COURSE GUIDE

AGR 501
STATISTICS AND RESEARCH METHODOLOGY

Course Team Prof. G.N. Akpa **(**Writer/Developer) - ABU
Prof. G. E. Jokthan (Course Editor) - NOUN
Prof. N. E. Mundi (Programme Leader) - NOUN
Dr. Sabo Elizabeth (Course Coordinator) - NOUN

NATIONAL OPEN UNIVERSITY OF NIGERIA

National Open University of Nigeria
Headquarters
University Village
Plot 91, Cadastral Zone,
Nnamdi Azikiwe Expressway
Jabi, Abuja

Lagos Office
14/16 Ahmadu Bello Way
Victoria Island, Lagos

e-mail: centralinfo@noun.edu.ng
website: www.nouedu.net

Published by
NOUN Press

Printed 2017

ISBN: 978-058-700-X

CONTENTS **PAGE**

INTRODUCTION

AGR 501 – is a three- credit unit course for students pursuing first degree in Agricultural Science. It will also be useful for lecturer, teachers and researchers in various other disciplines.

The course consists of 23 units covering basic concepts in statistical design in the field of agriculture. This course guide will gives you insight into the nature of the course, the materials you are going to use and how you are to use the materials for meaningful benefits.

You are encouraged to devote, at least, two hours to studying each unit. You are also advised to attempt the self-assessment exercises as they are necessary for proper understanding of the units. You are also requested to pay more attention to tutor-marked assignments (details are provided in a separates file).

There will be tutorial classes and you are also advised to use online training materials (YouTube) for further comprehension. Details of the time and location of the tutorials will be known to you; this is a great opportunity for you to have face to face contact with your course coordinator. Any area that is not clearly understood will be properly explained.

WHAT YOU WILL LEARN IN THIS COURSE

The overall aim of this course is to introduce you to the fundamental principles of research methodology in Agricultural Science. During the course, you will learn about the importance of experimental design, how to conduct a novel research and various computational approach as related to different statistical methodologies.

COURSE AIMS

The aim of the course is to give you a better understanding of experimental design, and how it applies to agriculture, especially in relation to Animal Science. The aim of the course will be achieved by the following:

- definition of research problem and its necessities
- developing hypothesis and its objectives
- helping you to understand the various principles of research design
- exposing you to questionnaire preparation and collection of data
- exposing you to measurement and evaluation in agricultural research
- helping you to understand the various statistical theories

- giving you insight into statistical methods and presentation of research findings in tabular and graphical forms

COURSE OBJECTIVES

To achieve the aims set out above, the course has a set out overall objectives. In addition, each unit also has specific objectives. The unit objectives are always given at the beginning of each unit. You should read them before you start working through the unit. You may always want to refer to them during your study of the unit so as to check on your progress. You should always look at the unit objectives after completing a unit. In this way, you can be sure that you have done what was required of you by the unit.

Below are the wider objectives of the course, as a whole. By meeting these objectives, you should have achieved the aims of the course as a whole. On successful completion of the course, you should be able to:

- define research problem and state its importance
- explain the importance of hypothesis testing
- explain the various assumptions involved in hypothesis testing
- describe research design and its procedures
- explain the merits and demerits of each of the research design
- know how to prepare questionnaire to fit your research objectives
- determine steps involved in data collection
- evaluate and validate measurement in agricultural(animal science) research
- explain various statistical theories with specific assumptions
- explain the importance of statistical methodology, its computation and applications
- present your research findings in tabular and graphical forms for easy consumption.

WORKING THROUGH THIS COURSE

To complete this course, you are required to read the study units as well as other related textbooks. Then each unit contains self-assessment exercises, and at certain points in the course, you are required to submit assignments for assessment purposes. At the end of the course, you are going to sit for a final extermination. The course should take you about 42 weeks to complete.

COURSE MATERIALS

In this course material you will find the following:

1. Course guide
2. Study unit
3. Textbooks and references
4. Presentation schedule

STUDY UNITS

There are 23 study units divided into seven modules in this course. They are as follows:

Module 1 Definition of Research Problem and its Necessities

Unit 1 Research Problem and Selecting the Problem
Unit 2 Research Hypothesis
Unit 3 Procedure for Hypothesis Testing
Unit 4 Advantages and Limitations of Hypothesis Testing

Module 2 Principles of Research Design

Unit 1 Meaning of Research Design
Unit 2 Concept Relating to Research Design
Unit 3 Basic principles of Experimental Design
Unit 4 Criteria for Defining a Problem

Module 3 Questionnaire Preparation and Collection of Data

Unit 1 Collection of Primary Data
Unit 2 Collection of Secondary Data
Unit 3 Guidelines for Constructing Questionnaire

Module 4 Measurement and Evaluation

Unit 1 Measurement in Research and in Scale
Unit 2 Sources of Error in Measurement
Unit 3 Data Collection Techniques

Module 5 Statistical Theory

Unit 1 Sampling Fundamentals in Agricultural Research
Unit 2 Sampling Distribution
Unit 3 Central Limit and Sampling Theory

Module 6 Statistical Methods

Unit 1 Guidelines and Explanations of Statistical Methods
Unit 2 Approaches to Data Compilations and Analysis
Unit 3 Computational Techniques of Experimental Design

Module 7 Presentation of Research Findings in Narrative, Tabular and Graphical Forms

Unit 1 Results Presentation through Tabular Form
Unit 2 Results Presentation through Graphical Form

The first module define research problem and its necessities which is composed of two units (1-2). Module 2 explains research hypothesis testing and various procedures involved in testing hypothesis while Module 3 discusses principles of research design. Module 4 deals with how to prepare questionnaire and data collection. Module 5 explains measurement and evaluation in agricultural research while Module 6 explains statistical theory in agricultural research. Module 7 concentrates on statistical methods. Research result presentation in tabular and graphical forms is enveloped in Module 7.

Each unit includes specific objectives and summaries of key issues and ideas. There are textbooks which will also provide additional information. The exercises in each unit have to be done to ensure that you are going through the course. Doing the exercises is also a way to gauge your progress and to reinforce your understanding of the material. Together with tutor-marked assignments, these exercises will assist you in achieving the stated learning objectives of the individual units, and the course as a whole.

ASSIGNMENT FILE

In this file, you will find the details of the work you must submit to your tutor for marking. The marks you obtain will form part of your total score for this course.

ASSESSMENT

There are two aspects to the assessment of the course. First is Tutor-Marked Assignment (TMA) and there is a written examination. You are advice to be sincere in attempting the exercises. In attempting the assignments, you are expected to apply the information, knowledge and techniques gathered during the course.
The assignment must be submitted to your tutor for formal assessment in accordance with the deadlines stated in the presentation schedule and the assignment file. The work submitted to your tutor for assessment will count for 30% of your total course mark. At the end of the course, you will need to sit for a final examination; this examination will count for 70% of your total course mark.

TUTOR –MARKED ASSIGNMENT

You are encouraged to submit your assignments as required. Each assignment counts for 10% of your marks for the course. You will be able to complete your assignment from the information gathered from reading the study unit and other recommended texts. However, it is desirable that you research more and read other references as this will give you a broader view point and may provide deeper understanding of the subject.

When you have completed each assignment, send it to your tutor. Make sure that each assignment reaches your tutor on or before the deadline given in the presentation schedule and assignment file. If for any reason, you cannot complete your work on time, contact your tutor to discuss the possibility of an extension. Extensions will not be granted after the due date, except for exceptional circumstances.

FINAL EXAMINATION AND GRADING

The final examination for this course will take three hours and have a value of 70% of the total course grade. The examination will consist of questions which reflect the types of self-assessment exercises and tutor-marked assignment you have previously encountered. All areas of the course will be assessed. Take time to reverse the entire course before the examination. The examination covers information from all parts of the course.

PRESENTATION SCHEDULE

Your course materials give you important dates for attending tutorials and timely completion and submission of your tutor marked assignments. Do remember that you are required to submit all your assignment by the due date. You should guard against falling behind in your work.

COURSE MARKING SCHEME

The following table lays out how the actual marking scheme is broken down.

Table 1: Course Marking Scheme

Assessment	Marks
Assignment 1-4 Final Examination	The best three of the four assignments count for 30% of course marks. 70% of overall course marks.
Total	100% of course marks

COURSE OVERVIEW

This table brings together the units, the number of weeks you should take to complete them and the assignment that follows them.

Table 2: Course Schedule

Units	Title of Work	Weeks activity	Assessment (End of Unit course Guide)
	Module 1		Assignment 1
1	What is research problem? Selecting the problem	1	
2	What is a research hypothesis?	1	Assignment 2
3	Procedure for hypothesis for testing	1	
4	Advantages and limitation of hypothesis testing		
	Module 2		
1	Meaning of research design	1	Assignment 3
2	Concept of relating to research design	1	

3	Basic principles of experimental design	1	
4	Criteria for defining a problem	1	
	Module 3		
1	Collection of primary data	1	
2	Collection of secondary data	1	
3	Guidelines for constructing questionnaire	1	Assignment 4
	Module 4		
1	Measurement in research and measurement in scale	1	Assignment 5
2	Sources of error in measurement	1	Assignment 5
3	Data collection techniques	1	
	Module 5		
1	Sampling fundamental in agricultural research	1	Assignment 6
2	Sampling Distribution	1	Assignment 6
3	Central limit and sampling theory	1	
	Module 6		
1	Guidelines and Explanations of statistical methods	1	Assignment 7
2	Approaches to data compilations and analysis	1	Assignment 7
3	Computational techniques of experimental design	1	
	Module 7		
1	Result presentation through tabular form	1	Assignment 8
2	Result presentation through graphical form	1	Assignment 8
	Revision		
	Examination	18	

HOW TO GET THE MOST FROM THIS COURSE

In distance learning, the study units replace the conventional university lecturer. This is one of the great advantages of distance learning; you can read and work through specially designed study materials at your own pace, and at a time and place that suit you best.

Each of the study units follows a common format. The first item is an introduction to the subject matter of the unit and how a particular unit is integrated with the 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. When you have finished the unit, you must go back and check whether you have achieved the objectives. If you make a habit of doing this you will significantly improve your chances of passing the course.

Self-assessment exercises are interspersed throughout the units. Working through these exercises will help you to achieve the objectives of the unit and prepare you for the assignments and the examination. You should do each exercise as you get to it in the study unit. There will also be numerous examples given in the study units; work through these when you come to them, too.

FACILITATORS/TUTORS AND TUTORIALS

There are 20 hours of tutorials (ten two-hour sessions) provided in support of this course. As soon as you are allocated a tutorial group, you will be notified of the dates, times and location of tutorials, together with the name and phone number of your tutor.

Your tutor will mark and comment on your assignments; he/she will keep a close watch on your progress and on any difficulties you may encounter and provide assistance to you during the course. You must mail your tutor-marked assignments to your tutor well before the due date (at least two working days are required). They will be marked by your tutor and returned to you as soon as possible.

Do not hesitate to contact your tutor by telephone, e-mail, or via the discussion board if you need help. The following might be circumstances in which you would find help necessary.

Contact your tutor if you:

- do not understand any part of the study unit
- have difficulty with the assignments/ exercises
- have a question or problem with your tutors comments on any assignment or with the grading of an assignment.

You should try your best to attend tutorials. This is the only chance to have face to face contact with your tutor and to ask questions. You can raise any problem encountered in the course of your study. To gain the maximum benefit from the tutorials, prepare a list of questions before hand, you will learn a lot from participating actively in the discussions.

SUMMARY

AGR 501– Statistics and Research Methods is a course that intends to introduce you to statistical methodologies ranging from identification of research problem, mapping out your research hypothesis, experimental designs, data analysis and finally, presentation of your research result in tables and graphs. Upon the completion of this course, you will be equipped with the basic knowledge of the principles and techniques involved in data analysis and also be able to:

- define a research problem
- state the merit and demerit of hypothesis testing
- highlight the major and minor advantages in using different research designs
- explain the steps involved in questionnaire preparation
- differentiate between primary and secondary data collection
- discuss why is necessary to minimize error and optimize efficiency in research designs
- enumerate the contribution of statistical theory in reduction of experimental noise
- discuss the various steps in presentation of data in graphs and tables.

To maximise gain from this course, you should try to apply the principles you have learnt especially in the area of experimental design and data analysis.

I wish you success and hope that you will find the course both interesting and useful. Good luck.

MAIN COURSE

CONTENTS PAGE

MODULE 1 DEFINITION OF RESEARCH PROBLEM AND ITS NECESSITIES

UNIT 1 RESEARCH PROBLEM AND SELECTING THE PROBLEM

CONTENTS

1.0 INTRODUCTION

The first unit is structured to enable you understand how research problems are been formulated and viable steps in solving the problems. The unit will, give you an insight into characteristics of research problem and gives you highlight into the major techniques in defining research problem and constructive synopsis of how to effectively plan your research, so as to arrive at a real time solution to your research problem.

2.0 OBJECTIVES

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

- define a research problem
- identify a research problem
- explain the necessities for defining a research problem
- state the main characteristics for selecting your research problem.

3.0 MAIN CONTENT

3.1 Definition of Research Problem

Research refers to search for knowledge. The difficulty in which a researcher experiences in the process of searching for knowledge is called a problem. Generally, a research problem refers to some difficulty which researchers experience in theoretical situations and want to find solution to the problem. The major component to establishing a research problem is when individual or organisation is faced with a problem of deciding on what course of action is the best. The minor component of research problem are stated below

i. There must be an individual or group that has some difficulty in achieving their set goals.
ii. There must be environments to which difficulty pertains.
iii. There must be some set objectives to be attained.

Therefore, a research problem stimulates a researcher to find the best solution for the given problem, so that optimally attained objectives could be established in the context of a given environment.

SELF- ASSESSMENT EXERCISE

Study your environment carefully. Generate a research problem and identify a need for the research.

3.2 Characteristics of a Research Problem

It is widely believed that all research problems are researchable in the general sense of it. Scientifically, research problems have two major characteristics which are researchable and non- researchable. A researchable research problem has the following listed attributes:

i. A discrepancy does not exist between desired and observed situation.
ii. We do not know why discrepancy exists.
iii. We do not know the best solution.

Non- researchable research problems have the following listed attributes*:*

i. A discrepancy exits between desired and observed situation.
ii. We know why the discrepancy exists.
iii. We know the best solutions.

A good discriminating example between researchable and non-researchable problem is that a veterinary pathologist wants fist-hand report of what animals think and feel in the 4 to 10 hours before their death. So, to answer this problem, the researcher should bear in mind that this is a non-researchable research because the animals are dead and cannot talk, so they cannot fill out survey or answer the questions. A researchable problem and the researcher must be able to answer the questions the research poses and have scientific reliable data source to solve the problem.

SELF ASSESSMENT EXERCISE

i. Compare and contrast researchable and non- researchable problems.
ii. Differentiate between delimitation and limitation.

3.3 Procedures for Problem Selection

Research problem undertaken for study must be carefully selected. Every researcher must map out strategies in solving the problems. The following points may be observed by researchers in selecting a research problem:

- controversial subjects should not became the choice of an average researcher
- a research with a lot of ambiguity should be avoided
- selection must be preceded by a preliminary study
- a subject which is overdone should not be chosen for it will be difficult to give informative conclusion in such a case

- from practicability of the research, all the necessary resources must be
- feasibility of the research must be ascertained
- efficient and industrious teams that are well equipped in skills should be involved
- there must be adequate training and personal predilections
- the research must be in light of your training and personal preferences.

SELF- ASSESSMENT EXERCISE

State four procedures for selecting research problem.

4.0 CONCLUSION

In this unit, you have learnt that it is important to know about the source of research problem as first step. Knowing about the attributes of a suitable problem and keeping them in mind will help you to be focused. Also, delimiting the research problem will help you narrow down a general idea to one that is manageable by you and the pros and cons of problem selection for effective planning of research that is feasible.

5.0 SUMMARY

The discussion presented so far has served to demonstrate the criteria for defining a problem which constitute a principal component in solving research problems. Revealed the importance of proper selection before observing the above mentioned points and zest for solving problem should take a prime position in the mind of the researcher

6.0 TUTOR-MARKED ASSIGNMENT

1. What is research problem and briefly explain its major and minor component.
2. Differentiate between researchable and non researchable problems.
3. What are the major points for consideration in a selecting research problem?

7.0 REFERENCES/FURTHER READING

Greer, S. (1977). “On the selection of problems.” In M. Bulmer (Ed.) *Sociological Research Methods: An Introduction*. London: Macmillan.

Chance, William A. (1975): *Statistical Methods for Decision Making.* Bombay: D.B. Taraporevala Sons & Co. Pvt. Ltd.

Hay, A. (1984). “Tactics in Presenting a Research Proposal”. 16(4): 332-333.

UNIT 2 RESEARCH HYPOTHESIS

CONTENTS

1.0 INTRODUCTION

A hypothesis is a tentative explanation for certain behaviors, phenomena, or events that have occurred or will occur. The hypothesis states the researcher's expectations concerning the relationship between the variables in the research problem as outlined in the first module. Formulating a research hypothesis requires research questions and generation of operational definitions for variables.

2.0 OBJECTIVES

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

- define the concept of hypothesis testing
- state the assumptions of parametric and non-parametric test of hypothesis.

3.0 MAIN CONTENT

3.1 Basic Concepts in Hypothesis Testing

A hypothesis is a statement that needs to be proved or disproved. It is also a statement that describes or explains a relationship among variables or they can be described as tentative statement of the expected relationships between two or more variables. The major reasons why hypothesis is important in our research are as follows:

i. It directs the investigation
ii. It allows the investigator to confirm relationship or no relationship
iii. It provides useful template for organising the results and conclusions
iv. It provides specific focus
v. It helps in summarizing the results.

Hypothesis has two major classifications which are null and alternate hypothesis. A null hypothesis is a statement that has no actual relationship between or among variables (H_0) or it can be refer to as hypothesis of no difference. For example, there is no significance difference in the yield response of maize with high and low quality protein. An alternate hypothesis is a statement that suggests a potential outcome that the researcher may expect (H_a). In testing a research hypothesis, the researcher must be mindful of type 1 and 11 error so as not to make false conclusions. When the researcher reject a true null hypothesis is called type I error while acceptance of false null hypothesis is called type II error.

SELF ASSESSMENT EXERCISE

Differentiate between type I and II errors in animal breeding research.

3.2 Test of Hypothesis

The hypothesis testing determines the validity of assumption with the intent to choose between two conflicting hypotheses about a value of population parameter. They are several tests of hypotheses which are subsumed into parametric and non- parametric test of hypotheses.

3.2.1 Basic assumption of parametric tests

i. Large sample size
ii. Observations must come from a normal population
iii. Parameters must have high dependence on the parent populations

3.2.2 Basic assumption of non- parametric test

i. It does not depend on assumption about the parameters of the parent population
ii. They assume only nominal or ordinal data
iii. More observations are needed

SELF ASSESSMENT EXERCISE

State the basic assumptions of parametric and non-parametric test.

