the average ranks will be close together and 'k' will be small
- (iii) If the H_A is true, the average ranks will differ and k will be large.

(iv) **Simplified 'K' Formula**

A simplified formula of K is:

$$K = -3(n+1) + \frac{12}{n(n+1)} \sum \frac{R_i^3}{t_i}$$

Example

	Treatment 1	Treatment 2	Treatment 3
Rank-sum R_i	3	3	4
Rank-sum square	9	9	16
$\frac{R_i^2}{t_i}$	$\frac{9}{1} = 9$	$\frac{9}{2} = 4.5$	$\frac{16}{1} = 16$

$$\sum \frac{R_i^2}{t_i} = 9 + 4.5 + 16 = 29.5$$

Using the new formula

$$K = -3(4+1) + \frac{12}{4(4+1)} \times 29.5$$

$$= -15 + 17.7 = 2.7$$

Note: Rank – sums R_i is always equal to $n(n+1)/2$

From our example $\sum R_i = 3 + 3 + 4 = 10$

$$\frac{n(n+1)}{2} = \frac{4(5)}{2} = \frac{20}{2} = 10$$

3.3 Use of Table of the Null Distribution 'K'

This table is only appropriate when 3 samples are being compared and each sample size is less than or equal to 5. The table is known as **critical values of k for the kruskal-wallis Test with 3 Independent samples**

(i) If the 3 sample sizes, $t_1 > t_2 > t_3$. That is t_1 is the largest sample and t_3 is the smallest sample. Assuming in another example 5 students underwent treatment 1, 2 treatment 2 and 4 treatment III, then you $t_1 = 5$; $t_2 = 4$ and $t_3 = 2$, and supposing our calculated k is 5.97, you look for the critical 'k' value

For $t_1 = 5$, $t_2 = 4$ and $t_3 = 2$ at 0.05 level of significant. This is 5.27 since the calculated 'k' value is higher (5.97), the H_0 is not retained but reject for the alternative since higher k values favour the alternative hypotheses.

3.3.1 Chi-Square Approximation to K

When there are more than 3 samples or when sample sizes are larger than 5 Chi-square approximation to 'k' is used. To use this table, you need to calculate only the degree of freedom which is $n-1$. In our example on 3 treatments for four students minus one (i.e. $4 - 1$). This is 3. You look at the column of the table of critical values of χ^2 and at 0.05 level of significance to get the corresponding value. This 7.815 since this is greater than our 'k' ($7.815 > 2.7$) the H_0 is rejected. Larger values favour H_A .

3.3.2 The Treatment of Ties

One of the assumptions of Kruskal-Wallis test is that the response variable is continuous. However, there are instances when there are ties among the response variable. In such a situation the assumption is violated. You are allowed to use such data however you need to calculate 'midrank'. If fewer than a quarter of the scores are involved in ties, the formula for calculating k ;

$$k = \frac{-3(n+1) + \frac{12}{n(n+1)} \sum \frac{R_i^2}{t_i}}{1 - \frac{\sum u_i(u_i-1)(u_i+1)}{n^3 - n}}$$

3.3.3 Multiple Comparisons When K is Significant

Kruskal Wallis tests significant difference among 3 or more groups. It only tells you that a difference exist and not the locus of the difference. To determine the locus of the difference, the groups/sample are compared in pairs. This is known as **multiple comparison**. For kruskal Wallis test this comparison is done using **Rank-Sum Test** for example, for a three sample (Test 1; Test 2 and Test 3) you will need to compare:

Test 1 and Test 2
Test 1 and Test 3
Test 2 and Test 3

The number of test is given by $k(k-1)/2$

Multiple comparison provide a method of searching through the data in an attempt to locate differences that might be interesting. This means carrying out 3 Rank-Sum test. Unfortunately this procedure introduces a complication. Since there are 3 tests (more than 1) the significant levels become distorted.

Remember that the significant level (α) is the probability of falsely rejecting the null hypothesis (H_0). This is also known as **type/error**. It then means that as the significant level is distorted due to many tests, so the possibility of making a type 1 error increases. Thus if you reject the H_0 at $\alpha=0.05$, you may set your α as high as 0.20 when carrying out the many multiple comparison.

Procedure for Performing Multiple Comparison 1- Choose α (the Per-experiment Error Rate)

Example:

suppose you set $\alpha = 0.05$ and you have 3 samples, you need to do $k(k-1)/2$ comparisons. That is $3(3-1)/2$ comparisons. This translates to $3 \times \frac{2}{2} = 3$ comparisons. The per – comparison error rate therefore should be $\frac{\alpha}{3} = \frac{0.05}{3} = 0.017 \cong 0.02$. The per-experiment error rate will be $3 \times 0.02 = 0.06$.

So you set your $\alpha = 0.06$ and perform each comparison using a significance level of 0.02.

2 – List the Sample in Order of Increasing Average Rank

Since you already know the rank-sums (R) of the samples, the average ranks are simply $\frac{R_i}{t_i}$. If in one 3 samples ABC sample B has the smallest average rank followed by A and C the largest, we order our samples as BAC.

3 – Rank-Sum Test

This is done by comparing the sample with the smallest average rank with the one with the largest. In the example we begin by comparing B and C first. As soon as we get a non-significant result we stop and draw a line joining the samples in the ordered list that have shown not to differ.

3.3.4 Interpretation of Computer Output from Krustal-Wallis Test

The outputs generated by the computer process are simple, tables titled “Ranks” and ‘Test statistics’.

Kruskal-Wallis Test

Ranks

	School	N	Mean Rank
Post-test Integrated Science Achievement Score	School 1 (GSS Garki)	16	30.19
	School 2 (GSS Wuse)	16	31.44
	School 3 (GSS T/Wada)	16	11.88
	Total	48	

Test Statistics^{a,b}

	Post-test Integrated Science Achievement Score
Chi-Square	19.691
df	2
Asymp. Sig.	.000

a. Kruskal Wallis Test

b. Grouping Variable: School

The piece of information you need from the output are chi-square (x^2) value, degrees of freedom (df) and significant level (Asymp. Sig). If α is less than 0.05 you conclude that there is a statistically significant difference in the continuous (response) variable across the 3 groups. You can then inspect the mean Rank for the three groups on the first table and determine the group with the highest overall ranking.

