



**NATIONAL OPEN UNIVERSITY OF NIGERIA**

**SCHOOL OF BUSINESS AND HUMAN RESOURCES**

**COURSE CODE: BHM 710**

**COURSE TITLE: BUSINESS STATISTICS**

**COURSE  
GUIDE**

**BHM 710: BUSINESS STATISTICS**

*Course Writers:*

**Dr. Onyeamachi Joseph Onwe**

National Open University of Nigeria, Lagos

*Content Editor:*

**Dr. D.I. Mai-Lafia**

Department of Economics

University of Jos, Nigeria

*Course Coordinator:*

**Mrs. O. Inua**

School of Business and Human Resource Management

National Open University of Nigeria

Victoria Island, Lagos

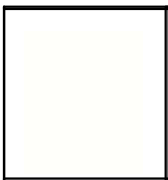
*Programme Leader:*

**Dr. D.I. Mai-Lafia**

School of Business and Human Resource Management

National Open University of Nigeria

Victoria Island, Lagos



NATIONAL OPEN UNIVERSITY OF NIGERIA



**NATIONAL OPEN UNIVERSITY OF NIGERIA**

**National Open University of Nigeria**

***Headquarters:***

14/16 Ahmadu Bello Way  
Victoria Island, Lagos

***Abuja Annex:***

5, Dar'el Salam Street, (off Aminu Kano Crescent)  
Wuse II, Abuja

e-mail: [centralinfo@nou.edu.ng](mailto:centralinfo@nou.edu.ng)

URL: [www.nou.edu.ng](http://www.nou.edu.ng)

National Open University of Nigeria 2009

First Printed 2009

ISBN

All Rights Reserved

Printed by .....

For

National Open University of Nigeria

## **Introduction**

The current business environment has continuously been leaning toward quantitative techniques, coupled with the emergence of computers. It has been noted that within the next few decades, no manager will be able to operate effectively without some knowledge of quantitative techniques employing the use of statistical methods. It is on this observation that this course has been designed for managers beginning their trip to managerial excellence. Business Statistics as a course enables managers perform the basic managerial functions of forecasting, planning, and controlling.

The course has been conveniently arranged for you in fifteen distinct but related units of study activities. In this course guide, you will find out what you need to know about the aims and objectives of the course, components of the course material, arrangement of the study units, assignments, and examinations.

## **The Course Aim**

The course is aimed at acquainting students with the principles of Business Statistics and the way can be useful in solving quantitative managerial problems. By so doing, it will bridge the gap between knowledge and application of statistical principles among today's managers. To ensure that this aim is achieved, some important background information will be provided and discussed, including:

1988	Definitions of Statistics
1989	Data and Data Collection
1990	Stages of Statistical Inquiry
1991	Expected Values
1992	Index Numbers

## **The Course Objectives**

On completion of the requirements of this course students and managers alike will be able to:

- 3 Appreciate the uses and importance of statistics in business decision making;
- 4 Know how the application of the principles of statistics can aid in the achievement of business objectives;
- 5 Understand how to make business decisions under uncertainties;

- 6 Be equipped with tools necessary in business forecasting;
- 7 Be equipped with tools needed for standardization of monetary values associated with contracts; and,
- 8 Understand and apply modern quantitative techniques in management decisions.

### **Composition of the Course Material**

The course material package is composed of:

- (a) The Course Guide
- (b) The Study Units
- (c) Self-Assessment Exercises
- (d) Tutor-Marked Assignments
- (e) References

### **The Study Units**

The study units are as listed below:

## **UNIT 1: DEFINITION OF STATISTICS, STATISTICAL DATA, STATISTICAL INQUIRY, AND USES OF STATISTICS**

### **Content**

- 1.0 Introduction
- 2.0 Objectives
- 3.0 Definitions, Data, Statistical Inquiry, and Uses of Statistics
  - 3.1 Definitions
  - 3.2 Data, Sources of Data, and Methods of Data Collection
  - 3.3 Stages of Statistical Inquiry
  - 3.4 Uses of Statistics in Business and Economics
  - 3.5 Self-Assessment Exercise
- 4.0 Conclusion
- 5.0 Summary
- 6.0 Tutor-Marked Assignment
- 7.0 References

## **UNIT 2: STATISTICAL POPULATION, SAMPLING TECHNIQUES, AND DETERMINATION OF SAMPLE SIZE**

**Content**

- 1.0 Introduction
- 2.0 Objectives
- 3.0 Statistical Population, Sampling, and Determination of the Sample Size
  - 3.1 Differentiating Population from a Sample
  - 3.2 Sampling Techniques
  - 3.3 Determination of the Sample Size
  - 3.4 Self-Assessment Exercise
- 4.0 Conclusion
- 5.0 Summary
- 6.0 Tutor-Marked Assignment
- 7.0 References

**UNIT 3: MEASURES OF CENTRAL TENDENCY****Content**

- 1.0 Introduction
- 2.0 Objectives
- 3.0 The Measures of Averages or Central Tendency
  - 3.1 Arithmetic Mean
  - 3.2 The Median
  - 3.3 The Mode
  - 3.4 Self-Assessment Exercise
- 4.0 Conclusion
- 5.0 Summary
- 6.0 Tutor-Marked Assignment
- 7.0 References

**UNIT 4: MEASURES OF DISPERSION AND SKEWNESS****Content**

- 1.0 Introduction
- 2.0 Objectives
- 3.0 Measures of Variation and Skewness
  - 3.1 Measures of Dispersion or Variation
  - 3.2 Measures of Skewness
  - 3.3 Self-Assessment Exercise
- 4.0 Conclusion
- 5.0 Summary

6.0 Tutor-Marked Assignment

7.0 References

## **UNIT 5: PRESENTATION AND ANALYSIS OF BUSINESS DATA: ANALYTICAL TOOLS (ONE)**

### **Content**

1.0 Introduction

2.0 Objectives

3.0 Preliminaries and Tools of Data Analysis

3.1 Preliminaries

3.2 Tools for Data Presentation and Analysis

3.3 Self-Assessment Exercise

4.0 Conclusion

5.0 Summary

6.0 Tutor-Marked Assignment

7.0 References

## **UNIT 6: PRESENTATION AND ANALYSIS OF BUSINESS DATA: ANALYTICAL TOOLS (TWO)**

### **Content**

1.0 Introduction

2.0 Objectives

3.0 The Z-Statistic and Non-Parametric Tools of Analysis

3.1 The Z-Statistic

3.2 The Non-Parametric Tools

3.3 Self-Assessment Exercise

4.0 Conclusion

5.0 Summary

6.0 Tutor-Marked Assignment

7.0 References

## **UNIT 7: PROBABILITY THEORIES AND BUSINESS APPLICATIONS (ONE)**

### **Content**

1.0 Introduction

2.0 Objectives

3.0 The Concept of Probability, Theory of Sets, and Definitions of Probability

3.1 The Concept of Probability

3.2 The Theory of Sets

3.3 Definitions of Probability

3.4 Self-Assessment Exercise

4.0 Conclusion

5.0 Summary

6.0 Tutor-Marked Assignment

7.0 References

## **UNIT 8: PROBABILITY THEORIES AND BUSINESS APPLICATIONS (TWO)**

### **Content**

1.0 Introduction

2.0 Objectives

3.0 Laws of Probability, Computation of Probabilities, the Bayes Theorem, and Expected Values

3.1 Laws of Probability

3.2 Computational Formula for Multiple Occurrence of Events

3.3 Joint, Marginal, Conditional Probabilities and the Bayes Theorem

3.4 The Bayes Theorem

3.5 Probability and Expected Values

3.6 Self-Assessment Exercise

4.0 Conclusion

5.0 Summary

6.0 Tutor-Marked Assignment

7.0 References

## **UNIT 9: COUNTING PRINCIPLES: COMBINATIONS AND PERMUTATIONS**

### **Content**

1.0 Introduction

2.0 Objectives

3.0 Definitions, Computation, and Applications of Combinations and Permutations

3.1 Definitions of Combinations and Permutations

3.2 Combinations and Permutations Formula



3.3 Applications of Permutations and Combinations to the Probability Theory  
3.4 Self-Assessment Exercise  
4.0 Conclusion  
5.0 Summary  
6.0 Tutor-Marked Assignment  
7.0 References

## **UNIT 10: DISTRIBUTION OF BUSINESS DATA: THE BINOMIAL AND POISSON DISTRIBUTIONS**

### **Content**

1.0 Introduction  
2.0 Objectives  
3.0 The Binomial and Poisson Distributions  
3.1 The Binomial Distribution  
3.2 Poisson Distribution  
3.3 Poisson Approximation to the Binomial Distribution  
3.4 Self-Assessment Exercise  
4.0 Conclusion  
5.0 Summary  
6.0 Tutor-Marked Assignment  
7.0 References

## **UNIT 11: THE NORMAL DISTRIBUTION AND CONFIDENCE LIMITS**

### **Content**

1.0 Introduction  
2.0 Objectives  
3.0 The Normal Distribution, Normal Approximation to Binomial Distribution, and Confidence Limits  
3.1 The Normal Distribution  
3.2 Normal Approximation to Binomial Distribution  
3.3 Confidence Limits  
3.4 Self-Assessment Exercise  
4.0 Conclusion  
5.0 Summary  
6.0 Tutor-Marked Assignment  
7.0 References

## **UNIT 12: INVENTORY CONTROL**

### **Content**

- 1.0 Introduction
- 2.0 Objectives
- 3.0 Inventory Costs, Inventory Control Systems, and Inventory Models
  - 3.1 Inventory Costs
  - 3.2 Important Terminologies in the Inventory Control Theories
  - 3.3 The Inventory Graph
  - 3.4 Inventory Control Systems
  - 3.5 Inventory Control Models
  - 3.6 Self-Assessment Exercise
- 4.0 Conclusion
- 5.0 Summary
- 6.0 Tutor-Marked Assignment
- 7.0 References

## **UNIT 13: DECISION ANALYSIS**

### **Content**

- 1.0 Introduction
- 2.0 Objectives
- 3.0 Decision Making Under Uncertain Conditions
  - 3.1 Certainty and Uncertainty in Decision Analysis
  - 3.2 Analysis of the Decision Problem
  - 3.3 Expected Monetary Value Decisions
  - 3.4 Decision Making Involving Sample Information
  - 3.5 Self-Assessment Exercise
- 4.0 Conclusion
- 5.0 Summary
- 6.0 Tutor-Marked Assignment
- 7.0 References

\

## **UNIT 14: BUSINESS FORECASTING**

### **Content**

- 1.0 Introduction
- 2.0 Objectives

- 3.0 Steps and Techniques of Forecasting
- 3.1 Steps in Forecasting
- 3.2 Types of Forecasts
- 3.3 Methods of Forecasting
- 3.4 Self-Assessment Exercise
- 4.0 Conclusion
- 5.0 Summary
- 6.0 Tutor-Marked Assignment
- 7.0 References

## UNIT 15: INDEX NUMBERS

### **Content**

- 1.0 Introduction
- 2.0 Objectives
- 3.0 Definitions, Classifications, and Applications of Price Indices
- 3.1 Definitions
- 3.2 Classifications and Application of Price Indices
- 3.3 Self-Assessment Exercise
- 4.0 Conclusion
- 5.0 Summary
- 6.0 Tutor-Marked Assignment
- 7.0 References

### Assignments

Each unit of the course has a self-Assessment exercise. You will be expected to attempt them as this will enable you understand the content of the unit.

### **Tutor-Marked Assignment**

The tutor-Marked Assignments at the end of each unit are designed to test your understanding and application of the concepts learned. It is important that these assignments are submitted to your facilitators for assessments. They make up 30 percent of the total grading score for the course.

### **Final Examination and Grading**

At the end of the course, you will be expected to participate in the final examinations as scheduled. The final examination constitutes 70 percent of the total grading score for the course.

### **Summary**

This course, BHM 710: Business Statistics, is ideal for today's global business environment. It will enable you apply statistical principles in such business functions as planning, implementing, controlling, forecasting, and evaluation. Having successfully completed the course, you will be equipped with the latest global knowledge on business decisions. I bet you will enjoy the course. Good luck.

**MAIN  
COURSE**

**BHM 710: BUSINESS STATISTICS**

*Course Writers:*

**Dr. Onyeamachi Joseph Onwe**  
National Open University of Nigeria, Lagos

*Content Editor:*

**Dr. D. I. Mai-Lafia**  
Department of Economics  
University of Jos, Nigeria

*Course Coordinator:*

**Mrs. O. Inua**  
School of Business and Human Resource Management  
National Open University of Nigeria  
Victoria Island, Lagos

*Programme Leader:*

**Dr. D.I. Mai-Lafia**  
School of Business and Human Resource Management  
National Open University of Nigeria  
Victoria Island, Lagos



**NATIONAL OPEN UNIVERSITY OF NIGERIA**



**NATIONAL OPEN UNIVERSITY OF NIGERIA**

**National Open University of Nigeria**

***Headquarters:***

14/16 Ahmadu Bello Way  
Victoria Island, Lagos

***Abuja Annex:***

5, Dar'el Salam Street, (off Aminu Kano Crescent)  
Wuse II, Abuja

e-mail: [centralinfo@nou.edu.ng](mailto:centralinfo@nou.edu.ng)

URL: [www.nou.edu.ng](http://www.nou.edu.ng)

National Open University of Nigeria 2009

First Printed 2009

ISBN

All Rights Reserved

Printed by .....

For

National Open University of Nigeria

## **BHM 710: BUSINESS STATISTICS**

### **UNIT 1: DEFINITION OF STATISTICS, STATISTICAL DATA, STATISTICAL INQUIRY, AND USES OF STATISTICS**

#### **Content**

- 1.0 Introduction
- 2.0 Objectives
- 3.0 Definitions, Data, Statistical Inquiry, and Uses of Statistics
- 3.1 Definitions
- 3.2 Data, Sources of Data, and Methods of Data Collection
- 3.3 Stages of Statistical Inquiry
- 3.4 Uses of Statistics in Business and Economics
- 3.5 Self- Assessment Exercise
- 4.0 Conclusion
- 5.0 Summary
- 6.0 Tutor-Marked Assignment
- 7.0 References

#### **1.0 Introduction**

In recent times, statistical concepts have been regarded as necessary in business and economic decisions. Hardly can one plan effectively without the knowledge of at least the basic principles of statistics. In this unit, you will be informed of the preliminaries in the study of statistics and how the concept can be applied in making business decisions.

#### **2.0 Objectives**

By the time you go through this unit, you will be able to:

1. Know the general definition of statistics
2. Understand what is meant by statistical data
3. Know the necessary stages of statistical inquiry
4. Know the uses of statistics in business and economic decisions

#### **3.0 Definitions, Data, Statistical Inquiry, and Uses of Statistics**

##### **3.1 Definitions.**

The Science of Statistics can be viewed as the application of scientific method in the analysis of numerical data for the purpose of drawing inferences that are useful in making rational decisions. The term, 'statistics' can be defined as numerical facts systematically collected and presented for analysis. Today's managers are interested in the numerical facts affecting their organizations, particularly, those facts affecting production figures, sales figures, financial forecasts, productivity, profitability, accountability, and several others.

The term 'statistics' do not always refer to data, but refer to technique or method of investigating problems by analyzing statistical facts. Statistical methods are essentially

therefore, a branch of mathematics. It requires an understanding of such mathematical terms as averages, trends, distributions, probabilities, and the like. Viewed in this way, one may define statistics as a range of mathematical techniques for analyzing problems *in the real world*. *As a matter of fact, mathematical statistics is of increasing importance* in a wide range of activities in the areas of industrial mass production, economics, politics, psychology, analysis of public opinion, agriculture, traffic studies, physics and engineering.

In these areas, data are used in testing hypotheses, and models can be developed for planning and forecasting purposes. Consider, for example, a case of sales forecast. It could be asserted that revenue generated from sales will depend on the amount of naira expenditure on radio advertisement. In mathematical terms, this can be stated as follows:

$R = f(\text{Advert Expenditures})$ ,

where R represents revenue from sales.

### 3.1.1 Types of Statistics.

There are basically two types of statistics:

1. Primary Statistics. This involves the analysis of primary (or internal) data.
2. Secondary Statistics. Involving the analysis of secondary (or external) data.

Statistics can be further classified into:

**Descriptive Statistics.** This can be defined as those methods involving the collection, presentation, and characterization of a set of data in order to properly describe the various features of that set of data. The numerical facts collected by us on any subject or behaviour of interest will help in describing the state of affairs existing at the time of investigation. For example, the annual salaries of workers in XYZ Company can be described in a tabular form as follows:

**Table 3.1 Annual Salaries for XYZ Company Employees**

Salary (N'000s)	Number of Workers	Average Age
60 – 70	30	18 years
71 – 81	28	23 years
82 – 92	18	28 years
93 – 103	14	32 years
<b>TOTAL</b>	<b>90</b>	

By inspection, you will observe that the company employs 90 workers, with annual salaries between 60 and 103 thousand naira. On the average, the workers are between 18 and 32 years of age.

**Inferential Statistics.** Inferential statistics can be defined as those methods that make possible the estimation of a characteristic of a population or the making of a decision concerning a population based only upon sample results. This is often referred to as *analytical statistics*. *It goes further than the descriptive statistics by analyzing the data so as to enable us make reliable and authentic decisions.* This decision might be to increase salaries, decrease salaries, or no action at all, for our example on the XYZ Company.



### 3.2 Data, Sources of Data, And Methods of Data Collection

#### 3.2.1 Data

Unbiased set of statistic observations or information are referred to as data. There are basically two types of data:

1. Qualitative Data involving random variables that yield categorized responses;
2. Quantitative Data involving random variables that yield numerical responses.

Figure 3.1 below is a summary of how the two types of data can be generated.

**Figure 3.1: Types of Data.**

Data Type	Question Type	Responses
Qualitative	Do you own company shares	Yes --- No ---
Quantitative:		
Discrete	How many Cigas do you smoke a day?	Number -----
Continuous	How tall are you?	Number -----

Note that discrete quantitative data are numerical responses which arise from a counting process, while continuous quantitative data are numerical responses which arise from a measuring process.

#### 3.2.2 Sources of Data.

Sources of data include:

1. Publications. These include published data from such government agencies as the Federal Office of Statistics (FOS), Nigeria Deposit Insurance Corporation (NDIC), Central Bank of Nigeria (CBN), and the World Bank. These are the major sources of secondary data.
2. Experimentation: Data can be obtained through experiments.
3. Survey: With appropriate questionnaire instrument, reliable data can be obtained through survey. Survey is the major source of primary data.

#### 3.2.3 Methods of Data Collection

There are five major methods of data collection:

1. Observation. This involves monitoring the situation under investigation by trained observers.
2. Inspection. Inspection requires test of objects (e.g. test of weight, tensile strength, blood pressure, etc).

3. Abstraction From Records – Analysis of past records (or secondary data).
4. Questionnaire: Data can be collected through responses to structured and unstructured questionnaires.
5. Interviewing. This involves a person-to-person administration of questionnaires.

### 3.3 Stages of Statistical Inquiry

The stages of statistical inquiry are as follows:

**Stage 1: Problem Statement.** A statistical inquiry must begin with explicitly identify and define the problem of interest. The inquiry should not be launched in general terms. As a case in point, an investigation into the problem of labour turnover may not equally affect all parts of the firm concerned. Inquiry may not be necessary in many departments where there is no problem of labour turnover. The problem area may be a particular department, or a particular process or product. A careful statement of the problem of interest will give the researcher terms of reference from which relevance data can be collected for analysis.

**Stage 2: Decision on the Best Approach.** In line with the terms of reference in stage 1, you need to determine how you will approach the problem. Statistical evidence may already be available from past research. Another person may have already faced the same problem, it will be a waste of time to repeat an investigation. Many good inquiries start with a literature Review in which one can read the published materials already available on the problem of interest. Other inquiries begin with a thorough survey of all past records available in house. As an example, the investigation of the labour turnover problem in a particular department might begin with examination of the in-house personnel files. The basic question would be: Why did employees in this department leave? Did the explanations these people gave fit into a pattern of behaviour that identifies the cause of the problem in focus? Was it working conditions, level and type of remuneration, supervisory problems, or what?

**Stage 3: Definition of the Extent of the Inquiry.** You need to state the extent of the inquiry. Shall it extend to the whole population? Should it be just a sample survey?

**Stage 4: Determination of the Instrument for the Inquiry.** You need to know whether questionnaire should be used as an instrument for the inquiry. A pilot inquiry may be necessary to test out the questionnaire in order to find out if there are unforeseen ambiguities in the questions.

**Stage 5: Data Collection:** Most statistical inquiries are in the form of interviews. It is essential to appoint interviewers, brief them adequately and ensure that they conduct the interviews in a proper manner. Any conclusion or inference drawn from a badly conducted series of interviews will be meaningless. Other inquiries do not involve interviews. In these cases, data are collected by enumerators who record facts as they become available.

**Stage 6: Editing and Classification of the Data.** The outcome of stage 5 will be a mass of raw data, in a very indigestible form. Most of the time, these need editing. The data are then tallied, tabulated and summarized in a useful form; in a form that they could be easily analyzed.

**Stage 7: Data Analysis.** At this stage, you are now in the position to analyse the data. The data analysis needs the use of those who are knowledgeable about the subject of interest and skilled in statistical methods.

**Stage 8: Data Presentation and Report Writing.** The outcome of any survey should be a set of proposals that would remedy the identified problem. In making these proposals, it you are required to present the data in a simple and convincing manner, as part of the report to the appropriate authority. Such report must quote the terms of reference of the inquiry. You should use such tools as tables, charts, diagrams to show the findings of the survey.

Your report should suggest the causes or reasons for the identified problem, and make authentic recommendations for the solutions. Statistical surveys are generally expensive, any the researcher will be judged by his or her cost effectiveness in demonstrating the true facts and recommending the most likely cure in the circumstances.

### **3.4 Uses of Statistics in Business and Economics**

It is often difficult to separate statistical data from statistical techniques or methodology. These are not mutually exclusive subject areas. Thus, though population statistics are usually collected and presented by social statisticians, they are of enormous importance to economists and business persons alike. Economists require data for most of their economic analysis, especially in model building and econometric analysis. In business terms, products suitable for a particular age group need to be available in increasing numbers as a bulge in the birthrate moves through its life cycle. A business person who is aware of the passing of the wave may decide to expand output as it approaches, and cut down on output as it passes. The National survey on income and expenditure is of interest to both the public-sector parastatals and the private-sector firms.

All published secondary statistics are part of management information, but many statistical management information are generated in house. These are generated both as raw data and in treated or refined form. You may have tables show output, sales, machine utilization, stock levels, personnel rolls, and the like. You may also have charts of production, productivity, sales, machine utilization, and the like; ratios of stock turnover, gross profit, net profit, working capital, and so on. Countless analysis will be made of products, customer trends, sales areas, sales periods, order size, distribution method, maintenance programmes, vehicle usage, cash flow, and so on. All such management information will require the manager to be aware of statistical techniques, familiar with statistical jargon and appreciative of its uses. Corporate objectives are seldom realized by chance; they have to be achieved through managerial controls which keep the firm abreast of trends.

### **3.5 Self -Assessment Exercise**

Explain briefly why statistical principles are important business decisions

### **4.0 Conclusion**

This unit serves as preliminary to the understanding of the importance of statistics in business and economic decisions. It has been able to inform you about the general definition of statistics as it is used in business. Other interesting issues presented by the

unit are in the areas of sources and methods of collecting data, stages of statistical inquiry, and the uses of statistics in business and economics.

### **5.0 Sum mary**

The Science of Statistics can be viewed as the application of scientific method in the analysis of numerical data for the purpose of drawing inferences that are useful in making rational decisions. The term, 'statistics' can be defined as numerical facts systematically collected and presented for analysis. Most managers are interested in the numerical facts affecting their organizations, especially, those facts affecting production figures, sales figures, financial forecasts, productivity, profitability, accountability, and several others. There are basically two types of statistics: Primary Statistics, which involves the analysis of primary (or internal) data and Secondary Statistics, involving the analysis of secondary (or external) data.

Unbiased set of statistic observations or information are referred to as data. There are basically two types of data:

1. Qualitative Data involving random variables that yield categorized responses; and,
2. Quantitative Data involving random variables that yield numerical responses.

Basic sources of data include: publications, experimentation, and survey. Data can be collected through:

*Observation. This involves monitoring the situation under investigation by trained observers.*

*Inspection, requiring test of objects (e.g. test of weight, tensile strength, blood pressure, etc).*

*Abstraction From Records – Analysis of past records (or secondary data).*

*Questionnaire: Data can be collected through responses to structured and unstructured questionnaires.*

*Interviewing, involving a person-to-person administration of questionnaires.*

There are basically eight stages of statistical inquiry, each of which must be completed for a statistical inquiry to be reasonable and useful. These stages begin with the statement of problem of inquiry and ends with data presentation and report writing.

### **6.0 Tutor-Marked Assignm ent**

What are the main stages in a statistical inquiry? Illustrate your stage with the following:

A proposal has been made to move a factory to new premises in a suburban area. Housing will be provided for staff wishing to move, and the inquiry is to establish the exact position of personnel if the move takes place. In the opinion of the board of directors, the move is viable from other business points of view.

### **7.0 References**

1. Onwe, O. J. (2007) Statistical Methods for Business and Economic Decisions: A Practical Approach (Lagos: Samalice)

2. Hanke, J. E. and Reitsch, A. G. (1991) Understanding Business Statistics (Homewood, IL: Richard Irwin)

## **UNIT 2: STATISTICAL POPULATION, SAMPLING TECHNIQUES, AND DETERMINATION OF SAMPLE SIZE**

### **Content**

- 1.0 Introduction
- 2.0 Objectives
- 3.0 Statistical Population, Sampling, and Determination of the Sample Size
- 3.1 Differentiating Population from a Sample
- 3.2 Sampling Techniques
- 3.3 Determination of the Sample Size
- 3.4 Self- Assessment Exercise
- 4.0 Conclusion
- 5.0 Summary
- 6.0 Tutor-Marked Assignment
- 7.0 References

### **1.0 Introduction**

In this unit, you will be looking at what statistical population is all about, sampling from the population, and how you can determine statistically a sample size from a given population. Our aim is to make you understand one of the preliminary stages of statistical analysis that is, sampling from a given population.

### **2.0 Objectives**

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

1. Distinguish between a statistical population and a sample from the population
2. Know the recommended sampling techniques
3. Take a sample from an entire population
4. Determine the needed sample size for a given study.

### **3.0 Statistical Population, Sampling, and Determination of the Sample Size**

#### **3.1 Differentiating Population from a Sample**

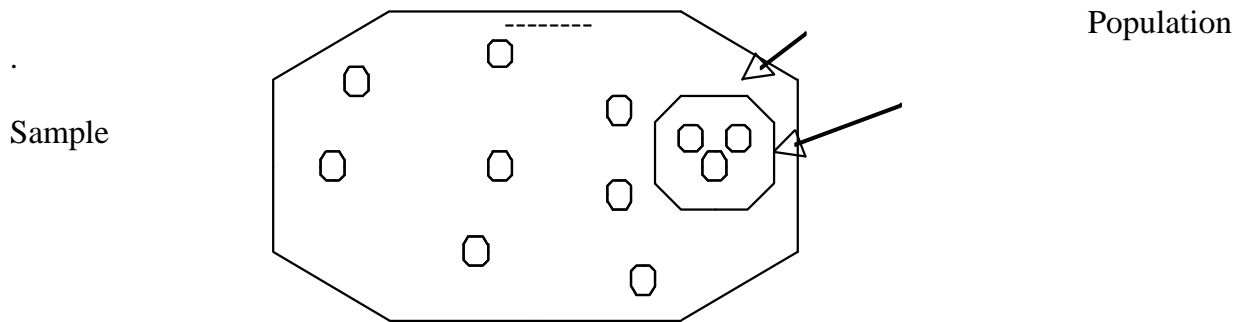
##### **3.1.1 Population.**

This depicts the totality of all behaviours in a given universe. It is synonymous to the universal set of all behaviours.

##### **3.1.2 Sample.**

A sample is a subset of a population. Statisticians use sample behaviours to draw inferences about the population behaviour. For a quick comparison, consider figure 3.1 below.

**Figure 3.1: Pictorial Representation of the Difference between Population and Sample.**



### 3.2 Sampling Techniques

There are two basic classifications of Sampling Techniques:

#### 3.2.1 Probabilistic Sampling Techniques including.

**1. Simple Random Sampling whereby every subject or respondent has the same chance** of being selected. Simple random sampling can be done either with replacement – where subjects are replaced after selection from the entire population or without replacement whereby subjects are withdrawn from the entire populations after being selected. For a reliable and practical sampling, it is advisable to apply a simple random sampling without replacement as this is capable of improving the chances of subjects yet to be part of the sample or yet to be selected.

**2. Systematic Sampling. This involves systematic selection of subjects from a given** population. In this selection process with a starting point (or subject) selected randomly without replacement, the required sample size is selected using a sampling interval.

This sampling interval is obtained by the formula:

$$\text{Sampling Interval} = \frac{N}{n},$$

Where N = Size of population  
n = the sample size.