3.3 Parametric Tests

The most frequently used parametric tests for agricultural research are:

1. z- test
2. t-test
3. Chi-square test
4. f-test

The tests are based on the assumptions of normality.
Z test is based on the probability distribution. It is a test statistics for evaluating the significances of several statistical measures.
Conditions for using z test are:

i. Judging the significance of difference between means of two independent
ii. samples.
iii. Testing binominal distribution on the presumption that tends to approximate
iv. normal distribution.

Testing the significance of median, mode, coefficient of correlation and several other statistical measures.
T –test is based on t- distribution and considered an approximate test in agricultural statistics for judging the significance of a sample mean. Also, it can be used to test the difference between the means of two samples in case of small sample when population variance is not known. It can be used for testing the significance of coefficients of simple and partial correlations. T-test applies in small case sample ($n<30$) when population variance is unknown.

For instance, a poultry geneticist in National Animal Production Research Institute wants to test whether the arithmetic mean of the bodyweight of sire line broiler breed, 3000g is significantly different from 10 birds selected to be the parent of the next generation with an average of 3200g and given the standard deviation as 500g.

Illustration

The hypothetical mean is N_0 =3200g and the hypotheses are as follows:
H_0: μ=3200g
H_1: $\mu \neq$ 3200g
The sample mean is $\bar{x}$ 3000g
Standard deviation(s) = 500g
Standard error, $SE = \frac{S}{\sqrt{n}}$
The calculated value of the t- statistic is

$$t = \frac{\bar{x} - \mu_0}{s/\sqrt{n}} = \frac{4000 - 4200}{500\sqrt{10}} = \frac{-200}{158.23} = -1.26$$

For α=0.05 and degree of freedom (10-1) = 9, the critical value is $t_{\alpha/2} = -$ 2.262, H_0 is not rejected with an α =0.05 level of significance. The sample mean of 3000g is not significantly different from 3200g.
Chi- square test is statistical measure use in sampling analysis for comparing a variance to a theoretical variance.
Conditions for using chi- square are:

i. To test the goodness of fit
ii. To check if categorical data shows dependency
iii. To compare between theoretical populations and actual data, when categories are used.
iv. To test the homogeneity or significance of population variance
v. To test the significance of association between two attributes

Illustration

Prof Akpa had drawn a sample of 10 jersey milking cows from a jersey herd in Shonga, Kwara state. The sum of square deviations from the mean of the given sample is 50. Test the hypothesis that the variance of the population is 5 at 0.05 level of significance.

Solution

n = 10; df (degree of free doom) df =n-1=10-1=9

$\sum\left(x_i - \overline{x}\right)^2 = 50$

$\sigma_s^2 = \frac{\sum\left(x_i - \overline{x}\right)^2}{n-1} = \frac{50}{9}$

N=10

∴ Take the null hypothesis as H$_{0;}$ $\sigma_p^2 = \sigma_s^2$ In order to test this hypothesis, we work out the x^2 value as:

$\chi^2 = \frac{\sigma_s^2}{\sigma_p^2}(n-1) = \frac{\frac{50}{9}}{5}(10-1) = \frac{50}{9}\ X\ \frac{1}{5}\ X\ \frac{9}{1} = 10$

The table value of χ^2 at 5 percent level for 9 d.f is 16.92. Since the calculated value is 10 which is less than the tabulated value. So we accept the null hypothesis and conclude that the variance of the population is 5.

Note: In order that we may apply the chi-square as a test of goodness of fit or to test the significance of association between two attributes. χ^2 is then calculated as follows.

Where;

$\chi^2 = \frac{O_{ij} - E_{ij}}{E_{ij}}$

O_{ij} = observed frequency of the cell in the ith row and jth column

E_{ij} = expected frequency of the cell in the ith row and jth column.

If two distributions (observed and theoretical) are similar, χ^2 =0, but in agricultural practices χ^2 is not equal to zero and as such we must know the sampling distribution of χ^2 so that we may find the probability of an observed χ^2 being given by a random sample from the hypothetical value.

F- test is based on f- distribution and is used to compare the variance of the two independent samples. This text is also used in analysis of variance (ANOVA) for judging the significance of more than two sample means. It can be used for judging the significance of multiple correlation coefficients. Test statistic, f, is calculated and compared with its probable value for accepting or rejecting the null hypothesis.

SELF ASSESSMENT EXERCISE

i. Hypothesis testing is important in our research, why?

ii. Compare and contrast parametric and non-parametric test of hypothesis.

4.0 CONCLUSION

It is clearly stated that for effective planning of a vibrant research, a well-articulated hypothesis has to be formulated.

5.0 SUMMARY

This unit has provided an opportunity to understand the importance of hypothesis testing in research planning

6.0 TUTOR- MARKED ASSIGNMENT

1. Differentiate between a null and alternate hypothesis.
2. State the basic mathematical model for a null and alternate hypothesis.
3. Explain comprehensively the tests of hypothesis.

7.0 REFERENCES/FURTHER READING

Diagnet, J. (1973). *Main Trends in Interdisciplinary Research*. London: George Allen and unwin ltd.

Sadhe, A.N. & Singh, A. (1980). *Research Methodology in Social Sciences*. Bombay; Himalaya Publishing House.

UNIT 3 PROCEDURE FOR HYPOTHESIS TESTING

CONTENTS

1.0 INTRODUCTION

This unit of this module deals with procedures or systematic steps involved in hypothesis testing. Hypothesis testing cannot be effective without the major procedures for validating the test.

2.0 OBJECTIVES

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

- list the various steps involved in hypothesis testing;
- define the procedure for testing hypothesis.

3.0 MAIN CONTENT

3.1 Steps Involved in Testing Hypothesis

Making a formal statement: the formal statement for null (H_0) and alternate (H_a) should be clearly stated considering the nature of the research problem. For instance, a dairy scientist wants to test whether the arithmetic mean of the sample, 4000kg is significantly different from 4200kg in lactation milk yields of 10 cows. Given the standard deviation as 500kg.

Illustration

The hypothetical mean is H_0 =4200kg and the hypotheses are as follows

H_0: μ=4200kg

H_1: $\mu \neq$ 4200kg

The sample mean is $\bar{x}$ = 4000kg
Standard deviation(s) = 500kg
Standard error, $SE = \frac{S}{\sqrt{n}}$
The calculated value of the t- statistic is:

$$t = \frac{\bar{x} - \mu_0}{s/\sqrt{n}} = \frac{4000 - 4200}{500\sqrt{10}} = \frac{-200}{158.23} = -1.26$$

For α=0.05 and degree of freedom (n-1) = 9, the critical value is $t_{\alpha/2}$= $-$2.262, H_0 is not rejected with an α =0.05 level of significance. The sample mean of 4000kg is not significantly different from 4200kg of population mean.

i. Selecting a significance level. The hypotheses are tested on predetermined level of significance and as such should be specified. In agricultural studies, 5% or 1% level is adopted depending on the purpose of the research.

Factors that affect the level of significance

a. Magnitude of difference between sample mean
b. Size of the sample
c. Variability of measurements within samples
d. Whether the hypothesis is directional or non-directional

ii. Deciding the distribution to use: This step involve determination of appropriate sampling distribution. The choice normally remains between normal and t distribution.
iii. Calculation of the probability: The researcher has to calculate the sample result, if it would diverge as widely as it has from expectations, if the null hypothesis were in fact true.
iv. Comparing the probability: calculated probability has to be compare with specified value for α, the significance level. If the calculate probability is equal to or smaller than α value in case of one tailed test and $\frac{\alpha}{2}$ in case of two tailed test, then accept the null hypothesis, then if it is greater, reject the null hypothesis.

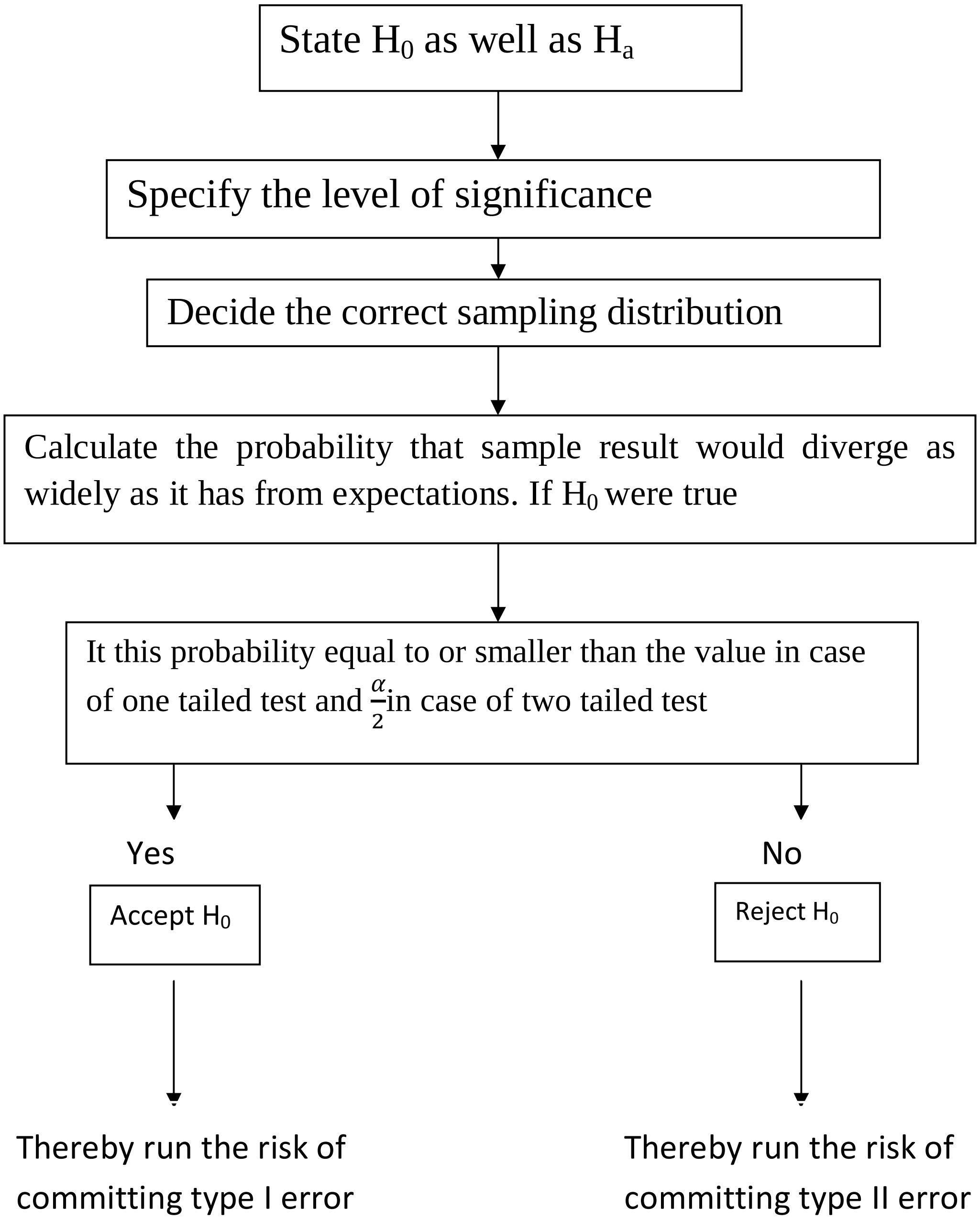
SELF-ASSESSMENT EXERCISE

State and explain the steps involved in hypothesis testing.

3.2 Flow Diagram for Hypothesis Testing

The general procedure for hypothesis testing can also be depicted in the form of a flow chart for better understanding as shown in Fig 1.

Flow Diagram for Hypothesis Testing



SELF ASSESSMENT EXERCISE

Draw a step-wise flow diagram that runs the risk of committing type I or II error.

4.0 CONCLUSION

In this unit, you have learnt how to construct a flow diagram for hypothesis testing. Thus, this will guide you in building computer algorithm in solving agricultural research problems.

5.0 SUMMARY

This unit showed us how we can solve agricultural research problem in a simplified flow chart.

6.0 TUTOR-MARKED ASSIGNMENT

Construct a flow chart for hypothesis testing, systematically.

7.0 REFERENCES/FURTHER READING

Clarke, G.M, (1994). *Statistics and Experimental Design: An Introduction for Biologists and Biochemists.* (3rd ed.). New York: Oxford University Press,.

Litle, T.M. & Hills, F.J. (1978). *Agricultural Experimentations*. New York: John Wiley and Sons

UNIT 4 ADVANTAGES AND LIMITATIONS OF HYPOTHESIS TESTING

CONTENTS

1.0 INTRODUCTION

This unit of the module deals with the major important advantages and limitation of hypothesis testing. For efficient planning of research, it is required of us to set a criterion of acceptance or rejection for decision making.

2.0 OBJECTIVE

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

- list and explain the important advantages and disadvantages of hypothesis testing.

3.0 MAIN CONTENT

3.1 Important Advantages of Hypothesis Testing

From our previous discussion on hypothesis testing, you would have observed the major selling point of hypothesis testing. The advantages are as follows:.

i. Hypothesis testing allows you to conclusively accept or reject your estimate. This is useful if you have large amount of estimates that you just care about rejecting or accepting them.

ii. Relatively simple to calculate: Setting your criterion for acceptance or rejection is simple to estimate.
iii. Suitable for comparing a treatment with the control
iv. It has the propensity to suggest new experiment and observations.
v. It is used to validate the adequacy of experiment.
vi. It is limited in scope and precision: the narrow nature of hypothesis makes it more testable.

SELF ASSESSMENT EXERCISE

Among the selling points of hypothesis testing, which of the factors is most important and why?

3.2 Limitations of the Tests of Hypotheses

1. The test does not explain why the difference exist (i..e. between means of the two samples). They simply indicate whether the difference is due to fluctuations of sampling.
2. The tests should not be used in mechanical fashion. It should be kept in view that testing is not decision- making.
3. Results of significance tests are based on probabilities and as such cannot be expressed with full certainty. When a test shows that a difference is statistically significant ($p<0.05$), then it simply suggests that the difference is probably not due to chance.
4. Statistical inferences based on the significance tests cannot be said to be entirely correct evidences concerning the reliability of the hypothesis.

This is specifically so in case of small samples where the probability of drawing erring inferences happens to be generally higher. All these limitations suggest that in problems of statistical significance, the inference techniques must be combined with adequate knowledge of the subject matter along with the ability of good judgment.

SELF ASSESSMENT EXERCISE

Discuss all the major advantage and limitations of hypothesis testing.

4.0 CONCLUSION

In this unit, knowing the advantage and disadvantage is a great step in conducting an effective research. It is interesting to know that hypothesis testing is more robust in agricultural research because of the strength in decision making.

5.0 SUMMARY

This unit has shown that for you to conduct an effective research, you must understand the pros and cons of hypothesis testing.

6.0 TUTOR MARKED ASSIGNMENT

List the advantages and limitation of hypothesis testing.

7.0 REFERENCES/FURTHER READING

Clarke, G. M. (1994). *Statistics and Experimental Design: An Introduction for Biologists and Biochemists. (*3rd ed.). New York: Oxford University Press.

Sharma, H. D. & Mukleryi. S. P. (1976). *Research in Economics and Commerce, Methodology and Surveys.* Valranasi: Indian Biographic Centre.

MODULE 2 PRINCIPLES OF RESEARCH DESIGN

UNIT 1 MEANING OF RESEARCH DESIGN

CONTENTS

1.0 INTRODUCTION

Research design is the overall plan for connecting the conceptual research problems to the empirical research. With the advent of simple to complex layout in research design, there is a need to carefully use the most robust tool to provide the best decision.

2.0 OBJECTIVES

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

- explain the meaning and significance of a research design
- discuss the quality of a good research design.

3.0 MAIN CONTENT

3.1 Definition of the Concept of Research Design

A research design is the conceptual structure within which research is conducted. It constitutes the blueprint for the collection, measurement and analysis of agricultural research data. As such the design maps out the computations and finally data analysis. More explicitly, the design decisions happen to be in respect of:

a) What is the study about?
b) Where is experimental site of the study?
c) What type of data is required?
d) Where can the required data be found?
e) What is the time period of the study?
f) What will be the sample design?
g) What techniques of data collection will be used?
h) How will the data be analysed?
i) In what style will the report be prepared?

3.1.1 Types of Research Design

a. Sampling Design
This deals with the statistical method of randomly selecting research items to be observed for the given study.

b. Observational Designs
This deals with the conditions under which the observations are to be taken.

c. Statistical Design
This concerns the research layout and how data observed from the field and collated are to be analyzed.

d. Operational Design
This deals with the techniques by which the procedures specified in the sampling, statistical and observational layouts can be carried out.

SELF ASSESSMENT EXERCISE I

If you are asked to conduct a surveillance study in your area of specialisation, how will you roll out your decision layout?

3.2 Need for Research Design

Research design is needed because it has the smooth sailing of the various research operations. Therefore, making research as efficient as possible yields maximal information and minimises expenditure of effort, time and money. It helps us map out informative variables in our data collection and analysis for our research project. Preparation of research design should be done with great care to minimise experimental error. Research design has a great bearing on the reliability of the results arrived at and as such constitutes the firm foundation of the entire edifice of the research work. Although, at times the need for a well thought out research design at times might not be realised due to parallax error or random error.

It is therefore imperative that an efficient and appropriate design must be established before igniting research operations, so as to isolate the inadequacies in the research design.

SELF-ASSESSMENT EXERCISE

i. Explain the strength of the need for research design in a successful research experiment.
ii. Research design must minimise bias and maximise reliability. Discuss.

3.3 Features of a Good Research Design

A good research design must have the tendency to minimise bias and maximise efficiency. Also a good research design must not be noisy. This implies that it must submit itself to all the assumptions necessary for maximising output. It is always assumed that design which gives the smallest experimental error is supposed to be the best design, in many investigations. A singular research design cannot serve the purpose of all types of research design appropriate for a particular research problem, usually involves the consideration of the following factors.

i. The means of obtaining information
ii. The availability and skills of researcher and its collaborators
iii. The objective of the problem to the studied
iv. The availability of time and money for the research work

Note, if the research study is an exploratory or a formative research, wherein the major emphasis is on discovery of ideas and insights, the research design most appropriate must be flexible enough to permit the consideration of many different aspects of a phenomenon.

4.0 CONCLUSION

There are several research designs and the research must decide in advance of data collection and analysis to which design would minimise error and maximise efficiency.

5.0 SUMMARY

This unit has provided a green light to understanding the importance of a research design towards a successful planning to accomplish the objectives of the investigation.

6.0 TUTOR-MARKED ASSIGNMENT

1. List and explain the various types of research design.
2. What are the appropriate selection factors for a good research design?
3. Explain the meaning and significance of research design.

7.0 REFERENCES/FURTHER READING

Ackoff, R. L. (1962). *Scientific Method*. New York: John Wiley & Sons.

Bowley, A. L. (1937). *Elements of Statistics*. (6th ed.) London: P.S. King and Staples ltd.

UNIT 2 CONCEPTS RELATING TO RESEARCH DESIGN

CONTENTS

1.0 INTRODUCTION

It is important to have residual knowledge of the key operational terms before conducting an experiment. Doing so, will minimize bias and maximize reliability of the research.