4.0 CONCLUSIONS

The Kruskal-Wallis test is non parametric, that is, it does not make any assumption on the nature of the underlying distributions (except

continuity). As many other non parametric tests, it will not use the values of the observations directly, but will first convert these values into ranks once these observations are merged into a single sample.

5.0 SUMMARY OF KRUSTAL WALLIS TEST

- You must have more than two independent samples
- Scores on the response variable must be ordinal
- Select significant level (α)
- Find n (total no of scores) and t_i (no of scores in the ith sample).
- Rank order all n scores (use mid ranks if ties occur).
- Find the sum of the ranks for each sample calculate k.
- If more than a quarter of the response scores are involve in ties use formulae for ties.
- If more than three samples are being compared or any of the sample size is larger than 5, use χ^2 approximation.
- Find the relevant critical value.
- Reject the H_0 if calculated k is larger than or equal to the critical k value.
- If H_0 is reject (i.e. there is significant difference use multiple comparison to locate locus of difference.

6.0 TUTOR-MARKED ASSIGNMENT

The number of problems solved by each of 4 people is shown below:

Subjects	Type of Problem		
	A	B	C
1	20	30	40
2	24	26	25
3	29	31	53
4	42	48	49

Are there any differences between the three types of problems A, B, C?
(Leach, 1979)

7.0 REFERENCES/FURTHER READINGS

Cohen, J. W. (1988) Statistical Power Analysis for the Behavioural Sciences (2nd edition), Hillsdale, NJ: Lawrence Erlbaum Associates

Leach, C., (1979), *Introduction to Statistics A Non-Parametric Approach for the Social Sciences*, New York: John Willy

Stevens, J. (1996), *Applied Multivariate Statistics for Social Sciences*, (3rd edition), Matiway, New Jersey: Lawrence Erlbaum Associates

Tabachnick, B. G. & Fidell, L. S. (1996), *Using Multivariate Statistics* (3rd edition), New York: HarpeCollins

<http://www.gower.k12.il.us/Staff/ASSESS>

www.apsu.edu/oconnort/3760/3760lect03a.htm

faculty.vassar.edu/lowry/webtext.html

www.uwsp.edu/psych/stat/14/nonparm.htm

writing.colostate.edu/guides/research/stats/pop2a.cfm

www.shodor.org/interactivate/lessons/IntroStatistics

www.statbasics.co.uk

UNIT 5: FRIEDMAN TEST

CONTENTS

- 1.0 Introduction
- 2.0 Objectives
- 3.0 Main Content
 - 3.1 Understanding Friedman Test
 - 3.2 Friedman 'Q' Test Statistic
 - 3.2.1 Steps in Calculating 'Q' Test Statistic
 - 3.2.2 Use of Table of Null Distribution of Q
 - 3.2.3 Use of Chi-Square Appropriate to Q
 - 3.2.4 Treatment of Ties
 - 3.2.5 Calculation of Correction for Ties (T)
 - 3.2.6 Interpretation of Computer Output for Friedman Test
 - 3.3 Friedman Test as a Means of Reflecting Agreement or Disagreement between subjects
 - 3.3.1 The Cochran Test
 - 3.3.2 Calculation of Cochran Q Test
- 4.0 Conclusions
- 5.0 Summary
- 6.0 Tutor-Marked Assignment
- 7.0 References/Further Readings

1.0 INTRODUCTION

In module 2 unit 2, you studied the case of collecting data from more than two groups at different times or events. This situation we referred to as repeated measure. In this unit you will study the non-parametric alternative. This is known as Friedman Test.

2.0 OBJECTIVES

At the end of this unit you should be able to:

- Calculate Friedman 'Q' test statistic
- Use table of Null distribution to make decisions
- Calculate Chi-square approximation to Q
- Relate Friedman's test to Cochran Test
- List the used of Cochran's test
- Interpret Computer output for Friedman's test

3.0 MAIN CONTENT

3.1 Understanding Friedman Test

Friedman Test is the non-parametric alternative to the one-way repeated measures ANOVA. It is used when you take the same sample of subjects or cases and you measure them at three or more points in time or under three different conditions. In this case the response variable must be either ordinal or categorical. Friedman Test is a counterpart of the Kruskal-Wallis test which you studied in the last unit except that it deals with several related samples. Let us consider an accreditation panel of three (3) Lecturers who visited three universities and were each expected to rate each of the university on a 10-point scale of adequacy of the activities (infrastructure, resources and teaching) in the universities. In this example the number of subjects/group(n) is 3. The number of samples (k) is also 3. Because each subject is to rate all the 3 universities, it is known as repeated measure.

3.2 Friedman ‘Q’ Test Statistic

3.2.1 Steps in Calculating ‘Q’ Test Statistic

Let us assume that the result of the lecturers rating of the universities is as presented in the table below:

Universities (Samples (k))				
Lecturers	1	4	5	6
(Subject/Group)	2	3	7	9
(n)	3	3	2	5

It is possible to determine if the rating of the samples (Universities) differ in any way among the subjects (Lecturers). The test statistic for Friedman test is the ‘Q’, Q is calculated through different steps:

Step 1: Rank the rating of each lecturer separated giving the lowest rank to the lowest score (how is this different in Kruskal-Wallis test?). Answer = We rank all the subjects rating together.

This is presented in table below with the ranks

Universities (k)						
Lecturers (n)	X	Rank	Y	Rank	Z	Rank
1	4	1	5	2	6	3
2	3	1	7	2	9	3

3	3	2	2	1	5	3
---	---	---	---	---	---	---

If the H_0 is true there will be no systematic tendency for any one of the universities to be given low (or high) scores by all the lecturers.

Step 2: Calculate the Rank sum for each University

Table below shows only the ranks and rank-sum

Universities (k = 3)				
		X Rank	Y Rank	Z Rank
Lecturers (n = 3)	1	1	2	3
	2	1	2	3
	3	2	1	3
Rank-Sum	R_1	4	5	9

What is the difference with Kruskal-Wallis (Answer - we work with total net average).

