NATIONAL OPEN UNIVERSITY OF NIGERIA

SCHOOL OF SCIENCE AND TECHNOLOGY

COURSE CODE: DAM363

COURSE TITLE: ECONOMIC STATISTICS

COURSE GUIDE

DAM363
ECONOMIC STATISTICS

Course Team Fagbamila Olaitan (Developer/Writer)
Dr. Bankole Abiola (Editor) - NOUN
Prof. Afolabi Adebanjo (Programme Leader) - NOUN
Greg Onwodi (Coordinator) - NOUN

NATIONAL OPEN UNIVERSITY OF NIGERIA

National Open University of Nigeria
Headquarters
14/16 Ahmadu Bello Way
Victoria Island
Lagos

Abuja Office
No. 5 Dar es Salaam Street
Off Aminu Kano Crescent
Wuse II, Abuja
Nigeria

e-mail: centralinfo@nou.edu.ng
URL: www.nou.edu.ng

Published By:
National Open University of Nigeria

First Printed 2011

ISBN: 978-058-047-7

All Rights Reserved

CONTENTS **PAGE**

Introduction

Economic Statistics is a two-credit first semester course. It is a course available to students offering Bachelor of Science (B.Sc) degree in computer and information communication technology.

Economic Statistics is applying statistics to economic problems. The scholar who specialises in the field of economic statistics is primarily an economic statistician and must be sound in the application of different statistical and economic tools to solve real life economic problems in the same way with his colleagues who specialise in other areas of economics and statistics. Indeed, there is no way we will make mention of real life economic problems without considering economic statistics as the means to solving the problems. The concept of applied statistics has various meanings to people in different fields. The concern of economic statisticians is with statisticians, economists, business people (men and women) and researchers (in any area of specialisation). Economic Statistics is a discipline which measures economic problems and provides solutions to them. The purpose underlying the study of Economic Statistics is to apply economic and statistical tools to solve real life economic problems.

What you will learn in this course

The course consists of units and a course guide. The course guide tells you briefly what the course is about, what course materials you will be using and how you can work with these materials. In addition, it advocates some general guidelines for the amount of time you are likely to spend on each unit of the course in order to complete it successfully.

It gives you guidance in respect of your Tutor-Marked Assignment which will be made available in the assignment file. There will be regular tutorial classes that are related to the course. It is advisable for you to attend these tutorial sessions. The course will prepare you for the challenges you will meet in economic statistics.

Course Aims

The aim of the course is to provide you with an understanding of economic statistics. It also aims to provide you solutions to real life economic problems.

Course Objectives

To achieve the aims set out, the course has a set of objectives. Each unit has specific objectives which are included at the beginning of the

unit. You should read these objectives before you study the unit. You may wish to refer to them during your study to check on your progress. You should always look at the unit objectives after the completion of each unit. By doing so, you would have followed the instructions in the unit.

Below are comprehensive objectives of the course as a whole. By meeting these objectives, you should have achieved the aims of the course as a whole.

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

- explain the concept of applied statistics
- extrapolate the application of applied statistics on real life economic problems
- describe measures of dispersion and probability theory
- explain the concept of estimation and hypothesis testing in statistical inference
- explain index numbers and its application to solve economic problems
- explain the concept of purchasing power parity and its application to evaluate the economic health of nations
- explain the concept of human development and its measurement
- describe the basic concept of corruption indexes, effects, problems and the solutions to the problems of corruption on the economy
- describe the concept of economic liberalisation and globalisation.

Working through this Course

To complete this course, you are required to read each study unit, read the textbooks and read other materials which may be provided by the National Open University of Nigeria.

Each unit contains self assessment exercises and at certain points in the course, you would be required to submit assignments for assessment purposes. At the end of the course, there is a final examination. The course should take you about 17 weeks to complete. Below, you will find listed, all the components of the course, what you have to do and how you should allocate your time to each unit in order to complete the course on time and successfully.
This course entails that you spend a lot of time to read. I would advise that you avail yourself the opportunity of attending the tutorial sessions where you have the opportunity of comparing your knowledge with that of other people.

The Course Materials

The main components of the course are the followings.

1. The course material
2. Study units
3. References/Further Reading
4. Assignments
5. Presentation Schedule

Study Units

Module 1 Introduction to Economic Statistics

Unit 1 Statistics as an Applied Tool to Economic Problems
Unit 2 Descriptive Statistics: Measures of Location
Unit 3 Descriptive Statistics: Measures of Dispersion
Unit 4 Introduction to Probability and Distribution

Module 2 Statistical Inference

Unit 1 Estimation
Unit 2 Hypothesis Testing

Module 3 Index Number

Unit 1 Basic Understanding of Index Number
Unit 2 CPI and other Forms of Index Number
Unit 3 Application of Index Number to Economic Problems
Unit 4 Measure of Macroeconomic Variables (Real and Nominal Values)

Module 4 Purchasing Power Parity

Unit 1 The Meaning of Purchasing Power Parity
Unit 2 Purchasing Power Parity as Exchange Rate Determination

Module 5 Human Development

Unit 1 Human Development Indices

Module 6 Measure of Corruption

Unit 1 Measure of Corruption as a Menace to the Economy

Module 7 Economic Liberalisation

Unit 1 The meaning of Economic Liberalisation

The first unit focuses on the history, basic concept of economic statistics, data classification, mode of data collection and various sampling techniques. The second unit deals with measures of location while the third focuses on measures of dispersion. The fourth unit is concerned with introduction to probability and probability distribution. Unit five treats estimation theory and confidence interval. Unit six focuses on test of hypothesis test statistics (TS) and error in hypothesis test.

Units seven, eight and nine are concerned with basic concepts of index numbers, CPI and other forms of index numbers and application of index numbers to economic problems. Unit ten is on real and nominal values as measurement of macroeconomic variables. Units eleven and twelve focus on the meaning of power purchasing parity (PPP) and PPP as an exchange rate determination.

The 13th unit is concerned with the concept of human development and human development indices. Unit 14 deals with the meaning of corruption, corruption indexes, effects of corruption on the economy and also, problems and solutions to corruption Unit 15 focuses on the concepts of liberalisation and globalisation and the effects of liberalisation and globalisation on developing countries.

Each unit consists of one or two weeks work and includes introductions, objectives, main content, self assessment exercises, conclusion, summary, tutor-marked assignments (TMAs),references/ further reading. The unit directs you to work on exercises related to the required reading. In general, these exercises test you on the materials you have just covered and thereby assist you to evaluate your progress and to reinforce your comprehension of the material. Together with the TMAs, these exercises will help you in achieving the stated learning objectives of the individual units and of the course as a whole.

Presentation Schedule

Your course materials have important dates for the early and timely completion and submission of your TMAs and attending tutorials. You should remember that you are required to submit all your assignments by the stipulated time and date. You should guard against falling behind in your work.

Assessment

There are three aspects to the assessment of the course. The first is made up of self assessment exercises, the second consists of the tutor-marked

assignments and the third is the written examination/end of course examination.

You are advised to do the exercises. In tackling the assignments, you are expected to apply information, knowledge and technique you gathered during the course. The assignments must be submitted to your facilitator for formal assessment in accordance with the deadlines stated in the presentation schedule and the assignment file. The work you submit to your tutor for assessment will count for 30% of your total course work. At the end of the course, you will need to sit for a final or end of course examination of about three hour duration. This examination will count for 70% of your total course mark.

Tutor-Marked Assignments (TMA)

The TMA is a continuous assessment component of your course. It accounts for 30% of the total score. You will be given four TMAs to answer. Three of these must be answered before you are allowed to sit for the end of course examination. The TMAs would be given to you by your facilitator and returned after you have done the assignment. Assignment questions for the units in this course are contained in the assignment file. You will be able to complete your assignment from the information and material contained in your reading, references and study units. However, it is desirable in all degree level of education to demonstrate that you have read and researched more into your references, which will give you a wider view point and may provide you with a deeper understanding of the subject.

Make sure that each assignment reaches your facilitator on or before the deadline given in the presentation schedule and assignment file. If for any reason you cannot complete your work on time, contact your facilitator before the assignment is due to discuss the possibility of an extension. Extension will not be granted after the due date unless there are exceptional circumstances.

Final Examination and Grading

The end of course examination for economic statistics will be for about three hours and it has a value of 70% of the total course work. The examination will consist of questions, which will reflect the type of self-testing, practice exercise and tutor-marked assignment problems you have previously encountered. All areas of the course will be assessed.

Revise the whole course after finishing the last unit and while waiting for the time to sit for the final examination. You might

find it useful to review your self assessment exercises, TMAs and comments written on them by your facilitator before the final examination. The final examination covers information from all parts of the course.

Course Marking Scheme

Assignment	Marks
Assignments 1-4	Four assignments, best three marks of the four at 10% each -30% of course marks.
End of course examination	70% of overall course marks.
Total	100% .

Facilitators/Tutors and Tutorials

There are 16 hours of tutorials provided in support of this course. You will be notified of the dates, times and location of these tutorials as well as the name and phone numbers of your facilitator, as soon as you are allocated a tutorial group.

Your facilitator will mark and comment on your assignments. Keep a close watch on your progress and other difficulties you might face and report these to your facilitator who will provide assistance to you during the course. You are expected to mail your Tutor-Marked Assignments to your facilitator before the scheduled date (at least two working days are required). The assignments will be marked by your tutor and returned to you as soon as possible.

Do not delay to contact your facilitator by telephone or e-mail if you need assistance.

The following might be circumstances in which you would find assistance necessary, hence you would have to contact your facilitator if:

- you do not understand any part of the study or the assigned readings
- you have difficulty with the self-tests
- you have a question or problem with an assignment or with the grading of an assignment.

You should endeavour to attend the tutorials. This is the only chance to have face to face contact with your course facilitator and to ask questions which are answered instantly. You can raise any problem encountered in the course of your study.

To gain much benefit from course tutorials, prepare a question list before attending them. You will learn a lot from participating actively in discussions.

Summary

Economic Statistics is a course that intends to provide solutions to real life economic problems. It also serves as a tool, which often enables the economists, to widen the frontiers of their analytical concerns to issues that have significant economic implications. Nevertheless, do not forget to apply the principles you have learnt to your understanding of economics and statistics. I wish you success in the course and I hope that you will find it comprehensive and interesting.

Course Code DAM363
Course Title Economic Statistics

Course Team Fagbamila Olaitan (Developer/Writer)
Dr. Bankole Abiola (Editor) - NOUN
Prof. Afolabi Adebanjo (Programme Leader) - NOUN
Greg Onwodi (Coordinator) - NOUN

NATIONAL OPEN UNIVERSITY OF NIGERIA

National Open University of Nigeria
Headquarters
14/16 Ahmadu Bello Way
Victoria Island
Lagos

Abuja Office
No. 5 Dar es Salaam Street
Off Aminu Kano Crescent
Wuse II, Abuja
Nigeria

e-mail: centralinfo@nou.edu.ng
URL: www.nou.edu.ng

Published By:
National Open University of Nigeria

First Printed 2011

ISBN: 978-058-047-7

All Rights Reserved

CONTENTS **PAGE**

MODULE 1 INTRODUCTION TO ECONOMIC STATISTICS

UNIT 1 STATISTICS AS AN APPLIED TOOL TO ECONOMIC PROBLEMS

CONTENTS

1.0 INTRODUCTION

Economic Statistics is a branch of applied statistics focusing on the collection, compilation and dissemination of statistics concerning the economy of a region, a country or a group of countries. Economic statistics is also referred to as a subtopic of official statistics, since most of the economic statistics are produced by official organisations (e.g. statistical institutes, intergovernmental organisations such as United Nations, European Union or OECD, central banks, ministries, etc.). Economic Statistics provides the empirical data needed in economic research (econometrics) and it is the basis for decision and economic policy making.

2.0 OBJECTIVES

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

- define statistics
- describe various sampling techniques
- explain various types and methods of data collection
- distinguish between the following: primary and secondary data, sample and population.

3.0 MAIN CONTENT

3.1 What is Statistics?

Statistics is the body of theory, concepts, methods and methodologies of collection, analysis, interpretation and representation of numerical data relating to socio-economic and political activities. The analysis of such data enables us to make valid conclusion and reasonable decisions about the present, near and distant future of any business activities

Statistics also plays a dual role, firstly: It is studied to learn how to gather together and summarise data in such a way as to make them readily comprehensible; this is called descriptive statistics. Secondly, it may be studied primarily with the aim of learning to draw conclusions from data; this is called inferential statistics.

The two must go hand in hand. First, we must collect data in such a way that meaningful results may be anticipated; then we describe the data, and lastly draw inferences from it.

The field of Statistics can be divided into two main parts:

1. **Mathematical Statistics:** This concerns the development of new methods of statistical inference and requires detailed knowledge of abstract mathematics for its implementation.
2. **Applied Statistics:** This concerns the application of the methods of mathematical statistics to specific subject areas, such as economics, psychology, business, public health, demography, biostatistics etc.

3.2 Collecting and Analysing Statistical Data

Statistical analysis begins with the collection of data. The way this is done is probably the most important aspect of the investigation because, any interpretation of the figures will be faulty if the data on which it is based is unreliable; when this happens, bias does exist.

3.2.1 Sources of Bias in Statistical Data

This exists when:

i.. using the data from responses to questions e.g. if information about the ages of a group of people is gathered by asking them to fill a questionnaire, over aged people may enter less than their true ages. In fact, people do lie, for many different reasons, in response to questionnaires administered to them

ii. using data that does not properly apply to the problem being investigated e.g. using published number of AIDS victims to investigate the number of people with AIDS. The figures on AIDS victims are based on the number of people who registered as having the disease. This does not include other categories of people who have the disease but are often guilty of declaring it.

3.2.2 Evaluation of Statistical Investigations

Statistical investigation is only good if the quality of the data on which it is based is good. If the data is faulty, any conclusion, however good the analysis is, will not be reliable. Then, the following questions should be asked about the data used in any investigation.

- Where did the data come from?
- Who is supplying it and why?
- How was the data collected?
- Is it all the relevant data or a sample?
- If a sample is used, how was the sample chosen?
- Is the data relevant to the investigation?
- Does the conclusion follow from the information?

3.2.3 Why do we study Statistics?

Statistics is basically applied for the purpose of describing data and making inferences.

It shows the interrelationship between variables in a clearer form and provides good measure of comparison.

It is also used to understand properly the numerical information in a more readable form.

Suppose that, an education authority needs to decide whether or not to build a new primary school. One factor in this decision must be the likely demand for schools in the future. Then knowledge, about the birth-rate, migration, the age profile of the population etc. can show a trend which can help in making judgments about future demands. This is not the only consideration, but a decision made without the statistical information is less likely to be a good one. The more you know about statistics, the better you are able to make reasonable decisions based on statistical evidence.

3.2.4 Data and Information

In many situations where we require information, it is impossible to collect all the necessary facts. We may wish to know, for instance, average annual salary for all females currently employed in Nigeria. It would be totally impractical to try and find each appropriate individual annual salary and then average them. Instead, we might select a suitable sample of females currently employed and average their annual salaries. This average would be our estimate for the overall average annual salary of females currently employed in Nigeria.

Running any business organisation requires statistical information to be obtained e.g. planning future sales targets requires past sales data to be obtained; making up employees wages requires data from clock cards to be obtained. Once statistical information is available, it can be used by managers to obtain the best results consistent with the objectives of the organisation.

3.2.5 Data

Data are raw facts, unassembled and frequently unrelated to one another. It relates to facts, events, transactions etc.

3.2.6 Information

If a data is processed in some ways, it becomes information. Then, information can be defined as a collection of related places of data.

Ways of transforming data into Information

(i) By bringing related pieces of data together
(ii) By summarising data
(iii) By tabulation and diagrammatic techniques
(iv) By statistical analysis
(v) Final report

3.2.7 Types of Data

There are two types of data, namely: primary and secondary data.

3.2.8 Primary Data

This data is especially collected for the purpose for which it is required. This is the data that one collects by himself. We may collect information from the whole population or use sampling by investigating only part of the population. Primary data helps us to understand any curiosity that it contains.

Methods of Collecting Primary Data

(i) **Direct Measurement:** Where the information we require can be measured or counted, we can arrange for ourselves or someone else to take the necessary measurements directly without recourse to a third party influence. This method has the distinct advantage that accurate measurements can be obtained. It can however, turn out to be very time-consuming and costly if misused. Furthermore, there are only certain measurements which can be taken directly.

(ii) **Experimentation:** This is used more by the scientists than the businessmen and is linked closely to the method of direct observation. It is a method where a particular treatment is given to the population or the people to determine the effect it has on the population e.g. applying fertilizer on different crops to determine their yields, determine the consumers' behaviour towards a new product due to advertisement etc.

(iii) **By Observation:** This involves watching or counting events as they happen and it is used primarily for traffic censuses and in work study. Collections of physical, chemical, biological, traffic flow etc. fall under this category. But in most cases, it would he

uneconomical and in certain circumstances inapplicable. For example, it would be ridiculous to follow people around every day to find out how often they clean their teeth.

(iv) **Questionnaire:** This method is mostly prominent in social sciences. This is where an interviewer has a set of questions (the questionnaire) which he or she administers to the respondents (persons chosen to be in the sample). This is a method often adopted in town centres or when interviewers visit the respondents' homes or places of work. It has the advantage that the results can be obtained quickly and reliably. Due to the high response rate obtained when using this method, it is usually considered the most effective way of obtaining data. This method is subdivided into three parts - by post, face-to-face interview and by phone.

3.2.9 Secondary Data

This is the data which had already been collected by us or some other agencies for a different purpose but which is relevant to our present investigation. It is the data taken from some other sources.

3.2.10 Sources of Secondary Data

i. Those produced by individual companies, local authorities, trade union, pressure groups, labour research, financial times etc.
ii. Through government statistics such as Federal Bureau of Statistics (FOS), annual abstracts, national income and expenditure register, employment gazette, business monitors etc.

3.2.11 Problems Associated with Secondary Data

i. The major problem with secondary data is that it has not been collected specifically for the purpose required. This means that if we are to use it, we must check to ensure we understand what the figures collected represent, that is, the population which we are investigating the figures of, to the accuracy we require and finally, the trustworthiness of the figures.
ii. Is the information actual, seasonally adjusted, estimated or projected?
iii. The reason for collecting the data may be unknown.
iv. The data may be incomplete.

3.2.12 Population and Sample

Non-statistician describes population as a group of people living in a delineated territory at a specified time. However, a statistical

population covers a much wider field and does apply to people and objects.

3.2.13 Population

This is the whole group of items or people we are concerned with. Then, if one examines a small part of the group it is called a sample. A population can be finite or infinite; a finite population is the one in which the number is countable otherwise, it is infinite e.g. the population consisting the number of bournvita tins produced in a day by Cadbury Nig. Plc is finite while the population consisting of all possible outcomes (heads or tails) in successive tosses of a coin is infinite.

3.2.14 Sample

If deciding that collection of statistical data from a whole population is inappropriate, time consuming and capital intensive, then we need to choose a sample. But the sample must be unbiased and chosen from the right population. How do we select our sample? Many techniques of sampling are in use. Meanwhile, before any technique can be considered, the situation, including limitations of cost and time, must be considered.

3.3 Methods of Sampling

There are four different methods of taking samples from population.

(i) Simple Random Sampling

A random sample is a sample selected in such a way that every member of the population has an equal chance of being selected. It is a definite method of selection aimed at eliminating bias as far as possible. This can be achieved by numbering every member of a population, putting their numbers in a bag, mixing the numbers together and drawing out numbers of those to be in the sample e.g. raffle or lottery. Also, a computer can be programmed to turn out a series of random numbers which will identify the items to be used in the sample. If a computer cannot be afforded then, a table of random numbers can be used.

(ii) Systematic Sampling

This is a method used where we sample each item at k^{th} interval. It is the sampling at regular intervals from an ordered sampling frame. Consider a sample of individuals drawn from a population of N individuals by:

(a) choosing one individual at random
(b) choosing every k^{th} individual thereafter, returning to the beginning of the list.

When you have reached the end, k could be chosen as $^{N}/n$.

If a list of all items in a population exists (e.g. electoral register) or if the population items are filed in order e.g. (invoices/credit notes) then systematic sampling may be used.

(iii) Multi-stage Sampling

If the simple or stratified random sampling is considered to be time consuming and capital intensive then, multistage sampling may be used. In multi-stage sampling, we select a few areas which we believe are representative of the population as a whole. We then take a random sample within each of these areas, when the population is well spread out, particularly geographically. Simple random sampling will result in a dispersed sample. As an enumerator, you would probably object to doing a grand tour of the area, asking questions of only one particular person in each place you visited. Some methods are needed to 'narrow the field down' to a smaller area. Multi-stage attempts to do this without adversely affecting the randomness of the result.

The first step is to divide the population into manageable, convenient groups or areas. The process may be taken one (or several) stages further to reduce the sample area to more manageable proportions. Finally, a simple random or systematic sample will be used to select the eligible respondents.

(iv) Cluster Sampling

This is a selection from randomly chosen groups of neighbouring individuals. It is easy to confuse this with multi-stage sampling as the first-stage appears to be identical. The major difference is that cluster sampling is used when the population has not been listed and it is the only way to obtain a sample. It is not used primarily as means of reducing costs.

(v) Stratified Sampling

This is the act of dividing the sample frame into non-overlapping subsets, with proportionate simple random sampling from each subset. In some situations, it is possible to improve on simple random sampling by stratification of the population. This is particularly true where the population is heterogeneous (i.e., made up of dissimilar groups) and the population can be stratified into homogeneous (i.e. similar) classes.

(vi) Quota Sampling

This is non-random sampling of targeted sets of individuals. This is done by specifying how many people or items within a certain group you want to be sampled (set a quota) and then collect data from anyone or anything fitting into the required category until the quota is filled. This method is widely used by interviewers encountered in town centres, streets and market research. The main advantage of this method is that, if a respondent refuses to answer the questions for any reason, the interviewer will just look for another person in the same category.

(vii) Judgment Sampling

Here, an expert, or a team of experts, uses personal judgement to select what a truly representative sample would be. It certainly cannot be called a random sample as it involves human judgement which could involve bias.

(viii) Convenience Sampling

The most important factor here is the ease of selecting the sample. No effort is made to introduce any element of randomness. An example of this is a quality controller who takes the first 50 items of the production line as his sample. This is a dangerous procedure as any fault in the process occurring after or before this could remain unnoticed until the next sample is taken.

(ix) Sampling with Probability Proportional to Size

This is a method of sampling in the first or second stage of a multi-stage sample. It is designed to ensure that the probability of selection of a sub-unit is proportional to the number of elements of the population in the sub-unit.

3.4 Variables

A variable is a measurement which varies from one individual to another or from one item to another e.g. the number of mistakes in a set of accounts, the weight of a steel bar, age, time etc.

3.4.1 Types of Variables

i. Discrete Variables

These are variables which can only take whole numbers or integer values e.g. the number of people who watch a particular programme on

the television, the number of people sitting for an examination etc. Generally, when we count things, the answers we get are whole numbers, these are discrete variables.

ii. Continuous Variables

These are variables, which can take any value within a certain range, so a decimal or fractional value can be obtained for continuous variables e.g. the length of a bar, the time to produce a particular item, the temperature of a place at a particular time, volume, average etc.

3.4.2 Types of Data

Discrete Data

It consists of numerical values in situations where one can make a list of the possible values e.g. 1, 2, 3, 4.

Continuous Data

It consists of numerical values in cases where it is not possible to make a list of outcomes e.g. height, mass, time, weight etc.

Data can also be classified as either Qualitative or Quantitative.

(a) Qualitative Data

i. Nominal

This consists of unordered 'either' or observations e.g. Dead or Alive, Male or Female, Cured or Not Cured, Blood group A, O, B or AB, Pregnant or Not Pregnant; Infected or Not Infected etc.

ii. Ordered Classifications

If there are more than two categories of classifications, it may be possible to order them in some ways: e.g. after treatment, a patient can be improved; the same or worse.

iii. Ranked Data

In some studies, it may be appropriate to use classifications that can be ranked e.g. patients with rheumatoid arthritis may be asked to order their preference for four dressing aids. Here, although numerical values are assigned to each aid, one cannot assume they behave as numbers. They are in fact only codes for best, second best,

third best and worst. Another example is that a feature of the patient may be classified as being either 'nil', 'mild', 'moderate', 'severe', 'very severe' on a rating scale may classify the patient as being in one of the grades I, II, III, IV or V.

It is however, important that the use of numbers to label categories or grades is not taken to imply that a truly quantitative measurement has been obtained; the numbers are being used purely for their ordinal relationship to each other.

(b) Numerical or Quantitative Data

i. Numerical Discrete

This consists of counts e.g. numbers of babies born to 1000 women in 3 years, number of customers who patronise 'Mr. Bigg's' daily, etc.

ii. Numerical Continuous

Such data are measurements that can in theory at least, take any value within a given range e.g. gestation period, ovulation period, birth weight of baby etc.

3.5 Errors

Statistical errors are majorly concerned with estimating an uncertainty. It is also concerned about the size, relative size and nature of errors. Two questions are normally asked concerning statistical errors.

(i) Are the errors always positive?
(ii) Do these errors cancel out?

3.5.1 Statistical Errors

In statistical investigation, attempt is always made to estimate values or quantities. Estimations such as the followings may occur.

(i) What is the average stock we are expecting this week from a particular company?
(ii) What is the profit that will be made in two years?
(iii) How many people will become Chartered Accountants in five years? These are always made before the actual occurrence of any event. But, whatever it is to be estimated, one is subject to getting the wrong answer or prediction.

3.5.2 Nature of Statistical Errors

(a) Bias

If the estimates of values are either consistently above or below the correct or target value, then the tendency to overestimate or underestimate is called bias.

(b) Precision

Many results are usually collected or obtained in most statistical investigations. Consistency of these results with one another would be an indication of how good the estimates are. The closer the results, the more precision they are said to have. Precision is the term used to describe how consistent the results are to one another.

(c) Accuracy

Accuracy is used to describe how close a result is to the true target value. It is however, important to note that precision and accuracy are not the same. A set of results could be very similar to one another, but despite that, it could be above or below the true value and hence may not be accurate.

3.6 History of Statistics

Some scholars pinpoint the origin of statistics to 1662, with the publication of *Natural and Political Observations upon the Bills of Mortality* by John Graunt. Early applications of statistical thinking revolved around the needs of states to base policy on demographic and economic data, hence its *stat-* etymology. The scope of the discipline of statistics broadened in the early 19th century to include the collection and analysis of data in general. Today, statistics is widely employed in government, business, and the natural and social sciences.

Because of its empirical roots and its focus on applications, statistics is usually considered to be a distinct mathematical science rather than a branch of Mathematics. Its mathematical foundations were laid in the 17th century with the development of probability theory by Blaise Pascal and Pierre de Fermat. Probability theory arose from the study of games of chance. The method of least squares was first described by Carl Friedrich Gauss around 1794. The use of modern computers has expedited large-scale statistical computation, and has also made possible new methods that are impractical to perform manually.

SELF ASSESSMENT EXERCISE

1. Briefly explain the following terms:

(i) Statistics error
(ii) Simple random sampling, systematic sampling

2. Differentiate between the following terms:

(i) Sample and population
(ii) Quantitative and quantitative data
(iii) Data and information

4.0 CONCLUSION

This unit has served to demonstrate the relevance of statistical analysis, how statistics can be used in many situations, and how you, as an analyst in private, business, and public life can perform useful statistical analyses.

5.0 SUMMARY

In this study, we have learnt that:

- economic Statistics is an applied statistics to solve economic problems
- sample can be distinguished from population and the difference is that population is the whole subject in consideration while sample is a subset of that population.
- the two sources of data collection are primary and secondary.
- there are different sampling techniques such as simple random sampling, systematic random sampling etc.

6.0 TUTOR-MARKED ASSIGNMENT

1. What are the advantages of primary data over secondary data?
2. How can error be minimised in a statistical study?
3. Explain with examples two types of variables.

7.0 REFERENCES /FURTHER READING

Anderson, D. R.; Sweeney, D. J. & Williams, T.A. (1986). *Statistics: Concepts and Applications*, pp. 5 - 9. West Publishing Company.

Chance,B. L. & Rossman, A. J. (2005). *Investigating Statistical Concepts, Applications, and Methods*. Duxbury Press. ISBN 978-0495050643.

Giovanini, E. (2008).*Understanding Economic Statistics*, OECD Publishing, ISBN 978-92-64-03312-2.

Moses, L. E. (1986). *Think and Explain with Statistics*, pp. 1 - 3. Addison-Wesley.

Willcox, W. (1938). *The Founder of Statistics.* Review of the International Statistical Institute 5(4):321-328.

UNIT 2 DESCRIPTIVE STATISTICS: MEASURE OF LOCATION

CONTENTS

1.0 INTRODUCTION

This is one of the methods of summary statistics for univarite data (i.e. data basically concerned with a single quantity).

These measures assume that the variables or the data provided concentrate at a point. They answer the arithmetic mean, geometric mean, harmonic mean, median and mode of the distribution.

2.0 OBJECTIVES

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

- explain measure of location
- calculate the mean, median and mode for grouped and ungrouped data
- enumerate the advantages and disadvantages of measure of location.

3.0 MAIN CONTENT

3.1 Arithmetic Mean

If we have a list of items or values for a discrete variable for ungrouped data, then the arithmetical mean is the simple average of the values. It is calculated by adding all the values together and dividing by the total number of values in our list.

This is represented as

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

Where n is the number of observations
If the observation occurs frequently then, it changes to

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

Where f is the frequency

Example 1

The heights in meters of 12 walnut plants, after twenty years growth are 4, 5, 6, 3, 7, 2, 8, 5, 3, 4.8, 4.5, 7.2. 5.2 The mean is

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

$$= \frac{4+5+6+3+7+2+5.5+4.8+4.5+7.2+5.1}{12}$$

$$= \frac{62}{12} = 5.17$$

Example 2

The number of employees in 7 commercial banks in Lagos is given as 4368, 4387, 4334, 4383, 4356, and 4369. Find the mean

Solution

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

$$= \frac{4368+4387+4697+4331+4383+4356+4369}{7}$$

$$= \frac{30591}{7} = 4370.14$$

Assumed Mean method

For an ungrouped data:

$$\text{Mean} = A + \frac{\sum d_i}{n}$$

Where A is the assumed mean

d_i is the deviation from assumed mean
n is the number of observations

Example 3

Using Example 2 above, compute the mean, using assumed mean method.

Solution

The value A must be chosen within the range of the given values for easy calculation.

Here, let A = 4369

d = xi –A

X	4368	4387	4397	4331	4383	4356	4369	$\sum d_i$
D	-1	18	28	-38	14	-13	0	8

$$\therefore \text{ Mean} = A + \frac{\sum d_i}{n}$$

$$= 4369 + \frac{8}{7}$$

$$= 4370.14$$

Mean for Grouped: Discrete Data

If the data is given as a frequency table with frequencies for individual values or if the observations occur frequently then;

$$\text{Mean} = \frac{\sum fx}{\sum f}$$

Where f is the frequency or number of observations

Example 4

The table below shows the number of chocolates bought in a super market per day

X	0	1	2	3	4	6
F	3	9	15	23	11	2

Find the mean?

Solution

Mean = $\bar{x}$

$$= \frac{0\times3+1\times9+2\times15+3\times23+4\times11+6\times2}{3+9+15+23+11+12}$$

$$= \frac{164}{63} = 2.603$$

Mean for Grouped Data with Class Intervals

Here, to calculate the mean we start by adjusting the interval so that they are the equivalent of continuous intervals. Having done this, we encounter the problem which always arises with grouped data, because we do not know the exact value for each result in an interval. This makes it necessary to approximate and this is done by assuming that all values in an interval are at the mid-point of the interval.

Example 5

The data below shows the weight (kg) of 20 students in a class. Find the mean weight.

Weight (kg)	46-50	51-55	56-60	61-65
Number of students	7	5	6	2

Solution

Mean = $\frac{\sum fx}{\sum f}$

Weight	f	x	fx
46-50	7	48	336
51-55	5	53	265
56-60	6	58	348
61-65	2	63	126
	20		1075

$= \frac{1075}{20}$

= 53.75kg

Assumed Mean Method of Estimating Arithmetic Mean

The technique involves the use of the formula

$$\text{Mean } (\bar{x}) = A + \frac{\sum fd}{\sum f}$$

Where'd' is the deviation from the assumed mean
'A' is the assumed mean, which may be taken within or outside the distribution. However, it is better to choose the value of A within the distribution.

Example 6

Use the data in example 6 above to compute the mean using assumed mean method.

Solution

Choose the assumed man to be 35.

Speed Vkm/h	f	Mid-Point x	d=x-A	fd
10-20	32	15	-20	-640
20-30	38	25	-10	-380
30-40	63	35	0	0
40-50	29	45	10	290
50-60	38	55	20	760
	200			30

$$\bar{x} = \text{A} + \frac{\sum fd}{\sum f}$$

$$= 35 + \frac{30}{200} = 35.15\text{km/h}$$

Coding Method of Estimating Arithmetic Mean

The technique involves the use of the formula:

$$\bar{x} = \text{A} + \left[\frac{\sum fp}{\sum f}\right]C$$

Where A is the assumed mean
C is the class width or size

$$P= \frac{x-A}{c}$$

Example 7

Using the data from the example above compute the mean, using coding method.

Solution

Speed Vkm/h	f	Mid-Point x	D=x-A	$P= \frac{x-A}{c}$	fp
10-20	32	15	-20	-2	-64
20-30	38	25	-10	-1	-38
30-40	63	35	0	0	0
40-50	29	45	10	1	29
50-60	38	55	20	2	76
	200				3

$$\bar{x} = A + \left[\frac{\sum fp}{\sum f}\right]C$$

Let the assumed mean be 35

A=35 , $\sum f = 200, \sum fp = 3$ c = 10

$$\bar{x} = 35 + \left[\frac{3}{200}\right]10$$

= 35+0.15
= 35.15km/h

Example 8

The lengths of steel bars produced by Eldorado Nig. Ltd over a period are given in the table below:

Length (m)	<5	5≤x≤15	10≤x≤15	15≤x≤20	20≤x≤25	≥25
Number produced	4	22	27	32	68	3

i. Estimate the mean using
ii. Simple method
iii. Assumed mean method
iv. Coding method

Compare your results

Solution

Length (m)	No. prod. F	Mid Class x	Fx	D= x-A	Df	P= $\frac{x-A}{c}$	fp
<5	4	2.5	10	-15	-60	-3	-12
5≤x≤15	22	7.5	165	-10	-220	-2	-44
10≤x≤15	27	12.5	337.5	-5	-135	-1	-27
15≤x≤20	32	17.5	560	0	0	0	0
20≤x≤25	68	22.5	1530	5	340	1	68
≥25	3	27.5	82.5	10	30	2	6
	200		2685		-45		-9

(i) Simple Method:

Mean= $\frac{\sum fx}{\sum f}$

= $\frac{2685}{156}$ = 17.21

(ii) Assumed Mean Method

Let the assumed mean A = 17.5

$\therefore$ Mean = A + $\frac{\sum fd}{\sum f}$

= 17.5 + $\left[\frac{-45}{156}\right]$ = 17.5 – 0.289 = 17.21

Coding method:

Let the Assumed Mean = 17.5, c = 5

Mean = A + $\left[\frac{\sum fp}{\sum f}\right]C$ = 17.5+ $\left[\frac{-9}{156}\right]5$ = 17.5 – 0.289 = 17.21

Using the three methods the mean is the same.

Advantages of Arithmetic Mean

i. It is easy to compute
ii. It has determinate exact value
iii. The calculation is clear and precise
iv. It uses all the items in the data in its calculation
v. It provides a good measure of comparison.
vi. It can be used for further statistical calculations or processing.

Disadvantages of Arithmetic Mean

i. It cannot be obtained graphically except if the graph is symmetric.
ii. If some values are missing it cannot be calculated.
iii. It can be distorted by extreme values in the distribution.
iv. The mean may be significantly affected by the inclusion of mistake in the observation.

Geometric Mean

This can be calculated using the formula below:

$$G.M = \sqrt[n]{x_1^{f_1} x_2^{f_2} \ldots . x_m^{f_m}}$$

When $n \sum f$

Example 9

Find the geometric mean of the following data 3, 2, 6, 4

Solution

$$G.M. = \sqrt[n]{3x2x6x4}$$

$$= \sqrt[4]{144} = 3.46$$

Harmonic Mean

Consider n observation $x_1, x_2, \ldots . x_n$

$$H.M. = \frac{n}{\sum_{i=1}^{n} \frac{1}{x_i}}$$

For Grouped data

$$\text{H.M} = \frac{n}{\sum_{i=1}^{n}\left[\frac{f_i}{x_i}\right]}$$

Example 10

Find the harmonic mean of the following data 3, 2, 6, 4

Solution

$$\text{H.M} = \frac{n}{\sum_{i=1}^{n}\frac{1}{x_i}} = \frac{4}{\frac{1}{3}+\frac{1}{2}+\frac{1}{6}+\frac{1}{4}} = \frac{4}{\frac{15}{12}} = 4 \text{ x } \frac{12^{4}}{15_{5}} = \frac{16}{5} = 3.2$$

Relationship between Arithmetic, Geometric and Harmonic Mean

The relationship that exists among them is H.M. ≤G.M.≤A.M. This inequality could be verified using a set of data.

Example 11

The number of times a patient takes a prescribed drug over a period of days is given in the data below; 3,5,2, 6. Compute the arithmetic mean, harmonic mean and geometric mean. Hence, compare your results.

Solution

Arithmetic Mean (A.M.) = $\frac{\sum x}{n}$

= $\frac{3+5+2+6}{4}$

= $\frac{16}{4}$ = 4

Geometric Mean (G.M.) = $\sqrt[n]{x_1, x_2, x_3, x_4}$

= $\sqrt[4]{3x5x2x6}$

= 3.66

$$\text{Harmonic Mean (H.M.)} = \frac{n}{\sum_{i=1}^{n} \frac{1}{x_i}} = \frac{4}{\frac{1}{3}+\frac{1}{5}+\frac{1}{2}+\frac{1}{6}}$$

$$= \frac{4}{\frac{36}{30}}$$

$$= 4 \text{ x } \frac{30}{36} = 3.33$$

From the results above it is obvious that H.M. <G.M.<A.M.

The equality holds i.e. H.M. = G.M. = A.M., only if all the items in the data are equal or the same.