2.0 OBJECTIVES

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

- explain and illustrate various forms of research design
- write out the appropriate mathematical models.

3.0 MAIN CONTENT

3.1 Important Concepts

Before describing the architectures of different research designs, it will be interesting to explain the various concepts relating to designs so that these may be better and easily understood. The major concepts are listed below.

1. **Dependent and Independent Variables**
   A concept which can take on different quantitative values is called a variable. As such subjects like weight, height, yield and income are all examples of variables. Quantitative attributes are quantified on the basis of the presence or absence of the concerned attributes. Phenomena or process which can take on quantitatively different values even in decimal points are called continuous variables. Note in research that variables are not continuous if they can only be

expressed as integer values or in discrete form. Age of an animal or plant is an example of continuous variables, but the number of animals is an example of a non-continuous variable. If one variable depends upon or is a consequence of the other variable, it is termed as dependent variable, and the variable that is antecedent to the dependent variable is termed as independent variable. For instance, if we say that the height of an animal depends upon age. That means that the height is a dependent variable and age is an independent variable.

2. **Extraneous variable:** Independent variables that are not related to the purpose of the study but may be affected by dependent variable are termed extraneous variables. Suppose a researcher wants to test the hypothesis that there is a relationship between body weight of dairy animals, breed and milk yield in the Maizube Dairy Farms. In this case, milk yield is dependent variable and breed is independent variable. Feeding may as well affect the bodyweight, but since it is not related to the purpose of the study undertaken by the researcher, it will be termed extraneous variable. However, whatever effect is noticed on dependent variable as a result of extraneous variable it is technically described as experimental error.

3. **Control:** The major importance of a good research design is to minimise the influence of random noise or effect of extraneous variable and maximise efficiency. It is used as restrain experimental conditions.

4. **Research Hypothesis**: When a prediction or an unverified relationship is to be tested by scientific methods, it is termed research hypothesis. The usual norm is that a research hypothesis must contain a dependent and independent variable.

5. **Treatment:** The different condition under which experimental and control groups are subject to is referred to as treatment. Example is a comparative analysis of feeding regime on four genetic groups of an animal.

6. **Experiment:** The process of examining the truth of a statistical hypothesis, relating to some research problem, is known as experiment. Experiment can be grouped as comparative and absolute experiment. For example we want to determine the impact of a feed type on the yield of milk, it is a case of absolute; but if we want to compare the impact of some other feed type, our experiment then will be termed as a comparative experiment.

7. **Experimental Unit**: The pre-determined plots or blocks where different treatments are imposed and upon which observations are made.

4.0 CONCLUSION

This unit highlighted the routine procedures that need to be in the mind of the researchers for a reliable result delivery output.

5.0 SUMMARY

This unit has given a synopsis of important concepts that influence error minimisation in research experimentation.

6.0 TUTOR-MARKED ASSIGNMENT

Explain the following terms:

1. Research hypothesis
2. Dependent and independent variables
3. Control
4. Experimental blocks

7.0 REFERENCES/FURTHER READING

Campbell, D.T. & Stanley, J. (1963). *Experimental and Quasi-Experimental Designs for Research on Teaching.* Boston, MA: Houghton Mifflin.

Spector, P.E. (1993). "Research designs." In M.L. Lewis-Beck(Ed.) *Experimental Design and Methods. International Handbook of Quantitative Applications in the Social Sciences*, Vol 3. London; Sage, 1-74.

UNIT 3 BASIC PRINCIPLES OF EXPERIMENTAL DESIGN

CONTENTS

1.0 INTRODUCTION

An experimental design declares how to obtain data or a set of rules used to choose samples from population. Also, in a central experiment, an experimental design describes how to assign treatments to experimental units, but within the frame of the design there must be an element of randomness of treatment assigned.

2.0 OBJECTIVES

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

- differentiate between formal and informal experimental design
- state the principles of experimental design.

3.0 MAIN CONTENT

3.1 Concept of Experimental Design

An experiment can be defined as a planned research conducted to obtain new facts or to confirm or refute the results of previous experiments. The process of observing, collecting or measuring data can be considered an experimental design.

Steps in planning an experiment

1. Introduction to the problem.
2. Statement of the hypothesis.
3. Description of the experimental design.
4. Collection of data.
5. Analysis of data.
6. Interpretation of results relative to the hypothesis.

All these basic steps have been explained in previous modules. Professor Fisher has enumerated three principles of experimental designs as follows:

1. The principles of replication.
2. The principle of randomisation.
3. Principle of local control.

Principle of replication: The experiment should be repeated more than once. Thus, each treatment is applied in many experimental units instead of once, this tends to increase the accuracy of the experiment. For example, suppose we are to examine the performance of two broiler breeds of chicken going by the principle of replication, we first divide the pens into several parts, and ensured that each of the breed is represented in all the components of the pens. We can then collect the data of the two breeds at a given age and draw conclusion by comparing the same. This then guarantee the reliability of the results.

Principle of randomisation: This provides protection when we conduct an experiment against the effect of extraneous factors. In other words, this principle indicates that we should design or plan our experiment in such a way that the variations caused by extraneous factors can all be combined under the umbrella of chance. For instance, if we rear one breed of broiler chickens, say in the first half of the parts of the poultry house and the other breed is reared in the other half, then it is just possible that the rearing method may be different in the first half in comparison to the other half. If this is so, our results would be biased and not realistic. In such a situation, we may assign the breeds of broiler chickens to be reared in different parts of the poultry house on the basis of some random sampling techniques to protect our research against extraneous factors.

Principle of local control: This is another principle of experimental design. The extraneous factor is made to vary deliberately over as wide as necessary and it needs to be done in such a way that the variability it causes can be measured and hence eliminated from experimental error. Using previous example on broiler breeds, we should first divide the poultry house into several homogenous parts, known as blocks and then each block is divided into parts equal to the number of treatments to be imposed. Then the treatments are randomly assigned to these parts of a block. In this way the principle of local control can eliminate the variability due to extraneous factor(s) from the experimental error.

3.2 Categories of Experimental Design

We can classify experiment design into two components viz: informal and formal experimental designs. Informal experimental designs are those of analysis based on differences in magnitude while formal experimental designs use precise statistical procedures for analysis. The informal experimental designs are:

a. Before and after without control design
b. After only with control
c. Before and after with control design: In this design two areas are selected and the dependent variable is measured in both the areas for an identical time-period before the treatment. The treatment is then introduced into the test area only and dependent variable is measured in both for an identical time period. The treatment effect is determined by subtracting the change in the dependent variable in the control area from the change in independent variable area. It has advantage over the other two informal experimental designs because it avoids extraneous variation of the test and control areas.

The formal experimental designs consist of completely randomised design, randomised complete block design, factorial design and Latin Square design.

a.) **Completely Randomised Design (CRD)**: This involves only two principles viz: the principle of replication and randomisation of experimental designs.

Characteristics

i. It is the simplest possible design and its procedure for analysis is also easier.
ii. Its subjects are randomly assigned to experimental treatments.
iii. It can also accommodate unequal replication due to its flexibility. Also, it is analysed by one-way analysis of variance.

Advantage of completely randomised design

i. CRD is flexible because it can be done even in a limited number of experimental subjects however equal number of subjects for each treatment is encouraged.
ii. Statistical analysis is simple and easy.
iii. Statistical analysis is manageable even though some of the experimental subjects or treatments are lacking.

Disadvantages

Difficult to establish homogeneity of characteristics among subjects either experimental or control sample. If researcher fails to establish homogeneity on the characteristics among subject it will lead to greater experimental error.

Assumptions:

i. Independent random samples (results of one sample do not affect other samples).
ii. Samples are taken from normal population(s) or distribution.

b) Randomised Complete Block Design (RCBD): It is an improvement over CRD. In the RCBD, the principle of local control can be applied along with other two principles of experimental design. Subjects are first divided into groups, known as blocks, such that within when an experiment is finished the data can be arranged for easy computing.

Probably this is the most used and useful of the experimental designs. The design features are:

i. Takes advantage of grouping similar experimental units into blocks or replicates.
ii. The blocks of experimental units should be as uniform as possible.
iii. The purpose of grouping experimental units is to have the units in a block as uniform as possible so that the observed differences between treatments will be largely due to "true" differences between treatments.

Randomisation Procedure

i. Each replicate is randomised separately.
ii. Each treatment has the same probability of being assigned to a given experimental unit within a replicate.
iii. Each treatment must appear at least once per replicate.

Advantages of the RCBD

1. Generally more precise than the CRD.
2. No restriction on the number of treatments or replicates.
3. Some treatments may be replicated more times than others.
4. Missing plots are easily estimated.
5. Whole treatments or entire replicates may be deleted from the analysis.
6. If experimental error is heterogeneous, valid comparisons can still be made.

Disadvantages of the RCBD

1. Error degree of freedom is smaller than that for the CRD (problem with a small number of treatments).
2. If there is a large variation between experimental units within a block, a large error term may result (this may be due to too many treatments).
3. If there are missing data, a RCBD experiment may be less efficient than a CRD

   NOTE: The most important item to consider when choosing a design is the uniformity of the experimental units.

How blocks are formed is critical to the effectiveness of the analysis.

- With field plots, blocks are laid out so that they are perpendicular to the maximum direction of change in the disturbing factor to be controlled.
- Wide border (discard) areas are used to overcome interference between neighboring plots (i.e. to maintain independence of responses) within blocks and between blocks.

  Time blocks may need discard times between “replication.

c) **Factorial Experiments:** A factorial experiment has two or more sets of treatments that are analysed at the same time. Note that treatments signify particular level of independent categorical variable, often called a factor. Therefore if two or more factors are examined in an experiment, it is a factorial experiment. There are two components namely main effect and interaction effect. The main effect of a factor is defined to be the change in response produced by a change in the level of a factor. Interaction effect is the multiplier or synergistic effect of the factors.

Characteristics

i. All combinations of factor levels are tested.
ii. The effect of a factor alone is called main effect
iii. Effect of different factors acting together is called interaction.
iv. They provide equivalent accuracy with less labour.
v. They permit various other comparisons of interest.
vi. In a factorial design, all possible combinations of the levels of the factors are investigated in each replication.
vii. Each treatment appears the same number of times in each block.
viii. It is analyse by two-way analysis of variance.
ix. Experimental unit are divided to treatments and blocks.

Advantages of Factorials

i. Factorial layouts are more efficient than one-factor-at-a-time experiments.
ii. A factorial design is necessary when interactions may be present to avoid misleading conclusions.
iii. Factorial designs allow the effects of a factor to be estimated at several levels of the other factors, yielding conclusions that are valid over a range of experimental conditions.
iv. The simplest type of factorial designs involves only two factors or sets of treatments.
v. There are levels of factor A, and *b* levels of factor B, and each replicate contains all *ab* treatment combinations.

Additional Concepts in Factorial Designs

Fixed Effect:

- the levels of a factor are pre-determined
- the inference will be made only on the levels used in the experiment

Random Effect:

- the levels of a factor are randomly chosen
- the inference will be drawn about a population, from which the factors are chosen

d) Latin Square Design: Is an experimental design frequently used in agricultural research. The conditions under which agricultural investigations are carried out are different from those in other studies because it allows for simultaneously blocking of two factors.

Characteristics

i. Random permutation of column.
ii. Random permutation of row.
iii. Treatments are assigned in Latin letters in a random fashion.
iv. The number of rows and columns has to correspond to the number of treatment levels.

4.0 CONCLUSION

For a desirable or robust reliability of experimental results, the researcher must use to most suitable experimental design to actualise its goal.

5.0 SUMMARY

This unit has introduced us to the pros and cons of experimental design and the characteristics to delineate between different experimental designs.

6.0 TUTOR-MARKED ASSIGNMENT

Explain and illustrate the following designs:

1. Latin square design
2. Simple factorial design
3. Completely randomized design
4. Randomized complete block design

7.0 REFERENCES/FURTHER READING

Box, G.E.P. (1978). *Statistics for Experimenters: An Introduction to Design, Data Analysis and Model Building.* New York: John Wiley and Sons.

Clarke, G.M. (1994). *Statistics and Experimental Design: An Introduction for Biologists and Biochemists.* (3rd). New York: Oxford University Press.

UNIT 4 CRITERIA FOR DEFINING A PROBLEM

CONTENTS

1.0 INTRODUCTION

This unit will focus on criteria for defining a problem. An essential characteristics of this problem requires the researcher to classify the criteria into two components namely: source of research problem and delimiting the research problem.

2.0 OBJECTIVES

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

- set out constructively the criteria for defining a research problem
- explain in details the sources of research problem
- differentiate between limitation and delimitation.

3.0 MAIN CONTENT

3.1 Source of Research Problem

There is no clear cut classification of sources of research problem in the field of agriculture. However, Greer (1977) suggests three broad classifications which remain important in agricultural research. He classified them as policy problem, problems of social philosophy and problems which are original to the development of social sciences disciplines. In other words, three important sources for problematic situations are experience, deductions from theory and relevant literature. The selection of research topic has the likelihood to arise from researcher's experience. In connection with the problem of its immediate society such as hunger, unemployment, crime and climate change. These factors will open up windows of research topic through a sort of intuitive process.

The second source of research problem relates to deductions that can be inferred from generalizations in the society that the researcher is familiar with. These deductions from crop and animal reaction to the environment are then fitted into empirical frame of reference through research. Examples from theoretical dispositions take the following form: Response to selection of broiler chickens for body weight traits in the tropical conditions of Nigeria. In this case, the researcher asks questions: Is there any improvement in bodyweight after each generation of selection? He then goes further to carry out a systematic research to ascertain whether empirical data confirm or reject the hypothesis and hence the theory.

Lastly, selection of a research problem may emanate from problems situated in the literature. An extensive review of literature opens up a lot of windows in an area that has not been exploited within a particular environment.

SELF ASSESSMENT EXERCISE

Explain Greer Classification of sources of research problem

3.2 Delimiting the Research Problem

Delimitation is a premediated limitation that the researcher narrows down while limitation is a weakness in the study which becomes apparent during or after its completion. The major categories of delimitations are:

i. The selection of sources
ii. The number of observations, subject or cases
iii. Time and geographic location

This will help the research to have focus. Also, the minor steps in delimiting the research problem could be accomplish through the following steps:

i. Read extensively the literatures related to your field of interest.
ii. Exchange ideas with experts (domestic and international) in your field through online media.
iii. Notice the difference between delimitation and limitation.

3.3 Necessity for Defining the Problems

There is a saying that a problem clearly stated is a problem half solved. Problem to be solved must possess some special attributes such as clarity, relevance and unambiguous. The need for defining a problem is a herculean task. However, systematic approach in solving the problem is highly desirable. The usual necessity is that the researcher should formulate a question and set out systematic approach for solving the problem.

The major techniques in defining the problem involve the following steps:

1. Statement of problem: The research statement should be specific, focused statements and question. Importantly, the feasibility of a particular solution has to be considered and the same should be kept in view while stating the problem.
2. Understanding the nature of the problem: The best way to have a soft landing is to discuss the problem with experts who have comprehensive understanding on the formations of the problem for proper guidance. The knowledge of the environment in which the problem was discovered is desirable.
3. Developing the ideas through discussions: Discussion concerning a problem often produces useful information through social media (Twitter, Facebook, Whatsapp) and colleagues working in similar direction.
4. Restructuring the research problem: Finally the researcher must restructure the research problem into a working proposition so as to help in development of working hypothesis.

SELF ASSESSMENT EXERCISE

List the major techniques in defining a research problem in your field of study.

4.0 CONCLUSION

The task of defining a research problem is a giant step in solving research problem. This problem must follow a sequential pattern such as resolving the ambiguities, thinking and rethinking process from an operational point of view so as to clear way for the development of working hypothesis and solving the problem itself.

5.0 SUMMARY

In this unit, you learnt the definition, characteristics and necessity of research problem. You also learnt about the technique in defining a research problem through a sequential pattern.

6.0 TUTOR-MARKED ASSIGNMENT

1. Explain briefly the second sources of research problem according to Greer 1977 classification with a reference example.
2. What are the techniques in defining a research problem?
3. Explain delimitation in a research problem

7.0 REFERENCES/FURTHER READING

Chance. W. A. & Bombay, D. B. (1975). S*tatistical Methods for Decision Making*. Taraporevala Sons and Company.

Ferber, R. (1948). *Handbook of Marketing Research*. New York: McGraw-Hill Inc.

Fisher, R.A. (1958). *Statistical Methods for Research Workers*. (13th ed.). New York: Hafner Publishing Company.

Sharma, H.D. & Mukheryi, S. P. (1976). *Research in Economics and Commerce, Methodology and Surveys.* Varanasi: Indian Biographic Centre.

Sharma, B.A.V. (1983). *Research Methods in Social Science*. New Delhi: Sterling Publishers Ltd.

MODULE 3 QUESTIONNAIRE PREPARATION AND COLLECTION OF DATA

UNIT 1 COLLECTION OF PRIMARY DATA

CONTENTS

1.0 INTRODUCTION

The first unit is designed to enable you understand the process involve in data collection. This begins after a research problem has been defined and research design has been laid out, while deciding about the method of data collection to be used for the study, the researcher should keep in mind two types of data viz, primary and secondary data.

2.0 OBJECTIVES

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

- determine the method of data collection
- state the procedures of data collection through questionnaire.

3.0 MAIN CONTENT

3.1 Observation Method

This is the most common method of primary data collection in behavioral sciences. In a way, we all observe things around us, but this sort of observation is subjective and not tool and method of data collection for the researcher when it serves a formulated research purpose.

Under observation method, the information is sought by way of investigator's direct observation without asking from the respondent. For instance, in a study relating to assessment of feed standards or non-conventional feedstuff evaluation, the investigator instead of asking the type of raw materials for feed ingredients used by the respondent, may himself look at the feed for assessment. The main bias is eliminated, if observation is done accurately. Secondly observation are more direct toward the scope of the research; it is not complicated by either the past behaviour or future intentions or attitudes. Lastly, this method is independent of respondent's willingness to respond and as such ration on the part of respondents as happens to be the cases in the interview or questionnaire method.

Limitation of Observational Method

i. It is an expensive method.
ii. The information provided by this method is very limited.
iii. Random factors may interfere in the observational task.

SELF-ASSESSMENT EXERCISE

In a way, observation method of primary data collection is subjective and not scientific. Explain.

3.2 Interview Method

The interview method of data collection involves presentation of oral-verbal stimuli and reply in terms of oral-verbal responses.

(a) **Personal interviews**: this requires a person known as interviewer asking questions generally in a face-in-to-face contact to the other respondents. This sort of interview may be in the form of direct personal investigation the interviewer has to gather the information marginally or personally from the sources concerned. This is also an on the spot data collection and has to meet people from whom data have to be collected. This method is particularly suitable for intensive investigations. The major exception to this method is that it may not be possible to contact directly the persons concerned on account of the extensive scope of enquiry, the direct personal investigation technique may not be used. In such scenarios, an indirect oral examination can be conducted under which the interviewer has to cross-examine other persons who are supposed to be knowledgeable about the problem under investigation and the information, obtained is recorded. Most of the agricultural commissions and co-on investigations make use of this method.