Step 3: Calculate the average total Rank score across the subjects

The average total rank score (\bar{R}) is given by:

$$\begin{aligned}
 \bar{R} &= \frac{n(k+1)}{2} \\
 &= \frac{3(3+1)}{2} \\
 &= \frac{3 \times 4}{2} \\
 &= \frac{12}{2}
 \end{aligned}$$

$$\bar{R} = 6$$

Note: If the H_0 is true, none of the 3 rank totals should be very different from this average value. (Question: what is the situation with our example? Answer: There is no rank total for any of the universities that is same as the \bar{R}).

Step 4: Calculate the square of the difference between the Rank-sum (R_i) and the average total Rank score (\bar{R}). The difference is given by $R_i - \bar{R}$. This is illustrated for our example in table below.

Universities (k = 3)				
		X Rank	Y Rank	Z Rank
Lecturers (n = 3)	1	1	2	3
	2	1	2	3
	3	2	1	3
Rank-sum (R_i)	4	5	9	$\bar{R} = 6$
$R_i - \bar{R}$	4 - 6	5 - 6	9 - 6	
Difference between Rank sum and average	= - 2	= - 1	= 3	

Step 5: Sum the square of the difference between Rank – sum and average Rank total

$$\text{i.e. } \sum (R_i - \bar{R})^2$$

$$\sum (R_i - \bar{R})^2 = 4 + 1 + 9 = 14$$

Note: We could use this summation as the test statistic for Friedman test.

Step 6: Calculate Q

When the sum of the square of the difference between the rank-sum and average rank total is multiplied by a factor $\frac{12}{nk(k+1)}$ we get the Friedman's test statistic Q.

$$\therefore Q = \frac{12}{nk(k+1)} \times \sum (R_i - \bar{R})^2$$

Considering our example

$$\begin{aligned}
Q &= \frac{12}{3 \times 3(3+1)} \times 4 \\
&= \frac{12}{9 \times 4} \times 14 \\
&= \frac{1}{3} \times 14 \\
Q &= 4.67
\end{aligned}$$

A simpler but equivalent formula is $Q = -3n(k+1) + \frac{12\sum R_i^2}{nk(k+1)}$

Note: The values of Q close to zero favour H_0 while large values favour H_A . At 0.05 α level, you need a Q value as large as 6.00. Q is never negative before rejecting H_0 .

3.2.2 Use of Table of Null Distribution of Q

In every good statistical book, there is a table that gives critical values of the null distribution of Q. This table is appropriate when there are 3 samples ($k = 3$) with sample size 'n' between 2 and 13 or when $k = 4$ with n between 2 and 8. As in our example ($n = 3$; $k = 3$), the critical Q value for $n = 3$ at 0.05 α level is 6.00. Our calculated value of 4.67 is not as large or larger than 6.00 we conclude that there is no reliable evidence of a difference between the universities. From the table 4.67 is at α level of 0.2. We can only reject our H_0 if we are using α level of 0.2.

3.2.3 Use of Chi-Square Appropriate to Q

When n and k not specified in the table of null distribution of Q is encountered the table of chi-square approximation to Q is used. This table is also found in most good statistics books.

To use this table you require to know the degree of freedom (df) for your sample. This is given as $k - 1$ (i.e. the number of samples minus one). For our example, the df is $3 - 1 = 2$. On the table to the left hand column is the df and at the top across the table (top row) is the significant level. For df = 2 at α 0,05, the critical χ^2 value is 5.991. If our calculated Q (4.67) is as large or larger than 5.991 we reject the H_0 .

3.2.4 Treatment of Ties

In calculating Friedman's Q statistics, you assume that there is no tie in each subject rating. However, in practice this may occur. In such a situation, **MIDRANK** technique is used. That means the tied scores get the average rank of the ties. When the ties are few, the use of the Q formulae, Null distribution of Q and χ^2 approximation of Q will give a reasonable accurate result. However when the ties are much, the calculation of Q is modified and the χ^2 approximation of Q used.

The formulae (which has correction factor for ties) is given as:

$$Q = \frac{0.3n(k+1) + \frac{12\sum R_1^2}{nk(k+1)}}{1 - \frac{T}{n(k^3 - k)}}$$

where \therefore n = number of subjects
k = number of samples
 R_1 = rank-sum
T = correction for ties

3.2.5 Calculation of Correction for Ties (T)

$$T = \sum (t_i^3 - t_i)$$

Consider three (3) Lecturers rating of three universities with ties as presented in the table below:

University			
Lecturers	X	Y	Z
1	4	5	6
2	3	7	7
3	3	3	3

There are ties in the rating of lecturers 2 and 3. Using the **midrank** the ranking is shown on table below.

University (Ranks)			
Lecturers	X	Y	Z
1	1	2	3
2	1	2.5	2.5
3	2	2	2

For lecture 2 there is a tie for the rating of universities Y and Z. These would have taken the ranks 2 and 3. The midrank is therefore $\frac{2+3}{2} = \frac{5}{2} = 2.5$. For lecturer 3, all the ratings for the 3 universities tied. These would have taking the ranks 1, 2 and 3. The midrank therefore is $\frac{1+2+3}{3} = \frac{6}{3} = 2$

For lecturer 1, there are 2 mid Ranks while for lecturer 2, there is only one midrank

∴

Lecturers	Midrank	1	2.5
1	No of scores at the value t_i	1	2
2	Midrank		2
	No of scores at the value t_i		3

$\sum (t_1^3 - t_1)$ for the 2 lecturers

$$\begin{aligned} \text{Lecturer 1} &= (1^3 - 1) + (2^3 - 2) \\ &= (1 - 1) + (8 - 2) \\ &= 0 + 6 = 6 \end{aligned}$$

$$\begin{aligned} \text{Lecturer 2} &= (3^3 - 3) \\ &= 27 - 3 = 24 \end{aligned}$$

$$\begin{aligned} T &= \sum (t_1^3 - t_1) \text{ for lecturer 2} + \sum (t_1^3 - t_1) \text{ lecturer 3} \\ &= 6 + 24 \\ &= 30 \end{aligned}$$

3.2.6 Interpretation of Computer Output for Friedman Test

The output generated by SPSS for Friedman's test is shown below. There are two major tables captioned 'Ranks' and 'Test Statistics'.

Friedman Test

Ranks

	Mean Rank
Post-test Score for Section A	2.07
Post-test Score for Section B	1.94
Post-test Score for Section C	1.99

The 'Rank' table as the name implied gives the ranks of the different related measurements.