Root Mean Square (R.M.S)

If n is the number of occurrence of a particular variable x i.e. x_1, $x_2 \ldots x_n$. Then the root mean square

$$\text{R.M.S.} = \sqrt{\frac{\sum_{i=1}^{n} x_i^2}{n}}$$

Example 12

Find the root mean square of the heights in ft. of 5 students in a class-3 2, 4, 6, 7

Solution

$$\text{R.M.S} = \sqrt{\frac{3^2+2^2+4^2+6^2+3^2}{5}}$$

$$= \sqrt{22.8} = 4.775$$

3.2 Median

This is the item that divides the set of data into two equal halves. It is the middle item. However, the data must be arranged in either ascending or descending order of magnitude. If a set of data has odd number of observation then, the median item is $\frac{(N+1)^{th}}{2}$ item where n is the number of observation. But for even number of items, the median is the average of the two middle numbers after ordering.

Example 13

The heights of girls in a class are 3, 2, 4, 1, 5, 3, and 6
Since this is an odd number of items i.e. 7

$\therefore$ The medianth item = $\frac{(N+1)^{th}}{2}$ item

$= \frac{7+1}{2} = 4$

$\therefore$ 1, 2, 3, 3, 4, 5, 6

Then, the median = 3

Example 14

Find the median of the following data

i.e. 8, 3, 5, 2, 6, 7, 3, 9

Solution

Ordering the data gives:
2, 3, 3, 5, 6, 7, 8, 9

This is an even number of items

Note: N = 8
$\therefore \quad \frac{N}{2} = \frac{8}{2} = 4$

Then, count 4 from the first and from the last items

2, 3, 3, 5, 6, 7, 8, 9
$\therefore$ Median = $\frac{5+6}{2}$

$= \frac{11}{2} = 5.5$

Median for Grouped Discrete Data

If frequency table is provided for a discrete variable where the table has frequencies for individual values, then compute the cumulative frequency. If the total number of values, n is odd, the median value is the value for which the cumulative frequency first exceeds or equals:

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

But if n is even, the median is the value for which cumulative frequency first exceed N/2. If the cumulative frequency equal N/2 for a value, this means that the median is half-way between this value and the next value.

Example 15

The scores in a test by 55 students are given in the table below:

No of students	0	1	2	3	4	5	6
Score	3	8	20	14	6	3	1

Find the median.

Solution

X	Frequency	Cumulative frequency
0	3	3
1	8	11
2	20	31
3	14	45
4	6	51
5	3	54
6	1	55

Since N is odd i.e. N = 55

$\therefore \quad \frac{N+1}{2} = \frac{55+1}{2} = 28$

Then, the value of $\frac{N+1}{2}$ is first exceeded by the figure for 2 marks

Then, the median is 2 marks

Example 16

The numbers of items sold in a supermarket over a period of 40 days are shown in the table below:

Items Sold	0	1	2	3	4	5
No of Days	1	3	16	11	5	4

Find the median of number of items sold.

X	Frequency	Cumulative frequency
0	1	1
1	3	4
2	16	20
3	11	31
4	5	35
5	4	40

Here N = 40 is even

$\therefore \frac{N}{2} = \frac{40}{2} = 20$

Then, the value of N/2 is half way between 2 and 3

$\therefore$ Median diameter is $\frac{2+3}{2} = \frac{5}{2} = 2.5$

Median for Grouped, Discrete Data, with Class Interval

This can be done using formula method i.e. interpolation and also by graphical method. But these methods may waste your precious time. The formula below is applied after identifying the median class that contains the middle observation.

$$\text{Median} = L_1 + \left[\frac{(N/2 - (\sum f)}{f_{median}}\right] xc$$

Where, L_1 = lower class boundary of the median class

N = number of items.

$(\sum f)_1$ = sum of all frequencies of the class/classes lower than the median class.

f_{median} = frequency of the median class

c = size of the median class.

Example 17

The table below shows the weekly profit in naira from a minimart

Weekly profit (N)	1-10	11-20	21-30	31-40	41-50	51-60
Frequency (f)	6	6	12	11	10	5

Find the median weekly profit.

Solution

Here, N = 50 i.e. even

$\therefore$ N/2 = 50/2 = 25

The class 31-40 first exceeded N/2 in the cumulative frequency column
$\therefore$ The median class is 31-40

Weekly profit	f	Class Boundaries	Cumulative Frequency
1-10	6	0.5-10.5	6
11-20	6	10.5-20.5	12
21-30	12	20.5-30.5	24
31-40	11	30.5-40.5	35
41-50	10	40.5-50.5	45
51-60	5	50.5-60.5	50
	50		

$L_1 = 30.5$, N = 50, $(\sum f)_1 = 6+6+12 = 24$, $f_{median} = 11$ c = 10

$$\therefore \text{Median} = L_1 + \left[\frac{(N/2 - (\sum f)_1}{f_{median}}\right] xc$$

$$= 30.5 + \frac{\left(N/2 - 24\right)}{11} x10$$

$$= 30.5 + \frac{10}{11} = 31.41$$

The Median weekly profit is N 31.41

Example 18

The annual salaries (N'000) of the employees of a company are summarised in the table below:

Salary	5≤x<10	10≤x<15	15≤x<20	20≤x<25	25≤x<30	30≤x<35
Freq.	22	30	41	15	17	15

Calculate the median salary

Using formula method

Solution

Here, the boundaries have been set, then, there is no need for constructing a new set of class boundaries.

Salary	Freq	Cumulative
5-10	22	22
10-15	30	52
15-20	41	93
20-25	15	108
25-30	17	125
30-35	15	140
	140	

N = 140
∴ N is even then, $^{N}/2 = {}^{140}/2 = 70$

Therefore the interval 15-20 exceeded N/2, which means that 15-20 is the median class.

$$N = 140,\ L_1 = 15,\ (\sum f)_1 = 22 + 30 = 52,\ f_{median} = 41,\ c = 5$$

$$\therefore \text{Median} = L_1 + \frac{\left({}^{140}/_{2} - 52\right)}{41} x5$$

$$= 17.20$$

The median salary is ₦17, 200

Advantages of Median

(i) It is easy to determine
(ii) It can be obtained graphically
(iii) It gives a clear idea of the distribution of the data
(iv) It can be calculated in some cases where mean and mode cannot
(v) It is a measure of rank or position.
(vi) It is unaffected by the presence of extreme values in the distribution

Disadvantages of Median

(i) The statistical properties are difficult to determine
(ii) It is not useful for further statistical calculations
(iii) It may require rearrangement of the data involved. This may be difficult if large population is required.
(iv) It may not be representative if there are few data.

3.3 Mode

This is the item that occurs most in a set of data or list of items. It is the item or value that has the highest frequency.

Example 19

Find the modal item in 2, 2, 3, 2, 3, 4, and 2

Solution

The item that occurs most is 2 and hence the mode = 2, bimodal values occur when two or more figures have the highest frequency. Then, the mode is obtained by taking the arithmetical mean of the two values.

Example 20

Find the mode of the data below 2, 3, 3, 4, 2, 6, 3, 3, 4, 4, and 4

Solution

The numbers 3 and 4 have the same frequency and this frequency is the highest

Then, Mode = $\frac{N}{2} = \frac{3+4}{2} = 3.5$

If a frequency table for a discrete variable is provided and that the table has frequencies for individual values, then the mode is the number that has the highest frequency on the table. But if two or more numbers have this highest frequency, the mode is obtained by taking the arithmetic mean of the values.

Example 21

The table below shows the number of accident on a road per day and this is observed for five consecutive days. Find the mode.

Days	1	2	3	4	5
Freq	10	11	18	3	2

Solution

The mode is simply 3 because it has highest frequency.

Mode for Grouped, Discrete Data with Equal Class Intervals

To determine a value for the mode here, we must at first determine which interval has the highest frequency. This is called the modal interval or modal class. Then the formula below is used.

$$\text{Mode} = L1 + \left[\frac{\Delta_1}{\Delta_1 + \Delta_2}\right] xc$$

Where, L_1 is the lower class boundary of the modal class
Δ_1 is the excess of modal frequency over the frequency of immediate lower class i.e. $\Delta_1 = f_m - f_{m-1}$
Δ_1 is the excess of modal frequency over the frequency of immediate Upper class i.e. $\Delta_2 = f_m - f_{m+1}$ c is the class width or size.

Example 22

The frequency distributions of storm duration (in minutes) are based on data appearing in the article "Lighting phenomenology in the Tampa Bay Area". Use the data to estimate the mode.

Storm duration	0-25	25-50	50-75	75-100	100-125
Single Peak(f)	3	6	17	15	9

Solution

Here, the class boundaries have been set. There is no need of constructing the class boundaries

The highest occurrence of storm duration is 17
Then, the modal class = 50-75

$L_1 = 50$, $\Delta_1 = 17 - 6 = 11$, $\Delta_2 = 17 - 5 = 12$, $c = 25$

$$\therefore \text{mode} = L_1 + \left[\frac{\Delta_1}{\Delta_1 + \Delta_2}\right] xc$$

$$= 50 + \left[\frac{11}{11+12}\right] x25$$

$$= 61.96$$

Example 23

The table shows the wage of the 50 employees in Adebowale Nig. Ltd, and the number in respective wage group.

Weekly Wage (N'000)	5-9	10-14	15-19	20-24	25-29	30-34
Number of Employee	6	11	15	8	6	4

Estimate the modal weekly wage using;
Formula Method
Graphical Method

Solution

Weekly Wage '000N	Number of employee (f)	Mid WW(x)	Class Boundaries
5-9	6	7	4.5-9.5
10-14	11	12	9.5-14.5
15-19	15	17	14.5-19.5
20-24	8	22	19.5-24.5
25-29	6	27	24.5-29.5
30-34	4	32	29.5-34.5
	50		

The modal class is 15-19

$$\therefore \quad \text{mode} = L_1 + \left[\frac{\Delta_1}{\Delta_1 + \Delta_2}\right] xc$$

$$= 14.5 + \left[\frac{15-11}{15-11+15-8}\right] x5$$

$$= 14.5 + \left[\frac{4}{11}\right] x5$$

$$= 16.32$$

Advantages of Mode

i. It can be obtained graphically
ii. It is not affected by extreme values
iii It is easy to understand and compute
iv. If a mode exists, it is certain to have a value that was actually observed
v. The data does not need any ordering before mode can be obtained.

Disadvantages of Mode

i. It can be a very poor average if the data is not symmetrical
ii. It is not an ideal measure of central tendencies
iii. It is not useful in further statistical processing
iv. It does not make use of all the values in the distribution
v. Its statistical properties are difficult to determine
vi. It may not be unique because two or more values may be equally frequent.

Example 24

In Wahum Nig. Ltd. dry cells are usually subjected to test before been transferred to sales department. The life spans (in hour) of 50 dry cells subjected to this test are shown in the table below:

38.2	46.7	46.3	35.1	30.5	41.3	49.2	37.9
36.9	41.3	30.2	53.9	42.8	53.2	53.9	45.5
45.7	47.5	42.4	36.6	40.6	54.4	42.3	43.8
48.2	49.9	46.3	38.9	46.9	41.6	33.8	45.2
41.3	43.5	52.4	38.2	41.6	36.3	35.6	50.3
38.3	51.2	46.9	38.4	32.6	32.9	37.4	54.0
39.4	37.5						

a. By using tally method construct a frequency distribution table having class interval of 30 but less than 35, 35 but less than 40 and so on for the data.

b. Estimate

(i) Arithmetic mean
(ii) Geometric mean
(iii) Harmonic mean

c.(i) Using formula method: determine the median and modal life spans.
(ii) Draw an ogive and histogram to verify your results.

Solution

a.

Class interval	Tally	Freq (f)	x	fx	Cum. Freq.
30-35	𝍸	5	32.5	162.5	5
35-40	𝍸𝍸 ||||	14	37.5	525	19
40-45	𝍸𝍸 |	11	42.5	467.5	30
45-50	𝍸𝍸 ||	12	47.5	570	42
50-55	𝍸 |||	8	52.5	420	50
		50		2145	

b. (i) Arithmetic Mean = $\frac{\sum fx}{\sum f}$

$= \frac{2145}{50} = 42.9$

(ii) Geometric Mean = $\sqrt[n]{32.5^5 x 37.5^{14}}\, 42.5^{11} x 47.5^{12} x 52.5^8$

$= \sqrt[5]{2.453186276 x 10^{81}}$

$= 42.44$

(iii) Harmonic Mean = $\frac{n}{\sum_{i=1}^{n}\left(\frac{f_i}{x_i}\right)}$ where; n $\sum f$

$= \frac{50}{\frac{5}{32.5}+\frac{14}{37.5}+\frac{11}{42.5}+\frac{12}{47.5}+\frac{8}{52.5}}$

$= \frac{50}{0.139+0.3733+0.2588+0.2588+0.2526+0.1524}$

$= \frac{50}{1.191} = 41.98$

c. (i) Median = $L_1 + \left[\frac{\left(N/2 - (\sum f)_1\right)}{f_{median}}\right] xc$

Here, the median is 40-45 i.e. it first exceeded

$^N/2 = {}^{50}/2 = 25$

$\therefore L_1 = 40, (\sum f)_1 = 14+5 = 19, c = 5$

$F_m = 11$

$\therefore$ Median = $40 + \frac{\left(50/2 - 19\right)}{11} x5$

$= 42.73$

Mode = $L_1 + \left[\frac{\Delta_1}{\Delta_1 + \Delta_2}\right] xc$

Here, the modal class is 35-40 i.e. it has the highest frequency

$\therefore L_1 = 35,\ \Delta_1 = 14-5 = 9,\ \Delta_2 = 14-11 = 3$

$C = 5$

$$\text{Mode} = 35 + \left[\frac{9}{9-3}\right]x5$$

$= 38.75$

SELF ASSESSMENT EXERCISE

Find the mean, median and mode of the following data
2, 2, 3, 2, 3, 4, 2 .

4.0 CONCLUSION

Measures of central tendency such as mean, mode and median were discussed with quite a number of illustrations for better understanding. Formulas and computation methods for calculation of the mean and others were developed and demonstrated for grouped and ungrouped data.

5.0 SUMMARY

In this unit we have learnt:

- how to calculate mean, median and mode for grouped and ungrouped data
- the advantages and disadvantages of mean, median and mode
- the relationships that exist among Harmonic, Arithmetic and Geometric mean. i.e. H.M.<G.M.<A.M.

6.0 TUTOR-MARKED ASSIGNMENT

1. The following figures were extracted from personal records of workers in a state civil service in Nigeria. The table shows the number of employees and the number of times they absented themselves from duty in a particular year.

No of days absent	1-5	6-10	11-15	16-20	21-25	26-30
No of employees	620	874	150	204	101	80

Using the above data estimate the

a. Arithmetic mean (using assumed mean method)
b. Geometric mean
c. Harmonic mean
d. Median of the distribution
e. Mode of the distribution

2. The following are the percentage yields from 40 runs of a chemical process, taken from J.S. Hunter's article "The Technology of Quality" (RCA Engineer, May/ June 1985)

65.6	67.5	68.2	73.5	72.7	69.0	70.8	69.9
65.6	67.8	68.3	70.6	70.4	72.0	70.9	71.4
66.2	67.8	74.2	69.3	69.1	70.0	69.5	68.3
66.8	68.0	70.7	72.8	72.6	68.9	69.8	69.7
67.2	68.0	69.5	70.6	71.2	71.7	71.3	69.7

a(i) Construct a frequency distribution from the data using approximate class interval.
(ii) Using your data, estimate Arithmetical mean,
Harmonic mean and Geometric mean. Hence, investigate H.M. $\leq$ G.M. $\leq$ A.M.
b. Draw a cumulative frequency curve from the data.

7.0 REFERENCES/FURTHER READING

Donal, R.P. & Edward B. O. *Business and Economics Statistics*, (Revised Edition). Business Publication Inc.

McClave, & Sincich, *Statistics*, (Eighth Ed). Prentice Hall.

Murray R. Spiegel: *Theory and Problems of Statistic.* Schaum's Outline Series, McGraw-Hill.

UNIT 3 DESCRIPTIVE STATISTICS: MEASURE OF DISPERSION

CONTENTS

1.0 INTRODUCTION

After obtaining a measure of location where, on the average, the distribution lies, such as the mean or median, we next begin to wonder about the variability of the distribution and look for some numbers that can be used to measure how spread out the data are. In many instances, the dispersion of the data is of at least as much interest as is the mean.

If measures of location are used to compare two sets of each data, there may be difference in the extent to which their values spread around the central or mean value. Measures of location cannot distinguish between the sets in such cases. Hence, measures of dispersion are applied.

2.0 OBJECTIVES

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

- explain what measures of dispersion are
- calculate, based on measures of dispersion and using various methods on both grouped and ungrouped data
- list the advantages and disadvantages of the measure of dispersion.

3.0 MAIN CONTENT

3.1 Definition of Dispersion

Dispersion is the degree to which a numerical data tend to spread about an average value. Measures of dispersion are referred to as the spread of a set of data. When mean or average is in focus from other variables, it is called measures of dispersion.

3.2 Range

The range of a set of numerical data is the difference between the highest and the lowest values. It is the 'spread' or 'width of the data.

Advantages of Range

a. It is easy to calculate
b. It is the simplest possible measure of spread
c. It may be useful for further statistical calculations.

Disadvantages of Range

a. It depends on extreme values. So it is susceptible to odd results
b. It only uses two values-the remaining data is ignored
c. It cannot be used to calculate open-ended distribution
d. It cannot be used with grouped data, since it ignores the intermediate values.

3.3 Interquartile Range

This is the difference between the upper quartile and lower quartile i.e.

$$Q_1=Q_3-Q_1$$

3.3.1 Semi-Interquartile Range or Quartile Deviation

This is the half of the interquartile range. This is otherwise known as quartile deviation.

$$Qs = ½ [Q_3-Q_1]$$

Example 1

The marks scored by 7 students in a test are:

3, 7, 2, 10, 2 and 6.

Find the range?

Solution

Range = 10-2=8

Computing Quartile for Grouped Data

$$Qi = Li\_\frac{\left[\frac{N}{4}i - \sum f_i\right]c}{f_q} \quad \text{where I = 1, 2,3,4.}$$

L_i = Lower class boundary of the quartile class
N = Number of observations
$\sum f_i$ = Cumulative frequency just before the quartile class.
c = Size of the quartile class
f_q = Frequency of the quartile class

Advantages of Quartile Deviation

a. It is easy to understand
b. It is unaffected by the presence of extreme values in the distribution
c. It is half the range of the middle 50% of values, so it tells us about the spread of the middle values.

Disadvantages of Quartile Deviation

a. It does not make use of all the items in the data for its computations
b. It is not suitable for further statistical analysis
c. It is rather awkward to calculate
d. It may not be representative if there are few data.

Example 2

The number of new firms commencing trade per month in a metropolitan district was recorded each month in a year as: 4, 2, 5, 2, 3, 6, 8, 3, 2, 4, 6

Determine the following (i) Range (ii) Interquartile Range
(iii) Semi-Interquartile Range

Solution

Re-order the data:

2, 2, 2, 3, 3, 4, 4, 5, 5, 6, 6, 8

Range = 8-2 =6
Interquartile Range = Q_3-Q_1
Q_3 is the number in $\frac{3}{4}x12 = 9 \Rightarrow Q_3 = 5$

Q_1 is the number in $\frac{1}{4}x12 = 3 \Rightarrow Q_1 = 2$
$\therefore Q_1 = Q_3 - Q_1$
= 5-2
= 3

Semi-Interquartile Range = ½ (Q_3-Q_1)
= ½ (5-2)
= 1.5

Example 3

The numbers of applicants for different posts advertised in a company are recorded in the table below:

Applicants	6-10	11-15	16-20	21-25	26-30
Frequency	10	13	42	24	11

Determine (i) Range (ii) Semi-interquartile Range

Solution

(i) Range = Upper class boundary of the highest-
Lower class boundary of the lowest class
= 30.5 -5.5
= 25

Applicants	Freq.	Cum. Freq.	Class Boundaries
6-10	10	10	5.5-10.5
11-15	13	23	10.5-15.5
16-20	42	65	15.5-15.5
21-25	24	89	20.5-25.5
26-30	11	100	25.5-30.5

The Q_1 position is ¼ of 100 = 25
Then, the 1st quartile class is 16-20

$$\therefore\ Q1 = L1 + \frac{\left(N/4 - \sum f_1\right)}{f_q}c \quad = 15.5 + \frac{\left(100/4 - 23\right)}{42}5$$

$$= 15.5 + \frac{\left(25 - 23\right)}{42}5 \quad = 15.74$$

Also Q_3 position is ¾ of 100 = 75

Then, the 3rd quartile class is (21-25)

$$\therefore\ Q_3 = L_3 + \frac{\left(3/4\,N - \sum f_3\right)}{f_3}c_3 \quad = 20.5 + \frac{\left(3/4\,X100 - 65\right)}{24}5 \ = 20.5 + 2.08$$

$\therefore\ Q_3 = 22.58$

Then, the Semi-interquartile range ½ $[Q_3 - Q_1]$
= ½ [22.58-15.74]
=3.42

Example 4

The lengths, to the nearest millimeter, of 100 screws were measured and recorded in the table below:

Length (mm)	25-29	30-34	35-39	40-44	45-49	50-54
Number of Screws	5	12	34	20	15	14

You are required to use your data to compute:

- Range
- Median
- Interquartile Range and hence, Quartile Deviation
- Prepare a cumulative frequency table for the data and hence, draw an ogive curve for the distribution.
- Using your curve, estimate median and quartile deviation.
- Compare your results with values obtained above.

Solution

Applicants	Freq.	Mid. Length x	Cum. Freq.	Class Boundaries
25-29	5	27	5	24.5-29.5
30-34	12	32	17	29.5-34.5
35-39	34	37	51	34.5-39.5
40-44	20	42	71	39.5-44.5
45-49	15	47	86	44.5-49.5
50-54	14	52	100	49.5-54.5
TOTAL	**100**			

(i) Range = Upper class boundary of the highest class- Lower class boundary of the least class
= 54.5 – 24.5 = 30

(ii) Median = $L1+\frac{\left(\frac{1}{2}N-\sum f_1\right)}{f_{median}}c$ = $34.5+\frac{\left(\frac{100}{2}-17\right)}{34}5$

= 39.35

(iii) $\frac{N}{4}=\frac{100}{4}=25$

Then, the 1st Quartile class is (30-34)

$\therefore Q_1 = L_1 + \frac{\left(\frac{1}{4}N-\sum f_1\right)}{f_q}c$ = $29.5+\frac{\left(\frac{100}{4}-5\right)}{12}5$

= 37.83

$\frac{3N}{4}=\frac{3x100}{4}=75$

Then, the 3rd Quartile class is (45-49)

$Q_3 = L_3+\frac{\left(\frac{3}{4}N-\sum f_3\right)}{f_3}c$ = $44.5+\frac{(\frac{3\times100}{4}-71)}{15}5$

= 45.83

Interquartile Range = $Q_3=Q_1$
= 45.83-37.83
=8
Quartile Deviation = ½ (8)
= 4

3.4 Mean Deviation (MD)

It measures the average spread of values from the arithmetic mean. Mean deviation is referred to as the average amount by which a value in a distribution differs from the arithmetic mean. It is otherwise known as mean absolute deviation.

i. For Ungrouped Data

$$\text{M.D.} = \frac{\sum / x - \bar{x} /}{n}$$

ii. For Grouped Data

$$\text{M.D} = \frac{\sum f \mid x - \bar{x} \mid}{\sum f}$$

Features of Mean Deviation (MD)

a. It is easy to understand
b. It is straightforward to calculate
c. It cannot be distorted by extreme values
d. It makes use of every value in the distribution
e. It cannot be used for further statistical processing.

Example 5

Find the mean-deviation of 8, 7, 11, 6, 16, 4, and 11

Solution:

$$\bar{x} = \frac{8+7+11+6+16+4+11}{7} = \frac{63}{7} = 9$$

$$= {}^{22}\!/_{2} = 11$$

Example 6

Calculate the mean-deviation of the distribution below:

Class	10-15	15-20	20-25	25-30
Frequency	3	2	4	1

Solution

Class	Freq.(f)	X	Fx	$\|x-\bar{x}\|$	$f\|x-\bar{x}\|$
10-15	3	12.5	37.5	6.5	19.5
15-20	2	17.5	35.0	1.5	3.0
20-25	4	22.5	90.0	3.5	14.0
25-30	1	27.5	27.5	8.5	8.5
TOTAL	**10**	**TOTAL**	**190.0**	**TOTAL**	**45**

Mean, $\bar{x} = \dfrac{\sum fx}{\sum f}$

$= \dfrac{190}{10} = 19$

Mean Deviation, M.D. = $\dfrac{\sum f\,|x-\bar{x}|}{\sum f} = \dfrac{45}{10} = 4.5$

3.5 Variance

Variance makes use of the sum of the squares of deviations from the mean. The more variation there is in the x-values, the larger will be the value of $\dfrac{\sum\left(x-\bar{x}\right)^2}{n}$

Or Variance = $\dfrac{\sum\left(x-\bar{x}\right)^2}{n-1}$

Note

i. Using the divisor n

It is appropriate in two cases:

a. When the values $x_1, x_2 \ldots x_n$ represent the whole population
b. When the values $x_1, x_2 \ldots x_n$ represent a sample from a population and we are concerned about the variation within the sample itself.

Then, the variance is denoted by:

$$\sigma_n^2 = \frac{1}{n}\left[\sum_{i=1}^{n}\left(x-\bar{x}\right)^2\right]$$

If $x_1, x_2, \ldots, x_n$ represent a sample of data then, σ_n^2 is called the sample variance.

If $x_1, x_2, \ldots, x_n$ represent the entire population then, σ_n^2 is called the population variance.

ii. Using the division n-1

This is used when $x_1, x_2 \ldots x_n$ represent a sample from a given population and we are concerned about the variation in the population. The sample variance is denoted by $S^2 or \sigma_{n-1}^2$

$$S^2 = \frac{N_{1S2_1} + N_{2S_{N_2}^2}}{N_1 + N}$$

3.6 Standard Deviation (SD)

This is the positive square root of the mean of the squared deviations from the arithmetic mean of the distribution. It is also the square root of the variance. Standard deviation is used widely in statistics as a measure of spread.

3.6.1 Features of Standard Deviation

i. It makes use of all the values in the distribution
ii. It is the best measure of dispersion
iii. It corrects the mathematical inaccuracy of the mean deviation
iv. It is the most prominent measure of dispersion
v. It can be used to determine the sample size to estimate the standard error of the sample mean in significance test.
vi. It is difficult to compute when a grouped distribution has unequal interval or open ended class.
vii. It is the basis for further statistical development
viii. It can be distorted by extreme values
ix. It is almost impossible to visualise what the standard deviation figure represents.

3.6.2 Computing Standard Deviation

(i) For Ungrouped Data

$$\sigma_n \sqrt{\frac{\sum \left(x - \bar{x}\right)^2}{n}}$$

(ii) For Grouped Data

$$\sigma_n = \sqrt{\frac{\sum f\left(x-\bar{x}\right)^2}{n}}$$

$$\sigma_n = \sqrt{\frac{\sum fx^2}{\sum f} - \frac{\left(\sum fx\right)^2}{\sum f}}$$

$$\sigma_n = \sqrt{\frac{\sum fx^2}{\sum f} - \bar{x}^2}$$

Example 7

The numbers of patients who visit a clinic on 5 work-days are as follows: 7, 6, 3, 4, and 10.

Compute the mean deviation and hence, standard deviation for the data.

Solution

Mean $\bar{x} = \frac{7+6+3+4+10}{5}$

$= \frac{30}{5} = 6$

X	Fx	$\lvert x-\bar{x}\rvert$	$f\lvert x-\bar{x}\rvert$
3	-3	3	9
4	-2	2	4
6	0	0	0
7	1	1	1
10	4	4	16
TOTAL		10	30

Mean Deviation = $\frac{\sum \lvert x-\bar{x}\rvert}{n}$

$= \frac{10}{5} = 2$

Standard Deviation = $\sqrt{\frac{\sum\left(x-\bar{x}\right)^2}{\sum f}}$

$= \sqrt{\frac{30}{5}}$

$= \sqrt{6} = 2.45$

This shows that the mean deviation tends to give an impression of smaller dispersion as compared to the standard deviation.

Example 8

The weights to the nearest kg. of a group of 50 students in Yaba College of Technology are given below:

65	70	60	46	51	55	59	63	68	53
47	53	72	63	67	62	64	70	57	56
73	56	48	51	58	63	65	62	49	64
53	59	63	50	48	72	67	56	61	64
56	52	49	62	71	58	53	69	63	59

(a) (i) Prepare a grouped frequency table with class interval 45-49, 50-54, 55-59, etc.

(ii) Calculate the standard deviation of the distribution

Solution

Class interval	Tally	Freq (f)	Mid Class (x)	Fx	$x-\bar{x}$	$(x-\bar{x})^2$	$f(x-\bar{x})^2$
45-49	𝍸 |	6	47	282	-12.1	146.41	878.46
50-54	𝍸 ||||	9	52	468	-7.1	50.41	453.69
55-59	𝍸𝍸 |	11	57	627	-2.1	4.41	48.51
60-64	𝍸𝍸 ||	12	62	744	2.9	8.41	100.92
65-69	𝍸 |	6	67	402	7.9	62.41	374.46
70-74	𝍸 |	6	72	432	12.9	166.41	998.46
TOTAL		**50**	**TOTAL**	**2955**	**TOTAL**		**2854.5**

Mean = $\bar{x} = \dfrac{\sum fx}{\sum f}$

$= \dfrac{2955}{50} = 59.1$

Standard Deviation = $\sqrt{\dfrac{\sum f\left(x-\bar{x}\right)^2}{\sum f}}$

$= \sqrt{\dfrac{2854.5}{50}} = 7.56$

Example 9

The table below shows percentage yields from 40 runs of a chemical process, taken from J.S. Hunter's article 'The technology of Quality' (RCA Engineer, May/June 1985).

Yields	50-54	55-59	60-64	65-69	70-74
Frequency	6	10	15	4	5

Estimate the standard deviation using a short-cut method.

Solution

Yield	F	X	d=$x-\bar{x}$	d^2	fd	fd^2
50-54	6	52	-10	100	-60	600
55-59	10	57	-5	25	-50	250
60-64	15	62	0	0	0	0
65-69	4	67	5	25	20	100
70-74	5	72	10	100	50	500
TOTAL	**40**	**TOTAL**		**250**	**-40**	**1450**

Let the Assumed Mean, $\bar{x}$ = 62

$$\bar{d} = \frac{\sum fd}{\sum f}$$

$$= \frac{-40}{40} = -1$$

$$\sigma = \sqrt{\frac{\sum fd^2}{\sum f} - \overline{d^2}}$$

$$= \sqrt{\frac{1450}{40} - (-1)^2} = \sqrt{35.25} = 5.94$$

Example 10

The table below shows the distribution of lengths of off-outs of timber over a period.

Length (xm)	0.2≤x<0.3	0.3≤x<0.4	0.4≤x<0.5	0.5≤x<0.6	0.6≤x<0.7
Frequency	6	18	10	12	2

Use the coding y = $\frac{x-0.45}{0.1}$ to facilitate the calculation of the standard deviation.

Solution

Length (xm)	Freq (f)	X	Y	Fy	y^2	fy^2
0.2≤x<0.3	6	0.25	-2	-12	4	24
0.3≤x<0.4	18	0.35	-1	-18	1	18
0.4≤x<0.5	10	0.45	0	0	0	0
0.5≤x<0.6	12	0.55	1	12	1	12
0.6≤x<0.7	2	0.65	2	4	4	6

Standard Deviation = $\sqrt{\frac{\sum fy^2}{\sum f} - \left(\frac{\sum fy}{\sum f}\right)^2}$

$= \sqrt{\frac{60}{48} - \left(\frac{-14}{48}\right)^2}$

$= \sqrt{1.25 - 0.0851}$

$= \sqrt{1.1649} = 1.08$

3.7 Comparison of Measures of Dispersion

The most commonly used measure of dispersion is the standard deviation and this is simply due to its further applications in statistical process. However, all the measures increase in size as the data values become more spread. The range is sometimes used as a ‘quick’ method for finding spread but it is only applicable to a few data values. The least used is the mean deviation. Finally, whenever mean is used as a measure of location, standard deviation is usually quoted.

3.8 Decile and Percentile

3.8.1 Decile

This is where the number of observations is divided into ten parts after ordering i.e. arrangement in ascending or descending order of magnitude. It is denoted by D_1, D_2, D_3,…, D_9.

3.8.2 Percentile

These are the numbers dividing the data into 100 parts after ordering. They are represented by P_1, P_2, P_3, … P_{99}. The values of 50^{th} percentile,

25^{th} percentile and 75^{th} percentile represent the median, lower and upper percentile or quartiles respectively.

The two can be represented mathematically as:

$$\text{Decile} = L_1 + \frac{\left[iN/10 - \sum f_1\right]}{f_d} xc$$

$$\text{Percentile} = L_1 + \frac{\left[iN/100 - \sum f_1\right]}{f_p} xc$$

Where; L_1 = the lower class boundary of either the decile or percentile class as the case may be

N = Number of items in the data

$\sum f_1$ = Cumulative frequency just before the decile or percentile class

f_d = Frequency of the decile class

f_p =Frequency of the percentile class

Example 11

The data below shows the heights of maize in (m) planted over a given period of time: 2,11,19,27,6,9,15,13. Using the data provided, estimate the 3^{rd}, 5^{th} and 8^{th} deciles and the 25^{th}, 50^{th} and 75^{th} percentiles.

Solution

(i) D_3 has the position = $\frac{3}{10}(n+1)^{th}$

$= \frac{3}{10}(9+1)^{th}$

$= 3^{rd}$ term

$\therefore D_3 = 9$

(ii) D_5 has the position = $\frac{5}{10}(9+1)^{th}$

$= 5^{th}$ term

(iii) D_8 has the position = $\frac{8}{10}(9+1)^{th}$

$= 8^{th}$ term

$D_8 = 20$

2, 6, (9,) 11, (13,) 15, 19, 20, 27

D_3 D_5

(i) P_{25} has the position = $\frac{25}{100}(n+1)^{th}$

$= \frac{25}{100}(9+1)^{th}$

$= 2.5^{th}$ term

$\therefore P_{25} = \frac{6+9}{2} = \frac{15}{2} = 7.5$

P_{75} has the position = $\frac{75}{100}(n+1)^{th}$

$= \frac{75}{100}(9+1)^{th}$

$\therefore P_{25} = \frac{19+20}{2} = \frac{15}{2} = 19.5$

$\therefore$ 2, [6, 9,] 11, 13, 15, [19, 20,] 27

$P_{25} = \frac{6+9}{2} = 7.5$ $P_{75} = \frac{19+20}{2} = 19.5$

Example 12

The data **represents** the **number** in '000 of households, in Nigeria receiving local authority help or home care services.

Class interval (age)	11-15	16-20	21-25	26-30	31-35	36-40	41-45
Freq. (f)	4	15	12	11	15	14	9

Using your data, determine

a. 7^{th} decile
b. 60th percentile

Solution

Prepare a class boundary and cumulative frequency table for the distribution.

	Class Boundaries	F	Cum freq.
11-15	10.5-15.5	4	4
16-20	15.5-20.5	15	19
21-25	20.5-25.5	12	31
26-30	25.5-30.5	11	42

31-35	30.5-35.5	15	57
36-40	35.5-40.5	14	71
41-45	40.5-45.5	9	80

Position of $D_7 = \left(\frac{7}{10}\right)\sum f$

$= \frac{7}{10}$ x 80

= 56

Then, the class interval for D_7 is (31-35)

$L_1 = 30.5$, N = 80, $\sum f_1 = 42$, $f_{d=} 15$

$\therefore$ Decile $= L_1 + \frac{\left[iN/10 - \sum f_1\right]}{f_d} xc$

$= 30.5 + \frac{\left[7x\frac{80}{10} - 42\right]}{15} x5$

= 35.17

Position of the 60th percentile = $\frac{(60)}{100}\sum f$

$= \frac{60}{100} x \frac{80}{1}$

= 48

The, the class interval for P_{60} is (31-35)

$\therefore$ $L_1 = 30.5$, N = 80, $\sum f_1 = 42$, $f_p = 15$

$\therefore$ Percentile $= L_1 + \frac{\left[iN/100 - \sum f_1\right]}{f_p} xc$

$= 30.5 + \frac{\left[60x\frac{80}{100} - 42\right]}{15} x5$

$= 30.5 + \frac{[48 - 42]}{15} x5$

= 32.5

3.9 Coefficient of Variation (COV)

This is a relative measure of dispersion and therefore otherwise called coefficient of relative dispersion. It is basically used for comparison i.e. enables us to compare the dispersion of two distributions in widely different units e.g. yearly output of two factories producing different commodities; change in temperature over a period of time in two countries; imported goods into two different countries etc.

The standard deviation alone does not measure relative variation. For example, a standard deviation of $ 1 would be considered large if it is describing the variability from store to store in the price of an ice cube tray. On the other hand, a standard deviation of $1 would be considered small if it is describing store-to-store variability in the price of a particular brand of freezer.

COV can be expressed mathematically as:

$$\text{COV} = \frac{\text{Standard Deviation}}{\text{Mean}} \times \frac{100}{1}$$
$$= \frac{\sigma}{x} \times \frac{100}{1}$$

Example 13

A standard deviation of ₦10 per day on a mean daily wage of ₦300 will be insignificant. However, a standard deviation of ₦10 on a mean of ₦ 40 daily wage is a very serious variation.

Solution

For the two situations:

A: $\bar{x} = 300$ $\sigma = ₦10$

$$\text{COV} = \frac{\sigma}{x} \times \frac{100}{1} = \frac{10}{300} \times \frac{100}{1}$$
$$= 3.33\%$$

B: $\bar{x} = 40$ $\sigma = ₦10$

$$\therefore \text{COV} = \frac{\sigma}{x} \times \frac{100}{1} = \frac{10}{40} \times \frac{100}{1}$$
$$= 25\%$$

This shows that the variability of data B is higher than that of A. This implies that the wage differences in data B are wider than that of A.

SELF ASSESSMENT EXERCISE

Consider the following data 8, 7, 11, 6, 16, 4, 11

Find (i) range (ii) mean-deviation (iii) variance and standard deviation

4.0 CONCLUSION

Measures of dispersion such as range, standard deviation, and variance were discussed with quite a number of illustrations. Formulas and computation methods for calculation of the range and others were developed and demonstrated for grouped and ungrouped data. Formulas were developed for a population and a sample.