For example, suppose the size of the population, N = 2000,  
and the desired sample size, n = 400, the sampling interval will be:

$$\frac{N}{n} = \frac{2000}{400} = 5$$

It follows that from the randomly selected starting point, every 5th subject will be selected as part of the sample until 400 subjects are selected

**3. Stratified Sampling. This gives a fair representation of various strata in a given** population. To illustrate this technique, assume a working population of 500 executives with:

10 General Managers  
 100 Managers  
 200 Assistant Managers  
 190 Supervisors

To obtain a representative stratified sample from this population, the stratification is as follows:

	Number	Percent	
General Managers	10	2%	
Managers	100	20%	
Asst. Managers	200	40%	
Supervisors	190	38%	
Total	<u>500</u>	<u>100%</u>	

Assume we want to obtain a stratified sample of 100 executives from the population, the sample will be made up of:

General Manager = 2% of 100 = 2  
 Managers = 20% of 100 = 20  
 Asst. Managers = 40% of 100 = 40  
 Supervisors = 38% of 100 = 38  
 Total 100

From this process, we would have a representative sample of 2 General Managers, 20 managers, 40 Asst. Managers and 38 Supervisors.

**4. Cluster Sampling.** This sampling technique is useful especially where ethnic or geographical representation is needed in a given study. In this technique, the researcher selects a geographical area at random. Every single subject in such area is then used as part of the sample.

For all practical purposes, and to ensure minimum bias with statistical analysis and results, it is advisable to use any of the above probabilistic sampling techniques.

### 3.2.2 Non-Probabilistic Sampling Techniques

These techniques include:

**1. Convenience Sampling (or Accidental Sampling)** is applied by researchers interested in having ideas of situations or phenomenon of interest. A marketing researcher for Nigerian Breweries, for example, may decide to station him or herself at a beer parlor in order to obtain the opinions of consumers of a particular brand of the brewery's beer. The intention is for the researcher to interview any of the consumers he may come in contact with accidentally. It is a good example of convenience or accidental sampling.

As with any other non-probabilistic sampling design, convenience or accidental sampling design may give rise to unreliable inferences and conclusions. This is however,



economical and simple to use.

**2. Judgment Sampling.** This is applicable to situations in which the researcher is guided by the belief that reference subjects will provide the required data or information for a particular research process. A researcher interested in studying the economic implications of Bank Consolidation in Nigeria, for example, is not likely to get reliable information from petty traders. Using his or her own judgment, the researcher would choose respondents from the banking industry or profession as this class of respondents would be in a better position to provide the relevant information on Bank Consolidation.

**3. Quota Sampling.** This is similar to stratified sampling technique as it selects representative subjects according to their percentage representation in the population. It is mostly used in cases whereby the characteristics of the population of interest can be easily identified. These characteristics are usually represented in the sampling process so the basic information about the population can be obtained. A typical example of quota sampling would be in the case of a population consisting of different groups such as, students, lecturers, and parents. In such a population, the researcher may want to classify the population according to size as follows:

Students	.....	60%
Lecturers	.....	20%
Parents	.....	20%
Total	.....	100%

Assume the researcher is interested in a sample size of 5000 subjects, then, using the quota sampling design, he or she would choose the sample of students, lecturers, and parents as follows:

Students	-----	60% of 5000 =	3000
Lecturers	.....	20% of 5000 =	1000
Parents	.....	20% of 5000 =	1000
Total	.....	100%	5000

Non-probabilistic sampling techniques are not advisable in practical research.

### 3.3 Determination of The Sample Size

Techniques for the determination of the sample size for a given research depends on the type of research as well as the size of the population under investigation. For most practical business studies, the population under investigation is usually of a finite or known size, N. In this case, the following formula, as advanced by Taro Yamane, has been suggested for use for determination of sample size.

$$n = \frac{N}{1 + Ne^2}$$

where n = desired sample size

N = the finite size of the population

e = maximum acceptable margin of error as determined by the researcher

1 = a theoretical or statistical constant

**Example:** Consider a researcher working on a population of 310 subjects. He would like to determine the sample size with a 5 percent margin of statistical error. According to the above formula, the required sample size would be:

$$n = \frac{N}{1 + Ne^2} = \frac{310}{1 + 310(0.05)^2} = 175$$

It follows that the required sample size would be approximately 175 subjects  
Other formulas that can be used, according to the research needs include:

**3.3.1 Sample Size for Mean Values:** The sample size for statistical estimation and tests relating to population mean or average values can be obtained by:

$$n = \frac{Z^2 s^2}{e^2}$$

where  $Z$  = the Z-statistic (or value) corresponding to the desired confidence level  
 $s$  = a pre-determined value of the population standard deviation  
 $e$  = the maximum acceptable margin of error

The value for the population standard deviation ( $s$ ) can be obtained from previous statistical studies of similar population. If such information is not available, a satisfactory value can be obtained through pilot surveys, using a reasonable sample of the population.

**Example:** Suppose you want to be 99 percent confident that the true value of the population mean for the PDP voters in Enugu State will be within 10 percent of the sample mean. By implication, the estimate of the true population mean by the sample would be in error only 10 percent. Your look at the normal distribution would reveal that the value of  $Z$  corresponding to 99 percent level of confidence or alternatively, 1 percent level of significance, is 2.58. If you obtained the information that a survey of similar population indicates that 0.8 is a realistic estimate of the population standard deviation, then, the sample size will be:

$$n = \frac{Z^2 s^2}{e^2} = \frac{(2.58)^2 (0.8)^2}{(0.1)^2} = 426$$

**3.3.2 Sample Size For Proportions:** The sample size for statistical estimation and tests of proportions is obtained by:

$$n = \frac{pq}{e^2} Z^2$$

where  $p$  = the approximate value of the proportion of success  
 $q = 1 - p$  = the estimated true proportion of failure.

Note that  $p$  is the population proportion to be estimated by taking the sample. If the value of  $p$  can be estimated based on past data or experience, then this value can be applied to the formula. In the absence of past data, you are advised to make use of the value of  $p$  that will make the quantity  $pq$  as large as possible, and such value is  $p = 0.50$ .

**Example:** Suppose that you as a researcher is interested in knowing the sample size needed to estimate the proportion of your customers who are females. You wish to be 99 percent confident, with a fairly small sampling error of 2.5 percent in estimating the true proportion. There are no prior information about the true proportion,  $p$ , so that, for a conservative estimate,  $p = 0.50$ .

With 99 percent level of confidence or 1 percent level of significance, the normal distribution table on  $Z$  would indicate the value of  $Z$  as 2.58. Using the applicable formula, we get:

$$n = \frac{Z^2 pq}{e^2} = \frac{(2.58)^2 (0.5)(0.5)}{(0.025)^2} = 2663$$

It follows that the required sample size would be approximately 2663 Subjects.

### 3.4 Self -Assessment Exercise

Discuss briefly the major differences between a sample and population

### 4.0 Conclusion

This unit has expanded your knowledge of sampling and sampling techniques. You also learned how you can compute or determine the sample size for a given population study. The technique to be applied in the determination of sample size will be determined by both the type of research you want to carry out as well as the size of your population of interest.

### 5.0 Summary

The term 'population' depicts the totality of all behaviours in a given universe. It is synonymous to the universal set of all behaviours. A sample is a subset of a population. Statisticians use sample behaviours to draw inferences about the population behaviour.

There are two basic classifications of sampling techniques: (i) Probabilistic sampling techniques including, simple random sampling; systematic sampling; stratified sampling; and, cluster sampling. (ii) Non-probabilistic sampling techniques including, convenience sampling; judgment sampling; and, quota sampling.

Three levels in the determination of sample size were presented. First is the sample size for finite population size. Second is the sample size for mean values. And, third is the sample size for studies involving population proportions.

### 6.0 Tutor-Marked Assignment

Suppose you wish to be 90 percent confident, with a fairly small sampling error of 2 percent in estimating the true proportion. There are no prior information about the true

proportion,  $p$ , so that, for a conservative estimate,  $p = 0.50$ . What would be your sample size?

### **7.0 References**

1. Onwe, O. J. (2007) Statistical Methods for Business and Economic Decisions: A *Practical Approach* (Lagos: Samalice)

## **UNIT 3: MEASURES OF CENTRAL TENDENCY**

### **Content**

- 1.0 Introduction
- 2.0 Objectives
- 3.0 The Measures of Averages or Central Tendency
  - 3.1 Arithmetic Mean
  - 3.2 The Median
  - 3.3 The Mode
  - 3.4 Self- Assessment Exercise
- 4.0 Conclusion
- 5.0 Summary
- 6.0 Tutor-Marked Assignment
- 7.0 References

### **1.0 Introduction**

In this unit, you will be discussing the statistical measures of central tendency including, the arithmetic mean, the median, and the mode. These measures are significant in management decisions involving averages.

### **2.0 Objectives**

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

1. Understand and apply the different measures of averages
2. Make decisions involving population averages.

### **3.0 The Measures of Averages or Central Tendency**

The three important measures of central tendency to be Discussed in this section, among others include:

1. Arithmetic Mean
2. Median
3. Mode

### **3.1 Arithmetic Mean**

The arithmetic mean is used in estimating the average of a given set of observations. We are interested in three levels of estimates:

- (i) Arithmetic Mean of observations of equal weights or levels of importance.
- (ii) Arithmetic Mean of weighted set of observations (weighted Arithmetic Mean)
- (iii) Arithmetic Mean for a grouped data (or set of observations)

#### **3.1.1 Arithmetic Mean f or Observation of Equal Weight**

The computation of arithmetic mean for a set of observations of equal weights or levels of importance follows the following process:

Let X represent a variable of which the observations that are of equal weights are given as:

$X_i (i = 1, 2, 3, \dots, n)$ . Then, the arithmetic mean of this set of observations is defined by:  $\bar{X} = \frac{X_1 + X_2 + X_3 + \dots + X_n}{n}$

where  $n$  = number of observations

### Example:

Let  $X_i = 2, 2, 4, 7, 3, 1$

$$\bar{X} = \frac{X_1 + X_2 + X_3 + X_4 + X_5 + X_6}{n} = \frac{1+2+2+3+4+7}{6} = \frac{19}{6} = 3.1667$$

Notice that the set of observations were presented in an array before computation of the mean. It is therefore advisable to always array a set of observations before processing them for subsequent analysis.

By an “array”, we mean the presentation of observations in either an ascending or descending order.

### 3.1.2 Weighted Arithmetic Mean

This computation of the arithmetic mean takes into account the weights assigned to individual observations. By definition, Let  $X$  represent a variable, with observations of different weights, and

Let  $X_i (i = 1, 2, 3, \dots, n)$  = individual observations on the variable  $X$

$w_i (i = 1, 2, 3, \dots, n)$  = weights attached to individual observations

$\bar{X}_w$  = weighted mean of the observations on  $X$

Then, 
$$\bar{X}_w = \frac{w_1 X_1 + w_2 X_2 + w_3 X_3 + \dots + w_n X_n}{w_1 + w_2 + w_3 + \dots + w_n}$$

### Example.

Assume three levels of examinations, Quiz, Mid-semester, and Final examination. The weights assigned to each of these examinations and scores for a particular participant is as summarized below:

Exam	Weight (w)	Score (x)
Quiz	1	65
Mid-Semester	3	50
Final Exam	6	45

What is the average score of the participant for the three Examinations?:

### Solution:

Using our definition,

$$\bar{X}_w = \frac{w_1 X_1 + w_2 X_2 + w_3 X_3 + \dots + w_n X_n}{w_1 + w_2 + w_3 + \dots + w_n}$$

We compute as follows:

Exam	Weight (w)	Score (xi)	w xi
Quiz	1	65	65
Mid-Semester	3	50	150
Final Exam	6	45	270
		Swi=10	Swixi=485

$$\bar{X} = \frac{\sum w x_i}{\sum w} = \frac{485}{10} = 48.5$$

The required average score is 48.5

### 3.1.3 Arithmetic Mean of A Grouped Data

When a set of observations is presented in a grouped form or in class limits, the computation of the arithmetic mean is done according to the following definitions:

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

Where  $\sum f = n$

$\bar{X}_g$  = Grouped mean of the set of data on variable x.

f = frequency of observations

x = mid- value of each class limit defined by:

$$x = \frac{\text{Lower Class Limit} + \text{Upper Class Limit}}{2}$$

for each class limit.

#### Example

The following is data on the daily wages paid to workers in a given factory:

Daily Wages (in N)	No. of Workers (f)
200 – 300	15
301 – 401	30
402 – 502	45
503 – 603	60
<b>Total</b>	<b>150</b>

Compute the average daily wages.

#### Solution

We re-tabulate the data as follows:

Daily Wages	f	X	fX
200 - 300	15	250	3750
301 - 401	30	351	10530
402 - 502	45	452	20340
503 - 603	60	553	33180
			<b><math>\sum fX = 67800</math></b>

Note: Computations of the values of X (ie, the mid-value of classes) were as follows:

$$\frac{200+300}{2} = \frac{500}{2} = 250$$

$$\frac{301+401}{2} = \frac{702}{2} = 351$$

$$\frac{402+502}{2} = \frac{904}{2} = 452$$

$$\frac{503+603}{2} = \frac{1106}{2} = 553$$

### 3.2 The Median

The average of a given set of data can also be estimated using the median. The median is a measure of central tendency which appears in the “middle” of an ordered sequence of values or observations. That is, half of the observations in a set of data are lower than the median value and half are greater than it. To compute the median from a set of raw data, we must first array the data. If we have odd number of observations, the median is represented by the numerical value of the  $(n+1)/2$  arrayed observation. On the other hand, if the number of observations in the sample is an even number, the median is represented by the mean or average of the two middle values in the ordered array.

#### Example

Consider the following raw data on hourly wage rate for six executive secretaries:

**Raw data (in ₦=N=)**

X1 = 950	X4 = 950
X2 = 300	X5 = 850
X3 = 1000	X6 = 750

Compute the median hourly wage rate for the six secretaries.

#### Solution.

Ordered Array:

$X_i = 300, 750, 850, 950, 950, 1000$

Since this involves an “even” number of observations, the median will be the average of the two middle values:  $\frac{850 + 950}{2} = \frac{1800}{2} = 900$

2

2

Thus, the median hourly wage rate is N900

As another example, consider the following ordered array of  $n = 5$  students’ mid-semester examination results:

$X_i = 64, 79, 88, 90, 94$



The median score will be the  $\frac{(n+1)}{2}$ th or  $\frac{(5+1)}{2}$ th

ordered observation, that is, 3rd ordered observation, which is 88, since this involves “odd” number of observations.

### 3.2.1 Median of A Grouped Data

The median value for a grouped data is computed according to the following definitions:

For a grouped data:

$$\text{Median} = L + \left( N - \frac{\sum f}{2} \right) \frac{C}{f_m}$$

Where L = Lower actual class limited of the median group, where the median group can be regarded as the group (or class limit) with the highest frequency.

N =  $\sum f$  = total number of observations in the given data

fb = total frequency of all classes below the median group

C = class width for the actual class limit.

$f_m$  = frequency of the median group

**Note: Class Width = Upper Class Limit – Lower Class Limit**

#### Example

Giving the following table on the share prices of a quoted company over a period of 60 days:

Price in =N=	No. of Days (f)
110-114	2
115-119	6
120-124	8
125-129	12
130-134	14
135-139	8
140-144	6
145-149	4

-----  
 $\sum f = 60$   
 -----

Find the median share price.

#### Solution:

We re-tabulate the data to get the needed computational values:

State Class Limits	Actual Class Limits	Frequency (f)
110-114	109.5-114.5	2
115-119	114.5-119.5	6
120-124	119.5-124.5	8
125-129	124.5-129.5	12
130-134	129.5-134.5	14
130-139	134.5-139.5	8
140-144	139.5-144.5	6
145-149	144.5-149.5	4
		<u><u><math>\sum f = 60</math></u></u>

From the re-tabulated data, the median group, that is, the group with the highest frequency (14) is (129.5-134.5).

By definition,

$$\text{Median Price} = L + (N - \frac{fb}{2}) \frac{C}{fm}$$

where  $L = 129.5$ ;  $N = 60$ ;  $fb = 12+8+6+2 = 28$ ;  $C = 5$ ;  $fm = 14$ .

$$\begin{aligned} \text{It follows that median price} &= 129.5 + \frac{(60-28)(5)}{14} \\ &= 129.5 + 10 \\ &= 129.5 + 0.71 \\ &= 130.21 \end{aligned}$$

Thus, the median share price is N130.21

### 3.3 The Mode

The mode is a quick measure of central tendency or average. The mode is the most typical or most frequently observed value in a given set of data.

For the set of data,  $X = 1, 2, 2, 4$ , the value with the highest frequency is 2. Hence, the mode for the given data is 2.

#### 3.3.1 Mode of A Grouped Data

The computational process for the mode of a grouped data is similar to that of the median. The process is as follows:

$$\text{Mode} = L + \frac{(fm - fL)C}{(fm - fL) + (fm - fh)}$$

where  $L$  = Lower actual class limit of the modal group

$fm$  = frequency of the modal group

$fL$  = frequency of the group before the modal group

$fh$  = frequency of the group after the modal group

$C$  = class width for the actual class limit

#### Example

Consider the following data on the sales by 90 sales representatives:

Sales (N'000s)	No. of Sales Reps.(f)
10-15	10
16-21	36
22-27	28
28-33	10
34-39	6

Calculate the modal value of sales.

#### Solution.

Re- tabulating the data, we get:

Stated Class Limits	Actual Class Limits	Frequency (f)
10-15	9.5-15.5	10
16-21	15.5-21.5	36
22-27	21.5-27.5	28
28-33	27.5-33.5	10
34-39	33.5-39.5	6

From the retabulated data,  $L = 15.5$ ;  $f_m = 36$ ;  $f_L = 10$ ;  $f_h = 28$ ;  $C = 6$ .  
It follows that:

$$\begin{aligned}
 \text{Mode} &= L + \frac{(f_m - f_L)C}{(f_m - f_L) + (f_m - f_h)} \\
 &= 15.5 + \frac{(36 - 10)(6)}{(36 - 10) + (36 - 28)} \\
 &= \frac{15.5 + 156}{34} \\
 &= 20.09
 \end{aligned}$$

Thus, the modal value of sales is  $N(20.09 \times 1000) = N20,090$  (since values are in thousands).

### 3.4 Self -Assessment Exercise

Explain why it is important for a manager to understand the different measures of averages or central tendency.

### 4.0 Conclusion

This unit has discussed in detail the important measures of averages in the business world. The measures include: (i) Arithmetic mean; (ii) Median; and (iii) Mode. The most accurate measure is that of the arithmetic mean.

### 5.0 Sum mary

The three important measures of averages in business decisions have been enumerated as arithmetic mean, the median, and the mode.

The arithmetic mean is the most accurate measure and is used in estimating the average of a given set of observations.

It can involve three levels of estimates:

- (i) Arithmetic Mean of observations of equal weights or levels of importance.
- (ii) Arithmetic Mean of weighted set of observations (weighted Arithmetic Mean)
- (iii) Arithmetic Mean for a grouped data (or set of observations)

Each of these were presented and exemplified in the discussions.

### 6.0 Tutor-Marked Assignm ent

The award of a contract is based on a contractor's scores on credibility test over five years in business. The scores are weighted as follows:

Year:	1	2	3	4	5
Weight:	1	3	5	7	9

If the scores of five of the contractors (Adamu, Okoli, Babalola, Udoh, and Okoh) are as shown below:

Contractor	Year				
	1	2	3	4	5
Adamu	62	62	50	61	48
Okoli	46	56	67	50	62
Babalola	51	54	60	55	58
Udoh	63	60	49	52	61
Okoh	58	62	60	52	70,

And given that a contract will be awarded to the contractor with the highest average score, which contractor do you think will win the contract?

## 7.0 References

1. Onwe, O. J. (2007) Statistical Methods for Business and Economic Decisions: A *Practical Approach (Lagos: Samalice)*
2. Hanke, J. E. and Reitsch, A. G. (1991) Understanding Business Statistics (Homewood, IL: Richard Irwin)

## **UNIT 4: MEASURES OF DISPERSION AND SKEWNESS**

### **Content**

- 1.0 Introduction
- 2.0 Objectives
- 3.0 Measures of Variation and Skewness
- 3.1 Measures of Dispersion or Variation
- 3.2 Measures of Skewness
- 3.3 Self- Assessment Exercise
- 4.0 Conclusion
- 5.0 Summary
- 6.0 Tutor-Marked Assignment
- 7.0 References

### **1.0 Introduction**

The second most important characteristics which describe a set of data is the amount of variation, scatter, or spread in the data. In this unit, we discuss in detail the various measures of dispersion and skewness. The purpose of these measures is to amplify the imperfect summary of any statistical distribution usually provided by the three measures of averages commonly used: the mean, the median, and the mode. These averages are inherently unsatisfactory because no single measure of average can tell you everything about a distribution, and the wider the dispersion of a given data around the average, the less satisfactory the average becomes. In order to improve your understanding of population averages, you need to know how wide the dispersion is around the average, and whether it is symmetrical (un-skewed) or asymmetrical (skewed). The first set of measures to be discussed here are measures of dispersion, and the second set measures of skewness.

### **2.0 Objectives**

Having gone through the discussions in this unit, you will:

1. Appreciate the issues concerning dispersions and skewness in business data
2. Understand the various ways you can determine the dispersion or standard deviation of a given set of business data.
3. Know how to make decisions involving instability in the business environment due to variations in business data.

### **3.0 Measures of Variation (Dispersion) and Skewness**

This section focuses on the relevant measures of variation and skewness in the behaviours or distribution of business data. Each of these measures is accompanied by relevant examples.

#### **3.1 Measures of Dispersion or Variation**

The common measures of dispersion include:

- (i) The Range
- (ii) The Quartile Deviation
- (iii) Mean Deviation

- (iv) Variance
- (v) Standard Deviation
- (vi) Coefficient of Variation

The variation or dispersion can be said to measure the degree of uniformity of observations in a given set of data. The greater the variation, the more un-uniform the observations in a given set of data

### 3.1.1 The Range

The Range (R) of a given set of ungrouped data can be determined from an ordered array as the difference between the highest observation and the lowest observation in a distribution.

Let  $X_h$  = Highest observation

$X_L$  = Lowest observation

Then,  $R = X_h - X_L$

Given the arrayed data:  $X = 2, 5, 8, 9, 12, 13, 18$ , the range will be:

$R = 18 - 2 = 16$ .

The range can be an unsatisfactory measure of dispersion because it is affected by extreme values or items which render it unrepresentative of majority of the set of data.

### 3.1.2 The Quartile Deviation

Unlike the range, quartile deviation does not take extreme values or items. Quartiles are the boundaries separating the items in a given distribution or set of data into quarters.

There are, therefore, three quartiles: the lower quartile (at the 25 percent mark); the *median (at the 50 percent mark)*; and, the upper quartile (at the 75 percent mark). To compute the quartiles of ungrouped data, you simply use:

$0.25(n + 1)$ , for the lower quartile

$0.50(n + 1)$ , for the median quartile

$0.75(n + 1)$ , for the upper quartile

For grouped data, you simply use:

$0.25n$  for the lower quartile

$0.5n$  for the median quartile

$0.75n$  for the upper quartile

### Example

Consider the following output distribution of the employees of a manufacturing company:

**Table 3.1: Output of Employees**

Units of Output	Number of Employees (f)
21 – 30	7
31 – 40	11
41 – 50	14
51 – 60	8
61 – 70	5

Table 3.1 indicates that there are 45 items or observations ( ie. total number of employees or sum of the frequencies, Sf).

Using these information, the quartiles are as follows:

$$\text{Lower quartile (Q1)} = 0.25n = 0.25(45) = 11.25\text{th item}$$

$$\text{Median quartile (Q2)} = 0.5n = 0.5(45) = 22.5\text{th item}$$

$$\text{Upper quartile (Q3)} = 0.75n = 0.75(45) = 33.75\text{th item}$$

The values of the quartile items are determined simply as follows:

**Lower quartile:** Since, according to table 3.1, there are 7 items in the first group (ie, group of 21 – 30), the quartile item is the  $(11.25 - 7) = 4.25\text{th}$  item of the second group. Thus,

$$\begin{aligned} \text{Value of the lower quartile (Q1)} &= 30 + \frac{(4.25) \times 10}{11} \\ &= 30 + 3.66 \\ &= 34 \text{ approximately.} \end{aligned}$$

Therefore, the value of the lower quartile is about 34 units.

In a similar process, the value of the median and upper quartiles can be determined, thus:

**Value of Median quartile:** The 22.5th item in the distribution is in the 41 – 50 group and is the  $(22.5 - 18) = 4.5\text{th}$  item out of 14 in the group (note that the figure 18 is the cumulative frequency of the first and second groups, and the figure 10 appearing in the calculations is the class interval of the distribution). The value of the median quartile (Q2) is therefore:

$$\begin{aligned} \text{Q2} &= \frac{40 + (4.5) \times 10}{14} \\ &= 40 + 3.21 = 43.21 \\ &= 43 \text{ units approximately.} \end{aligned}$$

**Value of the Upper quartile (Q3):** The 33.75th item in the distribution is in the third group, the group of (41 – 50), and since there are 32 items in the third group (the cumulative frequency), the median is the  $(33.75 - 32) = 1.75\text{th}$  item in the fourth group. The value of the upper quartile is therefore:

$$\begin{aligned} \text{Q3} &= \frac{50 + (1.75) \times 10}{8} \\ &= 50 + 2.19 = 52.19 \\ &= 52 \text{ units approximately.} \end{aligned}$$

**The quartile deviation referred to as the semi-interquartile range is defined as one-half the difference between the upper quartile and the lower quartile. Thus,**

$$\text{Quartile Deviation} = \frac{\text{Q3} - \text{Q1}}{2}$$

In this example, therefore, the quartile deviation is:

$$\frac{52 - 34}{2} = 9 \text{ units}$$

The distribution in table 3.1 can then be described as having a median value of 43 units and a quartile deviation around the median value of 9 units.

### 3.1.3 The Mean Deviation (MD)

The Mean Deviation can be defined simply by the following relationship:

$$MD = \frac{\sum |X - \bar{X}|}{n}$$

where  $\sum |X - \bar{X}|$  = sum of the absolute values of deviation from arithmetic mean  
 $n$  = number of observation

As an example, consider again the arrayed data,  $X = 2, 5, 8, 9, 12, 13, 18$ .

The mean deviation, MD, can be computed as follows:

$$= \frac{\sum |X - \bar{X}|}{n} = \frac{67}{7} = 9.57$$

By tabulation,

X	(X - $\bar{X}$ )	X - $\bar{X}$
2	-7.57	7.57
5	-2.57	2.57
8	-1.57	1.57
9	-0.57	0.57
12	2.43	2.43
13	3.43	3.43
18	8.43	8.43

$$\sum |X - \bar{X}| = 67$$

Thus,

$$MD = \frac{\sum |X - \bar{X}|}{n} = \frac{67}{7} = 9.57$$

### 3.1.4 The Variance

The Variance for a given set of an ungrouped data can be defined by:

$$\text{Variance} = S^2 = \frac{\sum X^2 - \frac{(\sum X)^2}{n}}{n-1}$$

where X represents the numerical values of the given set of an ungrouped data.

Continuing with our earlier example, where

$$X = 2, 5, 8, 9, 12, 13, 18$$

and by tabulation:

X	X <sup>2</sup>
2	4
5	25
8	64
9	81
12	144
13	169
18	324

$$\sum X = 67; \sum X^2 = 811;$$

$$S^2 = \frac{\sum X^2 - \frac{(\sum X)^2}{n}}{n-1} = \frac{811 - \frac{67^2}{7}}{7} = 641.29$$



$$\begin{aligned}\text{Thus, } S^2 &= \frac{\sum x^2 - (\sum x)^2 / n}{n-1} = \frac{811 - 641.29}{7-1} \\ &= \frac{169.71}{6} = 28.285\end{aligned}$$

Thus, the variance of the given set of ungrouped data is 28.285.

### 3.1.5 The Standard Deviation

Simply stated, the standard deviation is the most useful measure of variation. It can be defined as the square root of the variance for a given set of data.

Thus,

$$\text{Standard deviation} = S = \sqrt{S^2}$$

Or,

$$S = \sqrt{\frac{\sum x^2 - (\sum x)^2 / n}{n-1}}, \text{ for ungrouped data.}$$

The standard deviation for the last example is:

$$S = \sqrt{S^2} = \sqrt{28.285} = 5.318$$

### 3.1.6 Variance And Standard Deviation For A Grouped Data

The computation of variance and standard deviation for a grouped data is illustrated with the following example.

The Variance and Standard Deviation for a grouped data are defined by the following formulations:

$$\text{Variance} = S^2 = \frac{\sum fx^2 - (\sum fx)^2 / n}{n-1}$$

$$\text{Standard deviation} = \sqrt{S^2} = \sqrt{\frac{\sum fx^2 - (\sum fx)^2 / n}{n-1}}$$

#### Example.

The following data presents the profit ranges of 100 firms in a given industry.

Profits (N' millions)	No. of Firms (f)
10.15	8
16.21	18
22.27	20
28.33	12
34.39	15
40.45	17
46.51	10
	<u>f = n = 100</u>

We are required to compute the variance and standard deviation of profits within the industry.

