NATIONAL OPEN UNIVERSITY OF NIGERIA

APPLIED STATISTICS
ECO 452

FACULTY OF SOCIAL SCIENCES

COURSE GUIDE

Course Developers:
Dr. Adesina- Uthman Ganiyat Adejoke
Department of Economics, Faculty of Social Sciences,
National Open University of Nigeria.
and
Ogunjirin Olakunle
Yaba College of Technology
School of Liberal Studies
Department of Social Sciences.

Course Editor:

Dr. Ogunsakin Sanya
Department of Economics
Senior Lecturer,
Ekiti State University, Ado-Ekiti.

NATIONAL OPEN UNIVERSITY OF NIGERIA

National Open University of Nigeria

Headquarters

Plot 91, Cadastral Zone, University Village,
Nnamdi Azikiwe Expressway, Jabi, Abuja.

Introduction

Statistical economics is a branch of economic that deals with analysis of economic phenomenon.

What you will learn in the course

In this course, you will be introduced to various analytical tools in statistics; Here, you will be exposed to the underlying assumptions, formulae and calculations of the topics under consideration. Also, you will be taught the decision criteria under each topic.

Course Content

This course will expose you to different statistical tools that economist can apply in economic analysis. This course is built on the foundation of elementary statistics and elementary economics in the understanding of real life situation.

Course Aims

There are fourteen study units in the course and each unit has its objectives. You are advised to read through the objective of each them and bear them in mind as you through each of the unit. In addition these objective is the overall objective which includes;

- Exposing you to basic statistical tools that can be applied in economics,
- Apply these tools to real life situation,
- Expose the students to economic interpretation of all calculated coefficients

Course objectives

The over-all objectives of this course are;

- to expand the learning horizons of the students
- to understand how to apply statistical tool in economics

Working through This Course

You have to work through all the study units in the course. There are four modules and fourteen study units in all.

Course Materials

Major components of the course are:

1. Course Guide
2. Study Units
3. Textbooks
4. CDs
5. Assignments File
6. Presentation Schedule

Study Units

The breakdown of the four Modules and the 14 study units are as follows:

Module 1; Statistical Inference

Unit 1; sampling distribution defined

Unit 2; sampling distribution of proportion

Unit 3; sampling distribution of difference and sum of two means

Unit 4; probability distribution

Module 2: **Analysis of Variance and Analysis of Covariance**

Unit 1: one-way factor analysis of variance

Unit 2: two-way factor analysis of variance

Unit 3: analysis of covariance

Module 3: Multiple Regression Analysis

Unit 1: Estimation of multiple regressions

Unit 2: Partial correlation coefficient

Unit 3: Multiple correlation coefficient and coefficient of determination

Unit 4: Overall test of significance

Module 4: Time Series Analysis

Unit 1: time series and its components

Unit 2; Quantitative estimation of time series

Unit 3: price index

References and Other Resources

Attached to every unit is a list of references and further reading. Try to get as many as possible of those textbooks and materials listed. The textbooks and materials are meant to deepen your knowledge of the course.

Assignment File

In this file, you will find all the details of the work you must submit to your tutor for marking. The marks you obtain from these assignments will count towards the final mark you obtain for this course. Further information on assignments will be found in the Assignment File itself and later in this *Course Guide* in the section on assessment.

Presentation Schedule

The Presentation Schedule included in your course materials gives you the important dates for the completion of tutor-marked assignments and attending tutorials. Remember, you are required to submit all your assignments by the due date. You should guard against falling behind in your work.

Assessment

Your assessment will be based on tutor-marked assignments (TMAs) and a final examination which you will write at the end of the course.

Tutor Marked Assignments (TMA)

Every unit contains at least one or two assignments. You are advised to work through all the assignments and submit them for assessment. Your tutor will assess the assignments and select four which will constitute the 30% of your final grade. The tutor-marked assignments may be presented to you in a separate file. Just know that for every unit there are some tutor-marked assignments for you. It is important you do them and submit for assessment.

Final Examination and Grading

At the end of the course, you will write a final examination which will constitute 70% of your final grade. In the examination which shall last for two hours, you will be requested to answer three questions out of at least five questions.

Course Marking Scheme

This table shows how the actual course marking is broken down. Assessment	Marks
Assignments	Four assignments, best three marks of the four count at 30% of course marks
Final Examination	70% of overall course marks
Total	100% of course marks

CONTENT

Introduction

The course advanced statistics (ECO 410) is a first semester course which carries two credit units for fourth year level economics students in the School of Art and Social Sciences at the National Open University, Nigeria. The course is a very useful course to you in your academic pursuit, because it helps gain in-depth insight of the underlining statistical tools usually used by economists.

This course guide tells you what advanced statistics entails, what course materials you will be using and how you can work your way through these materials. It suggests some general guidelines for the amount of time required of you on each unit in order to achieve the course aims and objectives successfully. It also provides you some guidance on your tutor marked assignments (TMAs) as contained herein.

Course Content

This course is built on the foundation of what you have learnt in your elementary statistics. Topics covered include: statistical inference, probability distribution, analysis of variance, multiple regressions, time series, price index.

Working through the Course

To successfully complete this course, you are required to read the study units, referenced books and other materials on the course.

Each unit contains self-assessment exercises called Tutor Marked Assessment. At some points in the course, you will be required to submit assignments for assessment purposes. At the end of the course there is a final examination. This course should take about 15weeks to complete and some components of the course are outlined under the course material subsection.

Course Material

The major component of the course, What you have to do and how you should allocate your time to each unit in order to complete the course successfully on time are listed follows:

1. Course guide
2. Study unit
3. Textbook
4. Assignment file
5. Presentation schedule

Tutor-Marked Assignments (TMAs)

There are four tutor-marked assignments in this course. You will submit all the assignments. You are encouraged to work all the questions thoroughly. The TMAs constitute 30% of the total score.

Assignment questions for the units in this course are contained in the Assignment File. You will be able to complete your assignments from the information and materials contained in your set books, reading and study units. However, it is desirable that you demonstrate that you have read and researched more widely than the required minimum. You should use other references to have a broad viewpoint of the subject and also to give you a deeper understanding of the subject.

When you have completed each assignment, send it, together with a TMA form, to your tutor. Make sure that each assignment reaches your tutor on or before the deadline given in the Presentation File. If for any reason, you cannot complete your work on time, contact your tutor before the assignment is due to discuss the possibility of an extension. Extensions will not be granted after the due date unless there are exceptional circumstances.

Final Examination and Grading

The final examination will be of two hours' duration and have a value of 70% of the total course grade. The examination will consist of questions which reflect the types of self-assessment practice exercises and tutor-marked problems you have previously encountered. All areas of the course will be assessed

Use the time between finishing the last unit and sitting for the examination to revise the entire course material. You might find it useful to review your self-assessment exercises, tutor-marked assignments and comments on them before the examination. The final examination covers information from all parts of the course.

Course Marking Scheme

The table presented below indicate the total marks (100%) allocation.

Assessment	**Marks**
Assignment (Best three assignment out of the four marked)	30%
Final Examination	70%
Total	**100%**

Course Overview

The table presented below indicate the units, number of weeks and assignments to be taken by you to successfully complete the course, advanced statistics (ECO 410).

How to Get the Most from This Course

In distance learning the study units replace the 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.

Think of it as reading the lecture instead of listening to a lecturer. In the same way that a lecturer might set you some reading to do, the study units tell you when to read your books or other material, and when to embark on discussion with your colleagues. Just as a lecturer might give you an in-class exercise, your study units provides exercises for you to do at appropriate points.

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 and getting the best grade.

The main body of the unit guides you through the required reading from other sources. This will usually be either from your set books or from a readings section. Some units require you to undertake practical overview of events. You will be directed when you need to embark on discussion and guided through the tasks you must do.

The purpose of the practical overview of some certain practical issues are in twofold. First, it will enhance your understanding of the material in the unit. Second, it will give you practical experience and skills to evaluate economic propositions, arguments, and conclusions. In any event, most of the critical thinking skills you will develop during studying are applicable in normal working practice, so it is important that you encounter them during your studies.

Self-assessments are interspersed throughout the units, and answers are given at the ends of the units. Working through these tests will help you to achieve the objectives of the unit and prepare you for the assignments and the examination. You should do each self-assessment exercises as you come to it in the study unit.

The following is a practical strategy for working through the course. If you run into any trouble, consult your tutor. Remember that your tutor's job is to help you. When you need help, don't hesitate to call and ask your tutor to provide it.

1. Read this Course Guide thoroughly.
2. Organize a study schedule. Refer to the `Course overview' for more details. Note the time you are expected to spend on each unit and how the assignments relate to the units. Important information, e.g. details of your tutorials, and the date of the first day of the semester is available from study centre. You need to gather together all this information in one place, such as your dairy or a wall calendar.

Whatever method you choose to use, you should decide on and write in your own dates for working breach unit.

3. Once you have created your own study schedule, do everything you can to stick to it. The major reason that students fail is that they get behind with their course work. If you get into difficulties with your schedule, please let your tutor know before it is too late for help.
4. Turn to Unit 1 and read the introduction and the objectives for the unit.
5. Assemble the study materials. Information about what you need for a unit is given in the `Overview' at the beginning of each unit. You will also need both the study unit you are working on and one of your set books on your desk at the same time.
6. Work through the unit. The content of the unit itself has been arranged to provide a sequence for you to follow. As you work through the unit you will be instructed to read sections from your set books or other articles. Use the unit to guide your reading.
7. Up-to-date course information will be continuously delivered to you at the study centre.
8. Work before the relevant due date (about 4 weeks before due dates), get the Assignment File for the next required assignment. Keep in mind that you will learn a lot by doing the assignments carefully. They have been designed to help you meet the objectives of the course and, therefore, will help you pass the exam. Submit all assignments no later than the due date.
9. Review the objectives for each study unit to confirm that you have achieved them. If you feel unsure about any of the objectives, review the study material or consult your tutor.
10. When you are confident that you have achieved a unit's objectives, you can then start on the next unit. Proceed unit by unit through the course and try to pace your study so that you keep yourself on schedule.
11. When you have submitted an assignment to your tutor for marking do not wait for it return `before starting on the next units. Keep to your schedule. When the

assignment is returned, pay particular attention to your tutor's comments, both on the tutor-marked assignment form and also written on the assignment. Consult your tutor as soon as possible if you have any questions or problems.

12. After completing the last unit, review the course and prepare yourself for the final examination. Check that you have achieved the unit objectives (listed at the beginning of each unit) and the course objectives (listed in this Course Guide).

Tutors and Tutorials

There are some hours of tutorials (2-hours sessions) provided in support of this course. You will be notified of the dates, times and location of these tutorials. Together with the name and phone number of your tutor, as soon as you are allocated a tutorial group.

Your tutor will mark and comment on your assignments, keep a close watch on your progress and on any difficulties you might 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 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 units or the assigned readings
- You have difficulty with the self-assessment exercises
- You have a question or problem with an assignment, with your tutor's comments on an assignment or with the grading of an assignment.

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

Summary

On successful completion of the course, you would have developed critical thinking skills with the material necessary for efficient and effective use of statistical tools economics. However, to gain a lot from the course please try to apply anything you must have learnt in the course to practice by doing the calculation on paper yourself. We wish you success with the course and hope that you will find it both interesting and useful.

MODULE ONE; Statistical Inference

UNIT ONE; Sampling Distribution

CONTENT

1.0 INTRODUCTION

Generally statistical data are studied in order to learn something about the broader field which the data represents. In order to make statistical work meaningful, statistician generalize from what we find in the figure at hand to the wider phenomenon which they represent. In technical language we regard a set of data as a sample drawn from a larger “universe”. We analyze the data of the sample in order to draw conclusion about the corresponding universe or population.

In a sense universe actually exists and it is theoretically possible to study the universe completely. But in another sense the universe is broader and in a sense less tangible.

This unit happens to be one of the four units in this module, for proper understanding of the topics in this unit a thorough knowledge of elementary statistics is required.

2.0 OBJECTIVES

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

- Sample
- Population
- Sampling theory
- Parameter estimation
- Estimate sample mean, population mean etc.

3.0 MAIN CONTENT

3.1 Sampling Theory, Population and Sample Defined

Statistical inference is defined as the process by which on the basis of sample we draw conclusion about the universe from which sample is drawn. It can as well be defined as a process by which conclusion are drawn about some measure or attribute of a population based upon analysis of sample. Samples are taken and analyzed in order to draw conclusion about the whole population.

Sampling theory is a study of relationships existing between a population and samples drawn from the population. Sampling theory is also useful in determining whether the observed differences between two samples are due to chance variation or whether they are really significant.

In general, a study of the inference made concerning a population by using sample drawn from it together with indication of accuracy of such inferences by using probability theory is called statistical inference. Population of a variable X is usually defined to consist of all the conceptually possible values that the variable may assume. Some of these values may have already been observed, others may not have occurred, but their occurrence is conceivably possible. The number of conceptually possible values of a variable is called size of the population. This size varies according to the phenomenon being investigated.

A population may be finite, when it consists of a given number of values or it may be infinite, when it includes an infinite number of values of the variable.

In most cases values of population are hardly known, what we usually have is a certain number of values that any particular variable has assumed and which have been recorded in one way or the other. Such data form a sample from the population.

Sample refers to a collection of observation on a certain variable. The number of observations included in the sample is called the size of the sample.

The main object of the theory of statistics is the development of method of drawing conclusion about the population (unknown) from the information provided by a sample.

In order to facilitate the study of population and sample, statisticians have introduced various descriptive measures that is various characteristics values that describes the important features of the sample or the population. The most important of these characteristics are the mean, variance and the standard deviation. To distinguish between sample and populations statistician use the term parameter for the basic descriptive measure of population while statistics is usually used for the basic descriptive measure of a sample.

Table M1.1.1

Basic Descriptive Measure of Population and Sample

	Population parameters	**Symbol**	**Sample statistics**	**Symbol**
I	Population mean	μ	Sample mean	$\bar{x}$
Ii	Population variance	σx^2	Sample variance	S_{x2}
Iii	Population standard deviation	σ_x	Sample standard deviation	s_x

Note $E(x) = \mu = \frac{x_1 + x_2 + \ldots\ldots x_n}{n}$

SELF ASSESMENT EXERCISE

What are descriptive measures that can be used in describing a sample or population?

3.2 Sampling Distribution of Parameter & Sample Estimates

The population mean is usually referred to as the expected value of the population and it is conventionally denoted as E(x) or μ. But for a discrete random variable the expected value is computed by the sum of the product of value of X_1 multiplied by their various probabilities.

$$E\,(x) = \mu = \sum_{i=1}^{n} xf(X_1)$$

Where X_i is the probability of variable x.

The variance of a population is defined as the expected value of the squared deviations of the value of x from their expected mean value.

$$\text{Var}\,(x) = \sigma_x^2 = \sum \frac{(X - E(x))^2}{n} = \frac{\sum (x - \mu)^2}{n}$$

Where E (x) = population mean value

This shows the various ways in which the various value of random variable x is distributed around their expected mean values. The smaller the variance, the closer and cluster of the values of x around the population mean.

The standard deviation of a population is defined as the square root of the population variance. This is denoted as:

$$\sigma_x = \sqrt{\frac{\Sigma(x-(Ex))^2}{n}} = \Sigma\sqrt{\frac{(x - \mu)^2}{n}}$$

The standard deviation is a measure that describes how dispersed the values of x is around the population mean.

$\text{COV}(XY) = \Sigma(XY) - \Sigma X\, \Sigma Y$

Worked Example

Given the population 11, 12, 13, 14, 15 calculate the mean, standard deviation, and the variance of the given population.

Table M1.1.2

Table of Analysis for Sample Mean, Standard Deviation and Variance

X	**$X - \mu$** **$X – E(X)$**	**$(X - \mu)^2$** **$(X – E(x)^2$**
11	1 1 – 13 = 2	4
12	12 – 13 = 1	1
13	13 – 13 = 0	0
14	14 – 13 = 1	1
15	15 – 13 = 2	4
n = 5		10

$$\bar{x} \quad \mu = \frac{11 + 12 + 13 + 14 + 15}{5} = \frac{65}{5} = 13$$

$\text{Var}(X) = \Sigma (X – E(x))^2 = \Sigma (x - \mu)^2$

$\text{Var}(X) = \frac{10}{2} = 5$

$\delta_X = \sqrt{2}$

$\delta_X = 1.4142$

SELF-ASSESSMENT EXERCISE

Define standard deviation of a population

3.3 Estimation of Sample Statistics

As it has been said before now that, the term statistics is usually used in describing the features of a sample. The basic statistic of a sample corresponding to the parameters of the population are sample mean usually denoted by $\bar{x}$, sample variance denoted by S_x^2 and sample standard deviation denoted by S_x.

Sample mean is defined as the average value in the sample it is denoted by $\bar{x}$. The sample arithmetic mean is calculated by adding up the observation of the sample and then dividing by the total number of observations.

$$\bar{X} = \frac{\sum_{i=o}^{n} X}{n}$$

Sample variance as it has been said before now, it is a measure of dispersion of the value of x in the sample around their average value. This is denoted as

$$S_x^2 = \frac{\sum_{i=o}^{n}(x - \bar{x})^2}{n} = \frac{\Sigma x^2 - n\bar{x}^2}{n} = \frac{\Sigma x^2 - \bar{x}^2}{n}$$

The sample standard deviation is denoted by S_x this is taken to be the square root of the variance.

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

Covariance; this statistics usually involves two variable. The covariance is defined as the sum of the product of the deviation of variable x and y from the various means.

$$COV(XY) = \frac{\sum_{i=1}^{n}(x - \bar{x})(y - \bar{y})}{n}$$

Question

From the information of population supplied in the preceding subsection i.e. 11, 12, 13, 14, 15

Table M1.1.3

Sample Statistics Table of Analysis

Possible samples	**$\bar{x}$ = mean of each sample**
(11,12)	$\frac{11+12}{2}$ = 11.5
(11,13)	$\frac{11+13}{2}$ = 12
(11,14)	$\frac{11+14}{2}$ = 12.5
(11,15)	$\frac{11+15}{2}$ = 13
(12,13)	$\frac{12+13}{2}$ = 12.5
(12,14)	$\frac{12+14}{2}$ = 13
(12, 15)	$\frac{12+15}{2}$ = 13.5
(13,14)	$\frac{13+14}{2}$ = 13.5
(13,15)	$\frac{13+15}{2}$ = 14
(14,15)	$\frac{14+15}{2}$ = 14.5
n = 10	

$$\text{Sample mean} = \frac{11.5+12+12.5+13+12.5+13+13.5+13.5+14+14.5}{10}$$

$$= \frac{130}{10} = 13$$

All the information about the population and possible samples can be summarize in a frequency distribution as depicted in table 1.3 below.

Table M1.1.14
Table of Possible Samples

X	**F**
11	1
11.5	1
12	1
12.5	2
13	2
13.5	2
14	1
14.5	1
15	1

Variance of sample mean $= \frac{\sum_{i=1}^{n}(x - \bar{x})^2}{N}$

$$S_x^2 = \frac{(11\text{-}13)^2 + (11.5\text{-}13)^2 + (12\text{-}13)^2 + 2(12.5\text{-}13)^2 + 2\,(13\text{-}13)^2 + 2\,(13.5\text{-}13)^2 + (14\text{-}13)^2 + (14.5\text{-}13)^2}{9} + \frac{(15\text{-}13)^2}{9}$$

$$S_x^2 = \frac{(-2)^2+(1.5)^2+(-1)^2+2(-0.5)^2+2(0)^2+2(0.5)^2+(1)^2+(2)^2+(1.5)^2}{9}$$

$$S_x^2 = \frac{4 + 2.25 + 1 + (0.25)2 + 2(0) + 2\,(0.25) + 1 + 4 + 2.25}{9}$$

$$S_x^2 = \frac{4 + 2.25 + 1 + 0.5 + 0 + 0.5 + 1 + 4 + 2.25}{9}$$

$$S_x^2 = \frac{15.5}{9}$$

$S_x^2 = \quad 1.722$

$S_x^2 \cong 2$

$S_x = \sqrt{1.722}$

$S_x = 1.31233$

From the foregoing analysis it would be observed that given X_1, X_2 ……X_n of any random sample of size n from any infinite population with population mean u and σ^2 then with sample mean $\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x$ we have

(i) $E(\bar{x}) = \mu$

(ii) $Var (x) = \frac{\sigma 2}{n}$

SELF-ASSESSMENT EXERCISE

Define the sample variance of any given population

3.4 Estimators for Mean and Variance

Given that X_1, X_2, X_3 …. X_n is a random sample of size n from normal population with mean μ and variance σ^2 i.e. ($X \sim N(\mu, \sigma^2)$, then the statistics $\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x$

Therefore $Z = \frac{\frac{\bar{x} - \mu}{\sigma}}{\sqrt{n}} \sim N(0,1)$

This is a general case whereby sampling is specifically taken from a normal distribution.