Test Statistics^a

N	48
Chi-Square	.489
df	2
Asymp. Sig.	.783

a. Friedman Test

From the table of 'Test of Statistics' you need to consider the row named '**Asymp. Sig.**' Values equal to or less than 0.05 suggest significant difference in the measures. Values larger than 0.05 suggest no significant difference as in the result above (sig. level = .783).

3.3 Friedman Test as a Means of Reflecting Agreement or Disagreement between subjects

This is known as Kendall Coefficient of concordance – 'W'

The index of agreement $w = \frac{Q}{n(k-1)}$

Where n = number of subjects
K = number of sample
Q = Friedman statistic

3.3.1 The Cochran Test

Sometimes when extensive ties result in the ratings by subjects to 2 categories, the Friedman test turns into a test devised by Cochran. This is known as Cochran Q Test.

Assignment

What is the simpler Cochran Q Test formula?

In a two-sample case, Cochran Test reduces to McNemar Test

3.3.2 Calculation of Cochran Q Test

Let us consider a situation where the subjects were asked to agree or disagree with some programmes as presented on table below:

Programme				
Student	W	X	Y	Z
1	A	A	A	A
2	A	A	D	D
3	D	A	D	D
4	A	A	A	D
5	D	D	D	A

Where A = Agree
D = Disagree

To calculate Cochran Q you need to; **step 1**, calculate the total number of Agree for each programme (B) and square it (B^2). **Step 2**: Calculate total number of Agree by each student/subject (L) and square it (L^2). The result is presented in table below:

Programme						
Student	W	X	Y	Z	L_i	L_i^2
1	A	A	A	A	4	16
2	A	A	D	D	2	4
3	D	A	D	D	1	1
4	A	A	A	D	3	9
5	D	D	D	A	1	1
B_i	3	4	2	2	11=G	31
B_i^2	9	16	4	4	33	

Step 3:

Calculate the grand total of Agree (G). This is the sum of L which must be equal to the sum of B_i .

Step 4: Using the formula

$$\text{Cochran Q} = \frac{(k-1)\{k\sum B_1^2 - G^2\}}{KG - \sum L_1^2}$$

where k = number of sample

L = total number of Agree for each programme

B = total number of Agree for each subject

G = Grand total of L or B

From our example:

$$\begin{aligned} Q &= \frac{(4-1)\{4 \times 33 - 11^2\}}{4 \times 11 - 31} \\ &= \frac{3 \times \{132 - 121\}}{44 - 31} \\ &= \frac{33}{13} \end{aligned}$$

$$Q = \underline{2.54}$$

Note: You will obtain the same result even if you decide to use the disagree response in place of agree for the calculation.

4.0 CONCLUSIONS

Friedman's test is the counterpart of the Kruskal-Wallis test when there are several related samples. When only two samples are being compared it is like the sign test.

5.0 SUMMARY

In this unit you have learnt about the Friedman's test. You also learnt how to calculate Friedman's statistic. The relationship between Friedman and Cochran test was discussed. Finally you learnt how to interpret computer output for Friedman's analysis

6.0 TUTOR-MARKED ASSIGNMENT

Two schools are being accredited and 3 Officials are sent to rate the two schools. The rating of the officials are presented below:

Officials	Schools	
	A	B
1	4	5
2	3	7
3	3	2

Compute and interpret all relevant information using Friedman's test.

7.0 REFERENCES/FURTHER READINGS

Cohen, J. W. (1988) Statistical Power Analysis for the Behavioural Sciences (2nd edition), Hillsdale, NJ: Lawrence Erlbaum Associates

Stevens, J. (1996), Applied Multivariate Statistics for Social Sciences, (3rd edition), Matiway, New Jersey: Lawrence Erlbaum Associates

Tabachnick, B. G. & Fidell, L. S. (1996), Using Multivariate Statistics (3rd edition), New York: HarpeCollins

<http://www.gower.k12.il.us/Staff/ASSESS>

www.apsu.edu/oconnort/3760/3760lect03a.htm

faculty.vassar.edu/lowry/webtext.html

www.uwsp.edu/psych/stat/14/nonparm.htm

writing.colostate.edu/guides/research/stats/pop2a.cfm

www.shodor.org/interactivate/lessons/IntroStatistics

www.statbasics.co.uk

MODULE 6 TEST CONSTRUCTION STATISTICS

- Unit 1: Validity and Reliability of Scores
- Unit 2: Item Analysis
- Unit 3 Scaling Techniques

UNIT 1: VALIDITY AND RELIABILITY OF SCORES

CONTENTS

- 1.0 Introduction
- 2.0 Objectives
- 3.0 Main Content
 - 3.1 The Requirement of Validity
 - 3.1.1 Validity Defined
 - 3.1.2 Relationship between Validity and Reliability
 - 3.1.3 Method to assess validity

 - 3.2 Reliability Estimation Using a Split-half Methodology
 - 3.2.1 Steps in Calculating Split-half Reliability
 - 3.2.2 Estimating Reliability using the Kuder-Richardson Formula 20
 - 3.2.3 Estimating Reliability Using the Kuder-Richardson Formula 21
 - 3.2.4 Estimating Reliability Using Cronbach's alpha (α)
 - 3.2.5: Establishing Inter-rater Agreement
- 4.0 Conclusions
- 5.0 Summary
- 6.0 Tutor-Marked Assignment
- 7.0 References/Further Readings

1.0 Introduction

In your course on research methodology, you studied the different steps in test construction. First you determine the trait, ability, emotional state, disorder, interests, or attitude that you want to assess. Then you decide how you want to measure the construct you selected. This is followed by the creation of the items and the rating of the quality of the items by experts so that only the items with the highest ratings are retained. Your test is then ready to be tested on a sample of people. After administering the test to a good cross section of the people, test statistics are used to make decision of

the final items to be included in the test material. This module will consider some of these test construction statistics.

2.0 Objectives

At the end of this unit you should be able to:

- List the various forms of validity
- Estimate reliability using split-half
- Estimate reliability using Kuder-Richardson
- Estimate reliability using inter-rater reliability

3.0 Main Content

3.1 The Requirement of Validity

3.1.1 Validity Defined

Validity refers to the degree to which evidence and theory support the interpretations of test scores entailed by proposed uses of tests.