The method of information gathering through personal interviews is usually carried out in structured interviews, the alternate form of structured interview. Therefore, it is majorly characterised by a flexibility of approach to questioning. It also demands deep knowledge and greater skill on the part of the interviewer. Generally, in the case of descriptive studies, we use the technique of structured interview because it is more economical, providing a safe basis for generalisation and requiring relatively lesser skill on the part of interviewer.

SELF-ASSESSMENT EXERCISE

i. Compare and contrast the merit and demerit of structural and non-structural interview.
ii. Nonstructural interview is majorly characterised by flexibly. Explain.

3.3 Questionnaires

This form of data collection is quite popular, particularly in case of large or big enquiries. It is a common method that is being adopted by private individuals, research workers, private and public organisations, governments and students. The process involves sending or giving the questionnaire to the respondents concerned with a request to answer the question and return the questionnaire. A questionnaire consists of a number of questions printed or typed in a definite order on a form or set of forms.

Merits of Questionnaires

1. There is low cost even when the sample size is large.
2. It is free from bias of the interviewer; answers are in respondents own words.
3. Respondents have adequate time to give well thought out answers.
4. Respondents, who are not easily accessible, can also be reached conveniently through survey monkey.

Demerits of Questionnaire

1. Low rate of return of the duly filled in questionnaires.
2. It can be used only when respondents are educated and co-operating.
3. High possibilities of ambiguous replies.
4. Slow pace of response as compared to other form of data collection.

SELF-ASSESSMENT EXERCISE

Mention the merits and demerits of questionnaires.

4.0 CONCLUSION

Assessing the major methods of primary data collection, you must have observed that data collection through questionnaire is the most effective. This is validated by its flexibility, reliability and ability to absorb large sample size.

5.0 SUMMARY

In this unit, you have learnt about the various methods of primary data collection and their attributes in positive and negative forms.

6.0 TUTOR-MARKED ASSIGNMENT

(1) Explain how observation method is a scientific tool and method of data collection for researchers.
(2) What is a questionnaire?

7.0 REFERENCES/FURTHER READING

Oppenheim, A. N. (1966). *Questionnaire Design and Attitude Measurement*. New York: Basic Books.

Payne, S. (1951). *The Art of Asking Questions*. Princeton University Press.

UNIT 2 COLLECTION OF SECONDARY DATA

CONTENTS

1.0 INTRODUCTION

This unit will introduce you to characteristics of secondary data. Majorly secondary data are data that are already available, collected and analyzed by someone. When the researcher utilizes secondary data, then he has to look into various sources from where he can collect the necessary information.

2.0 OBJECTIVE

At the end of this unit, you should be able to highlight the characteristics of secondary data.

3.0 MAIN CONTENT

3.1 Sources of Secondary Data

Secondary data can be grouped into published or unpublished data. Usually published data are available in:

(a) Various publications of the central, state and local governments archives;
(b) Various publications of foreign Governments or international bodies and their subsidiary organisation;
(c) Technical and trade journals;
(d) Books, magazines of various associations connected with business, banks, educational consults, etc;
(e) Reprints prepared by research scholars from universities;

(f) Public records and statistics, historical documents, and other sources of unpublished data which are many and may be found in diaries, letters, unpublished biographies and auto biographies, private individuals and organisations.

3.2 Characteristics of Secondary Data

a. Reliability of data: The reliability can be tested by finding out the following things about the said data:

i. Who collected the data?
ii. What were the source(s) of the data?
iii. Were they collected by using proper methods?
iv. At what time were they collected?
v. Was there any bias of the compiler?
vi. What level of accuracy was desired?

b. Suitability of data: The data that are suitable for one enquiry may not necessarily be found suitable in another enquiry. Hence, if the available data are found to be unsuitable, they should not be used by the researcher. In this context, the researcher must be very carefully to scrutinise the units of collection used as at time of data collection.

c. Adequacy of data: If the level of accuracy achieved in data is found to be noisy for the purpose of present enquiry, they will be considered as inadequate and should not be used by the researcher. Another characteristic of inadequacy are related to an area which may be either narrower or wider than the area of the present enquiry.

Note that, it is very risky to use already available data. The already available data should be used by the researcher only when he finds them reliable, suitable and adequate.

Advantages of Secondary Data

i. It is economical
ii. It is time saving
iii. It helps to make primary data collection more specific
iv. It helps to improve the understanding of the problem

Disadvantages of Secondary Data

i. They are usually noisy because of large influence of environmental effects.
ii. Class boundaries may be different when units are homogenous

SELF-ASSESSMENT EXERCISE

State and explain the major delimitation of secondary data.

4.0 CONCLUSION

It is therefore concluded that secondary data are cheaper and can be quickly obtained.

5.0 SUMMARY

In this unit, you have learnt the attributes of secondary data. The major positive attribute is that they are cheaper and easily accessible while the negative attribute is that the distribution or dataset might be noisy.

6.0 TUTOR-MARKED ASSIGNMENT

1. What are the various factors to be tested in accessing the reliability of a secondary data?
2. Itemise the characteristics of secondary data.

7.0 REFERENCES/FURTHER READING

Pearl, J. (2000). *Causality: Models, Reasoning and Inference*. Cambridge. U.K: Cambridge University Press.

Salsburg, D. (2001). *The lasting Tasting Tea: How Statistics Revolutionised Science in the 20th Century*. New York: W.H Freeman and Company.

UNIT 3 GUIDELINES FOR CONSTRUCTING QUESTIONNAIRE

CONTENTS

1.0 INTRODUCTION

In this unit, 6 points have been identified which the researcher must pay attention to when constructing an appropriate and effective questionnaire.

2.0 OBJECTIVE

At the end of this unit, you should be able to construct the steps in questionnaire formulations.

3.0 MAIN CONTENT

3.1 Basic Steps in Questionnaire Construction

1. The researcher must keep in view the problem he is to study for it provides the starting point for developing the questionnaire. He/she must be clear about the various aspects of his research problem to be dealt with in the course of his research project.
2. Appropriate form of questions depends on the nature of information sought, sampled respondents and the kind of analysis intended.
3. Rough draft of the questionnaire should be prepared, giving due thought to the appropriate sequence of putting questions.
4. Researcher must invariably re-examine, and in case of need may revise the rough draft for a better one.
5. Pilot study should be undertaken for pre-testing the questionnaire. The questionnaire may be edited in the light of the results of the pilot study.

Questionnaire must contain simple but straight forward directions for the respondents so that they may not feel any difficulty in answering the questions.

4.0 CONCLUSION

For a reliable construction of questionnaires, these listed procedures have to be undertaken to validate the adequacy of the research.

5.0 SUMMARY

The major synopsis for effective and efficient questionnaire formulations has been listed.

6.0 TUTOR-MARKED ASSIGNMENT

In a research methodology class, how would you prepare a template for questionnaire formulations?

7.0 REFERENCE/FURTHER READING

Cody, R. P,. & Smith, J. (1997). *Applied Statistics and SAS Programming Language*. Practice Hall.

MODULE 4 MEASUREMENT AND EVALUATION

UNIT 1 MEASUREMENT IN RESEARCH

CONTENTS

1.0 INTRODUCTION

This unit gives an introduction to measurement in agricultural science. Technically speaking, measurement is a process of mapping aspects of a domain onto other aspects of a range to some rule of correspondence. The researcher should bear it in mind that by measurement, we simply mean the process of assigning numbers to objects or observations. The most widely used classification of measurement scales are: (a) nominal scale (b) ordinal scale (c) interval scale (d) ratio scale.

2.0 OBJECTIVES

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

- explain in details the concept of measurement
- explain measurement scale as related to agricultural design
- list various types of measurement scale.

3.0 MAIN CONTENT

3.1 Measurement in Agricultural Research

In the field of agricultural science, we are said to measure when we use some yardstick or criteria to determine weight of animals. We also measure when we judge animals based on their features. We, thus, measure animals and plants as well as abstract concept termed as random error. Measurement is a relatively complex and demanding task, especially when it concerns qualitative and /or abstract phenomena. By measurement we mean the process of assigning numbers to objects or observations. For example, height, weight, purity or biological age of dairy cows can be measured. These can be done directly with some standard unit of measurement such as kilograms, centimeters, etc. Thus, we can expect high accuracy with less noise in such quantitative measures, but if the concept is abstract and the measurements are not standardised, we are less confident about the accuracy of the results of measurements. Also, in measuring, we devise some form of scale in the range and then transform or map the properties of objects from the domain onto this scale; for example, in case we are to find the male to female ratio of calves in the dairy herd, then we may tabulate our birth type according to sex. In terms of set theory, this process is one of mapping the observed physical properties of all the birth type of dairy cows (the domain) onto a sex classification (the range). The rule of correspondence is: If the object in the domain appears to be male, assign to "0" and if female assign to "1". In this artificial or nominal way, categorical data (qualitative or descriptive) can be made into numerical data and if we thus code the various categories, we refer to the numbers we recorded as nominal data. *Nominal data* are numerical in name only, because they do not share any of the properties of the numbers we deal with in ordinary arithmetic.

3.2 Measurement Scale in Agricultural Research

The most widely used classifications of measurement scale are:

a. Nominal scale
b. Ordinal scale
c. Interval scale
d. Ration scale

a Nominal scale

This is a system of assigning number symbols to events in order to label them. Such numbers cannot be considered to be associated with an ordered scale for their order is of no consequence; the numbers are just convenient labels for the particular class of events and as such have no quantitative value. Nominal scales provide convenient ways of keeping track of animals, objects and events in agricultural measurement.

Research implies a statement of 'greater than' or 'less than' (an equality statement is also acceptable) without your being able to state how much greater or less. A percentile or quartile measure is used for measuring dispersion. Correlations are restricted to various rank order methods. Measures of statistical significance are restricted to the non-parametric methods.

b Ordinal scale

This is the lowest level of the ordered scale that is commonly used. Ordinal scale places events in order, but there is no attempt to make the intervals of the scale equal in terms of some rule. Rank orders represent ordinal scales and are frequently used in agricultural research relating to qualitative phenomena. An agricultural student's rank in his graduation class involves the use of an ordinal scale. Ordinal scales only permit the ranking of items from highest to lowest. Ordinal measures have no absolute values, and the real differences between adjacent ranks may not be equal.

Thus, the use of an ordinal scale in agricultural research is to measure dispersion. Product moment correlation techniques are appropriate and the generally used tests for statistical significance are 't' test and 'F' test.

c Interval scale

The interval scale is adjusted in terms of some rule that has been established as a basis for making the units equal. The units are equal so far as one accepts the assumptions on which the rule is based. The primary limitation of the interval scale is the lack of a true zero; it does not have the capacity to measure the complete absence of a trait or characteristics in plant and animals.

Interval scales provides more powerful measurement than ordinal scales for interval scale also incorporates the concept of equality of interval. As such more powerful statistical measures can be used with interval scales. Mean is the appropriate measure of central tendency, while standard deviation is the most widely used to measure dispersion. There is generally no measure of dispersion for nominal scale. Chi-square test is the most common test of statistical significance that can be utilized. For the measures of correlation, the contingency coefficient can be worked out.

Nominal scale is the least powerful level of measurement in agricultural research. It indicates no order or distance relationship and has no arithmetic origin. A nominal scale simply describes differences between agricultural items (plants, animals) by assuming them to categories. Nominal data are thus, counted data. In spite of all this, nominal scales are still very useful and are widely used in agricultural survey and other ex-post-facto research when data are being clarified by major sub-groups of the population.

d Ratio scale

The ratio scale has an absolute or true zero of measurement. The term ‘absolute zero’ is not as precise as it was once believed to be. For instance, the zero point on a centimeter scale indicates the complete absence of length or height. But an absolute zero of temperature is theoretically unobtainable and it remains a concept of existing only in the researcher’s mind. The number of minor traffic-rule violations and the number of incorrect letters in a page of typed script represent scores on ratio scale.

Ratio scale represents the actual amount of variables. Measures of physical dimensions such as weight, height of animals, etc are examples. Generally, all statistical techniques are useable with ratio scales and all manipulations are carried out with ratio scale values. Geometric and harmonic means can be used as measures of central tendency and coefficient of variation may also be calculated.

SELF ASSESSMENT EXERCISE

Differentiate between the four measurements scales in agricultural statistics.

4.0 CONCLUSION

In this unit, you should understand that measurement in agricultural research can take two forms which are qualitative and quantitative forms. The qualitative are descriptive in nature and majorly takes categorical forms while quantitative traits are continuous. Proceeding from the nominal scale (the least type of scale) to ratio scale (the most precise), relevant information is obtained increasingly. If the nature of the variable permits, you should use the scale that provides the most precise description as an agricultural researcher.

5.0 SUMMARY

In this unit, you should have it in mind that coding in agricultural research is one of the vital steps in error minimisation for effective research output. You have clearly understood the various measurement scales from the least precise to the most precise and their application in agricultural research.

6.0 TUTOR-MARKED ASSIGNMENT

1. Differentiate between qualitative and quantitative traits in animal breeding and genetics.
2. Briefly describe the measurement in agricultural research.
3. There is no measure of dispersion for nominal scales. Explain.
4. Explain in details the limitation of interval scale.
5. Discuss in details ordinal scale and its application in agricultural research.
6. Explain the concept of 'Absolute Zero.'

7.0 REFERENCES/FURTHER READING

Box, G. E. P. (1978). *Statistics for Experimenters: An Introduction to Design, Data Analysis, Model Building*. New York: John Wiley& Sons.

Clarke, G. M. (1994). *Statistics and Experimental Design: An Introduction for Biologists and Biochemists.* (3rd). New York: Oxford University Press,.

Cochran, W. G. & Cox, G. M. (1992). Experimental Designs, (2nd). New York: John Wiley & Sons,

UNIT 2 SOURCES OF ERROR AND TECHNIQUE OF DEVELOPING MEASUREMENT TOOLS

CONTENTS

1.0 INTRODUCTION

Minimising error is a key to effective research delivery in exploratory studies. As such the researcher must be aware of sources of error in measurement. The researcher should bear it in mind that there is no way you can totally eliminate errors from a research study, you can only minimise the error term.

2.0 OBJECTIVES

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

- determine the major sources of error in agricultural research
- know the various techniques in development of measurement tool.

3.0 MAIN CONTENT

3.1 Sources of Error in Measurement

Measurement should be precise and unambiguous in an ideal research study. This objective, however, is often not met in its entirety. As such you must be aware of the sources of error in measurement. The following are the possible sources of error in measurement.

a. Respondent

At times the respondent may be reluctant to express strong negative feelings or it is just possible that he may have very little knowledge but may not admit his ignorance. All this reluctance is likely to result in an interview of 'guesses'. Transient factors like fatigue, boredom, anxiety may limit the ability to respond accurately and fully.

b. Situation

Situational factors may come in the way of correct measurement. Any condition which places a strain on interview can have serious effects on the researcher-respondent rapport.

c. Measurer

The researcher can distort responses by rewarding or recording questions. His behaviour, style and looks may encourage or discourage certain replies. Careless mechanical processing may distort the findings. Errors may also creep in because of incorrect coding, faulty tabulation and/or statistical calculations, particularly in the data analysis stage in agricultural measurement.

d. Instrument

The possibility of error arising because of defective measuring instrument in agricultural measurement cannot be overlooked. The use of complex instrument that is beyond the comprehension of the researcher or that he never handles, response choice omissions are a few things that make the measuring instrument defective and may result in measurement errors. Another type of instrument deficiency is the poor sampling of agricultural items (such as plants, animals) concerned.

SELF-ASSESSMENT EXERCISE

Give examples to each of the factors listed above and discuss briefly any one of the factors?

3.2 Technique of Developing Measurement Tools

The technique of developing measurement tools involves a four-stage process, consisting of the following:

i. Concept development
ii. Specification of concept dimensions
iii. Selection of indicators
iv. Formation of index

i. Concept Development

This concept means that the researcher should arrive at an understanding of the major concepts pertaining to his study. This step of concept is more apparent in theoretical studies than in the more pragmatic research, where the fundamental concepts are often already established.

ii. Specification of Concept Dimensions

Here, the researcher needs to specify the dimensions of the concepts that he developed in the first stage. This task may either be accomplished by deduction i.e. by adopting a more or less intuitive approach or by empirical correlation of the individual dimensions with the total concept and/or the other concepts.

iii. Selection of Indicators

Once the dimensions of a concept have been specified, the researcher must develop indicators for measuring each concept element. Indicators are specific question scales, or other devices by which respondent's knowledge, opinion, expectation, etc are measured. As there is seldom a perfect measure of a concept, the researcher should consider several alternatives for the purpose. The use of more than one indicator gives stability to the scores and it also improves their validity.

iv. Formation of Index

This last step is combining the various indicators into an index, i.e. formation of an index. When we have several dimensions of a concept or different measurement of a dimension, we may need to combine them into a single index. One simple way for getting an overall index is to provide scale values to the responses and sum up the corresponding scores. Such an overall index would provide a better measurement tool than a single indicator because of the fact that an "individual indicator has only a probability relation to what we really want to know". This way we must obtain an overall index for the various concepts concerning the research study.

SELF-ASSESSMENT QUESTION

i. Discuss briefly in your understanding the four techniques of developing measurement tools.

4.0 CONCLUSION

The researcher must know that a better measurement depends on successfully meeting the entire problem that cause errors explained above. He must, to the extent possible try to eliminate all possible sources of error so that the final results may not be contaminated. The researcher must also bear in mind that providing a better measurement tools will give an accurate result and also limit errors.

5.0 SUMMARY

This unit has clearly explained the source of error that can occur in agricultural measurement which you must eliminate in order to get an accurate result. It has also explained the techniques of developing measurement tools and the best indicators to be used in agricultural research.

6.0 TUTOR-MARKED ASSIGNMENT

1. Give a detailed explanation on more sources of error in agricultural research measurement apart from the ones given to you.
2. What are indicators?

7.0 REFERENCE/FURTHER READING

Cochran, W.G. & Cox, E. M. (1992). *Experimental Designs*, (2nd). New York:.John Wiley & Sons.

UNIT 3 DATA COLLECTION TECHNIQUES

CONTENTS

1.0 INTRODUCTION

Research are carried out in different fields of study for various reasons which are to discover new phenomena, provision of further insights into existing principles and build models to explain known facts about natural processes. The primary motive for experimentation is data collection.

Data collection is one of the steps in scientific research commonly referred to as the scientific or research method. In module 4 we have extensively reviewed data collection but research question of how, why and when do we collect data need to be emphasized. This basic research questions will help the students in formulating the right hypothesis to answer the research problem.

2.0 OBJECTIVE

At the end of this unit, you should be able to clearly state why, how and when we need to collect data.

3.0 MAIN CONTENT

3.1 Why Do We Collect Data?

The major steps involved are:

1. To confirm existing hypothesis either fully or partially. Since, in our previous module we have defined hypothesis as an unverified

statement. Therefore, subjecting the hypothesis to a decision test is important to decide whether to accept or reject the null hypothesis.