#### Solution.

By definition,

$$\text{Variance} = \frac{\sum f x^2 - (\sum f x)^2 / n}{n-1}$$

$$\text{Standard Deviation} = \sqrt{S^2} = \sqrt{\frac{\sum fx^2 - (\sum fx)^2 / n}{n-1}}$$

The computational process is as follows:

Profits (N millions)	Frequency (f)	Mid-Value (x)	fx	$\sum fx$	$\sum fx^2$
10-15	8	12.5	100	156.25	1250
16-21	18	18.5	333	342.25	6160.5
22-27	20	24.5	490	600.25	12005
28-33	12	30.5	366	930.25	11163
34-39	15	36.5	547.5	1332.25	19983.75
40-45	17	42.5	722.5	1806.25	30706.25
46-51	10	48.5	485	2352.25	23522.50
	$\sum f = n = 100$			$\sum fx = 3044$	$\sum fx^2 = 104791$

#### SUMMARY:

$$\sum fx = 3044$$

$$\frac{(\sum fx)^2}{n} = \frac{(3044)^2}{100} = 92659.36$$

$$\sum fx^2 = 104791$$

It follows that:

$$\begin{aligned} \text{Variance} = S^2 &= \frac{\sum fx^2 - (\sum fx)^2 / n}{n-1} = \frac{104791 - 92659.36}{100-1} \\ &= \frac{12131.64}{99} = 122.54 \end{aligned}$$

$$\text{Standard Deviation} = \sqrt{S^2} = \sqrt{122.54} = 11.07$$

Thus, the required variance and standard deviation are 122.54 and 11.07 respectively.

### 3.1.7 The Coefficient of Variation

Unlike other measures of variability, the coefficient of variation is a relative measure. It is particularly useful when comparing the variability of two or more sets of data that are expressed in different units of measurements.

The coefficient of variation measures the standard deviation relative to the mean and is computed by:

$$\text{Coefficient of Variation} = CV = \frac{S}{\bar{x}} \times 100\%$$

The coefficient of variation is also useful in the comparison of two or more sets of data which are measured in the same units but differ to such an extent that a direct comparison of the respective standard deviations is not very helpful. As an example, suppose a potential investor is considering the purchase of shares in one of two companies, A or B, which are listed on the Nigerian Stock Exchange (NSE). If neither company offered dividends to its shareholders and if both companies were rated equally high in terms of potential growth, the potential investor might want to consider the volatility of the two stocks to aid in the investment decision.

Now, suppose each share of stock in Company A has averaged N50 over the past months with a standard deviation of N10. In addition, suppose that in this same time period, the price per share for Company B's stock averaged N12 with a standard deviation of N4. Observe that in terms of actual standard deviations, the price of Company A's shares seems to be more volatile than that of Company B. However, since the average prices per share for the two stocks are so different, it would be more appropriate for the potential investor to consider the variability in price relative to the average price in order to examine the volatility/stability of two stocks.

The coefficient of variation of company A's stock is

$$CVA = \frac{(\frac{SA}{XA})}{100\%} = \frac{(N10)}{50} \times 100\% = 20\%$$

That of Company B's is

$$CVB = \frac{(\frac{SB}{XB})}{100\%} = \frac{(N4)}{12} \times 100\% = 33.3\%$$

It follows that relative to the average, the share price of company B's stock is much more variable/unstable than that of Company A.

### 3.2 Measures of Skewness

The measures of skewness are generally called Pearson's first coefficient of skewness and Pearson's second coefficient of skewness. Measures of skewness are used in determining the degree of asymmetry of a distribution; a distribution which is not symmetrical is said to be skewed.

#### 3.2.1 The Pearson's No. 1 Coefficient of skewness

The formula used in calculating Pearson's No. 1 coefficient is:

$$Sk = \frac{\text{Mean} - \text{Mode}}{s}$$

Notice that the mean, the mode, and the standard deviation are all expressed in the units of the original data. When the difference between the mean and the mode is computed as a fraction as a fraction of the standard deviation (or average spread of the data around the mean), the original units cancel out in the fraction. The result will be a coefficient of skewness, a number which tells you the extent of the skewness in the distribution.

**Example:** Consider a set of data on monthly sales of a company's product, the mean of which was found to be N240,000; the mode found to be N135,000; and the standard deviation found to be N85,000. The Pearson's No. 1 Coefficient of skewness would be calculated as follows:

$$Sk = \frac{\text{mean} - \text{mode}}{s} = \frac{240,000 - 135,000}{85,000} = 1.24$$

Generally, a complete absence of skewness would have a coefficient of skewness equal to zero. In our example, since the mean was larger than the mode, we obtained a positive coefficient of skewness to the extent of 124% of the standard deviation.

### 3.2.2 The Pearson's No. 2 Coefficient of Skewness

This type of the Pearson's coefficient of skewness came as a result of the fact that a precise calculation of mode is difficult in many distributions. Hence, Pearson's No. 2 coefficient of skewness uses the difference between the mean and the median of the distribution instead of the difference between the mean and the mode. In calculation, you have the formula:

$$sk = \frac{3(\text{mean} - \text{median})}{s}$$

This formula should give you a more accurate measure of skewness than that of the Pearson's No. 1 formula.

### 3.3 Self -Assessment Exercise

What are the importance of the measures of variation and skewness in business?

### 4.0 Conclusion

This unit has informed you of the different measures of variation and skewness as they are related to business decision making processes. As presented, the measures of variation include, among others, the range, the variance, and the standard deviation. Those of skewness are the Pearson's No. 1 and No. 2 coefficients.

### 5.0 Sum mary

The common measures of dispersion include:

- (i) The Range
- (ii) The Quartile Deviation
- (iii) Mean Deviation
- (iv) Variance
- (v) Standard Deviation
- (vi) C oefficient of Variation

The Range (R) of a given set of ungrouped data is determined from an ordered array as the difference between the highest observation and the lowest observation in a distribution. It can however, be an unsatisfactory measure of dispersion because it is affected by extreme values or items which render it unrepresentative of majority of the set of data.

Unlike the range, quartile deviation does not take extreme values or items. Quartiles are the boundaries separating the items in a given distribution or set of data into quarters.

The mean deviation is simply a measure of the mean of the absolute values of deviations from the data average. It is however not an accurate measure of variations.

The standard deviation and coefficient of variation are regarded as better and more accurate measures of variation or data uniformity. The coefficient of variation is a relative measure. It is particularly useful when comparing the variability of two or more sets of data that are expressed in different units of measurements. The measures of skewness are generally called Pearson's first coefficient of skewness and Pearson's second coefficient of skewness. These measures are used in determining the degree of asymmetry of a distribution; a distribution which is not symmetrical is said to be skewed.

#### **6.0 Tutor-Marked Assignment**

A distribution of data about the monthly salaries of 90 sales representatives is found to have an arithmetic mean of N60,000, with a standard deviation of N15,000, and a coefficient of skewness of 0.92. Explain what these terms mean in describing the distribution of the sales reps' salaries.

#### **7.0 References**

1. Onwe, O. J. (2007) *Statistical Methods for Business and Economic Decisions: A Practical Approach (Lagos: Samalice)*
2. Hanke, J. E. and Reitsch, A. G. (1991) *Understanding Business Statistics* (Homewood, IL: Richard Irwin)

## **UNIT 5: PRESENTATION AND ANALYSIS OF BUSINESS DATA: ANALYTICAL TOOLS (ONE)**

### **Content**

- 1.0 Introduction
- 2.0 Objectives
- 3.0 Preliminaries and Tools of Data Analysis
- 3.1 Preliminaries
- 3.2 Tools for Data Presentation and Analysis
- 3.3 Self- Assessment Exercise
- 4.0 Conclusion
- 5.0 Summary
- 6.0 Tutor-Marked Assignment
- 7.0 References

### **1.0 Introduction**

It has been a general observation that researchers and research students alike often find it difficult presenting and analysing data obtained from surveys and related studies. It is on this reasoning that this unit discusses the basic principles and tools of data presentation. This will serve as an eye opener for students of business statistics.

### **2.0 Objectives**

Having gone through this unit, you should be able to:

1. Understand the preliminary activities required for the analysis of statistical data
2. Know the tools you can use in your data analysis
3. Add to your knowledge of data analysis

### **3.0 Preliminaries and Tools of Data Analysis**

This section discusses the needed preparations for analysis and presentation of statistical data. You will also look at the recommended tools of data presentation and data analysis.

### **3.1 Preliminaries**

Before a given data can be presented for analysis and interpretation, it must be edited and coded.

#### **3.1.1 Editing**

By editing, we mean the examination of the given data in order to detect errors that may cause inconsistency if they are used for analysis in their original form. Through editing, these errors can be corrected accordingly. There are two types of editing:

1. Field editing
2. Central editing

**1. Field Editing.** This is a process whereby the researcher makes his or her records complete and correct without adding subjective information to his or her sources. It

involves the presentation of collected information in a readable form such that all information gathered is properly reported.

**2. Central Editing ensures maximum consistency in information by correcting any inconsistency in the collected data, which might create problems in the analysis and interpretation of the results.** There are four possible errors that should be watched out for in central editing:

- (i) Arithmetic or Numeric Errors. These are errors that involve the wrong recording of the units in responses. Information may be wrongly reported in months when they are requested to be in years.
- (ii) Errors of Transposition. This error occurs when a response is entered in the wrong place. For example, a question that asks for the respondent's State of Origin may generate an answer about the ethnic group to which the respondent belongs.
- (iii) Errors of Inappropriate Response. These errors occur when a respondent gives a relevant response but not in the exact form that is required.
- (iv) Errors of Omission. These errors are difficult to edit and, in most cases, are interpreted to mean "no response"

### **3.1.2 Coding**

Coding enables the researcher group responses into limited number of classes or categories for ease of analysis. A set of rules are observed when grouping responses into classes or categories. The rules include:

1. Exclusiveness. This requires that a data item or response must be placed in one cell of a given category set. This is essentially relevant in a situation where a respondent or response fits into several categories.
2. Exhaustiveness. This is a requirement that all data categories or cells must be able to provide the necessary data for answering the research questions and testing the research hypotheses.

Research is generally meant to generate data for analysis, and this results in a large volume of statistical information mostly in its raw stage. For the generated data to be used for the objectives of a research, they have to be reduced to manageable dimensions through analysis.

***Two types of data analysis are in place:***

1. Descriptive analysis
2. Causal analysis

**Descriptive Analysis** deals with the study of such research variables as profiles of the respondents, organisations, groups or any other subjects. Descriptive analysis may be either qualitative, involving frequency distributions, measures of central tendency and dispersion.

**Causal Analysis** involves the use of more sophisticated statistical tools to draw inferences based on the research information. Such statistics as Chi-Squares student t and least-squares estimators are some of the important statistical tools that can be used

under this heading.

### 3.2 Tools for Data Presentation and Analysis

This section examines the relevant data presentation and analytical tools in business statistics. The tools will include:

Frequency distribution tools

Parametric tools

Non-parametric tools

#### 3.2.1 Frequency Distribution Tools

The frequency distribution tools basically comprise of the histogram and the pie chart.

##### 1. Histogram And Frequency Polygon

A histogram is a pictorial representation of a frequency distribution of a given grouped data. The histogram is a set of bars or blocks constructed from the grouped data. The height of each bar is represented by the frequency of the corresponding observation, and the horizontal axis of each bar is represented by the class width.

The Frequency Polygon is a diagram obtained by connecting the midpoints of the bars or blocks to form the histogram of the given set of grouped data

As an example, consider the frequency distribution of the monthly salary of 100 workers in a given company:

**Table 3.1. Frequency Distribution of Monthly salaries of 100 workers**

Monthly Salary (N'00s)	No. of Workers (f)
640 – 659	7
660 – 679	20
680 – 699	33
700 – 719	25
720 – 739	11
740 – 759	4
	<u>f = 100</u>

The frequency distribution is represented by a two-dimensional graph, with the vertical axis labeled “frequencies (f)” and the horizontal axis labeled. “Actual Class Limits”.

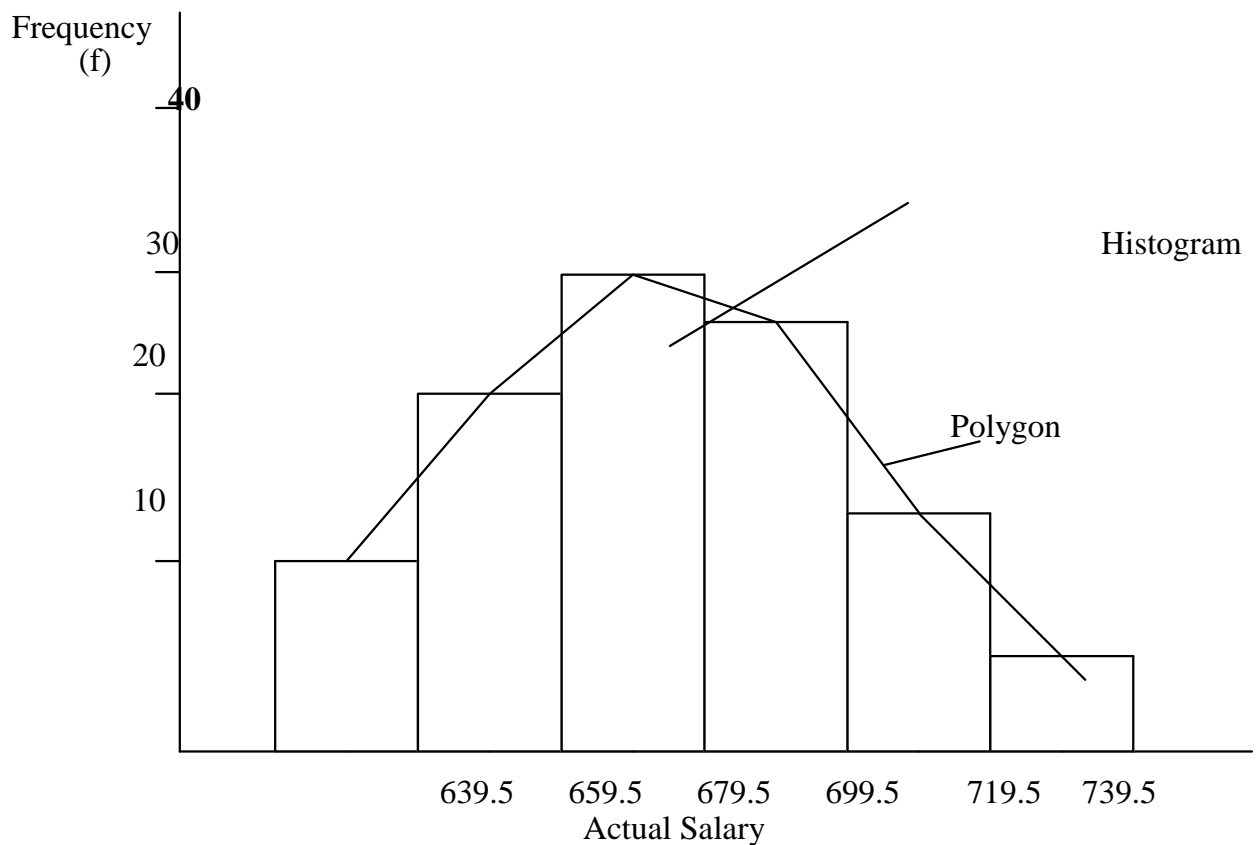
*The Actual Class Limits are obtained by subtracting a constant value, 0.5, from each lower class limit of the frequency distribution and adding same (0.5) to the corresponding upper Class Limits. This process is presented below.*



**Table 3.2: The Actual Frequency Distribution of Monthly Salaries of 100 workers**

Stated salary limits	Actual salary limits	No. of Workers
640-659	639.5-659.5	7
660-679	659.5-679.5	20
680-699	679.5-699.5	33
700-719	699.5-719.5	25
720-739	719.5-739.5	11
740-759	739.5-759.5	4

**Figure 3.1: The Graph of Histogram and Polygon**



### The Pie Chart

A pie chart shows the totality of the data being represented using a single circle (a “pie”). The circle is split into sectors, the size of each being drawn in proportion to the class frequency. For easy analysis, each sector is shaded differently.

As an example, consider Table 3.3 below, showing the non-managerial workforce employed at a given factory.

**Table 3.3: Non-Managerial Workforce**

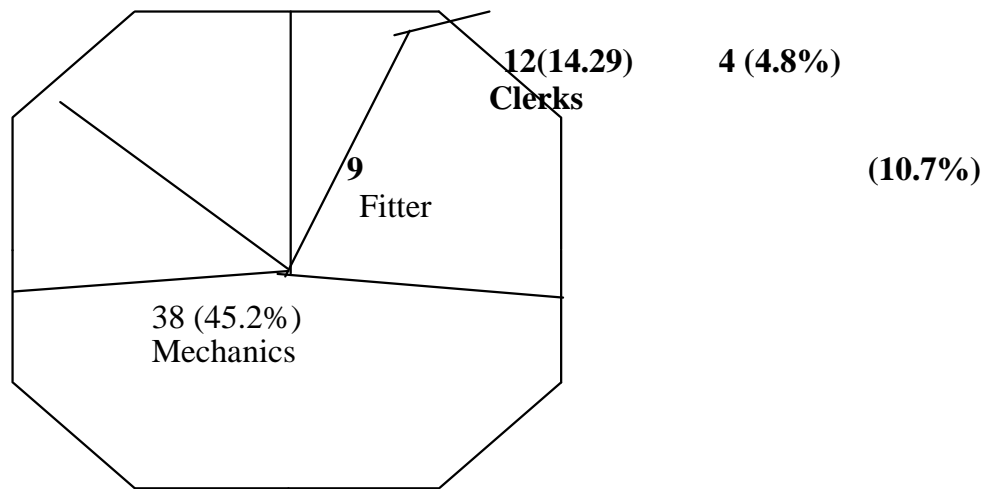
Job Description	Number Employed
Labourers	21
Mechanics	38
Fitters	9
Clerks	12
Draughtsmen	4
<b>Total</b>	<b>84</b>

The pie chart for these data is as shown in Figure 3.2

**Figure 3.2: Workforce Employed at a Factory**

**Draughtsmen**

**21(25%)  
Labourers**



### 3.2.2 Parametric Tools

Our major interest here is on the tools used in testing parametric hypotheses. That is, hypotheses concerning population parameters, such as the population mean,  $\mu$ , the population variance,  $\sigma^2$ , and the population standard deviation,  $\sigma$ . For practical purposes, we lay emphasis on the student t-statistic as a parametric tool, among such other tools as the Z-statistic.

#### The Student t-Statistic

The student t-statistic is used in testing hypotheses concerning the population mean or average, especially in cases involving a relatively small sample size; ( $n < 30$ ). Its application is better illustrated by an example. Before any illustration, it will be appropriate to review the process of hypotheses testing as it concerns the student t- statistic.

#### The Decision Values of $t$

The aim of any hypothesis testing is to either accept or reject a given null hypothesis. The decision to either accept or reject any hypothesis is based on two values of the test

statistic, in this case the t-statistic. These two values are referred to as the decision values, including:

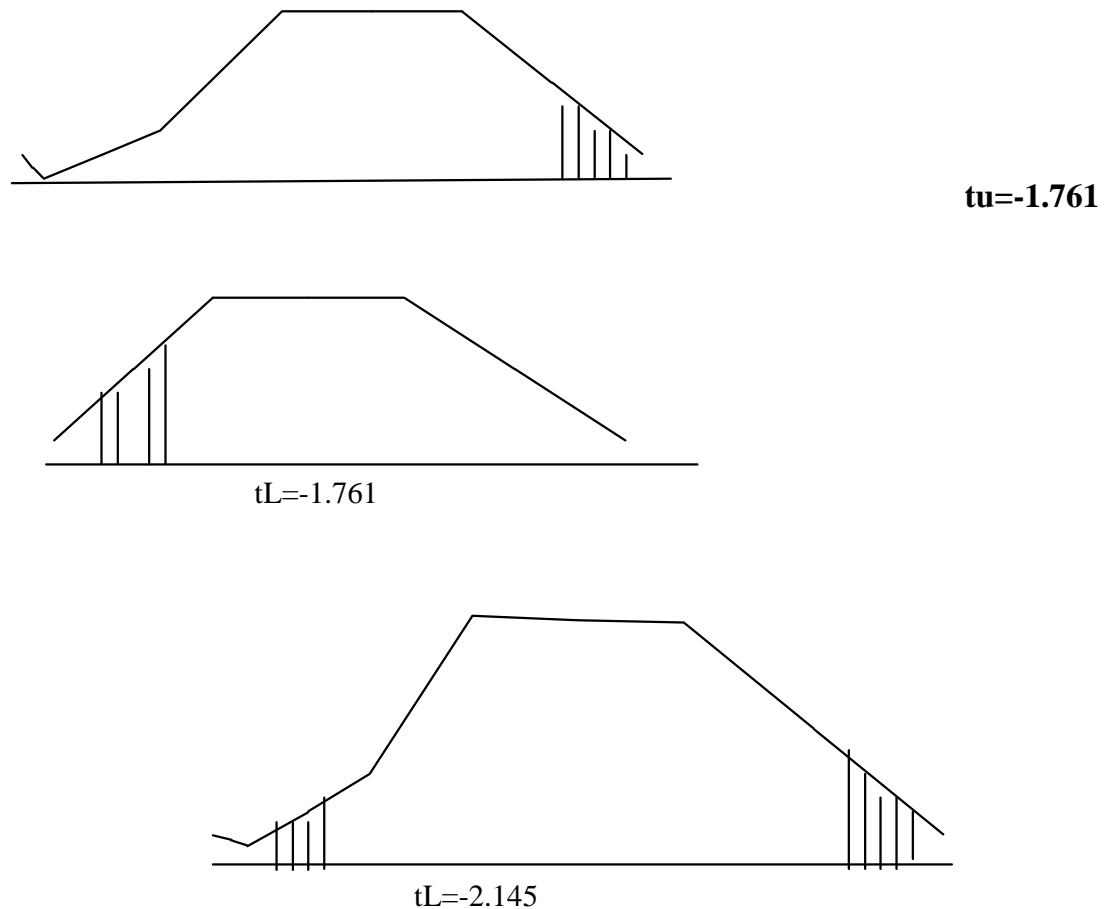
1. The critical or rejection value
2. The calculated or statistical value.

#### *The Critical or Rejection Value of t*

The critical value of t is obtained from the t distribution table, with known level of significance,  $\alpha$ , and the number of degrees freedom,  $n-1$ , where n represents the number of observations.

Assume  $n = 15$  observations. With 5% level of significance and  $n-1 = 15-1 = 14$  degrees of freedom obtained from the t-distribution table as  $t=1.761$ , for one-sided rejection region or  $t=2.145$ , for two-sided rejection regions (see figure 3.3)

**Figure 3.3: Rejection Values of t**



$t_u = 2.145$

#### *The Calculated Value of t ( $t_c$ )*

This is the value obtained using the statistical information needed for testing the stated hypotheses. To calculate this value, we simply apply the formula:

$$t_c = \frac{\bar{X} - \mu_0}{\frac{S}{\sqrt{n-1}}}$$

Where  $\bar{X}$  = sample mean  
 $\mu_0$  = the hypothesized population mean  
 $S$  = the standard error of the sample mean.  
 $S = \frac{S}{\sqrt{n-1}}$

Recall that  $S$  = Sample Standard deviation.

Having obtained the two values of the test-statistic, we compare them to know whether or not the null hypothesis should be rejected.

If  $t_c > t_u$ , reject  $H_0$  (the null hypothesis)

If  $t_c < t_u$ , accept  $H_0$  (the null hypothesis)

And

If  $t_c < t_L$ , reject  $H_0$  (the null hypothesis)

If  $t_c > t_L$ , accept  $H_0$  (the null hypothesis)

Where  $t_c$ ,  $t_u$ , and  $t_L$  refer to calculated, Upper, and Lower critical values of  $t$  respectively.

### Example

Upon examining the monthly billing records of a mail-order book company, the auditor takes a sample of 10 of its unpaid accounts. The accounts receivable were:

~~N4~~, ~~N5~~, ~~N7~~, ~~N7~~, ~~N9~~, ~~N10~~, ~~N11~~, ~~N12~~, ~~N18~~, N33.

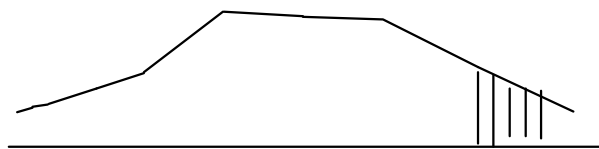
Based on the observed accounts, the auditor hypothesizes that on the average, the accounts receivable is greater than N15. Following this hypothetical belief, we want to test at one percent level of significance, the hypotheses:

$H_0: \mu = N15$  (null hypothesis)

$H_A: \mu > N15$  (Alternate hypothesis)

### Solution

First the inequality sign ( $>$ ) indicates the use of one-sided, upper rejection region in the decision of either to reject or accept the hypothesis. It follows that, using the table on  $t$ -distribution, the critical or rejection value of  $t$  is as indicated below, with  $\alpha = 0.01$  and  $n-1 = 10-1 = 9$  degrees of freedom



$$t_u = 2.821$$

The calculated value of  $t$  is obtained as follows:

$$t_c = \frac{\bar{X} - \mu_0}{\frac{S}{\sqrt{n-1}}}$$

Where  $\bar{X} = \frac{\sum X}{n} = \frac{166}{10} = 16.6$

$\mu_0 = 15$	
$H_0: \mu = 15$	
$H_1: \mu \neq 15$	
$\alpha = 0.05$	
$n = 10$	
$S = \sqrt{\frac{\sum X^2 - (\sum X)^2 / n}{n - 1}}$	
$\sum X$	$\sum X^2$
4	16
5	25
7	49
7	49
9	81
10	100
11	121
12	144
18	324
33	1089
$\sum X = 116$	$\sum X^2 = 1998$

$$\frac{(\sum X)^2}{n} = \frac{(116)^2}{10} = \frac{13456}{10} = 1345.6$$

$$S = \sqrt{\frac{\sum X^2 - \frac{(\sum X)^2}{n}}{n - 1}} = \sqrt{\frac{1998 - 1345.6}{9}} = \sqrt{72.49}$$

$$= 8.51$$

$$\text{Thus, } S_{\bar{X}} = \frac{S}{\sqrt{n}} = \frac{8.51}{\sqrt{10}} = 0.851$$

It follows that,

$$t_c = \frac{\bar{X} - \mu_0}{S_{\bar{X}}} = \frac{11.6 - 15}{0.851} = -3.995$$

### Decision

Since  $t_c(-3.995) < t_{\alpha/2}(2.82)$ , we accept the null hypotheses ( $H_0 = \mu = 15$ ), and conclude that on the average, the account receivable is not significantly greater than 15.

### 3.3 Self -Assessment Exercise

With simple examples, outline the two types of data analysis.

### 4.0 Conclusion

This unit discusses the preliminary requirements in data presentation and analysis. Two major types of analysis were of major focus including: (i) descriptive analysis; and, (ii) causal analysis. We were meant to understand that there are three major tools of data

analysis: (i) Frequency distribution tools; (ii) Parametric tools; and, (iii) Non-parametric tools. Among the tools discussed in the unit are the frequency distribution tools and one of the parametric tools, the student t-statistic.

## 5.0 Sum mary

Before a given data can be presented for analysis and interpretation, it must be edited and coded.

By editing, we mean the examination of the given data in order to detect errors that may cause inconsistency if they are used for analysis in their original form. Through editing, these errors can be corrected accordingly. There are Two types of editing were noted: (i) field editing; and, (ii) central editing.

There are basically two types of data analysis:

1. Descriptive analysis
2. Causal analysis

Descriptive Analysis deals with the study of such research variables as profiles of the respondents, organisations, groups or any other subjects. Descriptive analysis may be either qualitative, involving frequency distributions, measures of central tendency and dispersion. Causal Analysis involves the use of more sophisticated statistical tools to draw inferences based on the research information. Such statistics as Chi-Squares student t and least-squares estimators are some of the important statistical tools that can be used under this heading.

The basic tools of data analysis include:

1. Frequency distribution tools
2. Parametric tools
3. Non-parametric tools

## 6.0 Tutor-Marked Assignm ent

The daily sales of a marketing outlet has been sampled as follows:

Sales (X): 8, 12, 9, 16, 14, 18, 20, 26, 10, 17.

(N' million)

Test at 5 percent level of significance, the hypothesis:

HO:  $\mu = 15$

HA:  $\mu > 15$

## 7.0 References

1. Onwe, O. J. (2007) Statistical Methods for Business and Economic Decisions: A Practical Approach (Lagos: Samalice)
2. Hanke, J. E. and Reitsch, A. G. (1991) Understanding Business Statistics (Homewood, IL: Richard Irwin)

## **UNIT 6: PRESENTATION AND ANALYSIS OF BUSINESS DATA: ANALYTICAL TOOLS (TWO)**

### **Content**

- 1.0 Introduction
- 2.0 Objectives
- 3.0 The Z-Statistic and Non-Parametric Tools of Analysis
- 3.1 The Z-Statistic
- 3.2 The Non-Parametric Tools
- 3.3 Self- Assessment Exercise
- 4.0 Conclusion
- 5.0 Summary
- 6.0 Tutor-Marked Assignment
- 7.0 References

### **1.0 Introduction**

This unit is a continuation of our discussions on the tools of data analysis. It completes the presentation of parametric tools and ends the discussions with the non-parametric tools.