Worked Example

Given a random sample of 20 taken from a normal distribution with mean 90 and variance 25 find the probability that the mean is greater than 101.

Solution

$\bar{x} \sim 20 (90, 25)$

$Z = \frac{\frac{\bar{x} - \mu}{\sigma}}{\sqrt{n}} \sim N(0,1)$

$$= \frac{101 - 90}{\frac{25}{\sqrt{20}}}$$

$$= \frac{11}{\frac{25}{\sqrt{20}}}$$

$$= \frac{11}{\frac{25}{4.47213}}$$

$$= \frac{11}{5.590169}$$

$$= 1.967739$$

$$\cong 1.968$$

SELF-ASSESSMENT EXERCISE

What are the assumptions of a normal distribution?

3.5. The Role and Significant of Statistics in Social Sciences

It is interesting to know that accuracy, validity, reliability, objectivity, analysis, efficiency are all characteristics of the roles expected of statistical research in decision making and policy formulation for societal development. Do you know that social statistics are necessary in information gathering about socio-economics variables that are indices of economic growth and development? It started with what is known as the “statists” social research” and later grow to be known as “statistics”, a new term for quantitative evidence. Social sciences’ statistics is very significant because it assist in quantifying scientific developments and data on them therefore, making information on scientific studies more concise and precise. Social statistics is usually conducted to prove something for instance, how many women are affected by malaria compare to men in the society? How many people in the society are able to afford living in a duplex, flat, one-room apartment, face-to-face room or under the bridge? Consequently, it is significant to note that adequate cautions are usually put into stepwise data gathering, accuracy, and analysis for efficiency. The role of statistics in social sciences and its significant cannot be overemphasized.

Self-Assessment Question

Do you think that statistics in social sciences has role to play in societal problem solving?

4.0 CONCLUSION

It has been established that given a random sample of X_1, X_2, …. X_n with population mean μ and standard variance r^2.

(i) $\Sigma(\bar{x}) = \mu$

(ii) $Var(x) = \sigma^2/n$

5.0 SUMMARY

In this unit, we have attempted the definition of population, sample, sample distribution theory, so also estimation of parameter estimate and sample statistics had been attempted, so also it has been proved from our calculation that the mean of sample must always equal to the population mean it's representing and that the variance of the population and sample estimate are equal.

6.0 TUTOR MARKED ASSIGNMENT

Explain the descriptive measure of a sample statistics.

8.0 REFERENCE/FURTHER READING

- Adedayo, O.A. (2006): Understanding statistics. JAS publishers Akoka, Yaba.

- Dominick, S. and Derrick P. (2011): (Schaum outline series) Statistics and Econometrics (second edition) MCGRAW HILL, New York.

- Edward, E.L. (1983): Methods of statistical analysis in Economics and Business. HOUGHTON MIFFLIN COMPANY BOSTON.

- Esan F.O. and Okafor, R.O. (2010): Basis statistical method (revised edition) Toniichristo Concept Lagos.

- Koutsoyianis, A. (2003): Theory of Econometrics (second edition). Palgrave publishers Ltd (formerly Macmillan press Ltd), London and Basic stoke.

- Murray R. S. and Larry J. S. (1998): (Schaum outlines series). Statistics (Third edition) MCGRAW HILLS.

- Olufolabo, O.O. and Talabi, C.O (2002): Principles and practice of statistics. HASFEM Nig Enterprises, Shomolu, Lagos.

- Oyesiku, O.K. and Omitogun, O. (1999). Statistics for social and management sciences. Higher Education Books Publisher Lagos.

UNIT TWO; SAMPLING DISTRIBUTION OF PROPORTION

CONTENT

1.0 INTRODUCTION

This unit is an extension of unit one of this module. In this unit we are going to look at sampling distribution of proportion, sampling distribution of sum and difference and standard error. Since this unit is an offshoot of the unit one of this module, most of the statistical term used in unit one will be implied here.

2.0 OBJECTIVE

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

- Sampling distribution of proportion
- Sampling distribution of sum
- Sampling distribution of difference and
- Standard error

3.0 MAIN CONTENT

3.1 Sampling Distribution of Proportion Defined

Samples are usually embedded in a population, each time attribute is sampled, the concept of proportion is coming in. the estimation here is concentrating on the proportion of the population that has a peculiar characteristics. This sampling distribution is like of binomial distribution, where an event is divided into been a success represented with p or been a failure represented with q or 1 – p.

Given an infinite population consisting of sample size n. The sampling distribution of proportion is said to have a mean of np.

and variance

$$\text{var (p)} = \text{var (p)} = \frac{P\ (1-p)}{n} = \frac{pq}{n}$$

It is to be noted at this juncture that the sample proportion is also an unbiased estimator of the population proportion i.e. Σ(p) = P

Example

A coin is tossed 120 times, find the probability that head will appear between 45% and 55%.

Solution

From the above the prob(head) = ½ = p

Prob(not obtaining ahead) = ½ = q = 1 – p

45% of tosses = $\frac{45}{100} \times 120$

= 54

While 55% of tosses gives $\frac{55}{100} \times 120$

= 66

Mean μp = np = 120 x ½

= 60

S.D = $\sqrt{\text{npq}}$ = $\frac{0.25}{120}$ x 120

= 0.00208333

S.D. $\sqrt{\frac{0.25}{120}}$

= 0.4564

S.D. = $\sqrt{\text{npq}}$ = $\sqrt{120\ \left(\frac{1}{2}\right)\left(\frac{1}{2}\right)}$

= $\sqrt{30}$

= 5.477225575

Prob (54 < p < 78) = p ($\frac{54-60}{5.5}$ < z < $\frac{66-60}{5.5}$)

= p ($\frac{6}{5.5}$ < z < $\frac{6}{5.5}$)

= p (- 1.0909 < z < 1.091) = (0.3621) x 2

= 0.7242

SELF-ASSESSMENT EXERCISE

What is the symbolic definition of standard deviation of a sample proportion?

3.2 Standard Error

Standard error usually represented by S.E. is defined as the square root of the population variance written as $\sqrt{\text{var p}}$

note var(p) = $\frac{\text{pq}}{\text{n}}$ = $\frac{\text{p }(1-\text{p})}{\text{n}}$

∴ $\sqrt{\frac{\text{p }(1-\text{p})}{\text{n}}}$

From the example in subsection 3.2 above

p = ½ = q

n = 120

:. S.E $= \sqrt{\frac{0.5\ (0.5)}{120}}$

S.E $= \sqrt{\frac{0.25}{120}}$

S.E $= \sqrt{0.0020833}$

S.E = 0.0456

SELF-ASSESSMENT EXERCISE

What does S.E stands for?

4.0 CONCLUSION

During the course of our discussion of this unit we have talked about;

- Sampling distribution of proportion
- Standard error

5.0 SUMMARY

In the course of our discussion we defined the mean of a sampling distribution of proportion as np.

i.e. mean = np

variance (p) = $\underline{P(1\text{-}P)}$

$\sigma(p) = \sqrt{npq}$

6.0 TUTOR MARKED ASSIGNMENT

A coin is tossed 90 times, find the probability that tail will appear between 35% and 55%.

7.0 References

7.0 REFERENCE/FURTHER READING

- Adedayo, O.A. (2006): Understanding statistics. JAS publishers Akoka Yaba.

- Esan, F.O. and Okafor, R.O. (2010): Basis statistical method (revised edition) Toniichristo Concept, Lagos.

- Murray, R.S. and Larry, J. S. (1998): (Schaum outlines series). Statistics (Third edition) MCGRAW HILLS.

- Olufolabo, O.O. and Talabi, C.O. (2002): Principles and practice of statistics. HASFEM Nig Enterprises Shomolu Lagos.

- Oyesiku, O.K. and Omitogun, O. (1999): Statistics for social and management sciences. Higher Education Books Publisher Lagos.

UNIT THREE; SAMPLING DISTRIBUTION OF SUM AND DIFFERENCE OF TWO MEANS

CONTENT

1.0 INTRODUCTION

This unit is an extension of unit one and unit two of this module. In this unit we are going to look at sampling distribution of sum and difference of two means. Since this unit is an offshoot of the unit one of this module, most of the statistical term used in unit one will be implied here.

2.0 OBJECTIVE

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

- Sampling distribution of sum of two means
- Sampling distribution of difference and

3.0 MAIN CONTENT

3.1 Sampling Distribution of Difference of Two Means and Sum $(\bar{\mathbf{x}} - \bar{\mathbf{x}})$

If two independent random sample of sizes n_1 and n_2 are selected from 2 different population of size N_1 and N_2 with population means μ_1 and μ_2 respectively and population variance σ_1^2 and σ_2^2 respectively, then the *sampling distribution of the difference of two means* $(\bar{x}_1 - \bar{x}_2) = \mu_{p1} - \mu_{p2}$ and standard deviation of the sample distribution is written as

$$\sigma_{x1-x2} = \sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}$$

Also the sampling distribution of sum of means is as defined below:

$\mu_{p1+p2} = \mu_{p1} + \mu_{p2}$ and the standard deviation

$$\sigma^2_{p1+p2} = \sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}$$

SELF-ASSESSMENT EXERCISE

Sampling distribution of the difference of two means is defined as-------

3.2 Worked Example of Sampling Distribution of Sum of Two Means

Given that $p_1 = (30,50)$ and $p_2 = (40,70)$ show that $\mu_{p1+p2} = \mu_{p1} + \mu_{p2}$;

(ii) $\mu_{p1-p2} = \mu_{p1} - \mu_{p2}$ and (iii) $\sigma^2_{p1+p2} = \sigma^2_{p1} + \sigma^2_{p2}$ for a sample drawn from each other.

Solution

Sampling sum

Possible sample combination = (30, 40), (30,70) (50,40) (50,70)

Sample sum = 30 + 40 = 70; 30 + 70 = 100; 50 + 40 = 90; 50 + 70 = 120

$$\therefore \mu_{p1+p2} = \frac{70 + 100 + 90 + 120}{4}$$

$$\mu_{p1+p2} = \frac{380}{4}$$

$$\mu_{p1+p2} = 95$$

Considering the 1st population p_2 (30,50)

$\mu_{p1} = \frac{30+50}{2}$

$\mu_{p1} = \frac{80}{2}$

$\mu_{p1} = 40$

Considering the 2nd population (40, 70)

$\mu_{p2} = \frac{40+70}{2}$

$\mu_{p2} = \frac{110}{2} \quad = 55$

$\therefore \mu_{p1} + \mu_{p2} = 55 + 40$

$\mu_{p1} + \mu_{p2} = 95$

Note $\mu_{p1 + p2} = 95$

$\mu_{p1} + \mu_{p2} = 95$

$\therefore \mu_{p1 + p2} = \mu_{p1} + \mu_{p2}$

SELF-ASSESSMENT EXERCISE

What is the population and sample mean of P_1 = (70,90), P_2 = (60,80)

3.3 Worked Example of Sample Differences of Two Means

$\mu_{p1} - \mu_{p2} = 40 – 55$ from our calculation of means above

$\mu_{p1} - \mu_{p2} = 15$

Taking the differences of possible =sample $\mu_{p1 – p2}$

$$\mu_{p1 - p2} = \frac{(30-40) + (30-70) + (50-40) + (50-70)}{4}$$

$$\mu_{p1 – p2} = \frac{10 + -40 + 10 – 20}{4}$$

$$\mu_{p1 – p2} = \frac{-10 – 40 + 10 – 20}{4}$$

$\mu_{p1-p2} = \frac{-60}{4}$

$\mu_{p1-p2} = -15$

$\therefore \mu_{p1-p2} = \mu_{p1} - \mu_{p2}$

$-15 = -15$

(iii) σ^2_{p1+p2} = variance of 70,10, 90 & 120

Note population mean = 95

$$\sigma^2_{p1+p2} = \frac{\Sigma(x - \bar{x})^2}{n}$$

$$\therefore \sigma^2_{p1+p2} = \frac{(70-95)^2 + (100-95)^2 + (90-95)^2 + (120-95)^2}{4}$$

$$\sigma^2_{p1+p2} = \frac{-25^2 + 5^2 + -5^2 + 25^2}{4}$$

$$\sigma^2_{p1+p2} = \frac{625 + 25 + 25 + 625}{4}$$

$$\sigma^2_{p1+p2} = \frac{1300}{4}$$

$\sigma^2_{p1+p2} = 325$

Considering the population independently

σ^2_{p1} = variance of (30,50)

$$\sigma^2_{p1} = \frac{(30-40)^2 + (50-40)^2}{2}$$

Where 40 = μp_1 = mean of population 1

$$\sigma^2_{p1} = \frac{(-10)^2 + (10)^2}{2}$$

$$\sigma^2_{p1} = \frac{100 + 100}{2}$$

$$\sigma^2_{p1} = \frac{200}{2}$$

$= 100$

Considering the 2nd population

σ^2_{p2} = variance of (40,70)

$$\sigma^2_{p2} = \frac{(40 - 55)^2 + (70 - 55)^2}{2}$$

Where 55 = mean of population = μp_2

$$\sigma^2_{p2} = \frac{15^2 + 15^2}{2}$$

$$\sigma^2_{p2} = \frac{225^2 + 225^2}{2}$$

$$\sigma^2_{p2} = \frac{450}{2}$$

$\sigma^2_{p2} = 225$

$\sigma^2_{p1} + \sigma_{p2} = 225 + 100 = 325$

$\sigma^2_{p1 + p2} = 325$

SELF-ASSESSMENT EXERCISE

Sampling distribution of the difference of 2 mean $\bar{x}_1$ & $\bar{x}_2$ is usually written as?

4.0 CONCLUSION

In the course of our discussion of this unit you have learnt about

- Sampling distribution of difference of two means
- Sampling distribution of sum of two means

5.0 SUMMARY

In the course of our discussion on this unit we defined sampling distribution of the difference of two mean as μp_1 - μp_2 and standard deviation of the difference as

$$rx_1 - r_2 = \sqrt{\frac{r_1^2}{n_1} + \frac{r_2^2}{n_2}}$$

6.0 TUTOR MARKED ASSIGNMENT

Given the following population $p_1 = (10,20)$ $p_2 = (30,40)$ show that

(i) $\mu p_1 + p_2 = \mu p_1 + \mu p_2$

(ii) $\mu p_1 – p_1 = \mu p_1 - \mu p_{2.}$

7.0 REFERENCE/FURTHER READING

- Adedayo, O.A. (2006): Understanding statistics. JAS publishers, Akoka, Yaba.

- Esan F.O. and Okafor, R.O. (2010): Basic statistical method (revised edition) Toniichristo Concept, Lagos.

- Murray, R. S. and Larry J. S. (1998): (Schaum outlines series). Statistics (Third edition) MCGRAW HILLS.

- Olufolabo, O.O. and Talabi, C.O. (2002): Principles and practice of statistics. HASFEM Nig Enterprises Shomolu, Lagos.

- Oyesiku, O.K. and Omitogun, O. (1999): Statistics for social and management sciences. Higher Education Books Publishers, Lagos.

UNIT FOUR; PROBABILITY DISTRIBUTION

CONTENTS

1.0 INTRODUCTION

Probability Defined

For thorough understanding of this unit, it is assumed that you must have familiarized yourself with introductory statistics and unit one of this module. The main thrust of this unit is to introduce to you the concept of probability distribution, its discussion, calculation and interpretation of result. This unit is fundamental to the understanding of subsequent modules. This is because other unit and module will be discussed on the basis of the fundamentals concept explained here.

2.0 OBJECTIVES

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

i. Concept of probability

ii. Different probability distribution

iii. Calculate the different probability distribution

3.0 MAIN CONTENT

3.1 Probability Defined

Statisticians spends quality time measuring data and drawing conclusions based on his measurement sometimes, all the data is available to the statisticians and the measurement are bound to be accurate in such circumstances, it can be said that the statistician has perfect knowledge of the population.

There are a-times whereby this will not be the usual situation. In most cases, the statistician will not have the details he wants about the population and will be unable to collect the information he wants because of cost and labour involved.

However, because the entire population has not been examined, the statistician can never be completely sure of the result, so when quoting conclusion based on sample evidence, it is usual to state how confident the statistician is about his result. So you will often see estimates quoted with 85% confidence. This is simply talking about the probability that the estimate is right is 85%.

The probability of a value X of a random variable is usually referred to as the limiting value of the relative frequency of that value as the total number of observation on the variable approaches infinity, the value which the relative frequency assumes at the limit as the number of observations tends to infinity. This can be written as

$$P(x) = \lim_{n\rightarrow} \frac{f}{\sum Fx}$$

SELF-ASSESSMENT EXERCISE

What is another name that probability can be called?

3.2 Probability Distribution of a Random Variable

If a variable is discrete, if its value are distinct i.e. they are separated by finite distance. To each we may assign a given probability. If x is a discrete random variable which may assume the values X_1, X_2 …..X_n with respective probabilities $f(x_1)$, $f(x_2)$ ……, $f(x_n)$. Then the entire set of pairs of permissible value together with their respective probabilities is called probability distribution of a random variable x.

A random variable is a variable whose values are associated with the probability of being observed. A discrete random variable is one that can assume only finite and distinct value.

One of the discrete probability is the binomial distribution. This distribution is used to find the probability of X number of occurrences or success of an event, P(x) in n-trials of same experiment.

Binomial distribution is usually use to predict occurrence of events that are mutually exclusive in other words Binomial distribution is useful for problem that are concerned with determining the number of times an event is likely to occur or not occur during a given number of trials and consequently the probability of it occurring or not occurring.

Symbolically it is written as;

$$P(x) = {}^{n}C_{x}\, P^{x} q^{n-x}$$

Alternatively

$$P(x) = \frac{n!}{X!(n-x)}\; p^{x}\,(1-p)^{n-x}$$

Where P = probability of a success in a simple trial probability of one event

q = 1-P, probability of the alternative to the event (failure)

n = number of times the event can occur in number trials

x = number of successes in n-trials

Mean of the binomial distribution is $\mu = np$ and standard deviation is

$$\sigma = \sqrt{np\,(1-p)} \quad \text{or}$$

$$\sigma = \sqrt{npq}$$

Example: What is the probability of obtaining 3 heads in 5 toss of a balanced coin. (b) What is the probability of obtaining less than 3 heads in 5 toss of coin.