2. To contradict the data with reasons for the differences between or within the groups. This implies that significant variation between and within the groups could be easily quantified. Example, if the mean difference is significant, it means that the groups are different and if not significant, this implies that the groups are similar in nature.
3. To develop a new hypothesis: the hypothesis may concern some single value for the given data such as one sample sign test. To test if there will be no difference among two or more sets of data such as two-sample sign test, fisher-Irwin test and rank sum test, to test the relationship between variables such as rank correlation, Kendall's coefficient of concordance and other tests for dependence, testing concerning variation in he given data set i.e. test analogous to ANOVA such as Kruskal-Wallis test. Lastly, to test for randomness of a sample based on the theory of runs viz, one sample run test and to determine if categorical data shows dependency or if two classifications are independent viz, the chi-square test can as well be used to make comparison between theoretical populations and actual data when categories are used. As we know, hypothesis is classified as null hypothesis and alternate hypothesis. The hypotheses will include null and alternate hypotheses. The null hypothesis is made before data collection (H_0:T_1=T_2) and alternative hypothesis, which is worked out as the test statistics and compared with the table value. If the calculated is greater than the tabulated, reject the hypothesis (i.e. H_a: $T_1 \neq T_2$).

3.2 How Do We Collect Data?

We collect data for primary and secondary motives. The primary sources range from direct or indirect observations by the experimenter. The experimenter collects the data himself but in the instance where delegation is unavoidable, it must be to someone who knows the details of the experimental plan. The secondary sources are derived from already tabulated data. The investigator did not take part in data accumulation.

Also, when the researcher utilises secondary data, he has to look for various sources from where he can obtain them. In this case, he/she is not confronted with the problems that are usually associated with the collection of original data. Note that the researcher must be very careful in using secondary data. By the way of caution, the researcher before using secondary data must see that they possess following characteristics such as reliability of data, suitability of data, and adequacy of data.

SELF ASSESSMENT EXERCISE

Explain the reason why you need to collect data.

4.0 CONCLUSION

This unit has revealed the importance of how and why we collect data. Mostly, in agricultural science, we find that primary sources are being undertaken in several disciplines in agricultural science except in agricultural economics that mostly deal with secondary data.

5.0 SUMMARY

In this unit, you now have residual information on why and how we collect data which will enhance your knowledge of data collection.

6.0 TUTOR-MARKED ASSIGNMENT

1. Differentiate between primary and secondary data collection.
2. Confirmation of existing hypothesis is a major step in the need for data collection. Explain.

7.0 REFERENCE/FURTHER READING

Collins, C. A. & Seeney, F. M. (1999). *Statistical Experimental Design and Interpretation: An Introduction with Agricultural Examples*. New York: John Wiley & Sons..

MODULE 5 STATISTICAL THEORY

UNIT 1 SAMPLING FUNDAMENTALS IN AGRICULTURAL RESEARCH

CONTENTS

1.0 INTRODUCTION

Sampling is the process of obtaining information about an entire population before examining only a part of it. In most of the agricultural research work and surveys, the usual approach happens to make generalization or to draw inferences based on samples about the parameters of population from which the samples are taken.

2.0 OBJECTIVES

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

- state the importance of sampling
- explain some fundamental definitions of population parameters.

3.0 MAIN CONTENT

3.1 Need for Sampling in Agricultural Research

Sampling is used in agricultural practice for a variety of reasons such as:

1. Sampling can save time and money. A sample study is usually less expensive than a census study and produces results at a relatively faster speed.
2. Sampling may enable more accurate measurements. A sample study is generally conducted by trained and experienced researchers.
3. Sampling remains the only way when population contains many members.
4. Sampling remains the only choice when a test involves the destruction of the item (plant, animal) under study.
5. Sampling usually enables us to estimate the sampling errors and thus, assists in obtaining information concerning some characteristics of the population.

3.2 Some Fundamental Definitions

Before we talk about details and uses of sampling, it is appropriate that you are familiar with some fundamental definitions concerning sampling concepts and principles.

1. Universe/Population: From a statistical point of view, the term 'universe' refers to the total of the items or units in agriculture or any field of inquiry, whereas the term 'population' refers to the total of items about which information is desired. The attributes that are the object of study are referred to as characteristics and the units possessing them are called elementary units. The aggregate of such units is generally described as population. Thus, all units in agriculture or field of inquiry constitute universe and all elementary units (on the basis of one characteristic or more) constitute population. Often, we do not find any difference between population and universe and as such the two terms are taken as interchangeable. However, you must necessarily define these terms precisely.

The population or universe can be finite or infinite. The population is said to be finite if it consist of a fixed number of item (plant or animal) so that it is possible to enumerate it in its totality e.g., the population of dairy cows in a WAMCO farm. The symbol 'N' is generally used to indicate how many elements (or items) are there in case of a finite population. An infinite population is that population in which it is theoretically impossible to observe all the items. That is, we cannot have an idea about the total number of items. You should remember that no infinite population of physical objects does actually exist in spite of the fact that many of such populations appear to be very large. From a practical consideration, you then use the term infinite population for a population that cannot be enumerated in a reasonable period of time.

SELF ASSESSMENT EXERCISE

i. Differentiate between universe and population.
ii. Give more examples on finite and infinite population using animal science research as a case study.

2. Sampling frame

The elementary units or the group or cluster of such units may form the basis of sampling process in which case they are called sampling units. A list containing all such sampling units is known as sampling frame. Thus sampling frame consists of a list of items from which the sample is to be drawn. If the population is finite and item frame is in the present or past, then it is possible for the frame to be identical with the population. As such this frame is either constructed by a researcher for the purpose of his study or may consist of some existing list of the population.

3. Sampling design

A sample design is a definite plan for obtaining a sample from the sampling frame. It refers to the technique or the procedure the researcher would adopt in selecting some sampling units from which inferences about the population is drawn. Sampling design is determined before data are collected.

4. Statistic(s) and parameters

A statistic is a characteristic of a sample, where a parameter is a characteristic of a population. Thus, when we work out certain measures such as mean, median, mode or the like ones from samples, then they are called statistic(s) because they describe the characteristics of a sample. But when such measures describe the characteristics of population, they are known as parameter(s). For instance, the population mean (μ) is a parameter, whereas the sample mean ($\bar{x}$) is statistic. To obtain the estimate of a parameter from a statistic constitutes the prime objective of sampling analysis.

5. Sampling error

Sample surveys do imply the study of a small portion of the population and as such there would naturally be a certain amount of inaccuracy in the information collected. This inaccuracy may be termed sampling error or error variance. In other words, sampling errors are those errors which arise on account of sampling and they generally happen to be random variations (in case of random sampling) in the sample estimates around the true population values. Sampling error is inversely related to size of the sample i.e. sampling error decreases as the sample size increases and vice-versa. A measure of the random sampling error can be calculated for a given sample design and size and this measure is often called the precision of the sampling plan. Sampling error is usually worked out as the product of the critical value at a certain level of significance and the standard error.

As opposed to sampling errors, we may have non sampling errors which may creep in during the process of collecting actual information and such errors occur in all agricultural surveys whether census or sample.

6. Precision

Precision is the range within which the population average (or other parameter) will lie in accordance with the reliability specified in the confidence level as a percentage of the estimate $\pm$ or as a numerical quantity. For instance, if the estimate is Rs 4000 and the precision desired is $\pm$4%, then the true value will be no less than Rs 3840 and no more than Rs 4160. This is the range (Rs 3840 to Rs 4160) within which the true answer should lie. But if desired that the estimate should not deviate from the actual value by more than Rs 200 in either direction, in that case, the range would be Rs 3800 to Rs 4200.

Illustration

$$Actual value = 4000 - 200 = 3800 - lower \lim it$$
$$Actual value = 4000 + 200 = 4200 - upper \lim it$$

7. Confidence level and significance level

The confidence level or reliability is the expected percentage of times that the actual value will fall within the stated precision limits. Precision is the range within which the answer may vary and still be acceptable; confidence level indicates the likelihood that the answer will fall within that range, and the significance level indicates the likelihood that the answer will fall outside that range. You can always remember that if the confidence level is 95%, then the significance level will be (100-95) i.e. 5%; if the confidence level is 99%, the significance level is (100-99) i.e., 1% and so on.

8. Sampling distribution

We are often concerned with sampling distribution in sampling analysis. If we take certain number of samples and for each sample compute various statistical measures such as mean, standard deviation, etc then we can find that each sample may give its own value for statistic under consideration. All such values of a particular statistic, say mean, together with their relative frequencies will constitute the sampling distribution of the particular statistic, say mean. Accordingly, we can have sampling distribution of mean, or the sampling distribution of standard deviation or the sampling distribution of any other statistical measure. It may be noted that each item in a sampling distribution is a particular statistic of a sample. The sampling distribution tends to be quite close to the normal distribution if the number of samples is large. The significance of sampling distribution follows from the fact that the mean of a sampling distribution is the same as the mean of the universe. Thus, the mean of sampling distribution can be taken as the mean of the universe.

SELF-ASSESSMENT EXERCISE

i. What is the difference between confidence level and significance level?
ii. What is sampling distribution?

4.0 CONCLUSION

In conclusion, you must bear in mind that sample should be truly representative of population characteristics without any bias so that it may give a valid and reliable results. You should be familiar with fundamentals in sampling.

5.0 SUMMARY

In this unit, you have learnt the important of sampling and fundamentals or key words in sampling and their applications.

6.0 TUTOR-MARKED ASSIGNMENT

1. From your understanding, give the importance of sampling in agricultural research.
2. What do you understand by non-sampling error?
3. Explain the term ‘population’ with examples in its finite and infinite forms.
4. Differentiate between statistics and parameter.

7.0 REFERENCE/FURTHER READING

Keehi, R.O. (1999). *Design of Experiment: Statistical Principles of Research Design and Analysis.* New York: Dexbury Press,

UNIT 2 SAMPLING DISTRIBUTION

CONTENTS

1.0 INTRODUCTION

Sampling distribution is the process that involves probability of all possible means, proportions, t, f and chi-square distribution of all random samples of a given size that is taken from a population. This gives insight into the behaviour of the population parameters you are working with whether they are normally distributed or not.

2.0 OBJECTIVE

At the end of this unit, you should be able to discuss the application of different sampling distribution.

3.0 MAIN CONTENT

3.1 Importance of Sampling Distributions in Agricultural Research

Some important sampling distributions are as follows:

1. **Sampling distribution of Mean**
   Sampling distribution of mean refers to the probability distribution of all the possible means of random samples of a given size that we take from a population. If samples are taken from a normal population, N (μ σ_p), the sampling distribution of mean would be normal with mean $\sigma_{\bar{X}} = \mu$ and standard deviation =$\sigma_p \sqrt{n}$, where μ is the mean of the population, σ_p is the standard deviation of the

population and n means the number of items in a sample. But when sampling is from a population which is not normal (may be positively or negatively skewed), even then, as per the central limit theorem, the sampling distribution of mean tends quite closer to the normal distribution, provided the number of sample items is larger i.e. more than 30. In case you want to reduce the sampling distribution of mean to unit normal distribution i.e. N (0, 1), we can write the normal variant as:

$$z = \frac{\bar{x} - \mu}{\left(\sigma_p / \sqrt{n}\right)}$$

for the sampling distribution of mean. This characteristic of the sampling distribution of mean is very useful in several decision situations for accepting or rejection of hypothesis.

2. Sampling Distribution of Proportion

Like sampling distribution of mean, we can as well have a sampling distribution of proportion. This happens in case of statistics of attributes. Assume that we have worked out the proportion of defective parts in large number of samples, each with say 100 items (animals), that have been taken from an infinite population and plot a probability distribution of the said proportions, we obtain what is known as the sampling distribution of proportion. Usually the statistics of attributes correspond to the conditions of a binomial distribution that tends to become normal distribution as n becomes larger and larger. If p represents the proportion of defectives i.e. of successes and q of the proportion of non-defectives i.e. of failures (or q = 1-p) and if p is treated as a random variable, then the sampling distribution of proportion of successes has a mean = p with standard deviation.

$$= \sqrt{\frac{p.q}{n}}$$

Where n is the sample size. Presuming the binomial distribution approximates the normal distribution for large n, the normal variant of the sampling distribution of proportion Z= $\frac{\hat{p}-p}{\sqrt{(p.q)/n}}$, Where $\hat{p}$ (pronounced as p-hat) is the sample proportion of successes, can be used for testing of hypothesis.

3. Student's t-distribution

When population standard deviation (σ_p) is not known and the sample is of a small size (i.e. n $\leq$ 30), we use t distribution for the sampling distribution of mean and work out t variable as:

$$t = \frac{\bar{x} - \mu}{\sigma_{\bar{x}}}$$

where

$$\sigma_s = \sqrt{\frac{\Sigma(x_i - \bar{x})^2}{n}} - 1$$

i.e. the sample standard deviation, t-distribution is also symmetrical and is very close to the distribution of standard normal variant, z, except for small values of n. The variable t differs from z in the sense that we use sample standard deviation (σ_s) in the calculation of z. there is a different t distribution for the degrees of freedom for a sample of size n is n-1. A_s becomes approximately equal to the normal distribution. In fact for sample sizes of more than 30, the t distribution is so close to the normal distribution that we can use the normal to approximate the t-distribution.

But when n is small, the t-distribution is far from normal but when, t-distribution is identical with normal distribution. The t-distribution tables are available which give the critical values of t for different degrees of freedom at various levels of significance. The table value at a certain level of significance is compared with the calculated value of t from the sample data, and if the latter is either equal to or exceeds, we infer the null hypothesis cannot be accepted.

4. F distribution

If $(\sigma_{s1})^2$ and $(\sigma_{s2})^2$ are the variances of two independent samples of size n_1 and n_2 respectively taken from two independent normal population, having the same variance, $(\sigma_{p1})^2 = (\sigma_{p2})^2$, the ratio $F = \frac{(\sigma_{p1})^2}{(\sigma_{s2})^2}$ where $(\sigma_{s1})^2 = \sum \frac{(\bar{x}1i - \bar{x}\,1)^2}{\text{n1}} - 1$ has an F distribution with n_1-1 and n_2-1 degrees of freedom. F ratio is computed in a way that the larger variance is always in the numerator. Tables have been prepared for F distribution that gives critical values of F for various values of degrees of freedom for larger as well as smaller variances. The calculated value of F from sample data is compared with the corresponding table value of F and if the former exceeds the later, then we infer that the null hypothesis of the variances being equal cannot be accepted. We shall make use of the F ratio in the context of hypothesis testing and also in the context of ANOVA technique.

5. Chi-square (X^2) distribution

Chi-square distribution is encountered when we deal with collections of values that involve adding up squares. Variances of samples require us to add a collection of squared quantities and thus have distributions that are related to chi-square distribution. If we take each one of a collection of sample variances, divide them by the known population variance and multiply these quotients by (n-1), where n means the number of items in the sample, we shall obtain a chi-square distribution. Thus, $\left(\sigma_s^2 / \sigma_p^2 \right)$ would have the same distribution as chi-square distribution with (n-1) degrees of freedom. Chi-square distribution is not symmetrical and all the values are positive. One must know the degrees of freedom for using chi-square distribution. This distribution may also be used for judging the significance of difference between observed distributions depending upon the d.f. and χ^2 value is worked out as shown below:

Where;

$$\chi^2 = \frac{O_{ij} - E_{ij}}{E_{ij}}$$

O_{ij} = observed frequency of the cell in the i^{th} row and j^{th} column

E_{ij} = expected frequency of the cell in the i^{th} row and j^{th} column

Tables are there that give the value of χ^2 for a given d.f. which may be used with calculated value of χ^2 for relevant d.f. at a desired level of significance for testing hypothesis.

SELF-ASSESSMENT EXERCISE

Explain the student's t-distribution for large and small sample size with formulae.

4.0 CONCLUSION

You must have observed that for every sampling distribution there is a table of distribution for each and can be used for testing hypothesis either for accepting or rejecting the hypothesis.

5.0 SUMMARY

In this unit, you have learnt some important sampling distribution and their application for small or large population.

6.0 TUTOR-MARKED ASSIGNMENT

Explain different forms of sampling distributions in details.

7.0 REFERENCE/FURTHER READING

Leonard, S. & Anselm, L. (1973). *Field Research*. New Jersey: Prentice Hall Inc.

UNIT 3 CENTRAL LIMIT AND SAMPLING THEORY

CONTENTS

1.0 INTRODUCTION

In this unit, you should know that sampling theory is only applicable to random sample. It is also a study of relationship between population and samples drawn from the population when sampling is from a normal population, the means of samples drawn from such a population are themselves normally distributed. But when sampling is not from a normal population, the size of the sample plays a critical role. The significance of statistical theory is that it allows us to use sample statistics to make inferences about population parameters without the shape of the frequency distribution.

2.0 OBJECTIVES

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

- delineate between sampling and central limit theorem
- list and explain different types of sampling distribution.

3.0 MAIN CONTENT

3.1 Central Limit Theorem

When n is small, the shape of the distribution will depend largely on the shape of the parent population, but as n gets large (n> 30), the shape of the sampling distribution will become more like a normal distribution, irrespective of the shape of the parent population. The theorem which explains this sort of relationship is known as the central limit theorem. This theorem is by far the most important theorem in statistical inference. It assures that the sampling distribution of the mean approaches normal distribution as the sample size increases. In formal terms, we may say that the central limit theorem states that "the distribution of means of random samples taken from a population having mean μ and finite variance σ^2 approaches the normal distribution with Mean μ and variance $\frac{\sigma^2}{n}$ as n goes to infinity".

"The significance of the central limit theorem lies in the fact that it permits us to use sample statistics to make inferences about population parameters without knowing anything about the shape of the frequency distribution of that population other than what we can get from the sample".

SELF-ASSESSMENT EXERCISE
State the central limit theorem.

3.2 Sampling Theory

Sampling theory is the study of relationship existing between a population and samples drawn from the population. Sampling theory is applicable only to random samples. For this purpose the population or a universe may be defined as an aggregate of items about which knowledge is sought. The universe may be hypothetical or existent. In the former case the universe in fact does not exist and we can only imagine the items constituting it. Tossing a coin or throwing a dice is an example of hypothetical universe. Existent universe is a universe of concrete objects i.e. the universe where the items constituting it really exist. Sampling theory is designed to attain one or more of the following objectives.

i. **Statistical estimation**
Sampling theory helps in estimating unknown population parameters from the knowledge of statistical measures based on sample of parameter from statistic is the main objectives of the sampling theory. The estimate can either be a point estimate or it may be an interval estimate. Point estimate is a single estimate expressed in the form of a single figure, but interval estimate has two limits viz, the upper limit and the lower limit within which the parameter value may lie. Interval estimates are often used in statistical induction.

ii. Testing of hypothesis

The second objective of sampling theory is to enable us to decide whether to accept or reject hypothesis; the sampling theory helps in determining whether observed differences are actually due to chance or whether they are really significant.

iii. Statistical inference

Sampling theory in making generalization about the population/universe from the studies based on samples from it. It also helps in determining the accuracy of such generalization.