5.0 SUMMARY

In this unit, we have learnt that the following:

- Measures of dispersion is concerned with the spread of data
- Coefficient of variation is a relative measure of dispersion
- Mean deviation measures the average spread of values from the arithmetic mean
- Decile and percentiles are numbers dividing the data into 10 and 100 parts after ordering respectively.

6.0 TUTOR-MARKED ASSIGNMENT

1. The DTXL Production Company has produced the following data showing weekly productions for a period of 110 weeks.

Production output '000	$40 \leq x \leq 50$	$50 \leq x \leq 60$	$60 \leq x \leq 70$	$70 \leq x \leq 80$	$80 \leq x \leq 90$
Number of week	20	25	38	17	10

Use the data to estimate

(a) Range
(b) Interquartile Range
(c) Mean Deviation
(d) Standard Deviation

2. 50 electric bulbs were subjected to life span test by a manufacturing company. The life spans of the bulbs are as follows:

30	42	55	59	54	36	39	53	45	33
52	41	43	32	46	52	36	36	43	33
37	32	41	57	48	45	37	44	56	51
51	37	30	38	43	43	38	38	47	35
37	48	41	32	43	52	44	58	39	41

(a) By using tally method, prepare a frequency distribution table for the data using interval 30-34, 35-39.

(b) Find the mean, mean deviation, standard deviation for the distribution.

(c) Find (i) Q_1, Q_2, Q_3
(ii) Interquartile Range
(iii) $P_{12}, P_{25}, P_{75}, P_{80}$
(iv) D_1, D_2, D_6, D_9

7.0 REFERENCES/FURTHER READING

Donal, R.P. & Edward, B. O. *Business and Economics Statistics*, (Revised edition). Business Publication Inc.

McClave & Sincich: *Statistics*, (Eighth Ed). Prentice Hall.

Murray R. Spiegel: *Theory and Problems of Statistics*, Schaum's outline series. McGraw-Hill.

UNIT 4 INTRODUCTION TO PROBABILITY AND DISTRIBUTION

CONTENTS

1.0 INTRODUCTION

Business decisions and economics analysis often face an uncertain environment. Under uncertainty, one does not know which of several events or outcomes will occur in the future. Because of the importance of uncertainty, a precise terminology for dealing with uncertainty is needed. Probability theory is such a precise terminology: it is the language of uncertainty.

2.0 OBJECTIVES

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

- explain probability theory and probability distributions
- describe the approaches in assigning probabilities
- define the rules of probability
- itemise random variables
- explain conditional probability together with some other standard probability distributions such as Bernoulli, binomial etc.

3.0 MAIN CONTENT

3.1 Some Basic Definitions

- **A trial:** this is any process or attempt which generates an outcome which cannot be predetermined. A trial usually results into only one of the possible outcomes for example (I) a toss of a coin once, will lead to either a head (H) or a tail (I).
- **A random experiment:** this is an operation which when repeated, generates a number of outcomes which cannot be predetermined, for example, the toss of two coins at a time.
- **An outcome:** this is a possible result of a trial or experiment in a toss of two coins, an outcome could be any one of HH, HT, TH, and TT.
- **Sample space:** this is the collection of all possible outcomes. It is a set of all finite or countable infinite number of elementary outcomes. e.g. the sample space in a toss of two coins is given as S= (HH, HT, TH, TT).
- **An event:** this is a subset of a sample space, it consists of one or more possible outcomes e.g. when a die is tossed once. A= (set of even number) = (2, 4, 6).
- **Mutually exclusive events:** two events, A & B are said to be mutually exclusive, if the occurrence of A prevents the occurrence of B. This implies that the two units cannot occur together i.e. answer = O.
- **Mutually exhaustive events:** events A and B are said to be mutually exhaustive if they constitute the sample space i.e. $\sum \bar{a}_i = S$
- **Independent events:** two events A & B are said to be helped if the occurrence of A does not affect B. This implies that the two events can occur together.

3.2 Three Approaches to Defining Probability

Depending upon the type of event to be performed, there are various ways of assigning probabilities to the sample events in a sample space.

i. **Probabilities based on equally likely sample events** (A prior probability): This method assumes that the elementary outcomes of an expected event are equally likely.

i.e. **P (A)** $= \frac{N(A)}{N(S)}$

Where **P(A)** is the probability that event A will occur,
N (A) is the event space and N(S) is the sample space.

ii. Probabilities as relative frequency:

This approach defines probabilities as an idealisation of the proportion of time that a certain event will occur in respected trial of an event under the same condition. If what we expect is N and n(A) is the number of times that A occurs, then the relative frequency is R.F. = $\frac{n(A)}{N}$

iii. Subjective or personal probabilities: These are the probabilities assigned to an event based on subjective judgment from experience, information and behaviour. Such probabilities assigned arbitrarily are usually influenced by the bias and experience of the person assigning it.

In general, probability can be defined as the likelihood or chance that an event will occur. In other words, the probability of an event is the number of times the event occurs compared to the total possible equally likely time that the event can occur. It is also defined as the limiting value of a relative frequency of an event.

3.3 Probability of Event

The probabilities of an event are the sum of the probabilities of the simple events that make up the event.

Example

A coin is rolled 3 times, what are the probabilities of getting:

(i) 1 head (ii) 2 heads (iii) at least 2 heads

Solution

Let H & T represent H & T respectively, let the simple space be defined as S = (HHH, HTH, HHT, THH, TTH, HTT, THT, TTT).

(1) P (1head) = HTT, THT, TTH) = 3/8
(2) P (2heads) = (HHT, THH, HTT) = 3/8
(3) P (at least 2 heads)= P (2heads) + P (2heads) = $^{3}/_{8} + {}^{1}/_{8} = 0.5$

3.3.1 Rules of Probabilities

i. P (AUB) = P(A) +P(B) – P(UB)
ii. P(A)= 1-P(A)
iii. P(A) = 1 for certainly then o< P(A) ≤ 1

3.4 Conditional Probability

If E_1 & E_2 are 2 events, the probability that E_2 occurs given that E_1 has occurred is denoted by P (E_2/E_1) and it is called the conditional probability of E_2 given that E_1 has occurred. If the occurrence or non occurrence of E_1 does not affect the probability of occurrence of E_2 then Pr (E_2/E_1) = P (E_2) and we say that E_1 and E_2 are independent events otherwise they are dependent events.

Example

A bag contains 10 white balls and 15 black balls. Two balls are drawn in succession (a) with replacement (b) without replacement. What are the probabilities that;

i. the first ball is black and the second ball is white
ii. both are black
iii. both are of the same colour
iv. both are of different colours
v. the second is black given that the first is white.

Solution

Let W & B rep. white & black balls

(a) With replacement
Total no of balls 10+15=25

i. P(BnW) = P(B)xP(W) $= \frac{15}{25}\times\frac{10}{25} = 0.25$

ii. $P(B_1nB_2)$= $P(B_1)xP(B_2)= \frac{15}{25}\times\frac{15}{25} = 0.36$

iii. P (both black or both white)

$=P\left(P(B_1nB_2)+P(B_1nB_2)\right)=\left[\frac{15}{25}\right]^2+\left[\frac{10}{25}\right]^2=0.52$

iv. P (both are of diff colour) = P (BnW) + P (WnB)

$=\left(\frac{15}{25}\times\frac{10}{25}\right)+\left(\frac{10}{25}\times\frac{15}{25}\right) = 0.48$

v. $P(B/w) = \frac{P(BnW)}{P(W)} = \frac{0.24}{0.4} = 0.6$

b. without replacement

i. P(BW) = P(B)x P(w) $= \frac{15}{25}\times\frac{w}{24} = 0.25$

ii. $P(B_1 \cap B_2) = P(B_2) \times P(B_2) = \frac{15}{25} \times \frac{14}{24} = 0.35$

iii. P(both black or both white) = $P(B_1)\ P(B_2) + P(w)P(w_2)$

$= \frac{15}{25} \times \frac{14}{24} + \frac{10}{25} \times \frac{9}{24} = 0.35 + 0.15 = 0.50$

iv. P (both are of diff. whom) = $P(B)\ \beta(w) + P(w)PB/$

$= \frac{15}{25} \times \frac{10}{24} + \frac{10}{25} \times \frac{15}{24} = 0.25 + 0.25 = 0.50$

v. $P(B/W) = \frac{P(BnW)}{P(W)} = \frac{15}{25} \times \frac{10}{24} = 0.625$

3.5 Random Variables

A random variable is a function or rule that assigns numerical value to each simple event of a sample space e.g. a fair coin is tossed 3 times. The sample space

S = { $TTT, HTT, TTH, THT .....HHH$ }

Let x denote the number of heads which appears

Then $X_{(s)}$ = { 0,1,2,3 }.

3.5.1 Types of Random Variable

There are 2 types of random variable namely: discrete and continuous random variable.

(i) **Discrete random variable**: A discrete random variable is a rule or function that assigns only specified numerical value often integer to the simple event of a sample space.

(ii) **Continuous random variable: A** continuous **random variable** is a rule or function that may assign any numerical value (within some limits) to simple events.

3.6 Probability Distribution for R.V.

Probability can be defined as the arrangement of the events and their corresponding probabilities in a tabular form. However, the probability of any event depends on the type of variable involved.

3.6.1 Standard Probability Distribution

i. Bernoulli distribution

A random variable X that assigns only the value 0 or 1 is known as Bernoulli variable and a performance of an experiment with only 2 types of outcome is called Bernoulli trial. The pdf can be experienced as

$$f(x) = P^x q^{1-x},\ x = 0,1.$$

ii. Binomial distribution

In a sequence of n independent, Bernoulli trial with probability of success P on each trial, let X represent the no of successes. The discrete pdf of X is given by

$$f(x) = \binom{n}{x} p^x q^{n-x} \quad x = 0,1,2,3....n$$

Example

A fair coin is tossed 5 times, what are the probabilities of obtaining:

(i) 3 heads (ii) less than 2 heads (iii) 2 or more heads

Solution

L = 5, p = 1/2, q = ½

$$f(x) = \binom{n}{x} p^x q^{n-x}$$

i. $P(3) = \binom{5}{3}\left(\frac{1}{2}\right)^3\left(\frac{1}{2}\right)^{5-3} = 5/6$

ii. $P(x<2) = f(0)+f(1) = \binom{5}{0}\left(\frac{1}{3}\right)^0\left(\frac{1}{3}\right)^{5-0} f\binom{5}{1}\left(\frac{1}{2}\right)\left(\frac{1}{3}\right)^{1-1} = 6/32$

iii. $P(\geq 2) = 1 - P(x>2) = 1 - {}^6/_{32} = {}^{13}/_{16}$

iii. Poisson distribution

A discrete random variable X with probability function

$$f(x,\lambda) = \frac{\lambda^x e^{-\lambda}}{X} \qquad X = 0,1,2,3....$$

Where $\lambda > 0,$ is said to have a Poisson distribution with parameter λ .

Example

In a certain resort area, the no. of vacant rooms follow a Poisson distribution with the expected vacancy rate of 10 per night. Find the probability that in any one night the no. of vacant rooms will be (i) none (ii) one (iii) seven

Solution

$\lambda = 10, x = 0$

$$f(x,\lambda) = \frac{\lambda^x e^{-\lambda}}{X}$$

(i) $fx = \frac{\ell^{-10}10^0}{o} = 0.0000454$

(ii) λ = 10, x = 1

$p(x) = \frac{\ell^{-10}10^1}{} = 0.000454$

(iii) λ = 10 , x = 7

P(x) = $\frac{\ell^{-10}10^7}{7} = \frac{453.99929}{5040} = 0.090074.$

iv. Normal distribution

A continuous random variable X is said to have a normal distribution, if its probability density function is

$$f(x) = \frac{1}{\sigma\sqrt{2\Pi}}\ell^{-\frac{1}{2}\left(\frac{x-u}{\sigma}\right)^2} \qquad -\infty < x < \infty$$

v. Standard Normal distribution

A random variable Z is said to be a standardised normal distribution, if its probability density function is given by

$$f(z) = \frac{1}{2n}\ell^{-\frac{1}{2}z^2}$$

When u = o & σ^2 = 1

i.e. $z \approx N(0,1)$

Application

The Standard normal distribution table is used for determining the probability. of a value x drawn from a normal population with known u & standard deviation σ.

The standard normal variety Z is given by Z = $\frac{x-u}{\sigma}$

$P(x_1 \leq x \leq x_2) = P(z_1 \leq z \leq z_2)$

Where $Z_1 = \frac{x_1 - u}{\sigma}$ and $Z_2 = \frac{x_2 - u}{\sigma}$

For known value of u and σ
Z is also referred to as standard score.

SELF ASSESSMENT EXERCISE

1. Find the probabilities and state the approach used to assign the probability in the followings.

a. What is the probability that a coin, when tossed, will come down heads?
b. What is the probability that the next doctor you meet will be male?
c. Of the last 50 snacks supplied by a caterer, 5 got burnt, what are the probabilities that the next snacks will be burnt.

2. A coin is tossed 4 times, find the probabilities at obtaining
   i. heads ii. no head

4.0 CONCLUSION

In business problems, because nothing is certain or entirely predictable, a study of probability and its use in aiding decision making is essential.

5.0 SUMMARY

- We have been able to understand probability as a measure of likelihood that a particular event will occur or not.
- The three approaches of assigning probabilities to simple events in a sample space depend on the types of what we expect to be performed.

6.0 TUTOR-MARKED ASSIGNMENT

1. If the probability that a civil servant owns a car is 1/6 find the probability that:

a. 2 civil servants, A & B, selected at random each owns a car.
b. Of 2 civil servants, C&D selected at random, only one owns a car

c. Of three civil servants, X, Y& Z selected at random, only one owns a car.

2. A bag contains 5 white balls and 8 red balls, if 2 balls are picked at random without replacement, what are the probabilities of picking:

(a) 2 red balls (b) One red and a white ball in that order
(c) One red and one white

3. A particular piece of medical equipment is known to have 25% probability of failure during the 1st year of its life. A user of the equipment purchased one unit that contains 13 pieces of equipment.

What is the probability that: (a) none will fail
(b) at least 3 and at most 4 will fail (c) more than 2 will fail

4. Nicon Insurance is considering stating a policy for an ailment which affects 0.2 % of the population, if a random sample of 2000 individuals is selected find:

(i) the probabilities that exactly 3 have ailment
(ii) at most 6 and at least 3 have ailment

7.0 REFERENCES/FURTHER READING

Donal, R. P. & Edward, B. O. *Business and Economics Statistics*, (Revised Edition). Business Publication Inc.

McClave & Sincich: *Statistics*, (Eighth Ed). Prentice Hall.

Murray R. Spiegel: *Theory and Problems of Statistics*, Schaum's Outline series. McGraw-Hill.

MODULE 2 STATISTICAL INFERENCE

Unit 1 Estimation
Unit 2 Hypothesis Testing

UNIT 1 ESTIMATION

CONTENTS

1.0 INTRODUCTION

Statistical inferences involve the procedures of drawing conclusion about a population on the basis of the results obtained from a sample drawn from the population. Formal statistical inference involves using the tool of probability theory to quantify the reliability of some kinds of data based on conclusions.

2.0 OBJECTIVES

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

- explain estimation
- identify point and interval estimation
- explain confidence interval and its interpretation
- calculate Confidence Interval [C.I] for a population mean and the difference between two means.

3.0 MAIN CONTENT

3.1 Uses of Sampling for Making Inferences

(i) A manager is interested in knowing the average age of the employees.
(ii) An advertising executive wants to know the proportion of subscribers to a certain magazine that remembers a particular advertisement.
(iii) A sales manager wants to know the proportion of the population of subscribers to a certain magazine that remembers a particular advertisement. .
(iv) An educationist wants to know whether one educational procedure is better than another.

3.2 Sampling Theory

This is the study of relationship that exists between a population and samples drawn from the population. It is used in estimating unknown population parameters i.e. the mean μ and variance σ^2. It is also useful in determining whether the observed differences between two samples are due to chance variation or significant.

3.3 Estimation Theory

Estimation is basically concerned with estimated parameter from the sample statistics. An estimator is the procedure or rule usually expressed as a mathematical formula that tests how an estimate is determined.

3.3.1 Types of Estimator

An estimator is a formula or rule that we use to estimate the value of population parameters. There are two types namely: point estimate and interval estimate.

(i) Point Estimate

This is a numerical value obtained from a set of data, which is used as an estimate of an unknown parameter in a population e.g.

$\bar{x} = \frac{\sum x}{n}$ = is used as an estimate of sample mean

$S^2 = \frac{\sum (x_i - \bar{x})^2}{n-1}$ is used as an estimate of the population variance.

(ii) Interval Estimate

This consists of an interval that contains the parameter to be estimated within varying degree of conditions.

Estimators are usually judged using the criteria below.

(i) Unbiasedness
An estimator U is an unbiased estimator of population parameter θ if E (U) = θ

(ii) Consistency
If U is an unbiased estimator of θ and if Var (U) $\Rightarrow$ θ as the sample size increases, then U is a consistent estimator of θ.

(iii) Efficiency
If U and V are two unbiased estimators of θ with Var (U) < Var(V), then U is more efficient than V.

(iv) Sufficiency
Suppose $x_1,x_2,\ldots x_n$ is a random sample of size n and let T be an estimator of θ, if the conditional distribution of x_1, x_2, … x_n is given by T = t (for some given t) is independent of θ, the T is said to be sufficient for θ, if it contains all the relevant information available about θ from the given sample.

(v) Minimum Variance Unbiased Estimator
Let x_1,x_2, … x_n be a random sample of size n from f(x, θ). Let T = t $(x_1,x_2,\ldots x_n)$ be an unbiased estimator of θ such that:
E(T) = θ Var(T) <Var(T')

Where T is any other unbiased estimator of θ then T is said to be minimum variance unbiased estimator of θ.

3.4 Confidence Intervals

Confidence Interval for a parameter (or function of one or more parameters) is a data-based interval of numbers thought likely to contain the parameter (or function of one or more parameters) possessing a stated probability-based confidence or reliability. Based on a set of data, one might want to identify an interval of values that is likely to contain an unknown parameter and quantify in some sense "how likely" the interval is to cover the correct value.

A β% confidence interval for a population parameter is an interval obtained from a sample by some specified methods in such a way that, in repeated sampling, in the long run β% of the intervals thus obtained include the value of the parameter.]

Illustration

Consider a very simple research situation, in which a research worker wishes to estimate populations mean μ. A doctor wishes to study the effect of a new substance added to the diet of infants. He gives the diet to a group of 16 infants and measures the gains in weight over a one month period. He then, takes the arithmetic mean of the 16 gains in weight and finds, say $\bar{x}$ = 421.8g.

Normally, the specific sixteen infant gains are not the population which, as a research worker, the doctor wishes to study. He really wishes to know the mean gain in weight for infants in general. The population, then, consists of infants of the type whose diet comes under this doctor's supervision, and he wishes to estimate μ, the population mean.

The doctor's best point estimate for μ is his sample mean, or 421. 8g. But it is practically certain the μ is not exactly 421.8g, and the doctor wishes to get some ideas of where μ may reasonably be expected to be. Then, a confidence interval will be formed for μ.

If a higher confidence level is chosen, then, there is a greater confidence that the interval contain μ; but on the other hand, we pay for this higher level of confidence by having a longer interval.
To obtain a shorter interval and at the same time having one in which we have a high level of confidence; the sample size must be increased.

Interpretation of a Confidence Interval
To say that a numerical interval (a,b) is a 90% confidence interval for a parameter is to say that in obtaining it, one has applied methods of data collection and calculation that would produce intervals bracketing the parameter in about 90% of repeated applications.

3.5 Confidence Interval for a Population Mean

The following cases should be considered.

A sample (large or small) is taken from a normally distributed population, with known variance.

$$\text{C.I.} = \bar{x} \pm z_{\alpha} \frac{\sigma}{\sqrt{n}}$$

A large sample is taken from an unknown population distribution with known variance:

$$\text{C.I.} = \bar{x} \pm z_{\alpha} \frac{\sigma}{\sqrt{n}}$$

A large sample is taken from an unknown population distribution with unknown variance:

$$\text{C.I.} = \bar{x} \pm z_{\alpha} \frac{s}{\sqrt{n}}$$

A sample (large or small) is taken from a normally distributed population with unknown variance.

$$\text{C.I.} = \bar{x} \pm t_{n-1} \frac{s}{\sqrt{n}}$$

Where $\bar{x}$ is the sample mean
z_{α} is the confidence level
σ is the standard deviation
S is the sample standard deviation
n is the sample size
t_{n-1} is the confidence level with n-1 degree of freedom

Example 1

The standard deviation of bulbs manufactured by Solar Engineering Company is 5.6 if the mean life span of 64 bulbs were randomly selected from the lot as 60 days.

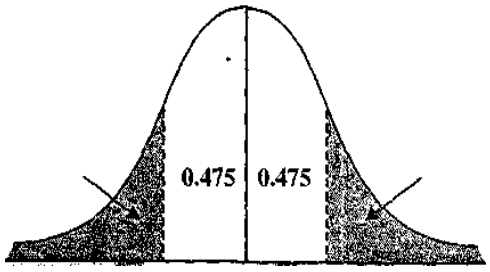
i. Construct the 95% confidence limit for the bulb.
ii. What is the minimum number of samples to be selected so that the error does not exceed 0.5?

Solution

Here, the population distribution is unknown, the variance is known since the standard deviation is given, and also a large sample i.e. n = 64> 30.

This is case 2.

(i) σ = 5.6, $\bar{x}$ = 60, n = 64



$\alpha/2$ =0.025 $\alpha/2$ = 0.025

-1.96

0 +1.96

α = 95% = 0.95 $\therefore$ 1-0.95 = 0.05

Then, $\frac{0.05}{2}$ = 0.025 $\Rightarrow$ 0.5-0.025 = 0.475

Check 0.475 under the body of your normal distribution table gives 1.96 $\Rightarrow z_\alpha$ = 1.96.

$\therefore$ The confidence interval C.I. = $\bar{x} \pm z_\alpha \frac{\sigma}{\sqrt{n}}$

$= 60 \pm 1.96 \text{ x } \frac{5.6}{\sqrt{6.4}}$

$= 60 \pm 1.372$

= [60-1.372, 60+1.372]

= [58.628, 61.372]

(ii) Error $\leq$ 0.5

$z_\alpha \frac{\sigma}{\sqrt{n}} = 0.5$

$\frac{1.96x5.6}{\sqrt{n}} \leq 0.5$

$0.5\sqrt{n} \leq 1.96 \text{ x } 5.6$

$\sqrt{n} \leq \frac{1.96x5.6}{0.5}$

$(\sqrt{n})^2 \geq (21.952)^2$

n $\geq$ 482

Therefore, for the error to be less than 0.5 n should be greater than 482.

Example 2

The random variable X has a normal distribution with mean μ. A random sample of 40 observations of x is taken and gives $\sum x$ = 83.3, $\sum x^2$ = 721.41. Find a 95% confidence interval for μ.

Solution

Here, the variance is unknown; the sample is taken from a normal distribution and n> 30 $\Rightarrow$ large.

This is case 4

$$\therefore \quad S^2 = \frac{\sum x^2 - \frac{(\sum x)^2}{n}}{n-1}$$

$$= \frac{721.41 - \frac{(83.3)^2}{40}}{40-1}$$

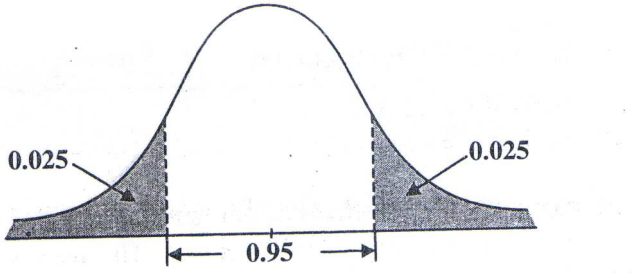
$$= 14.05 \qquad \Rightarrow \qquad s = 3.75$$

$$\therefore \text{ C.I} = \bar{x} \pm t_\alpha \frac{s}{\sqrt{n}}$$

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

$$= \frac{83.3}{40} = 2.0825$$

$t_\alpha = t_{95\%,n-1}$



$1 - 0.95 = 0.05 \qquad \therefore \qquad \frac{0.05}{2} = 0.025$

Then, 0.95 + 0.025 = 0.975

$\therefore \ t_\alpha = t_{95\%,n-1}$

$= t975.39$

This value cannot be obtained from the table and must be interpolated as:

2.04
= 2.04x0.1+2.02x0.9
= 2.022

C.I = $\bar{x} \pm t_{\alpha} \frac{s}{\sqrt{n}}$

= $2.0825 \pm 2.022 \text{ x } \frac{3.75}{\sqrt{40}}$

= 2.0825 ± 1.1989
= [2.0825-1.1989, 2.0825+1.1989]
= [0.8836, 3.2814]

Example 3

A medical practitioner is studying the effects of a new substance added to the diet of infants to improve their weights over a period of a month. The results of the effect on 9 infants chosen as samples are shown in the table below:

Infants	1	2	3	4	5	6	7	8	9
W_1 (Original weights)	256.1	302.2	287.1	289.5	300.3	307.4	320.2	296.1	288.9
W_2 (New weights)	299.5	352.3	387.4	378.2	368.2	329.5	386.9	396.2	336.8

Find a 95% symmetric confidence interval for the weight gain added if the distribution is assumed to be normally distributed.
Solution:

Infants	W_1	W_2	$X=W_2-W_1$	X_2
1	*256.1*	*299.5*	*43.4*	*1883.56*
2	302.2	352.3	*50.1*	*2510.01*
3	287.1	387.4	100.3	10060.09
4	289.5	378.2	88.7	7867.69
5	300.3	368.2	67.9	4610.41
6	307.4	329.5	22.1	488.41
7	320.2	386.9	66.7	4448.89
8	296.1	396.2	100.1	10020.01
9	288.9	336.8	47.9	2294.41
TOTAL			587.2	44183.48

Where, x is gain in weight over a period of 1 month.
Here, the variance is unknown, the distribution is normal and n = 9 < 30 (small sample). Then, this is case 4.

Sample mean $\bar{x} = \frac{\sum x}{n}$

$= \frac{587.9}{9} = 65.24$

$\therefore\ S^2 = \frac{\sum x^2 - \frac{(\sum x)^2}{n}}{n-1}$

$= \frac{44183.48 - \frac{(587.2)^2}{9}}{9-1}$

$= 733{,}993 \qquad \Rightarrow s = 27.09$

$\therefore\ = t_\beta = t_{95\%, n-1}$

$= t_{0.95+0.025,\ 9-1}$

$= t_{0.975,8}$

$= 2.31$

$\therefore$ C.I $= \bar{x} \pm t_\alpha \frac{s}{\sqrt{n}}$

$= 65.24 \pm \frac{2.31x27.09}{\sqrt{9}}$

$= 65.24 \pm 20.86$

$= [44.38, 86.1]$

Example 4

A machine is regulated to dispense liquid into cartons in such a way that the amount of liquid dispensed on each occasion is normally distributed with a standard deviation of 20ml. Find 99% confidence limits for the mean amount of liquid dispensed if a random sample of 40 cartons had an average content of 266ml.

Solution

Hence, the distribution is normal, n is large; the variance is known since the standard deviation is given. Then this is case 1.

n= 40 $\bar{x}$ = 266ml, σ = 20ml

$\therefore$ C.I $= \bar{x} \pm z_\alpha \frac{\sigma}{\sqrt{n}}$

Check 0.495 from the body of your normal distribution table gives: 2.58

$\therefore\ z_\alpha = 2.58$

$\text{C.I} = 266 \pm \frac{2.58x20}{\sqrt{40}}$

$= 266 \pm 8.159$

= [266-8.159,266+ 159]

= [257.841, 274.159]

Example 5

A random sample of 36 scores in statistics examination out of a total of scores obtained by 150 students showed a mean of 60 and a standard deviation of 6. Compute a 95% confidence interval for estimates of the mean of the 150 scores.

Solution

Here, n= 36, $N_p = 150$, $\bar{x} = 60$ $z_\alpha = 1.96$

$\therefore \text{C.I.} = \bar{x} \pm z_\alpha \sqrt{\frac{n_p - n}{n_p - 1}}$

$= 60 \pm = 1.96x \frac{6}{\sqrt{36}} \sqrt{\frac{150-36}{150-1}}$

$= 60 \pm 1.7144$

[60-1.7144, 60+1.7144]

= [58.2856, 61.7144]

3.5.1 Confidence Interval for Difference Between two Means: μ_1, μ_2

1. Known Variance

For samples of size n_1 and n_2 respectively, from a normal population with mean μ_1 and μ_2 and with variances σ^2, the statistic $\overline{x_1} - \overline{x_2}$ has a normal distribution with mean $\mu_1 - \mu_2$ and with variance

$$\sigma^2\left(\frac{1}{n_1} + \frac{1}{n_2}\right)$$

Therefore, the C.I. for difference between two means if the variance is known is:

$$\text{C.I} \left[\bar{x}_1 - \bar{x}_2 \pm z_\alpha \sigma_{\bar{x}} \sqrt{\frac{1}{n_1} + \frac{1}{n_2}} \right]$$

2. Unknown Variance

Where $S^2 = \frac{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2}{n_1 + n_2 - 2}$

$$s_1^2 = \frac{\sum x_1^2 - \frac{(\sum x_1)^2}{n_1}}{n_1}, s_2^2 = \frac{\sum x_2^2 - \frac{(\sum x_2)^2}{n_2}}{n_2}$$

$\therefore$ The confidence interval,

$$\text{C.I.} = \left[\bar{x}_1 - \bar{x}_2 \pm t_\alpha \left(\sqrt{\frac{1}{n_1} + \frac{1}{n_1}} \right] \right]$$

With n_1+n_2 -2 degree of freedom.

Example 6

If $\bar{x}_1 = 248.1$g
$\bar{x}_2 = 130.1$g,
$n_1 = 9$,
$n_2 = 16$,
$\sigma_{\bar{x}} = 56$

$$\text{C.I} = (\bar{x}_1 - \bar{x}_2) \pm z_\alpha \left(\sqrt{\frac{1}{n_1} + \frac{1}{n_2}} \right)$$

$$= (248.1\text{-}130.1) \pm 1.96 \text{x} 56 \text{x} \sqrt{\left(\frac{1}{9} + \frac{1}{16} \right)}$$

$= 118 \pm 45.7$
= (118-45.7, 11845.7)
= (72.3, 163.7)

Example 7

The tables below show the gain in weight of two different samples consisting of 16 and nine infants under supplemented and standard diet in a community. Table (a) is supplemented and table (b) is standard.

Table (a) Supplemented Diet

Infants Number	1	2	3	4	5	6	7	8
Gain in Weights	221	423	305	502	426	501	239	304
	9	10	11	12	13	14	15	16
	326	376	342	356	329	128	156	98

Table (b) Standard Diet

Infants number	1	2	3	4	5	6	7	8	9
Gain in Weights	220	205	232	182	183	176	124	192	180

Find a 95% confidence interval for the difference in the two mean weights.

Solution

Supplementary Diet		
Infant number	X_1 Gain in Weight	x_1^2
1	221	48841
2	423	17829
4	305	93025
5	502	252004
6	426	181476
7	501	251001
8	239	57121
9	304	92416
10	326	106276
11	376	141376
12	342	116964
13	356	126736
14	329	108241
1516	128	16384
	156	24336
	98	9604
Total	5032	1804730

$$s_1^2 = \frac{\sum x_1^2 - \frac{\left(\sum x_1\right)^2}{n_1}}{n_1 - 1}$$

$$= \frac{1804730 - \frac{(5032)^2}{16}}{16 - 1} = 14811.07$$

Supplementary Diet		
Infant number	X_1 Gain in Weight	x_1^2
1	220	48400
2	205	42025
3	232	53824
4	182	33124
5	183	33124
6	176	33489
7	124	15376
8	192	36864
9	180	32400
Total	1694	326478

$$\bar{x}_1 = \frac{5032}{16} = 314.5$$

$$\bar{x}_2 = \frac{1694}{9} = 188.22$$

$$S_2^2 = \frac{\sum x_2^2 - \frac{(\sum x_2)^2}{n_2}}{n_2}$$

$$= \frac{326478 - \frac{(1694)^2}{9}}{9 - 1} = 953.7$$

Pooled estimate of variance is

$$s^2 = \frac{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2}{(n_1 + n_2 - 2)}$$

$$= \frac{(16 - 1)(14811.07 + (9 - 1)(953.7)}{(16 + 9 - 2}$$

$$= 9991.12 \qquad \Rightarrow s = 99.96$$

Then, the confidence interval is:

$$\text{C.I.} = \left[(\bar{x}_1 - \bar{x}_2 \pm t_\alpha s \sqrt{\frac{1}{n_1} + \frac{1}{n_2}} \right]$$

$$\text{C.I.} = \left[314.5 - 188.22 \pm t_\alpha x 99.96 \sqrt{\frac{1}{16} + \frac{1}{9}} \right]$$

$$= [126.28 \pm t_\alpha x 41.65]$$

$$= t_\alpha = t_{95\%,n1+n2-2}$$
$$= t_{95\%,16+9-2}$$
$$= t_{0.975,23} = 2.07$$

$\therefore$ C.I. = (126.28 ± 2.07 x41.65)
= (126.28 ± 86.22)
= (126.28-86. 22, 126. 28+86.22)
= (40.06, 212.5)

$\hat{q}$ is the sample population for failure

n is the number of observation

$\hat{p} = \frac{r}{n'}$ r is the number of successes in a large sample of size n

Example 8

During a local government election, a random sample of 1,000 voters were interviewed, of whom 349 stated that they will support Alliance for Democracy party. Determine a 98% symmetric confidence interval for the proportion of Alliance for Democracy party in the population.

Solution

Here, n = 1000, r = 349

$$\hat{p} = \frac{r}{n}$$

$$= \frac{349}{1000} = 0.349$$

$$\hat{q} = 1 - \hat{p} = 0.651$$

z_α = 2.367(from the statistical table)

$$\therefore \text{ C.I.} = \hat{p} \pm z_{\alpha}\sqrt{\frac{\hat{p}\hat{q}}{n}}$$

$$= 0.349 \pm 2.3267 \text{ x } \sqrt{\frac{0.34x0.651}{100}}$$

$$= [0.349\text{-}0.0351,\ 0.349\text{+}0.0351]$$

$$= [0.3139,\ 0.3841]$$

SELF ASSESSMENT EXERCISE

What do you understand by the following?

(i) Point estimate (ii) Interval estimate

4.0 CONCLUSION

Estimation is a very useful tool for conveying from one person (statistician) to another person (the user) the information contained in a sample. If a single number estimate is required, a point estimate is used. If a measure of the precision of the estimate is required, the point estimate is converted into a confidence interval. This unit has dealt with the two types of estimation: point estimate and confidence interval.

5.0 SUMMARY

In this unit we have learnt that:

- point estimate is the single value computed using an estimator
- a good estimator must be unbiased, consistent, efficient and sufficient
- estimated parameter is obtained from sample statistics.

6.0 TUTOR-MARKED ASSIGNMENT

1. Packets of soap powder are filled by machinery. The weights of powder (to the nearest gram) in 32 packets chosen at random are summarised below:

Weight	999	1000	1001	1002	1003	1004
Packets	1	7	12	8	3	1

Find:

i. the amount by which the mean exceeds 1000g.
ii. the standard deviation.

iii. the random variable X has a normal distribution with mean μ. A random sample of 10 observations of X is taken and $\sum x_i = 83.3 \quad , \sum x_i^2 = 721.41$

Find

(a) a 95% confidence interval for μ

(b) a 99% confidence interval for μ

7.0 REFERENCES/FURTHER READING

Donald R.P. & Edward B. O. *Business and Economics Statistics,* (Revised Edition). Business Publications Inc.

McClave & Sincich: *Statistics*, (Eighth Ed). Prentice Hall.

Murray R. Spiegel: *Theory and Problems of Statistics,* Schaum's outline series. McGraw-Hill.

UNIT 2 HYPOTHESIS TESTING

CONTENTS

1.0 INTRODUCTION

A statistical hypothesis is merely a statement. As a statement, it may or may not be true. Perhaps supposition is a better word than hypothesis or statement.

2.0 OBJECTIVES

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

- explain basic concept of hypothesis testing
- identify the essence of test statistics in hypothesis testing
- explain critical and acceptance regions
- identify steps involved in hypothesis testing
- explain how to solve problems that are concerned with hypothesis testing.

3.0 MAIN CONTENT

3.1 What is Hypothesis?

A hypothesis is an idea or suggestion that is based on known facts and is used as a basis for reasoning or further investigation. Actually, a hypothesis is formulated with the hope of either accepting or rejecting it.

3.1.1 Types of Hypothesis

(i) Null Hypothesis

This is a statement that forms the basis of investigation in a significant test. A null hypothesis is usually formed to embody a status quo ‘‘pre-data” view of the parameter (s). It is typically symbolised as H_0. The null part of the phrase “null hypothesis” refers to the fact that null hypotheses are statements of no difference, or equality.

(ii) Alternative Hypothesis

This is a statement that stands in opposition to the null hypothesis; it specifies what forms of departure from the null hypothesis are of potential concern. It is symbolised as H_1 and has the same form as the corresponding null hypothesis except that the equality sign is replaced $\neq$,> or <.

In many applications, the alternative hypothesis embodies an investigator’s suspicions and/or hopes about the true state of affairs, amounting to a kind of research hypothesis that the investigator tends to establish.

A full specification of hypothesis test must be one of the followings:

(i) $H_0: \mu = a$; $H_1: \mu < a$ (one-sided) => lower tail test
(ii) $H_0: \mu = a$; $H_1: \mu > a$ (one-sided) => upper tail test
(iii) $H_0: \mu = a$; $H_1: \mu \neq a$ (two-sided) => two tail test

The sidedness of a test is the direction specified by the alternative hypothesis H_1 indicating the direction of a further choice of a new H_0 should the current one be rejected.

A health researcher wishes to study whether vitamin E deficiency affects the storage of vitamin A in rats. He feeds ten rats a diet deficient in vitamin E and then measures the amount of vitamin A in the liver of each rat and obtained the data in the table below:

Rat Number	1	2	3	4	5	6	7	8	9	10
Vitamin A	3142	2134	2037	1700	1350	2170	3080	2010	2123	3014

The researcher has done a great deal of research with rats, and he knows from his experience that with a normal diet, the average amount of vitamin A is very close to 3,200 and that the standard deviation is very close to 500. He wants to decide on the basis of his data, if very insufficient vitamin E has changed the mean amount of vitamin A, or whether the mean under the deficient diet is the same as for rats under the normal diet.