Solution

Probability of obtaining a head = 1/2 = p

Probability of not obtaining ahead = 1-p = q = ½

X = 3, n = 5

(a) $P(x) = \frac{n!}{X!(n-x)!}$ p.^{x} (1-p)$^{n-x}$

$= \frac{n!}{X!(n-x)!}$ p.^{x} q^{n-x}

$P(x) = \frac{5x4x3x2x1}{3x2x1\ (5-3)!}$ ½.^{3} ½$^{5-3}$

$P(x) = \frac{5x4x3x2x1}{3x2x1\ (5-3)!}$ ½.^{3} ½^{2}

$P(x) = \frac{10}{1}$ x $\frac{1}{8}$ x $\frac{1}{4}$

$P(x) = \frac{10}{32}$

P(x) = 0.3125

:. The probability of obtaining 3 heads from 5 tosses of coin = 0.3125

(b) Probability of obtaining less than 3 heads = P(0) + P(1) + P(2)

:. $P(0) = \frac{5!}{0!(5-0)!}$ ½^{0} . ½ $^{5-0}$

$= \frac{5x4x3x2x1}{5x4x3x2x1}$ .1 . $\frac{1}{32}$

= 1/32

= 0.03125

$P(1) = \frac{5!}{1!\ (5-1)!} \left[\frac{1}{2}\right]^{1} \left[\frac{1}{2}\right]^{5-1}$

$= \frac{5x4x3x2x1}{4x3x2x1x1}$ ½ . ½^{4}

$= \frac{5}{1}$ x ½ x $\frac{1}{16}$

$= \frac{5}{32}$ = 0.15625

$P(2) = \frac{5!}{2!\ (5-2)!} \left[\frac{1}{2}\right]^{2} \left[\frac{1}{2}\right]^{5-2}$

$= \frac{5x4x3x2x1}{2x3x2x1x1}$ ¼ . $\frac{1}{8}$

$P(2) = \frac{10}{1} \text{ x } \frac{1}{32}$

$P(2) = \frac{10}{32}$

$P(2) = 0.3125$

$P(<3) = P(0) + P(1) + P(2)$

$P(<3) = 0.03125 + 0.15625 + 0.3125$

$P(<3) = 0.49625$

Mean = μ = $np = 5 (½)$

$np = 5/2$

$np = 2.5$ heads

Standard deviation = $\sigma = \sqrt{npq}$

$\sigma = \sqrt{5 \left(\frac{1}{2}\right)\left(\frac{1}{2}\right)}$

$\sigma = \sqrt{\frac{5}{4}}$

$\sigma = \sqrt{1.25}$

S.D. = σ = 1.1180339887499

= 1.12 heads

SELF-ASSESSMENT EXERCISE

What do you understand by the word a random variable?

3.3 Poisson Distribution

Poisson distribution is another discrete probability distribution useful in describing the number of events that will occur in a specific period of time. It is usually used in determining the probability of a designated number of successes per unit of time. When the event or successes are independent and the average number of successes per unit of time remains constant. Symbolically it is written as;

$P(x) = \underline{\lambda^x e\text{-}^{\lambda}}$

X!

Where P(x) = probability of x number of successes

X = number of success (0,1,2 ….)

λ = average or mean number of success or event that occur in a given internal

e = natural logarithms base whose value equal 2.71828

note λ = mean & variance of poisson distribution

$\sigma = \sqrt{\lambda}$

Example

A study shows that an average number of 6 customers per hour stop for fueling at a filling station.

(a) What is the probability of 3 customers fuelling at any hour?

(b) What is the probability of less than 3 customers, fueling in any hour?

Solution

(a) note mean = variance = λ = 6

$e = 2.71828$

$x = 3$

$$p(x) = \frac{\lambda^x e^{-\lambda}}{x!}$$

$$P(x=3) = \frac{6^3.(2.71828)^{-6}}{3x2x1}$$

$$P(x=3) = \frac{216 \text{ x } 0.00248}{6}$$

$$P(x=3) = \frac{0.53568}{6}$$

$P(x=3) = 0.08928$

(b) $P(x < 3)$ = Prob(0) + prob(1) + prob(2)

$$P(x=0) = \frac{6^o \text{ x } 2.71828^{-6}}{0!}$$

$$P(x=0) = \frac{1 \text{ x } 0.00248}{1}$$

Note 0! = 1

$P(x=0) = 0.00248$

$$P(x=1) = \frac{6^1 \text{ x } 2.71828^{-6}}{1!}$$

$$P(x=1) = \frac{6 \text{ x } 0.00248}{1}$$

$P(x=1) = 0.01488$

$$P(x=2) = \frac{6^2 \text{ x } 2.71828^{-6}}{2!}$$

$$P(x=2) = \frac{36 \text{ x } 0.00248}{2\text{x}1}$$

$P(x=2) = 18 \text{ x } 0.00248$

$P(x=2) = 0.04464$

:. Prob $(x>3) = P(0) + P(1) + P(2)$

$P(x<3) = 0.00248 + 0.01488 + 0.04464$

$P(x<3) = 0.062$

Mean = variance = 6 $=\lambda$

S.D$=\delta. = \sqrt{6}$

S.D.$=\delta = 2.449489743$

SELF-ASSESSMENT EXERCISE

Define standard deviation of Poisson distribution?

3.4 Probability Distribution of a Continuous Variable (Normal Distribution)

If a variable is continuous, it can assume an infinite number of values within a given interval. An important feature of probability distribution is that the areas under these curve represents probabilities. The total area under the curve of a probability distribution, being the sum of individual probabilities is equal to unity (1).

The normal distribution as a continuous probability distribution and the most commonly used distribution in statistical analysis. The normal curve is bell-shaped and symmetrical about its mean. Usually, it extends indefinitely in both directions, but most of the area (probability) is clustered around the mean.

To find the probabilities for problems involving the normal distribution, first convert the x value into corresponding z value using

$$Z = \frac{X - \mu}{\sigma}$$

Where μ = mean value

σ = standard deviation

Then check up the corresponding value from the normal distribution table.

Example

Given that family incomes are normally distributed with μ = ₦14,000 and δ = 4000. What is the probability that a family picked a random will have;

(a) Between ₦13,000 and ₦16,000 ?

(b) Below ₦13,000?

(c) Above ₦16,000 ?

(d) Above ₦18,000?

$$Z = \frac{X - \mu}{\sigma}$$

(a) Here x = 13,000 & 16,000

When x = 13,000; $Z_1 = \frac{13,000 - 14,000}{4,000}$

$$Z_1 = - \underline{1,000}$$

$$4{,}000$$

$$Z_1 = -\ 0.25$$

When X = 16,000; $Z_2 = \frac{16{,}000 - 14{,}000}{4{,}000}$

$$= \frac{2000}{4000}$$

$$= 0.5$$

$Z_1 = 0.25$; $Z_2 = 0.5$

$Z_{T1} = 0.0987$; $Z_{T2} = 0.1915$

Where Z_{T1} and Z_{T2} represents the table value for Z_1 and Z_2

:. Prob $(13{,}000 \leq x \leq 16{,}000) = 0.0987 + 0.1915$

:. Prob $(13{,}000 \leq x \leq 1{,}6000) = 0.2902$

$$= 29\%$$

(b) Prob $(x < 13{,}000) = 0.5 – 0.0987$

$$= 0.4013$$

$$\cong 40\%$$

(c) Prob $(x > 16{,}000) = 0.5 – 0.1915$

$$= 0.3085$$

$$\cong 30.85\%$$

(d) Prob $(x > 18{,}000)$

$(x = 18{,}000)$

$$Z = \frac{18{,}000 – 14{,}000}{4{,}000}$$

$$Z = \frac{40{,}000}{4000}$$

$$Z = 1$$

$$Z_T = 0.3413$$

:. Prob $(x > 18{,}000) = 0.5 – 0.3413$

$$= 0.1587$$

Prob $(x > 18{,}000) = 15.8\%$

$\cong 16\%$

SELF-ASSESSMENT EXERCISE

Explain the attributes of a normal distribution curve

4.0 CONCLUSION

From our discussion so far you have learnt about:

- Probability
- Probability distribution
- Different probability distribution, the binomial, Poisson, and normal distribution.

5.0 SUMMARY

In the course of our discussion of this unit, we have defined the different probability distributions binomial distribution is defined as

$$P(x) = {}^{n}C_{x}\, P^{x}\, q^{n-x}$$

Alternatively

$$P(x) = \frac{n!}{X!(n-x)!}\, P^{x}\,(1-p)^{n-x}$$

Where P = Probability of success

$q = 1 – P$ = probability of failure

mean = np, S.D = $\sigma = \sqrt{\text{npq}}$

Poisson distribution is defined as

$$P(x) = \frac{\lambda^{x}\ e^{-\lambda}}{x!}$$

λ = mean = variance

$\sqrt{\lambda}$ = standard deviation

Normal distribution

$$Z = \frac{X - \mu}{\sigma}$$

6.0 TUTOR MARKED ASSIGNMENT

A study shows that 40% of the people entering a supermarket make a purchase. Using (a) binomial distribution, (b) Poisson distribution find the probability that out of 30 people entering the supermarket 10 or more will make a purchase.

7.0 REFERENCE/FURTHER READING

- Adedayo, A.O. (2006): Understanding Statistics. JAS Publishers, Lagos.

- Dominick, S. and Derrick, R. (2011): Statistics and Econometrics. (Schaum's outlines) McGraw Hill, New York.

- Esan, E.O. and Okafor, R.O. (2010): Basic Statistical Methods (Revised Edition) Tonichristo Concept.

- Koutsoyianis, A. (2003): Econometric Methods (second edition). Palgrave publishers Ltd (formerly Macmillan press ltd), London and basin stoke.

- Murray, R. S. and Larry, J. S. (1998): Statistics (Schaum outlines). McGraw Hill.

- Olufolabo, O.O. and Talabi, C.O. (2002): Principles and practice of statistics, HASFEM (NIG) ENTERPRISES, Somolu, Lagos.

- Oyesiku, O.K. and Omitogun, O. (1999): Statistics for social and management sciences. Higher Education Book Publishers Lagos.

- Owen, F. and Jones R. (1983): Statistics. Polytech Publishers Ltd. Stockport.

MODULE TWO: Analysis of variance and analysis of covariance

UNIT ONE: ONE-WAY ANALYSIS OF VARIANCE

CONTENTS

1.0 INTRODUCTION

A detailed knowledge and understanding of introductory statistics is assumed, it is also expected that students would have familiarized themselves with hypothesis testing. This unit is one of the four units in module 2 of the course.

2.0 OBJECTIVE

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

- Total sum of square
- Sum of square between groups

- Sum of square within the group
- Mean square

3.0 LOGIC OF ANALYSIS OF VARIANCE (ANOVA)

Analysis of variance (Anova) is usually used to test null hypothesis that the means of two or more populations are equal versus the alternative that at least one of the means is different. The null hypothesis (H_o) tested in the case of ANOVA is that the means of the population from which the sample is drawn are all equal i.e. H_o, $\mu_1 = \mu_2 = \mu_3 = \ldots\ldots\ldots = \mu_n$ while the alternative hypothesis says that H_o taken as a whole is not true i.e. H_1; $\mu_1 \neq \mu_2 \neq \mu_3$.

It is to be noted that each time ANOVA is used, all we are trying to do is to analyze or test the variances in order to test the null hypothesis about the means (i.e. H_o; $\mu_1 = \mu_2 = \mu_3$). The ANOVA procedure is based on mathematical theory that the independent sample data can be made to yield two independent estimate of the population variance namely;

(i) Within group variance (or error) this is variance estimate which deals with how different each of the values in a given sample is from other values in the same group.

(ii) Between group variance this is estimate that deals with how the means of the various samples differs from each other.

SELF-ASSESSMENT EXERCISE

State the null hypothesis of analysis of variance?

3.1 Assumptions of Anova

(i) Observations are independent and value of any of observation should not be related to the value of another observation.

(ii) Homogeneity of sample variance, it should be assumed that the variance are equal for all treatment populations.

(iii) The values in the population are normally distributed.

SELF-ASSESSMENT EXERCISE

State one assumption of analysis of variance

3.2 Steps involved in Anova Analysis

(i) Estimate the population variance from the variance between sample means (MSA)

(ii) Estimate the population variance from the variance within the samples (MSE)

(iii) Compute the fisher ratio. This is given as $F = \frac{MSA}{MSE}$

i.e. $F = \frac{\text{Variance of between the sample mean}}{\text{Variance of within the sample}}$

(iv) Compute the various degree of freedom i.e. the degree of freedom for between, within and total groups.

Degree of freedom for the sum between group is given as C – 1

Degree of freedom within group is writer as (r – 1) c

Total degree of freedom as r – 1

Where c = no of samples

R = no of observations

(v) The next thing is to obtain the critical value of F statistics using the F-table in the table, we have the horizontal row which is for degree of freedom of the sum between group numerator. While, the vertical column is meant for within group, check the between degree of freedom along the horizontal axis and within group along vertical axis. This can be checked at either at 0.05 (5%) level of significance or 0.01(1%) level of significance.

(vi) Compare the F- statistic value with the critical value if the calculated value is less than the tabulated value, accept the null hypothesis (H_o) and concluded that the difference is not significant. If the calculated value is greater the critical value reject Ho and accept H_i the alternative hypothesis and conclude that the difference is significant.

(vii) The result is expected to be summarized on an ANOVA table.

Table M2.1.1

Analysis of Variance Table

Sources of variation	**Sum of squares**	**Degree of freedom**	**Mean square**	**I ratio**
Between the means (examples by Factor A)	$SSA = r\Sigma(\overline{x_J} - \bar{\bar{x}})^2$	C – 1	MSA = $\frac{\text{SSA}}{\text{C – 1}}$	$\frac{\text{MSA}}{\text{MSE}}$
Within the sample (error or unexplained)	$SSE = \Sigma\Sigma\left(xij - \overline{x_{\iota J}}\right)^2$	(r – 1)c	MSE = $\frac{\text{SSE}}{\text{(r-1)c}}$	-
Total	$SST = \Sigma\Sigma\left(x_{ij} - \bar{\bar{\mathrm{x}}}\right)^2 =$ SSA + SSE	rc – 1	-	-

Where $\overline{x_J} =$ mean of sample j composed of r observations $= \frac{\Sigma x_{ij}}{\mathrm{r}}$

$$\bar{\mathrm{x}} = grand\ mean\ of\ all\ c\ samples\ = \frac{\Sigma_i \Sigma_j x_{ij}}{\mathrm{rc}}$$

SSA = Sum of square explained by factor A = $r\Sigma(\bar{\mathrm{x}} - \bar{\bar{x}})^2$

SSE = Sum of square of error unexplained by factor A = $\Sigma\Sigma(xij - x^{-})^2$

SST = Total Sum of squares = SSA + SSE = $\Sigma\Sigma\left(x_{ij} - \bar{\bar{x}}_J\right)^2$

Where c = no of samples

r = no of observations in each sample

SELF-ASSESSMENT EXERCISE

State the fisher ratio

3.4 Worked Example `

The information below relates to quantities of plastic produced by a plastic industry in 3 sections (morning, afternoon and evening) for 5 weeks. The production data are normally distributed with equal variance.

Table M2.1.2

Table showing production of a plastic industry

Weeks	**Morning (X_1)**	**Afternoon (X_2)**	**Evening (X_3)**
1	85	77	90
2	83	81	92
3	79	75	84
4	81	82	82
5	82	80	87

Is there any significant difference due to production session?

Test at 5% level of significance.

Solution

$H_o; \mu_1 = \mu_2 = \mu_3$

$H_i; \mu_1 \neq \mu_2 \neq \mu_3$

Note let the quantities produced in morning be represented by X_1, afternoon X_2, evening X_3.

$\Sigma X_1 = 410$

$$\bar{x}_1 = \frac{\Sigma X_1}{r} = \frac{410}{5} = 82$$

where r = number of weeks

$\Sigma X_2 = 395$

$\overline{x_2} = \frac{\Sigma X_2}{r} = \frac{395}{5} = 79 \cong 79$

$\Sigma X_3 = 435$

$\overline{x_3} = \frac{\Sigma X_3}{r} = \frac{435}{r} = \frac{435}{5} = 87 \cong 87$

$$\overline{x} = \frac{410 + 395 + 435}{(5)(3)}$$

$= \frac{1240}{15} = 82.66667 = 82.67$

$\cong 83$

SSA $= 5[(82 - 82.67)^2 + (79 - 82.67)^2 + (87 - 82.67)^{2]}$

$= 5[(-0.67)^2 + (-3.67)^2 + (4.33)^2]$

$= 5(0.4489 + 13.4689 + 18.7489)$

$= 5(32.667)$

$= 163.3335$

SSE $= \Sigma\Sigma\left(x_{ij} - \overline{x_j}\right)^2$

$= (85 - 82)^2 + (83 - 82)^2 + (79 - 82)^2 + (81 - 82)^2 + (82 - 82)^2 + (77 - 79)^2$

$+ (81 - 79)^2 + (75 - 79)^2 + (82 - 79)^2 + (80 - 79)^2 + (90 - 87)^2 + (92 - 87)^2$

$+ (84 - 87)^2 + (82 - 87)^2 + (87 - 87)^2$

$= (3)^2 + (1)^2 + (-3)^2 + (-1)^2 + 0^2 + (-2)^2 + (2)^2 + (-4)^2\ (3)^2 + (1)^2 + (3)^2 + (5)^2 +($-

$3)^2 + (-5)^2 + 0$

$= 9 + 1 + 9 + 1 + 0 + 4 + 4 + 16 + 9 + 1 + 9 + 25 + 9 + 25 + 0$

$= 122$

SST $= (85 - 82.67)^2 + (83 - 82.67)^2 + (79 - 82.67)^2 + (82 - 82.67)^2 +$

$(77 - 82.67)^2 + (81 - 82.67)^2 + (75 - 82.67)^2 + (82 - 82.67)^2 +$

$(80 - 82.67)^2 + (90 - 82.67)^2 + (92 - 82.67)^2 + (84 - 82.67)^2 +$

$(82 - 82.67)^2 + (87 - 82.67)^2$

$$= (2.33)^2 + (0.33)^2 + (-3.67)^2 + (1.67)^2 + (0.67)^2 + (-5.67)^2 + (1.67)^2 + (-7.67)^2 + (0.67)^2 + (2.67)^2 + (7.33)^2 + (9.33)^2 + (1.33)^2 + (0.67)^2 + (4.33)^2$$

$$= 5.4289 + 0.1089 + 13.4689 + 2.7889 + 0.4489 + 32.1489 + 58.8289 + 2.7889 + 0.4489 + 7.1289 + 53.7289 + 87.0489 + 1.7689 + 0.4489 + 18.7489$$

$$= 285.3335$$

Table M2.1.3

One-Way Analysis of Variance Table

Sources of variation	**Sum of squares**	**Degree of freedom**	**Mean square**	**I ratio**
Explained variation (between column)	$SSA = 163.3335$	3-1 =2	MSA = $\frac{163.335}{2}$ = 81.66675	$\frac{81.66675}{10.167}$ = 8.0325
Unexplained variation or error (within column)	$SSE = 122$	(5 – 1)3 = (4)3 = 12	MSE = $\frac{122}{12}$ = 10.167	
Total	285.3335	rc – 1 = 14	-	

Note Sum of Square

$SSA = r\Sigma(\overline{x_j} - \bar{\bar{x}})^2$

$$SSE = \Sigma\Sigma\left(xij - \overline{x_{\iota J}} \right)^2$$

$SST = \Sigma\Sigma\left(x_{ij} - \bar{\bar{\mathrm{x}}}\right)^2$

Degree of Freedom

Explained variation = c – 1 Where c = number of samples

Unexplained variation = (r – 1) c r = number of weeks

Total variation = rc – 1

MEAN SQUARE

$$MSA = \frac{SSA}{c-1}$$

$$MSE = \frac{SSE}{(r-1)c}$$

$$F\text{-ratio} = \frac{MSA}{MSE}$$

$F_{0.05}(2,12) = 3.88$ (Critical value)

Source: F distribution table

Decision

Accept Hi, reject H_o because $F_{cal} > F_{tab}$ which implies that there is significant difference between the mean of production sessions.

Self-assessment exercise

State the formulae for sum of square?

4.0 CONCLUSION

In the course of our study of one-way analysis of variance you must have learnt about;

- Explained variation
- Unexplained variation
- Total variation

5.0 SUMMARY

In the course of our discussion of one-way analysis of variation the following definitions were inferred

$SSA = r\Sigma(\bar{x}_j - \bar{\bar{x}})^2$

$$SSE = \Sigma\Sigma\left(xij - \overline{x_{\iota J}} \right)^2$$

$SSJ = \Sigma\Sigma\left(x_{ij} - \bar{\bar{\mathrm{x}}}\right)^2$

6.0 TUTOR MARKED ASSIGNMENT

Submit a one page essay on the definition of degree of freedom for explained variation, unexplained variation and total variation.