### **2.0 Objectives**

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

1. Learn additional tools of statistical analysis
2. Know more about hypotheses testing
3. Apply statistical tools in your analysis of data for business decisions.

### **3.0 The Z-Statistic and Non-Parametric Tools of Analysis**

This section completes the discussions on parametric tools of analysis with the Z-statistic and proceeds with the non-parametric tools such as the chi-square statistic.

#### **3.1 The Z-Statistic**

Like the t-statistic, Z-statistic is a parametric tool used in testing parametric hypotheses.

In this discussion, we are interested in the use of Z-statistic in testing hypotheses that involves percentages or proportion of subjects having particular responses to particular research question or issues.

#### **3.1.1 Tests of Hypotheses For Sample Proportions: One Sample**

For large samples, the applicable test-statistic for sample proportions is the Z-statistic.

The rejection value of Z for a given level of significant is obtained from distribution table. And the calculated value of Z for tests of proportions can be obtained by:

$$ZC = \frac{pS - P}{spS}$$

Where pS = the sample proportion or percentage

$$= \frac{\text{Number of successes in a sample}}{\text{Sample size}}$$

$P$  = Population proportion as indicated by the null hypothesis,  $H_0$ .

$$\text{sps} = \frac{\text{standard error of the sample proportion}}{= \frac{\sqrt{P(1-P)}}{n}}$$

### Example

Consider a supermarket that sells packaged men's shirts. The management learns from past experience that 15 percent of all shirts sold are returned to the supermarket by customers who complain that the shirts do not fit properly. In an attempt to correct this situation, the manufacturer of the shirts redesigned them and finds that, of the next 500 sales, 60 shirts were returned.

Problem is to test, at 5 percent level of significance, to see if there has been significant decrease in the population proportion of returns.

### Solution

We are required to test,

$$H_0: P = 0.15$$

$$H_A: P < 0.15 \quad \alpha = 0.05$$

Observe that the alternative hypothesis,  $H_A$ , calls for a one-sided test, using the lower rejection region. The rejection value of  $Z$  at  $\alpha = 0.05$  is  $-1.64$  (from the  $Z$ -distribution table) as indicated by figure 3.1 below.

**Figure 3.1: Rejection Value for  $Z$  ( $\alpha = 0.05$ )**



The calculated value of  $Z$  can be obtained as:

$$Z_c = \frac{PS - P}{\text{sps}}$$

where  $ps = \frac{60}{500} = 0.12$  = proportion of returns

$P = 0.15$  (from the null hypothesis,  $H_0$ )

$$\text{sps} = \frac{\sqrt{P(1-P)}}{n} = \frac{\sqrt{0.15(1-0.15)}}{500}$$



$$= \frac{\sqrt{0.15(0.85)}}{500} = 0.016$$

It follows that,

$$\frac{Z = 0.12 - 0.15}{0.016} = -1.875$$

### Decision

Since  $Z(-1.875) < Z_L(-1.64)$ , we reject  $H_0$  and conclude that there has been a significant decrease in the population proportion of returns as a result of the changes made in the design of the shirts.

### 3.2 The Non-Parametric Tools

There exists certain phenomena or variables in business statistics which can hardly be described quantitatively. The mode of gathering information on these variables requires the use of nominal and ordinal scales which do not meet standard requirements of parametric statistics. The valid inferential statistical tests for these type of variables is non-parametric tests.

Non-parametric test procedures involve either:

1. Those procedures whose test-statistic does not depend upon the form of the underlying population distribution from which the sample data were drawn;
- or
2. Those procedures which are not concerned with the population parameters;
- or
3. Those procedures for which the data are of little strength to warrant meaningful arithmetic operations.

*Conditions under which non-parametric statistics can be used are outlined as follows:*

1. When the hypothesis to be tested does not involve a population parameter.
2. When there are no assumptions of normality about the distribution of the variables.
3. When data are gathered from such weak measuring scales as ranking, frequency counts, and some subjective measuring scales.
4. When results are needed fast and no statistical sophistication is required.

There are two most commonly used non-parametric statistical methods in business statistics, including:

1. Chi-square ( $\chi^2$ ) statistic
2. Spearman Rank Correction ( $r_s$ )

At this level of discussion, however, we shall examine only the chi-square method.

### 3.2.1 The Chi-Square Method

The Chi-square test can only indicate whether or not a set of observed frequencies differ significantly from the corresponding set of expected frequencies and not the direction in which they differ. In practice, there are two types of Chi-square ( $\chi^2$ ) tests:

1. Test of Goodness-of-fit
2. Test of Independence and /or Homogeneity.

#### *The Test of Goodness-of-fit*

The test of goodness-of-fit is employed in situations whereby the researcher's objective is to find out whether or not a set of observed frequencies fits closely the theoretical or expected frequencies. The applicable formula for the test of goodness-of-fit is:

$$\chi^2 = \sum_{i=1}^r \frac{(f_{oi} - f_{ei})^2}{f_{ei}}$$

Where  $f_o$  = observed frequencies  
 $f_e$  = expected frequencies  
 $\chi^2$  = calculated value of  $\chi^2$

The number of degrees of freedom for  $r$  number of row entries is  $r-1$  for the test of goodness-of-fit. Given the number of degrees of freedom and the level of significance, the rejection value of  $\chi^2$  can be obtained from the  $\chi^2$ -distribution table.

Consider the following practical example:

In a study to determine customer preferences among two banking services: Current account and Savings account, 120 customers were asked to respond to the question, "please indicate which service you prefer most:"

Current Account ----- ? Savings Account----- ?

The responses indicate that 65 of the respondents preferred current account while 55 preferred savings account. This data can be summarized thus:

**Table 3.1: Number Preferring Banking Services**

Bank	Banking Service	Number preferring
Cc	Current Account	65
	Savings Account	55
	<b>Total</b>	120

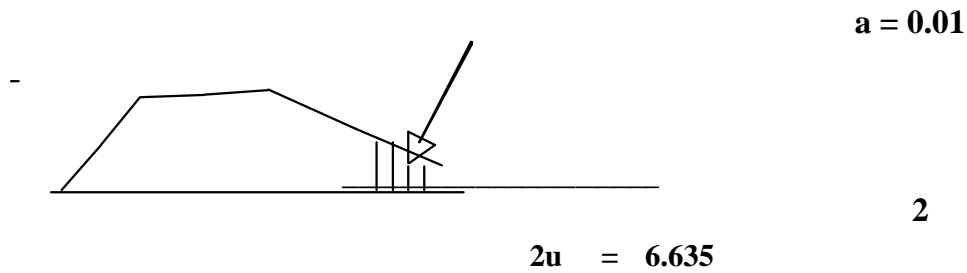
We want to test the following non-parametric hypotheses:

$H_0$  The customers do not show any preference for either Current Account or Savings Account.

$H_A$ : The customers show some preference for either Current Account or Savings Account.

We test the hypotheses at 1 percent level of significance (that is,  $\alpha = 0.01$ ). so that the rejection value, with  $r-1 = 2-1 = 1$  degree of freedom, where  $r$  is the number of row categories, that is, Current Account and Savings Account, is 6.635 from the  $\chi^2$ -distribution table. This rejection value is illustrate in figure 3.1 below.

**Figure 3.1: Rejection value for  $\chi^2$  (d.f = 1,  $\alpha = 0.01$ )**



The calculated value of  $\chi^2$

Bank	Banking Service NumNo. preferring (foi)	ExpExpected Preference (fei)
Curr	Current Account 65	60
Savi	Savings Account 55	60
Total	120	120

Note: Expected frequency (fei) =  $\frac{\text{Number of Respondents}}{\text{Number of rows}}$

$$\frac{120}{2} = 60$$

It follows that:

$$\begin{aligned} \chi^2 &= \sum \frac{(foi - fei)^2}{fei} \\ &= \frac{(65-60)^2}{60} + \frac{(55-60)^2}{60} \\ &= 0.42 + 0.42 \\ &= 0.84 \end{aligned}$$

### Decision

Since  $\chi^2 (0.84) < \chi^2_{critical} (6.635)$ , we do not reject  $H_0$ , implying that the customers do not show any preference for either Current or Savings Account. The observed pattern of 65 preferring Current Account and 55 preferring Savings Account is not statistically significant.

### Test of Independence And Homogeneity

A test of independence aims at ascertaining whether two or more variables are dependent upon each other. While tests of homogeneity aim at ascertaining whether the characteristics of two or more population variables are same.

Two variables are said to be independent (or not associated with each other) if the distribution of one is not related to the distribution of the other. Chi-square ( $\chi^2$ ) test of independence can therefore be used to test if the distribution of two variables in a population are independent of each other.

Two important assumptions of the Chi-square test of independence are worth mentioning:

1. The relevant data are randomly drawn from a population of interest.
2. Two criteria are used in the cross-classification of the observations, and each observation must belong to only one criterion. This cross-classification gives rise to what is referred to as an n-contingency table.

A Contingency table is the table in which the observed and expected frequencies with the various levels of two variables are presented. The table is named by the number of rows (r) and number of columns (c) it has, as an (r x c) contingency table. If the table has 2 rows and 3 columns, it will be referred to as a (2 X 3) contingency table. Note that in practice, the expected frequency is recorded in the same cell as the observed frequency. The expected frequency is, however, differentiated from the observed frequency by enclosing it in a bracket inside the cell.

As an example, assume that the director of Enugu State Chamber of Commerce is interested (for planning purposes) in learning more about the international participants in its annual International Trade Fares. From the Local Hotel Association, a list of past participants is obtained. The director plans to send questionnaires to the participants on the list to find out why they participated in the Trade Fares, how much they spent, how long they stayed, and what their future plans are. The director speculates that an offer to send each respondent a free gift will increase the rate of return of responses. To test this proposition, questionnaires were mailed to a random sample of 30 persons with the offer of the free gift. Questionnaires were also mailed to another random sample of 30 persons with no gift offer. The results are shown in the following (2 X 2) contingency table

**Table 3.2: A contingency Table of Numbers responding to Questionnaire**

	Questionnaire		
	Returned	Not Returned	
Gift Offered	22	8	30
Gift Not Offered	14	16	30
<b>Total</b>	<b>36</b>	<b>24</b>	<b>60</b>

The director's proposition can formally be hypothesized as follows:

Ho: The population proportion of questionnaire Returns is independent of the promise of a gift.

HA: The population proportion of questionnaire Returns is dependent upon the promise of a gift.

We want to test these hypotheses at 5 percent level of significant ( $\alpha = 0.05$ ) to see if there is a significant difference in the proportion of returns when the gift is offered.

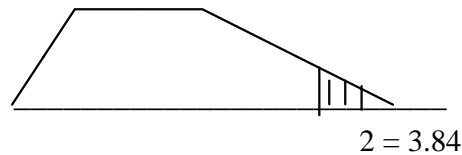
### Solution

From  $\chi^2$ -distribution Table, we observe the rejection value of  $\chi^2$ , with  $(r-1)(c-1) = (2-1)$

$(2-1) = 1$  degree of freedom, at  $\alpha = 0.05$  to be  $\chi^2_{0.05} = 3.84$

This can be illustrated as in figure 3.2 below

**Figure 3.2: Rejection value for  $\chi^2$  (df= 1;  $\alpha = 0.05$ )**



For tests of Independence and Homogeneity, the calculated value of Chi-square ( $\chi^2$ ) is obtained by:

$$\chi^2 = \sum_{ij} \frac{(f_{ij} - f_{ij}^e)^2}{f_{ij}^e}$$

Where the expected frequencies,  $f_{ij}^e$  can be obtained by the definition:

$$f_{ij}^e = \frac{R_{it} C_{jt}}{N}$$

Where,  $R_{it}$  = row total  
 $C_{jt}$  = column total  
 $N$  = the grand total

From the above table 3.5, the totals are

$$R_{1t} = 30; \quad R_{2t} = 30; \quad C_{1t} = 36 \\ C_{2t} = 24; \quad N = 60$$

It follows that the expected frequencies,  $f_{ij}^e$ , will be:

$$f_{11}^e = \frac{R_{1t} C_{1t}}{N} = \frac{(30)(36)}{60} = 18$$

$$f_{12}^e = \frac{R_{1t} C_{2t}}{N} = \frac{(30)(24)}{60} = 12$$

$$f_{21}^e = \frac{R_{2t} C_{1t}}{N} = \frac{(30)(36)}{60} = 18$$

$$f_{22}^e = \frac{R_{2t} C_{2t}}{N} = \frac{(30)(24)}{60} = 12$$

Note in our formulations that,  $f_{ij}^e$  = expected frequency of observation in the  $i$ th row and  $j$ th column, so that:

- $f_{11}^e$  = Expected frequency in the 1st row and 1st column
- $f_{12}^e$  = Expected frequency in the 1st row and 2nd column
- $f_{21}^e$  = Expected frequency in the 2nd row and 1st column
- $f_{22}^e$  = Expected frequency in the 2nd row and 2nd column

As mentioned earlier, the expected frequencies in our example can be presented along with the corresponding observed frequencies in the (2X2) contingency table as follows:

Questionnaire			
Gift Returned Not	Returned Total		
Offered 22	(18)	8 (18) 30	
Not offered 14	(18)	16 (12) 30	

Total 36	24	60
----------	----	----

The figures in brackets are the corresponding expected frequencies.

The calculated  $\chi^2$  can now be presented as follows:

$$\begin{aligned}\chi^2 &= \sum \frac{(f_{ij} - E_{ij})^2}{E_{ij}} \\ &= \frac{(22 - 18)^2}{18} + \frac{(14 - 12)^2}{12} + \frac{(6 - 12)^2}{12} \\ &= 0.89 + 1.33 + 0.89 + 1.33 = 4.44\end{aligned}$$

### Decision

Since  $\chi^2(4.44) > \chi^2_{0.05}(3.84)$ , we reject the null hypothesis,  $H_0$ , and infer that the population proportion of questionnaire returns in our example is dependent upon the promise of a gift.

### 3.3 Self -Assessment Exercise

State and test at 1 percent significance level any parametric hypothesis you think of.

### 4.0 Conclusion

This unit concludes our discussions of tools used in data presentation and analysis. The use of parametric tools was restricted to the chi-square statistic involving tests of goodness-of-fit and tests of independence and homogeneity.

### 5.0 Summary

Non-parametric test procedures involve either:

1. Those procedures whose test-statistic does not depend upon the form of the underlying population distribution from which the sample data were drawn;
- or
2. Those procedures which are not concerned with the population parameters;
- or
3. Those procedures for which the data are of little strength to warrant meaningful arithmetic operations.

The Chi-square test can only indicate whether or not a set of observed frequencies differ significantly from the corresponding set of expected frequencies and not the direction in which they differ. In practice, we observe two types of Chi-square ( $\chi^2$ ) tests:

1. Test of Goodness-of-fit
2. Test of Independence and /or Homogeneity.

The test of goodness-of-fit is employed in situations whereby the researcher's objective is to find out whether or not a set of observed frequencies fits closely the theoretical or expected frequencies. A test of independence, on the other hand, is aimed at ascertaining whether two or more variables are dependent upon each other. Tests of homogeneity aim at ascertaining whether the characteristics of two or more population variables are same.

### 6.0 Tutor-Marked Assignment

A sample survey of users of public libraries was conducted to investigate the reading habits of men and women. The results are as follows:

	Type of literature preferred	
	Fiction	Non – fiction
Men	132	102
Women	168	98
<b>Total</b>	<b>300</b>	<b>200</b>

Test at five percent level to see if there is any evidence that women show greater preference for fiction than men.

### 7.0 References

1. Onwe, O. J. (2007) Statistical Methods for Business and Economic Decisions: A *Practical Approach* (Lagos: Samalice)
2. Hanke, J. E. and Reitsch, A. G. (1991) Understanding Business Statistics (Homewood, IL: Richard Irwin)

## **UNIT 7: PROBABILITY THEORIES AND BUSINESS APPLICATIONS (ONE)**

### **Content**

- 1.0 Introduction
- 2.0 Objectives
- 3.0 The Concept of Probability, Theory of Sets, and Definitions of Probability
  - 3.1 The Concept of Probability
  - 3.2 The Theory of Sets
  - 3.3 Definitions of Probability
  - 3.4 Self- Assessment Exercise
- 4.0 Conclusion
- 5.0 Summary
- 6.0 Tutor-Marked Assignment
- 7.0 References

### **1.0 Introduction**

In this unit, we develop the basic principles of probabilistic analysis, with special emphasis on the applications of set operations, events, probability Laws, computation of probabilities, and their applications to business decisions. The aim is to enable you make effective decisions under uncertainties.

### **2.0 Objectives**

By the time you must have seriously gone through this unit, you should be able to:

1. Know the principles underlying the concept of probability
2. Know about sets and set operations
3. Define probability

### **3.0 The Concept of Probability, Theory of Sets, and Definitions of Probability**

#### **3.1 The Concept of Probability**

The basic elements of probability theory are the outcomes of the process or phenomenon under study. Each possible type of occurrence is referred to as an event. The collection of all the possible events is called the sample space.

A compound or joint event is an event that has two or more characteristics. For example, the event of a student who is “an economics major and obtained a B in exam or above average grade” is a joint or compound event since the student must be an economics major and have a B or above average grade.

The event “black ace” is also a compound event since the card must be both black and ace in order to qualify as a black ace.

Probability is a concept that most people understand naturally, since such words as “Chance,” “Likelihood,” “possibility” and “proportion are used as part of everyday speech. For example, most of the following which might be heard in any business situation are in fact statements of probability.

- a) “There is a 30% chance that this job will not be finished in time”.



- b) “There is every likelihood that the business will make a profit next year”.
- c) “Nine times out of ten he arrives late for his appointments”.
- In statistical sense, probability simply puts a well defined structure around the concept of everyday probability, enabling a logical approach to problem solving to be followed.

### 3.2 The Theory of Sets

A mathematical set is a collection of distinct objects, often referred to as elements or members. For example:

- (a) The employees of a company working in the Public Relations department could be represented as:

$PR = \{\text{Joseph, Adamu, Adebola, Nkom, Margerate}\}$

- (b) The location of shops for a big automobile parts dealer could be represented as:

$S = \{\text{Abuja, Enugu, Lagos, Aba, Onitsha, Kano, Ikot Ekpene}\}$

#### 3.2.1 Set Relations and Operations

##### 3.2.1.1 Set Relations

###### 1 Subsets.

A subset of a set, say A, is a set which contains some of the elements of set A. For instance:

If set  $A = \{a, e, I, o, u\}$ , then:

$X = \{a, e, i\}$  is a subset of A

$Y = \{e, i\}$  is a subset of A

$Z = \{I, o, u\}$  is a subset of A

###### 2. The number of A Set.

The number of a set A, written as  $n[A]$ , is defined as the number of elements in set A. For example:

If  $A = \{a, e, I, o, u\}$ , then  $n[A] = 5$  (that is, 5 elements in set A).

###### 3. Set Equality.

Two sets are said to be equal only when they have identical elements. For example:

If  $A = \{1, 2, 3\}$  and  $B = \{1, 2, 3\}$ , then Set A = Set B.

###### 4. Universal Set.

A universal set, denoted by U is a set containing different subsets of its elements. For example, a combination of different behaviours in a given population can be considered as a universal, while a selected sample of such behaviours are referred to as the subsets. A set of all English alphabets make up the universal set, while a set containing the vowels would be referred to as the subset.

###### 5. Complement of a Set.

The complement of a set A (denoted by  $A'$ ) contained in a given universal set, U, is the set of elements in the universal set that are not contained in set A. For example: If set A

represents the set of all skilled workers in a given universal set,  $U$ , then the complement of set  $A$ ,  $A'$ , is the set of unskilled workers who are members of the universal set.

## 6. Venn Diagrams.

Venn diagrams are simple pictorial representations of a set. They are useful for in demonstrating relationships between sets.

### 3.2.1.2 Operations on Sets.

There are two basic operations on sets, including:

1. Set Union
2. Set Intersection

#### 1. Set Union.

The union of two sets  $X$  and  $Y$ , denoted by  $X \cup Y$ , is defined as the set which contains all the elements in  $X$  and  $Y$ . For example, if  $X = \{1, 2, 3, 4\}$  and  $Y = \{3, 4, 5\}$ , then, without repetition of elements,  $X \cup Y = \{1, 2, 3, 4, 5\}$ , so that any element of  $X$  must be an element of  $X \cup Y$ ; similarly, any element of  $Y$  must be an element of  $X \cup Y$ .

#### 2. Set Intersection.

The intersection of two sets  $X$  and  $Y$ , denoted by  $X \cap Y$ , is defined as a set containing all the elements common to sets  $X$  and  $Y$ .

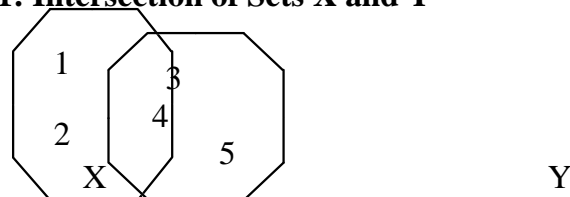
The union or intersection of three or more sets is a natural extension of the examples above and below.

As an example of intersection of two sets, consider the two sets  $X$  and  $Y$  above:

$X \cap Y = \{3, 4\}$ ; 3 and 4 are common to sets  $X$  and  $Y$ .

This can be represented by a venn diagram as follows:

**Figure 3.1: Intersection of Sets  $X$  and  $Y$**



The elements  $\{3, 4\}$  are contained in the circle common to sets  $X$  and  $Y$ .

### 3.2.2 Set Enumeration.

Set enumeration considers sets in terms of number of elements contained within the various areas defined by their union or intersection. Identifying the number of elements in these areas is known as set enumeration. As an illustration, consider the following enumeration problem.

Suppose an accounting firm currently employs 16 staff. Given that three staff have no formal qualifications, and of the seven staff who are graduates, 5 are also qualified as chartered accountants, it is possible to evaluate: (a) the number of staff who are non-graduates, chartered accountants and (b) the number of graduates who are not qualified chartered accountants.

The two values can be calculated as follows:

1. Since three staff have no formal qualifications, there should be  $16 - 3 = 13$  staff with at least one of the two qualifications that is, a graduate or a chartered accountants.
2. But there are 7 staff who are graduates, which implies that  $16 - 7 - 3 = 6$  must be non-graduate, qualified chartered accountants, which gives the answer to possibility (a) above.
3. In addition, since 5 staff are qualified chartered accounts and graduates, there would be  $7 - 5 = 2$  staff who are graduates only. This gives answer to possibility (b) above.

In an extended problem, the above approach is not structured enough for the solution. In the following discussions, we present a more structured procedure which solves the above problems and forms a basis for more logical approach for solving enumerations problems in general.

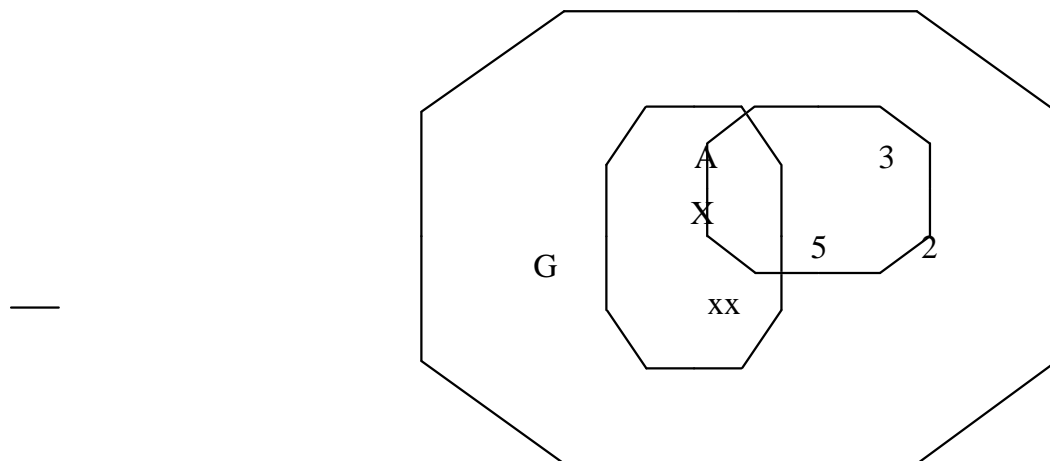
### 3.2.1 General Solutions to Enumeration Problems.

The above problem can be solved stepwise as follows:

**Step1:** we identify the attribute sets. The attribute sets in the problem at hand are 'Graduate status' (define this as set G) and 'Qualified chartered Accountants' (define this as set A). Note that there is a universal set involved, that is, 'Accounting Firm's staff'.

**Step 2:** we draw an outline Venn Diagram showing the sets that are involved:

**Figure 3.2: Universal Set of Accounting Firm's Staff**



Do the following to form the required Venn diagram:

- (i) Draw a big circle enclosing all figures and letters above. This becomes the universal set
- (ii) Draw a circle to enclose x and 5 and label it A to form Set A
- (iii) Draw another circle to enclose 5 and 2 and label it G to form Set .

Observe that the figure 3 stands alone in the universal set. And the figure 5 is at the intersection of Sets A and G.

The diagram contains four defined regions with values (2, 3, 5, and x), representing the number of elements in each of the regions.

**Step 3: Use the information given to fill in as much of the Venn diagram as possible.**

Observe that if two sets intersect within some defined universal set, there will be four defined distinct areas. In the present example: 'AnG'; 'A alone'; 'G alone'; and 'neither A nor G', technically defined as (A G).

Since there are total of 16 staff, the sum of the numbers in the four areas must be 16.

Interpreting the Venn diagrams above:

1. 5 staff are graduates and qualified chartered Accountants. This number is thus entered in the '(AnG)' area.
2. Since there are 7 graduates in all, there must be  $7 - 5 = 2$  graduates who are not chartered accountants. This number is thus entered in the 'G alone' area.
3. Three (3) of the staff had no qualifications at all. This number is thus entered in the 'neither A nor G or A G' area.

**Step 4: finally, we evaluate the number of elements in any remaining unknown areas.**

As indicated by the above diagram, x is an unknown number in the area 'A alone'. We need to solve for x as required, that is, the number of chartered accountants who are not graduates.

Putting x as the unknown number:

$$3 + 5 + 2 + x = 16 \text{ (the total number of staff)}$$

Solving for x, we get:  $x = 6$ .

**3.2.3 Notes on the General Enumeration Problem.**

Note in particular that there are:

- (a) 4 distinct areas for two attribute sets, and
- (b) 8 distinct areas for three attribute sets.

The general procedure for solving enumeration problems follows the following steps:

**STEP 1: Identify the attribute sets**

**STEP 2: Draw an outline Venn diagram**

**STEP 3: Use the information given to fill in as much of the diagram as possible**

**STEP 4: Evaluate the number of elements in unknown areas.**

Consider the following example:

A survey was carried out by a researcher, one of the aims being to discover the extent to which computers are being used by firms in a given area. 32 firms had both stock control and payroll computerized, 65 firms had just one of these two functions computerized, and 90 firms had a computerized payroll. If 22 firms had neither of these functions computerized, how many firms were included in the survey?

**Solution**

**STEP 1: The two attributes involved are computerized payroll, with set (say P), and computerized stock control, with set (say S).**

**STEP 2: Using standard notations,**

Let p = number of firms with a computerized payroll only;

$s$  = number of firms with a computerized stock control only;  
 $ps$  = number of firms with both payroll and stock control computerized; and  
 $x$  = number of firms with neither functions computerized,  
 construct a Venn diagram describing the situation.

**STEP 3: The following equations can be set up from the given information:**

$$ps = 32 \quad (1)$$

$$p + s = 65 \quad (2)$$

$$ps + p = 90 \quad (3)$$

$$x = 22 \quad (4)$$

**STEP 4: Substituting for  $ps = 32$  in equation 3, we get  $p = 58$**

Substituting for  $p = 58$  in equation, we get  $s = 7$ .

It follows that the number of firms included in the survey equals:

$$p + s + ps + x = 58 + 7 + 32 + 22 = 119.$$

### 3.3 Definitions of Probability

There are basically two separate ways of calculating probability including:

1. Calculation based on theoretical probability. This is the name given to probability that is calculated without an experiment, that is, using only information that is known about the physical situation.
2. Calculation based on empirical probability. This is probability calculated using the results of an experiment that has been performed a number of times. Empirical probability is often referred to as relative frequency or Subjective probability.

#### 3.3.1 Definition of Theoretical Probability

Let  $E$  represent an event of an experiment that has an equally likely outcome set,  $U$ , then the theoretical probability of event  $E$  occurring is written as  $Pr(E)$  and given by:

$$Pr(E) = \frac{\text{number of different ways that the event can occur}}{\text{number of different possible outcomes}} = \frac{n(E)}{n(U)}$$

Where  $n(E)$  = the number of outcomes in event set  $E$   
 $n(U)$  = total possible number of outcomes in outcome set,  $U$ .