The theory of sampling can be studied under two heads viz, the sampling of attributes and the sampling of variables and that too in the context of large and small samples. When we study some qualitative characteristics of the items in a population, we obtain statistics of attributes in the form of two attribute, present and the other class consisting of items wherein the attribute is absent.

i. The parameter value may be given and it is only to be tested if an observed ‘statistics’ is its estimate.
ii. The parameter value is not known and we have to estimate it from the sample.
iii. Examination of the reliability of the estimate i.e. the problem of finding out how far the estimate is expected to deviate from the true value for the population.

All the above stated problems are studied using the appropriate standard errors and the tests of significance. The theory of sampling can be applied in the context of statistics of variables (i.e. data relating to some characteristics concerning population which can be measured or enumerated with the help of some well-defined statistical unit in which case the objective happens to be:

i. To compare the observed and expected values and to find if the difference can be ascribed to the fluctuations of sampling.
ii. To estimate population parameters from the sample.
iii. To find out the degree of reliability of the estimate

The sampling theory for large samples is not applicable in some samples because when samples are small, we cannot assume that the sampling distribution is approximately normal. As such we require a new technique for handling small samples, particularly when population parameters are unknown. Sir William S. Gosset (Pen name student) developed a significance test, known as student's t-test, based on t distribution and through it made significant contribution in the theory of sampling applicable in case of small samples. Student's t-test is used when two conditions are fulfilled viz, the sample size is 30 or less and the population variance is not known.

While using t-test we assume that the population for which sample has been taken is normal or approximately normal, sample is a random sample, observation is independent, there is no measurement error and that in the case of two samples when equality of the two populations means is to be tested, we assume that the population variances are equal. For applying t-test, we work out the value of test statistics (i.e. 't') and then compare with the table value of t (based on 't' distribution) at certain level of significance for given degree of freedom. If the calculated value of't' exceeds the table value, we infer that the difference is significant, but if calculated value of t is less than or equal to the concerning table value of it, the difference is not treated as significant. The following are formulae commonly used to calculate the t value

i. To test the significance of the mean of a random sample

$t = \frac{\bar{x}-\mu}{\sigma_{\bar{x}}}$

where $\bar{x}$ = mean of the sample

μ= mean of the population

$\sigma_{\bar{x}}$ = standard error of mean worked out as under

$$\sigma_{\bar{x}} = \frac{\sigma_s}{\sqrt{n}} = \sqrt{\frac{\sum(x_i - \bar{x})^2}{n-1}} \Big/ \sqrt{n}$$

and the degrees of freedom = (n-1)

ii. To test the difference between the means of two samples

$$t = \frac{\bar{x}_1 - \bar{x}_2}{\sigma_{\overline{x_1}} - \bar{x}_2}$$

where

$\bar{x}1 = mean\ of\ sample\ one$

$\bar{x}2$ = mean of sample two

$\sigma_{\bar{x}1-\bar{x}2}$

$= standard\ error\ of\ difference\ between\ two\ samples\ mean\ worked$

$$\sigma_{\bar{x}_1} - \bar{x}_2 = \sqrt{\frac{\sum(x_{1i} - \bar{x}_1)^2 + \sum(x_{2i} - \bar{x}_2)^2}{n_1 + n_2 - 2}} \times \sqrt{\frac{1}{n_1} + \frac{1}{n_1}}$$

And the d.f $= n_1 + n_2 - 2$

iii. To test the significance of the coefficient of simple correlation

t = $\frac{r}{\sqrt{1-r^2}}\ X\ \sqrt{n-1}\ or\ t =\ {}^{r}\sqrt{\frac{n-1}{1-r^2}}$

where r = the coefficient of simple correlation and the d.f. = (n-2)

iv. To test the significance of the coefficient of partial correlation

$$t = \frac{1}{\sqrt{1-r_p^2}} \times \sqrt{n-k}\ \ or\ t = r_p\sqrt{\frac{n-k}{1-r_p^2}}$$

where p is any partial coefficient of correlation and the d.f. = (n-k), n being the number of pairs of observations and k being the number of variables involved.

v. To test the difference in case of paired or correlated samples data (in which case t test is often described as difference test)

$$t = \frac{\bar{D}-\mu_D}{\sigma_D}\sqrt{n}\ or\ t = \frac{\bar{D}-0}{\sigma_D}\sqrt{n}$$

where

Hypothesized mean differences (μ_D) is taken as zero (0)

$\bar{D}$ = mean of the differences of correlated sample items

σ_D = standard deviation of differences worked out as under

$$\sigma_D = \frac{\sum D_i^2 - \frac{\bar{D}}{n}}{n-1}$$

$$Di = Differences\ (i.e. Di = (Xi - Yi)$$

n= number of pairs in two samples and the d.f. = (n-1)

n = number of pairs in two samples and the d.f. = (n-1)

SELF ASSESSMENT EXERCISE

State the three (3) types of problems in case of sampling of attributes.

3.3 Concept of Standard Error

The utility of the concept of standard error in statistical induction is on the account of the following reasons:

1. The standard error helps in testing whether the difference between observed and expected frequencies could arise due to chance. Thus, the standard error is an important measure in significance tests or in examining hypotheses. If the estimated parameter differs from the calculated statistics by more than 1.96 times the S.E., the difference is taken as significant at 5 per cent level of significance. This, in

other words, means that the difference is outside the limits i.e. it lays in the 5 per cent area (2.5 per cent on both sides) outside the 95 per cent area of the sampling distribution.

2. The standard error gives an idea about the reliability and precision of a sample. The smaller the standard error the greater the uniformity of sampling distribution and hence, greater is the reliability of sample. Conversely, the greater the standard error greater is the difference between observed and expected frequencies. In such a situation the unreliability of the sample is greater.

   The size of standard error depends upon the sample size to a great extent and it varies inversely with the size of the sample. If double reliability is required in other words reducing standard error to ½ of its existing magnitude, the sample size should be increased four-field.

3. The standard error enables us to specify the limits within which the parameters of the population are expected to lie with a specified degree of confidence. Such an interval is usually known as confidence interval.

   Important feature for computing the standard errors concerning various measures based on samples are as indicated below:

a. In case of sampling of attributes:

i. Standard error of number of successes = $\sqrt{n.p.q}$

   where n = number of events in each sample
   p = probability of success in each event,
   q = probability of failure in each event

ii. Standard error of proportion of successes $\sqrt{\frac{(p.q)}{n}}$

iii. Standard error of the difference between proportions of two samples.

$$\sigma_{P1} - P_2 = \sqrt{pq\ \left(\frac{1}{n_1} + \frac{1}{n_2}\right)}$$

Where p= best estimate of proportion in the population and is worked out as under:

$$P = \frac{n_1P_1 + n_2P_2}{n_1 + n_2}$$

q = 1-p

n_1= number of events in sample one

n_2 = number of events in sample two

Note: Instead of the above formula, we use the following formula:

$$\sigma_{P1} - P_2 = \sqrt{\frac{P_1 q_1}{n_1} + \frac{P_2 q_2}{n_2}}$$

When samples are drawn from two heterogeneous population where we cannot have the best estimate of proportion in the universe on the basis of given sample data. Such a situation often arises in study of association of attributes.

b. In case of sampling of variable (large samples)

i. Stand error of mean when population standard deviation is known

$$\sigma_{\bar{x}} = \frac{\sigma_P}{\sqrt{n}}$$

where

σ_p = standard deviation of population

n = number of items in the sample

Note: This formula is used when n is 30 or less

ii. Standard error of mean when population standard deviation is unknown:

$$\sigma_{\bar{x}} = \frac{\sigma_s}{\sqrt{n}}$$

where

σ_s = standard deviation of the sample and is worked out as

$$\sigma_s = \sqrt{\frac{\sum(x_i - \bar{x})^2}{n-1}}$$

iii. Standard error of standard deviation when population standard deviation is known:

$$\sigma_s = \frac{\sigma_P}{\sqrt{2n}}$$

iv. Standard error of standard deviation when population standard deviation is unknown:

$$\sigma_s = \frac{\sigma_s}{\sqrt{2n}}$$

Where

$$\sigma_s = \sqrt{\frac{\sum(x_i - \bar{x})^2}{n-1}}$$

n = number of items in the sample

v. Standard error of the coefficient of simple correlation

$$\sigma_r = \frac{1 - r^2}{\sqrt{2n}}$$

Where r = coefficient of simple correlation
n= number of items in the sample

vi. Standard error of difference between means of two samples:

a. When two samples are drawn from the same population

$$\sigma_{\bar{x}1} - \bar{x}_2 = \sqrt{\left(\frac{(\sigma_{P1})^2}{n_1} - \frac{(\sigma_{P2})^2}{n_2}\right)}$$

(If σ_{p1} *and* σ_{p2} are not known, then in their places σ_{s1} *and* σ_{s2} respectively may be substituted)

c. In case of sampling of variables (small samples):

i. Standard error of mean when σ_p is unknown"

$$\sigma_{s1.2} = \sqrt{\frac{n_1(\sigma_{s1})^2 + n_2(\sigma_{s2})^2 + n_1(\bar{x}_1 - \bar{x}_{1.2})^2 + n_2(\bar{x}_2 - \bar{x}_{1.2})^2}{n_1 + n_2}}$$

Where

$$\bar{x}_{1.2} = \frac{n_1(\bar{x}_1)^2 + n_2(\bar{x}_2)^2}{n_1 + n_2}$$

Note: All these formulae apply in case of infinite population. But in case of finite population where sampling is done without replacement and the sample is more than 5% of the population, we must as well use the finite population multiplier in our standard error formulae. For instance, S.E. $\bar{x}$ in case of finite population will be as under:

$$SE_{\bar{x}} = \frac{\sigma_P}{\sqrt{n}} . \sqrt{\frac{N - n}{N - 1}}$$

It may be remembered that in cases in which the population is very large in relation to the size of the sample, the finite population multiplier is close to one and has little effect on the calculation of S.E. As such when sampling fraction is less than 0.5, the finite population multiplier is generally not used.

$$\sigma_{\bar{x}} = \frac{\sigma_S}{\sqrt{n}} . \frac{\sqrt{\frac{\sum(x_i - \bar{x})^2}{n - 1}}}{\sqrt{n}}$$

ii. Standard error or difference between two sample means when σ_p is unknown

$$\sigma_{\bar{x}_1} - \bar{x}_2 = \frac{n_1(\bar{x}_1 - \bar{x}_{1.2})^2 + n_2(\bar{x}_2 - \bar{x}_{1.2})^2}{n_1 + n_2 - 2} . \sqrt{\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}$$

SELF ASSESSMENT EXERCISE

Explain one of the reasons for utility of standard error in statistical inductions.

4.0 CONCLUSION

In statistical theories, you must have observed that the central limit theorem in statistical inference assures that as the sampling distribution of the mean approaches normal distribution the sample size increases. While, sampling theory is only applicable to random samples, standard error is considered the key to sampling theory.

5.0 SUMMARY

In this unit, you have learnt about statistical theories (central limit theorem, sampling theory and standard error), their importance in statistical induction and their application in agricultural research.

6.0 TUTOR-MARKED ASSIGNMENT

1. What is the significance of the central limit theorem?
2. Define the central limit theorem.
3. What is the difference between hypothetical and existent universe?
4. What are the objectives of sampling theory?

7.0 REFERENCE/FURTHER READING

Kothori, C.R. (1984). *Quantitative Techniques*. (2nd), New Delhi: Vikas Publishing House ltd.

MODULE 6 STATISTICAL METHODS

UNIT 1 GUIDELINES AND EXPLANATIONS OF STATISTICAL METHODS

CONTENTS

1.0 INTRODUCTION

Statistical methods are regarded as taskforce of statistical inference whose charge was to elucidate or map out designs for solving statistical problems such as significance testing and its alternatives. It is widely known that methods appropriate in one area may be in appropriate in another area. Choosing the right design is a key to accurate results.

2.0 OBJECTIVES

At the end of this unit, you should be able to;

- discuss the characteristics of different statistical methods
- explain Guidelines for application of the different statistical methods.

3.0 MAIN CONTENT

3.1 Basic Procedures of Statistical Methods

The under listed factors are the principal components of statistical methods

i. Design
ii. Population
iii. Sample
iv. Assignment
v. Measurement

Design

In setting out your layout, ensure you make clear at the outset what type of study you are doing. Do not cloak a study in one guise to try to give it the assumed reputation of another. For studies that have multiple goals, be sure to define and prioritize those goals. There are many forms of empirical studies in agricultural science, including case reports, controlled experiments, quasi-experiments, statistical simulations, surveys, observational studies and studies of studies (meta-analyses). Some are hypothesis testing. It is always assumed by researchers that some of the forms yield information that is more valuable or credible than others. Occasionally proponents of some form of research have its own strength, weakness and standards of practice.

Population

The interpretation of the results of any study depends on characteristics of the population intended for analysis. Define the population (participants, stimuli or studies) clearly. If control or comparison groups are part of the design, you need to present how they are defined. Animal science students sometimes think that a statistical population is the animal race. They also have difficulty distinguishing a class of objects versus a statistical population. Populations may be sets of potential observation on people, animals, objects or even research articles.

Sample

In sampling you have to describe the procedures and emphasise any inclusion or exclusion criteria. If the sample is stratified (e.g. by location or sex) describe fully the method and rationale and note the proposed sample size for each subgroup. Interval estimates for clustered and stratified random samples differ from those for simple random samples. Statistical software such as statistical package for social science (SPSS), general statistics (GENSTAT) and statistical analysis system (SAS) are now becoming available for these purposes. If you are using a convenience sample whose member are not selected at random, be sure to make the procedures clearer to your audience. Using a convenience sample in animal science does not automatically disqualify a study from publication, but it harms your objectivity to try to conceal this by implying that you used a random sample. Sometimes convenience sampling could be strengthened by explicitly comparing sample characteristics with those of defined population across a wide range of variables. Sample designs could be classified based on non-probability sampling and probability sampling. Non probability sampling is that sampling procedures which does not afford any basis for estimation of probability that each item in the population has of being included in the sample. In this type of sampling, the researchers of the inquiry purposively choose the particular unit of the universe for constituting a sample on the basis that the small mass that they so select out of huge one will be typical or representative of the whole; for instance, if the milk yield of Friesian cattle in Nigeria is to be purposively selected for intensive study on the principle that they can be representative, also known as random sampling or chance sampling under this sampling, every Friesian cattle in Nigeria has an equal chance of inclusion in the sample. Random sampling ensures the law of statistical regularity which states that if on the average the sample chosen is a random one, the sample will have the same composition and characteristics as the universe. This is the reason why random sampling is considered as the best technique of selecting a representative sample.

Assignment

This could be classified as random and nonrandom assignment. Random assignment should not be confused with random selection because it allows for the strongest possible causal inferences free of extraneous assumptions. It is a sensitive and robust method of controlling bias and lurking of variables. Random does not mean "haphazard". Randomization is a fragile condition, easily corrupted deliberately, as we see when a skilled magician flips a fair coin repeatedly to heads. In the field of agriculture, self-randomization is incapable of producing a random process; it is best not to trust the random behavior of a physical device unless you are an expert in the matter. It is safer to use the pseudorandom sequence from a well-designed computer generator or from published tables of random numbers.

In the case of non-random assignment, for some research questions, random assignment is not feasible. In such cases, we need to minimize effects of variables that affect the observed relationship between a causal variable and an outcome. Such variables are commonly called confounds or covariates. The researcher needs to attempt to determine the relevant covariates, measure them adequately and adjust for their effects either by design or by analysis. Example, in animal nutrition experiments, during the feeding trial, we always try to randomly balance the initial weight of the animals through averaging the treatment effect so that there will be similarity within the treatments or we use initial weight as a covariate in our model, if peradventure there is significant difference among the dietary treatments.

Measurement Variables

Defining the variables in the study, show how they are related to the goals of the study, and explain how they are measured. A variable is a method for assigning to a set of observations a value from a set of possible outcomes. If we declare the range of an association between body measurements to be positive real numbers and the domain of the observations is 1000, then, a value of 1000 is not illegal, it is an outlier.

Instrument

It is the most frequently used term in questionnaire form of data collection. It also summarizes the biometric properties of its scores with specific regard to the way the instrument is used in a population. Biometric properties include measures of validity, reliability and any other qualities affecting conclusion. If a physical apparatus is used, provide enough information (brand, model and design specifications) to allow another experimenter to replicate your measurement process. There are many methods for constructing instruments and biometrically validating scores from such measures. Traditional true score theory and item-response test theory provide appropriate frameworks for assessing reliability and internal validity. Signal detection theory and various coefficient of association can be used to assess external validity. The instrument needs to be tested for reliability. Reliability is a property of the scores on a test for a particular population. Also, another rule of thumb is that the instrument does not correlate with other key constructs. It is just as important to establish that a measure does not measure what it should not measure as it is to show that it does measure what it should.

Procedures

The major steps or guidelines in statistical methods are listed below.

i. Describe any anticipated sources of attrition due to noncompliance, drop out, death or other factors. This will indicate how such attrition may affect the generalizability of the results.
ii. Clearly describe the conditions under which measurement are taken (e.g. animal id, date of birth, sex, personnel who collected the data)

iii. Describe the specific methods used to deal with experimental bias, especially if you collected the data yourself.

Power and Sample Size

This provides information on sample size and the process that led to sample size decisions. Document the effect of size, sampling and measurement assumptions, as well as analytical procedures used in power calculations, because power computations are most meaningful when done before data are collected and examined. It is also important to show how effect –seize estimates have been derived from previous research and theory in order to dispel suspicions that they might have been taken from data used in the study constructed to justify a particular sample size. Once the study is analysed, confidence intervals replace calculated power in describing results.

SELF-ASSESSMENT EXERCISE

Mention the major steps in constructing the guidelines for statistical methods.

4.0 CONCLUSION

In this unit, we have looked at the factors that are principal in statistical methods for effective research delivery in research methodology. These factors provide a template in construction of state of the art research.

5.0 SUMMARY

This unit provided a synopsis for research layout, sampling selection and statistical framework for effective research delivery.

6.0 TUTOR-MARKED ASSIGNMENT

1. Briefly explain statistical method in experimental design.
2. Differentiate between probabilistic and non-probabilistic sampling.

7.0 REFERENCE/FURTHER READING

Lastrucci, C. (1967). *The Scientific Approach; Basic Principle of the Scientific Method*. Cambridge, Mass; Schentam Publishing Co. Inc.

UNIT 2 APPROACHES TO DATA COMPILATIONS AND ANALYSIS

CONTENTS

1.0 INTRODUCTION

This unit introduces you to data compilation/computations and analysis. Before presenting your results, there is a need to check if there is no missing data, attrition and non-response. Also checking for further techniques to ensure that the data analysed are not produced by anomalies in the data (e.g., outliers, points of high influence, nonrandom missing data, selection bias and attrition problems). This manipulation should be a standard component of all analyses.

2.0 Objectives

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

- screen and clean your data before analysis
- choose the right statistical tool and computer program for data analysis.