3.1.2 Relationship between Validity and Reliability

Though a reliable test may not be valid, one cannot have a valid measure without its being reliable

3.1.3 Method to assess validity

- **Face Validity:** In this method you simply look at the content of the measurement items and advance an argument that, on its face, the measure has validity.
- **Concurrent Validity:** The method here is to correlate new measure with a previously validated measure of the construct. You will recall that in an earlier module you had learnt about regression and correlation. You need to review these methods of correlation.
- **Predictive Validity:** This is the degree to which a measure predicts known groups in which the construct must exist
- **Construct Validity:** In this method a new measure is administered to the subjects along with a least two other measures (one of these measures should be a valid measure of a construct that is known conceptually to be directly related to the new measure, and another

measure should be known conceptually to be inversely related to the construct.

3.2 Reliability Estimation Using a Split-half Methodology

To use the split-half design you need to create two comparable test administrations. The items in a test are split into two tests that are equivalent in content and difficulty. Often this is done by splitting among odd and even numbered items. This assumes that the assessment is homogenous in content. Once the test is split, reliability is estimated as the correlation of two separate tests with an adjustment for the test length.

3.2.1 Steps in Calculating Split-half Reliability

Let us consider an example of a test of 40 items for 10 students. The steps for calculating the reliability are:

Step 1: Divide the original items into even (X) and odd (Y) halves as shown in the table below (columns 3 and 4 respectively).

Step 2: Calculate the Mean of the original 40 items and the split-half even and odd items

Step 3: Calculate the mean deviation for each student by subtracting the mean from each student score for the even and odd halves (x and y or columns 5 and 6 in the table)

Step 4: Find the squared deviations (x^2 and y^2) for each student (columns 7 and 8)

Step 5: Calculate the Standard Deviations (SD) for the even and odd halves. This is computed by

- squaring the deviation (e.g., x^2) for each student,
- summing the squared deviations (e.g., $\sum x^2$);
- dividing this total by the number of students minus 1 (N-1) and
- taking the square root.

$$SD_x = \sqrt{\frac{\sum x^2}{N-1}} \qquad SD_y = \sqrt{\frac{\sum y^2}{N-1}}$$

Step 6: Calculate the product of x and y (xy in column 9)

These calculations are summarized in table below.

Student	Score (40)	X Even (20)	Y Odd (20)	x	y	x ²	y ²	xy
A	40	20	20	4.8	4.2	23.04	17.64	20.16
B	28	15	13	-0.2	-2.8	0.04	7.84	0.56
C	35	19	16	3.8	0.2	14.44	0.04	0.76
D	38	18	20	2.8	4.2	7.84	17.64	11.76
E	22	10	12	-5.2	-3.8	27.04	14.44	19.76
F	20	12	8	-3.2	-7.8	10.24	60.84	24.96
G	35	16	19	0.8	3.2	0.64	10.24	2.56
H	33	16	17	0.8	1.2	0.64	1.44	0.96
I	31	12	19	-3.2	3.2	10.24	10.24	-10.24
J	28	14	14	-1.2	-1.8	1.44	3.24	2.16
MEAN	31.0	15.2	15.8			95.60	143.60	73.40
SD		3.26	3.99					

Step 7: Calculate the Pearson Product Moment Correlation Coefficient

The formula for Pearson Product Moment Correlation Coefficient is:

$$r_{xy} = \frac{\sum xy}{(N-1) (SD_x) (SD_y)}$$

Where:

- x is each student's score minus the mean on even number items for each student.
- y is each student's score minus the mean on odd number items for each student.
- N is the number of students.

$$r_{xy} = \frac{73.4}{9(3.26) (3.99)} = .63$$

The Spearman-Brown formula is usually applied in determining reliability using split halves. When applied, it involves doubling the two halves to the full number of items, thus giving a reliability estimate for the number of items in the original test.

$$\rho = \frac{2r_{xy}}{1+r_{xy}} = \frac{2(.63)}{1.63} = .77$$

3.2.2 Estimating Reliability using the Kuder-Richardson Formula 20

Kuder and Richardson devised a procedure for estimating the reliability of a test in 1937. It has become the standard for estimating reliability for single administration of a single form. Kuder-Richardson measures inter-item consistency. It is tantamount to doing a split-half reliability on all combinations of items resulting from different splitting of the test. When data is large using the KR20 is a challenging set of calculations to do by hand. It is easily computed by a spreadsheet or basic statistical package. The rationale for Kuder and Richardson's most commonly used procedure is roughly equivalent to:

- 1) Securing the mean inter-correlation of the number of items (k) in the test,
- 2) Considering this to be the reliability coefficient for the typical item in the test,
- 3) Stepping up this average with the Spearman-Brown formula to estimate the reliability coefficient of an assessment of k items.

ITEM (k)													X	$x=X-$	x^2
	1	2	3	4	5	6	7	8	9	10	11	12			
Student (N)	1=correct						0=incorrect						(Score)	mean (score-mean)	
A	1	1	1	1	1	1	1	0	1	1	1	1	11	4.5	20.25
B	1	1	1	1	1	1	1	1	0	1	1	0	10	3.5	12.25
C	1	1	1	1	1	1	1	1	1	0	0	0	9	2.5	6.25
D	1	1	1	0	1	1	0	1	1	0	0	0	7	0.5	0.25
E	1	1	1	1	1	0	0	1	1	0	0	0	7	0.5	0.25
F	1	1	1	0	0	1	1	0	0	1	0	0	6	-0.5	0.25

G	1	1	1	1	0	0	1	0	0	0	0	0	5	-1.5	2.25
H	1	1	0	1	0	0	0	1	0	0	0	0	4	-2.5	6.25
I	1	1	1	0	1	0	0	0	0	0	0	0	4	-2.5	6.25
J	0	0	0	1	1	0	0	0	0	0	0	0	2	-4.5	20.25
Σ	9	9	8	7	7	5	5	5	4	3	2	1	65	0	74.50
													mean		Σx^2
													6.5		74.50