He then formulates a null hypothesis which he proposes to test: i.e. the mean of a population of amounts of vitamin A in the liver of rats who are fed a diet deficient in vitamin E is 3,200. On the basis of his data, he wishes either to accept the null hypothesis or to reject the null hypothesis. His null hypothesis is equivalent to saying that the deficient diet has no effect, on the average, on the amount of vitamin A in the liver of rats i.e.

H_0:μ=3,200
H_1: μ<3,200

3.2 Test Statistics (TS)

This is the particular form of numerical data summarisation to be used in a hypothesis test. The formula for the test statistics typically involves the number appearing in the null hypothesis. The essence of test statistics in hypothesis test is a function of sampled data whose magnitude determines the acceptance or rejection of the H_0 under test.

The table below shows the distributions and their respective conditions:

Condition	Distribution/TS
Case 1 Test for mean, known variance, normal distribution or large sample i.e. >30	$Z = \frac{\bar{x} - \mu}{\sigma / \sqrt{n}}$
Case 2 Test for mean, large sample, variance unknown	$Z = \frac{\bar{x} - \mu}{s / \sqrt{n}}$
Case 3 Large mean, λ ,Poisson distribution	$Z = \frac{(\bar{x} + 0.5) - \lambda}{\sqrt{\lambda}}$ When 0.5 is the continuity

Case 4 Test for mean, small sample i.e n≤ 30, variance unknown	$T = \frac{\bar{x} - \mu}{s/\sqrt{\lambda}}$ With n-1 degree of freedom
Case 5 Test for proportion, large sample size	$2 = \frac{\bar{p} - p}{\sqrt{\frac{pq}{n}}}$

3.2.1 Rejection Region (RR) or Critical Region

The rejection region of a test is a set on the real number line specified in such a way that whenever the numerical value of the test statistics falls into this region, the H_0 is rejected and accepted if the numerical value of the test statistics falls outside the region.

Generally, rejection region or critical region refers to the set of values that lead to the rejection of H_o in favour of H_1 while, the set of values that leads to the acceptance of Ho is the acceptance region. The figures below show the region to be accepted or rejected or rejected depending on the sidedness of the test:

Two-tailed Test

Upper-tailed Tests

Lower-tailed Tests

3.2.2 Errors in Hypothesis Test

i. Types 1 Error

This is the possibility of deciding in favour of H_1 when in fact H_O is true. It is committed when a true hypothesis H_1 is rejected when it ought to have been accepted.

ii. Type II Error

This is the possibility of deciding in favour of H_0 when in fact H_1 is true. It is committed when an alternative hypothesis is rejected when it ought to have been accepted.

The content of the two definitions above is conveniently represented in the table below:

		Our Decision	
Reality	H_0 correct	We accept H_O correct	We reject H_0, Type I Error
	H_0 in correct	Type II error	Correct

3.2.3 Significance Level of a test: α

This is the probability that H_0 is nevertheless rejected in favour of H_1 when the population parameter has the value specified by H_O. It is the maximum probability of committing type one error when the null hypothesis holds. It is typically represented by α

3.2.4 Outlined Steps in Hypothesis Tests

1. Write down the two hypotheses e.g. $H_0 : \mu = \mu_0; \mu_0; H_1 : \mu < \mu_0$
2. Determine the appropriate test-statistic to be used and the distribution of the corresponding random variable
3. Determine the significance level of the test.
4. Decide on the distribution of the test statistics and the sidedness of the test i.e. whether single or double.
5. By using the significance level of the test sidedness of the test and the distribution of the test statistics, decide the boundary or boundaries of the rejection region.
6. Give the decision rule i.e. whether to accept or reject H_0

Example 11

A random sample of size 35 selected from a population whose distribution is normal with mean μ and variance 36 gives a sample mean 48. Test the hypothesis H_O: $\mu = 50$ against the sample mean 48. Test the hypothesis $H_0 : \mu < 50$ at 5% level of significance.

Solution

Here, $n > 30$ i.e. $n = 35$ and the variance σ^2 is known i.e. $\sigma^2 = 36 \Rightarrow \sigma = 6$

Step 1:

$H_O: \mu = 50$
$H_o: \mu < 50$

Step 2:

$$\text{T.S.} = \frac{\bar{x} - \mu}{\sigma/\sqrt{n}}$$
$$= \frac{48 - 50}{6/\sqrt{35}}$$
$$= 1.972$$

Step 3:

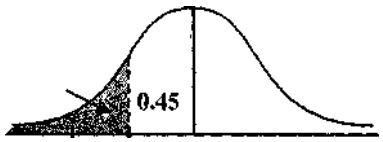
$\alpha = 5\%$

Step 4:

Since the alternative hypothesis is $H_0 : \mu < 50$ it is a lower tailed test.

Step 5:

0.05

1.972  -1.645 0

5% = 0.05 $\therefore 0.5 - 0.05 = 0.45$

Check this value under the body of your normal distribution table gives p-value

$\therefore$ p-value = -1.645
$\therefore$ The p-value = 1.645. Since this is a lower tailed test a minus sign is assigned to it.

Conclusion

The test-statistic falls within the rejection region, therefore, H_0 is rejected and H_1 is accepted. Then, we conclude that $\mu < 50$

Example 12

A random sample of size 40 selected from a normal population with mean μ and a standard deviation 4 gives a sample mean of 36, Test the hypothesis $H_0: \mu >= 30$ against H_1: $\mu > 30$ at 6% level of significance.

Solution

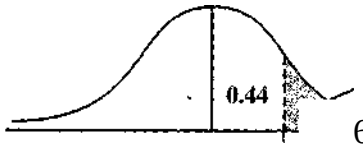
n=40, σ=4 $\bar{x} = 36$, $\alpha = 6\%$

The T.S = z = $\frac{\bar{x} - \mu}{\sigma / \sqrt{n}}$

$= \frac{36 - 30}{4 / \sqrt{40}}$

= 9.49

since H_1: $\mu > 30$ this is upper-tailed test



α = 6% = 0.06 $\therefore$ 0.5 -0.06 = 0.44

Check this value under normal table gives
$\therefore$ p-value = 1.555

Conclusion

The test-statistic falls within the rejection region. We therefore reject H_0 and conclude that u. > 30 at 6% significance level.

Example 13

A random sample of size 50 selected from a normal distribution N (μ, 4) population gives a sample mean of 20. Test the hypothesis H_0: μ = 21.5 against the alternative hypothesis H_1: $\mu \neq 21.5$ at 5% level of significance.

Solution

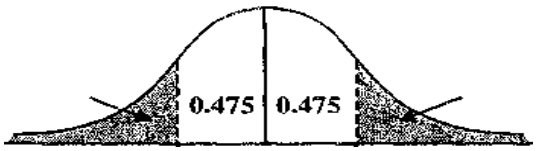
n= 50, $\sigma^2 = 4 \Rightarrow \sigma = 2$, $\bar{x} = 20$, $\alpha = 5\%$

The T.S. = z = $\frac{\bar{x} - \mu}{\sigma / \sqrt{n}}$

$$= \frac{20 - 21.5}{1/\sqrt{50}}$$
$$= -5.3$$

Since H_1: $\mu \neq 21.5$ this is a two-sided test

$\alpha/_2 = 0.025$ % = 0.025



-5.3 -1.96 0 +1.96 +5.3

Since it is a two-sided test, the RR consists of two parts i.e. the lower and the upper-tails. Then, α is shared between the two parts

α = 5% $\Rightarrow$ $\frac{\alpha}{2}$ = 0.025

$\therefore$ 0.5-0.025 = 0.475

Check this value under the body of your normal distribution table gives
$\therefore$ p-value = 1.96

Conclusion

The test-statistics fails within the rejection regions. We then reject H_0 and accept H_I and then conclude that $\mu \neq 21.5$.

Example 14

A random sample of 6 observations taking from a population whose distribution is normal with mean μ and unknown variance σ^2 gives the following observations 6, 4, 10, 2, 8, 5. Test the hypothesis H_O: μ = 6.5 against the alternative hypothesis H_I : μ <6.5 at 5% level of significance.

Solution

Variance unknown, n=6 i.e. small sample then t-distribution is used

$\sum x = 6 + 4 + 10 + 2 + 8 + 5 = 35$
$\sum x^2 = 6^2 + 4^2 + 10^2 + 2^2 + 8^2 + 5^2 = 245$
$\therefore \quad \bar{x} = \frac{\sum x}{n}$

$$= \frac{35}{6} = 5.833$$

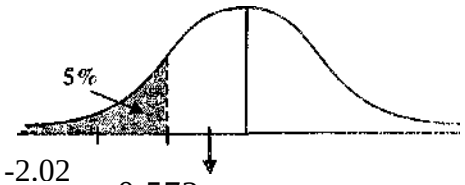
$$s^2 = \frac{\sum x^2 - \frac{(\sum x)^2}{n}}{n-1} = \frac{245 - \frac{(35)^2}{6}}{6-1}$$

$$= 8.167$$

$$\therefore \quad s = \sqrt{8.167} = 2.858$$

$$T.S = t = \frac{\bar{x} - \mu}{s/\sqrt{n}} = \frac{5.833 - 6.5}{\frac{2.858}{\sqrt{6}}}$$

$$= 0.572$$



-2.02 -0.572

5% = 0.05 ∴ 1-0.5 = 0.95

The p-value = $t_{\alpha,n-1}$

$= t_{5\%,n-1}$

$= t_{1-0.05,n-1}$

$= t_{0.95,6-1}$

$= t_{0.96,n-1}$

$= 2.02$ (since this is a lower tailed test minus is assigned to the p-value i.e. = -2.02)

Conclusion

The test-statistics falls within the acceptance region, we then accept H_0 and reject H_I and conclude that the mean $\mu = 6.5$

Example 15

The farm manager of "Adebowale Farm" wishes to determine whether eggs sold as size I have mean weight 70.0g, he weighs a sample of 200 eggs and his results are summarised by $\sum x = 13824$, $\sum x^2 = 957320$ where x is the weight of an egg in grams. Test whether there is significant evidence, at the 3% level, that the mean weight is not 70.0g.

Solution

n = 200 Unknown variance

H_o: μ = 70g

H_1: $\mu \neq 70$g

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

$$= \frac{13824}{200} = 69.12$$

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

$$= \frac{957320 - \frac{(13824)^2}{200}}{200-1}$$

$S^2 = 9.071$

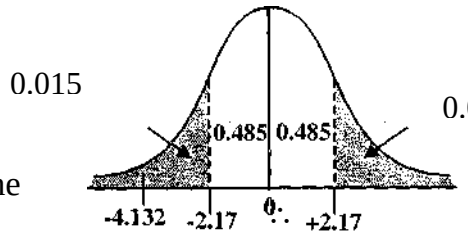
$\therefore$ s = 3.012

$$\text{T.S} = z = \frac{\bar{x} - \mu}{s/\sqrt{n}}$$

$$= \frac{69.12 - 70}{\frac{3.012}{\sqrt{200}}}$$

= -4.132

Since $\mu \neq$ 70g this is a 2-tailed test



The 0.5 – 0.015 = 0.485

$^{0.03}/_2 = 0.015$.

Check this value under the body of your normal distribution table gives:
p-value = 2.17 i.e. ±2.17

Conclusion

The test-statistics fails within the rejection region, we then reject H_0 and accept H_1 and therefore conclude that $\mu \neq$=70g i.e. the mean weight is not equal to 70g which means it could be greater than or less than 70.0g.

Example 16

The time taken for cessation of bleeding was recorded for a large number of persons whose fingers had been pricked. The mean time was found to be σ^2 minutes and the standard deviation was a^2 minutes. In an effort to determine whether pressure applied to the upper arm increases bleeding time, three persons had pressure equal to 20mm, applied to their upper arms and their fingers pricked. For these three persons, the times taken for bleeding to stop were 1.15, 1.75 and 2.50 minutes; test at 5% level of significance, whether the mean time is equal to 1.407 or not.

Solution

n = 3 (small sample), variance unknown;

$H_0: \mu = 1.407$

$H_1: \mu \neq 1.407$

$\sum x = 1.15 + 1.75 + 2.50 = 5.4$

$\sum x^2 = 1.15^2 + 1.75^2 + 2.50^2 = 10.635$

$\therefore$ Mean $\bar{x} = \dfrac{\sum x}{n}$

$= \dfrac{5.4}{3} = 1.8$

$s^2 = \dfrac{\sum x^2 - \dfrac{(\sum x)^2}{n}}{n-1}$

$= \dfrac{10.635 - \dfrac{(5.4)^2}{3}}{3-1}$

$S^2 = 0.4575$

$\therefore$ s = 0.6764

T.S = $t = \dfrac{\bar{x} - \mu}{s/\sqrt{n}} = \dfrac{1.8 - 1.407}{\dfrac{0.6764}{\sqrt{3}}} = 1.0064$

5% = 0.05 $\therefore$ 1-0.5 = 0.95

i.e p-value = $t_{\alpha, n-1}$

$= t_{5\%, n-1}$

$= t_{0.975,2}$

$= 4.30$

Conclusion

Since the test-statistics falls within the acceptance region, we therefore accept H_o and reject H_I and then, conclude that the mean time is equal to 1.407.

Example 18

In a public opinion poll, 1000 randomly chosen electors were asked whether they would vote for Party "A" at the next election and 357 replied "Yes". The leader of the party believes that the true proportion is 0.4. test, at the 8% level, whether he is overestimating his support.

Solution

n 100, r=357,

$$\therefore \quad \hat{p} = \frac{r}{n}$$

$$= \frac{357}{100} = 0.357$$

The hypothesis Ho: p=0.4

H_1: P< 0.4

$$\therefore \text{ z} = \frac{\hat{p} - p}{\sqrt{\frac{pq}{n}}}$$

$$= \frac{0.357 - 0.4}{\sqrt{\frac{0.4x0.6}{1000}}}$$

$$= -2.78$$

8% = 0.08 $\therefore$ 0.5-0.08 = 0.42

Check this value under the body of your normal distribution table gives p- value = 1.453

Conclusion

The null hypothesis is rejected and the alternative hypothesis is accepted, then the proportion p< 0.4; therefore the leader of the party is over estimating his support.

3.3 Tests on the Differences between Two Means

If the researcher asks the question, "Is the mean amount of vitamin A for rats on the **deficient** diet the same as the mean amount for rats on the **normal** diet?" He then decides to take 20 rats and divides them at random into two groups: to one group he gives the normal diet and to the other group he gives the deficient diet. He obtains two sets of data, each consisting of 10 amounts of vitamin A, and computes their mean $\bar{x}_1$ and $\bar{x}_2$ for the normal and deficient diet-group. He now calculates the difference between the computed means i.e. $\bar{x}_1 - \bar{x}_2$

The problem is to decide whether their difference is large enough to convince him that there really is a difference between amounts of vitamin A under the two diets, or whether the difference might have been caused by sampling variation.

The form of the hypothesis test for comparing the means of two population or difference of means depends on whether or not the population variances are known or not.

3.4 Population Variance Known

The hypothesis is:

H_o: $\mu_1 = \mu_2$ or H_o: $\mu_1 - \mu_2 = 0$
H_o: $\mu_1 < \mu_2$ or H_o: $\mu_1 - \mu_2 < 0$
H_o: $\mu_1 > \mu_2$ or H_o: $\mu_1 - \mu_2 > 0$
H_o: $\mu_1 \neq \mu_2$ or H_o: $\mu_1 - \mu_2 \neq 0$

Let X and Y be the two random variables with mean μ_x, μ_y and variances σ_x^2, σ_y^2 respectively. The random variance $\overline{X} - \overline{Y}$ has normal distribution with mean $\mu_x - \mu_y$ and variance

$$\frac{\sigma_x^2}{n_x} + \frac{\sigma_x^2}{n_y}$$

Hence, the test statistics T.S. = Z = $$\frac{\overline{X} - \overline{Y} - (\mu_x - \mu_y)}{\sqrt{\frac{\sigma_x^2}{n_x} + \frac{\sigma_y^2}{n_y}}}$$

Then, if $\mu_x - \mu_y = 0$

$$\therefore \quad \text{T.S.} = \frac{\bar{x} - \bar{y}}{\sqrt{\frac{\sigma_x^2}{n_x} + \frac{\sigma_y^2}{n_y}}}$$

3.5 Population Variance is common but unknown

(i) Large Sample Sizes

The Test Statistics T.S

$$Z = \frac{\bar{x} - \bar{y}}{\sqrt{S_p^2\left(\frac{1}{n_x} + \frac{1}{n_y}\right)}}$$

If the assumption that $\sigma_x^2 = \sigma_y^2$ cannot be made

Then the S = z = $\frac{\bar{x} - \bar{y}}{\sqrt{\frac{s_x^2}{} + \frac{s_y^2}{n_y}}}$

Where s_p^2 is the pooled estimate of the common variance

$$s_p^2 = \frac{(n_x - 1) + (n_y - 1)}{n_x + n_y - 2} or s_p^2 = \frac{n_x\sigma_x^2 + n_y\sigma_y^2}{n_x + n_y - 2}$$

$$s_x^2 = \frac{\sum x^2 - \frac{(\sum x)^2}{n_x}}{n_x - 1}, s_y^2 = \frac{\sum y^2 - \frac{(\sum y)^2}{n_y}}{n_y - 1}$$

(ii) Small Sample Sizes

Here, the Test Statics T.S. is given by

$$= \frac{\bar{x} - \bar{y}}{\sqrt{s_p^2\left(\frac{1}{n_x} + \frac{1}{n_y}\right)}}$$

And it has a t-distribution n_x+n_y-2 degree of freedom and s_p^2 is the pooled estimate of variance.

Example 19

A random sample of 85 observations is taken from a population with standard deviation 10.2 and the sample mean is 31.2. A random sample of 72 observations is taken from a second population with standard

deviation 15.8 and the sample mean is 35.5. Test, at the 1% level, whether the second population has a greater mean than the first.

Solution

Here n is large i.e. $n_1 = 85$, the standard deviation are known

$\therefore$ $n_1 = 85$, $\sigma_1 = 10.2$, $\overline{x_1} = 31.2$

$\therefore$ $n_2 = 72$, $\sigma_2 = 15.8$, $\overline{x_2} = 35..5$, $\alpha = 1\%$

The hypothesis is

Ho: $\mu_2 - \mu_1 = 0$

H1: $\mu_2 - \mu_1 > 0$

$\therefore$ T.S = z = $$\frac{\overline{x_2} - \overline{x_1} - (\mu_2 - \mu_1)}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}}$$

$$= \frac{35.5 - 31.2 - 0}{\frac{10.2^2}{85} + \frac{15.8^2}{72}}$$

= 1.985.

1% = 0.01 $\therefore$ 0.5 – 0.01 = 0.49

Check 0.49 under normal distribution table gives
p-value = 2.33

Conclusion

Since the T.S. is less than the p-value, then, we accept H_0 and reject H_1 and then conclude that the second population has the same mean as the first i.e. the second population does not have a greater mean than the first.

Example 20

A machine assesses the life of a Bic biro, by measuring the length of a continuous line drawn using the pen. A random sample of 80 pens of brand A have a total writing length of 96.84km. A random sample of 75 pens of brand B has a total writing length of 93. 75km. Assuming that the standard deviation of the writing length of a single pen is 0.15km for both brands, test at the 5% level, whether the writing lengths of the two brands differ significantly.

Solution

Here, the sample is large and the variances are known.

Brand A: $n_1 = 80$, $\sum x_1 = 96.84$ $\therefore$ $\bar{x}_1 = 1.2105$, $\sigma_1 = 0.15$km

Brand B: $n_2 = 75$, $\sum x_2 = 93.75$ $\therefore$ $\bar{x}_2 = 1.25$, $\sigma_2 = 0.15$km

The hypothesis is

H_o: $\mu_2 - \mu_1 = 0$

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

Then, the T.S. = z = $$\frac{\bar{x}_2 - \bar{x}_1 - (\mu_2 - \mu_1)}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}}$$

$$= \frac{1.2105 - 1.25}{\sqrt{\frac{0.15^2}{80} + \frac{0.15^2}{72}}}$$

$$= -1.62$$

Also, the p-value = 1.96

Conclusion

The T.S. falls within the acceptance region, then we accept H_0 and reject H_1 and then conclude that the writing lengths of the two brands do not differ.

Example 21

Christmas trees are sown in seed compost A and, after 6 years the resulting 105 Christmas trees have mean height of 0.0641m, with the corresponding value of s^2 being $0.0453m^2$. Christmas trees are also sown in seed compost B and grown in similar circumstances. After 6 years the 97 trees have mean height 0.578m, with the corresponding value of s^2 being $0.0712m^2$. Test whether there is significant evidence, at the 5% level, that taller trees are produced in seed compost A.

- Without assuming that the population variances are equal
- Assuming that the population variances are equal

Solution

Compost A: $n_1 = 105$, $\overline{x_1} = 0.641$, $s_1^2 = 0.0453$
Compost B: $n_2 = 97$, $\overline{x_2} = 0.578$, $s_2^2 = 0.0712$

The hypothesis is:

$H_0: \mu_1 - \mu_2 = 0$
$H_1: \mu_1 - \mu_2 > 0$

Here, the sample is large and the variances are known. If these variances are not equal then the test statistics is

(i) $$T.S = z = \frac{\bar{x}_1 - \bar{x}_2 - (\mu_1 - \mu_2)}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}}$$

$$= \frac{0.641 - 0.578 - 0}{\sqrt{\frac{0.0453^2}{105} + \frac{0.0712^2}{92}}}$$

$= +7.29$

p-value = 1.645 i.e. at 5% level of significance

Conclusion

Since the T.S. is greater than the p-value, we reject the null hypothesis H_o and accept the alternative hypothesis H_1 and then conclude that there is significant evidence that taller trees are produced in seed compost A.

(ii) Assuming the population variances are equal
Then, the pooled estimate of variances is given as:

$$s^2 = \frac{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2}{n_1 + n_2 - 2}$$

$$= \frac{(105 - 1)(0.0453) + (97 - 1)(0.0712)}{105 + 92 - 2}$$

$= 0.0592$

Then, the test statistics:

$$z = \frac{\bar{x} - \overline{x_2} - (\mu_1 - \mu_2)}{\sqrt{s^2\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}}$$

$$= \frac{0.641 - 0.578 - 0}{\sqrt{0.0592\left(\frac{1}{105} + \frac{1}{92}\right)}}$$

$$= \frac{0.063}{0.0347}$$

The p-value = 1.645

Conclusion

Since the T.S. is greater than the p-value, we reject the null hypothesis H_o and accept the alternative hypothesis H_1 and then conclude that there is significant evidence that taller trees are produced in seed compost A

Example 22

The table below shows the amounts of vitamin A in livers of rats under normal diet and vitamin E deficient diet.

Rat Number	1	2	3	4	5	6	7	8	9	10
Normal Diet	3950	3800	3450	3400	3700	3900	3800	3050	2700	2000
Deficient Diet	2650	3350	2450	2650	2650	3150	2900	1700	1700	2500

Test the hypothesis at 50% level of significance whether the data provides evidence that there is a difference between the mean amount of vitamin A for rats on the deficient diet and those on the normal diet.

Solution

The hypothesis is: H_o: $\mu_1 - \mu_2 = 0$; H_1: $\mu_1 - \mu_2 \neq 0$

Normal Diet: $n_1 = 10; \sum x_1 = 33{,}750; \sum x_1^2 = 117, 437{,}500$

Deficient Diet: $n_2 = 10; \sum x_2 = 25{,}700; \sum x_2^2 = 68{,}655{,}000$

$\therefore \overline{x_1} = 3{,}375; \overline{x_2} = 2{,}570; s_1^2 = 392{,}361; s_2^2 = 298{,}556$

$$s_1^2 = \frac{\sum x_1^2 - \frac{(\sum x_1)^2}{n_1}}{n_1 - 1} = \frac{117437500 - \frac{(33750)^2}{10}}{10 - 1} = 392{,}361$$

$$s_2^2 = \frac{\sum x_2^2 - \frac{(\sum x_2)^2}{n_2}}{n_2 - 1} = \frac{68655000 - \frac{(25700)^2}{10}}{10 - 1} = 289{,}556$$

The pooled estimate of variance is

$$s^2 = \frac{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2}{n_1 + n_2 - 2}$$
$$= \frac{(10-1)392361 + (10-1)289556}{10+10-2}$$
$$= 340{,}958$$

Since the variance is unknown and the sample is small then, the test-statistics will be chosen from t-distribution

$$\therefore \text{T.S} = \frac{\bar{x} - \bar{x_2} - (\mu_1 - \mu_2)}{\sqrt{s^2\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}}$$
$$= \frac{3375 - 2570 - 0}{\sqrt{340958\left(\frac{1}{10} + \frac{1}{10}\right)}} = \frac{805}{261.27}$$
$$= 3.08$$

p-value = $t_{5\%, n_1 + n_2 - 2}$
$= t_{0.975,10+10-2}$
$= t_{0.975,18} = 2.10$

Conclusion

Since the T.S. falls within the rejection region, we reject H_0 and accept H_1 and then, conclude that there is a difference between the mean amount of vitamin A for rats on deficient diet and those on the normal diet.

Example 24

In a supermarket, the daily sales of packets of ‘crispo’ cereal have mean 124.5 and variance 129.96. Following a television advertisement, the mean daily sales over a period of 12 days increased to 132.5. State any necessary assumptions and test at the 1% significance level whether the mean number of packets sold per day has increased.

Solution

Before Advert: $\bar{x}_1 = 124.5$; $\sigma_1^2 = 129.96$; $n_1 = 12$
After Advert: $\bar{x}_2 = 132.5$; $\sigma_2^2 = 129.96$; n2 = 12 ; $\alpha = 1\%$

The assumption is that the variance after advertisement is the same as the variance before advertisement. Hence, the variances are known and the sample is small.

$$\text{The T.S.} = z = \frac{\bar{x}_1 = \bar{x}_2 - (\mu_1 - \mu_2)}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}}$$

$$= \frac{124.5 - 132.5 - 0}{\sqrt{\frac{129.96}{12} + \frac{129.96}{12}}}$$

$$= -\frac{8}{4.654} = -1.72$$

The hypotheses are

H_0: $\mu_1 - \mu_2 = 0$

H1: $\mu_1 - \mu_2 < 0$ i.e. $\mu_2 > \mu_1$

But 1% = 0.01 $\therefore$ 0.5-0.01 = 0.49

Checking this under normal table gives
p-value = -2.33

Conclusion

Since the T.S. falls within the acceptance region, then, accept H_0 and reject H_1 and conclude that the mean number of packets sold per day has not increased.

SELF ASSESSMENT EXERCISE

(1) Describe the meaning of Type 1 error and Type II error.

(2) A sample of 40 observations is selected from one population. The sample mean is 102 and the sample deviation is 5. A sample of 50 observations is selected from a second population. The sample mean is 99 and the standard deviation is 6.Conduct the following test of hypothesis using the 5% significance level.

$$H_o : \mu_1 = \mu_2$$

$$H_i : \mu_1 \neq \mu_2$$

(i) Is this a one tail test or two tail test?

(ii) State the decision rule.

(iii) Compute the value of the test statistic.

What is your decision regarding H_0?

4.0 CONCLUSION

Hypothesis testing is a major area of statistical inference that is highly useful in conveying information about a population based on a sample. The hypothesis can be compared to a level of significance. If the p-value is smaller than α, the level of significance, the test results are considered to be significant and the null hypothesis is rejected; if larger, the null hypothesis is not rejected.

The rejection of a true null hypothesis is a Type I error. Not rejecting a false null hypothesis is a Type II error.

5.0 SUMMARY

In this unit we have learnt the followings.

- Hypothesis is an idea that is based on known fact.
- There are two types of hypothesis: null and alternate hypothesis
- The basic steps in hypothesis test are: statement of hypotheses, level of significance, test statistics, decision and conclusion.

6.0 TUTOR-MARKED ASSIGNMENT

1. Given the following hypothesis

$$H_o : \mu = 400$$
$$H_i : \mu \neq 400$$

2. For a random sample of 12 observations, the sample mean was 407 and the sample standard deviation 6. Using the 0.01 significant level:

a. state the decision rule
b. compute the value of the test statistic.

3. What is your decision regarding the null hypothesis?

7.0 REFERENCES/FURTHER READING

Donald, R. P. & Edward, B. O. *Business and Economics Statistics*, (Revised edition). Business Publications Inc.

Lehman, E. L. (1959). *Testing Statistical Hypothesis*. New York: John Wiley & Sons Inc.

MODULE 3 INDEXES NUMBERS

UNIT 1 BASIC UNDERSTANDING OF INDEX NUMBER

CONTENTS

1.0 INTRODUCTION

Index number shows how figures are compared by using ratios. One year is chosen as the reference year and scaled to 100. All other figures are scaled so that the ratio of the index number to 100 is the same as the ratio of that original value to the value in the year under reference.

2.0 OBJECTIVES

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

- explain an index number
- describe several uses of price index and the meaning of consumer price index
- construct indexes (price, value e.t.c.).

3.0 MAIN CONTENT

3.1 What Is An Index Number?

Index numbers are descriptive statistics. We are all aware that prices, for example, vary from time to time and from place to place. A price index number is useful to describe the way in which prices vary from one time to another or from one place to another. There are many other forms of index numbers besides price indexes. Industrial output, farm output, unemployment, employment, and highway accidents vary from time to time and from place to place. Any of these can be measured and described with an appropriate index number.

Index numbers are used in many situations. A firm choosing a new plant location may be interested in the price of consumer goods, which might be an important factor in determining the wage rates necessary for a given location. An organisation interested in promoting highway safety might be interested in an index of highway accidents to determine where a safety campaign should be instituted. A person seeking a retirement home might be interested in the price of consumer goods in various places to aid in determining where to live after retirement.

There are many other uses for index numbers.

3.2 Simple Price Index

A price index (*plural*: "price indices" or "price indexes") is a normalised average (typically a *weighted* average) of prices for a given class of goods or services in a given region, during a given interval of time. It is a statistical design which helps to compare how these prices, taken as a whole, differ between time periods or geographical locations.

Price indices have several potential uses. For particularly broad indices, the index can be said to measure the economy's price level or a cost of living. More narrow price indices can help producers with business plans and pricing. Sometimes, they can be useful in helping to guide investment.

A number of different formulas, at least hundreds, have been proposed as means of calculating price indexes. While price index formulas all use price and quantity data, they amalgamate this data in different ways. A price index generally aggregates using various combinations of base period prices (p_0), later period prices (p_t), base period quantities (q_0), and later period quantities (q_t). Price index formulas can be framed as comparing expenditures (expenditure is a price multiplied by a quantity) or taking a weighted average of price relatives (p_t / p_0).

3.3 Construction of Index Numbers

Now that a price index has been constructed, it may be helpful to state in symbols what has already been accomplished in the preceding section. Using the symbol *p* to stand for the price of a commodity, the subscript *n* to represent the year under consideration (called the *given* year), the subscript 0 to indicate the base year, and the symbol *q* to stand for the quantity of a commodity bought or sold, we can restate the index number calculations.

To combine these simple price indexes into a price index using the base year value of transactions as weights, the index is constructed as:

$$\frac{\sum \frac{p_n}{p_O}(p_O q_o)}{\sum p_O q_o} X100$$

Where the summation sign indicates a summation over all commodities included in the index. (For historical purposes, this particular form of price index is called the *Laspeyres price index,* named after the person influential in its development).

If some typical market basket is selected for weighting the simple price indexes, the index number is computed as:

$$\frac{\sum \frac{p_n}{p_O}(p_O q_t)}{\sum p_O q_t} X100$$

Where the subscript *t* denotes the typical quantity bought or sold for that commodity.

3.4 Value, Quantity and Geographical Indexes

There are several other useful indexes. For example, consider a retail store that sells two products: cross-country motorcycles and bicycles. The store is interested in looking at the changes in its volume of sales over the past several years. If prices have changed over this time period, a very important question is whether the firm is interested in looking at changes in the naira value of sales, or in changes in the number of physical units sold. The first question, the change in naira value of sales, is addressed by constructing a *value index.* The second question, changes in physical volume of sales, is addressed by constructing a *quantity index.* The firm's records yielded the information shown in Table 1.

Table 1: Sales of Motorcycles and Bicycles

	Bicycles			Motorcycles			
Year	Number sold	selling price (N)	sales revenue (N)	Number sold	selling price (N)	sales revenue(N)	Total Sales(N)
1976	300	100	30,000	50	20	1,000	31,000
1977	320	110	35,200	80	25	2,000	37,200
1978	260	90	23,400	130	35	4,550	27,950
1979	200	105	21,000	140	30	4,200	25,250
1980	270	120	32,400	120	35	4,200	36,600

3.5 Value Index

A value index is nothing more than the ratio of the naira volume of sales in the given year to the naira volume of sales in the base year (with the usual factor of 100, arising because index numbers are implicitly stated as percentages). For the data given inTable1 the value index is computed in Table 2. From Table 2 it is apparent that 1978 sales, for example, were 90.2 percent of 1976 sales.

Table 2: Value Index for a Motorcycle and Bicycle Shop (1976 = 100)

1976	31,00/3131,000x100 = 100.0
1977	37,200/31,000x100= 120.0
1978	27,950/31,000x100=90.2
1979	25,200/31,000x100=81.3
1980	36,600/31,000x100=118.1

In equation form, the value index is computed as

$$\frac{\sum p_n q_n}{\sum p_O q_o} x100$$

It is important to note that a value index combines changes in both prices and quantity of transactions into one figure.

3.6 Quantity Index

If the owner of the bicycle and motorcycle shop is interested in describing changes in the physical volume of sales, apart from the naira volume of sales, a quantity index should be computed. The first step in computing this index is to compute the simple index of quantities for

each of the individual commodities, bicycles and motorcycles. These simple indexes are computed in Table 3.

To combine these two simple quantity indexes into one *quantity index,* it is necessary to weight each commodity by its importance. For this problem, the store owner (or the statistical analyst) must make a decision about the appropriate importance of bicycles and motorcycles.

Table 3: Simple Quantity Indexes: Bicycles and Motorcycle (1976 = 100)

Year	Bicycles	Motorcycles
1976	300/300x100 = 100.0	50/50 x 100 = 100.0
1977	320/300x100 = 106.7	80/50 x 100 = 160.0
1978	260/300x100 = 86.7	130/50 x 100 = 260.0
1979	200/300x100 = 66.7	140/50 x 100 = 280.0
1980	270/300x100 = 90.0	120/50 x 100 = 240.0

Table 4: Quantity index for bicycle –motorcycle shop, based on judgmental weight (1976 = 100)

Year	Bicycles		Motorcycles		index
1976	0.9 x 100.0	+	0.1 x 100.0	=	100.0
1977	0.9 x 106.7	+	0.1 x 100.0	=	112.0
1978	0.9 x 86.7	+	0.1 x 260.0	=	104.0
1979	0.9 x 66.7	+	0.1 x 280.0	=	88.0
1980	0.9 x 90.0	+	0.1 x 240.0	=	105.0

1976, motorcycles contributed only about 3 percent of naira sales, while in 1979 motorcycle amounted to nearly 17 percent of naira sales. If the owner of the shop felt that typical sales figures for the store were 90 percent bicycles and 10 percent motorcycles, these weights could be used in constructing the quantity index. Of course, it is possible to use any other set of weights, such as 1976 values, 1980 values, or some average of values for various years. Using the store owner's judgmentally determined 90 percent versus 10 percent weights, the quantity index is shown in Table 4.

From this table, it is apparent that there has been a substantial decline in physical volume of sales from 1977 to 1978 and 1979. It is also apparent that the value index of 118.1 for 1980 does not represent changes in physical volume alone. Although the naira volume in 1980 was 118.1 percent of the naira volume in 1976 (Table 2), the physical

quantity of merchandise passing through the doors increased only to an index of 105.0.

The formula for computing a quantity index may be expressed as:

$$\sum \frac{q_n}{q_o} w \, X \, 100$$

Where the symbol *w* denotes the weight appropriate for a particular commodity, and the sum of the weights is one. If base year naira values are used as weights, the quantity index (a Laspeyres quantity index) is

$$\frac{\sum \frac{q_n}{q_O}(p_O q_o)}{\sum p_O q_o} X 100$$

Values for other years could be used as weights.

3.7 Geographical Price Index

Price indexes discussed so far reflect price changes over the time at the same location. Prices may also vary from place to place at the same time. For simplicity, let us consider price indexes for two cities, A and B, for the year 1981. This simplified example will assume only three commodities: transportation, housing and food. The example is further simplified by assuming that only one price exists for each commodity in each city. The raw data are shown in Table 5.

It is apparent by now that the construction of an index number requires weights. Suppose consumer expenditure studies in each city reveal the proportion of expenditures for each commodity in each city, as shown in Table 6. Suppose a soon-to-retire worker is interested in whether prices are higher in A or B. However, the worker needs to answer the question, "The price of what?" if the worker is interested.

Table 5: Data for Price Index in two Cities (price per unit)

City	Transportation	Housing	food
A	N58,9	N195	N1.89
B	N51.9	N235	N1.92

Table 6: Proportion of expenditure for each commodity

City	transportation	housing	food
A	0.30	0.45	0.25
B	0.35	0.25	0.40

In the price in city B of the "market basket" of city A, the expenditure proportions for city A are used to weight the price of city B relative to city A:

$$\text{Price index in city B for the market basket of city A} = \frac{\sum \frac{p_B}{p_A} w_A}{\sum w_A} X100$$

Where W_A is the relative importance of a commodity in city A, A similar definition, with A and B reversed, gives the price index in city A for the market basket of city B. These indexes are calculated below.

Price index in city B for market basket of city A:

$$\left[\left(\frac{51.9}{58.9}x0.30\right)+\left(\frac{235}{195}x0.45\right)+\left(\frac{1.92}{1.89}x0.25\right)\right]x100 = 106.1$$

Price index in city A for market basket of city B:

$$\left[\left(\frac{58.9}{51.9}x0.35\right)+\left(\frac{195}{235}x0.25\right)+\left(\frac{1.89}{1.92}x0.40\right)\right]x100 = 99.8$$

These calculations indicate that prices in city B are higher than prices in city A. More specifically, a resident of city A, who went to city B and purchased the "A" market basket, would pay 106.1 percent as much as for the same items in city A. However, a resident of city B who went to city A and purchased the "B" market basket would pay 99.8 percent as much as for the same items in city B. This seems to be saying that city B prices are about 6 percent higher than city A prices, but city A prices are about the same as city B prices! The apparent contradiction arises because the market basket differs from city to city.