If, for example, an ordinary six-sided die is to be rolled, the equally likely outcome set,  $U$ , is  $\{1,2,3,4,5,6\}$  and the event "even number" has event set  $\{2,4,6\}$ . It follows that the theoretical probability of obtaining an even number can be calculated as:

$$Pr(\text{even numbers}) = \frac{n(\text{even number})}{n(U)} = \frac{3}{6} = 0.50$$

#### Other Examples

A wholesaler stocks heavy (2B), medium (HB), fine (2H) and extra fine (3H) pencils which come in packs of 10. Currently in

stock are 2 packs of 3H, 14 packs of 2H, 35 packs of HB and 8 packs of 2B. If a pack of pencil is chosen at random for inspection, what is the probability that they are:  
 (a) medium (b) heavy (c) not very fine (d) neither heavy nor medium?

### Solutions

Since the pencil pack is chosen at random, each separate pack of pencils can be regarded as a single equally likely outcome. The total number of outcomes is the number of pencil packs, that is,  $2+14+35+8 = 59$ .  
 Thus,  $n(U) = 59$

$$\Pr(\text{medium}) = \frac{n(\text{medium})}{n(U)} = 35/59 = 0.593$$

$$\Pr(\text{heavy}) = \frac{n(\text{heavy})}{n(U)} = 8/59 = 0.136$$

(c) Note that the number of pencil packs that are not very fine is  $14+35+8 = 57$

$$\text{Therefore } \Pr(\text{not very fine}) = \frac{n(\text{not very fine})}{n(U)} = 57/59 = 0.966$$

(d) "Neither heavy nor medium" is equivalent to "fine" or "very fine" in the problem. There is  $2+14 = 16$  of these pencil packs.

$$\text{Thus, } \Pr(\text{neither heavy nor medium}) = \frac{n(\text{neither heavy nor medium})}{n(U)} = 16/59 = 0.271$$

### 3.3.2 Definition of Empirical (Relative Frequency) Probability.

If E is some event of an experiment that has been performed a number of times, yielding a frequency distribution of events or outcomes, then the empirical probability of event E occurring when the experiment is performed one more time is given by:

$$\Pr(E) = \frac{\text{number of times the event occurred}}{\text{number of times the experiment was performed}} = \frac{f(E)}{Sf}$$

Where  $f(E)$  = the frequency of event E

$Sf$  = total frequency of the experiment.

Put differently, the empirical probability of an event E occurring is simply the proportion of times that event E actually occurred when the experiment was performed.

For example, if, out of 60 orders received so far this financial year, 12 were not completely satisfied, the proportion,  $12/60 = 0.2$  is the empirical probability that the next order received will not be completely satisfied.

### Other examples

A number of families of a particular type were measured by the number of children they have, given the following frequency distribution:

Number of children:	0	1	2	3	4	5 or more
Number of families:	12	28	22	8	2	2

Use these information to calculate the (relative frequency) probability that another family of this type chosen at random will have:

- (a) 2 children      (b) 3 or more children      (c) less than 2 children

### Solutions

Here,  $S_f$  = total number of families = 74

$$(a) \quad \Pr(2 \text{ children}) = \frac{f(2 \text{ children})}{S_f} = \frac{22}{74} = 0.297$$

$$(b) \quad f(3 \text{ or more children}) = 8+2+2 = 12$$

$$\text{Thus, } \Pr(3 \text{ or more children}) = \frac{12}{74} = 0.162$$

$$(c) \quad \Pr(\text{less than 2 children}) = \frac{f(\text{less than 2 children})}{S_f} =$$

$$\frac{12 + 28}{74} = 0.541$$

### 3.4 Self -Assessment Exercise

With an appropriate example, discuss the four major steps that you can follow in solving a set enumeration problem.

### 4.0 Conclusion

The unit has exposed you to the basic principles of sets and how they can be used to solve business-related set enumeration problems. You learned the four steps that should be followed when solving set enumeration problems. The use of Venn diagrams is important in the application of these steps.

### 5.0 Summary

The basic elements of probability theory are the outcomes of the process or phenomenon under study. Each possible type of occurrence is referred to as an event. The collection of all the possible events is called the sample space. Probability is a concept that most people understand naturally, since such words as “Chance,” “Likelihood,” “possibility” and “proportion are used as part of everyday speech.

A mathematical set is a collection of distinct objects, often referred to as elements or members. *There are two basic operations on sets, including:*

1. Set Union
2. Set Intersection

Set enumeration considers sets in terms of number of elements contained within the various areas defined by their union or intersection. Identifying the number of elements in these areas is known as set enumeration.

The general procedure for solving enumeration problems follows for basic steps including:

**STEP 1: Identify the attribute sets**

**STEP 2: Draw an outline Venn diagram**

**STEP 3: Use the information given to fill in as much of the diagram as possible**  
**STEP 4: Evaluate the number of elements in unknown areas.**

There are two distinct definitions of probability: (i) definitions based on theory; and, (ii) empirical probability definitions.

Let E represent an event of an experiment that has an equally likely outcome set, U, then the theoretical probability of event E occurring is written as  $\Pr(E)$  and given by:

$$\Pr(E) = \frac{\text{number of different ways that the event can occur}}{\text{number of different possible outcomes}}$$

If E is some event of an experiment that has been performed a number of times, yielding a frequency distribution of events or outcomes, then the empirical probability of event E occurring when the experiment is performed one more time is given by:

$$\Pr(E) = \frac{\text{number of times the event occurred}}{\text{number of times the experiment was performed}} = \frac{f(E)}{Sf}$$

### 6.0 Tutor-Marked Assignment

A and B are two intersecting sets and a, b, ab, and x represent the usual symbols for the number of elements contained in the four defined areas of the associated Venn diagram.

Find the value of ab and a if:

$n[A] = 28$ ;  $a + b = 36$ ;  $x = 48$ ; the number of elements in the universal set is 96.

### 7.0 References

1. Francis, A (1998) Business Mathematics and Statistics, 5th edition (Great Britain: Ashford Colour Press).
2. Onwe, O. J. (2007) Statistical Methods for Business and Economic Decisions: A Practical Approach (Lagos: Samalice)



## **UNIT 8: PROBABILITY THEORIES AND BUSINESS APPLICATIONS (TWO)**

### **Content**

- 1.0 Introduction
- 2.0 Objectives
- 3.0 Laws of Probability, Computation of Probabilities, the Bayes Theorem, and Expected Values
- 3.1 Laws of Probability
- 3.2 Computational Formula for Multiple Occurrence of Events
- 3.3 Joint, Marginal, Conditional Probabilities and the Bayes Theorem
- 3.4 The Bayes Theorem
- 3.5 Probability and Expected Values
- 3.6 Self- Assessment Exercise
- 4.0 Conclusion
- 5.0 Summary
- 6.0 Tutor-Marked Assignment
- 7.0 References

### **1.0 Introduction**

This unit is an extension of unit 7. The unit will throw more light to the concept of probability by looking at the laws of probability, the computation of probability, and the application of probability theories in business and investment decisions.

### **2.0 Objectives**

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

1. Understand the laws of probability and how they can be used in estimating probability figures
2. Know the formula for estimating probabilities for multiple events
3. Understand the Bayes Theorem in relation to the computation of probabilities
4. Apply the probability theory in investment decisions.

### **3.0 Laws of Probability, Computation of Probabilities, the Bayes Theorem, and Expected Values**

#### **3.1 Laws of Probability**

There are four basic laws of probability:

1. Addition Law for mutually exclusive events
2. Addition Law for events that are not mutually exclusive
3. Multiplication Law for Independent events
4. Multiplication Law for Dependent events.

#### **3.1.1 Addition Law for Mutually Exclusive Events**

Two events are said to be mutually exclusive events if they cannot occur at the same time. The addition law states that if events A and B are mutually exclusive events, then:

$$\Pr(A \text{ or } B) = \Pr(A) + \Pr(B)$$

### Examples

The purchasing department of a big company has analysed the number of orders placed by each of the 5 departments in the company by type as follows:

**Table 3.1: Departmental Orders**

Type of order	Department					
	Sales	Purchasing	Production	Accounts	Maintenance	Total
Consumables	10	12	4	8	4	38
Equipment	1	3	9	1	1	15
Special	0	0	4	1	2	7
Total	11	15	17	10	7	60

An error has been found in one of these orders. What is the probability that the incorrect order:

- a) Came from maintenance?
- b) Came from production?
- c) Came from maintenance or production?
- d) Came from neither maintenance nor production?

### Solutions

a) Since there are 7 maintenance orders out of the 60,  $\Pr(\text{maintenance}) = \frac{7}{60} = 0.117$

b) Similarly,  $\Pr(\text{Production}) = \frac{17}{60} = 0.283$

c) Maintenance and production departments are two mutually exclusive events so that,

$$\begin{aligned}\Pr(\text{maintenance or production}) &= \Pr(\text{maintenance}) + \Pr(\text{production}) \\ &= 0.117 + 0.283 = 0.40\end{aligned}$$

d)  $\Pr(\text{neither maintenance nor production})$   
 $= 1 - \Pr(\text{maintenance or production}) = 1 - 0.4 = 0.6$

### 3.1.2 Addition Law for Events that are Not Mutually Exclusive Events

If events A and B are not mutually exclusive, that is, they can either occur together or occur separately, then according to the Law:

$$\Pr(A \text{ or } B \text{ or Both}) = \Pr(A) + \Pr(B) - \Pr(A) \cdot \Pr(B)$$

### Example

Consider the following contingency table for the salary range of 94 employees:

**Table 3.2: Contingency Table For The Salary range of 94 Employees**

Salary/month	Men	Women	Total
≥ N10,000 and above	20	37	57
Below N10,000	15	22	37
Total	35	59	94

What is the probability of selecting an employee who is a man or earns below N10,000 per month?

**Solution**

The two events of being a man and earning below N10,000 is not mutually exclusive. It follows that:

$$\begin{aligned}
 \Pr(\text{man or below N10,000}) &= \Pr(\text{man}) + \Pr(\text{below N10,000}) \\
 &\quad - \Pr(\text{man}) \cdot \Pr(\text{below N10,000}) \\
 &= \frac{35}{94} + \frac{37}{94} - \frac{(35)(37)}{(94)(94)} \\
 &= 0.372 + 0.394 - (0.372)(0.394) \\
 &= 0.766 - 0.147 \\
 &= 0.619 \text{ or } 61.9\%
 \end{aligned}$$

**3.1.3 Multiplication Law for Independent Events**

This Law states that if A and B are independent events, then:

$$\Pr(A \text{ and } B) = \Pr(A) \cdot \Pr(B)$$

As an example, suppose, in any given week, the probability of an assembly line failing is 0.03 and the probability of a raw material shortage is 0.1.

If these two events are independent of each other, then the probability of an assembly line failing and a raw material shortage is given by:

$$\begin{aligned}
 \Pr(\text{Assembly line failing and Material shortage}) &= (0.03)(0.1) \\
 &= 0.003
 \end{aligned}$$

**3.1.4 Multiplication Law for Dependent Events**

This Law states that if A and B are dependent events, then:

$$\Pr(A \text{ and } B) = \Pr(A) \cdot \Pr(B/A)$$

Note that  $\Pr(B/A)$  is interpreted as probability of B given that event A has occurred.

**Example**

A display of 15 T-shirts in a Sports shop contains three different sizes: Small, Medium and Large. Of the 15 T-shirts:

3 are small  
6 are medium  
6 are large.

If two T-shirts are randomly selected from the T-shirts, what is the probability of selecting both a small T-shirt and a large T-shirt, the first not being replaced before the second is selected?

**Solution**

Since the first selected T-shirt is not replaced before the second T-shirt is selected, the two events are said to be dependent events.

It follows that:

$$\begin{aligned}
& \Pr(\text{Small T-shirt and Large T-shirt}) \\
&= \Pr(\text{Small}) \cdot \Pr(\text{Large}|\text{Small}) \\
&= (3/15)(4/14) \\
&= (0.2)(0.429) \\
&= 0.086
\end{aligned}$$

### 3.2 Computational Formula for Multiple Occurrence of an Event

The probability of an event, E occurring X-times in n number of trials is given by the formula:

$$\Pr(E_n x) = C_{n,x} p^x q^{(n-x)}$$

Where  $C_{n,x} = \frac{n!}{x!(n-x)!}$

p = probability of success  
q = probability of failure  
p + q = 1

#### Example

Assume a drug store with 10 antibiotic capsules of which 6 capsules are effective and 4 are defective. What is the probability of purchasing the effective capsules from the drug store?

#### Solution

From the given information: The probability of purchasing an effective capsule is:

$$P = 6/10 = 0.60$$

$$\text{Since } p + q = 1; q = 1 - 0.60 = 0.40; n = 10; x = 6$$

$$\Pr(E_{10} 6) = \text{probability of purchasing the 6 effective capsules}$$

$$= C_{10,6} (0.6)^6 (0.4)^4$$

$$= 10! / (6!(10-6)!) (0.6)^6 (0.4)^4$$

$$= \frac{10 \cdot 9 \cdot 8 \cdot 7 \cdot 6!}{6! 4!} (0.0012)$$

$$= \frac{10 \cdot 9 \cdot 8 \cdot 7}{4!} (0.0012) = 0.252$$

Hence, the probability of purchasing the 6 effective capsules out of the 10 capsules is 25.2 percent

### 3.3 Joint, Marginal, Conditional Probabilities, and The Bayes Theorem

#### 3.3.1 Joint Probabilities

A joint probability implies the probability of joint events. Joint probabilities can be conveniently analysed with the aid of joint probability tables.

### 3.3.1.1 The Joint Probability Table

A joint probability table is a contingency table in which all possible events for a variable are recorded in a row and those of other variables are recorded in a column, with the values listed in corresponding cells as in the following example.

#### Example

Consider a research activity with the following observations on the number of customers that visit XYZ supermarket per day. The observations (or events) are recorded in a joint probability table as follows:

**Table 3.3: Joint Probability Table**

Age (Years)	Male (M)	Female (F)	Total		
Below 30 (B)	60	70	130		
30 and Above (A)	60	20	80		
Total	120	90	210		

We can observe four joint events from the above table:

Below 30 and Male (BnM) = 60

Below 30 and Female (BnF) = 70

30 and Above and Male (AnM) = 60

30 and Above and Female (AnF) = 20

Total events or sample space = 210

#### *The joint Probabilities*

The joint probabilities associated with the above joint events are:

$$\Pr(\text{BnM}) = \frac{60}{210} = 0.2857$$

$$\Pr(\text{BnF}) = \frac{70}{210} = 0.3333$$

$$\Pr(\text{AnM}) = \frac{60}{210} = 0.2857$$

$$\Pr(\text{AnF}) = \frac{20}{210} = 0.0952$$

### 3.3.2 Marginal Probabilities

The Marginal Probability of an event is its simple probability of occurrence, given the sample space. In the present discussion, the results of adding the joint probabilities in rows and columns are known as marginal probabilities.

The marginal probability of each of the above events:

Male (M), Female (F), Below 30 (B), and Above 30 (A) are as follows:

$$\Pr(M) = \Pr(BnM) + \Pr(AnM) = 0.2857 + 0.2857 = 0.57$$

$$\Pr(F) = \Pr(BnF) + \Pr(AnF) = 0.3333 + 0.0952 = 0.43$$

$$\Pr(B) = \Pr(BnM) + \Pr(BnF) = 0.2857 + 0.3333 = 0.62$$

$$\Pr(A) = \Pr(AnM) + \Pr(AnF) = 0.2857 + 0.0952 = 0.38$$

The joint and marginal probabilities above can be summarized in a contingency table as follows: \_\_

**Table 3.4: Joint And Marginal Probability Table.**

Age	Male (M)	Female (F)	Marginal Probability
Below 30 (B)	0.2857	0.3333	0.62
30 and Above (A)	0.2857	0.0952	0.38
Marginal Probability	0.57	0.43	1.00

### 3.3.3 Conditional Probability.

Assuming two events, A and B, the probability of event A, given that event B has occurred is referred to as the conditional probability of event A.

In symbolic term:

$$\Pr(A/B) = \frac{\Pr(A \text{ and } B)}{\Pr(B)} = \frac{\Pr(A) \Pr(B)}{\Pr(A)}$$

where  $\Pr(A/B)$  = conditional probability of event A

$\Pr(A \text{ and } B)$  = joint probability of events A and B

$\Pr(B)$  = marginal probability of event B

In general,

$$\Pr(A/B) = \frac{\text{Joint Probability of events A and B}}{\text{Marginal Probability of event B}}$$

### 3.4 The Bayes Theorem

Bayes theorem is a formula which can be thought of as “reversing” probability. That is, it finds a conditional probability, A/B given, among other things, its inverse, B/A.

According to the theorem, given events A and B,

$$\Pr(A/B) = \frac{\Pr(A) \Pr(B/A)}{\Pr(B)}$$

As an example in the use of Bayes theorem, if the probability of meeting a business contract date is 0.8, the probability of good weather is 0.5 and the probability of meeting

the date given good weather is 0.9, we can calculate the probability that there was good weather given that the contract date was met.

Let G = good weather, and m = contract date was met

Given that: Pr (m) = 0.8; Pr (G) = 0.5; Pr (m/G) = 0.9, we need to find Pr (G/m):

From the Bayes theorem:

$$\Pr (G/m) = \frac{\Pr (G) \Pr (m/G)}{\Pr (m)} = \frac{0.5(0.9)}{0.8} = 0.5625 \text{ or } 56.25\%$$

### 3.5 Probability and Expected Values

The expected value of a set of values, with associated probabilities, is the arithmetic mean of the set of values. If some variable, X, has its values specified with associated probabilities, P, then:

Expected value of X = E (X) =  $\sum PX$

#### Example

An ice-cream salesman divides his days into 'Sunny' 'Medium' or 'Cold'. He estimates that the probability of a sunny day is 0.2 and that 30% of his days are cold. He has also calculated that his average revenue on the three types of days is N220, N130, and N40 respectively. If his average total costs per day is N80, calculate his expected profit per day.

#### Solution

We first calculate the different values of profit that are possible since we are required to calculate expected profit per day, as well as their respective probabilities.

Given that Pr (sunny day) = 0.2; Pr (cold day) = 0.3

Since in theory, Pr (sunny day) + Pr (cold day) + Pr (medium day) = 1

It follows that:

Pr (medium day) = 1 - 0.2 - 0.3 = 0.5

The total costs are the same for any day (N80), so that the profit that the salesman makes on each day of the three types of day are:

**Sunny day:** N (220-80) = N140

**Medium day:** N (130-80) = N50

**Cold day:** N (40-80) = -N40 (loss)

We can summarize the probability table as follows:

Table 4.4 Probability Table

	Sunny	Medium	Cold
Profit (N)	140	50	-40
Probability	0.2	0.5	0.3

From this Table, the expected profit per day is:

$$E(\text{profit}) = N(0.2(140) + 0.5(50) - 0.3(40)) = N41.$$

### 3.6 Self -Assessment Exercise

Present a brief discussion of the laws of probability and their importance in estimation of probabilities of the occurrence of events.

### 4.0 Conclusion

You must have learned from this unit, the four laws of probability, classified under the addition laws and the multiplication laws. Of significance in this unit is its presentation of the computational formula for multiple events. You were also informed of three types of probabilities: (i) the joint probability; (ii) the marginal probability; and, (iii) the conditional probability. The application of these probabilities in the Bayes theorem was discussed in detail. The unit also related probabilities to expected values, giving you the background for a more complex application of probabilities to decision-making under uncertainty.

### 5.0 Summary

There are four basic laws of probability:

1. Addition Law for mutually exclusive events
2. Addition Law for events that are not mutually exclusive
3. Multiplication Law for Independent events
4. Multiplication Law for Dependent events.

The probability of an event, E occurring X-times in n number of trials is given by the formula:

$$\Pr(Enx) = C_n, x p^x q^{(n-x)}$$

A joint probability implies the probability of joint events. Joint probabilities can be conveniently analysed with the aid of joint probability tables.

The Marginal Probability of an event is its simple probability of occurrence, given the sample space.

Assuming two events, A and B, the probability of event A, given that event B has occurred is referred to as the conditional probability of event A.

The Bayes theorem presents a formula which can be thought of as “reversing” conditional probability. That is, it finds a conditional probability, A/B given, among other things, its inverse, B/A.

According to the theorem, given events A and B,

$$\Pr(A/B) = \frac{\Pr(A) \Pr(B/A)}{\Pr(B)}$$



The expected value of a set of values, with associated probabilities, is the arithmetic mean of the set of values. If some variable,  $X$ , has its values specified with associated probabilities,  $P$ , then: Expected value of  $X = E(X) = \sum PX$ .

### 6.0 Tutor-Marked Assignments

1. A firm has tendered for two independent contracts. It estimates that it has probability 0.4 of obtaining contract A and probability 0.1 of obtaining contract B. Find the probability that the firm:

- (a) obtains both contracts
- (b) obtains neither of the contracts
- (c) obtains exactly one contract

2. A financial manager studies an account receivable file to estimate the expected value of payments to be received in the coming month. The manager estimates that there is a probability of:

- 0.8 of receiving N5,200 from company A
- 0.2 of receiving N4,100 from company B
- 0.6 of receiving N8,700 from company C.

Compute the expected value of receipts.

### 7.0 References

1. Francis, A (1998) Business Mathematics and Statistics, 5th edition (Great Britain: Ashford Colour Press).

2. Onwe, O. J. (2007) Statistical Methods for Business and Economic Decisions: A Practical Approach (Lagos: Samalice)

## UNIT 9: COUNTING PRINCIPLES: COMBINATIONS AND PERMUTATIONS

### Content

- 1.0 Introduction
- 2.0 Objectives
- 3.0 Definitions, Computation, and Applications of Combinations and Permutations
- 3.1 Definitions of Combinations and Permutations
- 3.2 Combinations and Permutations Formula
- 3.3 Applications of Permutations and Combinations to the Probability Theory
- 3.4 Self- Assessment Exercise
- 4.0 Conclusion
- 5.0 Summary
- 6.0 Tutor-Marked Assignment
- 7.0 References

### 1.0 Introduction

This unit aims at introducing you to the principles of counting in statistics. You will learn the ways of selecting and arranging groups of items or elements. The process of selecting and arranging groups of items or elements is known as combinations and *permutations*. *This unit will enable you understand the meanings of these as well as their use in solving various types of probability problems associated with business decisions.*

### 2.0 Objectives

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

1. Understand the terms 'Combination' and 'Permutation'
2. Apply these terms to the computation of probabilities of occurrence of events

### 3.0 Definitions, Computation, and Applications of Combinations and Permutations

#### 3.1 Definitions of Combinations And Permutations

The term combination refers to a selection of distinct items, while permutation refers to a combination arranged in a particular form.

As examples, given the item W, X, Y, and Z to choose from:

WX is a combination (or selection) of two items. Other possible combination of two items include: WY, WZ, XY, XZ, YZ.

In addition, WX is a combination of two items containing two separate permutations:

WX and XW. WXY is a combination of three items containing six separate permutations: WXY, WYX, XWY, XYW, YWX, and YXW.

You should note that re-arranging items within a combination does not give a different combination. For instance, re-arranging WY to YW does not give a different combination. However, re-arranging items within a permutation will give a different permutation, for instance, you can create a different permutation by re-arranging WY to

YW, where YW becomes a different permutation. The order of arrangement is importance in permutations.

### Examples.

1. Four firms, A, B, C, and D recently submitted tenders for two jobs, Job 1 and Job 2. Each job must be given to a different firm. You are required to list the different possible ways that the job can be allocated.

### Solution.

The problem here is that of permutation so that we can define AB to imply that Firm A is allocated Job 1 and Firm B is allocated job 2. Here, AB have a different meaning from BA since BA will imply that Firm B gets Job 1 and Firm A gets job 2, which is different from the first allocations. You can then present the list to include:

AB	BA
AC	CA
AD	DA
BC	CB
BD	DB
CD	DC

It follows that the two jobs can be allocated in 12 different ways.

2. Three people are to be selected out of five people, 1, 2, 3, 4, and 5 in an organization. In how many ways can the three people be selected?

### Solution

This is a combination problem since the selection of persons 234 will be same as the selection 324 or 432. The different combinations are listed thus:

123	234
124	235
125	245
134	345
135	
145	

It follows that there are 10 different ways the three people out of five can be selected, according to the above list.

## 3.2 Combination and Permutation Formulas

### 3.2.1 Combination Formula

The number of different combinations of x items taken from n distinct items can be written as  $nC_x$ . This can be calculated using the formula:

$$nC_x = \frac{n!}{x!(n-x)!},$$

where  $n! = n(n-1)(n-2) \dots (2)(1)$ . Similarly for  $r!$

For example, the number of different combinations of 4 items taken from 6 distinct items would be:

$$\begin{aligned} {}^6C_4 &= \frac{6!}{4!(6-4)!} \\ &= \frac{6.5.4.3.2.1}{(4.3.2.1)(2.1)} \\ &= 15 \end{aligned}$$

Therefore, the number of different combinations of 4 items taken from 6 distinct items is 15.

### 3.2.2 Permutation Formula

The number of different arrangements (or permutations) of  $x$  items taken from  $n$  distinct items is defined by:

$${}^nPx = \frac{n!}{(n-x)!}$$

For example, the number of different arrangements (or permutations) of 2 items taken from 4 distinct items can be obtained as:

$$\begin{aligned} {}^4P_2 &= \frac{4!}{(4-2)!} \\ &= \frac{4!}{2!} = \frac{4.3.2.1}{2.1} = 12 \end{aligned}$$

It follows that the required number of different arrangements is 12. Recall our previous example on the allocation of two different jobs to four bidding firms.

### 3.3 Applications of Permutations and Combinations to The Probability Theory

Example number 1 and number 2 that follows are applications of permutations and combinations, respectively, to the probability theory.

#### Examples:

##### Example 1

1(a) How many ways can 3 different jobs be allocated to 5 men, where a man can do only one job?

1(b) What is the probability that man A will be doing job 1?

**Solutions**

1(a): The number of different ways of allocating the 3 different jobs to the 5 men is as follows:

$${}^5P_3 = \frac{5!}{3!} = 60 \text{ different ways}$$

1(b): The probability that man A does job 1 is determined as:

We first determine the number of permutations that satisfy, say AXX, where X refers to any man except man A. These X's can be filled by any 2 out of the remaining 4 men. This can be allocated in  ${}^4P_2 = \frac{4!}{2!} = 12$  different ways.

It follows that  $\Pr(\text{Man A does job 1})$

$$= \frac{\text{number of ways of allocating the remaining 2 jobs to 4 men}}{\text{number of ways of allocating the 3 jobs to the 5 men}}$$

$$= \frac{12}{60} = 0.2 \text{ or } 20 \text{ percent.}$$

The above calculation demonstrates the use of permutations in solving probability problems.

Example 2 below uses the following formula developed for the application of combinations for solving probability problems. The formula is summarized as follows:

The probability of an event, E, occurring x times out of n number of trials, written as  $\Pr(E_n, x)$  is formulated as follows:

$$\Pr(E_n, x) = {}^nC_x p^x q^{(n-x)}$$

Where p represents the probability of success, and q represents the probability of failure.

**Example 2**

An equipment test is repeated on three separate occasions. The probability that the test is successful on each occasion is 35 percent. What is the probability that out of the 3 tests, there are 2 successes.

**Solution**

In this question,  $n = 3$ ;  $x = 2$ ;  $p = 0.35$ ;  $q = 1 - 0.35 = 0.65$ .

$$\begin{aligned} \text{Thus, } \Pr(E_3, 2) &= {}^3C_2 (0.35)^2 (0.65)^1 \\ &= 3(0.1225)(0.65) = 0.231 \end{aligned}$$

Thus, the probability that out of the 3 tests, there are 2 successes is about 23.9 percent.

### 3.4 Self -Assessment Exercise

Briefly discuss how the term 'Combination' differs from the term 'Permutation'.

### 4.0 Conclusion

This unit has expanded your knowledge and applications of combinations and permutations as counting tools.

You were informed that the term combination refers to a selection of distinct items, while that of permutation refers to a combination arranged in a particular form. Formulas for the calculation of combinations and permutations were also presented and applied to computation of probabilities.

### 5.0 Sum mary

The term combination refers to a selection of distinct items, while permutation refers to a combination arranged in a particular form.

The number of different combinations of  $x$  items taken from  $n$  distinct items can be written as  $nC_x$ . This can be calculated using the formula:

$$nC_x = \frac{n!}{x!(n-x)!},$$

The number of different arrangements (or permutations) of  $x$  items taken from  $n$  distinct items is defined by:

$$nP_x = \frac{n!}{(n-x)!}.$$

### 6.0 Tutor-Marked Assignments

A haulage contractor has 2 loads to deliver to separate customers. He has 4 trailers, of which 2 are new. If the trailers are randomly chosen for any particular delivery:

- In how many ways can the trailers be allocated the deliveries?
- In how many ways can the trailers be allocated the deliveries, such that exactly one new trailer is used?
- What is the probability that exactly one new trailer is used?