P-values	0.9	0.9	0.8	0.7	0.7	0.5	0.5	0.5	0.4	0.3	0.2	0.1
Q-value	0.1	0.1	0.2	0.3	0.3	0.5	0.5	0.5	0.6	0.7	0.8	0.9
pq	0.09	0.09	0.16	0.21	0.21	0.25	0.25	0.25	0.24	0.21	0.16	0.09
Σpq	2.21											

Here, Variance

$$\sigma^2 = \frac{\Sigma x^2}{N-1} = \frac{74.5}{10-1} = 8.28$$

Kuder-Richardson Formula 20

$$\rho_{KR20} = \frac{k}{k-1} \left(1 - \frac{M(k-M)}{k\sigma^2} \right)$$

$$\rho_{KR20} = \frac{12}{12-1} \left(1 - \frac{2.21}{8.28} \right) = 0.80$$

p is the proportion of students passing a given item

q is the proportion of students that did not pass a given item

σ^2 is the variance of the total score on this assessment

x is the student score minus the mean score;

x is squared and the squares are summed (Σx^2);

the summed squares are divided by the number of students minus 1 (N-1)

k is the number of items on the test.

3.2.3 Estimating Reliability Using the Kuder-Richardson Formula 21

When item and data is large, or technological assistance is not available to assist in the computation of the large number of cases and items, the simpler, and sometimes less precise, reliability estimate known as Kuder-Richardson Formula 21 is an acceptable general measure of internal consistency. The formula requires only the test mean (M), the variance (σ^2)

²) and the number of items on the test (k). It assumes that all items are of approximately equal difficulty. (N =number of students)

For this example, the data set used for computation of the KR 20 is repeated.

Student (N=10)	X (Score)	$x = X - \text{mean}$ (score-mean)	x^2
A	11	4.5	20.25

$1/k^2$ in KR21 is a mathematical approximation of the ratio $1/pq/1^2$ in KR20. The formula simplifies the computation but will usually yield, as evidenced, a lower estimate of reliability. The differences are not great on a test with all items of about the same difficulty.

3.2.4 Estimating Reliability Using Cronbach's alpha (α)

This is the most commonly used reliability coefficient. It is based on the internal consistency of items in the tests. It is flexible and can be used with test formats that have more than one correct answer. The split-half estimates and KR20 are exchangeable with Cronbach's alpha. When examinees are divided into two parts and the scores and variances of the two parts are calculated, the split-half formula is algebraically equivalent to Cronbach's alpha. When the test format has only one correct answer, KR20 is algebraically equivalent to Cronbach's alpha. Therefore, the split-half and KR20 reliability estimates may be considered special cases of Cronbach's alpha.

It is important for you to note that knowing how to derive a reliability estimate whether using split halves, KR20 or KR21 and cronbach alpha is good but that knowing what the information means is the most important. A high reliability coefficient is no guarantee that the scores are well-suited to the outcome. It does tell you if the items in the assessment are strongly or weakly related with regard to student performance. If all the items are variations of the same skill or knowledge base, the reliability estimate for internal consistency should be high. If multiple outcomes are measured in one assessment, the reliability estimate may be lower. That does not mean the test is suspect. It probably means that the domains of knowledge or skills assessed are somewhat diverse and a student who knows the content of one outcome may not be as proficient relative to another outcome.

3.2.5: Establishing Inter-rater Agreement

This is an important type of reliability where scoring requires some judgment. It is an agreement among those who evaluate the quality of the items relative to a set of stated criteria. Preconditions of interrater agreement are:

- A scoring scale or rubric which is clear and unambiguous in what it demands of the student by way of demonstration.
- Evaluators who are fully conversant with the scale.

The end result is that all evaluators are of a common mind with regard to the items and that one mind is reflected in the scoring scale or rubric and that all evaluators should give the same or nearly the same ratings. The consistency of rating is called interrater reliability. Agreement levels should be at 80% or higher to establish a claim for interrater agreement.

After agreement have been established, list the ratings given to each student by each rater for comparison:

Student	Score: Rater 1	Score: Rater 2	Agreement
A	6	6	X
B	5	5	X
C	3	4	
D	4	4	X
E	2	3	
F	7	7	X
G	6	6	X
H	5	5	X
I	3	4	
J	7	7	X

Dividing the number of cases where student scores between the raters are in agreement (7) with the total number of cases (10) determines the rater agreement percentage (70%).

When there are more than two raters, the consistency of ratings for two at a time can be calculated with the same method. For example, if three teachers are employed as raters, rater agreement percentages should be calculated for:

- Rater 1 and Rater 2
- Rater 1 and Rater 3

Rater 2 and Rater 3

4.0 Conclusions

Other things being equal, the longer the test, the more reliable it will be when reliability concerns internal consistency. This is because the sample of behavior is larger

5.0 Summary

In this unit we have discussed the validity and reliability of test scores. We have also tried to learn different procedures for establishing validity and reliability.

6.0 Tutor-Marked Assignment

- 1) Study the following statements and indicate if the statement refers to validity or reliability or both. Specify the type of reliability or validity.
 - a. The correlation between a scholastic aptitude test and grade-point average is 0.74
 - b. The test was given twice to the same group, the coefficient of correlation between the scores is 0.87
 - c. Three Science Educators studied the items of the test for their relevance to the objective of the curriculum
 - d. My creativity test really measures creativity
- 2) A Head Teacher wishes to buy one of three sets of mathematics tests. As far as he can judge they are equally valid. The reliability of each of them is reported to be 0.75. Discuss other things the Head teacher must put into consideration if he has funds to buy only one of the sets

7.0 References/Further Readings

Cohen, J. W. (1988) *Statistical Power Analysis for the Behavioural Sciences* (2nd edition), Hillsdale, NJ: Lawrence Erlbaum Associates

Stevens, J. (1996), *Applied Multivariate Statistics for Social Sciences*, (3rd edition), Matway, New Jersey: Lawrence Erlbaum Associates

Tabachnick, B. G. & Fidell, L. S. (1996), Using Multivariate Statistics (3rd edition), New York: HarpeCollins
<http://www.gower.k12.il.us/Staff/ASSESS>

www.apsu.edu/oconnort/3760/3760lect03a.htm

faculty.vassar.edu/lowry/webtext.html

www.uwsp.edu/psych/stat/14/nonparm.htm

writing.colostate.edu/guides/research/stats/pop2a.cfm

www.shodor.org/interactivate/lessons/IntroStatistics

www.statbasics.co.uk

UNIT 2: ITEM ANALYSIS

CONTENTS

- 1.0 Introduction
- 2.0 Objectives
- 3.0 Main Content
 - 3.1 What is Item Analysis?
 - 3.2 Calculating Difficulty Index
 - 3.3 Calculating Discrimination Index
 - 3.4 Calculate Analysis of Response Options
 - 3.5 Interpreting the Results of Item Analysis
- 4.0 Conclusions
- 5.0 Summary
- 6.0 Tutor-Marked Assignment
- 7.0 References/Further Readings

1.0 INTRODUCTION

A good test is one with good items. But how do we know a good item and the items to be selected among the very many items constructed? To select a good item requires item analysis. There are many methods of item analysis. In this unit we will consider the various types of item analysis especially those very useful to a teacher.