7.0 REFERENCES/FURTHER READIDINGS

- Adedayo, O. A. (2006): Understanding Statistics: JAS Publishers, Akoka, Lagos.
- Dominick, S. and Derrick, R. (2011): Statistics and Econometrics, (Schaum Outlines) McGraw-Hill Company, New York.
- Edward, E.L. (1983): Statistical analysis in Economics and Business. Houghton Mufflin Company, Boston.
- Olufolabo, O.O. and Talabi, C.O. (2002): Principles and Practice of Statistics; HAS-FEM ENTERPRISES Somolu, Lagos.
- Owen, F. and Jones, R. (1978): Statistics, Polytech Publishers Ltd, Stockport.

UNIT TWO: TWO-WAY ANALYSIS OF VARIANCE

CONTENTS

1.0 INTRODUCTION

This unit is an extension of unit one, the difference between them is that, here, we can test for two (2) null hypothesis, one for factor A and the other for factor B.

2.0 OBJECTIVE

At the end of this unit, you should be able to test for two null hypothesis.

Ho; $U_{a1} = U_{a2} = U_{a3}$

Ho; $U_{b1\neq} U_{b2} \neq U_{b3}$

3.0 MAIN CONTENT

3.1 Two- Way Analysis of Variance Defined

For two way analysis, the set of observation involved are classified into two (2) factors or criteria; treatment factor or criteria and block or homogenous factor or criteria.

As we have discussed in one factor- analysis of variance, the total variation is divided or splitted into 3 components.

- Variation between treatment
- Variation between blocks and
- Residual or error variation

SELF-ASSESSMENT EXERCISE

State the divisions into which total variation is divided into?

3.2 Two-way Classification

Table M2.2.1

Two-way classification table

	Treatment (Factor A)		
		1 2 3 …………… t	Total
Block factor B	1	Y_{11} Y_{12} Y_{13} ………………… Y_{1j}	B_1
	2 3 4 5 “ “ “ “	Y_{21} Y_{22} Y_{23} ………………… Y_{2t}	B_2
	B	Y_{b1} Y_{b2} Y_{b3} …………………. Y_{bt}	B_b

3.3 The Formulars

(i) Column means is given by $\frac{\Sigma x_{ij}}{r}$

Row means of given $\frac{\Sigma x_{ij}}{c}$

Grand mean is given by $\bar{\bar{x}} = \Sigma \frac{\overline{x_{\iota\cdot}}}{r} = \Sigma \frac{\overline{x_{\iota}}}{c}$

The subscripted dot signifies that more than one factor is under consideration.

SST = $\Sigma\Sigma\left(x_{ij} - \bar{\bar{x}}_{J}\right)^2$

SSA = $r\Sigma(\overline{\mathrm{x\imath\jmath}} - \bar{\bar{x}})^2$ between column variation

SSB = $c\Sigma(\bar{x}_{\iota} - \bar{\bar{\mathrm{x}}})^2$ between row variation

SSE = SST – SSA – SSB

Degree of freedom of SSA = c – 1

Degree of freedom of SSB = r – 1

Degree of freedom of SSE = (r-1) (c – 1)

Degree of freedom of SST = rc – 1

Mean Square

$$MSA = \frac{SSA}{c - 1}$$

$$MSB = \frac{SSB}{r - 1}$$

$$MSE = \frac{SSE}{(r-1)(c-1)}$$

F- statistics

$$\text{F-ratio for factor A} = \frac{MSA}{MSE}$$

$$\text{F-ratio for factor B} = \frac{MSB}{MSE}$$

It is to be noted that; two (2) separate null hypothesis is considered.

(i) Ho; There is no difference between mean of treatment

(ii) Ho; There is no difference between mean of block.

SELF-ASSESSMENT EXERCISE

State the formulae for column mean?

3.4 Worked Example

Samples taken involving two (2) interactive factors A & B in a two analysis of variance experience gives the result below:

TableM2.2.2

Table showing interactive factors A and B

	Treatment A			
Block (B)	22	11	10	5
	13	10	8	6
	7	9	6	2

You are carry out a 2-way analysis of variance at 0.05 level of significance?

Solution

Hypothesis

1. Ho; $\mu_1 = \mu_2 \; \mu_3 = \mu_4$; H_1; $\mu_1 \neq \mu_2 = \mu_3 = \mu_4$
2. Ho; $\mu_1 = \mu_2 = \mu_3$; H_1; $\mu_1 \neq \mu_2 \neq \mu_3$

Table M2.2.3

Two-Way Classification Table

	Treatment A				**Total**	**Sample mean**
Block B	22 13 7	11 10 9	10 8 6	5 6 1	48 37 23	$\bar{x}_1 = 12$ $\bar{x}_2 = 9.25$ $\bar{x}_3 = 5.75$
Total	42	30	24	12	108	$\Sigma\bar{x}_i = 27$
Sample mean	42/3 $x._1 = 14$	30/3 $x._2 = 10$	24/3 $x._3 = 8$	12/3 $x._4 = 4$		$\bar{\bar{x}} = 9$

SST = $\Sigma\Sigma\left(x_{ij} - \bar{\bar{x}}_j\right)^2$

$(22 – 9)^2 = (13)^2 = 169;$

$(13 – 9)^2 = (4)^2 = 16;$

$(7 – 9)^2 = (-2)^2 = 4;$

$= 189$

$(11 – 9)^2 = (2)^2 = 4;$

$(10 – 9)^2 = (1)^2 = 1;$

$(9 – 9)^2 = (0)^2 = 0;$

$= 5$

$(10 – 9)^2 = (1)^2 = 1$

$(8 – 9)^2 = (-1)^2 = 1$

$(6 – 9)^2 = (-3)^2 = 9$

$= 11$

$(5 – 9)^2 = (-4)^2 = 16;$

$(6 – 9)^2 = (-3)^2 = 9;$

$(1 – 9)^2 = (-8)^2 = 64$

$= 89$

:. SST = 189 + 5 + 11 + 89 = 294

SSA = $r\Sigma(\bar{x} - \bar{\bar{x}})^2$ where r = no of column

= $3 [(14 – 9)^2 + (10 – 9)^2 + (8 – 9)^2 + (4-9)^2]$

= $3 [5^2 + (1)^2 + (-1)^2 + (-5)^2$

= 3 (25 + 1 + 1 + 25)

= 3 (52)

= 156

SSB = $c\Sigma(\overline{x_\iota} - \bar{\bar{x}})^2$ Where c = number of row

= $4[(12 – 9)^2 + (9.25 – 9)^2 + (5.75 – 9)^2]$

= $4 [(3)^2 + (0.25)^2 + (-3.25)^2]$

= 4 (9 + 0.0625 + 10.5625)

= 4 (19.625)

= 78.5

SSE = SST – SSA – SSB

= 294-156 – 78.5

= 59.5

Degree of Freedom

SSA = c – 1 = 4 -1 = 3

SSB = r – 1 = 3 – 1 = 2

SSE = (r-1) (c – 1) = (3 – 1) (4 – 1) = (2) 3 = 6

SST = rc – 1 = (4 x 3) – 1 = 12 – 1 = 11

Mean Square

$$MSA = \frac{SSA}{c-1} = \frac{156}{4-1} = \frac{156}{3} = 52$$

$$MSB = \frac{SSB}{r-1} = \frac{78.5}{3-1} = \frac{78.5}{2} = 39.25$$

$$MSE = \frac{SSE}{(r-1)(c-1)} = \frac{59.5}{(3-1)(4-1)} = \frac{59.5}{(2)(3)} = \frac{59.5}{6} = 9.916666667$$

F-ratio

$$\frac{MSA}{MSE} = \frac{52}{9.916667} = 5.243697303$$

$$\frac{MSB}{MSE} = \frac{39.25}{9.9166667} = 3.95798318$$

Table M2.2.4

Two-ways / Two Factor Analysis of Variance

Sources of variation	**Sum of squares**	**Degree of freedom**	**Mean square**	**E ratio**
Explained variation by factor A (between column)	$SSA = 156$	C – 1 = 3	MSA = 52	$\frac{MSA}{MSE}$ = 5.24370
Explained variation by factor B (between rows)	$SSB = 78.5$	r – 1 = 2	MSB = 39.25	$\frac{MSB}{MSE}$ = 3.95798
Unexplained variation or error	$SSE = 59.5$	(r – 1)(c-1) = 6	MSE= 9.91667	-
Total	294	11	-	-

Decision Criteria Test 1

1. Factor A Critical Value

$F_{3,6}$ = 4.76

Because $F_{cal.} > F_{tab.}$ Reject H_o and accept H_1 meaning that the mean of factor A are not equal.

Test II

2. Factor B Critical Value

$F_{2,6}$ = 5.14

Since $F_{cal.} < F_{tab.}$ Accept H_o and reject H_1 conclude that the mean of factor B are all equal.

SELF-ASSESSMENT EXERCISE

State the decision criteria for accepting or rejecting hypothesis?

4.0 CONCLUSION

In the course of our discussion on two-way analysis of variance, we have learnt about:

(i) Sum of square of Factor A

(ii) Sum of square of Factor B

(iii) Sum of square of the error term

(iv) Mean square of Factor A

(v) Mean square of Factor B

(vi) F-ratio of both Factor A and Factor B

(vii) Sum of Square of total variation.

5.0 SUMMARY

In our discussion the following definition were inferred to:

(i) SST = $\Sigma\Sigma\left(x_{ij} - \overline{\overline{x_J}}\right)^2$

(ii) SSA = $r\Sigma(\overline{\mathrm{x}}\mathrm{j} - \bar{\bar{x}})^2$

(iii) SSB = $c\Sigma(\overline{x_\iota} - \bar{\bar{\mathrm{x}}})^2$

(iv) SSE = SST – SSA – SSB

(v) MSA = $\frac{SSA}{c - 1}$

(vi) MSB = $\frac{SSB}{r-1}$

(vii) MSE = $\frac{SSE}{(r-1)(c-1)}$

(viii) F-ratio for

Factor A = $\frac{MSA}{MSE}$

Factor B = $\frac{MSB}{MSE}$

6.0 TUTOR MARKED ASSIGNMENT

Submit a one page essay on the definition of MSE, SST and F-ratio.

7.0 REFERENCES/FURTHER READING

- Adedayo, O.A. (2006): Understanding Statistics: JAS Publishers, Akoka, Lagos.

- Dominick, S. and Derrick, R. (2011): Statistics and Econometrics, (Schaum's Outlines) McGraw-Hill Company, New York.

- Edward, E. L. (1983): Statistical analysis in Economics and Business. Houghton Mifflin Company. Boston.

- Olufolabo, O.O. and Talabi, C.O. (2002): Principles and Practice of Statistics; HAS-FEM ENTERPRISES Somolu, Lagos.

- Owen, F. and Jones, R. (1978): Statistics, Polytech Publisher Ltd, Stockport.

UNIT THREE: ANALYSIS OF COVARIANCE

CONTENTS

1.0 INTRODUCTION

In general, research is conducted for the purpose of explaining the effect of the independent variable on the dependent variable, and the purpose of research design is to provide a structure for the research. In the research design, the researcher identifies and controls independent variable that can help to explain the observed variation in the dependent variable which in turn reduces error variables (unexplained variation).

In addition to controlling and explaining variation through research design, it is also possible to use statistical control to explain the variation in the dependent variable, statistical control is usually used when experimental control is difficult, if not impossible, can be achieved by measuring one or more variable in addition to the independent variable of primary interest and by controlling the variation attributed to these variables through statistical analysis rather than through research design. The analysis procedure employed in this statistical control is analysis of covariance (ANCOVA).

2.0 OBJECTIVE

At the end of our discussion on analysis of covariance, you should be able to;

- Define analysis of variance
- Define covariate
- Define adjusted Y_{is}
- Develop table of analysis of covariance
- Calculate the various terms that may be needed on the computation of ANCOVA Table

3.0 MAIN CONTENT

3.1 Analysis of Variance Defined

Analysis of covariance is an extension of the one-way analysis of variance that added quantitative variable (covariate) when used, it is assumed that their inclusion will reduce the size of the error variance and thus increase the power of the design. Analysis of covariance (ANCOVA) is a statistical test related to analysis of variance (ANOVA). It tests whether there is a significant difference between groups after controlling for variance explained by a covariate.

A covariate is a continuous variable that correlates with the dependent variable. This means that you can, in effect, "partial out" a continuous variable and run an ANOVA on the result.

This is one way that you can run a statistical test with both categorical and continuous independent variables.

The purpose of analysis of covariance is to remove one or more unwanted factor or variables in the analysis. A variable whose effect one wishes to eliminate by means of a covariance analysis called a covariate sometimes called concomitant variable.

ANCOVA works by adjusting the total sum of square, group sum of squares and error sum of square of the independent variable to remove the influence of the covariate.

ASSUMPTIONS OF ANALYSIS OF COVARIANCE

- Variance is normally distributed
- Variance is equal between group
- All measure are independent
- Relationship between dependent variable and the covariate as linear
- The relationship between the dependent variable and the covariate is the same for all groups.

Self-assessment exercise

Why analysis of covariance?

3.2 Estimation of ANCOVA

Hypothesis for ANCOVA

- H_o and H_i need to be stated slightly different for an ANCOVA than a regular ANOVA.

H_o: the group means are equal after controlling for the covariate

H_i: the group means are not equal after controlling for the covariate

Below are the lists of notation for the calculation of ANCOVA.

$$\mathbf{S_{yy}} = \sum_{i=1}^{n}\sum_{j=1}^{n}(\mathbf{Y} - \bar{\bar{\mathbf{Y}}})^{2} = \sum_{i-1}^{n}\sum_{j=1}^{n} \mathbf{Y}^{2} - \Sigma\mathbf{Y}^{2}/\ \mathbf{an}$$

$$\mathbf{S_{xx}} = \sum_{i=1}^{n}\sum_{j=1}^{n}(\mathbf{X} - \bar{\bar{\mathbf{X}}})\mathbf{2}$$

$$= \sum_{i-1}^{n}\sum_{j=1}^{n}(\mathbf{X}^{2} - \Sigma\mathbf{Yij}^{2}/\ \mathbf{an})$$

$$\mathbf{S_{xy}} = \sum_{i=1}^{n}\sum_{j=1}^{n}(\mathbf{X_{ij}} - \bar{\bar{\mathbf{X}}})\ (\mathbf{Y_{ij}}\ \_\ \bar{\bar{\mathbf{Y}}})$$

$$= \sum_{i=1}^{n}\sum_{j=1}^{n} \mathbf{X_{ij}\ Y_{ij}\ -\ \Sigma X_{ij}\ \Sigma Y_{ij}\ /\ an}$$

$$= \sum_{i=1}^{n}\sum_{j=1}^{n} \mathbf{X_{ij}\ Y_{ij}\ -\ \Sigma X_{ij}\ \Sigma Y_{ij}\ /\ an}$$

$$\mathbf{T_{yy}} = \sum_{i=1}^{n} (\bar{\mathbf{Y}}_{\mathbf{i}} - \bar{\bar{\mathbf{Y}}}_{\mathbf{i}})^{\mathbf{2}}$$

$$= \sum\nolimits_{\mathbf{i=1}} \left(\frac{Y_i^2}{n} - \frac{\Sigma Y^2}{an} \right)$$

$$\mathbf{T_{xx}} = \sum_{i=1} (\bar{\mathbf{X}}_{\mathbf{i}} - \bar{\bar{\mathbf{X}}}_{..})^{\mathbf{2}}$$

$$= \Sigma\ \frac{X_i^2}{n} - \frac{\Sigma X^2}{an}$$

$$\mathbf{T_{xy}} = \sum_{i=1} (\bar{\mathbf{X}}_{\mathbf{ij}} - \bar{\bar{\mathbf{X}}}_{\mathbf{n}})\ (\bar{\mathbf{Y}}_{\mathbf{ij}} - \bar{\bar{\mathbf{Y}}}_{..})$$

$$= \Sigma\ \bar{X}\ \bar{Y} - \frac{\Sigma X \Sigma Y}{an}$$

$$\mathbf{E_{yy}} = \sum_{i=1}^{n}\sum_{j=1}^{n} (\mathbf{Y_{ij}} - \bar{\mathbf{Y}})^{\mathbf{2}}$$

$$= S_{yy} - T_{yy}$$

$$\mathbf{E}_{\boldsymbol{XX}} = \sum_{i=1}^{n}\sum_{j=1}^{n} (\mathbf{X_{ij}} - \bar{\boldsymbol{X}})^{\mathbf{2}}$$

$$= S_{XX} - T_{XX}$$

$$\mathbf{E_{xy}} = \sum_{i=1}^{n}\sum_{j=1}^{n} (\mathbf{X_{ij}} - \bar{\mathbf{X}}_{\mathbf{i}})\ (\mathbf{Y_{ij}} - \bar{\mathbf{Y}}_{\mathbf{i}})$$

$$= S_{xy} - T_{xy}$$

S = T + E

Where $\bar{X}$ = mean of X

$\bar{\bar{X}}$ = Grand mean of X

$\bar{\bar{Y}}$ = Grand mean of Y

a = variable involved

n = no of observations

Where the symbols S,T and E are used to denote sum of square and cross product for total, treatment and error respectively.

Table M2.3.1

Analysis of Covariance for a Single Factor Experiment with One Covariate

Source of variation	**Df**	**Sum of square and product**			**Adjusted Regression Y**	**df**	**Mean square error (MSE)**
		X	XY	Y			
Treatment	a – 1	T_{xx}	T_{xy}	T_{yy}			
Error	a (n-1)	E_{xx}	E_{xy}	E_{yy}	$SSE = E_{yy} - \frac{(E_{xy})^2}{E_{xx}}$	a(n-1)-1	$\frac{SSE}{a(n-1)-1}$
Total	(an-1)	S_{xx}	S_{xy}	S_{yy}	$SS^1E = S_{yy} - \frac{(S_{xy})^2}{S_{xx}}$	an-2	
Adjusted Treatment					$SS^1E - SSE$	a-1	$\frac{SS^1E - SSE}{a - 1}$

$$F_o = \text{Fstatistics} = \frac{Exy^2/Exx}{MSE}$$

$$F_c = \frac{(SS'E - SSE) / (a-1)}{SSE / (a(n-1)-1)}$$

Which is distribute as

$F_{a-1,a(n-1)-1}$

Decision criteria

Reject Ho if Fc > $F_{\propto 1}$, a(n-1)-1

Worked Example

A soft drink distributor is studying the effectiveness of delivery methods. Three different types of truck have been developed, and an experiment is performed in the company's laboratory. The variable of interest is the delivery time in minute (Y): however, delivery time is also strongly related to the case volume delivered (X). Each truck is used four times and the data below are obtainable.