## **7.0 References**

1. Francis, A (1998) Business Mathematics and Statistics, 5th edition (Great Britain: Ashford Colour Press).
2. Onwe, O. J. (2007) Statistical Methods for Business and Economic Decisions: A *Practical Approach* (Lagos: Samalice)

## **UNIT 10: DISTRIBUTION OF BUSINESS DATA: THE BINOMIAL AND POISSON DISTRIBUTIONS**

### **Content**

- 1.0 Introduction
- 2.0 Objectives
- 3.0 The Binomial and Poisson Distributions
  - 3.1 The Binomial Distribution
  - 3.2 Poisson Distribution
  - 3.3 Poisson Approximation to the Binomial Distribution
  - 3.4 Self- Assessment Exercise
- 4.0 Conclusion
- 5.0 Summary
- 6.0 Tutor-Marked Assignment
- 7.0 References

### **1.0 Introduction**

This unit acquaints you with the meaning and practical applications of binomial and Poisson distributions. The aim is to enable you learn the formulae for calculating probabilities associated with these distributions. The distributions are very important in graphical presentation of behaviours in generated business and economic data. The distributions are also significant in data analysis using statistical methods.

### **2.0 Objectives**

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

1. Understand what Binomial and Poisson distribution are all about
2. Know how to apply Binomial and Poisson distributions in the analysis of business data
3. Know more about the Binomial probability formula and its applications
4. Know the Poisson distribution formula and its applications
5. Establish the relationship between Poisson and Binomial distributions

### **3.0 The Binomial and Poisson Distributions**

This section defines the terms 'Binomial' and 'Poisson' distributions, presents their formulas, their applications, and establishes their relationships. You need to pay attention to the specific examples and applications.

### **3.1 The Binomial Distribution**

A binomial distribution is used to describe the number of successes obtained when a number of identical trials of an experiment are performed. As an illustration, suppose a salesperson makes one call per day and considers the call successful if he or she sells goods worth N10,000. For a five-day working week, the salesperson can make either of the following successful calls: 0, 1, 2, 3, 4, or 5. If he or she tabulates his or her successful calls per week for 50-week year, the following frequency table will be obtained:



**Table 3.1 The Frequency of Successful Phone calls**

Number of Successful Calls (X)	Number of Weeks (f)
0	10
1	25
2	10
3	2
4	2
5	1
<b>Total</b>	<b>50</b>

0	10
1	25
2	10
3	2
4	2
5	1

**Total****50**

Table 3.1 is referred to as binomial distribution as it describes the number of successes in a number of identical trials in the phone call experiment.

**3.1.1 The Binomial Probability Formula**

Businesspeople are often interested in the probability of a number of successes in business decisions. The probability of  $x$  number of successes in  $n$  number of trials can be obtained by the formula:

$$\Pr(E_{n,x}) = {}^nC_x p^x (1-p)^{(n-x)},$$

where  $x = 0, 1, 2, 3, \dots, n$

**Examples**

It has been observed, from a manufacturing industry, that the probability that a machine will need adjustments during a day's production run is 20 percent. If there are 6 of these machines running on a particular day, find the probability that:

- (a) no machine will need adjustments
- (b) one machine will need adjustment
- (c) more than one machine will need adjustment.

**Solutions**

(a): Let  $x = 0$  (no machine will need adjustments);  $n = 6$  machines running on a particular day;  $p =$  Probability of success = probability that a machine will need adjustments during a day's production run = 0.20 (as given). It follows that  $q = 0.80$ .

Using the binomial probability formula:

$$\Pr(E_{6,0}) = {}^6C_0 (0.2)^0 (0.8)^{6-0} = (0.8)^6 = 0.262 = 26.2\%$$

Thus, the probability that no machine will need adjustments in a particular day is 26.2 percent.

(b): Let  $x = 1$  (one machine will need adjustments)

Then:  $\Pr(E_{6,1}) = {}^6C_1 (0.2)^1 (0.8)^{6-1} = 6(0.2)(0.8)^5 = 0.393 = 39.3\%$ . Thus, the probability that one machine will need adjustments is about 39.3 percent.

$$\begin{aligned}
 \text{(c): } \Pr(x > 1) &= \Pr(\text{more than 1 machine will need adjustment}) \\
 &= 1 - \Pr(x = 1 \text{ or } x = 0) \\
 &= 1 - \Pr(0 \text{ or } 1) \\
 &= 1 - [\Pr(0) + \Pr(1)] \\
 &= 1 - [0.262 + 0.393] = 0.345 = 34.5\%
 \end{aligned}$$

The probability that more than one machine will need adjustments is about 34.5 percent.

### 3.1.2 The mean and variance a Binomial Distribution

By definitions, given a binomial distribution with number of trials =  $n$  and probability of success =  $p$  at each trial:

The Mean ( $\mu$ ) =  $np$

The Variance ( $s^2$ ) =  $np(1 - p)$

#### Example

A bottling company takes a sample of 20 items in a crate of identical products. If 10 percent of these items are defective, what will be the mean and variance of the defective items in the sample?

#### Solutions

$n = 20$  and  $p = 0.10$ , that that:

The Mean ( $\mu$ ) =  $np = 20(0.10) = 2$

Thus the mean number of defective items is 2 items.

The variance ( $s^2$ ) =  $np(1 - p) = 20(0.10)(0.90) = 1.8$

Thus the variance is 1.8 number of items.

### 3.2 Poisson Distribution

A Poisson distribution describes the number of events occurring within some given intervals. This can be illustrated by the following example: If the number of calls coming to a telephone switchboard per minute is counted and recorded over a number of successive minutes, we may observe a Poisson distribution that looks like the following:

Number of calls received

Per minute interval (x): 0 1 2 3 4 5

Number of minute

Intervals (f): 11 10 21 4 1 6

An important characteristic of the Poisson distribution is that the events (the calls received in this case) must occur at random. The events must be independent of one another, and 'rare' as well, implying that at any particular point in the interval, the probability of an event occurring must be very low.

Consider another example of a Poisson distribution:

Let an event be an accident on a given highway  
 Let the interval be one month.

We can form a Poisson distribution as follows:

Number of accidents (x):	0	1	2	3	4	5	6 or more
Number of months (f):	0	3	10	13	6	2	1

### 3.2.1 Poisson Probability Formula

Given a Poisson situation, one can calculate the probability of any number of events occurring in the defined interval, if of course the mean number of events per interval is known. Thus, given a Poisson situation with mean number of events per interval,  $m$ , the probability of  $x$  number of events occurring can be defined by:

$$\Pr(x) = \frac{e^{-m} m^x}{x!}$$

where the letter 'e' is a special mathematical constant with an approximate value 2.718.  $x$  can take any one of the values: 0, 1, 2, 3, 4, ...,  $n$ . Exponential tables give the values of  $e^{-m}$  for various values of  $m$ .

### Examples

Suppose that a Poisson –distributed events had a mean of 1.72 within some given intervals, what is the probability that:

- (a) no event will occur in the interval
- (b) 2 events will occur in the interval

### Solutions:

- (a) The probability that no event will occur in the interval:  
 From the formula,  $x = 0$ ;  $m = 1.72$ . Thus,

$$\Pr(0) = \frac{e^{-1.72} 1.72^0}{0!} = e^{-1.72} = 0.1791 \text{ (from the exponential Tables)}$$

thus, the probability that no event will occur in the interval is about 18 percent.

- (b) The probability that two events will occur in the interval:

Here,  $x = 2$ ;  $m = 1.72$ . Thus,

$$\Pr(2) = \frac{e^{-1.72} 1.72^2}{2!} = 0.1791(1.4792) = 0.265$$

It follows that the probability of 2 events occurring in the interval is about 27 percent.

### Other Examples

Assume that customers arrive randomly at a shopping centre at an average rate of 5 percent per minute. Assume also that customer arrivals form a poisson distribution, calculate the probability that:

- (a) no customers arrive at any particular minute
- (b) exactly 1 customer arrives in any particular minute
- (c) 1 or more customers arrive in every 30-second period.

### Solutions

Notice that for questions (a) and (b), the interval for the poisson distribution is 1 minute, with a mean of 5. For question (c), the interval is 30 seconds, so that the mean must be adjusted to 2.5. Thus:

$$(a) \Pr(0) = \frac{e^{-5} 5^0}{0!} = 0.0067$$

Thus, the probability that no customer arrives in a particular minute is about 0.7 percent.

$$(b) \Pr(1) = \frac{e^{-5} 5^1}{1!} = 0.0335$$

It follows that the probability that exactly 1 customer arrives in any particular minute is about 3.4 percent

$$(c) \Pr(1 \text{ or more}) = 1 - \Pr(0) = 1 - e^{-2.5} = 1 - 0.0821 = 0.9179$$

Thus, the probability that 1 or more customers arrive in every 30 seconds is about 92 percent.

### 3.3 Poisson Approximation to the Binomial Distribution

In a given binomial distribution where the number of trials or sample size,  $n$  is large (that is, greater than 30), and the probability,  $p$  is small (less than 0.1 or 10 percent), the Poisson distribution can be used as an approximation to the binomial distribution for calculating probabilities. With these conditions, the mean number of events per interval would be defined as  $np$ . Note that:

- (a) This approximation is used as it is generally easier to calculate poisson probabilities (for large values of  $n$ ) than binomial probabilities;
- (b) The difference between poisson and binomial probabilities is very small under the above conditions. The larger the value of  $n$  and the smaller the value of  $p$ , the better is the approximation.

In the following example, we illustrate the use of this approximation:

Suppose that the items produced from given machine have been discovered to be 1 percent defective. If the items are boxed in lots of 200, what is the probability that a given box has 2 or more defective items?

### Solution

This is a binomial problem since  $n = 200$  and  $p = \Pr(\text{defective}) = 1 \text{ percent} = 0.01$ . For the Poisson approximation, therefore, mean ( $m$ ) =  $np = 200(0.01) = 2.0$ .

$$\begin{aligned}
\text{Thus, Pr(2 or more defectives)} &= 1 - \text{Pr}(0 \text{ or } 1 \text{ defectives}) \\
&= 1 - \{\text{Pr}(0) + \text{Pr}(1)\} \\
&= 1 - \{e^{-2} + e^{-2}(2)\} \text{ (exponential table)} \\
&= 1 - \{0.1353 + (0.1353)(2)\} \\
&= 1 - \{0.1353 + 0.2706\} \\
&= 0.5941
\end{aligned}$$

It follows that a given box has 2 or more defective items is about 59 percent.

If we were to apply the binomial probability formula:

$$\text{Pr}(E_n, x) = {}^nC_x \{p^x (1-p)^{n-x}\},$$

We get:

$$\text{Pr}(0) = \text{Pr}(E_{200}, 0) = {}^{200}C_0 (0.01)^0 (0.99)^{200} = 0.1340$$

$$\text{Pr}(1) = \text{Pr}(E_{200}, 1) = {}^{200}C_1 (0.01)^1 (0.99)^{199} = 0.2707$$

Thus,  $\text{Pr}(2 \text{ or more}) = 1 - (0.1340 + 0.2707) = 0.5953$ .

The above Poisson approximation (0.5941) is therefore very close to the probability (0.5953).

### 3.4 Self -Assessment exercise

Explain how you can apply the Binomial and Poisson distributions in business decisions.

### 4.0 Conclusion

You have learned that Binomial and Poisson distributions have important applications to business decisions especially as these decisions involve probabilities. The concepts learned can enable you in decision making involving uncertainties.

The formulas for the Binomial and Poisson distributions were presented, with relevant examples and applications. You also learned how Binomial and Poisson distributions are interrelated by approximation of one to the other.

### 5.0 Sum mary

You can summarise the discussions as follows:

1. A binomial distribution is used to describe the number of successes obtained when a number of identical trials of an experiment are performed. Businesspeople are often interested in the probability of a number of successes in business decisions. The probability of x number of successes in n number of trials can be obtained by the formula:

$$\text{Pr}(E_n, x) = {}^nC_x p^x (1-p)^{(n-x)},$$

where  $x = 0, 1, 2, 3, \dots, n$

2. A Poisson distribution describes the number of events occurring within some given intervals. Given a Poisson situation, one can calculate the probability of any number of events occurring in the defined interval, if of course the mean number of events per interval is known. Thus, given a Poisson situation with mean number of events per interval,  $m$ , the probability of  $x$  number of events occurring can be defined by:

$$\Pr(x) = \frac{e^{-m} m^x}{x!}$$

where the letter 'e' is a special mathematical constant with an approximate value 2.718.  $x$  can take any one of the values: 0, 1, 2, 3, 4, ...,  $n$ . Exponential tables give the values of  $e^{-m}$  for various values of  $m$ .

3 In a given binomial distribution where the number of trials or sample size,  $n$  is large (that is, greater than 30), and the probability,  $p$  is small (less than 0.1 or 10 percent), the Poisson distribution can be used as an approximation to the binomial distribution for calculating probabilities. With these conditions, the mean number of events per interval would be defined as  $np$ .

## 6.0 Tutor-Marked Assignments

**1. A given Minibus has 7 passenger seats and, on a routine run,**  
it is estimated that any passenger seat will be filled with a  
probability of  $p = 0.42$ :

- (a) What is the mean and variance of the binomial distribution of the passengers on a routine run?
- (b) What is the probability that:
  - (i) there will be no passengers;
  - (ii) there will be just 1 passenger;
  - (iii) there will be at least 3 passengers.

**2. A firm produces half-metre diameter rubber hose. It is**  
estimated that on the average, there are 0.4 flaws per 10  
metre length. Assuming flaws occur randomly, what is the  
probability that:

- (a) there are no flaws in a 10-metre length;
- (b) there is more than 1 flaw in a 10-metre length;
- (c) there is more than 2 flaws in 20-metre length.

## 7.0 References

1. Francis, A (1998) Business Mathematics and Statistics, 5th edition (Great Britain: Ashford Colour Press).
2. Onwe, O. J. (2007) Statistical Methods for Business and Economic Decisions: A Practical Approach (Lagos: Samalice)

## **UNIT 11: THE NORMAL DISTRIBUTION AND CONFIDENCE LIMITS**

### **Content**

- 1.0 Introduction
- 2.0 Objectives
- 3.0 The Normal Distribution, Normal Approximation to Binomial Distribution, and Confidence Limits
- 3.1 The Normal Distribution
- 3.2 Normal Approximation to Binomial Distribution
- 3.3 Confidence Limits
- 3.4 Self- Assessment Exercise
- 4.0 Conclusion
- 5.0 Summary
- 6.0 Tutor-Marked Assignment
- 7.0 References

### **1.0 Introduction**

Normal distribution has been seen as the most important distribution in ~~business~~ **statistics**. It is the name assigned to a type of distribution of continuous data that occurs frequently in practice. It is a distribution of 'natural phenomena' such as: weights of people or objects; heights of people or objects; lengths; and so on. In this unit, we discuss the importance of having a set of business data to be normally distributed. You will learn about confidence limits for sample means as well as confidence limits for sample proportions.

### **2.0 Objectives**

By the time you must have gone through the activities in this unit, you should be able to:

1. Understand the characteristics of a normal distribution
2. Understand the application of normal distributions to analysis involving probabilities of events.
3. Know the normal approximation to Binomial distributions
4. Know how to compute confidence limits for sample means and sample proportions.

### **3.0 The Normal Distribution, Normal Approximation to Binomial Distribution, and Confidence Limits**

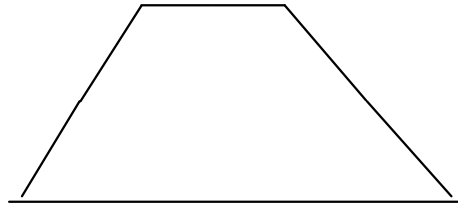
#### **3.1 The Normal Distribution**

The major characteristics of a normal distribution include:

- (a) It has a symmetric (frequency) curve about the mean of the distribution. That is, one-half of the curve is a mirror-image of the other.
- (b) The majority of the values tend to cluster about the mean, with the highest frequency at the mean itself.
- (c) The frequencies of the values taper away (symmetrically) either side of the mean, giving the curve a 'bell shape'.

A typical normal distribution curve is illustrated in figure 3.1 below

**Figure 3.1: A Normal Distribution Curve**



### **3.1.1 Normal Distribution and Probabilities**

There are information needed for evaluating probabilities for ranges of values for a Normal distribution. These include:

- (i) The mean ( $\mu$ ); and,
- (ii) The standard deviation ( $s$ ) of the distribution in question.

The method of calculating probabilities associated with normal distributions requires knowledge of the use of:

- (i) Z-Scores; and,
- (ii) Z-values from the standard normal distribution table

#### ***The Z-Scores and Z-Values***

The Z- score for any value of  $x$  of a normal distribution, having mean,  $\mu$ , and standard deviation,  $s$ , is determined by:

$$Z = \frac{x - \mu}{s}$$

This process is often referred to as “standardising the x-value.

Assuming we know the lengths of steel pins, for example, produced by a particular machine are distributed Normally, with  $\mu = 20$  cms and  $s = 0.1$  cms, and we want to know the probability that a randomly selected pin is less than 20.1 cms in length, that is,  $\Pr(x < 20.1)$ , we will first find the Z-score as:

$$Z = \frac{20.1 - 20}{0.1} = 1$$

It follows that  $\Pr(x < 20.1)$  has been normalised to  $\Pr(Z < 1)$

Solving the problem, we get:

First, and from the normal distribution table, we observe that

$\Pr(Z = 1) = 0.3413$ . This implies that the probability that a Z-score will lie between 0 and 1 is 0.3413. And since the sum of probabilities must equal 1, it follows that  $\Pr(Z < 0) = 0.05$ . Thus,

$\Pr(Z < 1) = \Pr(Z < 0) + \Pr(0 < Z < 1) = 0.5 + 0.3413 = 0.8413$ .



As another illustration, it has been determined that weights of bags of potatoes are normally distributed with mean = 5 kg and standard deviation = 0.2 kg. The potatoes are delivered to the seller, 200 bags at a time:

- (a) What is the probability that a bag selected at random will weigh more than 5.5 kg?
- (b) How many bags, from a single delivery, will be expected to weigh more than 5.5 kg?

### Solutions

(a) Question (a) requires the probability that a bag selected at random will weigh more than 5.5 kg. That is,  $\Pr(x > 5.5)$ , where  $x$  refers to weight of the bag.

$$\text{Standardising } x = 5.5 \text{ gives: } Z = \frac{5.5 - 5}{0.2} = 2.5$$

Using the standard normal table, we observe that  $\Pr(Z = 2.5) = 0.4938$ , and,  $\Pr(Z < 2.5) = 0.5 + 0.4938 = 0.9938 = \Pr(x < 5.5)$ . It follows that  $\Pr(x > 5.5) = 1 - 0.9938 = 0.0062$ . Thus, the probability that a bag selected at random will weigh more than 5.5 kg is about 0.6 percent.

(b) Using the result of question (a), we observe that the proportion of bags expected to weigh more than 5.5 kg is 0.062. Recall that each delivery consists of 200 bags, so that we expect  $0.0062 \times 200 = 1.24$  bags to weigh more than 5.5 kg. This implies that only 1 bag will be expected to weigh more than 5.5 kg.

### 3.2 Normal Approximation to the Binomial Distribution

Recall that it has been noted that when a binomial distribution has a large sample size,  $n$ , it can be difficult to calculate the probabilities associated with it. Recall also that when the probability,  $p$ , is small, the Poisson distribution can be used as an approximation to the binomial distribution.

In the same token, the Normal distribution can be used as an approximation to the binomial distribution when  $n$  is large and when  $p$  is not too small or large. In this situation, we apply the following relationships:

Mean of Normal distribution ( $\mu$ ) =  $np$

Standard deviation of Normal distribution ( $s$ ) =  $\sqrt{np(1 - p)}$

For instance, if a binomial distribution is of size,  $n = 100$  and

$p = 0.35$ , the normal approximation could be applied with:

$$\mu = np = 100(0.35) = 35$$

$$\text{and, } s = \sqrt{np(1 - p)} = \sqrt{100(0.35)(0.65)} = 4.77$$

### Example

Information had it that, from past records, about 40 percent of a firm's orders are for export. Their record for exports is 48 percent in one particular financial quarter. If the firm expects to satisfy about 80 orders in the next financial quarter, what is the probability that they will break their previous export record?

**Solution**

Letting an export order to be a 'success,' we form a binomial distribution with:

Trial = an order

Trial Success = an order for export

Number of trials = 80

Therefore,  $n = 80$  and  $p = \text{Pr}(\text{success}) = \text{Pr}(\text{export order}) = 40\% = 0.40$ . Since  $n$  is large and  $p$  is not too small or large, we can apply the Normal approximation with:

Mean of the Normal distribution ( $\mu$ ) =  $np = 80(0.4) = 32$

And standard deviation ( $s$ ) =  $\sqrt{np(1-p)} = \sqrt{80(0.4)(0.6)} = 4.38$

It follows that the number of orders for exports (say  $x$ ) has an approximate Normal distribution with mean 32 and standard deviation 4.38.

In order that the firm breaks the previous records, it will need more than 48 percent of the 80 orders.

Thus, it will need  $0.48 \times 80 = 38.40$  for export.

Therefore, we are looking for  $\text{Pr}(x > 38.4)$ .

Standardising the value 38.4, we get:

$$Z = \frac{38.4 - 32}{4.38} = 1.46.$$

We need to obtain, therefore,  $\text{Pr}(Z > 1.46)$ . The standard distribution table gives  $\text{Pr}(0 < Z < 1.46) = 0.4279$ . Thus,

$\text{Pr}(Z > 1.46) = 0.5 - 0.4279 = 0.0721$ . By interpretation, there is about 7 percent probability that the previous export record will be broken by the firm in the next financial quarter.

**3.3 Confidence Limits**

Confidence limits specify a range of values within which some unknown population value (mean or proportion) lies, with a given degree of confidence. Confidence limits are calculated based on sample results.

**3.3.1 Confidence Limits (or Interval) for a sample mean**

Given a random sample of a population, a confidence interval for the (unknown) population mean is defined by:

$$\bar{X} \pm Z(s/\sqrt{n})$$

where

$\bar{X}$  = the sample mean

$s$  = sample standard deviation

$n$  = sample size

$Z$  = confidence factor (1.64 for 90%; 1.96 for 95%; 2.58 for 99%, etc from the Z-table)

$s/\sqrt{n}$  = standard error of the sample mean

As an illustration, suppose a sample of 100 transactions yielded a mean gross value of N2,000 and a standard deviation of N20. We want to compute a 95 percent confidence interval as follows:

$$\begin{aligned} 95\% \text{ confidence interval} &= \bar{X} \pm Z(s/\sqrt{n}) \\ &= 2000 \pm (1.96)(20/\sqrt{100}) \\ &= 2000 \pm (1.96)(2) \\ &= 2000 \pm 3.92 \end{aligned}$$

Thus, the 95% confidence interval for the mean of the population is:  $1996.08 < \mu < 2003.92$ .

This implies that there is a 95 percent probability that the population mean,  $\mu$ , is between 1996.08 and 2003.92.

### 3.3.2 Confidence Limits for a Proportion

Given a random sample of a population, a confidence interval for the population proportion,  $P$ , can be obtained as:

$$\bar{p} \pm Z\sqrt{p(1-p)/n},$$

where  $p$  = sample proportion

$n$  = sample size

$Z$  = the confidence factor

$\sqrt{p(1-p)/n}$  = standard error of the proportion

#### Example

The same fuel was tested on 21 similar cars under identical conditions. Fuel consumption was found to have a mean of 41.6 kpg with a standard deviation of 3.2 kpg (kilometers per gallon). Only 14 of the cars were found to completely satisfy current exhaust emission guidelines. Calculate a 99 percent confidence interval for the percentage of similar cars that completely satisfy current exhaust emission guidelines.

#### Solution

Sample proportion,  $p = 14/21 = 0.667$

Value of  $Z$  at 99% confidence level = 2.58

$$\begin{aligned} \text{Thus, } \bar{p} \pm Z\sqrt{p(1-p)/n} &= 0.667 \pm 2.58\sqrt{(0.667)(0.333)/21} \\ &= 0.667 \pm (2.58)(0.103) \\ &= 0.667 \pm 0.265 \end{aligned}$$

Therefore, the 99 percent confidence interval as required can be written as:  
 $0.402 < P < 0.932$ .

By interpretation, this implies that at 99 percent level of confidence, the percentage of similar cars that completely satisfy current exhaust emission guidelines is between 40.2 and 93.2.

### 3.4 Self -Assessment Exercise

Discuss briefly the distinguishing features of a normal distribution

### 4.0 Conclusion

This unit has exposed you to the meaning and characteristics of normal distribution. It has also shown you how you can approximate normal distribution to Binomial distribution. You learned how to get the confidence limits for both sample mean and sample proportions.

### 5.0 Sum mary

The major characteristics of a normal distribution include:

- (a) It has a symmetric (frequency) curve about the mean of the distribution. That is, one-half of the curve is a mirror-image of the other.
- (b) The majority of the values tend to cluster about the mean, with the highest frequency at the mean itself.
- (c) The frequencies of the values taper away (symmetrically) either side of the mean, giving the curve a 'bell shape'.

The information needed for evaluating probabilities for ranges of values for a Normal distribution include: (i) the mean ( $\mu$ ); and, (ii) the standard deviation ( $s$ ) of the distribution in question.

The Z- score for any value of  $x$  of a normal distribution, having mean,  $\mu$ , and standard deviation,  $s$ , is determined by:

$$Z = \frac{x - \mu}{s}$$

This process is often referred to as “standardising the x-value.

Confidence limits specify a range of values within which some unknown population value (mean or proportion) lies, with a given degree of confidence. Confidence limits are calculated based on sample results.

### 6.0 Tutor-Marked Assignm ents

1. The specification for the length of an engine is a minimum of 99mm and a maximum of 104.4mm. A batch of parts is produced, with normal distribution, a mean of 102mm, and a standard deviation of 2mm. Parts cost N100 to make. Those that are too short have to be scrapped; those too long are shortened, at a further cost of N50. You are required:
  - (a) to find the percentage of parts which are (i) undersize, (ii) oversize;
  - (b) to find the expected cost of producing 1,000 usable parts;
  - (c) to calculate and to explain the implications of changing the production method so that the mean is halfway between the upper and lower specification limits (the standard deviation remains the same).

2(a) A company produces batteries whose lifetimes are normally distributed with a mean of 100 hours. It has been observed that 90 percent of the batteries last at least 40 hours:

- (i) estimate the standard deviation of the lifetimes
- (ii) what percentage of the batteries will not last 70 hours?

(b) A company produces electronic calculators. From past experience, the company knows that 90 percent of the calculators will be in good condition and 10 percent will be faulty if the production process is working satisfactorily. An inspector randomly selects 5 calculators from the production line every hour and carries out a rigorous check:

- (i) what is the probability that a random sample of 5 will contain at least 3 defective calculators?
- (ii) a sample of 5 calculators is found to contain 3 defectives, do you consider the production process to be working satisfactorily?

## **7.0 References**

1. Hanke, J. E. and Reitsch, A. G. (1991) Understanding Business Statistics (Homewood, IL: Richard Irwin)

2. Onwe, O. J. (2007) Statistical Methods for Business and Economic Decisions: A *Practical Approach* (Lagos: Samalice)

## **UNIT 12: INVENTORY CONTROL**

### **Content**

- 1.0 Introduction
- 2.0 Objectives
- 3.0 Inventory Costs, Inventory Control Systems, and Inventory Models
  - 3.1 Inventory Costs
  - 3.2 Important Terminologies in the Inventory Control Theories
  - 3.3 The Inventory Graph
  - 3.4 Inventory Control Systems
  - 3.5 Inventory Control Models
  - 3.6 Self- Assessment Exercise
- 4.0 Conclusion
- 5.0 Summary
- 6.0 Tutor-Marked Assignment
- 7.0 References

### **1.0 Introduction**

The aim of an inventory control system is to minimise costs and establish:

- (a) the optimum amount of stock to be ordered
- (b) the period between orders

A firm's inventory is defined as the total stocks of various kinds including: basic raw materials; partly finished goods and materials; sub-assemblies; office and workshop supplies; and finished goods.

An inventory must be carried by a firm for various reasons:

- 1. Anticipating normal demand
- 2. Taking advantage of bulk-purchase discounts
- 3. Meeting emergency shortages
- 4. As natural part of the production process
- 5. Absorbing wastages and unpredictable fluctuations

### **2.0 Objectives**

Having gone through this unit, you will be able to:

- 1. Understand the necessary processes for effective inventory control
- 2. Be familiar with inventory costs, inventory control systems, and inventory models
- 3. Manage inventory effectively, using statistical techniques.

### **3.0 Inventory Costs, Inventory Control Systems, and Inventory Models**

In this section, you will be taken through the concepts of inventory and costs associated with it; the inventory control systems; and, the inventory models. Working through the discussions will help you very much in your inventory management activities.

### **3.1 Inventory Costs**

Like any other business decision variables there are costs associated with inventories. These include:

#### **3.1.1 Ordering ( Replacement) Costs**

These are such costs as transportation costs, clerical and administrative costs associated with the physical movement of the purchased external goods. Where the goods are manufactured internally, there are alternative initial costs to be borne with each production run referred to as set-up costs

#### **3.1.2 Holding (Carrying) Costs**

These are:

- (a) Storage costs in terms of staffing, equipment maintenance, and handling;
- (b) Storage overheads (heat, light, rent, and the like);
- (c) Cost of capital tied up in inventory;
- (d) Insurance, security and pilferage;
- (e) Deterioration or breakages.