2.0 OBJECTIVES

At the end of this unit you should be able to:

- Define item analysis
- List the different types of item analysis
- Calculate difficulty index
- Calculate discrimination index
- Calculate response options
- Interpret the results of item analysis

3.0 MAIN CONTENT

3.1 What is Item Analysis?

Item analysis is a process of examining class-wide performance on individual test items. There are three common types of item analysis which provide teachers with three different types of information:

- ***Difficulty Index*** - Teachers produce a difficulty index for a test item by calculating the proportion of students in class who got an item correct. (The name of this index is counter-intuitive, as one actually gets a measure of how easy the item is, not the difficulty of the item.) The larger the proportion, the more students who have learned the content measured by the item.
- ***Discrimination Index*** - The discrimination index is a basic measure of the validity of an item. It is a measure of an item's ability to discriminate between those who scored high on the total test and those who scored low. Though there are several steps in its calculation, once computed, this index can be interpreted as an indication of the extent to which overall knowledge of the content area or mastery of the skills is related to the response on an item. Perhaps the most crucial validity standard for a test item is that whether a student got an item correct or not is due to their level of knowledge or ability and not due to something else such as chance or test bias.
- ***Analysis of Response Options*** - In addition to examining the performance of an entire test item, teachers are often interested in examining the performance of individual distracters (incorrect answer options) on multiple-choice items. By calculating the proportion of students who chose each answer option, teachers can identify which distracters are "working" and appear attractive to students who do not know the correct answer, and which distracters are simply taking up space and not being chosen by many students. To eliminate blind guessing which results in a correct answer purely by chance (which hurts the validity of a test item), teachers want as many plausible distracters as is feasible. Analyses of response options allow teachers to fine tune and improve items they may wish to use again with future classes.

3.2 Calculating Difficulty Index

Let us consider the steps for calculating difficulty index using the example of 25 students who took a test. The items are multiple choice types.

- Step 1 Determine the number of students choosing each item right and the number that sat for the test.

In our example there are 16 students who choose option B (the correct option) out of the 25 students who took the test.

Step 2 Calculate the proportion of students who got the item right.

Proportion that got item right = Number that got the item right/number that took the test

$$\begin{aligned} &= 16/25 \\ &= 0.64 \end{aligned}$$

This proportion is called the **difficulty index**. Difficulty Indices range from .00 to 1.0.

3.3 Calculating Discrimination Index

The steps for calculating discrimination index are:

Step 1 Sort your tests by total score and create two groupings of tests- the high scores, made up of the top half of tests, and the low scores, made up of the bottom half of tests. If normal class of 30 students, divide class in half; If you have a large sample of around 100 or more, you can cut down the sample you work with by taking top 27% (27 out of 100) and bottom 27% (so only dealing with 54, not all 100). Imagine this information for our example, 13 students had total scores to place them in the high group and 12 students in the low group.

Step 2 Determine the number of students in each group that had the correct option for each item. Imagine for our example 10 out of 13 students in the high group and 6 out of 12 students in the low group got the particular item correct

Step 3 For each group, calculate a difficulty index for the item.
Difficulty Index for High Group = $10/13 = .77$
Difficulty Index for Low Group = $6/12 = .50$

Step 4 Calculate the Discrimination Index

Subtract the difficulty index for the low scores group from the difficulty index for the high scores group to get the discrimination index for the item. For our example:
Discriminatory Index = $.77 - .50 = .27$
Discrimination Indices range from -1.0 to 1.0.

3.4 Calculate Analysis of Response Options

Step 1 Determine the number of students who choose each answer option. In our example, for a particular item, 4 students choose option A, 16 choose option B, 5 choose option C and 0 choose option D.

Step 2 Calculate the Analysis of Response Option
This is the proportion of students choosing each response option. For each answer options divide the number of students who choose that answer option by the number of students who took the test. From our example:

$$\begin{aligned} \text{A. } & 4/25 = .16 \\ * \text{B. } & 16/25 = .64 \\ \text{C. } & 5/25 = .20 \\ \text{D. } & 0/25 = .00 \end{aligned}$$

3.5 Interpreting the Results of Item Analysis

In our example, the item had a difficulty index of .64. This means that sixty-four percent of students knew the answer. If a teacher believes that .64 is too low, he or she can change the way they teach to better meet the objective represented by the item. Another interpretation might be that the item was too difficult or confusing or invalid, in which case the teacher can replace or modify the item, perhaps using information from the item's discrimination index or analysis of response options.

The discrimination index for the item was .27. The formula for the discrimination index is such that if more students in the high scoring group chose the correct answer than did students in the low scoring group, the number will be positive. At a minimum, then, one would hope for a positive value, as that would indicate that knowledge resulted in the correct answer. The greater the positive value (the closer it is to 1.0), the stronger the relationship is between overall test performance and performance on that item. If the discrimination index is negative, that means that for some

reason students who scored low on the test were more likely to get the answer correct. This is a strange situation which suggests poor validity for an item.

The analysis of response options shows that those who missed the item were about equally likely to choose answer A and answer C. No students chose answer D. Answer option D does not act as a distracter. Students are not choosing between four answer options on this item, they are really choosing between only three options, as they are not even considering answer D. This makes guessing correctly more likely, which hurts the validity of an item.

4.0 CONCLUSIONS

Item analysis is a process of examining class-wide performance on individual test items.