3.0 MAIN CONTENT

3.1 Data Computations and Analysis

The enormous variety of modern quantitative methods in agricultural science leaves researchers with the nontrivial task of matching analysis and design to the research question. It should be noted that the underlying assumptions required for the analysis and reasonable given the data.

Examine residuals carefully. Do not use distributional tests and statistical indexes of shape (e.g. skewness, kurtosis) as a substitute for examining your residuals graphically. Using a statistical test to diagnose problems in model fitting has several shortcomings. First, diagnostic significance tests based on summary statistic (such as tests for homogeneity of variance) are often impractically sensitive; our statistical tests of models are often more robust than our statistical test of assumptions. Secondly, statistics such as skewness and kurtosis often fail to detect distributional irregularities in the residuals. Lastly, statistical tests depend on sample size, and as sample size increases the tests often will reject innocuous assumptions. In general, there is no substitute for graphical analysis of assumption. Modern statistical packages SPSS, SAS, GENSTAT and R offer graphical diagnostics for helping to determine whether a model appear to fit data appropriately. Most users are familiar with residual plots for linear regression modeling. Fewer analyst are aware that John Tukey's paradigmatic equation, *data = fit + residual*, applies to a more general class of models and has broad implications for graphical analysis of assumptions. Stem and leaf plots, box plots, histograms, dot plots, auto and cross correlation plots all serve at various time for displaying residuals, whether they arise from analysis of variance (ANOVA), nonlinear modeling, factor analysis, latent variable modeling, multidimensional scaling, hierarchical linear modeling, or other procedures.

3.2 How to Analyze Data Using SPSS, SAS and JMP

Most of the new computer software has an inbuilt graphical user interface which make data analysis simpler through "point and click" mode. Data collected from the field has to be fed into a spreadsheet for data manipulation and cleaning. Cleaned data are now imported to the statistical package interface.

In the computer or statistical package interface there are a lot of inbuilt menu depending on the software you are using for your analysis. IBM SPSS version 20, several icons such as file, edit, view, data, transform, analyze, direct marketing, graphs, utilities, add-ons, window and help menu exist. In SAS, icons such as file, edit, view, tools, run, solutions, window and help menu exist. In JMP, you find icons such as file, tables, design of experiment, analyse, graph, tools, view, window and help-menu. Click on the yellow folder or envelope to import your data into the interface. In SAS interface, click solutions- analysis –analyst-statistics-then you choose the appropriate statistic based on your design (where you can define your variables as dependent, independent, fixed or random) - lastly you click on run to display your statistical output. In SPSS menu, click analyze, the interface will display different kinds of statistical designs, then click the design that fit into your experimental layout to define your criteria for analysis such as your independent, dependent, covariates, fixed and random variables. Lastly you will click ok to print out your statistical output.

In JMP statistical package, after importation of data into the analysis interface, click analyse and choose the type of design for your variable selection and definition. Lastly click ok.

SELF-ASSESSMENT EXERCISE

List the modern statistical packages that are mostly used in the field of agriculture.

4.0 CONCLUSION

In this unit, you have learnt how to compute and analyse data in different statistical software.

5.0 SUMMARY

This unit revealed the major steps in data manipulation and analysis using different sophisticated statistical software.

6.0 TUTOR-MARKED ASSIGNMENT

1. Explain the major steps in testing for outliers.
2. Write the full meaning of SPSS, SAS and GENESTAT.

7.0 REFERENCE/FURTHER READING

Hoshmand, R. (1994). *Experimental Research Design and Analysis. A Practical Approach for the Agricultural and Natural Sciences*. London: CRC Press.

UNIT 3 COMPUTATIONAL TECHNIQUES OF EXPERIMENTAL DESIGN

CONTENTS

1.0 INTRODUCTION

This unit is structured to solve basic examples relating to statistical methods. Since the theoretical examples has been reflected in the previous modules. It is important to note that different statistical designs have different models as a unique characteristics or features. This will also build our skills in the computation of various designs.

2.0 OBJECTIVES

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

- explain the mathematical steps involved in the calculation of t-test,
- analysis variance (f-test), change over design, factorial and nested design.

3.0 MAIN CONTENT

3.1 t-test (When Sample Size is Less than 30)

Ten South-African Jersey dairy cows are in lactation in Sebore farms. The arithmetic mean of the sample is 4000kg and the population mean is

4200kg with standard deviation of 500kg. Test whether the sample mean is significantly different from the population mean.

Solutions

The hypothetical mean is $\mu_0 = 4200$kg

H_0: $\mu = 4200$kg

H_a: $\mu = 4200$kg

Sample mean is $\bar{y} = 4000$kg

Sample standard deviation is s = 500kg

$$S/\sqrt{n} = 500/\sqrt{10}$$

The calculated value of the t-statistics is

$$t = \frac{y - \mu_0}{s/\sqrt{n}} = \frac{4000 - 4200}{500/\sqrt{10}} = -1.26$$

To calculate the critical value $= -t_{\alpha/2} = -2.26$. Then take a decision rule. Since the calculated t=-1.26 is less than the critical value $-t_{\alpha/2} = -2.26$, Ho is not rejected with an $\alpha = 0.05$ level of significance. This implies that the sample mean is not significantly different from 4200kg.

3.1.1 z-test (When the Sample Size is more than 30)

Example: Two groups of 40 broiler chickens were fed different dietary rations (A and B) to determine the best group that will perform optimally under Zaria conditions. At the end of the finisher phase at 8 weeks of age the following sample mean and variances were calculated in kg.

	Treatment A	Treatment B
Mean ($\bar{Y}$)	3.20	4.50
Variance (S^2)	0.25	0.36
Size (n)	40	40

The hypotheses for a two-sided test are:

H_0: $\mu_1 - \mu_2 = 0$

H_1: $\mu_1 - \mu_2 \neq 0$

The pooled standard error or the estimate:

$$S_{(\bar{y}_1 - \bar{y}_2)} = \sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}} = \sqrt{\frac{0.25}{40} + \frac{0.36}{40}} = 0.12$$

The calculated value of z static is:

$z = \frac{y_1 - y_2}{S_{(y_1 - y_2)}} = \frac{3.20 - 4.50}{0.12} = -10.57$ Since the Z value = -10.57 is more than the critical value ($-z_{\propto/2} = -z_{0.025} = -1.96$), the null hypothesis is rejected suggesting that ration B was able to optimally improve the growth performance of broiler chickens.

3.2 Anova for Completely Randomised Design

Illustration: An experiment was conducted to test the nutritive value of roselle seed cake as a replacement for groundnut cake at different inclusion levels in the dietary treatments on the bodyweight of broiler chickens at starter phase.

The model:

$Y_{ij} = \mu + T_i + \varepsilon_{ij}$

Y_{ij} = Observation in the ith treatment

μ = Overall mean

T_i = The effect of ith treatment

ε_{ij} = Random error $(0, \sigma^2)$

	TR1 (0%)	TR2 (10%)	TR3 (20%)	
	370 400 380 380 370	390 350 380 390 380	390 440 330 300 300	TOTAL
$\sum$	1900	1890	1760	5550
N	5	5	5	15
$\bar{y}$	380	372	352	370

For computations of sum of squares

1) $\sum i \sum j Yij$ = (370+400+……+300)=5550

2) $correction\ for\ the\ mean$

C.F = $\frac{(\sum i \sum j Yij)^2}{N} = \frac{(5550)^2}{15} = \frac{30802500}{15}$ = 2053500

3) Total (corrected) sum of squares:
$SS_{TOT} = (\sum i \sum j Yij^2$ - C.T
= 2072700-2053500
= 19200

4) Treatment sum of square
$SS_{TRT} = \sum \frac{(\sum jYij)^2}{ni} - C$
$= \frac{1900^2}{5} + \frac{1890^2}{5} + \frac{1760^2}{5}$
= 722000 + 714420 + 619520 – C
= 2055940 – 2053500
= 2440

5) Residual sum of square:
$SS_{RES} = SS_{TOT} - SS_{TRT}$ = 19200-2440
= 16760

ANOVA Table

Source	SS	Df	ms	F
Treatment Residual Total	2440 16760 19200	3-2=2 15-3=12 14	1220 1396.67	0.87

$F = \frac{MS_{TRT}}{MS_{RES}} = \frac{1220}{1396.67} = 0.87$

The power of test is calculated as $F_{0.05,\ 2,\ 12}$ = 3.89. The calculated F statistics (0.87) is less than the critical value (3.89), supporting the conclusion that dietary treatments are similar.

3.3 Anova for Randomised Complete Block Design

Example: An experiment was conducted to test the effect of three dietary treatments on bodyweight of lactating Friesian cattle.

Treatment	I	II	III	IV	Σ	Treatment average
T1	826.00	865.02	795.00	850.00	3336.02	834.00
T2	827.05	871.99	720.97	860.00	3280.05	820.00
T3	753.00	803.99	737.00	821.98	3116.00	779.00
Block	2406.05	2540.99	2253.00	2532.00	9731.00	
Block average	802.00	847.00	751.00	844.00	811	

Illustration: RCBD Layout

Blocks

Treatment	I	II	III	IV
T_1	Y_{11}	Y_{12}	Y_{13}	Y_{14}
T_2	Y_{21}	Y_{22}	Y_{23}	Y_{24}
T_3	Y_{31}	Y_{32}	Y_{33}	Y_{34}

The model:

$Y_{ij} = \mu + T_i + B_J + \varepsilon_{ij}$

$Y_{ij} =$ Observation in the ith treatment

$\mu =$ Overall mean

$T_i =$ The effect of ith treatment

B_J = Effect of the ith block

$\varepsilon_{ij} =$ Random error $(0, \sigma^2)$

Mathematical Computations of Sum of Squares

1. Total sum of squares

$$\sum_i \sum_j Y_{1,k} = (826.00 + 865.02 + - - - 821.98)$$
$$= 9731.99$$

2. Correction for the mean

$$C.T = \frac{\left(\sum_i \sum_j Y_{ij}\right)^2}{(x)\,(y)} = \frac{(9731.99)^2}{12} = 7892636$$

3. Total (correct) sum of squares

$$SS_{TOT} = \sum_i \sum_j Y_{i,j}{}^2 - C.T = \left(826.00^2 + 865.02^2 + ... - 821.98^2\right) - 7892636$$
$$= 7921058 - 7892636$$
$$= 28422$$

4. Block sum of square

$$SS_{Block} = \sum_j \frac{\left(\sum_i Y_{1,j}\right)^2}{n} - C.T$$
$$= \frac{2406.05^2}{3} + \frac{2540.99^2}{3} + \frac{2253.00^2}{3} + \frac{2532.00^2}{3} - 7892636$$
$$= 7910913 - 7892636 = 18277.26$$

The ANOVA table

Source	SS	df	ms	f
Block	18277.26	3	6092.42	10.51
Treatment	6667.36	2	3333.68	5.75
Residual	3477.38	6	579.56	
Total	28422	11		

The critical value of F for testing treatments for 2 and 6 degrees of freedom and the level of significant $(F_{0.05,2,6} = 5.14)$. Since the calculated $F = 5.75$ is greater than the critical value, this implies that dietary treatment has different effect on body weight.

3.4 Factorial Experiment

An experiment was conducted to determining the effect of enzyme additions supplementation in feed on feed conversion ratio of broiler chickens.

Allzyme	0.01		0.02	
Roxazyme	0	10	0	10
	0.6	0.6	0.5	0.7
	0.5	0.5	0.5	0.7
	0.5	0.6	0.9	0.9
	0.5	0.5	0.5	0.7
	0.5	0.6	0.5	0.7
Sum	2.6	2.8	2.9	3.7
Average	0.52	0.56	0.58	074

Model for factorial experiment
The model:

$Y_{ij} = \mu + A_i + B_J + (AB)_{ij} + \varepsilon_{ijk}$

$Y_{ijk} =$ Observation of k in the ith treatment

$\mu =$ Overall mean

$A_i =$ The effect of level i of factor A

B_J = The effect of level j of factor B

$(AB)_{ij}$ = Interaction of factor A and B with levels ij

$\varepsilon_{ij} =$ Random error $(0, \sigma^2)$

The mathematical sum of squares are calculated

Total Sum of squares

$\sum_i \sum_j \sum_k Y_{1,k} = (0.6 + 0.5........ + 0.7) = 12$

Correction for the mean

$$C.T = \frac{\left(\sum_i \sum_j \sum_k Y_{1,k}\right)^2}{rxc} = \frac{(12)^2}{20} = 7.2$$

Total sum of squares

$SS_{Total} = \sum_i \sum_j \sum_k (Y_{i,j})^2 - C.T$

$= 0.6^2 + 0.5^2 + \ldots + 0.7^2 - 7.2$

$= 7.52 - 7.2$
$= 0.32$

Sum of squares for Allzyme

$$SS_{Allzyme} = \sum_i \frac{\left(\sum_i \sum_j \sum_k Y_{1,k}\right)^2}{nr} - C.T$$

$$= \frac{(2.6 + 2.8)^2}{10} + \frac{(2.9 + 3.7)^2}{10} - 7.2$$

$= 2.9 + 4.4$
$= 7.3 - 7.2 = 0.10$

Sum of squares for Roxazyme

$$SS_{Roxazyme} = \sum_j \frac{\left(\sum_i \sum_j \sum_k Y_{1,k}\right)^2}{nc} - C.T$$

$$= \frac{(2.6 + 2.9)^2}{10} + \frac{(2.8 + 3.7)^2}{10} - C.T$$

$= 3.025 + 4.23 - 7.2$
$= 7.255 - 7.2 = 0.06$

Sum of squares for interaction

$$SS_{alzme \times Roxazyme} = \sum_i \sum_j \frac{\left(\sum_k Y_{1,k}\right)^2}{n} - SS_A - SS_B - C.T$$

$$= \frac{(2.6)^2}{5} + \frac{(2.5)^2}{5} + \frac{(2.9)^2}{5} + \frac{(3.7)^2}{5}$$

$= 7.34 - 0.10 - 0.06 - 7.2$
$= 0.02$

Residual sum of squares

$$SS_{residual} = SS_{TOT} - SS_{alzme} - SS_{Roxazyme} - SS_{alzXrox}$$

$= 0.32 - 0.10 - 0.06 - 0.02$
$= 0.14$

ANOVA Table

Source	SS	df	ms	f
Alzyme	0.10	1	0.10	11.11
Roxazyme	0.06	1	0.06	6.67
Alzymex Roxazyme	0.02	1	0.02	2.22
Residual	0.14	16	0.009	

Total	0.32	19	0.02

The critical value for value for $\alpha = 0.05$ is $F_{0.05}1{,}16 = 4.49$. The computed F value is 2.22 from the interactions. The calculated F value is less than the critical value. So this implies that there is no significant difference between the two enzyme types.

3.5 Hierarchical or Nested Design

An experiment was conducted to determine the effects of sire and dam on variability of daily gain of their offspring in Japanese quail.

Sire	Dam	Offspring	daily Total gain	Sum per sire	Sum per Dam
1	1	1	1.2		
1	1	2	1.2		
					2.4
1	2	3	1.2		
1	2	4	1.3		
1	3	5	1.1		
					2.5
1	3	6	1.2	7.2	
					2.3
2	4	7	1.2		
2	4	8	1.2		
					2.4
2	5	9	1.1		
2	5	10	1.2		
					2.3
2	6	11	1.2		
2	6	12	1.1	7.0	
					2.3
3	7	13	1.2		
					2.4
3	7	14	1.2		
3	8	15	1.3		
					2.6
3	8	16	1.3		
3	9	17	1.2	7.4	
					2.4
3	9	18	1.2		

4	10	19	1.3			2.6
4	10	20	1.3			
4	11	21	1.4			2.8
4	11	22	1.4			
4	12	23	1.3			
4	12	24	1.3	29.6	8.0	2.6
Sum			29.6	29.6	29.6	29.6

Solutions

The structure of the design is that a sire mated to several dams and each dam producing several offspring.

$x =$ number of sire =4

$y =$ number of dam per sire =3

$n =$ number of chicks per dam

Mathematical computations

1. Total sum

$$\sum_i \sum_j \sum_k Y_{1,k} = (1.2 + 1.2 + 1.2 + ... + 1.3) = 29.6$$

2. Correction total

$$C.T = \frac{\left(\sum_i \sum_j \sum_k Y_{1,k}\right)^2}{xyn} = \frac{(29.6)^2}{24} = 36.51$$

$$xyn = 4 \times 3 \times 2 = 24$$

3. Sum of squares for sires

$$SS_{Sire} = \sum_i \frac{\left(\sum_i \sum_j \sum_k Y_{1,k}\right)^2}{nx} - C.T \ \frac{1}{6}\left[(7.2)^2 + ... + (8)^2\right] - 36.51$$

$$= 0.09$$

4. Sum of squares for dam within sires

$$SS_{Sire}(dam) = \sum_i \sum_j \frac{\left(\sum_k Y_{1,k}\right)^2}{n} - SS_{Sire} - C.T$$

$$= \frac{1}{2}\left[(2.4)^2 + (2.5)^2 + ... + (2.6)^2\right] - 0.09 - 36.51$$

$$= 0.04$$

5. Total sum of squares

$$SS_{TOT} = \sum_i \sum_j \sum_k \left(Y_{i,j}\right)^2 - C.T$$

$$= (1.2)^2 + (1.2)^2 + ... + (1.3)^2 - C.T$$

$$= 0.15$$

6. Sum of squares for sire within dam

$$SS_{calves}(dam) = SS_{TOT} - SS_{Sire} - SS_{dam}(sire)$$
$$= 0.15 - 0.09 - 0.04$$
$$= 0.02$$

ANOVA Table

Source	SS	df	ms	f
Sire	0.09	3	0.03	6.22
Dam within sire	0.04	8	0.005	3.00
Chicks within dam	0.02	12	0.002	
Total	0.15	23		

The estimation of variance components are presented in the following table.

Source	E (Ms)	Variance Component	Percent
Sires	$\sigma^2 + 2\sigma^2{}_{dams} + 6\sigma^2{}_{Sires}$	0.004	56.63
Dams within sires	$\sigma^2 + 2\sigma^2{}_{dams}$	0.002	21.69
Chicks within Dams	σ^2	0.002	21.69
Total		0.008	100.00

The basic assumption is that the effects of sires and dams are random. The random error for sire is the mean square for dams within sire, and the random error for dams is the mean square for chicks within dams. The critical value for sire is $F_{0.05,}3{,}8 = 4.07,$ and the critical valve for dams within sires is $F_{0.05,}8{,}12 = 2.85.$ since the calculated F values are greater than the critical valves and thus the effects of sires and dams are significant on the daily gain of Japanese quail chicks.

SELF –ASSESSMENT EXERCISE

Under what conditions do you use t and z tests?

4.0 CONCLUSION

In this unit, you have learnt the mathematical computation of statistical methods through sequential steps of resolving your sum of squares which is finally processed in your ANOVA table.

5.0 SUMMARY

This unit revealed the mathematical computations in different statistical layout. For you to apply a particular statistical method several conditions surrounding the utilisation of such method has to be met.