#### **3.1.3 Stock out Costs**

These are costs associated with running out of stock. These include penalty payments, loss of goodwill, idle manpower and machine, and the like.

### **3.2 Important Terminologies in the Inventory Control Theories**

The followings are some important terminologies used in inventory control theories:

**3.2.1 Lead Time.** This is the time between ordering of goods and their replenishment. Orders may be internal (requiring a production run) or external.

**3.2.2 Economic Ordering Quantity (EOQ).** This refers to the external order quantity that minimises total inventory costs.

**3.2.3 Economic Batch Quantity (EBQ).** This refers to the size of the internal production run that minimises total inventory costs.

**3.2.4 Safety Stock.** This is a term used to describe the stock held to cover possible deviations in demand or supply during the lead time. It is sometimes referred to as buffer or minimum stock.

**3.2.5 Maximum Stock.** This is a level used as an indicator above which stocks are deemed to be too high.

**3.2.6 Reorder Level.** This is the level of stock, which when reached, signals replenishment order.

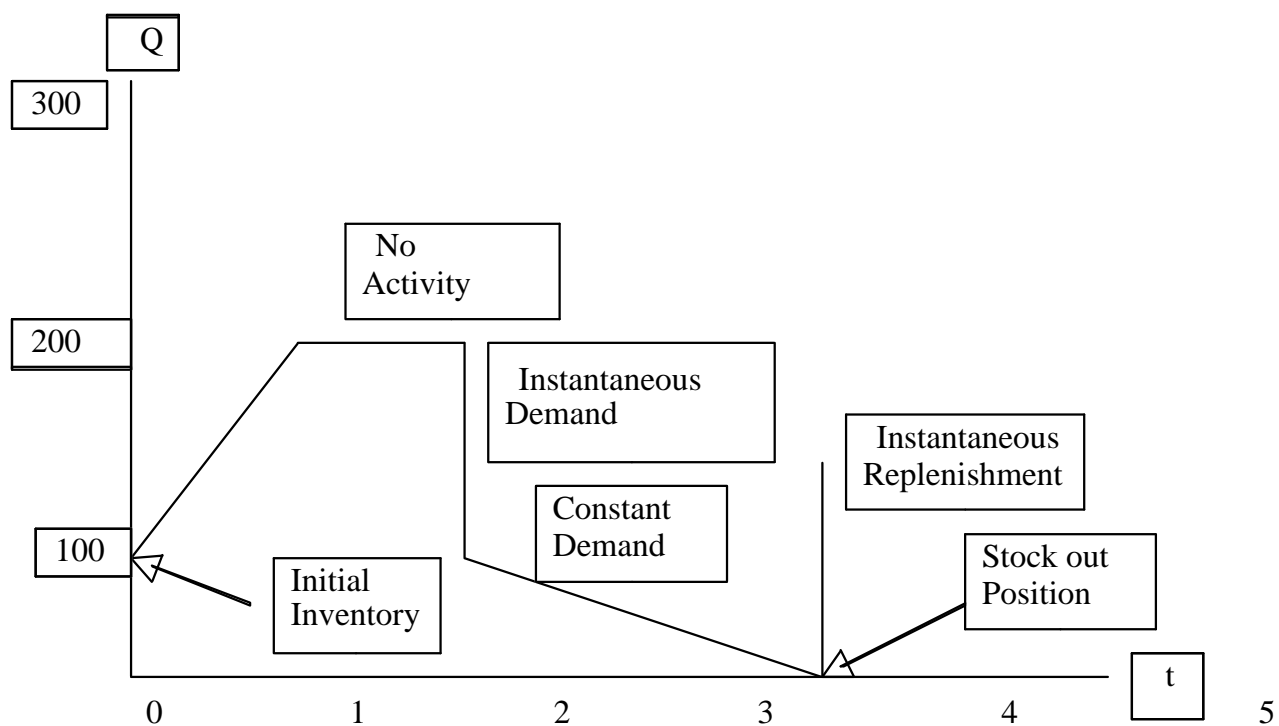
**3.2.7 Reorder Quantity.** This is the level of replenishment order.

### 3.3 The Inventory Graph

The purpose of inventory graph is to present the inventory control problems in graphical terms. It plots the relationship between quantity of stock held ( $Q$ ) and time ( $t$ ).

Figure 3.1 presents a general inventory graph with various features. It shows an initial inventory of 100 items, replenished by a further 100 items continuously over a given time period. Observe as indicated that for the next time period, there was no activity, but at time period 2, 100 items were demanded, followed, over the next two periods, by a continuous demand which used up the last 100 items. This stock out position led to the delivery of additional 150 items.

**Figure 3.1: The General Inventory Graph**



### 3.4 Inventory Control Systems

There are two basic inventory control systems:

#### 3.4.1 Re-order Level System

This is the most commonly used control system. It generally results in lower stocks. The system also enables items to be ordered in more economic quantities and is more responsive to fluctuations in demand than the second system discussed below.

The system sets the value of three important levels of stock as warning or action triggers for management:



(i) Re-order Level. This is an action level of stock which leads to the replenishment order, normally the Economic Order Quantity (EOQ). For a particular time period, the re-order level is computed as follows:

$$L_{ro} = \text{maximum usage per period} \times \text{maximum lead time (in periods)}$$

(ii) Minimum Level. This is a warning level set such that only in extreme cases (above average demand or late replenishment) should it be breached. It is computed as follows:

$$L_{min} = \text{Re-Order Level} - (\text{normal Usage} \times \text{Average lead time})$$

(iii) Maximum Level. This is another warning level set such that only in extreme cases (low levels of demand) should it be breached. It is computed by:

$$L_{max} = \text{Re-order Level} + \text{EOQ} - (\text{Minimum usage} \times \text{Minimum lead time})$$

### Examples

Suppose for a particular inventory, there exists:

- (a) the weekly minimum, normal and maximum usage of 600, 1000, and 1400 respectively;
- (b) the lead time which vary between 4 and 8 weeks (average = 6 weeks); and,
- (c) the normal ordering quantity (EOQ) of 20,000.

It follows that:

$$\text{The Re-order Level (} L_{ro} \text{)} = 1400 \times 8 = 11,200 \text{ units}$$

$$\text{Minimum Stock Level (} L_{min} \text{)} = 11,200 - 1000 \times 6 = 5,200 \text{ units}$$

$$\begin{aligned} \text{Maximum Stock Level (} L_{max} \text{)} &= 11,200 + 20,000 - 600 \times 4 \\ &= 28,800 \end{aligned}$$

### 3.4.2 The Periodic Review System

The periodic review system sets a review period, at the end of which the stock level of the given item is brought up to a predetermined value. This system has the following advantages:

- (i) It enables stock positions to be reviewed periodically so that the chances of obsolete stock items are minimized.
- (ii) Economies of scale is possible when many items are ordered at the same time or in the same sequence.

### 3.5 Inventory Control Models

Two basic inventory control models are currently in use. These include:

1. The Basic Model
2. The Adapted Basic Model (with gradual replenishment)

The models are as discussed below.

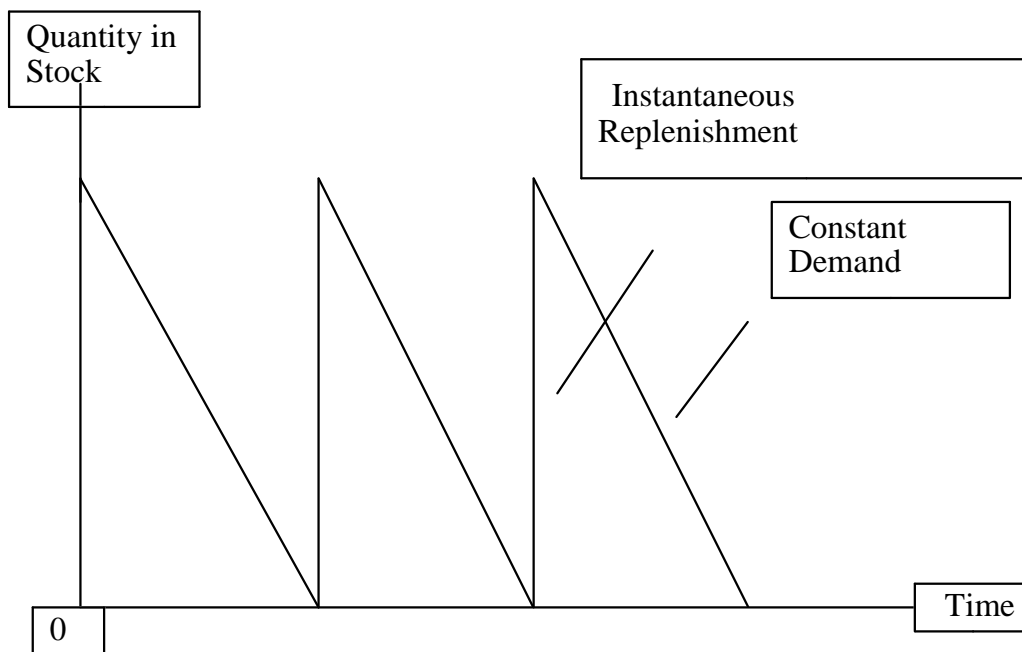
### 3.5.1 The Basic Model

The basic inventory control model is based on the following assumptions:

- (i) The rate of demand (that is, the number of items demanded per year),  $D$ , is constant and continuous over a given period, and no excess demands.
- (ii) The ordering cost ( $C_o = N/\text{circle}$ ) is constant and independent of the quantity ordered.
- (iii) Only one type of stock item is considered and its price ( $P = N/\text{item}$ ) is constant.
- (iv) The holding cost ( $C_h = N/\text{item}$ ) is the cost of carrying one article in stock for one year.
- (v) The quantity ordered per circle ( $q$ ) is supplied to store instantaneously whenever the inventory level becomes zero.

Figure 3.2 illustrates the standard inventory graph for the basic model:

**Figure 3.2: Inventory Graph for the Basic Inventory Control Model**



For the basic model, the total annual inventory cost is minimised when the Economic Ordering Quantity (EOQ) takes the following value:

$$EOQ = \sqrt{2DC_o/C_h}$$

Where  $D$  = annual demand

$C_o$  = Order Cost per circle

$C_h$  = Holding Cost per item

The total annual inventory cost,  $C$ , is defined by:

$$C = \text{Total Ordering Cost} + \text{Total Holding Cost}$$

$$= \text{number of orders per year} \times \text{order cost} + \text{average inventory level} \times \text{holding cost per item}$$

$$C = D/q \times C_o + q/2 \times C_h$$

Notice from the above equation that as  $q$  gets larger:  
Annual ordering cost becomes smaller; and annual holding cost becomes larger.

The basic model also involves the calculation of the following statistics:

(a) Number of orders per year  $= \frac{\text{Yearly Demand}}{\text{EOQ}}$

**EOQ**

(b) Length of Cycle (days)  $= \frac{\text{Number of Days per year}}{\text{Number of Orders per year}}$

(c) Average Inventory Level  $= \frac{\text{EOQ}}{2}$

## 2 Examples

A commodity has a steady rate of demand of 2,000 units per year. Placing an order costs N200 and it cost N50 to hold a unit for a year:

- Estimate the Economic Order Quantity (EOQ)
- Find the number of orders placed per year
- What is the length of the inventory circle?

## Solutions

**Note that from the given information,  $D = 2,000$ ;  $C_o = 200$ ; and,  $C_h = 50$**

- (a) The Economic Order Quantity is determined by:

$$\text{EOQ} = \sqrt{2DC_o/C_h} = \sqrt{2(2000)(200)/50}$$

$$= \sqrt{16000} = 126.491$$

Thus, the economic order quantity is about 127 units.

(b) Number of orders per year  $= \frac{\text{Yearly Demand}}{\text{EOQ}}$

$$= \frac{2000}{126.491} = 15.81$$

Thus, the number of orders per year is approximately 16 orders.

(c) The Cycle length  $= \frac{\text{Number of Days per Year}}{\text{Number of Orders per Year}}$

$$= \frac{365}{15.81} = 23.1$$

The cycle length is about 23 days.

### 3.5.2 The Adapted Model

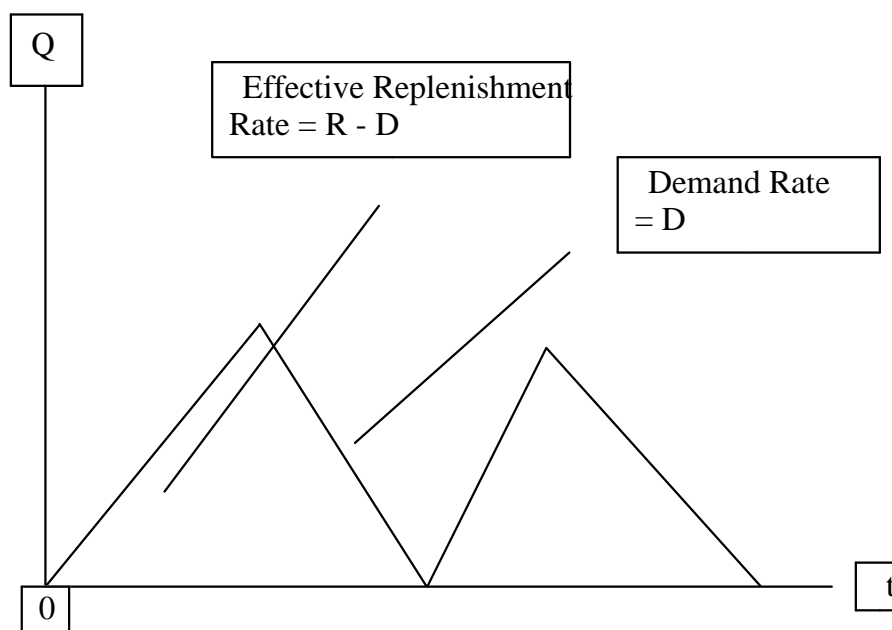
The adapted Model assumes gradual replenishment unlike the basic model. This assumption is on the premises that when stocks are received from the production line, it is very likely that finished items are received continuously over a period of time. Stock is therefore, subject to gradual replenishment. Like the basic model, this model also assumes constant demand rate,  $D$ , and the two cost factors: the order cost per cycle,  $C_o$ , and the holding cost per item,  $Ch$ .

The adapted model is often referred to as the production run model whereby:

- (i) A production run is started every time the inventory level decreases to zero, and stops when  $q$  items have been produced or supplied. The run lasts for time  $t$  and is known as the run time.
- (ii) The quantity ordered per cycle is referred to as the run size, and items are supplied at the rate of  $R$  per annum.
- (iii) The effective replenishment rate is defined as  $R - D$  items.

In figure 3.3, we present the inventory control graph for the adapted model.

**Figure 3.3: The Inventory Control Graph for the Adapted Model**



The value of the run size that minimises inventory cost for the adapted model is defined by:

$$\text{Economic Batch Quantity (EBQ)} = \sqrt{\frac{2DC_o}{Ch\{1 - D/R\}}}$$

where  $D$  = Annual Demand

$R = \text{Annual Production rate}$   
 $C_o$  and  $C_h$  are as defined earlier.

The adapted model uses the following optimal statistics:

(a)  $\text{Number of runs per year} = \frac{\text{Yearly Demand}}{\text{EBQ}}$

$\text{Length of cycle} = \frac{\text{Number of days per year}}{\text{Number of runs per year}}$

(c)  $\text{Run time (days)} = \frac{\text{EBQ} \times \text{Number of days per year}}{\text{Annual production rate}}$

(d)  $\text{Peak Inventory Level} = \text{Effective Replenishment rate} \times \text{run time}$

(e)  $\text{Average inventory level} = \frac{1}{2} \times \text{Peak inventory level}$

The following example illustrates the use of the adapted inventory control model:

A manufacturing activity requires a continuous supply of 3000 items per year from store, replenished by production runs, each of which operates at the constant rate of 5000 items per year. Each production run has a set-up cost of N300, and the holding cost per item per annum is N25. Compute the Economic Batch Quantity (EBQ), and find the number of runs per year, the length of cycle, the run time, the peak inventory level, and the average inventory level.

### Solutions

From the given information:  $D = 3000$ ;  $R = 5000$ ;  $C_o = 300$ ; and,  
 $C_h = \text{N}25$ .

Thus,

$$\begin{aligned} \text{EBQ} &= \sqrt{2DC_o / \{C_h(1 - D/R)\}} \\ &= \sqrt{2 \times 3000 \times 300 / \{25 \times (1 - 3000/5000)\}} \\ &= \sqrt{1,200,000} = 1095.445 \end{aligned}$$

It follows that the Economic Batch Quantity (EBQ) is approximately 1096 units

The number of runs per year =  $3000/1096 = 2.74$

The length of cycle =  $365/2.74 = 133.21$  days

The run time =  $\frac{1096 \times 365}{5000} = 80.01$  days.

$$\begin{aligned}
\text{Peak Inventory Level} &= \text{Effective Replenishment Rate (R - D)} \\
&\quad \times \text{Run time} \\
&= (5000 - 3000) \times 80.1 \\
&= 160,200 \text{ units}
\end{aligned}$$

$$\text{Average inventory level} = \frac{1}{2} \times 160,200 = 80,100 \text{ units.}$$

### 3.6 Self -Assessment Exercise

With the aid of a diagram, discuss the basic inventory control model.

### 4.0 Conclusion

This unit has enriched you with the essential issues in inventory controls including, the inventory costs, the inventory control systems, and the inventory control models. The basic point is that inventory costs, inventory control systems, and inventory control models are the necessary tools for effective management of company inventories.

### 5.0 Sum mary

The major issues of discuss in this unit can be summarised as follows:

1. There are costs associated with inventories including: (i) ordering costs; (ii) holding costs; and, (iii) stock out costs.
2. The important terminologies used in inventor y control theories are: (i) the Lead Time; (ii) Economic Order Quantity (EOQ); (iii) Economic Batch Quantity (EBQ); (iv) Safety Stock; (v) Maximum Stock; (vi) Re-order Level; and, Re-order Quantity.
3. There exists two inventory control systems, including (i) the Re-Order Level System; and, (ii) the Periodic Review System.
4. Two inventor y control models are in place: (i) The Basic Model; and (ii) the Adapted Basic Model.

### 6.0 Tutor-Marked Assignm ent

1. A manufacturer has been given a contract to supply 20,000 items per annum. The ordering cost is estimated at N120. The price per unit is N65, and the holding costs are 15 percent of the unit price. Using the basic inventory model:
  - (a) Calculate the Economic Order Quantity (EOQ)
  - (b) Compute the length of the inventory cycle and the number of runs per year.
2. An inventory system follows the adapted basic inventory control model. The demand rate is constant at 16,000 items per year, the unit price of the items is N250, the holding cost is 10 percent of the unit price per year, and the set-up cost is N110 per run. If the production rate is 10,000 items per year, calculate :
  - (a) the optimum run size (EBQ)
  - (b) the run time
  - (c) the cycle length
  - (d) the average inventory level

## **7.0 References**

1. Francis, A (1998) Business Mathematics and Statistics, 5th edition (Great Britain: Ashford Colour Press).
2. Onwe, O. J. (2007) Statistical Methods for Business and Economic Decisions: A *Practical Approach* (Lagos: Samalice)

## **UNIT 13: DECISION ANALYSIS**

### **Content**

- 1.0 Introduction
- 2.0 Objectives
- 3.0 Decision Making Under Uncertain Conditions
  - 3.1 Certainty and Uncertainty in Decision Analysis
  - 3.2 Analysis of the Decision Problem
  - 3.3 Expected Monetary Value Decisions
  - 3.4 Decision Making Involving Sample Information
  - 3.5 Self- Assessment Exercise
- 4.0 Conclusion
- 5.0 Summary
- 6.0 Tutor-Marked Assignment
- 7.0 References

### **1.0 Introduction**

Decision analysis is the modern approach to decision making both in economics and in business. It can be defined as the logical and quantitative analysis of all the factors influencing a decision. The analysis forces decision makers to assume some active roles in the decision-making process. By so doing, they rely more on rules that are consistent with their logic and personal behaviour than on the mechanical use of a set of formulas and tabulated probabilities.

The primary aim of decision analysis is to increase the likelihood of good outcomes by making good and effective decisions. A good decision must be consistent with the information and preferences of the decision maker. It follows that decision analysis provides decision-making framework based on available information on the business environment, be it a sample information, judgmental information, or a combination of both.

### **2.0 Objectives**

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

1. Understand what uncertainty in decision making is all about
2. Know how decision analysis under certainty differs from decision analysis under uncertainty.
3. Make business decisions using sample information.

### **3.0 Decision Making Under Uncertain Conditions**

Our discussions in this section focus on decision analysis under both conditions of certainty and conditions of uncertainty. The aim is to develop the necessary steps to be taken when making decisions under uncertainty.

### **3.1 Certainty and Uncertainty in Decision Analysis**

Most decision-making situations involve the choice of one among several alternatives actions. The alternative actions and their corresponding payoffs are usually known to the



decision-maker in advance. A prospective investor choosing one investment from several alternative investment opportunities, a store owner determining how many of a certain type of commodity to stock, and a company executive making capital-budgeting decisions are some examples of a business decision maker selecting from a multitude of a multitude of alternatives. The decision maker however, does not know which alternative will be best in each case, unless he/she also knows with certainty the values of the economic variables that affect profit. These economic variables are referred to, in decision analysis, as states of nature as they represent different events that may occur, over which the decision maker has no control.

The states of nature in decision problems are generally denoted by  $s_i$  ( $i = 1, 2, 3, \dots, k$ ), where  $k$  is the number of or different states of nature in a given business and economic environment. It is assumed here that the states of nature are mutually exclusive, so that no two states can be in effect at the same time, and collectively exhaustive, so that all possible states are included within the decision analysis.

The alternatives available to the decision maker are denoted by  $a_i$  ( $i = 1, 2, 3, \dots, n$ ), where  $n$  is the number of available alternatives. It is also generally assumed that the alternatives constitute a mutually exclusive, collectively exhaustive set.

When the state of nature,  $s_i$ , whether known or unknown, has no influence on the outcomes of given alternatives, we say that the decision maker is operating under *certainty*. *Otherwise, he/she is operating under uncertainty.*

Decision making under certainty appears to be simpler than that under uncertainty. Under certainty, the decision maker simply appraises the outcome of each alternative and selects the one that best meets his/her objective. If the number of alternatives is very high however, even in the absence of uncertainty, the best alternative may be difficult to identify. Consider, for example, the problem of a delivery agent who must make 100 deliveries to different residences scattered over Lagos metropolis. There may literally be thousands of different alternative routes the agent could choose. However, if the agent had only 3 stops to make, he/she could easily find the least-cost route.

Decision making under uncertainty is always complicated. It is the probability theory and mathematical expectations that offer tools for establishing logical procedures for selecting the best decision alternatives. Though statistics provides the structure for reaching the decision, the decision maker has to inject his/her intuition and knowledge of the problem into the decision-making framework to arrive at the decision that is both theoretically justifiable and intuitively appealing. A good theoretical framework and commonsense approach are both essential ingredients for decision making under uncertainty.

To understand these concepts, consider an investor wishing to invest N100,000 in one of three possible investment alternatives, A, B, and C. Investment A is a Savings Plan with returns of 6 percent annual interest. Investment B is a government bond with 4.5 percent annual interest. Investments A and B involve no risks. Investment C consists of shares of mutual fund with a wide diversity of available holdings from the securities market.

The annual return from an investment in C depends on the uncertain behaviour of the mutual fund under varying economic conditions.

The investors available actions ( $a_i$ ;  $i = 1, 2, 3, 4$ ) are as follows

- a1: Do not invest
- a2: Select investment A the 6% bank savings plan.
- a3: Select investment B, the 4.5 % government bond.
- a4: Select investment C, the uncertain mutual fund

Observe that actions a1 to a3 do not involve uncertainty as the outcomes associated with them do not depend on uncertain market conditions. Observe also that action a2 dominates actions a1 and a3. In addition, action a1 is clearly inferior to the risk-free positive growth investment alternatives a2 and a3 as it provides for no growth of the principal amount.

Action a4 is associated with an uncertain outcome that, depending on the state of the economy, may produce either a negative return or a positive return. Thus there exists no apparent dominance relationship between action a4 and action a2, the best among the actions involving no uncertainty.

Suppose the investor believes that if the market is down in the next year, an investment in the mutual fund would lose 10 percent returns; if the market stays the same, the investment would stay the same; and if the market is up, the investment would gain 20 percent returns. The investor has thus defined the states of nature for his/her investment decision-making problem as follows:

- s1: The market is down.
- s2: The market remains unchanged.
- s3: The market is up.

A study of the market combined with economic expectations for the coming year may lead the investor to attach subjective probabilities of 0.25, 0.25, and 0.50, respectively, to the states of nature, s1, s2, and s3. The major question is then, how can the investor use the foregoing information regarding investments A, B, and C, and the expected market behaviour serves as an aid in selecting the investment that best satisfies his/her objectives? This question will be considered in the sections that follow.

### **3.2 Analysis of the Decision Problem**

In problems involving choices from many alternatives, one must identify all the actions that may be taken and all the states of nature whose occurrence may influence decisions. The action to take none of the listed alternatives whose outcome is known with certainty may also be included in the list of actions. Associated with each action is a list of payoffs. If an action does not involve risk, the payoff will be the same no matter which state of nature occurs.

The payoffs associated with each possible outcome in a decision problem should be listed in a payoff table, defined as a listing, in tabular form, of the value payoffs associated with all possible actions under every state of nature in a decision problem.

The payoff table is usually displayed in grid form, with the states of nature indicated in the columns and the actions in the rows. If the actions are labeled  $a_1, a_2, \dots, a_n$ , and the states of nature labeled  $s_1, s_2, \dots, s_k$ , a payoff table for a decision problem appears as in table 3.1 below. Note that a payoff is entered in each of the  $nk$  cells of the payoff table, one for the payoff associated with each action under every possible state of nature.

**Table 3.1: The Payoff Table**

		STATE OF NATURE			
ACTION	$s_1 \ s_2 \ s_3 \ \dots \ s_k$				
$a_1$					
$a_2$					
$a_3$					
.					
.					
.					
$a_n$					

### Example

The managing director of a large manufacturing company is considering three potential locations as sites at which to build a subsidiary plant. To decide which location to select for the subsidiary plant, the managing director will determine the degree to which each location satisfies the company's objectives of minimising transportation costs, minimising the effect of local taxation, and having access to an ample pool of available semi-skilled workers. Construct a payoff table and payoff measures that effectively rank each potential location according to the degree to which each satisfies the company's objectives.

### Solution

Let the three potential locations be sites A, B, and C. To determine a payoff measure to associate with each of the company's objectives under each alternative, the managing director subjectively assigns a rating on a 0 – to – 10 scale to measure the degree to which each location satisfies the company's objectives. For each objective, a 0 rating indicates complete dissatisfaction, while a 10 rating indicates complete satisfaction. The results are presented in table 3.2 below:

**Table 3.2: Ratings for three alternative plant sites for a Manufacturing Company**

COMPANY OBJECTIVE	ALTERNATIVE		
	Site A	Site B	Site C
Transportation Costs	6	4	10
Taxation Costs	6	9	5
Workforce Pool	7	6	4

To combine the components of payoff, the managing director asks himself, what are the relative measures of importance of the three company objectives I have considered as components of payoff? Suppose the managing director decides that minimising transportation costs is most important and twice as important as either the minimization of local taxation or the size of workforce available. He/she thus assigns a weight of 2 to the transportation costs and weights of 1 each to taxation costs and workforce. This will give rise to the following payoff measures:

$$\text{Payoff (Site A)} = 6(2) + 6(1) + 7(1) = 25$$

$$\text{Payoff (Site B)} = 4(2) + 9(1) + 6(1) = 23$$

$$\text{Payoff (Site C)} = 10(2) + 5(1) + 4(1) = 29$$

### 3.3 Expected Monetary Value Decisions

A decision-making procedure, which employs both the payoff table and prior probabilities associated with the states of nature to arrive at a decision is referred to as the *Expected Monetary Value decision procedure*. Note that by *prior probability* we mean probabilities representing the chances of occurrence of the identifiable states of nature in a decision problem prior to gathering any sample information. The expected monetary value decision refers to the selection of available action based on either the expected opportunity loss or the expected profit of the action.

Decision makers are generally interested in the optimal monetary value decisions. The optimal expected monetary value decision involves the selection of the action associated with the minimum expected opportunity loss or the action associated with the maximum expected profit, depending on the objective of the decision maker.

The concept of expected monetary value applies mathematical expectation, where opportunity loss or profit is the random variable and the prior probabilities represent the probability distribution associated with the random variable.

The expected opportunity loss is computed by:

$$E(L_i) = \sum_j L_{ij}P(s_j), \quad (i = 1, 2, \dots, n)$$

where  $L_{ij}$  is the opportunity loss for selecting action  $a_i$  given that the state of nature,  $s_j$ , occurs and  $P(s_j)$  is the prior probability assigned to the state of nature,  $s_j$ .