Table M2.3.2

Table Showing Delivery Method of a Distributor

Truck Types					
1		**2**		**3**	
Y	X	Y	X	Y	X
27	24	25	26	40	38
44	40	35	32	22	26
33	35	46	42	53	50
41	40	26	25	18	20
$\Sigma Y_1 = 145$	$\Sigma Y_1 = 139$	$\Sigma Y_2 = 132$	$\Sigma Y_2 = 125$	$\Sigma Y_3 = 133$	$\Sigma Y_3 = 134$

Solution

$\overline{Y}_1 = \frac{145}{4} = 36.25$ $\qquad$ $\overline{X}_1 = \frac{139}{4} = 34.75$

$\overline{Y}_2 = \frac{132}{4} = 33$ $\qquad$ $\overline{X}_2 = \frac{125}{4} = 31.25$

$\overline{Y}_3 = \frac{133}{4} = 33.25$ $\qquad$ $\overline{X}_3 = \frac{134}{4} = 33.5$

$$\overline{\overline{X}} = \underline{139 + 125 + 134}$$

12 = 33.167

$H_o = T_1 = T_2 = \ldots. = T_n = 0$

$H_i = T_1 \neq T_2 \neq \ldots.. \neq T_n = 0$

$$\mathbf{S_{yy}} \quad = \sum_{\mathbf{i=1}}^{\mathbf{n}} \sum_{\mathbf{j=1}}^{\mathbf{n}} (\mathbf{Y_{ij}2 - \Sigma Y^2/an})$$

a = 3

n = 4

$S_{yy} = 27^2 + 44^2 + 33^2 + 41^2 + 25^2 + 35^2 + 46^2 + 26^2 + 40^2 + 22^2 + 53^2 + 18^2 – 410^2 /3x4$

= 729 + 1936 + 1089 + 1681 + 625 + 1225 + 2116 + 676 + 1600 + 484 + 2809 +324 - $(410)^2$ / 12

S_{yy} = 15,294 – 168,100 /12

S_{yy} = 15294 – 14,008.33

S_{yy} = 1,285.6711

$S_{xx} = 24^2 + 40^2 + 35^2 + 40^2 + 26^2 + 32^2 + 42^2 + 25^2 + 38^2 + 26^2 + 50^2 + 20^2 – (398^2 /(3x4))$

S_{xx} = 576 + 1600 + 1225 + 1600 + 676 + 1024 + 1764 + 625 + 1444 + 676 + 2500 + 400 – (158404/12)

S_{xx} = 14,110 – 13,200.333

S_{xx} = 909.6666711

$$\mathbf{S_{xy}} \quad = \sum_{\mathbf{i=1}}^{\mathbf{n}} \sum_{\mathbf{j=1}}^{\mathbf{n}} (\mathbf{XY - \Sigma X \Sigma Y/an})$$

S_{xy} = (27x24) + (44x40) + (33x35) + (41x40) + (25x26) + (35x32) + (46x42) + (26x25) + (40x38) + (22x26) + (53x50) + (18x20) – ((410) (398) /12)

S_{xy} = 648 +1760 + 1,155 + 1640 + 650 + 1120 + 1932 + 650 + 1520 + 572 + 2650 + 360 – (163180 /12)

S_{xy} = 14,657 – 163,180 /12

$S_{xy} = \ 14{,}657 – 13{,}598.333$

$S_{xy} = 1{,}058.67$

$$T_{yy} = \sum_{i=1}^{n} \frac{Yi}{a} - \frac{(\Sigma Y)^2}{an}$$

$$T_{yy} = \frac{145^2 + 132^2 + 133^2}{4} - \frac{410^2}{3x4}$$

$$T_{yy} = \frac{21{,}025 + 17{,}424 + 17{,}689}{4} - \frac{168{,}100}{12}$$

$$T_{yy} = \frac{56{,}138}{4} - \frac{168100}{12}$$

$T_{yy} = \ 14{,}034.5 – 14{,}008.33$

$T_{yy} = \ 26.1667$

$$T_{xx} = \sum_{i=1}^{n} (\bar{X}\,\mathbf{2} - \frac{(\Sigma X)^2}{an})$$

$$T_{xx} = \frac{\Sigma X^2}{n} - \frac{(\Sigma X)^2}{an}$$

$$T_{xx} = \frac{139^2 + 125^2 + 134^2}{4} – \frac{398^2}{3 \times 4}$$

$$T_{xx} = \frac{19{,}321 + 15{,}625 + 17{,}956}{4} – \frac{158{,}404}{12}$$

$$T_{xx} = \frac{52902}{4} - \frac{158{,}404}{12}$$

$T_{xx} = \ 13{,}225.5 \ - 13{,}200.333$

$T_{xx} = 25.1667$

$$\mathbf{T_{xy}} = \sum_{\mathbf{i=1}}^{\mathbf{n}} \bar{\mathbf{X}}_{\mathbf{ij}} \ \bar{\mathbf{Y}}_{\mathbf{ij}} – (\mathbf{\Sigma}\ \bar{\bar{\mathbf{X}}}\mathbf{\Sigma}\ \bar{\bar{\mathbf{Y}}})$$

$$T_{xy} \quad = \sum_{i=1}^{n} \frac{XiY}{n} - \frac{\Sigma X \Sigma Y}{an}$$

$$T_{xy} = \frac{(145 \text{ x } 139) + (132 \text{ x } 125) + (133 \text{ x } 134)}{4} - \frac{(410)(398)}{12}$$

$$T_{xy} = \frac{20{,}155 + 16{,}500 + 17{,}822}{4} - \frac{163{,}810}{12}$$

$$T_{xy} = \frac{54{,}477}{4} - \frac{163{,}810}{12}$$

$$T_{xy} = 13{,}619.25 – 13598.333$$

$$T_{xy} = 20.91667$$

$$E_{yy} = S_{yy} – T_{yy}$$

$$E_{yy} = 1285.6667 – 26.1667$$

$$E_{yy} = 1259.5$$

$$E_{xx} = S_{xx} – T_{xx}$$

$$E_{xx} = 909.667 - 25.1667$$

$$E_{xx} = 884.5$$

$$E_{xy} = S_{xy} – T_{xy}$$

$$E_{xy} = 1058.67 – 20.9167$$

$$E_{xy} = 1037.753$$

$$SS^1E = S_{yy} - \frac{(S_{xy})^2}{S_{xx}}$$

$$SS^1E = 1285.67 – \frac{(1{,}058.67)^2}{909.667}$$

$$SS^1E = 1285.67 – \frac{1{,}120{,}782.169}{909.667}$$

$$SS^1E = 1285.67 – 1{,}232.08$$

$SS^1E = 53.59038$

$SS^1E \cong 53.59$

with (an – 2) df = 12 – 2 = 10df

$$SSE = E_{yy} = \frac{(E_{xy})^2}{E_{xx}}$$

$$SSE \ = 1259.5 - \frac{(1037.753)^2}{884.5}$$

$$SSE \ = 1259.5 - \frac{1,076,931.912}{884.5}$$

SSE = 1259.5 – 1217.560104

SSE = 41.939896

SSE = 41.94

with a (n-1)-1) df = 3(4-1) – 1

= 3(3) – 1

= 9 – 1

= 8 d.f.

SS^1E – SSE = 53.59 – 41.94

= 11.65

with a – 1 df = 3 – 1 = 2 .d.f.

All the above calculations can be summarized in an ANCOVA Table, as presented below

Table M2.3.3

Analysis of Covariance (ANCOVA) Table

Source of variation	**d.f**	**Sum of square and product**			**Adjusted Regression**	**d.f**	**Mean Square Error**
		X	XY	Y			
Treatment	(3-1) 2	25.1667	20.91667	26.1667			
Error	3(4-1) 9	884.5	1029.753	1259.5	41.94	3(4-1)-1 8	5.2425
Total	(12-1) 11	909.667	1058.65	1285.67	53.59	(12-2) 10	
Adjusted Treatment					11.65	2	5.825

$$F_{statistics} = Fc = \frac{SS^1E - SSE \mid (a\text{-}1)}{SSE \mid a\,(n\text{-}1) - 1} = \frac{11.65/2}{53.59/8}$$

$$F_c = \frac{5.825}{6.69875}$$

F_c = 0.869565217

F_c = 0.9

F_{tab} = $F_{2,8}$ = 4.446

From the above $F_c > F_{tab}$

reject H_o , accept H_i ,: the mean of the delivery time are not equal.

The estimate $\hat{B}$ of the regression can be compute from

$$\hat{B} = \frac{Exy}{Exx} = \frac{1037.7533}{884.5}$$

$\hat{B}$ = 1.1732265461

Test of hypothesis can be carried out on this too, by using the test statistic.

H_o: $\hat{B}$ = 0

$$F_c = \frac{(Exy)^2 /(Exx)}{MSE}$$

$$F_c = \frac{(1037.753)^2 / 884.5}{5.2425}$$

$$F_c = \frac{1,217.5594}{5.2425}$$

F_c = 232.2478588

$F_{0.05,1,8}$ = 5.32

Decision

Since $F_c > F_{tab}$ reject H_o and accept H_i, it simply implies that the exists a linear relationship between the delivery time and volume delivered.

The adjusted treatment can be computed as;

Adjusted $Y_1 = \overline{Y}_1 - \hat{B}(\overline{X}_1 - \overline{\overline{X}})$

$Y_2 = \overline{Y}_2 - \hat{B} (\overline{X}_2 - \overline{\overline{X}})$

$Y_3 = \overline{Y}_3 - \hat{B}(\overline{X}_3 - \overline{\overline{X}})$

Where $\overline{\overline{X}}$ = grand mean of $X_{iz} = \overline{X}_1 + \overline{X}_2 + \overline{X}_3 = \overline{\overline{X}}$

$\overline{X}_{1,}\overline{X}_{2,}\ \overline{X}_3$ = the respective mean of x

$\overline{Y}_{1,}\overline{Y}_{2,}\ \overline{Y}_3$ = respective mean of Y

Adjusted $Y_1 = \overline{Y}_1 - \widehat{B}\ (\overline{X}_1 - \overline{\overline{X}})$

$= 36.25 - (1.173265461)\ (34.75 - 33.167)$

$= 36.25 - 1.16422\ (1.5833)$

$= 36.25 - 1.857631204$

$= 34.3923688$

$\cong\ 34.40$

Adjusted $Y_2 = \ \overline{Y}_2 - \widehat{B}\ (\overline{X}_2 - \overline{\overline{X}})$

$Y^2 = 33 - 1.173265461\ (31.25 - 33.167)$

$Y^2 = 33 - (1.16422)\ (-1.917)$

$Y^2 = 33 + 2.249149889$

$Y^2 = 35.24914989$

$Y^2 \cong 35.249$

Adjusted $Y_3 = \ \overline{Y}_3 - \widehat{B}\ (\overline{X}_3 - \overline{\overline{X}})$

$Y_3 = 33.25 - 1.173265461\ (33.5 - 33.167)$

$Y^3 = 33.25 - 1.173265461\ (0.33)$

$Y^3 = 33.25 - 0.387177602$

$Y^3 = 32.8628224$

$Y^3 \cong 32.86$

SELF-ASSESSMENT EXERCISE

Define $\mathbf{S_{yy}}$**?**

4.0 CONCLUSION

In the course of our discussion on analysis of covariance you have learnt about the following:

- Definition of analysis of covariance
- Estimation of analysis of covariance

- Computation of analysis of covariance table
- Adjustment of the dependent variables

5.0 SUMMARY

In the course of our discussion the following were inferred.

$$\mathbf{S_{yy}} = \sum_{i=1}^{n}\sum_{j=1}(\mathbf{Y^2} - \frac{(\mathbf{EY})^2}{\mathbf{an}} = \Sigma\Sigma\ (\mathbf{Y} - \bar{\bar{\mathbf{Y}}})$$

$$\mathbf{S_{xx}} = \sum_{i=1}^{n}\sum_{j=1}(\mathbf{x} - \bar{\bar{\mathbf{X}}}) = \Sigma\Sigma\ (\mathbf{X^2} - \frac{(\Sigma\mathbf{X})^2}{\mathbf{an}}$$

$$\mathbf{S_{xy}} = \sum_{i=1}^{n}\sum_{j=1}(\mathbf{x} - \bar{\bar{\mathbf{X}}})(\mathbf{Y} - \bar{\bar{\mathbf{Y}}}) = \Sigma\Sigma\ (\mathbf{X_{ij}}\ \mathbf{Y_{ij}} - \frac{\Sigma\mathbf{Xi}\Sigma\mathbf{Y})}{\mathbf{an}}$$

$$\mathbf{T_{yy}} = \sum^{n}(\bar{\mathbf{Y}} - \bar{\bar{\mathbf{Y}}})^2 = \sum_{i=1}\{\frac{\mathbf{Y}}{\mathbf{n}} - \frac{\Sigma\mathbf{Y^2}}{\mathbf{an}}\}$$

$$T_{xx} = \Sigma\ (\bar{X}_1 - \bar{\bar{X}}) = \frac{\Sigma x}{n} - \frac{\Sigma x^2}{an}$$

$$T_{xy} = \Sigma\ (\bar{X}_1 - \bar{\bar{X}}) = (\bar{Y}_1 - \bar{\bar{Y}}) = \Sigma XY - \frac{\Sigma X\Sigma Y}{an}$$

$E_{yy} = S_{yy} – T_{yy}$

$E_{xx} = S_{xx} – T_{xx}$

$E_{xy} = S_{xy} – T_{xy}$

6.0 TUTOR MARKED ASSIGNMENT

Submit a one page discussion on the definition of analysis of covariance and its assumption.

7.0 REFERENCES

- Damodar N. G., Dawn C. P. and Sangetha, G. (2012): Basic Econometrics. Tata McGraw Hill Education Private Ltd. New Delhi India.
- Dominick, S. and Derrick, R. (2011): Statistics and Econometrics, (Schaum’s Outlines) McGraw-Hill Company, New York.

- Kuotsoyanis, A. (2003): Theory of Econometrics (second edition).Palgrave publishers Ltd (formerly Macmillan publishers Ltd), Houndmills, Basingstoke, New York.
- www.youtube.com

MODULE 3: Multiple Regression Analysis

UNIT ONE: MULTIPLE REGRESSIONS

CONTENTS

1.0 INTRODUCTION

Multiple Regressions Defined

In introductory statistic, simple linear regression is one of the topics discussed. Regression equation is an expression by which you may calculate a typical value of a dependent variable say Y, on the basis of the values of independent variable(s).

Multiple regression model attempts to expose the relative and combine importance of the independent variables on dependent variables.

Multiple regression models is one among the commonly used tools in research for the understandings of functional relationship among multi-dimensional variables. The model attempts to expose the relative and combine effect of the independent variable on the dependent variable.

For your success in this course of study it is required that you have a thorough knowledge of simple regression model, hypothesis testing among others.

2.0 OBJECTIVE

At the end of our discussion on multiple regression you should be to;

(i) Regress the independent variable on the dependent variable

(ii) Understand parameter estimates involved

(iii) You should know how to calculate the values of b_o, b_1, b_2, … bn

(iv) Test of significance

Coefficient of multiple determinations

Test of overall significance of the regression

Partial correlation coefficient

3.0 MAIN CONTENT

3.1 Multiple Regression and Assumptions Defined

Multiple regression analysis is usually used for testing hypothesis about the relationship between a dependent variable Y and two or more independent variable X and for prediction or forecasting. Three variable linear regression models is usually written as:

$$Y = b_o + b_1X_1 + b_2X_2 + \mu$$

Where Y = dependent variable

b_o = intercept

b_1, b_2, bn = partial correlation coefficient or regression coefficient

μ = error term or residuals

Assumptions of Multiple Regressions

Multiple regression models has the following assumptions

i. Randomness

ii. Normality

iii. Measurement error

iv. Independent of μ and x_s

v. Correct specification of model

vi. Multi-colinearity

vii. Homoscedascity

viii. Linearity

ix. Same number of cases and variables

SELF-ASSESSMENT EXERCISE

Define multiple regression model of four variables?

3.2 Estimation of the Parameters of the Multiple Regression (b_o, b_1 …b_n)

For the purpose calculation and because of the parameters involved deviation method of calculating regression will be used. The parameters involve are define as stated below:

$$\hat{b}o = \overline{Y} - \hat{b}_1 \overline{X}_1 - \hat{b}_2 \overline{X}_2$$

$$\hat{b}_1 = \frac{(\Sigma x_1 y)(\Sigma x_2^2) - (\Sigma x_2 y)(\Sigma x_1 x_2)}{(\Sigma x_1^2)(\Sigma x_2^2) - (\Sigma x_1 x_2)^2}$$

$$\hat{b}_2 = \frac{(\Sigma x_2 y)(\Sigma x_1^2) - (\Sigma x_1 y)(\Sigma x_1 x_2)}{(\Sigma x_1^2)(\Sigma x_2^2) - (\Sigma x_1 x_2)^2}$$

bo = Calculated bo

$\hat{b}_1$ = Calculated b_1

$\hat{b}_2$ = Calculated b_2

$\overline{Y} = \frac{\Sigma Y}{n}$

$\bar{X} = \frac{\Sigma X}{n}$

$x_1 = X_1 - \bar{X}_1$

$y = Y - \bar{Y}$

$x_2 = X_2 - \bar{X}_2$

Self-assessment exercise

Define b_2?

3.3 Worked Example

The table below shows the value of expenditure on clothing, total expenditure and the price of clothing.

Table M3.1.1

Table Showing Expenditure on Clothing, Total Expenditure and Price of Clothing

	1990	**1991**	**1992**	**1993**	**1994**	**1995**
Price of clothing (X_2	3.5	9.8	8.3	7.6	9.3	7.7
Total expenditure (X_1	3.5	2.2	2.7	1.6	2.8	4.6
Value of expenditure clothing (Y)	2.0	1.5	1.7	1.6	2.7	3.5

Find the least square regression equation of Y on X_1 and X_2.

Table M3.1.2

Multiple Regression Table

Years	**Y**	**X_1**	**X_2**	**y**	**x_1**	**x_2**	**x_1x_2**	**x_1^2**	**x_2^2**	**x_1y**	**x_2y**	**y^2**
1990	20	35	35	-1.7	6	-42	-252	36	1764	-10.2	71.4	2.89
1991	15	22	98	-6.7	-7	21	-147	49	441	46.9	-140.7	44.89
1992	17	27	83	-4.7	-2	6	-12	4	36	9.4	-28.2	22.09
1993	16	16	76	-5.7	-13	-1	13	169	1	74.1	5.7	32.49
1994	27	28	93	5.3	-1	16	-16	1	256	-5.3	84.3	28.09
1995	35	46	77	13.3	17	0	0	289	0	226.1	0	176.89
n = 6	$\Sigma Y = 130$	$\Sigma X_1 = 174$	$\Sigma X_2 = 462$	$\sum y=0$	$\Sigma_{x1} = 0$	$\Sigma_{x2} = 0$	-414	548	2498	341	-7	307.34

$\bar{Y} = \frac{\Sigma Y}{n} = \frac{130}{6} = 21.7$

$\bar{X}_1 = \underline{\Sigma X_1} = \underline{174} = 29$

$n \qquad 6$

$$\bar{X}_2 = \frac{\Sigma X_2}{n} = \frac{462}{6} = 77$$

$$y = Y - \bar{Y}$$

$$x_1 = X_1 - \bar{X}_2$$

$$x_2 = X_2 - \bar{X}_2$$

$$\hat{b}_1 = \frac{(\Sigma x_1 y)(\Sigma x_2^2) - (\Sigma x_2 y)(\Sigma x_1 x_2)}{(\Sigma x_1^2)(\Sigma x_2^2) - (\Sigma x_1 x_2)^2}$$

$$\hat{b}_1 = \frac{(341)(2498) - (-7)(-414)}{(548)(2498) - (-414)^2}$$

$$\hat{b}_1 = \frac{851{,}818 - 2898}{1368904 - 171396}$$

$$\hat{b}_1 = \frac{848{,}920}{1197508}$$

$$\hat{b}_1 = 0.70891$$

$$\hat{b}_1 \cong 0.71$$

$$\hat{b}_2 = \frac{(\Sigma x_2 y)(\Sigma x_1^2) - (\Sigma x_1 y)(\Sigma x_1 x_2)}{(\Sigma x_1^2)(\Sigma x_2^2) - (\Sigma x_1 x_2)^2}$$

$$\hat{b}_2 = \frac{(-7)(648) - (341)(-414)}{(548)(2498) - (-414)^2}$$

$$\hat{b}_2 = \frac{-4536 + 141174}{1368904 - 171396}$$

$$\hat{b}_2 = \frac{136{,}638}{1197508}$$

$$\hat{b}_2 = 0.1141$$

$$\hat{b}_2 \cong 0.11$$

$\hat{b}_0 = \overline{Y} - \hat{b}_1 \overline{X}_1 - \hat{b}_2 \overline{X}_2$

$\hat{b}_o = 21.7 – 0.71(29) – 0.11 (77)$

$\hat{b}_o = 21.7 – 20.59 – 8.47$

$\hat{b}_o = - 24.370634778 + 21.7$

$\hat{b}_o \cong - 7.36$

The regression of Y on X_1 and X_2 is as written below

$$\hat{Y} = -7.36 + 0.71X_1 + 0.11X_2$$

The equation above is the multiple regression of value of expenditure on clothing, on price of clothing and total expenditure.

In multiple regression analysis of four parameters are of great importance from both the equation and the result, these are;

(i) Partial correlation coefficient (bn)

(ii) Multiple correlation coefficient (R^2)

(iii) Coefficient of determination (R^2)

(iv) Test of significance

All these will be dealt with in the subsequent unit of this module.

SELF-ASSESSMENT EXERCISE

Define the mean deviation of Y?