This differing market basket makes geographical price indexes particularly difficult to construct. A commodity that is very important in one city, such as home heating fuel in Fairbanks, is less important in another city.

SELF ASSESSMENT EXERCISE

1a) What is index number?
b) Why is it necessary periodically to change commodities and weight for the index?
2. What are the potential uses of price index?

4.0 CONCLUSION

Index numbers are very essential statistical tools to measure economic problems, and its construction revolves around the art and common sense of the use of number.

5.0 SUMMARY

In this unit we have learnt that:

- index numbers are the measurement of economy in terms of price, values, quantity and geographic index. We also learnt their construction such as:

- Price index= $\frac{\sum \frac{p_n}{p_O}(p_O q_o)}{\sum p_O q_o} X100$

- Value index= $\frac{\sum p_n q_n}{\sum p_O q_o} x100$

- Quantity index= $\frac{\sum \frac{q_n}{q_O}(p_O q_o)}{\sum p_O q_o} X100$

6.0 TUTOR-MARKED ASSIGNMENT

1. The price index for 1975 stood at 289.2. Does that necessarily mean there had been a recent price increase? Discuss.
2. The wholesale price of 40 pound box of red delicious apples were given below for six years

Years	Price
1970	49
1971	78
1972	56

1973	65
1974	79
1975	56

Construct a price index for these apples, 1970=100

7.0 REFERENCES/FURTHER READING

Chance, W. A. (1966). "A Note on the Origins of Index Numbers". *The Review of Economics and Statistics,* Vol. 48, No. 1. Feb., pp. 108-10.

Diewert, W. E. (1993). "Index Numbers" in *Essays in Index Number Theory*. Ed. W. E. Diewert and A. O. Nakamura. Vol 1. Elsevier Science Publishers.

McCulloch, J. H. (1982). *Money and Inflation: A Monetarist Approach* 2e, Harcourt Brace Jovanovich / Academic Press.

Triplett, J. E. (1992). "Economic Theory and BEA's Alternative Quantity and Price Indices". *Survey of Current Business*.

UNIT 2 CONSUMER PRICE INDEX (CPI) AND OTHER FORMS OF INDEX NUMBERS

CONTENTS

1.0 INTRODUCTION

Price index is one of the important microeconomics variables, such as Consumer Price Index, GDP Deflator, and Producer Price Index

2.0 OBJECTIVES

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

- explain CPI and other microeconomics variables
- state the important uses of CPI from economist and statistical standpoint
- outline the uses of other microeconomics variables
- explain the effect of quality changes on CPI
- apply CPI to deflate a time series.

3.0 MAIN CONTENT

3.1 Important Price Measures

Many economic research organisations publish statistical data about many aspects of our economy. Many of these data are expressed as index

numbers. From a personal standpoint, the index number that directly affects most of us is the Consumer Price Index, which is often loosely called the “cost of living” index.

3.2 Consumer Price Index

“The Consumer Price Index (CPI) is a statistical measure of changes, over a time, in the price of goods and services . . . typically purchased by urban consumers.” The index is broken down into major expenditure groups, such as food and beverages, housing, apparel and upkeep, transportation, and others. These major expenditure groups are then weighted to form the overall index. The Consumer Price Index, in a form roughly comparable to current data, has been published since 1921. Until 1978, the CPI measured price changes for urban wage earners and clerical workers. Beginning in 1978, there have been two versions of the CPI. The CPI-W measures price changes for urban wage earners and clerical workers, consistent with decades of CPI data. The second version, CPI-U, measures price changes for all urban consumers.

The CPI-U is based on buying habits of about 80 percent of the civilian population of the United States. The primary groups of consumers included in the CPI-U, but excluded in CPI-W, are the self-employed, professional, white-collar, salaried workers, retired persons, and the unemployed. Because of these differences, the CPI-W covers only about half of the consumers covered by the CPI-U.

From a statistical standpoint, one of the most important uses of the CPI is to deflate another economic time series. From an economist's point of view, an extremely important use of the CPI is to measure the effect of economic policy on price levels.

3.3 Quality Changes in the CPI

One aspect that the index number should incorporate, if possible, is changes in the quality of goods included in the index. Unfortunately, quality is a very difficult aspect to measure. Are 1977 automobiles "better" than 1967 automobiles? How are quality changes in gas mileage and safety features measured? The index statistician needs to be aware of such changes in product composition while the index is being compiled. The Federal Bureau of Statistics (FBS) explains its procedures: One of the most difficult conceptual problems faced in compiling a price index is accurate measurement and treatment of quality change because products and consumption patterns are constantly changing. For example, with each model change of an automobile, the FBS faces the problem of separating the price rise from the increase in price due to quality change.

Quality change in a new model of an item should not be reflected as a price change, since the index measures the cost to consumers of purchasing a constant market basket of goods and services of constant quality through time. Ideally, estimates would be obtained for each naira value quality change resulting from a change in the model or item priced. This estimate would reflect how much consumers value the quality change.

However, this direct measuring of quality change is extremely difficult since measurement of the value consumers place on quality change is rarely possible. Therefore, to adjust for quality change, FBS uses an indirect method to measure the quality change by evaluating the additional cost associated with producing the change in quality. For new automobile features, this estimate is based on all costs incurred in manufacturing plus the established company mark-up to the selling price of passenger cars. This estimate of costs applies to all new features that are installed as standard equipment, that is, features on cars in the same or comparable series. For all items that replace or modify some previously existing features, the estimate is based on the difference in cost between the old and the new features. In other words, the estimate of cost for new items is computed for both the new and the old features. The difference between these values is used as the estimate of quality change.

Adjustments for quality change in the CPI "new car" index include structural and engineering changes that affect safety, environment, reliability, performance, durability, economy, carrying capacity, manoeuvrability, comfort, and convenience. Although, antipollution equipment on automobiles originally did not increase quality because the utility to the purchaser is difficult to determine, these devices do improve quality for consumers in general and therefore an increase in physical quality for the individual consumer. Consequently, quality adjustments are made for pollution controls to automobiles.

Quality adjustments exclude changes in style or appearance, such as chrome trim, unless these features have been offered as options and purchased by a large proportion of customers. Also, new technology sometimes results in better quality at reduced or no increase in cost. When no satisfactory value has been developed for such a change, it is ignored, and prices are compared directly.

3.4 Applying the CPI to Deflate a Time Series

Consider Table 1 on Mercantile Stores example below. These data were reported in constant (1967) naira with a brief word description regarding the deflation of the actual sales to constant naira sales. Selecting a

small series of quarterly sales figures for an example, the deflation of a part of this time series is demonstrated in Table 1. The Apparel and Upkeep component of the CPI was used for the price index. This is not a perfect choice, because Mercantile does not deal in upkeep (dry cleaning, etc.). It also omits household furnishings (bedding, etc.) which are a part of mercantile business. Each index value used for deflation is the average of the index for the three months in that quarter.

The reader should note that sales in Column 1 are affected by changes in quantities *and* prices, whereas sales in Column 3 reflect changes in quantity only. The sales in Column 1 are measures of values (i.e. prices X quantities), whereas sales in Column 3 reflect quantities, but in naira units. In general, sales in *current* naira (as in Column 1) constitute a time series of *values,* and sales in constant naira (as in Column 3) constitute a time series reflecting *physical sales,* not naira sales.

Table 1: Mercantile Stores Quarterly Sales and Apparel and Upkeep Price Index

Year	Quarter	1	2	3
		Reported sales N (current naira millions)	price index (1967 =100)	=1/2x100 Sales (1967 naira; N millions)
1972	1	N87.4	121.3	N72.1
	2	N100.0	121.9	N82.0
	3	N112.4	122.7	N91.6
	4	N158.5	124.3	N127.5
1973	1	N102.6	124.7	N82.3
	2	N116.3	126.4	N92.0
	3	N123.9	128.1	N96.7
	4	175.3	129.9	N134.9

3.5 Producer Price Indexes

Prices paid by consumers are measured by the Consumer Price Index. Before it is purchased by a consumer, the price of a consumer item is measured by various Producer Price Indexes. As an example, follow wheat through its production processes to flour, bread, and then consumption. The price of wheat is included in the “Producer Price Index for crude materials for further processing.” After the wheat has been processed into flour, its price is included in the “Producer Price Index for intermediate goods.” After the flour has been processed into bread, its price is included in the "Producer Price Index for finished goods.” When the bread is purchased by a consumer, its price is in-

cluded in the CPI. The Producer Price Indexes are published monthly, by Federal Bureau of Statistics. They "measure average changes in prices received in primary markets of the United States by producers of commodities in all stages of processing." The FBS emphasises three Producer Price Indexes, one for each of the three stages of production. Producer Price Indexes are also available for commodity groupings, such as farm products, rubber and plastic products, and many other groupings. With UK Producer Price Indexes, a manufacturer can compare changes in the prices paid or received with the national average price changes for similar commodities. For example, there is a separate index available for the intermediate material "prepared paint." This index shows the prices (relative to the base year) for prepared paint at the first important commercial transaction after manufacture, the price paid to a paint manufacturer. A manufacturer of another commodity, such as household furniture, might be interested in comparing its prices (relative to the same base year) for prepared paint with the national index. Thus, the prepared paint represents a selling price for the paint manufacturing industry and a purchase price for the household furniture industry.

In 1978 the FBS changed the name of the old Wholesale Price Index to the Producer Price Index for all commodities. In using the various Producer Price Indexes, data are available for many past years. However, historical publications use the term Wholesale Price Index instead of the current title, Producer Price Index.

3.6 GDP Deflator

The GDP deflator can be viewed as a measure of general inflation in the domestic economy. Inflation can be described as a measure of price changes over the time. The deflator is usually expressed in terms of an index, i.e. a time series of index numbers. Percentage changes on the previous year are also shown. The GDP deflator reflects movements of hundreds of separate deflators for the individual expenditure components of GDP. These components include expenditure on such items as bread, investment in computers, imports of aircraft and exports of consultancy services.

3.6.1 Uses of the GDP Deflator Series

The GDP deflator allows for the effects of changes in price (inflation) to be removed from a time series, i.e. it allows the change in the volume of goods and services to be measured. The resultant series can be used to express a given time series or data set in real terms, i.e. by removing price changes.

3.6.2 Measurement in National Accounts

In most systems of national accounts the GDP deflator measures, the ratio of nominal (or current-price) GDP to the real (or chain volume) measure of GDP. The formula used to calculate the deflator is:

$$\text{GDP deflator} = \frac{\text{Nominal GDP}}{\text{Real GDP}} \times 100$$

Dividing the nominal GDP by the GDP deflator and multiplying it by 100 would then give the figure for real GDP, hence deflating the nominal GDP into a real measure.

It is often useful to consider implicit price deflators for certain subcategories of GDP, such as computer hardware. In this case, it is useful to think of the price deflator as the ratio of the current-year price of a good to its price in some base year. The price in the base year is normalised to 100. For example, for computer hardware, we could define a "unit" to be a computer with a specific level of processing power, memory, and hard drive space and so on. A price deflator of 200 means that the current-year price of this computing power, is twice its base-year price - price inflation. A price deflator of 50 means that the current-year price is half the base year price - price deflation.

Unlike some price indexes, the GDP deflator is not based on a fixed basket of goods and services. The basket is allowed to change with people's consumption and investment pattern. (Specifically, for GDP, the "basket" in each year is the set of all goods that were produced domestically, weighted by the market value of the total consumption of each good.) Therefore, new expenditure patterns are allowed to show up in the deflator as people respond to changing prices. The advantage of this approach is that the GDP deflator reflects up to date expenditure patterns. For instance, if the price of chicken increases relative to the price of beef, people would likely spend more money on beef as a substitute for chicken. A fixed market basket measurement would miss this change.

In practice, the difference between the deflator and a price index like the Consumer Price Index (CPI) is often relatively small. On the other hand, with governments in developed countries increasingly utilising price indexes for everything from fiscal and monetary planning to payments to social programme recipients, the even small differences between inflation measures can shift budget revenues and expenses by millions or billions of naira.

3.7 Indexes of Quantity Output

One of the more important indexes of industrial output for the entire economy is the Federal Reserve Board index of quantity output, more commonly called the index of industrial production.

Table 2: GNP in Current and Constant Naira, and Implicit Deflator

	1	2	3
Year	GNP , Current Naira (billion)	GNP, constant Naira (billion)	Implicit Deflator (1) ÷ (2) x 100 (1972 =100)
1960	N506.0	N736.8	N68.7
1961	N523.3	N755.3	N69.3
1962	N563.8	N799.1	N70.6
1963	N594.7	N830.7	N72.7
1964	N635.7	N874.4	N74.3
1965	N688.1	N925.9	N76.8
1966	N753.0	N981.0	N79.0
1967	N796.3	N1,007.7	N82.6
1968	N868.5	N1,051.8	N86.7
1969	N935.8	N1,078.8	N91.4
1970	N982.4	N1,075.3	N96.0
1971	N1,063.4	N1,107.5	N100.0
1972	N1,171.1	N1,171.1	N105.8
1973	N1,306.3	N1,235.0	N116.0
1974	N1,412.9	N1,217.8	N105.8
1975	N1,528.8	N1,202.3	N116.0
1976	N1,702.2	N1,273.0	N127.2
1977	N1,899.6	N1,340.5	N133.7
1978	N141.7	N1,399.2	N152.1

It measures changes in the physical volume or quantity of output of manufacturers, minerals, and electric and gas utilities. The index does not cover production on farms, in the construction industry, in transportation, or in the various trade and service industries. The index reflects output changes at all stages of production.

As with other indexes we have examined, the index of industrial production is broken into many components. For example, there are separate indexes available for automotive products, which are further broken into subcategories. Output indexes are also available for car-peting and furniture, appliances, TV, and furniture, primary metals,

fabricated metal products, machinery, transportation equipment, lumber and so on. With this kind of detailed information, a firm is able to compare its own output over time with the appropriate industry index of output, to determine how its own growth compares with the national average. From an investor's standpoint, the index also provides information about the relative growth of various product groups or industries. As are other indexes, the index of industrial production is revised periodically to reflect changes in the economy. For example, a revision in 1971 incorporated information to match the index series to bench marks obtained in the *Annual Survey of Manufacturers.* An earlier revision was completed in 1962.

Another useful index of output can be constructed from gross national product data. The GNP is "the market value of the output of goods and services produced by the nation's economy."[9] This series is one of the more widely used measures of the output of the entire economy- The time series of GNP can easily be expressed as an index, simply by dividing the value for each year by the value of GNP for the base year. This computation, showing 1967 as the base year, is shown in Table 3. This index is a value index, rather than a measure of the aggregate physical output of the economy.

Table 3: GNP as a Value Index and as a Quantity Index

Year	1	2	3	4
	GNP (Billion of current naira)	GNP value index (1) ÷796.3x100	GNP(billion of 1972 naira	GNP Quantity index (3)÷ 1007.7 x100
1960	N506.0	N63.54	N736.8	N73.12
1961	N523.3	N65.72	N755.3	N74.95
1962	N563.8	N70.80	N799.1	N79.30
1963	N594.7	N74.68	N830.7	N82.77
1964	N635.7	N79.83	N874.4	N86.44
1965	N688.1	N86.41	N925.9	N91.88
1966	N753.0	N94.56	N981.0	N97.35
1967	N796.3	N100.00	N1,007.7	N100.00
1968	N868.5	N109.07	N1,051.8	N104.38
1969	N935.8	N117.52	N1,078.8	N107.06
1970	N982.4	N123.37	N1,075.3	N106.71
1971	N1,063.4	N133.54	N1,105.5	N109.90
1972	N1,171.1	N147.07	N1,171.1	N116.22
1973	N1,306.3	N164.05	N1,235.0	N122.56
1974	N1,412.9	N177.43	N1,217.8	N120.85
1975	N1,528.8	N191.99	N1,202.3	N119.31
1976	N1,702.2	N213.76	N1.273.0	N126.33

1977	N1,899.6	N238.54	N1,340.5	N133.03
1978	N141.7	N267.19	N1,399.2	N138.85

As price levels change (inflation or deflation), the value index becomes a very poor measure of actual changes in output. GNP constant naira values are used to produce an output index. This computation is demonstrated in Table3.

Just as there were various price indexes to measure various price changes, we have not seen two quantity indexes to measure different kinds of output changes. The GNP in constant naira, expressed as an index, is a much broader-based output index than the Federal Reserve Board index of industrial production. Two important sectors of the economy covered by GNP, but omitted in the Federal Reserve Board index, are agriculture and services.

3.8 Productivity Indexes

Productivity is important to any economy. Labour productivity is really nothing more than output per labour-hour. If a measure of output is available, this output may be divided by the labour input (usually expressed in hours). This quotient, expressed as an index, is a measure of changes in labour productivity.

To construct a productivity index for a particular firm (or industry), it is necessary only to divide a measure of output for that industry (such as a quantity index) by the amount of input of the factor of production whose productivity is being computed. As an example, the owner of the bicycle-ski shop could easily compute the productivity of the labour force of the shop by dividing the output index from Table 4 by the number of labour hours used in each of the years.

Table 4: Labour productivity for the bicycle-ski shop (1976 = 100)

Year hours	**1**	**2**	**3**	***Labor productivity* index**
	output index	Labour Hour	(1) ÷ (2)	(3) ÷ 0.0200 x 100
1976	N100.0	N5,000	N0.0200	N100.0
1977	N112.0	N5,200	N0.0215	N107.5
1978	N104.0	N5,400	N0.0193	N95.5
1979	N88.0	N4,600	N0.0191	N95.5
1980	N105.0	N4,800	N0.0219	N109.5

Figure 4 also shows the index of labour productivity for the five years for which data are available. This index could then be compared with

published indexes of labour productivity to provide information useful to the manager of the firm.

SELF ASSESSMENT EXERCISE

1. What does it mean to deflate to time series to account for price change?
2. Differentiate between CPI and GDP deflator.

4.0 CONCLUSION

The index numbers are essential statistical and economic measures which are used to check general persistent rise and fall in prices of commodities of any economy at a particular time and location.

5.0 SUMMARY

In this unit we have learnt that:

- statistical point of view of CPI is used to deflate another economic time series while economic point of view of CPI is used to measure the effect of economy policy on price level.
- GDP is a measure of general price changes over the time in the domestic economy.

$$\text{i.e. GDP deflator} = \frac{\text{Nominal GDP}}{\text{Real GDP}} \times 100$$

6.0 TUTOR-MARKED ASSIGNMENT

1. What is the CPI?
2. What is the meaning of GDP deflator?

7.0 REFERENCES/FURTHER READING

Donald B.O. & Edward R. P. (1981). *Business and Economics Statistics.* Revised ed.

Doody,F. S. (1965). *Introduction to the Use of Economic Indicators.* New York: Random House, Inc.

Fisher, I. (1923). *The Making of Index Numbers.* New York: Houghton Mifflin Co.

Moore, G. H., & Shishkin, J. (1967). *Indicators of Business Expansions and Contractions.* New York: National Bureau of Economic Research, Columbia University Press.

Mudgett, B. D. (1951). *Index Numbers.* New York: John Wiley & Sons, Inc.

UNIT 3 APPLICATIONS OF INDEX NUMBERS TO ECONOMICS PROBLEMS

CONTENTS

1.0 INTRODUCTION

Two basic types of data are needed to construct the CPI: price data and weighting data. The price data are collected for a sample of goods and services from a sample of sales outlets in a sample of locations for a sample of times. The weighting data are estimates of the shares of the different types of expenditure as fractions of the total expenditure covered by the index. These weights are usually based upon expenditure data obtained for sampled decades from a sample of households.

The index is usually computed yearly, or quarterly in some countries, as a weighted average of sub-indices for different components of consumer expenditure, such as food, housing, clothing, each of which is in turn a weighted average of sub-sub-indices at the most detailed level, the elementary aggregate level.

The index reference period, usually called the base year, often differs both from the weight-reference period and the price reference period. This is just a matter of rescaling the whole time-series to make the value for the index reference-period equal to 100. Annually revised weights are a desirable but expensive feature of an index, for the older the

weights the greater is the divergence between the current expenditure pattern and that of the weight reference period.

2.0 OBJECTIVES

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

- explain index numbers
- describe the steps to be considered before the construction of index number
- explain how to deflate a series using a retail price index.

3.0 MAIN CONTENT

3.1 What are Index Numbers?

This can be defined as a statistical method of measuring a variable or a group of related variables with respect to its value at a particular period of time. This specified period of time is called the base period. Products, such as petroleum, cash crops etc. are examples of group related variables.

If the above mentioned variables are with respect to price, then it is called price index. Also, if they are with respect to quality or volume, they are called quantity index. Although, a typical index number may be concerned with whole sale prices, number of exported good, volume of production e.t.c.

Advantages of Index Number

Actually, indices are widely used as summaries of change in financial contexts by agents, government and industries.
Index number is imperative in business and economic situation because of the following reasons.

- It measures the changes in prices, production, standard and cost of living etc.
- It helps in comparing related items from year to year, from country to country etc.
- It enables us to predict or forecast business and economic conditions at a specified time.
- It provides or shows a clear picture of the changes in a variable with respect to time.
- It provides general information.

Steps to be considered before the construction of an index number

i. The significance of the index.
The reasons for putting up the index must be considered and carefully defined.

ii. Choice of the items in the basket.
It is always difficult while constructing index number to include all items due to time and implications. Hence, a sample must be taken and must not be biased.

iii. Choice weights.
Weight means the importance or significance attached to the construction of index number and hence, its choice must be considered.

iv. Choice of base period should be a period when things are stable and normal e.g. the base period must not be chosen when there is existence of isolated natural occurrences such as war, flood earthquake, famine, poor weather e.t.c.

3.2 Method of Computation

The nature, data availability and the scope of index number determine the method to be used for computation.

Price Relative Index (PRI*)*

This is the percentage ration of the price of an item or a commodity in current period to the price in earlier specified or reference period.

$$PRI = \frac{P_n}{P_O} X 100$$

Where P_n is the price for the current period
P_o is the price for the base period.

Quantity Relative Index

This is the percentage ratio of the quantity of an item or a commodity in current period to the quantity in earlier specified or reference period.

$$QRI = \frac{q_n}{q_O} X100$$

Where q_n is the quantity for the current period
q_o is the quantity for the base period.

Example 1

The prices of a television set in 1996 and 1997 were ₦15, 000 and ₦20, 000 respectively. Calculate:

The price relative index

- using 1996 as the base year
- using 1997 as the base year

Solution

P_{1996} = ₦15, 000 P_{1997} = ₦20,000

$$PRI = \frac{P_n}{P_O} \times 100$$

$$PRI = \frac{P_{1997}}{P_{1996}} \times 100 = \frac{20,000}{15,000} \times 100$$
$$= 133.3\%$$

This means there has been an increase of 33.3% over the 1996 price.

$$PRI = \frac{P_{1996}}{P_{1997}} \times 100 = \frac{15,000}{20,000} \times 100$$
$$= 75\%$$

This means that there is a decrease of 25% in 1996 when 1997 was computed.

Example 2

The table shows the cost of a ‘mudu’ of four types of food items in a given year.

Calculate the price relative using 1990 as the base year.

Food items	1990	1991	1992
yam flour	60	70	85
gari	80	90	120
beans	120	180	210
rice	200	250	300

Solution

Since the base period is 1990. Then the price of all the food items is 100%.

For 1991

$$\text{Yam Flour} = \frac{70}{60}\times 100 = 116.7\% \qquad \text{Gari} = \frac{90}{80}\times 100 = 112.5\%$$

$$\text{Beans} = \frac{180}{120}\times 100 = 150\% \qquad \text{Rice} = \frac{250}{200}\times 100 = 125\%$$

For 1992

$$\text{Yam Flour} = \frac{85}{60}\times 100 = 141.7\% \qquad \text{Gari} = \frac{120}{80}\times 100 = 150\%$$

$$\text{Beans} = \frac{210}{120}\times 100 = 175\% \qquad \text{Rice} = \frac{300}{200}\times 100 = 150\%$$

Food items	1990	1991	1992
yam flour	100	116.7	141.7
gari	100	112.5	150
beans	100	150	175
rice	100	125	150

3.3 Simple Price Index (SPI)

This can be defined as the ratio of the prices of a group of related items in a given year to the sum of the price of the items in a base year. This can be expressed mathematically as:

$$\text{SPI} = \frac{\sum P_n}{\sum P_o}\times 100$$

$$\text{QPI} = \frac{\sum q_n}{\sum q_o}\times 100$$

Example 3

Compute the simple price index of the data in the table below using 1995 as the base year.

SOAP TYPE	YEAR		
	1995	1996	1997
Lux	₦25	₦30	₦40
Delta	₦40	₦45	₦60
Tura	₦35	₦50	₦65
Premier	₦20	₦25	₦30
Total	120	150	195

Solution

1995

$$SPI = \frac{\sum P_n}{\sum P_o} X 100$$

$$SPI = \frac{120}{120} X 100 = 100\%$$

1996

$$SPI = \frac{\sum P_{1996}}{\sum P_{1995}} X 100 = \frac{150}{120} x100 = 125\%$$

1997

$$SPI = \frac{\sum P_{1997}}{\sum P_{1995}} X 100 = \frac{195}{120} x100 = 162.5\%$$

Year	1995	1996	1997
SPI	100%	125%	162.5%

Note: Simple price index is also called aggregate price index

Disadvantages of Simple Price Index

- It can be affected by the items' extreme values
- It is based on assumption that all items have equal weights since only prices are considered.

Simple Relative Price Index (SRPI)

This is basically designed to control the effects or inadequacies of the simple price index which is affected by their extreme values and units. The method is also referred to as simple average of relative.

$$SRPI = \frac{1}{N}\sum \frac{p_n}{p_o}$$

$$SRQI = \frac{1}{N}\sum \frac{q_n}{q_o}$$

Example 4

The table below shows the prices of three items from 1980 to 1983.

ITEMS	YEAR			
	1980	1981	1982	1983
vegetable oil	240	260	280	320
cotton seed oil	380	390	400	450
groundnut oil	340	370	400	420

Using the simple relative price index method to compute a price index of oil in 1983 using;

(i) 1980 as the base year
(ii) 1980-1982 as the base year

Solution

(i) Using 1980 as the base year

$$\text{PI of Vegetable oil} = \frac{320}{240} X 100 = 133.3\%$$

$$\text{PI of Cotton seed} = \frac{450}{380} X 100 = 118.42\%$$

$$\text{PI of Ground nut oil} = \frac{420}{340} X 100 = 123.53\%$$

$$SPRI = \frac{133.3 + 118.42 + 123.53}{3}\% = \frac{375.25}{3}\% = 125.1\%$$

(ii) Using 1980 -1982 as the base year

$$\text{For Vegetable oil} = \frac{240 + 260 + 280}{3} = 260$$

$$\therefore \text{ Price Relative Index} = \frac{320}{260}\% = 123.1\%$$

$$\text{For Cotton seed oil } = \frac{380 + 390 + 400}{3} = 390$$

$$\therefore \text{Price Relative Index} = \frac{450}{390} \text{ x } 100 == 115.4\%$$

$$\text{For Groundnut oil} = \frac{340 + 370 + 400}{3} = 370$$

$$\therefore \text{Price Relative Index} = \frac{420}{370} \text{ x } 100 = 113.5\%$$

$$\therefore \text{Simple Relative Price Index} = \frac{[123.1 + 115.4 + 113.5]\%}{3} = 117.3\%$$

Year	1980	1980-1982	1983
SRPI	125.1%	117.3%	100%

Disadvantages of SRPI

- It does not take into consideration the relative importance of each item.

3.4 Weighted Price Index

This corrects the inadequacies and limitations of simple aggregate and simple average relative methods since both methods do not take consideration of the relative importance of the various items and assume equal weight of items.

Advantages of Weighted Price Index

- It caters for the quantity of each commodity in order to weigh the price of each item.

- It uses the quantities in the base year as the weight or the quantity in the current year as the weight.

Weighted Mean of Relatives

This is otherwise called weighted average of relatives and the method involves computing index relative for each of the given item and using the weights given to determine a weighted mean of relative for the data. This can be mathematically expressed as:

$$\text{WAI (Weighted Average Index)} = \frac{\sum W_i I_i}{\sum W_i}$$

Where WI is the weighting factor.

'I' depends on either the price of the commodity or the quantity involved.

$$\text{Then} \quad I = \frac{P_n}{P_o} \text{ x } 100 \text{ (for price)}$$

$$I = \frac{q_n}{q_o} \text{ x } 100 \text{ (for quantity)}$$

Illustration

Imagine you are running a café and that you offer your customers when they order coffee, a choice from four different locations in Lagos i.e. Victoria Island, Festac Town, Yaba and Ikeja.

You charge the same amount irrespective of the coffee chosen; your task is to ascertain how much it is likely to cost you to provide this service to your customers. Once this is known, you will be able to decide how much you should charge a customer for coffee.

Your coffee suppliers volunteer the information that the four coffees sell in the ratio 4: 6 : 2 : 1 and cost ₦48, ₦15, ₦72, ₦28 per 100g respectively.

This information is sufficient to establish the likely average cost of a cup of coffee. The data values are the price that you have to pay for 100 g, the ratios can be considered to be relative frequencies or, in this context, 'weights'.

Then, the

$$\text{WAI (Weighted Average Index)} = \frac{\sum W_i I_i}{\sum W_i} = \frac{4x40 + 6x15 + 2x72 + 1x28}{4 + 6 + 2 + 1}$$

$$= \frac{454}{13} = 34.9$$

Interpretation

It is going to cost on average ₦34.92 to purchase 100g of the customer's choice of coffee.

Example 5

The costs of salad in consecutive years are given in the table below:

Vegetables	Year		Weights
Tomato	₦38	₦43	0.45
Lettuce	₦56	₦66	0.25
Spring Onion	₦17	₦22	0.05
Peppers	₦92	₦102	0.15
Sweet Corn	₦33	₦48	0.10

Calculate an index for the cost of salad vegetables in 1992 using 1991 as the base year and interpret your result

Solution

$$I = \frac{P_n}{P_o} x \ 100$$

Vegetables	P_o	P_n	$P_n/P_0=I_i$	W_i	W_iI_i
Tomato	38	43	1.132	0.45	0.5094
Lettuce	56	66	1.179	0.25	0.29475
Spring Onion	17	22	1.294	0.05	0.0647
Peppers	92	102	1.109	0.15	0.16635
Sweet Corn	33	48	1.455	0.10	0.1455
			Total	1.00	1.1807

$$\text{WAI (Weighted Average Index)} = \frac{\sum W_i I_i}{\sum W_i} = \frac{1.1807}{1.00} = 1.1807$$

Interpretation

This shows that on average, price index for 1992 is 1.18 indicating that price has increased by 0.1807 i.e 18.07% in 1992 compared to 1991.

Example 6

A company uses four raw materials in its production process. The cost per kg of each is given in the table below:

Material	Cost per kg. in 2003	Cost per kg. in 2004	Weight
A	72	82	22
B	96	98	30
C	42	105	15
D	88	120	33

You're required to determine the weighted average index in 2004 using 2003 as the base year and hence, interpret your result.

Solution

Vegetables	P_o	P_n	$P_n/P_0=I_i$	W_i	W_iI_i
A	72	82	1.1389	22	25.0558
B	96	98	1.0208	30	30.624
C	42	105	2.5	15	37.5
D	88	120	1.3939	33	44.9988
			Total	100	138.1786

$$\text{WAI (Weighted Average Index)} = \frac{\sum W_i I_i}{\sum W_i} = \frac{138.1786}{100} = 1.382$$

This implies that on average, price index for 2004 is 1.382, showing that price has increased by 0.382 i.e. 38.2% compared to base year which is 2003.

3.5 Laspeyres Index (LI)

i. Laspeyres Price Index (LPI)

LPI fixes the content and monitors the subsequent changes in price using the formula:

$$LPI = \frac{\sum P_n q_o}{\sum P_o q_o} \times 100 \quad 1$$

Where, P_n is the price for the current year
P_0 is the price for the base year
q_0 is the quantity for the base year
$\sum P_n q_0$ = the total cost of basket of items in the base year
$\sum P_0 q_0$ = the total value of the same basket in the given year

ii. Laspeyres Quantity Index (LQI)

$$LQI = \frac{\sum P_o q_n}{\sum P_o q_o} \times 100$$

3.6 Pasche Index (PI)

i. Pasche's Price Index (PPI)

PPI recognises and updates the contents of the basket and asks how much today's basket would have cost in earlier years. It under-estimates the price changes and requires more labour in data collection for computing the changes in price.

$$PPI = \frac{\sum P_n q_n}{\sum P_o q_n} \times 100$$

Where, $\sum P_n q_n$ = cost of items in the given year at the given values
$\sum P_0 q_0$ = the price of the items bought in the given year assuming the base year's price is used.

ii. Pasche's Quantity Index (PQI)

$$PQI = \frac{\sum P_n q_n}{\sum P_n q_o} \times 100$$

Example 7

The data given are the retail prices, in naira, of dairy products in 2003 and 2004 and the amount of each item purchased weekly by Adebowale family.

ITEMS	2003		2004	
	Price	*Quantity*	*Price*	*Quantity*
milk	15	30 dozen	20	28 dozen
eggs	82	42 dozen	88	40 dozen
butter	145	30 dozen	150	28 dozen
cheese	230	2 dozen	270	25 dosen

Using the data calculate:

a. Laspeyre's price index
b. Pasche's price index using 2003 as the base year

Solution

ITEMS	2003		2004		P_nq_0	P_0q_0	P_nq_n	P_0q_n
	P_o	q_0	P_n	q_n				
milk	15	30	20	28	600	450	560	420
eggs	82	42	88	40	3696	3444	3520	3280
butter	145	30	150	28	4500	4350	4200	4060
cheese	230	2	270	25	540	460	6750	5750
				Total	9336	8704	15030	13510

$$\text{Pasches Price Index} = \frac{\sum \text{Pnqn}}{\text{P0q0}} \text{ x } 100 = \frac{15030}{13510} x100 = 111.25\%$$

$$\text{Laspeyres Price Index} = \frac{\sum \text{Pnqn}}{\sum \text{P0q0}} \text{ x } 100 = \frac{9336}{8704} \text{ x } 100 = 107.26\%$$

3.7 Comparing Laspayres and Pasche

Actually, the two indices may be used, but there are advantages and disadvantages attached to each. This is because the advantages of one are automatically the disadvantages of the other.

Laspayres	Pasche
i. There is a general tendency to overestimate price changes.	i. Pasche's index underestimates the price changes
ii. The weights are taken from the base period figures and therefore, the only new data to be collected each time a new index is to be produced, will be the new prices.	ii. The Pasche's index requires new prices and weights each time, resulting in a larger data collection operation.
iii. It requires less re-calculation each time because the denominator remains constant.	iii. The denominator does not remain constant.
iv. It is very easy to make comparisons from year to year	iv. It can only be compared with the base year because the weights (and hence the denominator) change for every index.
v. The result may be less accurate than that of Pasche in terms of Current consumer patterns	v. It tends to give a better result in terms of current consumer patterns because it uses current period weighing.
vi. The units employed must be consistent as well as correct if the values are to be accurate	vi. As in Laspayres

3.8 Fisher's Ideal Index (FII)

This is otherwise known as ideal index number since it takes care of the underestimation of Pasche and overestimation of Laspeyre's indices respectively. It can be obtained by computing the geometric mean of both Pasche's and Laspeyre's indices.

$$\text{FII} = \sqrt{\text{Pasche Index x Laspeyre Index}}$$

3.9 Marshall-Edgeworth Index (MEI)

MEI uses the weighted aggregate for a particular year

$$\text{Marshall - Edgeworth Price Index} = \frac{\sum \text{Pn}(q0 + qn)}{\sum \text{P0}(q0 + qn)} \text{x100?}$$

$$\text{Marshall - Edgeworth Quantity Index} = \frac{\sum \text{Pn}(q0 + qn)}{\text{P0}(q0 + qn)} \text{ x } 100$$

Example 8

Commodities	1995		1996		1997	
	Price	*Qty*	*Price*	*Qty*	*Price*	*Qty*
A	38	48	32	66	40	58
B	56	36	60	43	68	36
C	34	60	38	54	46	52
D	27	45	33	46	29	42

Given the commodities A.B.C. and D and their respective prices and quantity, use the data above to calculate the followings.

- Laspeyre's Price Index (LPI)
- Pasche's Price Index (PPI)
- Fisher's Ideal Index (FII)
- Marshall -Edge Worth Price Index (MEPI)

You are required to use 1995 as the base year

Solution

Year	P_nq_0	P_0q_0	P_nq_n	P_0q_0	$P_n(q_0+q_n)$	$P_{0(}q_0+q_n)$
1996	1536	1824	2112	2508	3648	4332
	2160	2016	2580	2408	4740	4424
	2280	2040	2052	1836	4332	3876
	1485	1215	1518	1242	3002	2457
Total	7461	7095	8262	7994	15723	15089

Year	P_nq_0	P_0q_0	P_nq_n	P_0q_0	$P_n(q_0+q_n)$	$P_{0(}q_0+q_n)$
1997	1920	1824	2320	2204	4240	4028
	2448	2016	2448	2016	4896	4032
	2760	2040	2392	1768	5152	3808
	1305	1215	1218	1134	2523	2349
Total	8433	7095	8378	7122	16811	14217

For 1996

$$\text{LPPI} = \frac{\sum \text{Pnqn}}{\sum \text{P0q0}} \text{ x100} = \frac{7461}{7095} \text{x } 100 = 105.16\%$$

$$\text{PPI} = \frac{\sum \text{Pnqn}}{\sum \text{P0qn}} \text{ x } 100 = \frac{8262}{7994} x100 = 103.35\%$$

$$\text{FII} = \sqrt{\text{PPI x LPI}} = \sqrt{103.35 \text{ x } 105.16} = 104.25\%$$

$$\text{MEWI} = \frac{\sum \text{Pn(q0 x qn)}}{\sum \text{P0(q0 x qn)}} \text{ x } 100 = \frac{15{,}723}{15{,}089} \text{ x } 100 = 104.2\%$$

For 1997

$$\text{LPPI} = \frac{\sum \text{Pnq0}}{\sum \text{P0q0}} \text{ x } 100 = \frac{8433}{7095} \text{ x } 100 = 118.89\%$$

$$\text{PPI} = \frac{\sum \text{Pnqn}}{\sum \text{P0qn}} \text{ x} 100 = \frac{8378}{7122} \text{ x} 100 = 117.636\%$$

$$FIL = \sqrt{PPIxLPI} = \sqrt{118.89 \text{ x } 117.636} = 118.26\%$$

$$\text{MEWI} = \frac{\sum \text{Pn(q0 x qn)}}{\sum \text{P0(q0 x qn)}} \text{ x} 100 = \frac{16{,}811}{14{,}217} \text{ x } 100 == 118.25\%$$

Example 9

Using the data in the example above to calculate the weighted quantity index using

- Laspeyre's Quantity Index (LQI)
- Pasche's Quantity Index (PQI)
- Fisher's Ideal Index (FII)
- Marshall -Edge Worth Quantity Index (MQI)

Use 1995 as the base year.