The expected profits for each action is computed in a similar way:

$$E(p_i) = \sum_j p_{ij}P(s_j)$$

where  $p_{ij}$  represents profits for selecting action  $a_i$

**Example**

By recording the daily demand for a perishable commodity over a period of time, a retailer was able to construct the following probability distribution for the daily demand levels:

**Table 3.3: Probability Distribution for the Daily Demand**

$s_j$	$P(s_j)$
1	0.5
2	0.3
3	0.2
4 or more	0.0

The opportunity loss table for this demand-inventory situation is as follows:

**Table 3.4: The Opportunity Loss Table**

Action, Inventory		State of Nature, Demand		
		$s_1(1)$	$s_2(2)$	$s_3(3)$
	$a_1(1)$			3
6	$a_2(2)$		2	0
3	$a_3(3)$		4	2
0				

We are required to find the inventory level that minimises the expected opportunity loss.

**Solution**

Given the prior probabilities in the first table, the expected opportunity loss are computed as follows:

$$E(L_i) = \sum_{j=1}^3 L_{ij}P(s_j), \text{ for each inventory level, } i = 1, 2, 3.$$

The expected opportunity losses at each inventory level become:

$$E(L_1) = 0(0.5) + 3(0.3) + 6(0.2) = \text{N}2.10$$

$$E(L_2) = 2(0.5) + 0(0.3) + 3(0.2) = \text{N}1.60$$

$$E(L_3) = 4(0.5) + 2(0.3) + 0(0.2) = \text{N}2.60$$

It follows that in order to minimize the expected opportunity loss, the retailer should stock 2 units of the perishable commodity. This is the optimal decision.

**3.4 Decision Making Involving Sample Information**

In discussing prior probabilities, recall it was noted that prior probabilities are acquired either by subjective selection or by computation from historical data. No current information describing the probability of occurrence of the states of nature was assumed to be available.

In many cases, observational information or other evidence are available to the decision maker either for purchase or at the cost of experimentation. For example, a retailer whose business depends on the weather may consult a meteorologist before making decisions, or an investor may hire a market consultant before investing. Market surveys carried out before the release of a new product represent another area in which the decision maker may seek additional information. In each of these examples, the decision maker attempts to acquire information relative to the occurrence of the states of nature from a source other than that from which the prior probabilities were computed.

When such information are available, Baye's Law can be employed to revise the prior probabilities to reflect the new information. These revised probabilities are referred to as posterior probabilities.

By definition, the posterior probability represented symbolically by  $P(s_k/x)$  is the probability of occurrence of the state of nature  $s_k$ , given the sample information,  $x$ . This probability is computed by:

$$P(s_k/x) = \frac{P(x/s_k)P(s_k)}{\sum_i P(x/s_i)P(s_i)}$$

The probabilities,  $P(x/s_i)$  are the conditional probabilities of observing the observational information,  $x$ , under the states of nature,  $s_i$ , and the probabilities  $P(s_i)$  are the prior probabilities.

The expected monetary value decisions are formulated in the same way as before, except that the posterior probabilities are used instead of prior probabilities. If the objective is to minimize the expected opportunity loss, the quantity is computed for each action  $a_i$ . The expected opportunity loss in this case is computed by:

$$E(L_i) = \sum_j L_{ij}P(s_j/x) \quad i = 1, 2, 3, \dots, n$$

### Example

It is known that an assembly machine operates at a 5 percent or 10 percent defective rate. When running at a 10 percent defective rate, the machine is said to be out of control. It is then shut down and readjusted. From past experience, the machine is known to run at 5 percent defective rate 90 percent of the time. A sample of size  $n = 20$  has been selected from the output of the machine, and  $y = 2$  defectives have been observed. Based on both the prior and sample information, what is the probability that the assembly machine is in control (running at 5 percent defective rate)?

### Solution

The states of nature in this example relates to the assembly machine defective rates. Thus the states of nature include:

$s_1 = 0.05$ , and  $s_2 = 0.10$  with the assumed prior probabilities of occurrence of 0.90 and 0.10. We are required to use these prior probabilities, in line with the observed sample information, to find the posterior probability associated with the state of nature,  $s_1$ .

In this problem, the “experimental information,  $x$ ” is the observation of  $y = 2$  defectives from a sample of  $n = 20$  items selected from the output of the assembly machine. We need to find the probability that the experimental information,  $x$ , could arise under each state of nature,  $s_i$ . This can be done by referring to the binomial probability distribution table found in the appendix.

Under the state of nature  $s_1 = 0.05$ , we obtain:

$P(x/0.05) = P(n = 20, y = 2/0.05) = 0.925 - 0.736 = 0.189$  (from the binomial distribution table)

Under the state of nature,  $s_2 = 0.10$ , we obtain:

$P(x/0.10) = P(n = 20, y = 2/0.10) = 0.677 - 0.392 = 0.285$  (from the binomial distribution table).

We now employ the Baye’s Law to find the posterior probability that the machine is in control ( $s_1$ ) based on both the prior and experimental information. To make the work easy, we use the Columnar approach to the use of Baye’s Law as illustrated below:

**Table 3.5: Columnar Approach to Use of Baye’s Law**

	(1) State of Nature, $s_i$	(2) Prior, $P(s_i)$	(3) Experimental Information, $P(x/s_i)$	(4) Product, $P(s_i)P(x/s_i)$	(5) Posterior, $P(s_i/x)$
$s_1$	0.05	0.90	0.189	0.1701	0.86
$s_2$	0.10	0.10	0.285	0.0285	0.14
		1.00		0.1986	1.00

Looking at column (4), we observe the product of the entries in columns (2) and (3). These values measure the joint probabilities. The sum of the entries in column (4) is the term in the denominator of the formula for Baye’s Law and measures the marginal probability of observing the experimental information,  $x$ . The posterior probabilities, column (5), are obtained by taking each entry in column (4) and dividing by the sum of the entries in column (4).

Even though we found that 10 percent of the items in the sample is defective (that is, 2 out of the 20 items is defective), the posterior probability that the machine is running at the 10 percent defective rate (running out of control) is only 0.14, which is a little greater than the prior probability that the machine is out of control (0.10). It follows that the probability that the machine is not running out of control is 0.86.

### 3.5 Self -Assessment Exercise

How does decision making under certainty differ from decision making under uncertainty?

#### **4.0 Conclusion**

This unit has exposed you to the basic activities in making decisions under uncertainty. The unit stresses the fact that in problems involving choices from many alternatives, you must identify all the actions that may be taken and all the states of nature whose occurrence may influence decisions. Associated with each action is a list of payoffs. If an action does not involve risk, the payoff will be the same no matter which state of nature occurs.

#### **5.0 Summary**

Most decision-making situations involve the choice of one among several alternative actions. The alternative actions and their corresponding payoffs are usually known to the decision-maker in advance. The decision maker however, does not know which alternative will be best in each case, unless he/she also knows with certainty the values of the economic variables that affect profit. These economic variables are referred to, in decision analysis, as states of nature as they represent different events that may occur, over which the decision maker has no control.

A decision-making procedure, which employs a payoff table and prior probabilities associated with the states of nature to arrive at a decision is referred to as the *Expected Monetary Value decision procedure*. *By prior probability we mean probabilities* representing the chances of occurrence of the identifiable states of nature in a decision problem prior to gathering any sample information. The expected monetary value *decision refers to the selection of available action based on either the expected opportunity loss or the expected profit of the action.*

#### **6.0 Tutor-Marked Assignment**

For each of the following business decision-making problems, list the actions available to the decision maker and the states of nature that might result to affect the payoff:

- (i) The replacement of manually operated packaging machines by a fully automated machine;
- (ii) The leasing of a computer by a commercial bank to process checks and handle internal accounting;
- (iii) The assignment of seven secretaries to seven executives; and,
- (iv) The investment of a company's pension fund.

#### **7.0 References**

Onwe, O. J. (2007) *Statistical Methods for Business and Economic Decisions: A Practical Approach (Lagos: Samalice)*



## **UNIT 14: BUSINESS FORECASTING**

### **Content**

- 1.0 Introduction
- 2.0 Objectives
- 3.0 Steps and Techniques of Forecasting
  - 3.1 Steps in Forecasting
  - 3.2 Types of Forecasts
  - 3.3 Methods of Forecasting
  - 3.4 Self- Assessment Exercise
- 4.0 Conclusion
- 5.0 Summary
- 6.0 Tutor-Marked Assignment
- 7.0 References

### **1.0 Introduction**

Forecasts are based on past performances. In other words, future values are predicted from past values. This assumes that the future will be basically the same as the past and present, implying that the relationships underlying the phenomenon of interest are stable overtime.

Forecasting can be performed at different levels, depending on the use to which it will be put. Simple guessing, based on previous figures, is occasionally adequate. However, where there is a large investment at stake, structured forecasting is essential.

Any forecasts made, however technical or structured, should be treated with caution, since the analysis is based on past data and there could be unknown factors present in the future. However it is often reasonable to assume that patterns that have been identified in the analysis of past data will be broadly continued, at least into the short-term future. In this unit, we look at various steps and techniques of business forecasting with a view to educating you on effective forecasting techniques.

### **2.0 Objectives**

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

1. Know the steps in forecasting
2. Know the various types and techniques of forecasting
3. Make business forecasts

### **3.0 Steps and Techniques of Forecasting**

The starting point in business forecasting is to lay down the necessary steps to be followed. Of importance also is the choice of the appropriate methodology to be used. This sections educates you on these with appropriate examples.

#### **3.1 Steps In Forecasting.**

We outline the basic steps in forecasting as follows:

Step 1. Garther past data: daily, weekly, monthly, yearly.

Step 2. Adjust or clean up the raw data against inflationary factors. Index numbers can be used in deflating inflationary factors.

Step 3. Make forecasts from the “refined” data

Step 4. When the future data (which is been forecast) becomes available, compare forecasts with actual values, By so doing, one will be able to establish the error due to forecasting.

### 3.2 Types of Forecasts.

The basic types forecast are outlined below:

**3.2.1 Short-term Forecasts.** These are forecasts concerning the near future. They, are characterized by few uncertainties and therefore more accurate than distant future forecasts

**3.2.2 Long – term forecasts.** These concern the distant future. They are characterized by more uncertainties than short – term forecasts.

**3.2.3 Extrapolation.** These are forecasts based solely on past and present values of the variable to be forecast. In this case, future values are extrapolated from past and present values.

**3.2.4 Forecasts Based on Established Relationships between the variable to be forecast and other variables**

### 3.3 Methods of Forecasting

There are two generally used methods of forecasting:

1. Moving Averages
2. Trend lines or least squares.

#### 3.3.1 Moving Averages

Moving averages can be used to generate the general picture (or trend) behind a set of data or time series. The general pattern generated can be used to forecast future values. Note that a time series is a name given to numerical data described over a uniform set of time points. Time series occur naturally in all spheres of business activity.

The method of moving Averages can be illustrated by the following example.

A monthly sales data is given:

Sales (N)	Jan.	Feb.	March.	April.	May.	June
Past (Actual)	50	55	70	50	45	90

Using a 3 – period moving averages, the forecast values are:

$$\frac{50 + 55 + 70}{3} = 58 \quad (\text{Feb})$$

$$\frac{55 + 70 + 50}{3} = 58 \quad (\text{March})$$

$$\frac{70 + 50 + 45}{3} = 55 \quad (\text{April})$$

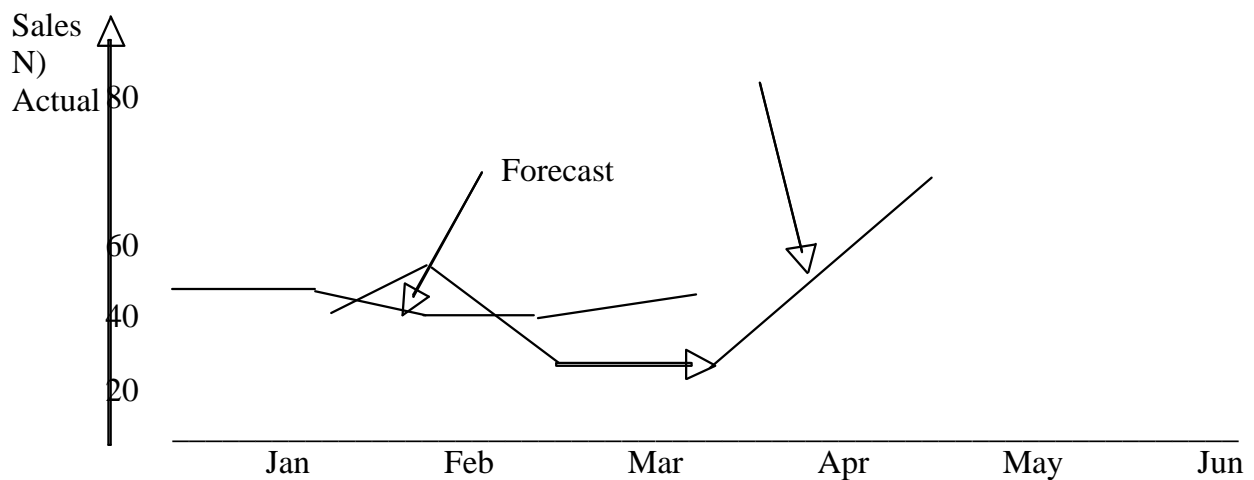
$$\frac{50 + 45 + 90}{3} = 62 \quad (\text{May})$$

We can thus summarize the forecast sales as follows:

Forecast sales (N)	Jan	Feb	March	April	May	June	
Future (forecast)	-	58	58	55	62	-	-

These can be presented in a graph as follows:

**Figure 3.1: Graph of Sales Forecasts.**



### 3.2.1 Least Squares or Trend Lines

The idea behind the use of trend lines in forecasting is based on the assumption that the general picture underlying a given set of data can be reasonably approximated by a straight line. Such a straight line can be extended backwards or forward for predicting past or future values.

#### Example

Suppose the line AB in the following straight line reasonably approximates a set of data for 1995 – 2000

**Figure 3.2: Trend line**

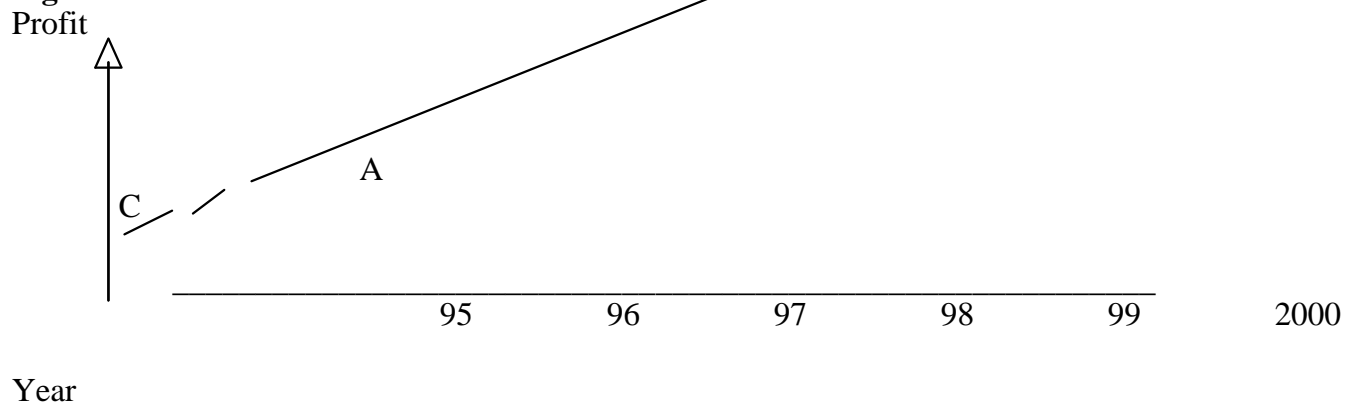


Figure 3.2 indicates that we can forecast profit backwards for years below 1995, using the dotted line AC. Similarly, profits can be forecast for years beyond 2000, using the dotted line BD.

The basic task in using a trend line for forecasting is to determine a line similar to line AB in figure 3.2: then forecasting backwards or forwards is a straight forward activity. The most effective way of determining such a line is the Least-Squares method.

### 3.3.3 The Least – Squares Method.

The least – squares method provides a sound mathematical basis for choosing the best trend line; of all possible trend lines for a given set of time series. This method provides an equation ( with its numerical coefficients) so that the value corresponding to any given year (or period) can be determined by substituting the given year (or period) into the equation.

For example, consider the following periodic data:

Table 5.1

Time Series Data	
Year (Period)	Output
(t)	(y)
1990	50
1991	80
1992	90
1993	49
1994	75
1995	58
1996	82
1997	73
1998	95

With t representing period and y representing output, the equation showing the relationship between time and output (or the estimated trend line) is given below:

$$Y = \hat{a} + b t$$

The Least – Squares method is then used to determine the numerical values of the parameters,  $\hat{a}$  and  $b$

We assume:

- t = 1 in 1990
- t = 2 in 1991
- t = 3 in 1992
- t = 4 in 1993
- t = 5 in 1994
- t = 6 in 1995
- t = 7 in 1996
- t = 8 in 1997
- t = 9 in 1998

Table 3.1 can thus be rewritten as:

t	y
1	50
2	80
3	90
4	49
5	75
6	58
7	82
8	73
9	95

The formulas for the least – squares estimates are as follows:

$$\hat{a} = \bar{Y} - b \bar{t}$$

where  $\bar{Y} = \frac{\sum Y}{n}$ ;  $\bar{t} = \frac{\sum t}{n}$ ;  $n$  = number of pairs of observations

$$b = \frac{\sum ty - n \bar{t} \bar{y}}{\sum t^2 - n (\bar{t})^2}$$

Using the given data and second formula, we get:

t	y	ty	t <sup>2</sup>
1	50	50	1
2	80	160	4
3	90	270	9
4	49	196	16
5	75	375	25
6	58	348	36
7	82	574	49
8	73	584	64
9	95	855	81
45	652	3412	285

$$\begin{aligned} \sum ty &= 9(3412) = 30708 \\ \sum t y &= 45(652) = 29340 \\ \sum t^2 &= 9(285) = 2565 \\ (\sum t)^2 &= (45)^2 = 2025 \\ \frac{\sum y^2}{n} &= \frac{652^2}{9} = 72.44 \\ \frac{\sum t^2}{n} &= \frac{45^2}{9} = 5 \end{aligned}$$

It follows that:

$$\begin{aligned} b &= \frac{30708 - 29340}{2565 - 2025} = \frac{1368}{540} \\ &= 2.53 \\ \hat{a} &= \bar{y} - b \bar{t} \\ &= 72.44 - 2.53 (5) \end{aligned}$$

$$= 72.44 - 12.65$$

$$= 59.79$$

The least – squares line becomes:

$$Y = 59.79 + 2.53 t.$$

This equation can be used any time to forecast the value of any given year, provided the numerical value of the year is appropriately identified.

For example, let us forecast the value of output, Y, for year 2003.

Following the systematic process, the year 2003 is associated with the numerical value,  $t = 14$ , so that for  $t = 14$ ,

$$\begin{aligned} Y_{2003} &= 59.79 + 2.53 (14) && \text{(by substitution)} \\ &= 59.79 + 35.4 \\ &= 95.21 \end{aligned}$$

therefore, the forecast value for output in year 2003 is 95.21

### 3.4 Self -Assessment Exercise

With appropriate examples, explain the major steps in business forecasting.

### 4.0 Conclusion

This unit has exposed you to the principles of business forecasting. Of significance in the discussions were the necessary steps in forecasting, types of forecasts, and methods of forecasting. A thorough review of these presentations will significantly improve your forecasting abilities.

### 5.0 Summary

We outline the basic steps in forecasting as follows:

Step 1. Gather past data: daily, weekly, monthly, yearly.

Step 2. Adjust or clean up the raw data against inflationary factors. Index numbers can be used in deflating inflationary factors.

Step 3. Make forecasts from the “refined” data

Step 4. When the future data (which has been forecast) becomes available, compare forecasts with actual values. By so doing, one will be able to establish the error due to forecasting.

There are basically three types of forecasts including, (i) short-term forecasts; (ii) long-term forecasts; and (iii) extrapolation. There are two generally used methods for forecasting: (i) Moving Averages; and, (ii) Trend lines or least squares.

The idea behind the use of trend lines in forecasting is based on the assumption that the general picture underlying a given set of data can be reasonably approximated by a straight line. Such a straight line can be extended backwards or forward for predicting past or future values.

The least – squares method provides a sound mathematical basis for choosing the best trend line; of all possible trend lines for a given set of time series. This method provides an equation (with its numerical coefficients) so that the value corresponding to any given

year (or period) can be determined by substituting the given year (or period) into the equation.

#### 6.0 Tutor-Marked Assignment

The data given below represent the annual gross revenue (in N' millions) obtained by a Telephone company over the periods 1987 – 1996:

#### Annual Gross Revenues

	Year	Gross Revenue (N' million)
	1987	
13.0	1993	14.1
	1994	15.7
	1995	17.0
	1996	18.4
	1997	20.9
	1998	23.5
	1999	26.2
	2000	29.0
	2001	32.8

- Plot the data on a graph paper
- Fit a least – squares trend line to the data and plot the line on your graph
- 2006?

#### 7.0 References

Onwe, O. J. (2007) Statistical Methods for Business and Economic Decisions: A Practical Approach (Lagos: Samalice)

## UNIT 15: INDEX NUMBERS

### Content

- 1.0 Introduction
- 2.0 Objectives
- 3.0 Definitions, Classifications, and Applications of Price Indices
  - 3.1 Definitions
  - 3.2 Classifications and Application of Price Indices
  - 3.3 Self- Assessment Exercise
- 4.0 Conclusion
- 5.0 Summary
- 6.0 Tutor-Marked Assignment
- 7.0 References

### 1.0 Introduction

It is becoming extremely important to consider indices in managerial functions of forecasting and planning. Index numbers are often used in standardisation of monetary values over time. Managers find standardised economic values to be a reliable way of contractual agreements as monetary variations can result from chronic inflationary pressures. In this unit, you will look at the meaning of index numbers and they way they are used in standardising monetary and/or economic values.

### 2.0 Objectives

By the time you read through this unit, you will be able to:

1. Ascertain the meaning and importance of index numbers
2. Know the various classifications of Price Index
3. Apply the concept of index numbers to everyday business and economic decisions.

### 3.0 Definitions, Classifications, and Applications of Price Indices

#### 3.1 Definitions.

An index number measures the percentage change in the value of some economic commodity over a period of time. Index numbers are expressed in terms of a base of 100. Two or more time periods are involved in the computation of an index number. One of these time periods represents the base time period. The value of the base time period stands as the standard point of comparison, while the values at the other time periods are used to show the percentage in value from the standard value of the base period.

Suppose we wish to compare the average monthly salaries of workers in a given company for:

1980, 1985, 1990, 1995, 2000 and 2002, using 1980 as the base year. The salary index for each of the six years can be calculated using the following relationship:

$$I_k = \frac{\text{Average monthly salary in year K}}{\text{Average monthly salary in 1980}} \times 100$$

The average monthly salaries and their computed index number for each of the six years are presented in table 3.1, referred to as Index Time Series.



**Table 3.1: Index Time Series of Monthly Salaries**

Year	Average monthly salaries(N)	Salary Index(1980 = 100)
1980	1,200	100%
1985	2,000	166.7
1990	2,800	233.33
1995	3,600	300
2000	5,000	416.7
2002	6,500	541.7

As can be observed from the above, each wage index is a percentage that indicates the percentage of 1980 salaries that were earned in a given year. For example, the salary index in 1990 indicates that the monthly salary in 1990 was 233.3% of the 1980 monthly salary.

### 3.2 Classifications and Applications of Price Index

This section focuses on the basic classifications of Price Indices including, the simple aggregate price index, and the weighted aggregate price index.

#### 3.2.1 The Simple Aggregate Price Index.

The Simple Price Index can be defined as the ratio of the sum of commodity prices in a given current period to the sum of same commodity prices in the selected base year (or period).

Symbolically,

$$\text{Simple Price Index, PI} = \frac{\sum P_c}{\sum P_o} \times 100,$$

where  $P_c$  = current year's prices.

$P_o$  = base year's prices.

Note that, given the Quantity Index (QI) defined by:

$$QI = \frac{\sum Q_c}{\sum Q_o} \times 100,$$

the corresponding Value Index (VI) is computed by:

$$VI = \frac{\sum P_c Q_c}{\sum P_o Q_o} \times 100$$

The simple aggregate price index is commonly used in comparing two sets of prices from a wide variety of items.

#### Example.

The average consumer prices, in Naira per Kg, for some staple food items in 1970 and 1990 are given in the following table 3.2. We are required to find the Aggregate Price Index for 1990, using 1970 as the base year.

**Table 3.2: Average Consumer Prices for Selected Staple Food Items**

Item	1970	1990
Sugar	10 40	
Wheat Flour	11 20	
Butter	71 99	
Ground beef	91 186	
Frying Chicken	39 88	

**Solution**

The required aggregate price index for 1990 is:

$$I_{1990} = \frac{40+20+99+186+88}{10+11+71+91+39} \times 100 = \frac{433}{222} \times 100 = 195.06$$

This implies that the prices of the five items are higher by 95.06 percent in 1990 than in 1970.

**3.2.2 Weighted Aggregate Price Index.**

The weighted aggregate price index is the ratio of an aggregate of weighted commodity prices for a given year, K, to an aggregate of the weighted prices of the same commodities in some base year, expressed as a percentage.

Symbolically,

$$I_k = \frac{\sum_{i=1}^n P_{ki} \cdot q_{0i}}{\sum_{i=1}^n P_{0i} \cdot q_{0i}} \times 100$$

**Examples**

An investor's holding in the shares of four companies in 1994 and 1998 is shown in Table 3.3 below:

**Table 3.3 Investor's Holding of Four Companies**

Company	No of shares		Price per share ( N )	
	1994	1998	1994	1998
A	350	400	0.50	1.25
B	200	180	1.25	3.75
C	140	200	6.25	12.50
D	130	150	12.50	18.75

Using 1994 as base year, calculate:

- (f) a simple aggregate price index and
- (g) a weighted aggregate price index for the investor's holding of shares.

**Solutions:**

(a) The simple aggregate index is:

$$I_{1998} = \frac{P_{1998}}{P_{1994}} \times 100 = \frac{400+180+200+150}{350+200+140+130} = 1.134$$

$$I_{1998} = \frac{P_{1998}}{P_{1994}} \times 100 = \frac{400+180+200+150}{350+200+140+130} \times 100$$

$$= 1.134 \times 100$$

$$= 113.4$$

(b) The weighted aggregate price index is:

$$I_{1998} = \frac{P_{1998} q_{1998}}{P_{1994} q_{1994}} \times 100$$

$$= \frac{1.25(400)+3.75(180)+12.50(200)+18.75(150)}{0.50(350)+1.25(200)+6.25(140)+12.5(130)} \times 100$$

$$= \frac{500+675+2500+2812.5}{175+250+875+1625} \times 100$$

$$= \frac{6487.5}{2925} = 2.218$$

$$= 221.79$$

**3.3 Self -Assessment Exercise**

Enumerate what you think are the advantages of index numbers in business decision-making processes.

**4.0 Conclusion**

This unit has expanded your application of statistical concepts by introducing you to the concepts of price indices. The knowledge of these concepts are necessary for business decisions involving realistic monetary values.

**5.0 Summary**

The discussions have defined an index number as a measure of the percentage change in the value of some economic commodity over a period of time. Index numbers are expressed in terms of a base of 100.

Two or more time periods are involved in the computation of an index number. One of these time periods represents the base time period. The value of the base time period stands as the standard point of comparison, while the values at the other time periods are used to show the percentage in value from the standard value of the base period.

Two basic classifications of price indices, including the simple aggregate price index, and the weighted aggregate price index, were of major interest in the unit.

The Simple Price Index was defined as the ratio of the sum of commodity prices in a given current period to the sum of same commodity prices in the selected base year (or period).

Symbolically,

$$\text{Simple Price Index, PI} = \frac{\sum P_c}{\sum P_o} \times 100,$$

where  $P_c$  = current year's prices.

$P_o$  = base year's prices.

While the weighted aggregate price index is the ratio of an aggregate of weighted commodity prices for a given year, K, to an aggregate of the weighted prices of the same commodities in some base year, expressed as a percentage.

Symbolically,

$$I_k = \frac{\sum_{i=1}^n P_{ki} q_{ki}}{\sum_{i=1}^n P_{oi} q_{oi}} \times 100$$

## 6.0 Tutor-Marked Assignments

Given the following hypothetical data on the primary market operation on the Nigeria Stock Exchange for the last quarter of 1999 and 2000.

**Table 3.4: Summary of New Issues of Companies**

Type of offer	Volume ( N million )				Value ( N 'million )	
	1999		2000		1999	2000
Right Issues	350	400	378	480		
Private Placement	200	180	260	360		
Debenture Stock	140		200		70	220

Using 1999 as the base year, compute an appropriate:

- Simple aggregate price index.
- Quantity index.
- Value index.
- Weighted aggregate price index.

For the above investments.

$$\text{HINT: } P_i = \frac{P_i q_i}{q_i}$$

## 7.0 References

Onwe, O. J. (2007) Statistical Methods for Business and Economic Decisions: A Practical Approach (Lagos: Samallice)