4.0 CONCLUSION

In the course of our discussion on multiple regression you have learnt about

- Definition of multiple regression
- Assumptions of multiple regression
- Regression coefficients
- Estimation of Multiple regression equation

5.0 SUMMARY

In this unit, multiple regression model is given as

$Y = B_o + B_1X_1 + B_2X_2 + \mu$

$\hat{B} = \bar{Y} - \hat{B}_1 \bar{X}_1 - \hat{B}_2 \bar{X}_2$

$$\hat{B}_1 = \frac{(\Sigma x1y)(\Sigma x_2^2) - (\Sigma x_2y)(\Sigma x_1x_2)}{(\Sigma x_1^2)(\Sigma x_2^2) - (\Sigma x_1x_2)^2}$$

$$\hat{B}_2 = \frac{(\Sigma x_2y)(\Sigma x_1^2) - (\Sigma x_1y)(\Sigma x_1x_2)}{(\Sigma x_1^2)(\Sigma x_2^2) - (\Sigma x_1x_2)^2}$$

Where $\hat{B}_1$ measures the change in Y for a unit change in X_1 while holding X_2 constant

B_2 measure change in Y per units change in X_2 holding X_1 constant

6.0 TUTOR MARKED ASSIGNMENT

i. The simplest possible multiple regression model is a ---------

ii. Given that $Y = B_1X_1 + B_2X_2 + B_3X_3 + \mu$ where $X_1 = 1$ this is an example of –

7.0 REFERENCES

- Damodar, N. G., Dawn, C. P., and Sangetha, G. (2012): Basic Econometrics. Tata McGraw Hill Education Private Ltd. New Delhi, India.
- Dominick, S. and Derrick, R. (2011): Statistics and econometric (Schaum outline) (2nd edition) McGraw Hill, New York.
- Oyesiku, O.K. and Omitogun, O. (1999): Statistics for Social and Management Sciences: Higher Education Books Publishers Lagos.
- Oyesiku, O.O., Abosede, A.J., Kajola, S.O, and Napoleon, S.G.(1999): Basics of Operation research. CESAP Ogun State University. Ago-Iwoye, Ogun State.

UNIT TWO: PARTIAL CORRELATION COEFFICIENT

CONTENTS

1.0 INTRODUCTION

It is assumed that you must have read unit 1 of this module that talks about multiple regression, a detailed understanding of this will be assumed. this unit is building on the unit 1 of this module. This unit will be dealing with thorough explanation of the parameters involved in the regression analysis.

2.0 OBJECTIVE

At the end of our discussion, you should be able to understand the following concepts such as;

- Partial regression co-efficient
- Estimation of partial regression co-efficient

3.0 MAIN CONTENT

3.1 Partial Correlation Defined

Partial correlation coefficient measures the correlation between the dependent variable and independent variables in the model. The regression coefficient B_1 and B_2 are known as partial regression or partial slope coefficient. B_1 measures the change in the mean value of Y per unit change in X_1 after removing the influence of X_2 or holding X_2 constant, this gives the direct effect or net effect of a unit change in X_1 on the value of Y.

B_2 coefficient measures the change in the mean value of Y per unit change in X_2 holding X_1 constant, this gives the direct effect or net effect of a unit change in X_2 on the mean of Y.

SELF-ASSESSMENT EXERCISE

Define partial correlation?

3.2 Estimation and Explanation of Partial Correlation Coefficient

Correlation coefficient (r) is defined as a measure the degree of linear association between two variable for three variable regression model we can compute 3 correlation coefficients $r_{yx1.x2}$, $r_{yx2.x1}$, r_{x1x2}

Where;

$R_{yx1\ .x2}$ = partial correlation coefficient between Y and X_1 holding X_2 constant

$r_{yx2\ .x1}$ = partial correlation coefficient between Y and X_2 holding X_1 constant

r_{x1x2} = partial correlation coefficient between X_1 and X_2 holding Y constant

The formular for these correlation coefficients are

$$r_{yx1} = \frac{\Sigma x_1 y}{\sqrt{\Sigma x_1^2}\sqrt{y^2}}$$

$$r_{yx2} = \frac{\Sigma x_2 y}{\sqrt{\Sigma x_2^2}\sqrt{\Sigma y^2}}$$

$$r_{x1x2} = \frac{\Sigma x_2 x_1}{\sqrt{\Sigma x_2^2}\ \sqrt{\Sigma x_1^2}}$$

$$r_{yx1.x2} = \frac{r_{yx1} - (r_{yx2})(r_{x1x2})}{\sqrt{(1 - r^2 x_1 x_2)}\ \sqrt{1 - r^2_{yx2}}} = \frac{r_{yx1} - (r_{yx2})(r_{x1x2})}{\sqrt{(1 - r^2_{x1x2})\ (1 - r^2_{yx2})}}$$

$$r_{yx2.x1} = \frac{r_{yx2} - (r_{yx1})(r_{x1x2})}{\left(\sqrt{1 - r^2_{x1x2}}\right)\sqrt{1 - r^2_{yx1}}} = \frac{r_{yx2} - (r_{yx1})(r_{x1x2})}{\sqrt{(1 - r^2_{x1x2})(1 - r^2_{yx1})}}$$

$$r_{x1x2.y} = \frac{r_{x1x2} - (r_{yx1})\ (r_{yx2})}{\left(\sqrt{1 - r^2_{yx1}}\right)\left(\sqrt{1 - r^2_{yx2}}\right)}$$

$$r_{x1x2y} = \frac{r_{x1x2} - (r_{yx1})\ (r_{yx2})}{\left(\sqrt{1 - r^2_{yx1}}\right)\ \left(\sqrt{1 - r^2_{yx2}}\ \right)}$$

Partial correlation coefficients range in value between -1 and +1. This value(s) is usually used to determine the relative importance of the different explanatory variables in a multiple regression.

SELF–ASSESSMENT EXERCISE

Define $r_{x1x2.y}$

3.3 Worked Examples

Here values of our correlation coefficient will be derived from the table in our unit of this moodle1i.e from the table of analysis.

Table M3.2.1

Correlation Coefficient Table

Years	**Y**	**X_1**	**X_2**	**Y**	**x_1**	**x_2**	**x_1x_2**	**x_1^2**	**x_2^2**	**x_1y**	**x_2y**	**y^2**
1990	20	35	35	-1.7	6	-42	-252	36	1764	-10.2	71.4	2.89
1991	15	22	98	-6.7	-7	21	-147	49	441	46.9	-140.7	44.89
1992	17	27	83	-4.7	-2	6	-12	4	36	9.4	-28.2	22.09
1993	16	16	76	-5.7	-13	-1	13	269	1	74.1	5.7	32.49
1994	27	28	93	5.3	-1	16	-16	1	256	-5.3	84.3	28.09
1995	35	46	77	13.3	17	0	0	289	0	226.1	0	176.89
n = 6	ΣY =130	ΣX_1 = 174	ΣX_2 = 462	∑y=0	Σ_{x1} = 0	Σ_{x2} = 0	-414	648	2498	341	-7	307.34

$\sum x_1y$ =341; $\sum x_2y$ = -7;$\sum x_1^2$ = 648;$\sum x_2^2$; $\sum y^2$=307.34; $\sum x_1x_2$= -414

$$r_{yx1} = \frac{\Sigma x_1 y}{\sqrt{\Sigma x1^2}\,\sqrt{\Sigma y^2}}$$

$$= \frac{341}{\sqrt{648} \text{ x } \sqrt{307.34}}$$

$$= \frac{341}{\sqrt{648 \text{ x } 307.34}}$$

$$= \frac{341}{\sqrt{199,156.32}}$$

$$r_{yx1} = \frac{341}{446.2693357}$$

r_{yx1} = 0.764112549

r_{yx1} ≅ 0.76

r^2_{yx1} = 0.583867987

r^2_{yx1} = 0.58

$$r_{yx2} = \frac{\Sigma x_2 y}{\sqrt{\Sigma x^2}_2\,\sqrt{\Sigma y^2}}$$

$$r_{yx2} = \frac{-7}{(\sqrt{2498})\ (\sqrt{307.34})}$$

$$r_{yx2} = \frac{-7}{}$$

$$(\sqrt{2498 \text{ x } 307.34}$$

$$r_{yx2} \quad = \quad \frac{-\ 7}{\sqrt{767{,}7353.32}}$$

$$r_{yx2} \quad = \quad \frac{-\ 7}{876.2050673}$$

$$= -\ 0.007988997395$$

$r_{yx2} = -\ 0.008$

$r^2_{yx2} = 0.0000638 = 0.000064$

$$r_{x1x2} \quad = \quad \frac{\Sigma x_1 x_2}{(\sqrt{\Sigma x^2}_2)\ (\sqrt{\Sigma} x_1{}^2)}$$

$$r_{x1x2} \quad = \quad \frac{-\ 414}{(\sqrt{2498)}\ \ (\sqrt{648})}$$

$$r_{x1x2} \quad = \quad \frac{-\ 414}{\sqrt{2498 \text{ x } 648}}$$

$$r_{x1x2} \quad = \quad \frac{-\ 414}{\sqrt{1{,}618{,}704}}$$

$$r_{x1x2} \quad = \quad \frac{-\ 414}{1{,}272.282987}$$

$r_{x1x2} = \ -\ 0.325399305$

$r_{x1x2} = -\ 0.33$

$r^2_{x1x2} = 0.105884707$

$r^2_{x1x2} = 0.11$

$$\text{Thus } r_{yx1.x2} \quad = \quad \frac{r_{yx1} - (r_{yx2})\ (r_{x1x2})}{(\sqrt{1 - r^2_{x1x2}})\ \ (\sqrt{1 - r^2_{yx2}})}$$

$$r_{yx1.x2} \quad = \quad \frac{0.76 - (\ -\ 0.008)\ (-\ 0.33)}{\sqrt{(1 - 0.11)(1 - 0.000064)}}$$

$$r_{yx1.x2} \quad = \quad \frac{0.76 - 0.00264}{}$$

$$\sqrt{(0.89)(0.999936)}$$

$$r_{yx1.x2} = \frac{0.75736}{\sqrt{0.88994304}}$$

$$r_{yx1.x2} = \frac{0.75736}{0.943367924}$$

$$r_{yx1.x2} = 0.802825685$$

$$r_{yx1.x2} \cong 0.802$$

$$r_{yx1.x2} = 80.2\%$$

$$r_{yx2.x1} = \frac{r_{yx2} - (r_{yx1})\,(r_{x1x2})}{\sqrt{(1 - r^2_{x1x2})}\,(\sqrt{1 - r^2_{yx1}})}$$

$$r_{yx2.x1} = \frac{-\,0.008 - (\,0.76)\,(-\,0.33)}{\sqrt{(1-0.11)(1-0.58)}}$$

$$r_{yx2.x1} = \frac{-\,008 + 0.2508}{\sqrt{(0.89)(0.42)}}$$

$$r_{yx2.x1} = \frac{0.2428}{\sqrt{0.3738}}$$

$$r_{yx2.x1} = \frac{0.2428}{0.61\ r_{yx2.x1} \quad 1391854}$$

$$r_{yx2.x1 =} \ 0.397126651$$

$$r_{yx2.x1} \cong 0.40$$

$$r_{yx2.x1} = 40\%$$

Therefore, from the calculations above it shows that X_1 explains more than X_2 and $X_1$is more important in explaining variation in Y.

SELF-ASSESSMENT EXERCISE

Define r_{x1x2}?

4.0 CONCLUSION

In the course of our discussion on partial correlation coefficient you must have learnt about the following:

- Partial correlation definition
- Partial correlation between the dependent variable and the independent variable such as; $r_{yx1.x2}$ = partial correlation coefficient between variable y and x_1 holding variable x_2 constant

$r_{yx2\ .x1}$ = partial correlation coefficient between variable y and x_2 holding x_1 constant.

$r_{x1x2.y}$ = partial correlation between variable y and x_1 variable x_2 holding variable y constant

5.0 SUMMARY

In the course of our discussion the following formulas were made use of;

$$r_{yx1} = \frac{\Sigma x_1 y}{\sqrt{\Sigma x_1^2 \, \Sigma y^2}}$$

$$r_{yx2} = \frac{\Sigma x_2 y}{\sqrt{\Sigma x_2^2 \, \Sigma y^2}}$$

$$r_{x1x2} = \frac{\Sigma x_2 x_1}{\sqrt{\Sigma x_2^2 \, \Sigma_{x1}^2}}$$

$$r_{yx1.x2} = \frac{r_{yx1} - (r_{yx2})(r_{x1x2})}{\sqrt{(1 - r^2_{x1x2})(1 - r^2_{yx2})}}$$

$$r_{yx2\ .x1} = \frac{r_{yx2} - (r_{yx1})(r_{x1x2})}{\sqrt{(1 - r^2_{x1x2})(1 - r^2_{yx1})}}$$

6.0 TUTOR MARKED ASSIGNMENT

Given that $Y = B_o + B_1 X_1 + B_2 X_2 + \mu$ the partial regression coefficient is given by------

REFERENCES /FURTHER READINGS

- Damodar, N. G., Dawn, C. P., and Sangetha, G.(2012): Basic Econometrics. Tata McGraw Hill Education Private Ltd. New Delhi, India.

- Dominick, S. and Derrick, R. (2011): Statistics and econometric (Schaum outline) (2nd edition) McGraw Hill. New York.

- Oyesiku, O.K. and Omitogun, O. (1999): Statistics for Social and Management Sciences: Higher Education Books Publishers, Lagos.

- Oyesiku, O.O., Abosede, A.J., Kajola, S.O. and Napoleon, S.G. (1999): Basics of Operation research. CESAP Ogun State University, Ago-Iwoye, Ogun State.

UNIT THREE: Multiple Correlation Co-efficient and Coefficient of Determination

CONTENTS

1.0 INTRODUCTION

This unit is an extension of unit one and two of this module. This unit requires thorough knowledge of unit 1 and unit two. In this unit we are going to look at multiple Correlation Coefficients (R) and multiple coefficient of determination (R^2).

2.0 OBJECTIVE

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

- Estimate multiple correlation coefficient (r)
- Estimate coefficient of determination
- Interprete your answer i.e. statistical interpretation

3.0 MAIN CONTENT

3.1 Multiple Correlation Coefficient (R) Coefficient of Determination (R^2) Defined

Multiple correlation coefficients represented by R measures the degree of linear association between two or more variables. Say variable Y and the entire explanatory variable jointly. Its value can be positive or negative; multiple correlation coefficients is always taken to be positive. In practice the multiple correlation coefficients is of little importance. The more meaningful coefficient is the coefficient of determination R^2 or r^2. Coefficient of determination (R^2) is defined as the proportion of the total variation in Y explained by the multiple regression of Y on X_1 and X_2. It measures goodness of fit of the regression equation. In a three variable model we are always interested in knowing the proportion of the variation in Y explained by each of the explanatory variable X_1 and X_2. The coefficient of determination is denoted by R^2 or r^2. Because of the relative importance of coefficient of determination (R^2) we concentrate more on the coefficient of determination (R^2).

SELF-ASSESSMENT EXERCISE

The most important coefficient is --------

3.2 Estimation of Coefficient of Determination

Conceptually, it is often written as:

$$\text{Coefficient of determination} = r^2 = R^2 = \frac{\hat{b}_1 \Sigma y_1x_1 + \hat{b}_2\Sigma yx_2}{\Sigma y^2}$$

$$R = r = ryx_1x_2 = \sqrt{\frac{r^2yx_1 + r^2yx_2 - 2ryx_1 . ryx_2 . rx_1x_2}{1 - r^2_{x1x2}}}$$

The value of R^2 lies between 0 and 1, if it is 1, the fitted regression line explains 100% of the variation in Y, on the other hand, if it is 0, the model does not explain any of the variation in Y. typically, however, R^2 lies between these two extremes values. The fit is said to be better, the closer R^2 is to 1.

Self-assessment exercise

The coefficient of determination usually lies between------ and ------

3.3 Worked Example

From our calculation in unit 1 & 2 of this module especially the table in unit 1, we are going to derive our values from the table in unit 1.

Table M3.3.1

Multiple Regression Table

Years	**Y**	**X_1**	**X_2**	**Y**	**x_1**	**x_2**	**x_1x_2**	**x_1^2**	**x_2^2**	**x_1y**	**x_2y**	**y^2**
1990	20	35	35	-1.7	6	-42	-252	36	1764	-10.2	71.4	2.89
1991	15	22	98	-6.7	-7	21	-147	49	441	46.9	-140.7	44.89
1992	17	27	83	-4.7	-2	6	-12	4	36	9.4	-28.2	22.09
1993	16	16	76	-5.7	-13	-1	13	169	1	74.1	5.7	32.49
1994	27	28	93	5.3	-1	16	-16	1	256	-5.3	84.3	28.09
1995	35	46	77	13.3	17	0	0	289	0	226.1	0	176.89
n = 6	ΣY =130	ΣX_1 = 174	ΣX_2 = 462	∑y=0	$\Sigma_{x1} = 0$	$\Sigma_{x2} = 0$	-414	548	2498	341	-7	307.34

$$R^2 = \frac{0.59\ (341) + 0.094(-7)}{307.34}$$

$$R^2 = \frac{201.19 - 0.658}{307.34}$$

$$R^2 = \frac{200.532}{307.34}$$

$$R^2 = 0.652476085$$

$$R^2 = 0.65$$

$$R^2 \cong\ 65\%$$

This implies that the explanatory variable (x_1 and x_2) can only account for 65% variation in variable Y i.e. both x_1 and x_2 contributes 65.2% to the explanation of the variation in Y.

$$r_{yx1x2} = \sqrt{\frac{r^2yx_1 + r^2yx_2 – 2yx_1 \, . \, ryx_2 \, . \, rx_1x_2}{1 – r^2_{x1x2}}}$$

$$r_{yx1x2} = \sqrt{\frac{0.058 + 0.000064 – 2\ (0.7641).\ (0.008)\ .0.11}{1 – 0.11}}$$

$$r_{yx1x2} = \sqrt{\frac{0.580064 – 2\ (15282)\ (-\ 0008)\ 0.11}{0.89}}$$

$$r_{yx1x2} = \sqrt{\frac{0.580064 – 0.001344816}{0.89}}$$

$$r_{yx1x2} = \sqrt{\frac{0.578719184}{0.89}}$$

$$r_{yx1x2} = \sqrt{0.650246274}$$

$r_{yx1x2} = 0.806378493$

$r^2 = 0.650246274$

$r^2 \cong 0.65\%$

SELF-ASSESSMENT EXERCISE

When r^2=0.85, what is the economic interpretation of this?

4.0 CONCLUSION

In the course of our discussion of this unit, you have learnt about the following:

- Concept of multiple correlation

- Coefficient of determination
- Estimation of R^2 & r
- Interpretation of r & r^2

5.0 SUMMARY

In our discussion of this unit we defined coefficient of determination R^2 as

$$R^2 = \frac{\hat{b}_1 \Sigma y_1 x_1 + \hat{b}_2 \Sigma y x_2}{\Sigma y^2}$$

$$r = \sqrt{\frac{r^2 yx_1 + r^2 yx_2 - 2ryx_1 \,.\, ryx_2 \,.\, rx_1x_2}{1 - r^2_{x1x2}}}$$

The closer the r^2 is to 1, the better

6.0 TUTOR MARKED ASSIGNMENT

i. The measure of proportion or percentage of variation in Y explained by the explanatory variable $x_1 \ldots x_n$ jointly is given by -----------

ii. Multiple coefficient of determination measures------------

7.0 REFERENCES

- Damordar, N. G., Dawn, C. P. and Sangeetha, G. (2012): Basic Econometrics (5th edition) Tata McGraw Hill Education Private Limited. New Delhi, India.

- Dominick, S. and Derrick, R. (2011): Statistics and econometrics (Schaum outline) (2nd edition). McGraw Hill, New York.

- Oyesiku O.K. and Omitogun O. (1999): Statistics for social and management science (2nd edition). Higher Education Books Publisher, Lagos.

- Oyesiku, O.O., Abosede, A.J., Kajola, S.O. and Napoleon, S.G. (1999): Basics of Operation research. CEAP OSU, Ago-Iwoye. Ogun State.