Solution

Year	P_0q_n	$q_n(P_0+P_n)$	$q_{0(}P_0+P_n)$	Year	P_0+P_n	$q_n(P_0+P_n)$	$q_{0(}P_0+P_n)$
1996	70	4620	3360	1997	78	4524	3744
	116	4988	4170		124	4464	4464
	72	3888	4320		80	4160	4800
	60	2760	2700		56	2352	2520
Total	318	16256	14550	Total	338	15500	15528

For 1996

$$PQI = \frac{\sum Pnqn}{\sum P0q0}\ 100 = \frac{8262}{7461} x\, 100 = 110.7\%$$

$$LQI = \frac{\sum Pnqn}{\sum P0qn} x\, 100 = \frac{7994}{7095} x100 = 112.7\%$$

$$FII = \sqrt{PQI\ x\ LQI} = \sqrt{110.7\ x\ 112.7} = 111.7\%$$

$$MEWQI = \frac{\sum qn(P0\ x\ Pn)}{\sum q0(P0\ x\ Pn)} x\, 100 = \frac{16,256}{14,550} x\ 100 = 111.7\%$$

For 1997

$$PQI = \frac{\sum Pnqn}{\sum P0q0}\ x\ 100 = \frac{8378}{8433} x\, 100 = 99.4\%$$

$$LQI = \frac{\sum P0qn}{\sum P0qn}?\ \ x\ 100 = \frac{7122}{7095} x100 = 100.4\%$$

$$FII = \sqrt{PQI\ x\ LQI} = \sqrt{99.4\ x\ 100.4} = 99.90\%$$

$$MEWQI = \frac{\sum qn(P0\ x\ Pn)}{\sum q0(P0\ x\ Pn)} x\, 100 = \frac{15,500}{15,528} x\ 100 = 99.8\%$$

3.10 Deflating a Series Using the Retail Price Index (RPI)

When a series of figures is concerned with ‘money’ value and recorded through time, then it will be affected by inflation and so changes can be misleading. In order to overcome this, ‘real prices’ or ‘money value’ are determined by deflating the original series. This can be done by effectively changing the money values at one point in time, the base time, and so making the figures directly comparable.

Example 10

The table below shows the average weekly pay and the Retail Price Index (RPI), from 1995 through 1999.

Year	Average Weekly Wage '000 (₦)	RPI
1995	72.30	168.2
1996	76.42	189.63
1997	83.96	198.24
1998	104.35	210.24
1999	115.76	230.34

By deflating the series, determine the rise in purchasing power of an individual from 1995 to 1999.

Solution

The real value of earning in relation to 1995 can be computed as

$$= \text{Average wage for the year x} \frac{\text{RPI for 1995}}{\text{RPI for the year}}$$

and so on.

For 1995 a deflated figure is $= 72.30 \text{ x } \frac{168.2}{168.2} = 72.30$

1996 a deflated figure is $= 76.42 \text{ x } \frac{168.2}{189.63} = 67.78$

1997 a deflated figure is $= 83.96 \text{ x } \frac{168.2}{198.24} = 71.24$

1998 a deflated figure is $= 104.35 \text{ x } \frac{168.2}{210.42} = 83.41$

1999 a deflated figure is $= 115.76 \times \frac{168.2}{203.34} = 74.79$

The results are properly represented in the table below:

Year	Average weekly wage' 000 (N)	RPI	'Real' Weekly Wage (1995)
1995	72.30	168.2	72.30
1996	76.42	189.63	67.78
1997	83.96	198.24	71.24
1998	104.35	210.24	83.41
1999	115.76	230.34	84.53

Methods of Computation of Index Numbers

$$\text{Then, the purchasing power} = \frac{84.53 - 72.30}{72.30} \times 100 = \frac{12.23}{72.30} \times 100 = 16.92\%$$

3.11 Record Prices and Record Earnings

If an item has been sold for a record price, it does not make the item the most valuable at that period. This may be as a result of inflation. The best way of judging the real value of a particular item or thing is to ask what else could be bought with the same amount of money. If this has increased, then we say the original item has increased in value.

3.12 Retail Price Index (RPI)

This is often used to know the relative change in value of an item. This is based on a weighted combination of the prices of a 'basket' of commodities. Then, a commodity holds its value if its apparent value rises in step.

Example 11

The information in the table below gives the January prices of chocolate for the years 1995-2000 together with the changing retail price index values, where January 1995 is considered as the base year.

Year	RPI	Chocolate Price (in N)
1995	100.0	79.5
1996	107.2	98
1997	110.3	98
1998	115.3	87
1999	123.1	95
2000	132.2	99

- Is there any year that chocolates were more expensive?
- What was the real price of chocolates given the cost of living?

Solution

The RPI has risen from 100 to 132.2 over the 5 year period

$$\text{Percentage Increase} = \frac{\text{Increase}}{\text{Actual value}} \text{x}100 = \frac{132.2 - 100}{100} \text{x}100 = 32.2\%$$

This means that prices generally have risen by nearly one third during this period, (this is called inflation). If the individual salary has increased exactly in proportion to the RPI, would the person be able to buy more chocolates with the money earned in 2000 than he could in 1995? The question is simply answered by converting the chocolates price into indices, with the 1995 price being indexed to 100.

SELF ASSESSMENT EXERCISE

1. List and explain the two types of data needed in the construction of CPI.
2. A manufacturing company makes use of five raw materials all of which have risen steeply in prices over the past few years. Demand for the product has also risen. Relevant figures, comparing a single week in 2001 with corresponding week in 2002 are given below:

Raw	2001		2002	
	Price	*Quantity*	*Price*	*Quantity*
A	17	1300	36	1620
B	23	25	*52*	26
C	40	992	56	1420
D	22	120	32	135
E	50	80	78	140

Calculate two index numbers, one to show the extent of the price increases and one to show the extent of the increase in consumption.

4.0 CONCLUSION

The method to be used for the computation of an index number is determined by the nature, data availability and scope of index number.

5.0 SUMMARY

In this unit we have learnt the followings.

- The two basic types of data needed for the construction of CPI are price and weighted data
- The index reference period is different from the weight reference period and the price reference period.
- Index numbers are used as summaries of changes in financial contexts by agent, government and industries.
- Various methods can be used in the computation of index number for example, Laspeyres, Pasches, fisher e.t.c.
- The advantages of Laspeyres method of index are the disadvantages of Pasches methods.

6.0 TUTOR-MARKED ASSIGNMENT

1. What are the steps to be considered before the construction of index number?
2. A shopping bowl consists of burger, meat pie and Scotch egg. In year 2000, the bowl contains 300g of burger, 350g of meat pie and 200g of Scotch egg with the prices per l00g being ₦ 120, ₦ 60 and ₦ 40 respectively. In year 2002, the bowl contains 350g of burger, 420g of meat pie and 180g of Scotch egg with the prices per l00g being ₦ 140, ₦ 65 and ₦ 45 respectively.

Using year 2000 as being 100, determine:

(a) Laspeyre's price index
(b) Pasche's price index
(c) Marshall- Edge worth's index
(d) Fisher's Ideal index

3. The costs, in pence, of the ingredients of sandwich are given in the table below: Calculate an index for the cost of sandwich ingredients in year 2 based on year 1 and interpret your result.

Item	Year 1		Year 2	
	Price	*Quantity*	*Price*	*Quantity*
Bread	47	*2*	48	4
Butter	56	*4*	58	8
Cheese	148	3	132	2
Tomato	18	7	14	3

Also, using year 1 as the base year compute

(a) Laspeyre's quantity index
(b) Pasche's quantity index,
(c) Fisher's ideal index
(d) Marshall-Edge worth's quantity index.

4. An index indicating price changes in energy supplies has been recorded from 1989-1993, the figures having base year 1989.

Year	Price Index	RPI
1989	106.2	223.5
1990	115.8	262.8
1991	126.7	292.8
1992	129.3	300.7
1993	136.4	314.7

(a) Convert the price index to base year 1989.
(b) Deflate the index using the RPI figures given.
(c) What is the deflated percentage price change from 1989 to 1983?

7.0 REFERENCES/FURTHER READING

Doody,F. S. (1965). *Introduction to the Use of Economic Indicators.* New York-Random House, Inc.

Fisher, I. (1923). *The Making of Index Numbers.* New York: Houghton Mifflin Co.

Moore, Geoffrey H., & Shishkin, Julius. (1967). *Indicators of Business Expansions and Contractions.* New York: National Bureau of Economic Research, Columbia University Press.

Mudgett, B. D. (1951). *Index Numbers.* New York: John Wiley & Sons, Inc.

UNIT 4 MEASUREMENT OF MACROECONOMIC VARIABLES

CONTENTS

1.0 INTRODUCTION

The distinction between nominal and real values is of key importance in measuring macroeconomic variables, such as wages, interest rates, income, wealth, and gross domestic product. Nominal variables are measured without adjusting for any changes in the value of the currency in which the variable is measured. Real variables are adjusted for changes in the currency's value. Simply put, real variables are adjusted so that increases due solely to the ballooning effects of inflation are removed from their values.

In order to capture meaningful changes in microeconomics, economists rely upon inflation-adjusted or real variables. Nominal values often provide misleading information. For example, for several years after the breakup of the former Soviet Union, many newly independent transition nations reported nominal GDP values that increased from year to year, often significantly. However, when adjusted for the effects of inflation, the real GDP figures showed a much different picture. Real GDP per capita is a key measure of economic growth for a country. It represents the inflation-adjusted value of goods and services available on average

per citizen in a nation. This often serves as a proxy measure for average standard of living, although there are problems associated with this interpretation. The GDP deflator, which is a form of price index, is used to convert nominal gross domestic product values to real GDP.

2.0 OBJECTIVES

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

- define real and nominal values
- differentiate between real and nominal values
- identify the pitfalls between real and nominal values
- change nominal to real values.

3.0 MAIN CONTENT

3.1 Review of Real and Nominal Values

A nominal value is expressed in the country's currency at the time of measurement. A real value is a nominal value that has been adjusted for changes in the value of the currency. Suppose for example, the nominal price of bread bun was ₦ 20 in 1980 and ₦ 1.50 in 2005. In the twenty-five years between 1980 and 2005, the purchasing power of the naira has changed because of changes in the average price level. In order to measure the real price of bread, the nominal price of bread would need to be adjusted by a price index that captures the changes in the average price level. Economists do not consider nominal variables to be of great interest because they include the distorting effects of higher prices and this factor masks whether actual real changes have occurred. Thus, economists regularly use prices indexes to adjust the nominal values and forming real values.

Note: One of the most important macroeconomic variables used by economists and policy makers is real GDP. Because economic well-being is measured, in part, by the goods and services individual citizens can purchase, it is important to convert nominal values for gross domestic product into real values that reflect the purchasing power of the currency. The GDP deflator is a price index used for this purpose. You should note that the GDP deflator is similar in nature to the CPI, although the goods covered and formulas are somewhat different.

3.2 Nominal and Real Variable

The first row of Table 1 shows the average price of a new house, which rose from N2500 in 1960 to N124700 in 2000. Are houses currently 50 times as expensive as in 1960? Not when we allow for inflation which also raised incomes and the ability to buy houses.

The second row of Table 1 shows the retail price index, using 2000 as the base year, inflation led to substantial increase in the RPI, during 1960-2000. The third row of table 1 calculates an index of rent house prices, expressed in 2000 in the top and bottom row.

To calculate the rent price of house in 1960, by expressing them at 2000 prices, we take the nominal price of N2, 500 and multiply by (100)/7.4) to allow for subsequent inflations, yield N33, 800. Rent prices have roughly tripled since 1960 (from N33, 800 to N124, 700). Most of the increase in nominal house prices in the top row of Table 1 was due to inflations.

	1960	1980	2000
House price (000s)	2.5	27.2	124.7
RPI (2000= 100)	7.4	39.3	100.0
Rent price of house (2000)	33.8	69.2	124.7

3.2.1 Difference between Real or Relative Prices

The distinction between nominal and real variables applies to all variable measured in money values. It does not apply to units of output, such as 4000 carpets per annum, which relate to physical quantities. Whatever the inflation rate, 4000 carpets are 4000 carpets. However, we do not know whether ₦100 is a large or a small number until we know the general price level for goods.

The argument comes over to prices themselves. The nominal price of solver has risen a lot since 1970. To calculate the index of the rent price of solver, divide an index of nominal solver prices by the real price index of economic scarcity. They show whether the price of a commodity rose faster than its price in general. Hence, real prices are sometimes called relative price. Consider the price of television over the last 30 years. TV prices, measured in pounds, have hardly changed. The RPI has risen a lot. The real price of TVs has fallen. Advances in technology reduced the cost of producing television sets. Because the real price has fallen, many households now have several. It is misleading to base our analysis on nominal values of variables.

3.3 The Purchasing Power of Money

The PPM is an index of the quantity of goods that can be bought for ₦1.
To distinguish real and nominal variable we say that real variable measure nominal variables as if the purchasing power of money had been constant.
Another way to express this idea is to distinguish nominal variable in current naira and rent variables in constant naira.

Table 1 described real prices of houses measured in 2000 naira. We could of course have used 1960 naira instead. Although the level of the real price index for houses would have been different, it would have grown at exactly the same rate as the final row of table 1.

3.4 Real and Nominal Interest Rates: Inflated and Present Values

Consider the table below. The 1st column of table 2 shows rental receipts in actual naira. The interest rate of 10 percent tells us how many actual naira we earn by lending ₦1 for a year.

At a nominal interest rate of 10 percent, ₦100 percent today accumulates to ₦110 by next year. But we want to know how many goods that ₦110 will then buy.

Suppose the nominal interests rate is 10 percent but inflation is 6 percent; leaving ₦1 for a year gives ₦1.10, but after a year it costs ₦1.06 to buy goods we could have bought for ₦1 holding with ₦1.10 to spend next year, our purchasing power rises by only 4 percent.
The real interest rate is 4 percent.

Thus:

Real interest rate = nominal interest rate –inflation rate

To use this formula, by a second example: normal interest rates are 17% and inflation is 70%. Leaving ₦100 for a year, you get ₦117, but it will cost you ₦120 to buy goods you could have bought today for N100. You are worse off by 3% by delaying purchases for a year and lending your money at the apparently high rate of 17%. Real interest rates are negative. The real interest rate is 3%, in real terms it costs you to be a lender.

The nominal interest rate does not compensate for higher prices of goods you ultimately wish to buy.

Table 2

Year	Rental (₦)	Se Value (₦ -)	Present Value (₦)
1	4000		3640
2	4000	+ 10,000	11620
Asset price			15260

3.5 What Determines the Real Interest Rate?

Two forces lead to positive real interest rates. First, people are impatient. Given the choice of an equal number of goods tomorrow or today, we would rather have them today. To delay spending on goods and services, savers have to be bribed with a positive real interest rate that lets them consume more goods in the future if they postpone consumption and lend today.

Second, there must be a way of earning positive real reforms or borrowers would never borrow. Borrowers pay positive real interest rate because they can buy capital goods that provide a stream of return more than sufficient to meet the interest cost.

Importance to consumer and the productivity of physical capital are the two forces that lead us to expect a positive interest rate.

3.6 Nominal and Real GNP

Nominal and real variables are also used in terms of GNP to measure economic performance.

Nominal GNP: measure GNP at the prices prevailing when income was earned.

Real GNP or GNP at constant price adjusts for inflation by measuring GNP in different years at the prices prevailing at some particular data known as the base year.
Consider table 3 which presents a simple hypothetical example of a white economy: Nominal GNP rises from ₦600 to ₦ 1440 between 1980 as the base by valuing output quantities in 2002 using 1980 prices.

Real GNP rises only from ₦600 to ₦860. This rise of 43% in real GNP gives a value by the economy as white.

		1980	2002
Quantity	apples	100	150
	chickens	100	140
Price (₦)	apples	2	4
	chickens	4	6
Value in current(₦)	apples	200	600
	chickens	400	840
	Nominal GNP	600	1440
Value in 1980	apples	200	300
	chickens	400	560
	Real GNP	600	860

3.7 Changing Nominal to Real Values

To transform a series into real terms, two things are needed: the nominal data and an appropriate price index. The **nominal** data series is simply the data measured in current naira and gathered by a government or private survey. The appropriate price index can come from any number of sources. Among the more prominent price indexes are the Consumer Price Index (CPI), the Producer Price Index (PPI), the Personal Consumption Expenditure index (PCE) and the GDP deflator.

Common price indexes measure the value of a basket of goods in a certain time period, relative to the value of the same basket in a base period. They are calculated by dividing the value of the basket of goods in the year of interest by the value in the base year. By convention, this ratio is then multiplied by 100.

Generally speaking, statisticians set price indexes equal to 100 in a given base year for convenience and reference. To use a price index to deflate a nominal series, the index must be divided by 100 (decimal form). The formula for obtaining a real series is given by dividing nominal values by the price index (decimal form) for that same time period.

But how does this simple formula remove price fluctuations from actual changes in a variable's overall value? Economic variables measured in dollar values like GDP, exports, construction contract values, venture capital and retail sales are calculated from the product of the quantity sold and the selling price. Analysts want to get their hands around the changes in quantity sold and disregard changes in prices because it is the quantity of goods and services consumed by households that affect well-

being, not the prices. In effect, the percentage change in real values over a given time period should mirror the percentage change in quantity.
Table 1 provides three scenarios that show how to correct the data for price fluctuations.

In each scenario, price and quantity are multiplied together to arrive at a nominal value in 1990 and 1995. Then the 1995 nominal value is divided by the ratio of the 1995 price index and the 1990 price index to arrive at a real value (or the 1995 value in 1990 dollars).

Scenario	Period	Price	Quantity	Nominal Value	Deflating Nominal to Real	Real Value
1. Price rises 50%, quantity stays same	1990	100	12	1,200		1,200
	1995	150	12	1,800	1,800/ (150/100) =	1,200
2. Price stays the same, quantity rises 50%	1990	100	12	1,200		1,200
	1995	100	18	1,800	1,800/ (100/100) =	1,800
3. Price rises 20%, quantity rises 25%	1990	100	12	1,200		1,200
	1995	120	15	1,800	1,800/ (120/100) =	1,500

The Mechanics of Each Scenario

Scenario 1: Prices rise 50 percent from 1990 to 1995 but the quantity stays the same. Result: The nominal value increases 50 percent, but the real value remains the same.

Scenario 2: The price remains constant but quantity increases by 50 percent.
Result: The real value rises by 50 percent.

Scenario 3: The price rises 20 percent and quantity rises 25 percent.
Result: After deflating the 1995 value to 1990 dollars, the real value rises 25 percent.

3.8 Illustration and Notation on Real Values and Nominal Values

The simplest case of a *bundle* of commodities (goods) is one that has only one commodity. In that case, output or consumption may be measured either in terms of money value (nominal) or physical quantity (real). Let *i* designate that commodity and let:

- P_i = the unit price of *i*, say, ₦5
- Q_i = the quantity of *i*, say, 10 units.

The nominal value of the bundle would then be price multiplied by quantity:

- nominal value of *i* = P_i x Q_i =N5 x 10 = N50.

Given *only* the nominal value and price, derivation of a *real value* is immediate:

- real value of bundle *i* = P_i x Q_i/P_i = Q_i = 50/5 = 10.

The price "deflates" (divides) the nominal value to derive a real value, the quantity itself.

Similarly, for a series of years, say five, *given only* nominal values of the goods and prices in each year *t*, a *real value* can be derived for each of the five years:

- real value of bundle *i* in year *t* = nominal value of $Q_i^t/P_i^t = Q_i^t$.

This example generalises for nominal values relative to real values across different years for which P, a **price index** comparing the general price level across years, is available. Consider a nominal value (say of an hourly wage rate) in each different year *t*. To derive a real-value series from a series of nominal values in different years, divide nominal value in each year by P^t, the price index in that year. By definition then:

- real value in year *t* = nominal value in year *t*/P^t.

Numerical Example

If for years 1 and 2 (say 20 years apart) the nominal wage and P are ₦10 and ₦16 and 1.00 and 1.333, respectively. The real wages are ₦10 (= 10/1.00) and ₦12 (= 16/1.333) respectively.

The real wage so constructed in each different year indexes the amount of commodities in that year that could be purchased relative to other years. Thus, in the example, the price level increased by 33 percent, but the real wage rate still increased by 20 percent, permitting a 20 percent

increase in the quantity of commodities the nominal wage could purchase.

The generalisation to a commodity bundle from the single-good illustration above is to a bundle of quantities of *different* commodities and different years. This has practical use, because **price indexes** and the **National Income and Product Accounts** are constructed from such bundles of commodities and their respective prices.

A sum of nominal values for each of the different commodities in the bundle is also called a *nominal value*. A bundle of *n* different commodities with corresponding prices and quantities for each year *t* defines:

- nominal value of that bundle in year t = P_1^t x Q_1^t + . . . + P_n^t x Q_n^t.

From the above:

- P_t = the value of a price index in year *t*.

The nominal value of the bundle over a series of years and corresponding P^t define:

- real value of the bundle in year t = Q^t = nominal value of the bundle in year t/P^t.

Alternatively, multiplying both sides by P^t:

- nominal value of the bundle in year $t = P^t$ x Q^t.

So, every nominal value can be **dichotomised** into a price-level part and a real part. The real part Q^t is an index of the quantities in the bundle.

An illustration of a *nominal-value* sum is **nominal GDP**. An illustration of a *real-value* sum (or quotient) is **real GDP**.

3.9 Pitfall between Real and Nominal Value in Terms of Cash Flows

Cash flows expected to occur in the future, expressed in real terms, are often equated with current values.

In the valuation of the cash flows we have two choices: cash flows are all denominated in real or in nominal terms. In the first case, the cash flows are expressed in the currency of purchasing value as of the time of occurrence of the flow and in the second case they are expressed in

currency of purchasing value as of the day of the evaluation. For example, a kilowatt/hour of electricity to be sold five years from now, in nominal terms would be valued at the actual price it is expected to cost five years from now, taking into account inflation and technological advancements. In real terms, it should be valued at that price, deflated by the expected inflation in electricity prices over those years. Unfortunately, to value it in real terms requires the forecast of the future price, including the rate of technological change, the competitive situation and the rate of inflation on electricity prices, not on the overall inflation level. Most evaluations, to avoid these complex issues, value future cash flows at today's prices and call it a real value. Nevertheless, today's price is neither a real nor nominal value. Do you expect personal computers to cost the same five years from now? If you were to express that future value in dollars of today's purchasing power, it would be far different than today's current value

SELF ASSESSMENT EXERCISE

1. Distinguish between nominal and real variables
2. Distinguish between the concepts of nominal and real income, using a graphical approach.

4.0 CONCLUSION

Real and nominal value denotes the difference between money valued with inflation adjustments, and valued as currency of the current time. Due to inflation, the same amount of money at some time in the future may not have as much value, or purchasing power, as it does presently. Essentially, although the face value may remain the same, that amount may not be as useful as it once had been.

5.0 SUMMARY

We have learnt that real interest rates differ from nominal interest rates. The expected real interest rate is equal to the nominal interest rate, minus the expected rate of inflation during the time of the investment. Real value and nominal value are also used in other fields where an item has a stated measurement or amount, while it in fact is somewhat different. For example, 56Kbps modems are only capable of transferring up to the speed of 53.3Kbps. Voltage, such as in electric currents and batteries, also has variance (in some cases plus and/or minus).

6.0 TUTOR-MARKED ASSIGNMENT

1. Review how a price index is calculated and how CPI differs from GDP deflator

2. Compute real GDP from nominal GDP
3. Provide a general formula that connects real and nominal percentage changes to changes in the rate of inflation
4. Distinguish between nominal GDP and real GDP and define the GDP deflator.

7.0 REFERENCES/FURTHER READING

Amartya, S. (1979). "The Welfare Basis of Real Income Comparisons: A Survey," *Journal of Economic Literature*, 17(1), pp. 1-45.

Diewert, W.E. "Index numbers," *The New Palgrave: A Dictionary of Economics*, v. 2, pp. 767-80.

O'Donnell, R. (1987). "Real and Nominal Quantities," *The New Palgrave: A Dictionary of Economics*, v. 4, pp. 97-98 (Adam Smith's early distinction vindicated).

Usher, D. (1987). "Real income," *The New Palgrave: A Dictionary of Economics*, v. 4, pp. 104

MODULE 4 PURCHASING POWER PARITY

Unit 1 The Meaning of Purchasing Power Parity
Unit 2 Purchasing Power Parity as Exchange Rate Determination

UNIT 1 THE MEANING OF PURCHASING POWER PARITY (PPP)

CONTENTS

1.0 INTRODUCTION

An important concept to understand when evaluating the economic health of nations and the relative dynamics between international markets is the idea of Purchasing Power Parity (PPP). The basic theory asserts that the prices of common goods between two countries should be equal once prices have been converted to a common currency.

Distilled to its basic form, purchasing power parity is a ratio that displays the relative price level differences across two countries for similar products or group of products.

PPP is used as a first step in making inter-country comparisons based in real terms of gross domestic product (GDP) and its component expenditures. GDP is commonly used as an economic indicator for size, growth, and health of a nation. PPP allows countries to be viewed through a common reference point. Taken as a long-term theory, one should expect a convergence of all prices for common goods around the world in order for equilibrium to take affect and mitigate cost arbitrage opportunities.

2.0 OBJECTIVES

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

- describe the basic concepts of PPP and how the theory of PPP works
- identify the major uses of PPPs
- calculate PPPs (the practical application of PPP and the flaws surrounding the application)
- identify the problems associated with PPP.

3.0 MAIN CONTENT

3.1 What is Purchasing Power Parity?

Purchasing Power Parity (PPP) is a theory which states that exchange rates between currencies are in equilibrium when their purchasing power is the same in each of the two countries. This means that the exchange rate between two countries should equal the ratio of the two countries' price level of a fixed basket of goods and services. When a country's domestic price level is increasing (i.e., a country experiences inflation), that country's exchange rate must depreciate in order to return to PPP.

3.2 Building the Theory of Purchasing Power Parity

To begin to understand PPP, one can start at how the theory is developed from the law of one price and then translated or generalised to the concepts of absolute purchasing power parity and relative purchasing power parity.

3.2.1 Law of One Price

The foundation of purchasing power parity is grounded in the law of one price. The theory states that barring frictional or complicating factors such as tariffs, taxes, and transportation costs, the price of an internationally traded good in one country should achieve the identical

price in another country, once the price is adjusted to a common currency.

To provide an illustrative context, a frequent example of purchasing power parity manifests itself in the form of the McDonald's Big Mac TM. Quite often, the economist will publish a light-hearted survey concerning the prices of a Big Mac TM from different countries in order to: (1) evaluate exchange rates based upon differences in the prices of this nearly ubiquitous good, (2) get a comparative sense of economic output and health across countries.

As an example, consider a basic ingredient to the Big Mac TM like a beef patty. Assume that the price for a beef patty in the United States is represented as $P_{bp}^{\$}$ and the price of the same beef patty in France is represented as P_{bp}^{E}, the law of one price can be expressed mathematically as follows:

$$P_{bp}^{\$} = E * P_{bp}^{E} \qquad \text{……..(1)}$$

Here, E is the exchange rate between the euro/dollar. Assume the cost of a pre-formed beef patty in the U.S. is $1.00 per patty. If the euro/dollar exchange rate is 0.845316 (the spot rate as of 12/07/2005), then the price for the same beef patty in France should be €0.85. An arbitrage opportunity could arise if the pre-formed patty sold for anything less than €0.85 in France. A shrewd trader could then purchase pre-formed beef patties in France and resell them in the U.S. for a tidy profit until supply and demand forces took hold and drove down the prices in the U.S. market while driving up the prices in the French market. This process would repeat itself until price convergence is achieved.

Thus, the law of one price holds true within this iterative pricing cycle for a discrete product like the beef patty.

3.2.2 Absolute PPP

Using the intuition built by the law of one price for a discrete product, one can apply the principle across an aggregate of products and prices. Or put in another way, one can imagine a common basket of goods that can be traded and prices compared across two countries- this is also known as the consumer price index (CPI). By using price indices, one can rewrite equation (1) to make a relative comparison of overall price levels between France and the U.S., P^{E} *and* $P^{\$}$:

$$P^{E} = E_{E/\$}\ xP^{\$} \qquad \text{…….(2)}$$

Once again $E_{E/\$}$, is the exchange rate between the euro and the dollar. If PPP holds true, then equation (2) can be rearranged to derive the form of absolute PPP:

$$\frac{P^E}{P^\$} x \frac{1}{E_{E/\$}^{PPP}} = 1 \qquad \text{……..(3.1)}$$

Or

$$\frac{P^E}{P^\$} = E_{E/\$}^{PPP} \qquad \text{……(3.2)}$$

The left-hand side of equation 3.1 can also be referred to as the real exchange rate or the exchange rate that has been adjusted for relative pricing levels. In the alternate form of equation 3.2, one can see that for purchasing power parity to hold, a very strict definition of the basket of goods must be met or the composition of the CPI must be identical.

However, in empirical usage, one finds international price indices are varied in goods and different weights assigned to those goods. Another complication associated with international indices is that they must be constructed from similar base years.

The adaptation of the law of one price to PPP brings a particular nuance or wrinkle. In the beef patty arbitrage example cited above, the trader would take advantage of the price differences between two markets and the exchange rate was assumed constant. Here, there is an attempt to hold the price levels constant and it is the exchange rate that fluctuates. This becomes the basis of the idea that PPP is a determinant of the exchange rates.

3.2.3 Relative PPP

One sees from the discussion above, that in order to control a constant price differential between the two country's indices, one can focus on measuring the relative purchasing power parity. Using a relative PPP allows for a different basket of goods and varying weights to be applied towards the goods within the CPI. This is because relative PPP states that fluctuations in the pricing levels will be related to the fluctuations in the exchange rates. Changes in relative exchange rates are attributed to varying inflation rates across countries. Thus, relative PPP can be expressed as follows:

$$\frac{E_{E/\$}^2 - E_{E/\$}^1}{E_{E/\$}^1} = \Pi_E - \Pi_\$ \qquad \text{……(4.1)}$$

Or

$$\%\Delta E_{E/\$} = \%\Delta P^{E} - \%\Delta P^{\$} \qquad \text{.....(4.2)}$$

In equation 4.1, one sees that the percent change in exchange rates over a given range of time will be equal to the differences in inflation of the euro, Π_E, and the dollar, $\Pi_\$$

Put in a slightly different manner, equation 4.2 expresses the differences in percent changes in price levels in France and United states, also known as changes in relative inflation, as direct determinant in the relative changes in exchange rates between the two countries. It is the calculation of relative PPP that many economists and theorist will anchor their empirical tests in order to prove out the validity of PPP.

3.3 PPP in the Long-Run

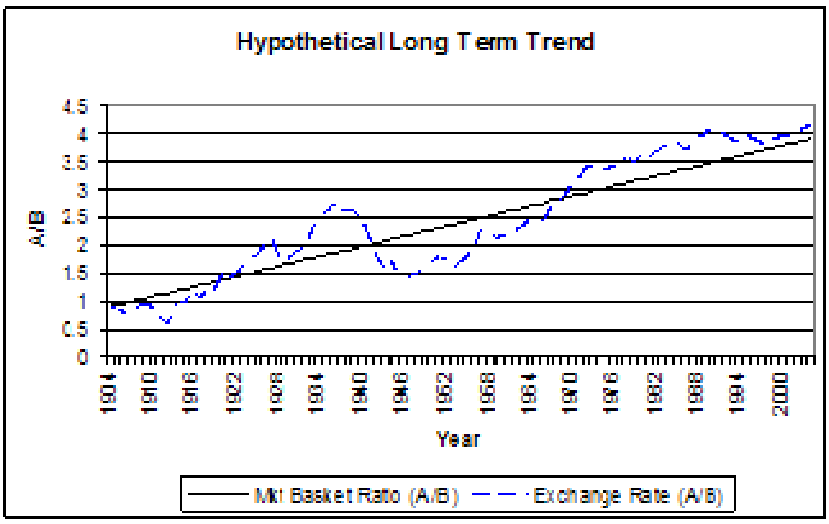
Economists have conceived of an alternative way to interpret or apply the PPP theory to overcome the empirical testing problem. The trick is to think of PPP as a "Long-Run" theory of exchange rate determination rather than a short-run theory. Under such an interpretation, it is no longer necessary for PPP to hold at any point in time. Instead, the PPP exchange rate is thought to represent a target towards which the spot exchange rate is slowly drawn.

This long-run interpretation requires an assumption that importers and exporters cannot respond quickly to deviations in the cost of market baskets between countries. Instead of immediate responses to price differences between countries, by engaging in arbitrage, buying at the low price and selling high, traders respond slowly to these price signals. Some reasons for the delay include imperfect information (traders are not aware of the price differences), long-term contracts (traders must wait till current contractual arrangements expire), and/or marketing costs (entry to new markets require research and setup costs). In addition, we recognise that the exchange rate is not solely determined by trader behaviour. Investors, who respond to different incentives, might cause persistent deviations from the PPP exchange rate even if traders continue to respond to the price differences.

When there is a delayed response, PPP no longer needs to hold at a particular point in time. However, the theory does imagine that eventually, traders will adjust to the price differences, buying low and selling high, causing an eventual adjustment of the spot exchange rate towards the PPP rate. However, as adjustment occurs, it is also quite

possible that the PPP exchange rate also continues to change. In this case, the spot exchange rate is adjusting towards a moving target.
How long will this adjustment take? In other words, how long is the long-run? The term itself is generally used by economists to represent some "unspecified" long period of time. It might be several months, several years or even several decades. Also, since the target, the PPP exchange rate, is constantly changing, it is quite possible that it is never reached. The adjustment process may never allow the exchange rate to catch up to the target, even though it is constantly chasing it.

Consider the chart below, the best way to see what the long-run PPP theory suggests is to consider the adjoining diagram. The figure presents constructed data (i.e., made up) between two countries, A and B. The solid black line shows the ratio of the costs of market baskets between the two countries over a long period of time, a century between 1904 and 2004. It displays a steady increase, indicating that prices have risen faster in country A relative to country B. The dotted blue line shows a plot of the exchange rate between the two countries during the same period. If PPP were to hold at every point in time, then the exchange rate plot would lie directly on top of the market basket ratio plot. The fact that it does not, means PPP did not hold all of the time. In fact PPP held only at times when the exchange rate plot crosses the market basket ratio plot, on the diagram. This happened only twice during the century - not a very good record.



3.4 Problems with the PPP Theory

The main problem with the PPP theory is that the PPP condition is rarely satisfied within a country. There are quite a few reasons that can

explain this and so, given the logic of the theory, which makes sense, economists have been reluctant to discard the theory on the basis of lack of supporting evidence. Below, we consider some of the reasons PPP may not hold.

3.4.1 Transportation Costs

The theory of the law of one price assumed that transportation costs were negligible and PPP requires the same assumption. However, getting products to and from different markets can add significant costs to goods. This will cause additional divergence in any price comparison when the product is applied to the basket of goods in a CPI. There is research that estimates transportation costs add 7 percent to the price of U.S. imports of meat, 6 percent to the import of price of dairy products, and 16 percent to the import price of vegetables.

3.4.2 Tariffs and Taxes

Once again, for PPP to hold, another key assumption underlying the law of one price is the abstraction of taxes and tariffs. There can be no imposition of unequal taxes or tariffs in goods imported or exported across countries, otherwise this would tend to cause further separation of true price levels between markets. A noted scholar on PPP, Cassel (1918), stated that "If trade between two countries is more hampered in one direction than in the other, the value of the money of the country whose export is relatively more restricted will fall, in the other country, beneath the purchasing power parity." Export restrictions were particularly relevant during Cassel's time due to their widespread use during the First World War. The effect of export restrictions translates into the currency of a country with high export tariffs to be undervalued relative to PPP of another country with lower export tariffs. Conversely, a country with higher import restrictions will be overvalued on PPP basis.

During the early nineties, Japan imposed quotas and tariffs on beef imports that went as high as 70 percent. Likewise, Korea imposed a 30 percent tariff as well as quantity restrictions from 1989 through 1994. Trade barriers such as these erected an obstruction to natural price equalisation of beef across the world markets, not to mention the impact upon the domestic prices of beef within Japan and Korea. Imposed barriers to trade may be a partial explanation to the overvaluation of the yen and the war against the US dollar until the late 1990s.

Taxes also have the same effect as tariffs upon deviations away from PPP. Many prices on goods that make up the composition of the consumer price index are inclusive of taxes or value added taxes. For

instance, countries with higher rates of taxation (in comparison to the U.S.) on like goods such as beef patties will appear to have an overvalued currency relative to the U.S. dollar.

3.4.3 Costs of Non-Tradable Goods

The theory of PPP asserts that after the barriers to trade are expunged, the cost of a good such as the Big Mac TM should be the same across all countries in which the Big Mac TM is traded. However, the local price of a Big Mac TM constitutes more than simply the ingredients by which it is composed. Also embedded in the local price is the cost associated with selling and marketing. Even within the same country, the cost of the Big Mac TM can vary depending upon costs to lease or rent the restaurant space and pay for utilities such as power, water, and heat. Rents and utilities are examples of non-tradable goods. While it is possible for titles and deeds to be traded, the location of a property cannot. Thus it may be cheaper to rent a space for a restaurant in Boise, Idaho than in Manhattan, NY, however, it is a useless comparison if one desires to sell Big Macs TM to the population in Manhattan.