5.0 SUMMARY

In this unit we have discussed the three common types of item analysis which provide teachers with three different types of information. We also calculated each item analysis and interpreted the results.

6.0 TUTOR-MARKED ASSIGNMENT

What type of validity is indicated in each of the following statements?

1 Scores on typing test (TT) correlated +.62 with chief Typist's rating after one year on the job.

2 Achievement test is based on our analysis of two most popular biology programme recognised by WAEC for school certificate examination

7.0 REFERENCES/FURTHER READINGS

Cohen, J. W. (1988) Statistical Power Analysis for the Behavioural Sciences (2nd edition), Hillsdale, NJ: Lawrence Erlbaum Associates

Stevens, J. (1996), Applied Multivariate Statistics for Social Sciences, (3rd edition), Matiway, New Jersey: Lawrence Erlbaum Associates

Tabachnick, B. G. & Fidell, L. S. (1996), Using Multivariate Statistics (3rd edition), New York: HarperCollins

<http://www.gower.k12.il.us/Staff/ASSESS>

www.apsu.edu/oconnort/3760/3760lect03a.htm

[**faculty.vassar.edu/lowry/webtext.html**](http://faculty.vassar.edu/lowry/webtext.html)

www.uwsp.edu/psych/stat/14/nonparm.htm

[**writing.colostate.edu/guides/research/stats/pop2a.cfm**](http://writing.colostate.edu/guides/research/stats/pop2a.cfm)

www.shodor.org/interactivate/lessons/IntroStatistics

[**www.statbasics.co.uk**](http://www.statbasics.co.uk)

UNIT 3: SCALING TECHNIQUES

CONTENTS

- 1.0 Introduction
- 2.0 Objectives
- 3.0 Main Content
 - 3.1 Guttman Scaling
 - 3.2 Rasch Scaling
 - 3.3 Rank-Order Scaling
 - 3.4 Constant Sum Scaling
 - 3.5 Paired Comparison Scale
- 4.0 Conclusions
- 5.0 Summary
- 6.0 Tutor-Marked Assignment
- 7.0 References/Further Readings

1.0 INTRODUCTION

Some times we need to measure or order attributes with respect to quantitative attributes or traits. This is known as Scaling. Scaling provides a mechanism for measuring abstract concepts. Respondents evaluate two or more objects at one time and objects are directly compared with one another as part of the measuring process. In this unit we will consider the different types of scaling.

2.0 OBJECTIVES

At the end of this unit you should be able to:

- List the different types of scaling
- Describe the different types of scaling
- Differentiate among the different types of scaling
- Give the advantages of scaling

3.0 MAIN CONTENT

3.1 Guttman Scaling

This can also be referred to as a cumulative scoring or scalogram analysis. The intent of this survey is that the respondent will agree to a point and their score is measured to the point where they stop agreeing. For this

reason questions are often formatted in dichotomous yes or no responses. The survey may start out with a question that is easy to agree with and then get increasingly sensitive to the point where the respondent starts to disagree. You may start out with a question that asks if you like music at which point you mark yes. Four questions later it may ask if you like music without a soul and which is produced by shady record labels only out to make money at which point you may say no. If you agreed with the first 5 questions and then started disagreeing you would be rated a 5. The total of questions you agreed to would be added up and your final score would say something about your attitude toward music.

3.2 Rasch Scaling

This probabilistic model provides a theoretical basis for obtaining interval level measurements based on counts from observations such as total scores on assessments. This analyzes individual differences in response tendencies as well as an item's discrimination and difficulty. It measures how respondents interact with items and then infers differences between items from responses to obtain scale values. This model is typically used to analyze data from assessments and to measure abilities, attitudes, and personality traits.

3.3 Rank-Order Scaling

This gives the respondent a set of items and then asks the respondent to put those items in some kind of order. The "order" could be something like preference, liking, importance, effectiveness, etc. This can be a simple ordinal structure such as A is higher than B or be done by relative position (give each letter a numerical value as in A is 10 and B is 7). You could present five items and ask the respondent to order each one A-E in order of preference. In Rank-Order scaling only $(n-1)$ decisions need to be made.

3.4 Constant Sum Scaling

With this ordinal level technique respondents are given a constant sum of units such as points, money, or credits and then asked to allocate them to various items. For example, you could ask a respondent to reflect on the importance of features of a product and then give them 100 points to allocate to each feature of the product based on that. If a feature is not important then the respondent can assign it zero. If one feature is twice as important as another then they can assign it twice as much. When they are done all the points should add up to 100.

3.5 Paired Comparison Scale

This is an ordinal level technique where a respondent is presented with two items at a time and asked to choose one. This is the most widely used comparison scale technique. If you take n brands then $[n(n-1)/2]$ paired comparisons are required. A classic example of when paired comparison is used is during taste tests. For example you could have a taste test in which you have someone try both Coke and Pepsi and then ask them which one they prefer.

7.0 CONCLUSIONS

Scaling is very important in Survey research. However, you must make sure the correct scaling is used during study.

8.0 SUMMARY

In this unit you have learnt the different types of scaling. You also learnt the differences among the different types of scaling and the advantage of each.

6.0 TUTOR-MARKED ASSIGNMENT

- 1) Compare and Contrast the different types of scaling.
- 2) An English teacher is interested in finding out the most commonly used method for teaching reading. Which of the scaling methods would you suggest to the English teacher. Given reason for your choice.

7.0 REFERENCES/FURTHER READINGS

Cohen, J. W. (1988) Statistical Power Analysis for the Behavioural Sciences (2nd edition), Hillsdale, NJ: Lawrence Erlbaum Associates

Stevens, J. (1996), Applied Multivariate Statistics for Social Sciences, (3rd edition), Matlway, New Jersey: Lawrence Erlbaum Associates

Tabachnick, B. G. & Fidell, L. S. (1996), Using Multivariate Statistics (3rd edition), New York: HarpeCollins

<http://www.gower.k12.il.us/Staff/ASSESS>

www.apsu.edu/oconnort/3760/3760lect03a.htm

faculty.vassar.edu/lowry/webtext.html

www.uwsp.edu/psych/stat/14/nonparm.htm

writing.colostate.edu/guides/research/stats/pop2a.cfm

www.shodor.org/interactivate/lessons/IntroStatistics

www.statbasics.co.uk