UNIT FOUR: TEST OF SIGNIFICANCE

CONTENTS

1.0 INTRODUCTION

This unit completes this module, so it is required that thorough knowledge of unit one to unit three is very germane. It is important to test for the significance of the value of the regression, Coefficients, and the level of prediction or explanation given by the regression equation.

2.0 OBJECTIVE

At the end of this unit the student(s) should be able to calculate and understand:

- The calculation of F.statistics (Fcal)
- Check the corresponding tabulated value of F.statistics through its degree of freedom.
- Compare the F.statistics and Ftab
- Interprete your answer

3.0 MAIN CONTENT

3.1 Test of Significance Defined

Test of significance is a procedure by which sample results are used to verify truity of falsity of a null hypothesis. The key idea behind test of significance is that of a test statistics (estimator and the sampling distribution of such a statistics under the null hypothesis).

The decision to accept or reject H_o is made on the basis of the test statistics obtained from the data at hand.

The overall significance of the regression can be tested with the ratio of the explained to the unexplained variance. This follows an F-distribution with k – 1 and n – k degree of freedom, where n is the number of observations and k is the number of parameters estimated.

The joint hypothesis can be tested by the analysis of variance (Anova).

SELF-ASSESSMENT EXERCISE

The decision to accept or reject Ho depends on ---------

3.2 Estimation of Test of Significance

The F-statistics or F-ratio for the test of significance can be written as:

H_o: $\mu_1 = \mu_2$; H_i: $\mu_1 \neq \mu_2$

$$F_{k-1,n-k} = \frac{\Sigma y/(k-1)}{\Sigma e^2/(n-k)} = \frac{R^2/(k-1)}{(1-R^2)(n-k)}$$

Also Anova table can as well be used for test of significance.

Table M3.4.1

Anova Tables for 3-Variables Regression

Source of variation	**Sum of squares**	**DF**	**Mean sum of square**
Due to Regression (ESS)	$\hat{B}_1\Sigma yx_1+\hat{B}_2\Sigma yx_2$	2	$\dfrac{\hat{B}_1\Sigma yx_1+\hat{B}_2\Sigma yx_2}{2}$
Due to Residual (RSS)	$\Sigma\mu_i^2$	n – 3	$\sigma^2 = \dfrac{\Sigma\mu_i^2}{n-3}$
Total	Σy_i^2	n – 1	

Note

$\Sigma y_i^2 = \hat{B}_1\Sigma yx_1 + \hat{B}_2\Sigma yx_2 + \Sigma\mu_1^2$

TSS = ESS + RSS

$$\text{F-.ratio} = \frac{(\hat{B}_1\Sigma yx_1 + \hat{B}_2\Sigma yx_2)/2}{\Sigma\mu^2/n-3}$$

$$E\ \frac{\Sigma\hat{\mu}^2}{n-3} = E\,(\hat{\sigma}^2) = \sigma^2$$

3.3 Summary of F-Statistic

Table M3.4.2

F-STATISTICS TABLE

Null hypothesis Ho	Alternative Hypothesis H_1	Criteria Region Reject Ho if
$\sigma_1^2 = \sigma_2^2$	$\sigma_1^2 > \sigma_2^2$	$\dfrac{S_1^2}{S_2^2} > F\propto$, ndf, ddf($f_{tab}$)
$\sigma_1^2 = \sigma_2^2$	$\sigma_1^2 \neq \sigma_2^2$	$\dfrac{S_1^2}{S_2^2} > F\propto/2$, ndf, ddf ($f_{tab}$) or $<F_{1-}\propto/2$, ndf, ddf

If the calculated F-ratio (Fc) exceeds the tabular value of F (F_{tab}) at the specified level of significance and degree of freedom, the hypothesis is accepted that the regression parameters are not all equal to zero and that R^2 is significantly different from zero.

If the null hypothesis is true, it gives identical estimates of true σ^2. This statement should not be surprising because if there's a trivial relationship between y and x_1 and x_2 the source of variation in Y will be due to the random forces usually represented by e_i or μ_1. If however, the null hypothesis is false, that is x_1 and x_2 actually influence Y; the equality will not hold. Here, the ESS will be relatively larger than the RSS taking due account of their respective degree of freedom. Therefore, the F-.ratio provides a test of the null hypothesis that the true slope coefficients are simultaneously zero.
DECISION CRITERIA; If the F-ratio calculated exceeds the critical F-value from the table at the $\propto$ percent level of significance we reject Ho; otherwise do not reject it. Alternatively if the F-cal of the observed F is sufficiently low accept Ho.

SELF-ASSESSMENT EXERCISE

State the decision criteria for test of significance?

3.4 Worked Example

Ho; = $\mu_1 = \mu_2 = 0$

note that $$F_{k-1,n-k} = \frac{\Sigma\hat{Y}^2/k-1}{\Sigma\mu^2/n-k} = \frac{R^2/(k-1)}{(1-R^2)(n-k)}$$

From the above it is glaring that calculation of $\Sigma\mu^2$ is required, so there's need to generate a new table apart from the one generated in unit one of this modules as to get the value for our $\Sigma\mu^2$.

Table M3.4.3

Test of Significance Table

Years	**Y**	X_1	X_2	$\hat{Y}$	**e**	e^2
1990	20	35	35	21.1294	- 1.1294	1.2755
1991	15	22	98	12.1294	- 4 1294	17.0519
1992	17	27	83	20.694	- 3.694	13.6456
1993	16	16	76	13.6094	2.3906	5.7150
1994	27	28	93	22.2194	4.7806	22.8541
1995	35	46	77	31.3994	3.6006	12.9643
n = 6	130	174	462		0	73.506436

Note: The trend equation = $Y = - 2.6706 + 0.59x_1 + 0.09x_2$

The $\hat{Y}$ is arrived at by substituting various values of x_1 and x_2 into the trend equation

$e = Y - \hat{Y}$ (ie column 2 – column 5)

$y_1^2 = 307.34$ (from unit one)

Method I

$$F_{3-1,6-3} = \frac{\frac{307.34}{3-1}}{\frac{73.506432}{6-3}} = \frac{\frac{307.34}{2}}{\frac{73.506432}{3}}$$

$$F_{cal} = F_{2,3} = \frac{153.67}{24.502144} = 6.271696061$$

Method II

Table M3.4.4

Anova Table for 3-Variance

Sources of variation	**Sum of squares**	**DF**	**MSS**
ESS	200.56	2	100.28
RSS	73.506436	3	24502
Total	274.066436	5	

$ESS = \hat{B}\Sigma y_1x_1 + \hat{B}_2\Sigma y_1x_2 = 0.59\ (341) =- 0.09\ (-7)$

$= 201.19 – 0.63$

$= 200.56$

$RSS = \Sigma e_1^2 = 73.506436$

$$F_{cal} = \frac{ESS/df}{RSS/df} = \frac{100.28}{24.502}$$

$= 4.09272$

$F_{cal} \cong 4.1$

$F_{tab} = F_{2,3} = 9.55$

$F_{tab} = 9.55$

Since $F_{cal} < F_{tab}$, accept H_o and reject H_1.

SELF-ASSESSMENT EXERCISE

What will be the decision criteria if $F_{cal} > F_{tab}$?

4.0 CONCLUSION

From our discussion on this unit you have learnt about:

- Definition test of significance
- Estimation of test of significance
- The interpretation of resulting

- Estimation through ANOVA table and otherwise
- Derivation of the error term u_i or e_i

5.0 SUMMARY

In the course of our discussion the following formulars where discussed

$$F_{k-1,n-k} = \frac{\Sigma\hat{Y}^2/(k-1)}{\Sigma e^2/(n-k)}$$

$$\Sigma y_1^2 = \widehat{b\iota}_1\Sigma y_1 x_1 + \widehat{b\imath}_2\Sigma y_1 x_2$$

$$\Sigma \mu_1^2 = \Sigma(Y-\hat{Y})^2$$

6.0 TUTOR MARKED ASSIGNMENT

Given the regression model $Y = B_1 + B_2X_2 + B_3 X_3+U$, how would you state the null hypothesis to test for test for significance of x_1 and x_2 on Y.

7.0 REFERENCES

- Damodar, N. G., Dawn, C. P. and Sangetha, G. (2012): Basic Econometrics. Tata McGraw Hill Education Private Ltd. New Delhi, India.

- Dominick, S. and Derrick, R. (2011): Statistics and econometrics (schaum outline) (2nd edition) McGraw Hill, New York.

- Oyesiku, O.K. and Omitogun, O. (1999): Statistics for Social and Management Sciences: Higher Education Books Publishers, Lagos.

- Oyesiku, O.O., Abosede, A.J., Kajola, S.O., and Napoleon, S.G.(1999): Basics of Operation research. (CESAP Ogun State University), Ago iwOye, Ogun State.

MODULE 4: Time series analysis

Unit 1: Time series and its components

Unit 2; Quantitative estimation of time series

Unit 3: Price index

UNIT ONE: TIME SERIES

CONTENTS

1.0 INTRODUCTION

In all the social sciences, and particularly economics and business, the problem of how condition changes with the passage of time is of utmost importance. For study of such problems, the appropriate kind of statistical information consist of data in the form of time series, figures which shows the magnitude of a phenomenon month after month or year after year. The proper methods for treating such data and thus summarizing the experience which they represent are indispensable part of the practicing statistician equipment.

2.0 OBJECTIVE

At the of this unit, you should be able to

- Understand or define time series
- Understand component part of time series
- Understand methods of estimating time series
- Estimation and graphical representation of the trend

3.0 MAIN CONTENT

3.1 Time Series Defined

Time series refers to sequence of observations that gives information on how data has been behaving in the past.

You might wonder why we should spend so much effort constructing series showing what has happened in the past. This is history and should we not rather be looking to the future? As you know the twentieth century is age of planning: government plans the economy for many years ahead; public corporation plan output and investment; most state plan to keep the rate of inflation down to an acceptable level.

Good planning is usually based on information and this is where the time series comes into its own. It provides information about the way in which economic and social variable have been behaving in the recent past, and provides an analysis of that behaviour that planner cannot ignore. Naturally, if we are looking into the future, there is certain assumption we have to make, the most important of which is that the behavioural pattern that we have found in the past could continue into the future. In looking to the future there are certain pattern that we assume will continue and it is to help in the determination of these pattern that we undertake the analysis of the time series.

Time series is usually ordered in time or space. Time series is denoted by sequence (Y_t) where Y_t is the observed value at time t.

Essentially, time series is usually applied to economic and business problems whose purpose of analyses data is to permit a forecast to the future both in the long term and short term. It may be used as an essential aid to planning. Example of time series data are volume of sales, the character and magnitude of its cost of production etc. population figure, price level, demand of a commodity.

SELF-ASSESSMENT EXERCISE

The essence of time series is forecast. (true/false)

3.2.0 Components of Time Series

The nature or variation or type of changes in times series can be categorise into:

- Secular trend or long term movement
- Seasonal variation
- Cyclical variation
- Irregular or residual variation

3.2.1 Secular Trend

This refers to the general direction in which the graph of time series appears to be going over a long period of time. This explains the growth or decline of a time series over a long period. Time series is said to contain a trend if the mean or average of series changes systematically with time. The trend could be upward or downward, this could take any of the shape below.

GRAPHICAL REPRESENTATION OF SECULAR TREND

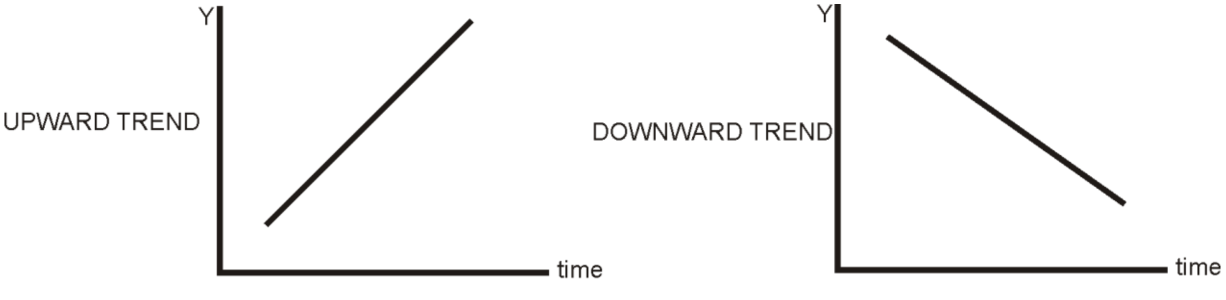

Fig. M4.1.1

3.2.2 Seasonal Variation

This refers to short term fluctuation or changes that occur at regular intervals less than a year. It is usually brought about by climatic and social factor(s), it is usually because of an event occurring at a particular period of the year. Examples of these are sale of card during valentine period, sale of chicken during xmas, new year or any festive period(s).

GRAPHICAL REPRESENTATION OF SEASONAL VARIATION

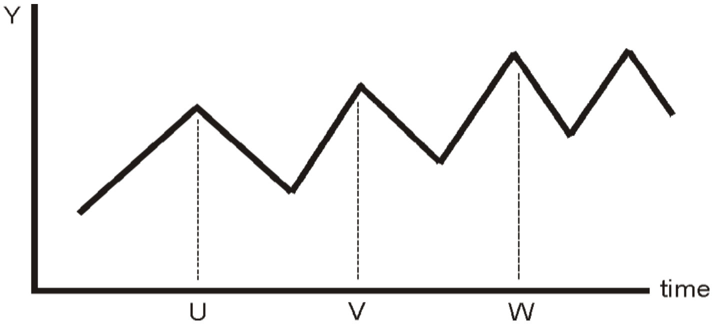

Fig. M4.1.2

3.2.3 Cyclical Variation

This refers to long term variations about the trend usually caused by disruption in services or socio-economic activities, cyclical variations are commonly associated with economic cycles, successive boom and slumps in the economy. A good example of this is business cycle.

GRAPHICAL REPRESENTATION OF CYCLICAL VARIATION

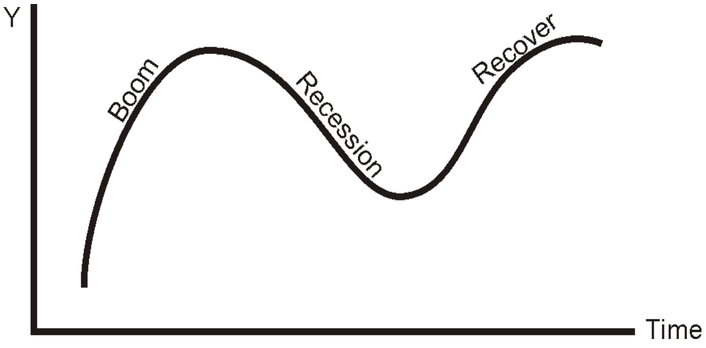

Fig. M4.1.3

3.2.4 Irregular Variation

This refers to time series movement that are not definite this is usually caused by unusual or unexpected and unpredictable events such as strike, war, flood, disasters. Here, there's no definite behavioural pattern.

GRAPHICAL REPRESENTATION OF IRREGULAR VARIATION

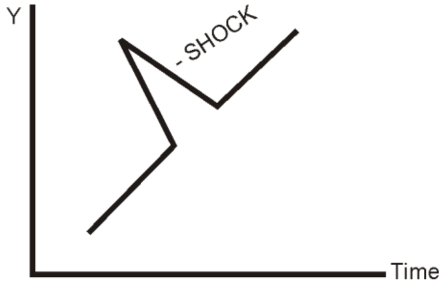

Fig. M4.1.4

SELF-ASSESSMENT EXERCISE

The trend of secular trend can either be upward or downward. (true/false)

3.3.0 Measurement of Trend

Basically trend values of a time series can be estimated by any of the following methods:

- Free hand
- Least square method
- Moving average and
- Semi average method

3.3.1 Free Hand Method

This method involves the drawing a scattered diagram of the values with time as the independent variable on the x-axis and then drawing the trend line by eye. This method is condemned because it is subjective and inaccurate method of obtaining a Trend line.

GRAPHICAL REPRESENTATION OF FREE HAND METHOD

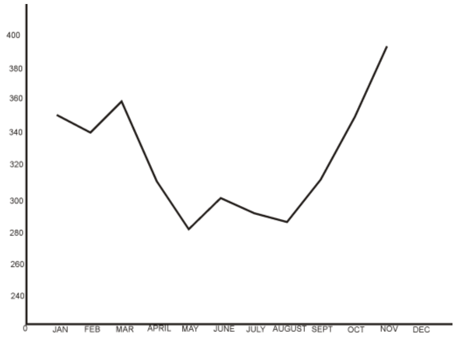

Fig. M 4.1.5

Other quantitative methods will be dealt with in the next unit

4.0 CONCLUSION

In the course of our discussion on time series analysis you have learnt about

- Time series
- Time series data
- Component of time series
- Free hand trend measurement
-

5.0 SUMMARY

Majorly time series decomposes itself into the following;

- Secular trend or long term movement
- Seasonal variation
- Cyclical variation
- Irregular or residual variation

6.0 TUTOR MARKED ASSIGNMENT

Table Showing Sales of a Chemist

Years	**Sales**
2000	85
2001	96
2002	108
2003	123
2004	98

Make a freehand sketch of the above information

7.0. REFERENCES/FURTHER READINGS

- Adedayo, O.A. (2006): Understanding Statistics. JAS Publishers, Akoka, Lagos.

- Dawodu, A.F. (2008): Modern business Statistics 1. NICHO Printing Works, Agbor, Delta State.

- Esan, E.O. and Okafor, R.O. (2010): Basis Statistical Method. Tony Chriisto Concept, Lagos.

- Olufolabo, O.O. & Talabi, C.O. (2002): Principles and Practice of Statistics HAS-FEM (NIG) ENTERPRISES Somolu Lagos.

- Owen, F. and Jones, R. (1978): Statistics. Polytech Publishers Ltd. Stockport.

- Oyesiku, O.K. and Omitogun, O.(1999): Statistics for social and Management Sciences. Higher Education Books Publisher, Lagos.

UNIT TWO; ESTIMATION OF TIME SERIES

CONTENTS

1.0 INTRODUCTION

This unit is an extension of unit one of this module, here, you are going to learn more about estimation of time series data, also, a thorough understanding of unit one of this module is required for proper understanding of this module.

2.0 OBJECTIVE

At the of this unit, you should be able to

- estimate any time series data,
- Understand methods of estimating time series,
- Estimate and do the graphical representation of the trends.

3.0 MAIN CONTENT

3.1.0 Least Square Method of Estimation

As defined in Preceding unit of this module, Time series refers to sequence of observations that gives information on how data has been behaving in the past. Estimation has to do with how time series are calculated, in this sub section we shall talk about three methods of estimation or measurement. These are method of least square, moving average and semi average method.

3.1.1 Least Square Method

This method is a statistical technique usually used in calculating the line of best fit or line of goodness that measures the goodness of fit of the curve, this is usually independent of human judgments, it makes an assumption that the trend line is a straight one. The least square formular is given as;

$$Y = a + bx + e$$

Where a = intercept

b = slope of the curve

e = error term

Formular 1

Trend equation of the least square method is given as

$$\Sigma Y = na + b\Sigma x$$

$$\Sigma Y = a\Sigma x + b\Sigma x^2$$

Where Σ = summation term derived from the data of the problem at hand

Σx = sum of X values

ΣY = sum of Y values

Σxy = sum found by multiplying each Y by corresponding X value and adding the Products

n = no of items involved in the whole time series

The least square estimates of a and b are the solution to the normal equation above which can be solve simultaneously.