6.0 TUTOR-MARKED ASSIGNMENT

Twenty one crossbred Holstein X Bunaji cows of different genetic constitution are in lactations. The arithmetic mean of the sample, 5200kg is significantly different from 3600kg. The sample standard deviation is 1200kg. Test the hypothesis at 95% confidence interval.

7.0 REFERENCE/FURTHER READING

Little, R.; Stroup, W. & Wolfinger, R. D. (1996). *Agricultural Experimentation*. New York: John Wiley and Sons.

MODULE 7 PRESENTATION OF RESEARCH FINDINGS IN NARRATIVE, TABULAR AND GRAPHICAL FORMS

UNIT 1 RESULTS PRESENTATION THROUGH TABULAR FORM

CONTENTS

1.0 INTRODUCTION

In this unit, you will be acquainted with guidelines for construction of tables and its interpretation. The unit will explain the introductory part and will briefly take you through the classification of tables. Go through all the sections carefully, as they are designed to make you understand the course better. For instance, pay due attention to your self-assessment exercises; this will enhance your understanding of the content of the unit, so that the stated objectives for the unit can be achieved.

2.0 OBJECTIVES

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

- differentiate between table types
- itemise the guidelines for construction of tables.

3.0 MAIN CONTENT

3.1 Guidelines for Construction of Tables

a. The title of the table should be self-explanatory and reflect the content of the table.
b. Tables should follow table format presentation.
c. Remember to mention the units .e.g. yielding.
d. In table heading, use singular form not plural.
e. Data presentation should be done in rows and columns.
f. Long rows and columns should be divided into subheadings.
g. Use subtotals where applicable.
h. Tables are used to show exact values.
i. Data should be presented with a few decimal points (e.g. 2 to 3 decimal places) for ease of calculation; data are presented in tables.

3.1.1 Advantage of Tabular Presentation

i. It allows the process of summarisation of complex data.
ii. It allow the interaction of data.
iii. It allows for process of essential information gathering.

3.2 Types of Tables

Tables are classified into one-way, two-way, and three-way. This helps to bring out the features and interactions under study.

i. **One-way Table**: It consists of serial numbers. Example, 1, 2, 3….X, Y…t. Information from this type of table is only in one direction.
ii. **Two-way Table**: It involves two set of data. Example, Age and breed of rabbits. The two-way table classification gives a more straight forward idea of the interaction (i.e. how the effect of one influences the other)

	B_1	B_2	B_3	Total
A_1	X_1	X_2	X_3	$\sum A_1$
A_2	X_2	X_3	X_1	$\sum A_2$
A_3	X_3	X_1	X_2	$\sum A_3$
Total	$\sum B_1$	$\sum B_2$	$\sum B_3$	$\sum A_i B_j$

iii. **Three-way Table**: It gives the effect of two variables on the third variable. Thus, it is called second order interaction. Beyond the three way classification, the statistical classification is called a many fold case. It is not important anymore because of the complexities.

	C_1		C_2		Total
	B_1	B_2	B_1	B_2	
A_1	X_1	X2	X3	X4	Σ A1
A_2	X_2	X3	X2	X3	Σ A2
A_3	X_3	X4	X1	X2	Σ A3
A_4	X_4	X1	X4	X1	Σ A4
Total	ΣC_1B_1	ΣC_1B_2	ΣC_2B_1	ΣC_2B_1	$\Sigma A_iC_jB_k$

Disadvantage of Table Presentation

i. It is less attractive to reader eyes
ii. It lacks ability for detecting serious compromises to data integrity
iii. It does not have the capacity to detect outliers

SELF ASSESSMENT EXERCISE

Highlight the various guidelines you need to put into consideration in table presentation.

4.0 CONCLUSION

Looking at the basic guide in constructing your tables, you must have observed that in order to enhance the clarity of your values, table presentation should follow an acceptable global format.

5.0 SUMMARY

In this unit, you learnt about basic guide for construction of tables and their advantages and disadvantages.

6.0 TUTOR-MARKED ASSIGNMENT

1. List sequentially the guidelines for construction of tables.
2. Briefly explain different types of tables.

7.0 REFERENCE/FURTHER READING

Iman, R.L. (1994). A Data Based Approach to Statistics. California: Duxbery Press, Belmont.

UNIT 2 RESULTS PRESENTATION THROUGH GRAPHICAL FORM

CONTENTS

1.0 INTRODUCTION

This unit introduces the reader to several attractive features of figures. Figures make our result presentation more simplex and easy to comprehend. Always bear in mind, that charts make your work to look colorful. Note that in all figures, you should include graphical representation of interval estimates.

2.0 OBJECTIVES

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

- present your data using figures
- determine the features of different graphs.

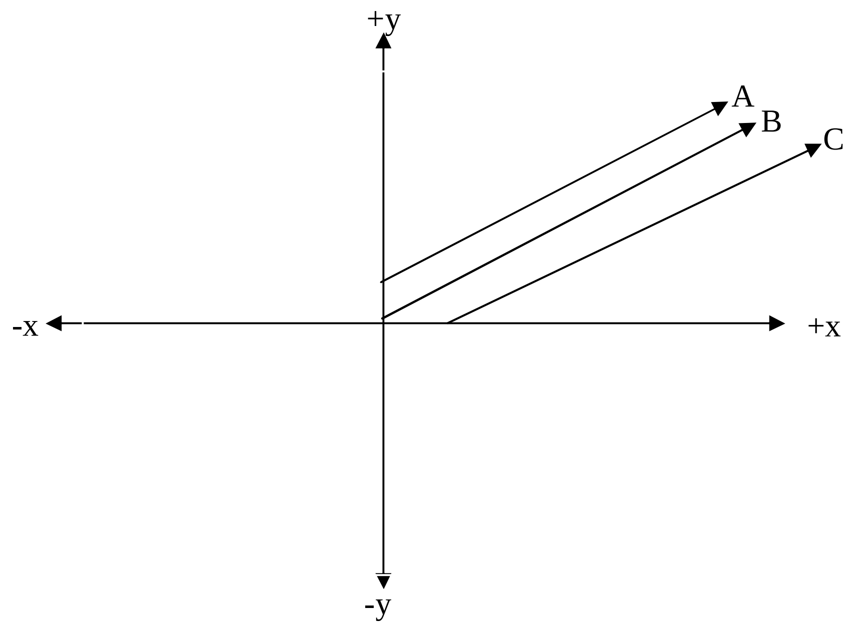
3.0 MAIN CONTENT

3.1 Graphical or Diagrammatical Representation of Data

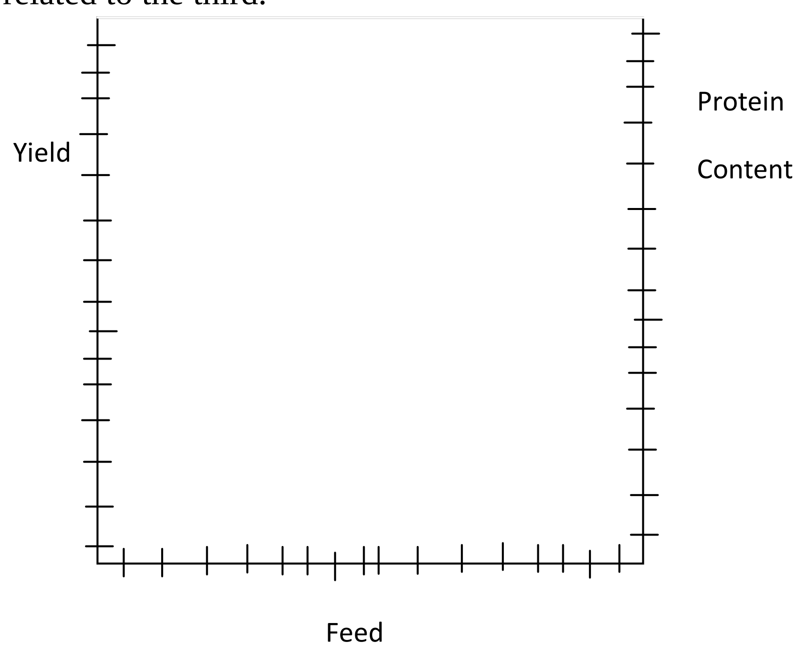
Data when represented in the form of a graph or diagram becomes easier to conceive, memorise and compare. Graphs are lines relating dependent and independent variables. The relationship of two variables can be graphed in the form of positive and negative variations.

Advantages of graphs

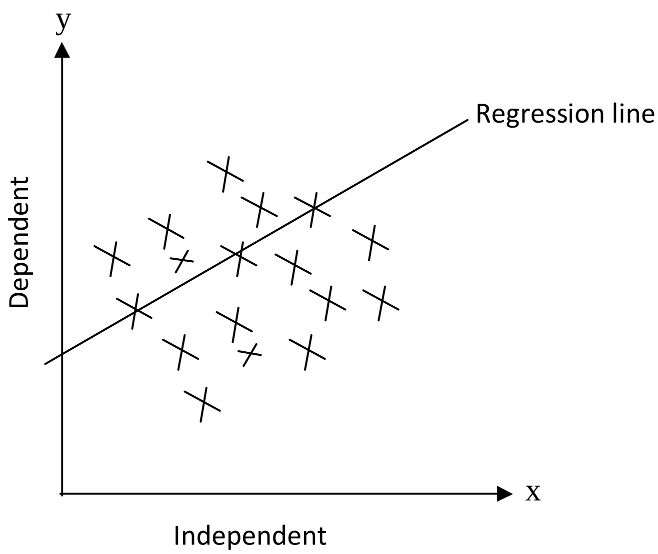
Graphs are frequently useful in interpreting significant interactions



More than one variable, A, B, C, can be represented in one graph for easy comparison. Three variables can be represented in one graph if two of them are related to the third.

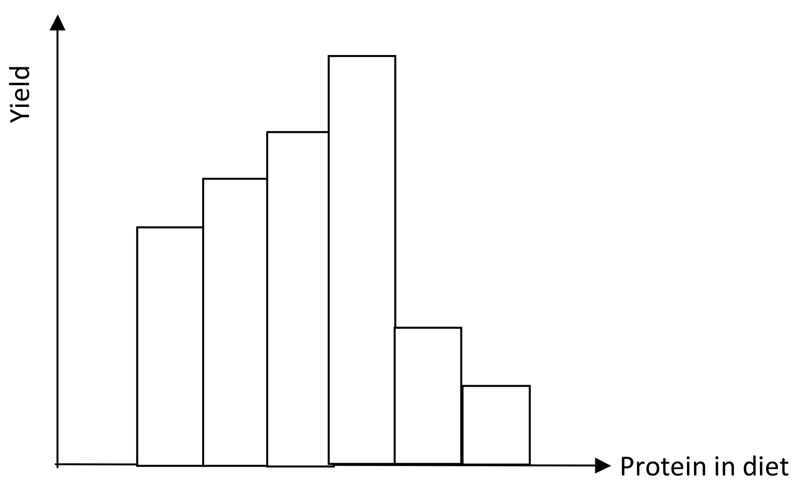


Scattered Diagrams – Used when studying bivariate population

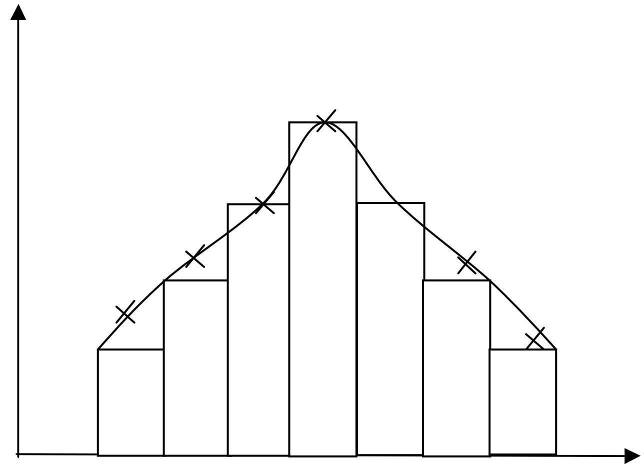


Here, the average representative points fall on the regression line. Other forms of graphs can be in the form of linear response, quadratic, cubic, quartic or quintile response. Higher degrees of response are too complicated.

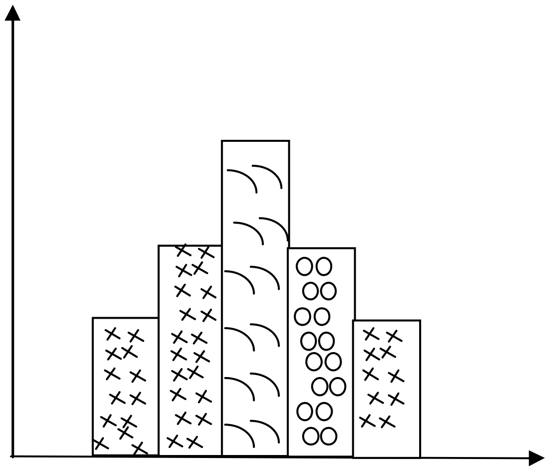
Histograms
Histograms are variables represented in the form of solid bars.



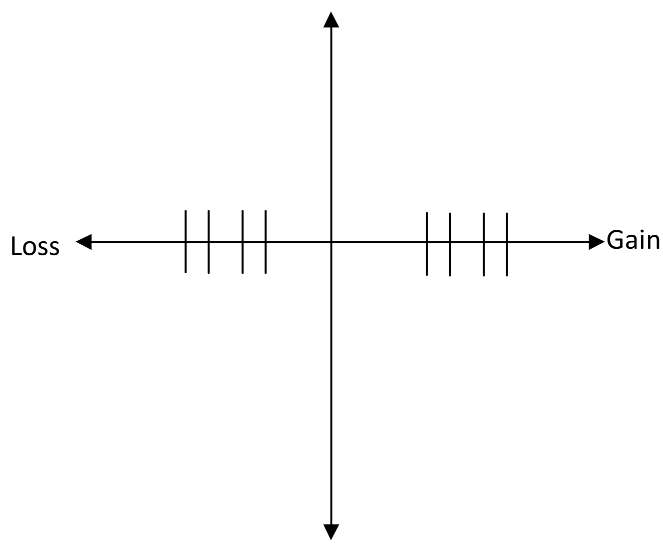
Frequency polygons are obtained by joining the mid-point of the bars from the top. For smaller intervals, the polygon is in the form of a curve called **frequency curve**.



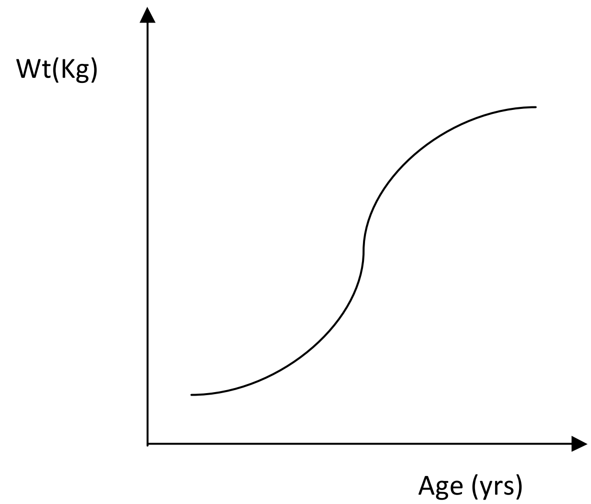
When more than one variable are used in a histogram, the bars could be distinguished by using different shades.



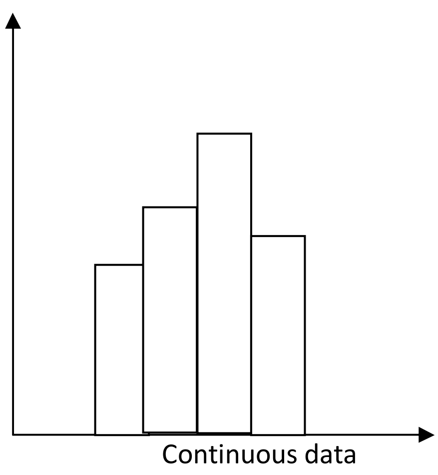
Two variables can be represented on both sides of a graph.

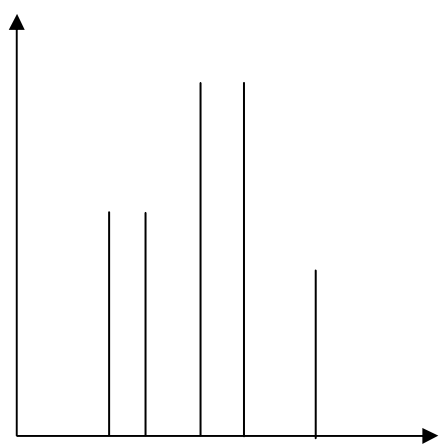


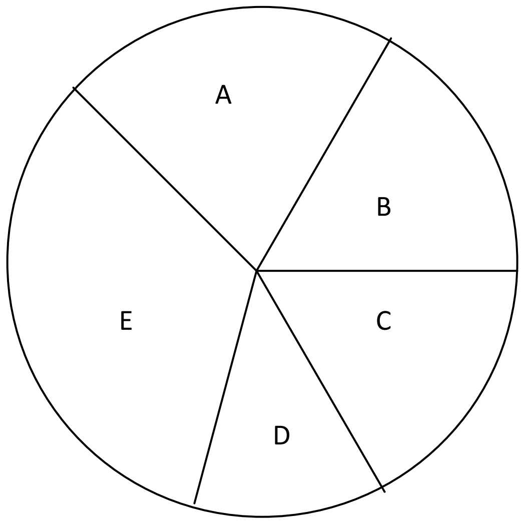
Sometimes, frequency curves are drawn from classified table. The frequency polygon or curve obtained is called **sigmoid curve** because it is S-shaped. It is also called **an ogive**. An example of this is the growth curve in animals.



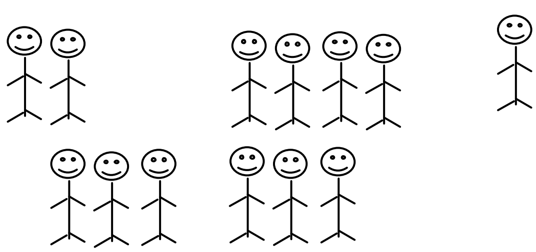
NB: In drawing histograms of discrete data, lines instead of bars are used and frequency polygons are not drawn from it.







Pie Chart
In Pie Chart the proportion of various components can be seen at a glance.



Pictograms
This makes use of sketches or shapes to represent data

= 1,000 animals

SELF ASSESSMENT EXERCISE

List and explain the basic features of a graph.

4.0 CONCLUSION

In this unit, you have learnt how to use different graphs for your data representation.

5.0 SUMMARY

In this unit, you have learnt the major importance of figures such as attracting the reader's eye and help convey global results.

6.0 TUTOR-MARKED ASSIGNMENT

1. Explain the advantages and disadvantages of figures.
2. Draw and list five figure types.

7.0 REFERENCE/FURTHER READING

McClave, J. R. & Dietrich, F.H. (1987). Statistics (3rd), Boston M.A.: Duxbury Press.