Another component incorporated in the prices is the cost to serve, market, and sell the Big Mac TM, which may be significantly different from country to country. Wages have always been understood to be a non-tradable good in economic terms. There are inherent barriers to people's ability to move across borders to take advantage of wage differences.

Costs of non-tradable goods can have a significant impact on pricing. Ong, (1997), estimates that non-traded goods account for 94 percent of a Big Mac's price.

3.4.4 Productivity

A theory developed by Balassa and Samuelson (1964), states that the non-tradable goods have a systemic impact on the deviation from PPP due to its correlation with the productivity of a nation. By incorporating non-tradable components into the goods that comprise the CPI, the inherent effect is to also incorporate the productivity differences across regions or countries. An underlying assumption of the Balassa-Samuelson theory is that high per-capita wages are an indicator of higher productivity. Thus, countries with high-wages will be overvalued relative to countries with lower-wages. There are further supportive arguments that differences in productivity are most pronounced in the traded goods sector. Therefore, high-productivity in traded goods will manifest in high-wages across all sectors - tradable goods and services - as companies within a country compete for workers. Higher wages paid

for services in high-income countries relative to lower-income countries will cause higher prices for services in the high-income countries.

There is a "trickle-down" effect of high prices in tradable goods to services within a high-income country. Thus, even if there were identical prices for tradable goods across countries, there would still be an overvaluation of currency in the high-income country.

3.4.5 Government Spending

Another contributing factor for deviations from PPP is differences in government spending on non-tradable goods across different countries. There is evidence that governments spend more on non-traded goods than the private sector does. For instance, if the U.S. government increased expenditures relative to another government, the price of non-traded goods in the U.S. will increase along with the overall price levels, akin to the above. Had PPP held prior to the expenditure, the dollar would now be overvalued relative to its PPP level. This phenomenon is particularly prevalent in high income nations.

3.4.6 Current Account Deficits

There is another related notion that purports a similar logic of the "trickle-down" price effects of non-tradable goods. Krugman, a noted economist, theorises that governments who establish current account deficits are spending more on traded goods than other countries. Hence, this will result in a drop in the price of non-traded goods in the deficit country. Here, had PPP held previously, there will be a situation where the deficit country's currency would be undervalued compared to its "healthier" neighbours.

3.4.7 Pricing to Market

Another contributing factor to deviations from PPP is a firm's ability to mark their products at different price points across different markets. Basic business theory states that profit maximisation can be achieved by pricing as high as is appropriate considering the elasticity of demand for a product. For instance, pharmaceutical companies often charge higher prices in the U.S. than in other countries for the same drug. This is partially attributed to higher incomes found in the U.S. and the corresponding willingness-to-pay. Another important factor that allows pricing to market is the ability to resell the product to another country. There are heavy restrictions on selling drugs within the U.S. and there has been increased debate on whether people can purchase prescription drugs from countries such as Canada.

3.4.8 Perfect Information

The law of one price assumes that individuals have good, even perfect, information about the prices of goods in other markets. Only with this knowledge will profit-seekers begin to export goods to the high price market and import goods from the low priced market. Consider a case in which there is imperfect information. Perhaps some price deviations are known to traders but other deviations are not known. Or maybe only a small group of traders know about a price discrepancy and that group is unable to achieve the scale of trade needed to equalise the prices for that product. (Perhaps they face capital constraints and cannot borrow enough money to finance the scale of trade needed to equalise prices). In either case, traders without information about price differences will not respond to the profit opportunities and thus prices will not be equalised. Thus, the law of one price may not hold for some products which would imply that PPP would not hold either.

SELF ASSESSMENT EXERCISE

1. What are PPPs?
2. What are the major uses of PPPs?

4.0 CONCLUSION

Purchasing Power Parity is a powerful tool that provides a common lens by which to view the economic health and condition of different countries.

5.0 SUMMARY

In this unit, we have learnt the followings.

- There have been numerous studies to prove the various theories surrounding PPP and its efficacy along particular time horizons.
- There has been a long standing belief that PPP is an effective tool to understand real exchange rates and their natural equilibrium states in the long-run.
- There is a consensus that the speed of convergence is relatively slow compared to the rampant shifts to exchange rates in the short-term.

6.0 TUTOR-MARKED ASSIGNMENT

1. How are PPPs calculated?
2. How are PPPs calculated for GDP?
3. What products are included in the basket of goods and services used for the calculation of PPPs?

7.0 REFERENCES/FURTHER READING

Engel, C. (1999). "Accounting for U.S. Real Exchange Rate Changes," *Journal of Political Economy*, V107, pp.507-38.

Frenkel, & M. Goldstein.eds. (1990). *International Policy Coordination and Exchange Rate Fluctuations.* Chicago: University of Chicago Press, pp. 159-87.

Griffiths, A. & Stuart. Wall. *Applied Economics*, (Tenth Ed). Prentice – Hall.

Gustav,C. (1918). "*Abnormal Deviations in International Exchanges," The Economic Journal*.

Krugman, P. R. "Equilibrium Exchange Rates," in W. H. Branson, Jacob A.

Michael, R. P. & Patricia, S. P. (2003). "*Burgernomics: A Big Mac TM Guide to Purchasing Power Parity*". *The Federal Reserve Bank of St. Louis*., pp. 9-12.

Ong. Li Lian, (2003). *The Big Mac Index: Applications of Purchasing Power Parity*. New York: Palgrave Macmillan.

Pakko & Pollard, pp. 17.

Suranovic, "Problems with Purchasing Power Parity".

Suranovic, S. (1997-2006). "International Finance Theory and Policy: Purchasing Power Parity," *The International Economics Study Center*.

UNIT 2 PPP AS AN EXCHANGE RATE DETERMINATION

CONTENTS

1.0 INTRODUCTION

The Purchasing Power Parity (PPP) theory uses the long-term equilibrium **exchange rate** of two currencies to equalise their **purchasing power**. It is based on the **law of one price**: the theory states that, in ideally efficient markets, identical goods should have only one price.

The PPP relationship becomes a theory of exchange rate determination by introducing assumptions about the behaviour of importers and exporters in response to changes in the relative costs of national market baskets. Recall the story of the law of one price, when the price of a good differed between two countries' markets, there was an incentive for profit-seeking individuals to buy the good in the low price market and resell it in the high price market. Similarly, if a market basket, containing many different goods and services, costs more in one market than another, we should likewise expect profit-seeking individuals to buy the relatively cheaper goods in the low cost market and resell them in the higher priced market. If the law of one price leads to the equalisation of the prices of goods between two markets, then it seems reasonable to conclude that PPP, describing the equality of market baskets across countries, should also hold.

However, adjustment within the PPP theory occurs with a twist compared to adjustment in the law of one price story. In the law of one price story, goods arbitrage in a particular product was expected to affect the prices of the goods in the two markets. The twist included in the PPP theory is that arbitrage, occurring across a range of goods and

services in the market basket, will affect the exchange rate rather than the market prices.

2.0 OBJECTIVES

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

- discuss the PPP theory and real exchange rate
- calculate based on PPP
- explain equilibrium story in an economic model.

3.0 MAIN CONTENT

3.1 The PPP Equilibrium Story

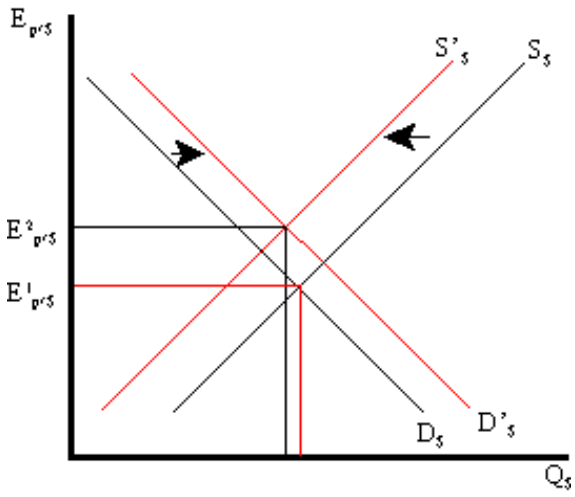
To see why the PPP relationship represents equilibrium, we need to tell an equilibrium story. An equilibrium story in an economic model is an explanation of how the behaviour of individuals will cause the equilibrium condition to be satisfied. The equilibrium condition is the PPP equation developed below,

$$E_{P/\$}^{PPP} = \frac{CB_P}{CB_\$} \qquad \ldots\ldots(1)$$

The endogenous variable in the PPP theory is the exchange rate. Thus, we need to explain why the exchange rate will change if it is not in equilibrium. In general, there are always two versions of an equilibrium story, one in which the endogenous variable ($E_{p/\$}$ here) is too high, and one in which it is too low.

PPP Equilibrium Story 1

Let's consider the case in which the exchange rate is too low to be in equilibrium. This means that,



$$E_{P/\$} < \frac{CB_P}{CB_\$} \Rightarrow CB_\$ E_{P/\$} < CB_P \qquad \text{.....(2)}$$

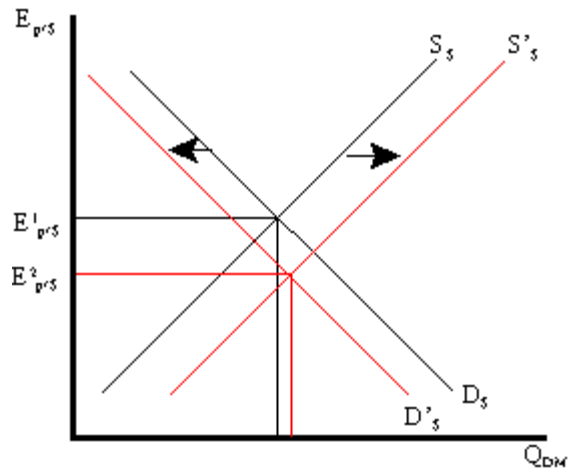
where $E_{p/\$}$ is the exchange rate that prevails on the spot market and, since it is less than the ratio of the market basket costs in Mexico and the US, it is also less than the PPP exchange rate. The right-hand side of the expression is rewritten to show that the cost of a market basket in the US evaluated in pesos, $CB_\$E_{p/\$}$, is less than the cost of the market basket in Mexico also evaluated in pesos. Thus, it is cheaper to buy the basket in the US, or, more profitable to sell items in the market basket in Mexico.

The PPP theory now suggests that the cheaper basket in the US will lead to an increase in demand for goods in the US market basket by Mexico, and, as a consequence, will increase the demand for US dollars on the foreign exchange market. Dollars are needed because purchases of US goods require US dollars. Alternatively, US exporters will realise that goods sold in the US can be sold at a higher price in Mexico. If these goods are sold in pesos, the US exporters will want to convert the proceeds back to dollars. Thus, there is an increase in US dollar demand (by Mexican importers) and an increase in peso supply (by US exporters) on the Forex. This effect is represented by a rightward shift in the US dollar demand curve in the adjoining diagram. At the same time, US consumers will reduce their demand for the pricier Mexican goods. This will reduce the supply of dollars (in exchange for pesos) on the Forex which is represented by a leftward shift in the US dollar supply curve in the Forex market.

Both the shift in demand and supply will cause an increase in the value of the dollar and thus the exchange rate, $E_{p/\$}$, will rise. As long as the US market basket remains cheaper, excess demand for the dollar will persist and the exchange rate will continue to rise. The pressure for change ceases once the exchange rate rises enough to equalise the cost of market baskets between the two countries and PPP holds.

PPP Equilibrium Story 2

Now let us consider the other equilibrium story, that is, the case in which the exchange rate is too high to be in equilibrium. This implies that,



$$E_{P/\$} > \frac{CB_P}{CB_\$} \Rightarrow CB_\$ E_{P/\$} > CB_P \qquad \text{..... (3)}$$

The left-hand side expression says that the spot exchange rate is greater than the ratio of the costs of market baskets between Mexico and the US. In other words the exchange rate is above the PPP exchange rate. The right-hand side expression says that the cost of a US market basket, converted to pesos at the current exchange rate, is greater than the cost of a Mexican market basket in pesos. Thus, on average US goods are relatively more expensive while Mexican goods are relatively cheaper.

The price discrepancies should lead consumers in the US, or importing firms, to purchase less expensive goods in Mexico. To do so, they will raise the supply of dollars in the Forex in exchange for pesos. Thus, the supply curve of dollars will shift to the right as shown in the adjoining diagram. At the same time, Mexican consumers would refrain from purchasing the more expensive US goods. This would lead to a reduction in demand for dollars in exchange for pesos on the Forex.

Hence the demand curve for dollars shifts to the left. Due to the demand decrease and the supply increase, the exchange rate, $E_{p/\$}$, falls. This means that the dollar depreciates and the peso appreciates.

Extra demand for pesos will continue as long as goods and services remain cheaper in Mexico. However, as the peso appreciates (the $ depreciates) the cost of Mexican goods rises relative to US goods. The process ceases once the PPP exchange rate is reached and market baskets cost the same in both markets.

3.2 Adjustment to Price Level Changes under PPP

In the PPP theory, exchange rate changes are induced by changes in relative price levels between two countries. This is true because the quantities of the goods are always presumed to remain fixed in the market baskets. Therefore, the only way that the cost of the basket can change is if the goods' prices change. Since price level changes represent inflation rates, this means that differential inflation rates will induce exchange rate changes according to the theory.

If we imagine that a country begins with PPP, then the inequality given in equilibrium story 1, $CB_{\$}E_{P/\$} < CB_{P}$, can arise if the price level rises in Mexico (peso inflation), if the price level falls in the US ($ deflation), or if Mexican inflation is more rapid than US inflation. According to the theory, the behaviour of importers and exporters would now induce a dollar appreciation and peso depreciation. In summary, an increase in Mexican prices relative to the change in US prices (i.e., more rapid inflation in Mexico than in the US) will cause the dollar to appreciate and the peso to depreciate according to the purchasing power parity theory.

Similarly, if a country begins with PPP, then the inequality given in equilibrium story 2, $CB_{\$}E_{P/\$} > CB_{P}$, can arise if the price level rises in the US ($ inflation), the price level falls in Mexico (peso deflation) or if US inflation is more rapid than Mexican inflation. In this case, the inequality would affect the behaviour of importers and exporters and induce a dollar depreciation and peso appreciation. In summary, more rapid inflation in the US would cause the dollar to depreciate while the peso would appreciate.

3.3 Purchasing Power Parity Theory and Real Exchange Rate

The absolute purchasing power parity theory suggests that, in the long run,

$$P = P^*/W, \qquad \text{……..(4)}$$

whereby P denotes the price of a given basket of goods in the domestic country and P* denotes the corresponding price abroad; W is the nominal exchange rate between the domestic country and the foreign country, expressed as the price of the domestic currency in units of the foreign currency, as is usually the case for euro exchange rates for example. A nominal appreciation of the domestic currency is therefore reflected by a rise in W. The relative purchasing power parity theory, however, is characterised by the long-run validity of

$$P = c(P^*/W) \qquad \text{……..(5)}$$

Whereby c represents a constant. If growth rates are flagged with a "^", (5) results in

$$\hat{W} = \hat{P}^* - P \qquad \text{……(6)}$$

According to (5), the purchasing power ratio between the domestic country and the foreign country is constant if the relative purchasing power parity theory holds. According to (6), the rate of change in the nominal exchange rate is equal to the inflation differential. If the countries in question are members of a monetary union, the nominal exchange rate is permanently fixed at W = 1. In such cases, the absolute purchasing power parity theory implies that domestic and foreign price levels should be identical in the long run, P = P* according to (4), and the relative purchasing power parity theory suggests that the inflation differentials according to (6) cancel out in the long run, ^ P* = ^ P.
In the majority of cases, the real exchange rate R is used as an indicator of an economy's price competitiveness. This may be defined as

$$R = W (P/P^*) \qquad \text{……(7)}$$

Equation (7) shows that the concept of a real exchange rate may also be applied to countries within a monetary union, where it is defined as the price ratio between the domestic country and the foreign country (R = P/P*). Equations (4), (5) and (7) show that absolute purchasing power parity is achieved if the real exchange rate is R =1 or In(R) = 0 and that relative purchasing power parity is achieved if R = c or In(R) = In(c).

Since the purchasing power parity theory is a long-run concept, it only applies if In (R) is stationary (relative version) or if In (R) is stationary at 0 (absolute version).

The indicator of the price competitiveness of the German economy is essentially no more than a trade weighted real exchange rate whereby the weights of individual trading partners correspond to their significance in German foreign trade.

3.4 Purchasing Power Parity Theory as a Theoretical Concept

There are two versions of purchasing power parity theory – an absolute and a relative one. Absolute purchasing power parity theory states that a basket of goods costs the same domestically and abroad if the goods prices are converted into a common currency. In other words, absolute purchasing power parity theory postulates that the purchasing power of money is equal between countries. By contrast, relative purchasing power parity theory does not compare domestic and foreign levels of purchasing power, but rather focuses on changes in this purchasing power. Relative purchasing power parity theory therefore states that the inflation rate differentials between two countries or regions are offset through inverse changes in the nominal exchange rate so that the purchasing power ratio between the two remains constant. It therefore follows that the validity of absolute purchasing power parity theory implies the validity of relative purchasing power parity theory, but not vice versa.

3.5 Calculation Based on PPP

The simplest way to calculate purchasing power parity between two countries is to compare the price of a "standard" good that is in fact identical across countries. Every year *The Economist* magazine publishes a light-hearted version of PPP: in its "Hamburger Index" that compares the price of a McDonald's hamburger around the world. More sophisticated versions of PPP look at a large number of goods and services. One of the key problems is that people in different countries consume very different sets of goods and services, making it difficult to compare the purchasing power between countries.

Consider the following information about movie video tapes sold in the US and Mexican markets.

Price of videos in US market ($P_{\$}^{v}$)	$20
Price of videos in Mexican market (P_{p}^{v})	p150
Spot exchange rate ($E_{p/\$}$)	10 p/$

The dollar price of videos sold in Mexico can be calculated by dividing the video price in pesos by the spot exchange rate as shown,

$$\frac{P_p^v}{E_{p/\$}}\left[\frac{\text{peso / video}}{\text{peso / \$}} = \frac{\text{peso}}{\text{video}} \times \frac{\$}{\text{peso}} = \frac{\$}{\text{video}}\right] = \frac{150}{10} = \$15/\text{video}$$

To see why the peso price is divided by the exchange rate (rather than multiplied), notice the conversion of units shown in the brackets. If the law of one price held, then the dollar price in Mexico should match the price in the US. Since the dollar price of the video is less than the dollar price in the US, the law of one price does not hold in this circumstance.

The next question to ask is what might happen as a result of the discrepancy in prices. Well, as long as there are no costs incurred to transport the goods, there is a profit-making opportunity through trade. For example, US travelers in Mexico who recognise that identical video titles are selling there for 25% less might buy videos in Mexico and bring them back to the US to sell. This is an example of "goods arbitrage." An arbitrage opportunity arises whenever one can buy something at a low price in one location and resell at a higher price and thus make a profit.

Using basic supply and demand theory, the increase in demand for videos in Mexico would push the price of videos up. The increase supply of videos in the US market would force the price down in the US. In the end, the price of videos in Mexico may rise to, say, 180 pesos while the price of videos in the US may fall to $18. At these new prices the law of one price holds since**,**

$$\frac{P_p^v}{E_{p/\$}} = \frac{180}{10} = \$18 = P_{\$}^{v}$$

The idea between the laws of one price is that identical goods selling in an integrated market, where there are no transportation costs or differential taxes or subsidies, should sell at identical prices. If different prices prevailed then there would be profit-making opportunities by buying the goods in the low price market and reselling it in the high price market. If entrepreneurs acted in this way, then the prices would converge to equality.

Of course, for many reasons the law of one price does not hold even between markets within a country. The prices of beer, gasoline and stereos will likely be different in New York City than in Los Angeles. The prices of these items will also be different in other countries when converted at current exchange rates. The simple reason for the discrepancies is that there are costs to transport goods between locations, there are different taxes applied in different states and different countries, non-tradable input prices may vary, and people do not have perfect information about the prices of goods in all markets at all times. Thus, to refer to this as an economic "law" does seem to exaggerate its validity.

SELF ASSESSMENT EXERCISE

1. What are the drawbacks to using exchange rates to convert GDP to a common currency for making international comparisons (e.g. of production or productivity)?
2. Can average exchange rates be used as proxies for PPPs?

4.0 CONCLUSION

PPP between two countries is the comparison of the price of a standard good that is identical across countries. It lumps items together into broad class, not putting quantity into consideration.

5.0 SUMMARY

In this unit we have learnt that:

- purchasing power parity is of course an imperfect device for determining things such as GDP, as the exchange rate will vary based on the basket item used for the index.
- the problem can be minimised by considering large sample commodities rather than one or two.

6.0 TUTOR-MARKED ASSIGNMENT

1. Should PPPs always be used rather than exchange rates in making international comparisons?
2. Can PPPs at the level of GDP be used to determine whether a currency is undervalued or overvalued?
3. Does PPP determine exchange rates in the short term?

7.0 FURTHER READINGS/REFERENCE

Engel, C. (1999). "Accounting for U.S. Real Exchange Rate Changes," *Journal of Political Economy*, V107, pp.507-38.

Griffiths, A, & Wall, Stuart. *Applied Economics*, (Tenth Ed). Prentice Hall

Gustav C. (1918). "*Abnormal Deviations in International Exchanges," The Economic Journal*.

Krugman, P. R. (1990)."*Equilibrium Exchange Rates*," in William H. Branson, Jacob A Frenkel, & Morris Goldstein, eds., *International Policy Coordination and Exchange Rate Fluctuations*. Chicago: University of Chicago Press, pp. 159-87.

Ong, Lillian (2003). *The Big Mac Index: Applications of Purchasing Power Parity*. New York: Palgrave Macmillan.

Pakko, M. R. & Pollard, P. S.(2003). pp. 17.

Pakko, M.R. & Pollard, P. S. (2003). "*Burgernomics: A Big Mac TM Guide to Purchasing Power Parity*". *The Federal Reserve Bank of St. Louis*., pp. 9-12.

Suranovic, S. (1997-2006). "*International Finance Theory and Policy*: *Purchasing Power Parity," The International Economics Study Center*.

Suranovic, S. "*Problems with Purchasing Power Parity*".

MODULE 5 HUMAN DEVELOPMENT

Unit 1 Human Development Indices

UNIT 1 HUMAN DEVELOPMENT INDICES

CONTENTS

1.0 INTRODUCTION

The Human Development Index (HDI) is a composite index which is designed to measure human well being in different countries. The index combines measures of expectancy, school enrolment, literacy, and income to provide a broader-based measure of well-being and development than income alone. Since its publication, this index has become widely cited and is commonly used as a way of ranking the quality of life in different countries.

2.0 OBJECTIVES

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

- describe human development and human development index
- compute human development index and physical quality of life index
- distinguish the relationship between economic growth and human development.

3.0 MAIN CONTENT

3.1 The Meaning of Human Development

The UNDP incorporated Sen's view in its first *Human Development Report* in 1990. According to it, human development goes far beyond income and growth to cover all human capabilities—the needs, aspirations and choice of the people. It defined human development as "a process of enlarging people's choices" that is created by expanding human capabilities. Income is *one* of the choices but it is not the *only* choice. Rise in income is not the same thing as the increase in human capabilities. Besides higher income, poor people put a high value on adequate nutrition, access to safe drinking water, better medical facilities, better school for their children, affordable transport, adequate shelter, secure livelihood and productive and satisfying jobs.

Human development is a broad and comprehensive concept. It is as much concerned with economic growth as with its distribution, as with basic human needs as with variety of human aspirations, as with the distress of the rich countries, and as with the human deprivation of the poor. The Report also explained the relationship between economic growth and human development. It emphasised, that there is no automatic link between the two. Economic growth is important because no society has been able to sustain the well-being of its people without continuous growth; so economic growth is essential for human development. However, human development is equally important because it is healthy and educates people who contribute more to economic growth through productive employment and increase in income. Thus, human development and economic growth are closely connected. In reality, economic growth is a means to an end, and the end is human development. Policy makers should, therefore, pay more attention to the quality of growth so as to support all-round human development.

3.2 Human Development Indices

Economists have tried to measure social indicators of basic needs by taking one, two or more indicators for constructing composite indices of human development. We study below the Physical Quality of Life Index (PQLI) of Morris and the Human Development Index (HDI) as developed by the United Nations Development Programme (UNDP).

3.2.1 Physical Quality of Life Index (PQLI)

The Physical Quality of Life Index was the most serious challenge to GNP per capita as the index of development. It was invented by M.D. Morris in 1979. He constructed a composite Physical Quality of Life Index (PQLI)

relating to 23 developing countries for a comparative study. He combines three component indicators of infant mortality, life expectancy at age one and basic literacy at age 15 to measure performance in meeting the most basic needs of the people. This index represents a wide range of indicators such as health, education, drinking water, nutrition and sanitation. The PQLI shows improvement in the quality of life when people enjoy the fruits of economic progress with increase in life expectancy (LE), fall in infant mortality rate (IMR) and rise in basic literacy rate (BLR).

Each indicator of the three components is placed on a scale of zero to 100 where zero represents an absolutely defined *worst* performance and 100 represents an absolutely defined *best* performance. The PQLI index is calculated by averaging the three indicators giving equal weight to each and the index is also scaled from 0 to 100.

If the indicators of life expectancy and basic literacy rate are *positive, the best* performance is shown as the *maximum* and the *worst* as the *minimum.* Infant mortality rate, being a *negative* indicator, for this the *best* indicator is shown as the *minimum* and the *worst* as the *maximum.* To find out the achievement level of the positive variable, its minimum value is deducted from its actual value and the balance is divided by the difference (range) between maximum value and minimum value i.e.

$$\text{Achievement Level} = \frac{\text{Actual Value} - \text{Minimun Value}}{\text{Max. Value} - \text{Min. Value}}$$

To find out the achievement level for a negative indicator, its actual value is deducted from its maximum value and the balance is divided by the difference (range; between maximum value and minimum value i.e.

$$\text{Achievement Level} = \frac{\text{Max. Value} - \text{Actual Value}}{\text{Max. Value} - \text{Min. Value}}$$

For life expectancy and Infant mortality rate, there is no natural maximum and minimum value. However, there is the need to select the right values.

According to Morris, each of the three indicators measures results and not inputs such as income. Each is sensitive to distribution effects. It means that an improvement in these indicators signifies an increase in the proportion of people benefiting from them. But none of the indicators depends on any particular level of development. Each indicator lends itself to international comparison. Taking Gabon's infants' mortality rate of 229 per thousand live births as the worst rate in 1950, Morris sets it at 0. At the

upper end, the best achievement is set at 9 per thousand for the year 2000. Again, taking Vietnam's life expectancy at age one as 38 years in 1950, Morris sets it at 0 of the life expectancy index. The upper limit is set at 77 years for men and women combined for the year 2000. Lastly, the basic literacy rate at 15 years is taken as the literacy index. This set of values is presented in Table 1.

Table 1: Maximum and minimum values of component indicators

Dimension	Maximum	Minimum	Range
Infant mortality rate	229	9	220
Life expectancy at age one	77	38	39
Basic literacy rate	180	0	100

On this basis, Morris presents the following correlation of (N=150).

(N = 150)	Infant Mortality	Life Expectancy
Life expectancy A age one	-0.1-919	-+
Literacy	-0.919	0.897

The coefficient of correlation between life expectancy at age one and infant mortality is of a high degree and negative. Similarly, the correlation between literacy and infant mortality rate *i.e.*, with literacy, infant mortality rate declines. The coefficient between literacy and life expectancy shows a high degree of-positive correlation *i.e.*, with literacy, the life expectancy also increases. Morris regards life expectancy at age one and infant mortality as very good indicators of the physical quality of life. So are literacy and life expectancy. In fact, the literacy indicator reflects the potential for development. We present in Table 2 the PQLI performance and GNP per capita of two LDCs and two developed countries.

Table 2: PQLI Performance and GNP per Capita Growth Rate

Country	PQLI			Average annual GNP Per capita growth Rate
	1950	1960	1970	
India	14	30	40	1.8
Sri Lanka	65	75	80	1.9
Italy	80	87	92	5.0
USA	89	91	93	2.4

Source: **Morris D. Morris and M.B. McAlpin, *Measuring the Conditions of India's Poor*, 1982.**

The above table reveals that India, which Morris calls a "basket case", exhibited slow but not insignificant improvement in its PQLI from 14 to 40 over a period of two decades from, 1950 to 1970, despite its low growth in average GNP per capita of 1.8. On the other hand, Sri Lanka's PQLI was much higher than India's during this period, though its average GNP per capita was almost the same. Of the two developed countries, both Italy and USA had very high PQLI. But Italy's average GNP per capita was more than double the USA. In this connection, Morris observes that there is no automatic link between GNP per capita and PQLI. In fact, the presence or absence of social relations, nutritional status, public health, education and family environment determine a society's PQLI. Further, it takes considerable time to build institutional arrangements that can generate and sustain a high PQLI.

Its Limitations

- The PQLI tries to measure “quality of life” directly rather than indirectly. However, it has its limitations.
- Morris admits that PQLI is a limited measure of basic needs. It supplements but does not supplant the GNP. It fails to dislodge GNP from its lofty perch. It does not explain the changing structure of economic and social organisation. It, therefore, does not measure economic development. Similarly it does not measure total welfare.
- Morris has been criticised for using equal weights for the three variables of his PQLI which undermine the value of the index in a comparative analysis of different countries.
- According to Meier, “Non-income factors captured by the PQLI are important, but so are income and consumption statistics and distribution-sensitive methods of aggregation *that are ignored by it.*

Conclusion

Despite these limitations, the PQLI can be used to identify particular regions of underdevelopment and groups of society suffering from the neglect or failure of social policy. It points towards that indicator where immediate action is required. The state can take up such policies which increase the PQLI rapidly along with economic growth.

3.2.2 Construction of PQLI

On the basis of the values of the component indicators given in Table 1, we can construct the PQLI in the following manner.

$$IMRI = \frac{229 - Actual\ IMR}{220}$$

$$LEI = \frac{Actual\ LIfeExpec\tan cy - 38}{39}$$

$$BEI = \frac{Actual\ LIfeacy\ Lervel - 0}{100}$$

We calculate the PQLI for India on the basis of 2001 Census data for these variables: IMR = 67, LE = 65 years, and BL = 65%.

$$IMRI = \frac{229 - 69}{220} = 0.74$$

$$LEI = \frac{65 - 38}{39} = 0.69$$

$$BEI = \frac{65 - 0}{100} = 0.65$$

$$PQLI = \frac{IMRI + LEI + BLI}{3} = \frac{0.74 + 0.69 + 0.65}{3}$$

$$= \frac{2.08}{3} = 0.69$$

Thus the Physical Quality of Life Index for India in 2001 was 0.69.

3.3 Human Development Index (HDI)

Lord Meghnad Desai and Nobel Laureate Amartya Sen invented the Human Development Index and UNDP incorporated it into its first Human Development Report in 1990.

Since then, the UNDP has been presenting the measurement of human development in terms of HDI in its annual report. The HDI is a composite index of three social indicators: life expectancy, adult literacy and years of schooling. It also takes into account real GDP per capita. Thus, the HDI is a composite index of three fundamental dimensions: living a long and healthy life, being educated and having decent standard of living.

The HDI value of a country is calculated by using the following three indicators. *Longevity*, as measured by life expectancy at birth. *Educational attainment*, *as* measured by a combination of adult literacy (two-thirds weight) and combined primary, secondary and tertiary enrolment ratio (one-third weight). *Decent Standard of living*, as measured by real GDP per capita based on purchasing power parity in terms of dollar (PPP$).

Before the HDI is calculated, an index is created for each of these dimensions: Life Expectancy Index, Education Index and GDP Index. To

calculate these indices, minimum and maximum values or goal posts are chosen for each indicator.

Table 3: Goalposts for Calculating the HDI

Indicator	Max, Value	Mim value
Life expectancy at birth (yes)	85	25
Adult literacy rate (%)	100	0
Combined gross enrolment ratio (%)	100	0
GDP per capita (PPP US$)	40,000	100

Performance in each dimension is expressed as a value between 0 and 1 by applying the following formula:

$$Dimension\ Index = \frac{Actual\ Value - Min.Value}{Max.Value - Min.value}$$

The *HDI* is then calculated as a simple average of the three dimension indices.

The *HDI value* for each country indicates the distance it has travelled towards the maximum possible value of 1 and how far it has to go to attain certain defined goals: an average life span of 85 years, access to education for all and a decent standard of living. The DHI ranks countries in relation to each other. A country's *HDI rank* is within the world distribution *i.e.*, it is based on its HDI value in relation to each developed and developing country for which the particular country has travelled from the minimum HDI value of 0 towards the maximum HDI value of 1. Countries with an HDI value below 0.5 are considered to have a low level of human development, those from 0.5 to 0.8 a medium level and those above 0.8 a high level. In the HDI, countries are also ranked by their GDP per capita.
The *Human Development Report*, 2004 presented the HDI values, HDI rank, and real GDP per capita ranks for the year 2002 relating to 177 developed and developing countries. Table 4 shows HDI values and HDI and HDI ranks for some of the countries.

Table 4: Human Development Index for Selected Countries, 2002

Country	HDI Value	HDI Rank	GDP per capita Rank minus HDI Rank
1. High Human Development			
Norway	0.956	1	1
Australia	0.946	3	9
USA	0.939	8	4
Japan	0.938	9	6
United kingdom	0.936	12	8
France	0.932	16	0
Costa Rica	0.834	45	14
2. Medium Human Development			
Russian Federation	0.795	57	3
Malaysia	0.793	59	2
Mauritius	0.785	64	15
China	0.745	94	5
Sri Lanka	0.740	96	16
India	0.595	127	10
Bhutan	0.536	134	0
Nepal	0.504	140	11
3. Low Human Development			
Pakistan	0.497	142	7
Uganda	0.493	146	4
Zimbabwe	0.491	147	25
Kenya	0.488	148	1
Nigeria	0.466	151	15
Tanzania	0.407	162	12
Zambia	0.389	164	3

Source: Human Development Report, 2004.

Of the 177 countries for which the HDI was calculated, 55 were in the high development category (with an HDI value of 0.80 or more) 86 medium category (0.5 to 0.79); and 36 in the low category (less than 0.50) Norway, Australia and USA led the HDI rankings in the high HD category. In the medium category, Bulgaria led with HDI rank of 56, Sri Lanka 96, India, 127, Bhutan 134, Bangladesh 138, and Nepal 140. In the Low category, Pakistan led with 142, Uganda 146, Zimbabwe 147, Nigeria 151, Tanzania 162 and Zambia 164. Thus the DHI reveals wide disparities in global human development. For instance, Norway's HDI value of 0.956 was more than three times of Sierra Leone’s of 0.273 which was at the bottom.

The HDI reveals that countries can have similar GDP per capita levels but different HDI values or similar HDI values but very different GDP per capita levels. Thus the HDI ranking of countries differ significantly from their ranking by GDP per capita. Countries whose GDP rank is higher than their HDI rank have considerable *potential* for distributing the benefits of higher incomes more equitably. But they have been less successful in channeling economic prosperity into better lives from their people. Of the 177 countries in 2002, there were 71 such countries whose HDI rank was lower than their GDP per capita rank. Prominent among them were Algeria (-103), India (-10), USA *(-4)*, Pakistan (-7) and Zimbabwe (-25). On the other hand, countries whose HDI rank is higher than their GDP rank, suggest that they have effectively *made use* of their incomes to improve the lives of their people. There were 106 such countries in 2002. Prominent among them were Cuba (39) and Tajikistan (45). It is said that the DHI led to the dethronement of GDP per capita. As a matter of fact, these two concepts do not measure the same thing. The HDI tries to measure the level of human capabilities and the set of choices available to people. On the other hand, GNP per capita is an indicator of well being, utility or welfare, the subjective enjoyment people get from consumption. Thus the HDI is an alternative measure of development. It supplements rather than supplants GNP per capita measure of development and provides different information from GNP per capita.

Its Limitations

- The HDI is not free from certain limitations.
- It is a *crude index* which attempts to catch in one simple number a complex reality about human development and deprivation, according to Prof. Amartya Sen.
- The three indicators are not the only indicators of human development. There can be others like infant mortality, nutrition, etc.
- The HDI measures relative rather than absolute human development so that if all countries improve their HDI value at the same weighted rate, the low human countries will not get recognition for their improvement.
- The weighting scheme for calculating the four components of HDI seems arbitrary.
- Even, giving equal (l/3rd) weight to each of the very different three indices for calculating the HDI, it is still arbitrary to the extent that one component index has a different variance than another, and equal weights seem unsatisfactory and unjustified. A country having high HDI may shift the focus from the high inequality, unemployment and poverty found within it.

Conclusion

Despite these weaknesses, by measuring average achievements in health, education and income, the HDI provides a better picture of the state of a country's development than its income alone.

3.3.1 Constructing the Human Development Index (HDI)

The HDI is based on three indicators: longevity, as measured by life expectancy at birth; educational attainment, as measured by a combination of adult literacy (two-thirds weight) and combined primary, secondary and tertiary enrolment ratios (one-third weight); and standard of living, as measured by real per capita(PPP$):

For the construction of the index, fixed minimum and maximum values have been set for each of these indicators:

- Life expectancy at birth: 25 years and 85 years for calculating the Life expectancy Index.
- Adult literacy: 0% and 100% for calculating the Education Index. Combined gross enrolment ratio (0% and 100%)
- Real GDP per capita (PPP$): $100 and $40,000 (PPP US$) for calculating GDP Index.

For any component of the HDI, individual indices can be computed by applying the formula:

$$Dimension\ Index = \frac{Actual\ Value - Min.Value}{Max.Value - Min.value}$$

1. **Life Expectancy Index.** If the life expectancy at birth of a country is 78 years, then the life expectancy index for that country would be

$$Life\ Expec\tan cy\ Index = \frac{78-25}{85-25} = \frac{53}{60} = 0.884$$

2. **Education Index.** The education Index is *th*e combination of adult literacy index and gross enrolment index. If the adult literacy rate of this country is 92, then its adult literacy index would be

$$Adult\ Literacy\ \ Index = \frac{92-0}{100-0} = 0.920$$

If the combined gross enrolment in this country is 60, then its gross enrolment index would be

$$\text{Gross Enrolment Index} = \frac{60-0}{100-0} = 0.600$$

Education Index = 2/3 (Adult Literacy Index) + 1/3 (Gross Enrolment Index) = 2/3 (0.920) + 1/3 (0.600) = 0.813

3. GDP Index. The GDP per capita (PPPUS$) of this country is $8,840, then the GDP index would be

$$GDP\ Index = \frac{\log(8{,}840) - \log(100)}{\log(40{,}000) - \log(100)} = \frac{\log 8740}{\log(100)} = 0.748$$

4. Human Development Index. The HDI is a simple average of the Life Expectancy Index, Education Index and adjusted GDP per capita (PPP$) Index.