Formular 2

The general formular is as given below

$$b = \frac{n\Sigma XY - \Sigma X\Sigma Y}{n\Sigma X^2 - (\Sigma X)^2}$$

$$a = \overline{Y} - \hat{b}\,\overline{x}$$

where $\overline{Y} = \frac{\Sigma Y}{n}$

$$\overline{X} = \frac{\Sigma x}{n}$$

Worked Example

Given the 7weeks information below about the sales of a company

Table M 4.1.1

Table Showing the Sales of a Company

Wk	**Sales**
1	15
2	25
3	38
4	32
5	40
6	37
7	50

Let X represents the weeks

Y represent the sales value

Table M4.1.2

Least Square Method Table of Analysis

X	**Y**	**XY**	**X^2**
1	15	15	1
2	25	50	4
3	38	114	9
4	32	128	16
5	40	200	25
6	37	222	36
7	50	350	49
$\sum X = 28$	**$\sum Y = 237$**	**$\sum XY = 1079$**	**$\sum X^2 = 140$**

$$\widehat{B} = \frac{n\Sigma XY - \Sigma X \Sigma Y}{n\Sigma X^2 - (\Sigma X)^2}$$

$n = 7$ $\qquad$ $\overline{X} = \frac{28}{7} = 4$

$\Sigma x = 28$ $\qquad$ $\overline{Y} = \frac{237}{7} = 33.857$

$\Sigma y = 237$

$\Sigma xy = 1079$

$\Sigma x^2 = 140$

$$\widehat{B} = \frac{7(1079) - 28\ (237)}{7\ (140) - 28^2}$$

$$\widehat{B} = \frac{7553 - 6636}{980 - 784}$$

$$\widehat{B} = \frac{917}{196}$$

$\widehat{B} = 4.67857$

$a = \overline{Y} - \widehat{B}\ \overline{x}$

a = 33.857 – (4.67857)4

= 33.857 – 18.7142

= 15.1427

The trend equation will be:

Y = 15.1427 + 4.6785x

This trend equation can be used in forecasting into future sales of the company, for example future sales value for the 10th and 12th week can be known by simply substituting the week's value into the trend equation.

i.e. for the 10th week we have;

Y = 15.1427 + 4.6785(10)

Y = 15.1427 + 46.785

Y = 61 – 9277

Y ≅ 62

For the 12th week

Y = 15.1427 + 4.6785(12)

Y = 15.1427 + 56.142

Y = 71.2847

Y = 71

3.2.0 Moving Average Method

A moving average is a simple arithmetic mean. We select a group of figures at the start of the series e.g. 3,4,5,7 and average them to obtain our first trend figure. Then you drop the first figure and include the next item in the series to obtain a new group. The average of this group gives the second trend figure. You continue to do this until all figures in the series is exhausted.

There is no doubt that the trend eliminates the large scale fluctuations found in the original series moving average smoothing is a smoothing technique used to make the long-term trend of a time series cleared.

Example 2

The table below contained information about the actual sales of a company

Table M4.1.3

Table Showing the Sales of a Company

Month	**Sale (units)**
Jan	350
Feb	340
Mar	360
April	310
May	280
June	300
July	270
August	260
Sept	310
Oct	350
Nov	370
Dec	390

Prepare a 3 month moving average forecast

Solution

Table M4.1.4

3- Month Moving Average Method Table of Analysis

Months	**Sales**	**3months Moving total**	**3months moving average trend**
Jan	350		
Feb	340	1050	350
Mar	360	1010	336.7
April	310	950	316.7
May	280	890	296.7
June	300	850	283.3
July	270	830	276.7
Aug	260	840	280
Sept	310	920	306.7
Oct	350	1030	343.3
Nov	370	1110	370
Dec	390		

Column 1 on the table represents the months

Column 2 represents the sale's figure

Column 3 is arrived at by adding the sales figure in 3s i.e

Jan + Feb + Mar = 1050

Feb + Mar + April = 1010

Mar + April + May = 950

Column 4 is arrived at by dividing the column 3 by the n which happen to be the moving average. This Rs called the trend.

3.2.1 GRAPHICAL REPRESENTATION OF MOVING AVERAGE TREND

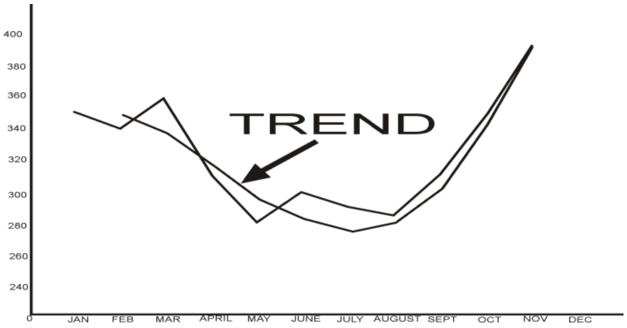

Fig. M4.1.6

Example 2

From the time series data below determine the trend on sales of a company

Table M4.1.5

Table showing sales of a company per quarter

Quarter

Years	**1**	**2**	**3**	**4**
1982	600	820	400	720
1983	630	840	420	740
1984	670	900	430	760

Prepare a 4-quarter moving average

Solution

Table M 4.1.6

4-point moving average table of analysis

Year	**Quarter**	**Sales**	**4 point moving total**	**4-point average moving or 4 quarterly average**	**2 point total or centre total**	**Moving average (trend)**
1982	1	600	-	-	-	-
	2	820	-	-	-	-
			2540	635	-	
	3	400			1277.5	638.75
			2570	642.5	-	-
	4	720			1290	645
			2590	647.5	-	-
1983	1	630			1300	650
			2610	652.5	-	-
	2	840			1310	655
			2630	657.5	-	-
	3	420			1325	662.5
			2670	667.5	-	-
	4	740			1350	675
			2730	682.5		
1984	1	670			1367.5	683.75
			2740	685	-	-
	2	900			1375	687.5
			2760	690	-	-
	3	430	-	-	-	-
	4	760	-	-	-	-

Column 1 represent years

Column 2 represents quarter periods

Column 3 represents sales values

Column 4 is arrived at by adding the sales value in 4s

Column 5 is derived by dividing column 4 by the nos of quarters

Column 6 is the total of column 5 when taken in 2s

Column 7 is arrive at by dividing column 6 by 2

GRAPHICAL REPRESENTATION OF FOUR QUARTER MOVING AVERAGE TREND

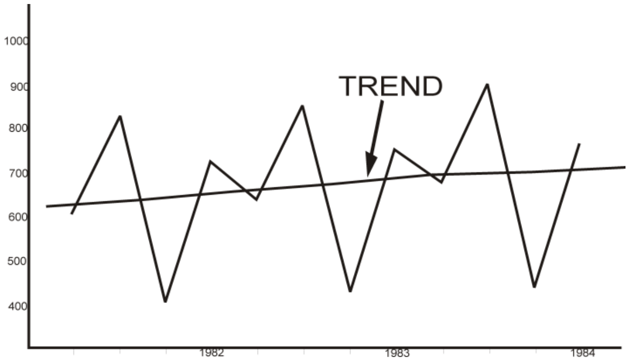

Fig. M4.1.7

3.3.0 Semi Moving Average Method

This method is usually used to estimate trends by separating or dividing that data into two equal parts and averaging the data each part, thus, obtaining two points on the graph of time series. A trend is then drawn between these two points and trend value can be determined. If the number of years is odd, the middle year is deleted and the group can then be divided into two equal parts.

Example

Table M4.1.8

Semi- Moving Average Method Table of Analysis

Years	Quarter	Y sales	X	Semi Average Total	Semi average method trend
1992	1	600	- 6		
	2	820	- 5		
	3	400	- 4		
	4	720	- 3	4010	668.33
1993	1	630	- 2		
	2	840	- 1		
	3	420	1		
	4	740	2		
1994	1	670	3		
	2	900	4		
	3	430	5		
	4	760	6	3,920	653.3

Column 4 represents the total of the 1st half and 2nd half.

Column 5 is arrived at by dividing the column 4 by 6 this represents the trend value, when plotted in a graph it gives the trend line.

3.3.1 GRAPHICAL REPRESENTATION OF SEMI MOVING AVERAGE TREND

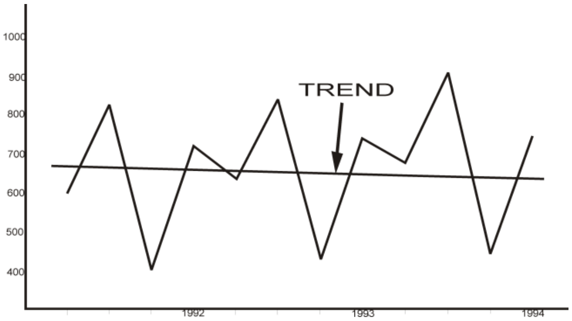

Fig. M4.1.8

SELF-ASSESSMENT EXERCISE

State the least square equation of a time series data?

4.0 CONCLUSION

In the course of our discussion on estimation of time series, you have learnt about

- least square method
- moving average method
- semi average method.

5.0 SUMMARY

The least square trend equation is written as

$$Y = a + b x + e$$

Where a = intercept $= \overline{Y} - \hat{b}\,\overline{x}$

$$\hat{b} = \text{slope} = \frac{n\Sigma XY - \Sigma x(\Sigma Y)}{n\Sigma x^2 - (\Sigma x)^2}$$

For moving average develop the following

- n – moving total
- determine the moving average
- plot the trend value to know the trend line

For semi average

- divide the data into 2 equal part
- when you have an odd data given, eliminate or delete the data in the middle
- get the half way total of each division
- divide the half way total by n depending on data supplied

6.0 TUTOR MARKED ASSIGNMENT

Table 4.1.9

Table Showing the Number of Prescriptions Dispensed by a Chemist

Year	**Quarters**			
	1	**2**	**3**	**4**
2000	-	-	60	71
2001	69	67	62	69
2002	73	66	62	68
2003	72	66	65	67
2004	75	-	-	-

Prepare a 4-point moving average of the above information?

7.0 REFERENCES/FURTHER READINGS

- Adedayo, O. A. (2006): Understanding Statistics. JAS Publishers, Akoka Lagos.

- Dawodu, A.F. (2008): Modern business Statistics 1. NICHO Printing Works. Agbor, Delta State.

- Esan, E.O. and Okafor, R.O. (2010): Basic Statistical Method. Tony Christo Concept, Lagos.

- Olufolabo, O.O. & Talabi, C.O. (2002): Principles and Practice of Statistics HAS-FEM (NIG) ENTERPRISES Somolu Lagos.

- Owen F. and Jones, R. (1978): Statistics. Polytech Publishers Ltd, Stockport.

- Oyesiku, O.K. and Omitogun, O. (1999): Statistics for social and Management Sciences. Higher Education Books Publisher, Lagos.

UNIT THREE: PRICE INDEX

CONTENTS

1.0 INTRODUCTION

In introductory statistics a lot of meaning has be given to the average, this (average) has been confirmed not to be necessarily representative of the data it describes. statisticians have constructed a device that attempt to measure the magnitude of economic change over time, a device called index number. This device is also used for international comparison of economic data. This device called index number is what this unit shall be looking at, we shall examine the basic principles by which index numbers are constructed.

2.0 OBJECTIVE

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

- Define index number
- Calculate the index number through different methods
- Use(s) of index number
- Relevance of index number

3.0 MAIN CONTENT

3.1 Index Number Defined

In statistical analysis of one very large and important class of problems, we must combine different set of data into a single measure e.g. we may wish to study the behaviour of wholesale prices and to do this, we calculate an index number which describes the changes, not in the various individual prices in which we are interested but in the group of prices taken as a whole.

The relevance of this statistical device is shown by the fact that governmental and other agencies devotes very substantial amount of money every year to the work of collecting appropriate data performing the necessary calculations for the construction of index numbers. The most widely known of these measure is the consumer price index or cost of living index.

In general, index numbers are used in the study of prices (wholesale, retail, form, export etc), output (manufacturing mining). The purpose of such measures is to get a summary of a whole range of similar activities, thereby, one will be able to investigate problem on relatively broad basis.

SELF-ASSESSMENT EXERCISE

What is the basis for an index number?

3.2 Computation of Index Number

Under this subsection we are going to look at the different index number available, how the index number can be calculated through different methods.

Price Relative Index Number

This method is usually in use where just one commodity is involved. It measures the rate of change in single commodity.

$$\text{Price relative} = \frac{P_n}{P_o} \text{ x } 100$$

Where P_n refer to price of current year and P_o represents the price of the base period or reference period.

Simple Price Index Number

Simple price index number is defined as the sum total of the price of related items divided by the sum total of its base or reference period.

$$\text{SPI} = \frac{\Sigma P_n}{\Sigma P_o} \text{ x } 100$$

Weighted Price Index Number

Here the concept weight is introduced to index number. These weights indicates the importance of the particular commodity depending on whether we use base year, given year or typical year quantities denoted by Q_o,$Q_{n.}$ We are going to look at the works and Marshall edge-worth Laspeyres, Paasche and Fisher on index number.

Laspeyres gave its own index number as

$$\text{LPI} = \frac{\Sigma P_n q_o}{\Sigma P_o q_o} \text{ x } \frac{100}{1}$$

$$\text{LQI} = \frac{\sum \text{PoQn x } 100}{\sum \text{PoQo}}$$

Paasche gave its own as

$$\text{PPI} = \frac{\Sigma P_n Q_n}{\Sigma P_o Q_n} \text{ x } 100$$

$$\mathbf{PQI} = \frac{\sum \mathbf{P_n Q_n \text{ x } 100}}{\sum \mathbf{P_n Q_o}}$$

Fisher’s Ideal Price Index

Fisher defined its own index number as the square root of the works of both Paashe and Laspeyres.

$$F = \sqrt{(\text{Laspeyres Index})\text{x } (\text{Paasche Index})}$$

$$= \sqrt{\left(\frac{\Sigma P_n Q_o}{\Sigma P_o Q_o}\right)\left(\frac{\Sigma P_n Q_o}{\Sigma P_o Q_n}\right) \text{x } 100}$$

Marshall Edge Worth Price Index

Marshall edge-worth defined its own index as

$$\text{MEPI} = \frac{\Sigma P_n (Q_o + Q_n)}{\Sigma P_o (Q_o + Q_n)} \text{ x } 100$$

$$\text{MEQI} = \frac{\Sigma Q_n (P_o + P_n)}{\Sigma Q_o (P_o + P_n)} \text{ x } 100$$

Where MEPI = Marshall edge-worth price index and

MEOI = Marshall edge-worth quantity index

3.3.0 Worked Example

Given the following about Open University, you are to compute the various price index numbers for 1990 using 1986 as base year.

Table M4.2.1

Table Showing Information about Open University

Commodity	**Quantities**		**Prices**	
	1986	**1990**	**1986**	**1990**
A	30	70	75	360
B	40	100	160	300
C	50	150	250	960
D	15	33	180	291

Solution

From the above 1986 values is the base year which represents the P_o values and P_n value is represented by 1990 values.

Table M4.2.2

Laspeyre, Paasche and Fisher's Table of Analysis

Price			**Quantities**					
Comm.	**1986**	**1990**	**1986**	**1990**				
	P_o	P_n	Q_o	Q_n	P_nQ_n	P_oQ_o	P_nQ_o	P_oQ_n
A	75	360	30	70	25,200	2,250	10,800	5,250
B	160	300	40	100	30,000	6,400	12,000	16,000
C	250	960	50	150	144,000	12,500	48,000	37,500
D	180	291	15	33	9,603	2,700	4,365	5,940
	665	**1,911**			**208,803**	**23,850**	**75,165**	**64,690**

$$SPI = \frac{\Sigma P_n}{\Sigma P_o} \text{ x } 100$$

$$= \frac{1911}{665} \text{ x } 100$$

$$= 287.36$$

This is simply implying that cost of the commodity had risen by 287.4% between 1986 and 1990.

$$LPI = \frac{\Sigma p_n q_o}{\Sigma p_o q_o} \text{ x } 100$$

$$LPI = \frac{75{,}165}{23{,}850} \text{ x } 100$$

$$LPI = 315.157$$

$$LQI = \frac{\Sigma P_o Q_n}{\Sigma P_o \, Q_o}$$

LQI = $\frac{64,690}{23,850}$

LQI = 271.2369

≅ 271%

Using Laspeyres price index it is showing the rate of rise in price as by 315.16% between 1986 and 1990.

Where LPI = Laspeyres Price Index

LPI = Laspeyres Quantity Index

Paasche method

PPI = $\frac{\Sigma P_n Q_n}{\Sigma P_o \, Q_n}$ x 100

PPI = $\frac{208,803}{64,690}$ x 100

PPI = 322.77%

PQI = $\frac{\Sigma p_n q_n}{\Sigma p_n q_o}$ x 100

PQI = $\frac{208,803}{75,165}$

PQI = 277.79

PQI ≅ 278%

Using Paasche price index the rate of increase in price is 322.8% between 1986 and 1990.

Fisher's Ideal Price Index

FPI = $\sqrt{\text{LPI x P. P I}}$

= $\sqrt{315.16 \text{ x } 322.77}$

= $\sqrt{101724.1932}$

= 318.942

FQI = $\sqrt{\text{LQI x PQI}}$

= $\sqrt{271.2369 \text{ x } 277.79286}$

FQI = $\sqrt{75{,}347.67419}$

FQI = 274.495

≅ 275%

Marshall Edge-Worth Index Number

TableM4.2.3

Marshal Edge-Worth Table of Analysis

P_o	P_n	Q_o	Q_n	$P_o+ P_n$	$Q_o+ Q_n$	$P_n(Q_o+ Q_n)$	$P_o(Q_o+ Q_n)$	$Q_o(P_o+ P_n)$	$Q_n(P_o+P_n)$
75	360	30	70	435	100	36,000	7,500	13,050	30,450
160	300	40	100	460	140	42,000	22,400	18,400	46,000
250	960	50	150	1201	200	192,200	50,000	60,500	181,500
180	291	15	33	471	48	13,968	8,640	7,065	15,543
						283,968	**88,540**	**99,015**	**273,493**

Using Marshall edge-worth price index

M.E. Price Index $= \frac{\Sigma P_n (Q_o + Q_n)}{\Sigma P_o (Q_o + Q_n)} \text{ x } 100$

$= \frac{283{,}968}{88{,}540} \text{ x } 100$

$= 320.722$

≅ 320.7%

M.E Quantity Index $= \frac{\Sigma Q_n (P_o + P_n)}{\Sigma Q_o (P_o + P_n)} \text{ x } 100$

MEQI $= \frac{273{,}493}{99{,}015}$

MEQI = 276.213

MEQI = 276%

SELF-ASSESSMENT EXERCISE

Define Fisher’s Ideal formular for price index?

4.0 CONCLUSION

In the course of our discussion on this unit you have learnt about

- Definition of price index
- Calculation about:

  Simple price index

  Weighted price index; where we talked about

  Laspeyres index number

  Paasche index number

  Fisher's ideal index number

  Marshall edge-worth index number

5.0 SUMMARY

Below is the summary of all the price indices we talked about in this unit.

Price relative = $\frac{P_n}{P_o}$ x 100

Simple price index = $\frac{\Sigma P_n}{\Sigma P_o}$ x 100

Laspeyres index = $\frac{\Sigma P_n Q_o}{\Sigma P_o Q_o}$ x100

Paasche index = $\frac{\Sigma P_n Q_n}{\Sigma P_o Q_o}$ x100

Fisher's ideal price index = $\sqrt{(\text{Laspeyre Index})\text{x } (\text{Paasche's Index})}$

Marshall Edge-worth Index

$$\text{MEPI} = \frac{\Sigma P_n\,(Q_o + Q_n)}{\Sigma P_o\,(Q_o + Q_n)}$$

$$\text{MEQI} = \frac{\Sigma Q_n\,(P_o + P_n)}{\Sigma Q_o\,(P_o + P_n)}$$

6.0 TUTOR MARKED ASSIGNMENT

Explain the weighted price index?

7.0 REFERENCES

- Adedayo, O.A. (2006): Understanding statistics. JAS Publishers, Lagos.

- Dawodu, A.F. (2008): Modern Business Statistics NIICHO Printing Works, Agbor, Delta State.

- Edward, E.L. (1983): Statistical Analysis in Economic and Business. (2nd edition) Houghton Mifflin Company, Boston.

- Esan, E.O. and Okafor, R.O. (2010): Basic statistical methods (Revised Edition) Toniichristo Concept, Lagos.

- Olufolabo, O.O. and Talabi, C.O. (2002): Principle and practice of statistics. HASFEM (NIG) Enterprises, Lagos.

- Owen, F. and Jones, R. (1978): Statistics. Polytech publishers Ltd, Stockport.

- Oyesiku, O.K. and Omitogun, O. (1999): Statistics for social and management sciences. (2nd edition) HEBP, Lagos.