It is derived by dividing the sum of these three indices by 3.

$$HDI = \frac{0.884 + 0.813 + 0.784}{3} = 0.815\,.$$

SELF ASSESSMENT EXERCISE

1. What is human development index?
2. What are the three component indicators for measuring quality life index?

4.0 CONCLUSION

Human development is a broad and comprehensive concept which is concerned with basic human capabilities i.e. basic human needs as with variety of human aspirations.

5.0 SUMMARY

We have been able to see that HDI is a composite index which combines measures of expectancy, school enrolment, literacy and income to provide a broader based measure of basic needs, their economic growth and development.

Remember that the composite HDI indices used to measure human capabilities are PQLI and HDI as developed by the UNDP , where PQLI shows improvement in the quality of life when people enjoy the fruits of

economic progress with increase in life expectancy. HDI is a composite index which takes into accounts real GDP per capital.

6.0 TUTOR-MARKED ASSIGNMENT

1. Economic growth and human development are closely connected. Explain?
2. What are the limitations of physical quality life index and human development index?
3. What are the three component indicators combined by human development index?

7.0 REFERENCES/FURTHER READING

Anand, S. & A. Sen (2000). "*The Income Component of the Human Development Index''*, *Journal of Human Development*, 83-106.

Anand, S. & Ravillion, M. (1993). "*Human Development in Poor Countries: on the Role of Private Incomes and Public Services''* *Journal of Economic Perspectives* 7, 133-150

Haq, Mahbubul (2005). '*The Birth of the Human Development Index' Reflections on Human Development.*

UNDP. 2002. '*A Technical Note: Calculating the Human Development Indices' Human Development Report.*

MODULE 6 MEASURES OF CORRUPTION

Unit 1 Corruption as a Menace to the Economy

UNIT 1 CORRUPTION AS A MENACE TO THE ECONOMY

CONTENTS

1.0 INTRODUCTION

An important aspect of economic analysis lies in its wealth of tools. These tools often enable economists to widen the frontiers of their analytical concerns to issues that have significant economic implications, even though they may not be strictly economic in nature. Corruption is an example of such an issue. The magnitude of corruption has varied among different societies and in different historical epochs. In contemporary times, this issue has assumed such a significant dimension as a social problem in some LCDs that discussion of it can not be ignored.

2.0 OBJECTIVES

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

- explain the meaning of corruption at all levels
- identify and determine corruption
- appreciate the effects and demands of corruption
- proffer solutions to the problems of corruption.

3.0 MAIN CONTENT

3.1 What Is Corruption?

Corruption is a word used to cover a wide range of morally offensive or criminal acts; thus, its precise definition is not easy. Gray and Kaufman define corruption as the abuse of public office for private gain. Thus, it refers to the use of position of power to seek personal advantage either by performing an act or omission of the expected performance/duty. A typical dictionary definition refers to corruption as various offences by public officials including bribery, the sale of offices, gaining of public contracts to favoured firms and individuals and granting of land or franchise in return for monetary rewards.

There is no comprehensive and universally accepted definition of corruption. The origin of the word is from the Latin *corruptus* (spoiled) and *corrumpere* (to ruin; to break into pieces). The working definitions presently in vogue are variations of "the misuse of a public or private position for direct or indirect personal gain".

This definition does point to some critical factors that come to mind when the word corruption is mentioned. Nevertheless, it is still limited. For example, the reward of the public official may be non-monetary, but this would not make the act less corrupt. A public official may deliberately connive at a violation of another person's rights by

remaining silent. This again is corruption. Indeed, without defining it exactly, most people recognise corruption when they see it.

3.2 Manifestation of Corruption

Corruption is a worldwide phenomenon which exists in all countries. In the United States, for instance, corrupt economic practices is referred to as ‘underground economy’ and in 1982, it was estimated to have accounted for about 14% of the GDP. More recently, corrupt practices bordering on crime have been uncovered in large corporations like Enron, World Com and Tyco. In the former Soviet empire, corruption was referred to as *na leva*, meaning ‘on the left’ (OLE) economy. It has been shown that during the Communist era, OLE was capable of supplying the Soviet people with their basic needs outside of the official system. The transitional period to a liberalised economic system witnessed many corrupt activities especially during the process of privatisation. ‘Smart’ businessmen engaged in ‘sharp’ practices and this led to the emergence of a minority of very rich individuals — referred to as ‘Mafia’ capitalists. In Italy, the underground economy is known to employ young people off the official record so as to escape giving them officially mandated benefits.

In recent times, Transparency International, a Berlin-based international non-governmental organisation (NGO), has been measuring corruption index. The Transparency International Corruption Index (TICI) is a measure of perception of corruption as seen by business people, risk analysts and the general public. TICI ranges from one to ten, with the index of ten reflecting 'highly corrupt'. Countries over the world have been rated in terms of their intensity of corruption. In 2003, Nigeria was rated the second most corrupt country, second only to Bangladesh while Finland was the most incorrupt, followed by Singapore, Britain, Hongkong, Germany and United States. Between 1998 and 2001, TICI, focusing on African countries showed that corruption tended to be less in Botswana, Namibia and South Africa while it was consistently high in Nigeria, Uganda, Kenya and Cameroon.

Given the pervasiveness of corruption especially in African countries, the big question is: why are people corrupt? We shall address this question theoretically by considering an economic model of corruption next.

3.2.1 An Economic Model of Corruption

Corruption can be analysed within the neo-classical utility maximising framework. In this context, an economic agent is assumed to be

maximising some net utility from engaging in corrupt activities. She/He, therefore, has a positive demand for corruption.

The individual's demand for corruption (Di^C) is determined by several factors. These factors include net gains from corruption (G), costs of undertaking corruption (C), expected punishment and probability of its occurrence (P^e), economic incentives for corruption (V), the level of income (Y^1), distribution of income (Y^d) and the level of individual education (E). Symbolically, therefore, the demand equation for corruption can be written as:

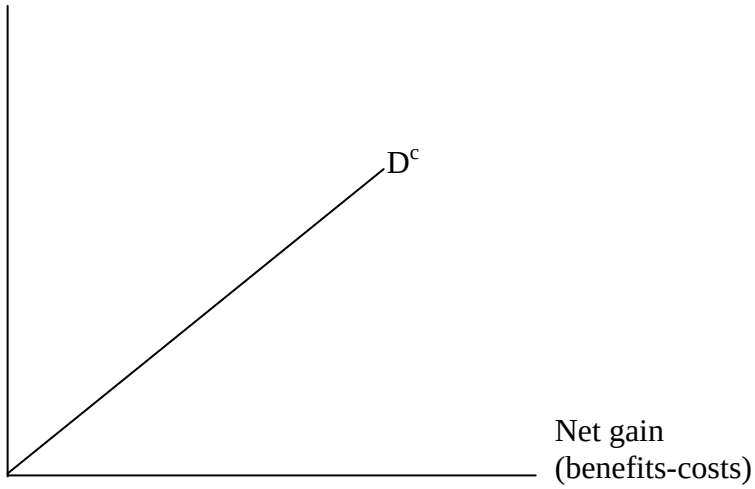
- $Di^C = f [G, C, P^e, V, Y^1, Y^d, E]$

Each of these determinants affects the demand for corruption, especially in the African context.

3.2.2 Net Gains from Corruption

Net gains from corruption are the difference between the benefits and direct costs of corruption. If the net gain is positive, corruption will take place. It then follows that the higher the net gain from corruption, the higher the degree of corruption in any given society. Under this circumstance, the demand for corruption behaves like a normal good.

Demand for corruption is, therefore, a direct function of net gains. **Dc** starts from the origin, thereby implying that, if net gains from corruption are zero, there would be no corruption. As net gains from corruption increase, more of it will be demanded by individuals in a given society, *ceteris paribus.* It must be noted that the gains from corruption may be monetary, non-monetary or psychological. Thus, some of the gains can be quantified but a good number of them cannot.



3.2.3 Costs of Engaging in Corrupt Activities

Like any other economic enterprise, corruption requires resources (inputs). These inputs constitute the cost of corruption and comprises direct and indirect (opportunity) costs.

Consider the following examples

The chairman of the tender's board, who is given the responsibility of awarding major contracts, wants to amass personal wealth by corrupt means. He will have to incur the costs of using his time and resources to lobby other members of the board and the contractors to get them to agree to his proposals. These are direct costs. The opportunity costs are the options which are given up to pursue corrupt activities. These might include negotiating a legitimate bank loan.

The corrupt official will weigh these costs *vis-a-vis* the benefits. The higher the costs of corruption, the less the demand for corruption will be In general, therefore, demand for corruption varies inversely with its costs and on a cost-corruption panel, demand for corruption is, therefore, downward-sloping. (Sketch this graph for yourself).

3.2.4 Expected Punishment

Since corruption is often a criminal act, it is punishable directly by law. The punishment can also be indirect by social disapproval and disgrace. A rational corrupt official will, therefore, assess the potential punishment, if and when apprehended, on two bases: (a) the severity, and (b) the probability.

3.2.5 Economic Incentives to be Corrupt

In an economic system, positive incentives may be created for corruption. This is often the case when governmental controls or interventions in the economic system create bureaucratic inefficiency. Consider the case of official efforts to conserve foreign exchange.

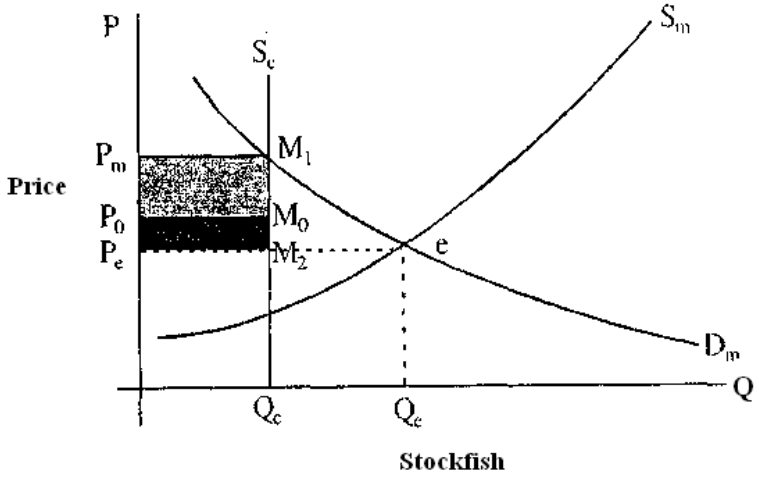


Fig. 2: Economic incentives for corruption: case of allocation of import licenses for stockfish

One of the ways of doing this is by allocation of import licenses for the importation of a commodity. Fig.2 illustrates this in the case of stockfish.

Normal profit is earned at **e** if there is no administrative control in the importation of stockfish. Now, if the government restricts importation to $\mathbf{Q_c}$ by the allocation of import licenses, people would be willing to pay $\mathbf{P_m}$. This would result in a monopoly profit of $\mathbf{P_eP_mM|M_2}$ for holders of import licenses. This in turn will generate the incentive for bribery or another form of corrupt activity to obtain the license. If a typical importer incurs a bribery cost of $\mathbf{P_ePoMoM_2}$ to obtain an import license, there would still be a profit of $\mathbf{P_mM|MoP_0}$- Since this type of import control is common in several African countries, positive incentives exist for corruption. When it is recalled that the control of one item may necessitate the control of another, the tendency for corruption to multiply is easily appreciated.

Furthermore, inefficiencies in the bureaucratic system tend to encourage mutual collaboration in corruption between the officials and the clients. To obtain an urgent passport in a congested office would be virtually impossible if the applicant does not bribe the official to speed it up. In general, therefore, administrative inefficiency in the economic system tends to generate corrupt practices because of positive incentives to bribe corrupt officials. Where corruption is extensive in an economic system, productive workers tend to be rewarded with inadequate or even negative incentives.

3.2.6 Level of Individual's Income

In an economy where individual earnings are inadequate to meet basic needs and social responsibilities, there is indirect pressure to be corrupt. This situation is common in many LDCs. Consider a top official on a salary of N2.7 million per annum. Ordinarily, this salary should be enough for family needs. But some cultures expect such an official to spend a huge amount on social functions, to educate relations, etc. These social responsibilities create pressure on officials to live beyond their means. One way of obtaining more money is through corrupt practices.
Generally, poverty is a great driver of corruption. Thus, the poorer the individual or the society, the more the incidence of corruption, *ceteris paribus.* The deepening poverty in Africa has been associated with rising incidence of corruption.

3.2.7 Income Distribution

Where distribution of income is highly inequitable, those on the lower scale often find it difficult to meet their most basic needs. The pressure to escape poverty may lead to a corrupt life. This may be abetted by the rich few who will bribe the poor to, for example, sell their votes. In this way, the rich few can use political power to maintain a system of economic inequality by corrupting the masses. This experience is quite common in most African societies with highly skewed income distribution.

3.3 Economic Effects of Corruption

Corruption would not be considered a social problem worthy of analytical and policy attention if its effects on the economic system were neutral. As we shall show presently, corruption significantly retards developing economies. Let us discuss some of these negative impacts briefly.

3.3.1 Increase in Economic Crime

Our theoretical analysis relating demand for corruption to its net gains, costs of operations and expected punishment offer a straightforward explanation for increase in criminal behaviour. This is because an economically rational official who is corrupt will always undertake related criminal activities to minimise the costs or punishment arising from corrupt behaviour. This goes towards explaining why some law enforcement agents are bribed, public buildings burnt down in order to burn the records of corrupt transactions, and murders committed to keep someone from revealing incriminating evidence. The economic costs of

these related crimes, especially those of arson and murder, can be enormous if quantified.

3.3.2 Reduction in the Supply of Public Goods

Extensive corruption may lead to a reduction in the supply of public goods. Take the case of police protection as a typical example. Suppose corruption is allowed to permeate the rank and file of the police force. This would mean that innocent citizens would have no effective protection against crimes; therefore, murderers, extortionists, rapists, etc. would thrive.

The same effect could apply to the political system whose essential role is to legislate for the public good. A corrupt political leadership can saddle the nation with making anti-developmental legislation since their intention would be that of self-protection. Here, political power becomes an instrument for reducing the cost and maximising the gains from corruption.

3.4 Corruption Indexes

Corruption indexes are measures that summarise one or more aspects of corruption-related phenomena. The characters that are the object of measurement may be aspects of corruption proper, or they may represent occurrences that are understood to be correlated with corruption, such as “public integrity” or various measures of the quality of governance. Most times, such indexes are based on perceptions of the phenomena. This is the case of Transparency International Corruption Perception Index, the result of the aggregation of several other indexes, and the Global Corruption Barometer that Transparency International commissioned to Gallup, focusing more on petty corruption.

Other indexes are based on descriptions of broad aspects of governance, and as such are more tenuously linked with the corruption phenomenon.

The three types of assessment of corruption may present some degree of simultaneity. A high public opinion exposure to media coverage on scandals may cause an artificial increase on perception-based indexes. For example, Italy’s ranking in the 1995 Transparency International CPI was below Mexico and Colombia, and just above Thailand, India and The Philippines. Such an unrealistic ranking is probably at least in part explained by the spate of attention that the corruption issue had received in Italy in the early 1990’s, corresponding to the so called “Tangentopoli” judicial inquiries. Perception-based corruption indexes could also influence corruption perception, given the wide dissemination they receive from the media, raising the possibilities that they influence

the very same perceptions on which they are based. Judicial activity may also be influenced by the perception of corruption, either of the informal type, or as expressed by an index, wherever the judiciary has degrees of freedom in allocating its effort.

3.5 Solutions to Corruption

Corruption everywhere is usually frowned upon as immoral and criminal. Public condemnation of corruption in Nigeria, for instance, dates back to the colonial era. In its Civil Service General Orders, the colonial government stipulated that any employee found guilty of corruption would be dismissed. The Nigerian Criminal Code treats corruption as a criminal offence punishable under the law. When the late General Murtala Mohammed took over power in 1975, he purged the country's public service of ‘'corrupt’ officials through mass dismissals.

As a follow up, the Code of Conduct and Public Complaints Commission were set up with branches throughout all the states of the Federation to fight corruption. The 1979 and 1999 Constitutions of Nigeria make it mandatory for public office holders to declare their assets before taking up office and at the end of their tenure. In 1982, President Shehu Shagari set up the Ethical Reorientation Committee to study the problem of a drastic fall in ethical standards among Nigerians in order to usher in an ethical revolution. In 1999, the government set up the Independent Corrupt Practices Commission (1CPC) and the Economic and Financial Crimes Commission (EFCC) to combat corruption.

It is a matter for concern that despite these attempted solutions, corruption is perceived to be on the increase in Nigeria. Indeed, ‘corruption paradox’ whereby its incidence varies with anti-corruption measures seems to hold sway. Most people agree that it poses a fundamental threat to the survival of the Nigerian economy. This view may not be far from the truth if corruption is found to be endemic or systemic.

One of the reasons for the failure of existing policies is the lack of economic focus. Corruption is seen mainly in moralistic and legalistic terms; hence the proposed solutions are morally and/or legally oriented. Our theoretical analysis, however, suggests that corruption is essentially an economic problem and, consequently, effective solutions must be looked for in the economic realm. Besides, the extant solution approaches tend to treat the symptoms rather than the fundamental causes of corruption. Success in tackling corruption needs to be both preventive and therapeutic.

3.5.1 Reduction of Net Gains to Corruption

Since net gains are what lure people into corruption, these can be reduced or completely eliminated by stepping up the costs of corruption. Such costs should include the public condemnation of ill-gotten wealth and the plugging of legal loopholes that corrupt officials tend to exploit. A non-police intelligence network comprising proven patriots can help in checking corrupt officials. To encourage efficiency, such patriotic volunteers should be rewarded by government through a ‘whistle blowing’ programme. Thus, whoever 'blows the whistle’ to alert authorities of corruption deals can be rewarded handsomely.

3.5.2 Increase in the Probability of Punishment

More resources should be given to law enforcement agents to enable them to improve their efficiency in apprehending and punishing corrupt officials. It follows logically that the higher the probability of getting caught and punished after committing the crime, the lower will be the incidence of corruption. In particular, stiff penalties should always be meted out to corrupt officials as a deterrent.

3.5.3 Elimination of Corruption Incentives and Uneconomic Subsidies

Incentives for corruption can be eliminated or minimised in the economic system through less administrative controls of the market process. Where such controls are inevitable, efficient administrative machinery should be set up in order to prevent the operation of the 'underground economy'. There is need to review the system of subsidies to eliminate those that encourage corruption.

3.5.4 Fighting Poverty

Since corruption thrives on absolute poverty, one of the solutions lies in the reduction of poverty. The most effective way of doing this is to adopt the basic needs approach to development. This strategy ensures that basic human needs like adequate food, water, shelter, medical facilities, etc. are provided for the masses. This will render them less susceptible to the corrupting influence of the rich or corrupt politicians. A market-based approach to ensuring provision of basic needs is to promote productive employment described below.

3.5.5 Education

Efforts should be made to raise the literacy rate to enable citizens to gain knowledge about their rights in society. The basic school curricula

should include moral instruction, ethics and civil rights. Corruption should be a subject in the curricula of schools with emphasis on accountability.

3.5.6 Private Provision of Public Goods

Experience has shown that governments' efforts to provide public goods are often prone to corruption. This can be avoided by contracting the provision of public goods like electricity, water supply, roads, etc, to private limited liability companies. Corruption would be minimised under this arrangement because of their relative efficiency compared to the government. To avoid exploitation of the public by these private monopolies, an effective regulatory mechanism should, however, be put in place.

3.5.7 Promoting Productive Employment

Productive employment means getting a job or means of livelihood that yields sustainable income streams for meeting people's basic needs. A high rate of unemployment, underemployment or low-wage employment tends to exacerbate corruption. A programme for promoting job creation or self-employment should receive the priority attention of governments. Having a legitimate means of livelihood will tend to enlist support of the public for the fight against corruption. This is because productively employed/ engaged people have a stake in living in a corruption-free environment.

3.5.8 Efficient Bureaucracy and Economic Management

Improvements in bureaucratic efficiency as well as general management efficiency will tend to eliminate corruption. This is because merit and productivity would be the yardsticks for reward. Improvement in productive efficiency, especially of the industrial system, can indirectly reduce the incentives to smuggling since product quality would be enhanced and the prices reduced.

3.5.9 Reduction in Income and Wealth Inequality

Measures to reduce distributional inequality such as taxation and income transfer programmes to the poor, would minimise corruption. If the income gap is drastically reduced, the urge to get rich by corrupt means can be minimised correspondingly.

3.5.10 Corruption Monitoring

Agencies set up by government to monitor corruption face many problems that have undermined their effectiveness. If government is serious about fighting corruption, agencies such as ICPC, EFCC, etc, should not only be staffed with highly dedicated and motivated professionals but should be adequately funded and resourced. They should in turn fight corruption not only through the demand (therapeutic) but through the supply strategies. From the supply perspective, its focus should be on undertaking rigorous research-probing the fundamental causes of corruption and setting up appropriate internal corruption monitoring mechanism. Such a monitoring system can highlight the shortcomings of an externally located monitoring such as Transparency International. It would also contribute to enriching the perspective of TICI findings.

SELF ASSESSMENT EXERCISE

Look at the statements below and decide if they are *true or false.*

1. Corrupt administrative practices are more likely to be found in a state which takes on functions that it does not have the capacity to carry out.
2. Shortages of essential commodities because of inadequate foreign exchange are likely to lead to unofficial market arrangements sometimes operating outside the law.
3. Socialist societies rarely have unofficial markets.
4. Unofficial markets are a rational economic response to price control during times of shortages.
5. Where import tariffs are low and demand for imports high, there is less likely to be unofficial markets in imported commodities.

4.0 CONCLUSION

Corruption has been known to exist in various forms in both developed and developing societies as well as in all ideological systems.

5.0 SUMMARY

In this unit we have learnt the followings.

- Demand for corruption is determined by a complex set of economic forces. These include net gains from corruption, costs of engaging in corrupt acts, expected punishments, existence of

economic incentives for corruption, level of individual income, income distribution and level of education

- The effects of corruption include an increase in economic crime, reduction in the supply of public goods, unemployment , increased costs of investments, balance of payments deficits, ineffective economic policy, income inequality, low utilisation of manpower, loss of human capital including lives, foreign debt burden, increased capital flight, reduced inflow of foreign direct investment and political instability, increased poverty and reduced economic growth.
- Solutions to corruption should focus on the moralistic or legalistic as well as therapeutic angles
- The economic approaches worth considering include: reduction in net gains of corruption by increasing its cost of operation, increase in the probability and severity of punishment, elimination of corruption incentives in the socio-bureaucratic system; fighting absolute poverty by provision of basic needs for the masses; provision of mass education to increase awareness to social rights; promotion of productive employment; efficient economic and bureaucratic management; reduction in income inequality.

6.0 TUTOR-MARKED ASSIGNMENT

1. Identify and discuss the major economic determinants of demand for corruption.
2. To what extent can one correctly assert that corruption poses a threat to your country's development?
3. Analyse the economic policies that can be adopted as solutions to the problem of corruption in your country. What problems can you anticipate in the implementation of these policies?

7.0 REFERENCES/FURTHER READING

African Development Bank (ADB), *African Development Report 2001*, Abidjan: ADB, Chapter 3.

Becker, G. S. & Laudes, W. M. (1974). *Essays in Economics of Crime and Punishment.* New York: Columbia University Press for National Bureau of Economic Research.

Ehrlich, (1973). 'Participation in Illegitimate Activities: A Theoretical and Empiric Investigation', *Journal of Political Economy*, Volume 81, pp. 521-56.

Odekunle, F. (1983). (ed.) *Nigeria: Corruption and Development.* Ibadan University Press, Ibadan.

Rose-Ackerman, S. (1978). *Corruption: A Study in Political Economy*. New York: Academic Press.

MODULE 7 ECONOMIC LIBERALISATION

Unit 1 The Meaning of Economic Liberalisation

UNIT 1 THE MEANING OF ECONOMIC LIBERALISATION

CONTENTS

1.0 INTRODUCTION

Economic Liberalisation is a very broad term that usually refers to fewer government regulations and restrictions in the economy in exchange for greater participation of private entities; the doctrine is associated with neoliberalism. The arguments for economic liberalisation include greater efficiency and effectiveness that would translate to a "bigger pie" for everybody.

Most first world countries, in order to remain globally competitive, have pursued the path of economic liberalisation: partial or full privatisation of government institutions and assets, greater labour-market flexibility, lower tax rates for businesses, less restriction on both domestic and foreign capital, open markets, etc.

In developing countries, economic liberalisation refers more to liberalisation or further “opening up” of their respective economies to foreign capital and investments. Three of the fastest growing developing economies today; China, Brazil and India, have achieved rapid economic growth in the past several years or decades after they have “liberalized” their economies to foreign capital.

2.0 OBJECTIVES

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

- define liberalisation as an economic concept
- explain liberalisation and its benefits to the economy
- identify the objectives of competition policy and law
- identify the impact of liberalisation on the workforce.

3.0 MAIN CONTENT

3.1 The Definition of Liberisation

Liberalisation means the opening up of markets within and between countries to promote "free trade", i.e. reducing barriers to trade such as tariffs (import and export duties) and other forms of regulation. Liberalisation also, in many circumstances, means freedom to exploit, due to this reduction in regulations and standards that are universally adhered to.

3.2 What Economic Liberalisation and Globalisation Means for the Masses

The subject of “Economic Liberalisation” and “Globalisation” has attracted the attention of the people both at home and abroad. A layman speculates “Economic Liberalisation” as a subsidy granted by the Government to big houses. Economic liberalisation means modernisation, upgrading of technology, large production, development of sunrise industries like electronics and improvement in the working of public sector enterprises so that the levels of productivity may be improved. All this would come about with the cordial principle of liberalisation.

Under “Economic Liberalisation”, a review of the various economic controls has been made and counter productive controls are being dispensed with. The industrial licensing policy, trade policy, taxation policy or, in a broader context, the entire fiscal policy, monetary policy, agricultural policy etc. are being cast in a manner so as to sub serve the goals of higher production and productivity.

These liberalisations are designed to initiate measures aimed at increasing international competitiveness of developing country’s goods and services and hence their exports. They also encourage in flow of foreign exchange into their private foreign direct investment. The economic Liberalisation policy has provided for encouraging private foreign direct investment in setting up joint ventures with their business

houses in the field of trading with equity contributions of up to 51% in these ventures.

The shift of emphasis in the economic liberalisation is towards greater reliance on the private sector for increasing the production in future. For this purpose, economic controls and regulations have been gradually relaxed and will be further relaxed with the passage of time. The liberalisation policy emphasises economics of large-scale production, modernisation and competition for production in the economy.

Although the process of economic liberalisation had commenced in 1973, but this process gathered momentum and it is difficult to predict the future shape of things in this regard. The direction is now more certain and the speed is faster towards economic Liberalisation in developing country. Broadly speaking, liberalisation policy is a move from the public sector towards the private sector. It emphasises production and productivity maximisation goals, rather than the mere idealistic talk of socialistic pattern of society.

The true meaning of the term “globalisation,” its core contents, and its impact on the economies and societies in both developed and developing countries, have been a vigorously contested subject for a long time now. This debate reminds one of the stories of the elephant and the seven blind men. Each person touched a separate part of the pachyderm’s anatomy, and depending upon the contours of that part, described the complete animal, with his imagination taking over where his visual faculty left off. The message of this allegory is not that the narratives were inherently flawed in their analytical conception. Rather, that globalisation is dynamic, so complex and still unfolding that one account of it could be at significant variance with another, and yet both might be offering true but partial accounts of the overall phenomenon.

Indeed, in contemporary international discourse, this term is often bandied about as a new phenomenon. In reality, globalisation has long been a feature of the international system, whether it relates to transmigration of people, cross-fertilisation of ideas, or even trans-mutation of technologies. What has changed distinctively in recent decades is the pace and scope of globalisation. And as its most powerful manifestation, advanced technology runs through the increasing jigsaw of the international system as a lubricant, imparting its liquidity as well as dynamism. Thus, the distinctive aspect of contemporary globalisation is the increase in its liquidity, which makes it a phenomenon, capable of tremendous progress but also of considerable harm.

Further, despite the controversy, most experts agree that at its core, globalisation denotes the international integration of the basic factors of production, i.e. labour, capital, technology, and entrepreneurship, and their output, i.e. goods and services. As national economies become increasingly integrated into the global economic matrix, it blurs the traditional boundaries between local, national and trans-national, and impinges crucially upon the latitude, jurisdiction, and sovereignty of state actors in the international system.

By comparison, the concept of "economic liberalisation" is relatively simpler to comprehend, even if not easy for governments to implement and sustain over the time. In essence, economic liberalisation denotes an attitudinal change along with changes in the structures and processes of an economy that was hitherto "closed" to the international markets. In attitudinal terms, this means change in the role of the government from one of a "chief controller" to that of "facilitator" and an "arbiter" of competing economic groups and interests within a country. The structural and procedural derivatives of liberalisation signify the entire gamut of changes that result from the deliberate policies of governments to permit a greater role for market forces in the functioning of economy.

The economic liberalisation policy is within the parameters of change allowed in the economic management of the nation. It will help in improving the growth performance of the economy and will reduce poverty and unemployment in the country.

3.3 Basic Assumptions

That free trade, competition, and the laws of supply and demand will regulate the market sufficiently well to provide economic growth, employment and prosperity for all.
Inefficiency will be reduced, as companies have to be efficient to stay competitive.

Participation in the global economy is expected to raise the living standards in poor countries as MNC investment brings jobs, monetary income and an increase in consumer goods.

"Comparative advantage" is a theory which states that nations should specialise in producing what they are best at producing, and that they should then trade with other nations.

Export-orientated growth should be encouraged.

3.3.1 Basic Problems

Quite clearly, liberalisation has not brought prosperity or employment for all. This may be partly because it has not been applied fairly but it is also because not everyone starts on an equal footing. In a world with reduced trade barriers, small companies cannot hope to compete fairly with the economic might, technical expertise and marketing power of huge Multinational Corporations (MNCs). Thus true competition gives way to monopolisation in many cases.

3.4 Trade Liberalisation and Competition Policy

The process of economic liberalisation started in several countries across the world during 1980s and 1990s. This was when most of the countries adopted policies of deregulation, privatisation and trade liberalisation. This was also the time when the World Trade Organisation (WTO) came into being. The WTO is a bundle of several agreements on goods and services, with various rules on how they are produced and traded, ensuring that there is healthy competition in the global market place. However, this is not enough, and therefore, at times, there was strong demand for competition rules to be integrated in the international trading system.

Trade and economic liberalisation aided competition in the market, by increasing the basket of goods and services with better quality and lower prices. Yet, anti-competitive practices undertaken by economic players or induced by inappropriate government policies have negated the gains of liberalisation. With the adoption of market oriented reforms, there were several players in the market resulting in competition. But at the same time, many abuses too creep into the system. In order to balance the system, price controls and market regulations were put in place. But by and large they proved ineffective. This is why countries adopted competition and economic regulatory laws to promote a healthy market and economic democracy.

The main objective of competition policy and law is to preserve and promote competition as a means of ensuring the efficient allocation of resources in an economy. This should eventually result in fair prices and adequate supplies for consumers and, it is hoped, faster growth and a more equitable distribution of income. While going through a phase of privatisation and deregulation, often monopoly power is transferred from the public to the private sector which harms the interests of consumers, especially the poor. This situation requires an appropriate regulatory mechanism. International anti-competitive practices can also be harmful to small and developing countries without effective competition laws. Cross-border mergers and acquisitions that lead to

market dominance and the restrictive practices that some trans-national corporations engage in, also further necessitate competition rules. Competition policy also promotes good governance in the corporate sector as well as in governments by diminishing the opportunities for rent-seeking behaviour and the corruption that often accompanies it. Competition law and regulatory tools are invoked mainly to take care of firms' behaviour and market failures.

Governments often intervene when markets fail but in the absence of a clearly defined competition policy and regulatory mechanisms, the intervention can be arbitrary and serve vested interests rather than the poor. It regulates the economic activity within the country by protecting business and consumers against abuses of economic power and promoting efficiency and consumer welfare.

Further, competition law provides necessary safeguard against arbitrary decision-making, apart from dealing with anti-competitive business practices by such investors. Investment through the M&A route is another area where there is significant interaction between investment and competition policy.

3.5 Impact of Liberalisation on the Workforce

Considering the Indian Economy: the liberalisation of the Indian economy started in 1991, with the government's attempts to decontrol the economy and open Indian markets to the rest of the world. Its perceived or projected impact was certainly faster economic progress for the country. Its needs were debated and re-debated for long by everyone in the intelligentsia. All those who considered themselves as custodians of values, culture, and nationalism, voiced its likely impact. All those who considered themselves as saviours of the workforce, debated its impact on likely unemployment. For long, it remained an important topic at top B & E schools for group discussions. However, during this phase of debating the pros and cons of liberalisation, one thing which I did not hear, was "How will it impact the workforce?"
Today, more than a decade after the start of liberalisation, this impact is being debated.

Visible outcomes

(i) Accelerated movement of the workforce to so called new economy and geographic dislocations.
(ii) More choices/options (to save, spend, or store).
(iii) Higher speed of change (products, people, or possibilities).
(iv) Survival of the powerful (money, manipulation, muscle). Here the word powerful has replaced 'the fittest'.
(v) Higher degree of information.

Each one of these areas had visible outcomes and has affected/impacted the workforce both positively and adversely in the following areas.

a. Competency Level

The old manual skill based worker dominance is changing to knowledge based worker dominance. It is changing the whole process of supervision and governance. Further, rate of redundancy or competency is increasing. Required skill sets change with changing technology. In order to remain current and employable, the workforce is constantly on the look out for latest skill sets. As a result, their attention is divided in doing the job and acquiring new skills. This is leading to the demand for being associated with newer technologies and unidirectional job rotations. This has also led to job-hopping at faster rate and lack of workforce with mature or high levels of competency.

At the same time, it is becoming a knowledgeable world today; more and more people are becoming domain experts. Along with high competency levels in specific domains, the workforce today is extremely flexible and has phenomenal capacity to adapt or shift its domain expertise. Outsourcing the most talked of and used phenomenon today has resulted from economies of volumes and domain expertise in specific areas.

The workforce today is more informed and is also well-versed with the latest trends and technological value-adds, it is thereby in a position to provide better services and efforts. However, faster rate of technology adoption than assimilation has created a sharp divide in the competency base of haves and have-nots. We are coming into an era where “a specialist in a specialised field is transforming into a generalist who is capable of performing specialist activity.”

b. Work Habits

Changing work environment, personal demands and social fabrics are causing changes in work habits. Further, the fear of losing a job amongst knowledgeable workers makes them work more as individuals rather than teams. Availability of information through net-surfing is making knowledgeable workers to adopt cut and paste methodology rather than being genuine and creative.

Due to shorter service spans, the habit of managing the situation rather than managing job with mid and long-term perspective is creeping in. Demand to “prove fast or move” by employers due to high competition and cost pressures coupled with inadequate business monitoring scenario (where overall enterprise management, strict process and legal

controls are still not mature) at times force the workforce to adopt shortcuts.

On the positive note, the cut and paste habits can be seen as a habit of plug and play components. This enables the informed and intellectual workforce to produce innovative creations.

In this intensely competitive era, the rat race for being "the first mover" has reached its peak. This has created desire amongst the workforce to be more innovative than repetitive when it comes to their work. Habit of reinventing the wheel is disappearing. Research reports have revealed that this "well informed and current in every aspect of the workforce" is able to do a mask and portray themselves in an extremely positive light, thus easily faking expertise and experiences at selection. This, at times, results in poor performance and is a great source of dismay for employers. Today's workforce is ready to don the mantle of "Entrepreneur" and go all out for its goals. Delegation and participative management with flattering organisational structures are becoming today's mantras to suit current work habits. Though this brand of "entrepreneurship" is welcome, it comes coupled with increased levels of indiscipline and disrespect to organisational systems or procedures and policies.

c. Psychological Impact

Liberalisation has impacted longevity of business and product cycles adversely. This has led to quicker starting and winding up of businesses. Automation has improved productivity and made many jobs redundant. Thus the workforce is under a constant threat of:

- Loss of jobs—high rate of downsizing, right sizing;
- Faster rate of skill redundancy (increasing need of new skill sets);
- Hire and fire policies;
- Continuous struggle to prove one's worth;
- Transfer to new geographic locations;
- Negative changes in remuneration and benefits;
- Constantly changing peers, supervisors and subordinates (working with new people).

Even the stars of recent past are highly insecure and unsure of themselves. This all has given rise to high degree of stress and stress related psychological disorders.
Liberalisation has resulted in a wholly new "call centre industry" that has opened its doors to the younger generation. Here employees and employers face a totally different set of challenges.

SELF ASSESSMENT EXERCISE

1. What do you understand by the following terms:

(a) Liberalisation?
(b) Globalisation?

2. What are the main objectives of competition policy and law?

4.0 CONCLUSION

Liberalisation and globalisation make developing countries dynamic, strong and progressive. Competition policy plays an important role in removing barriers to market entry for foreign investors.

5.0 SUMMARY

In this unit we have learnt that;

- liberalisation brings about free trade, efficient competition , faster economic growth, employment, privatisation, high standard of living and improvement on poverty level
- globalisation leads to international integration of labour, capital, land entrepreneurship and technology together with their output
- liberalisation has both positive and negative impact on the workforce in the following area: competency level, work habit e.t.c.

6.0 TUTOR-MARKED ASSIGNMENT

1. Why do many countries partake in liberalisation?
2. What is the impact of liberalisation on the workforce in Nigeria

7.0 REFERENCES/FURTHER READING

Bhagwati, J. (2002). Free Trade Today, *Princeton: Princeton Univ. Press.*

Dickerson O. M. (1951). *The Navigation Acts and the American Revolution.* Philadelphia: University of Pennsylvania Press140.

Landsburg, S E. *Price Theory and Applications.* Sixth Edition

Mankiw. G. *Macroeconomics.* Fifth Edition.

Pugel, T. A. *International Economics* 13th edition.

Stockamn, A. C. *Introduction to Economics*. Second Edition.