

INTRODUCTION TO STATISTICS

Stat 192

(For Distance Education Learners)

Compiled By

**Zewditu Girma
Yoseph Kanaa
Mulugeta Ayalew**

Department of Statistics and Mathematics

UNITY UNIVERSITY COLLEGE

May 2003

Preface

This book “***Introduction to Statistics*** (Stat192)” is prepared for distance learners and is intended to give basic statistical concepts. Effective decision-making in business and economic problems rely on Statistics. Reliable data must be collected and analyzed in order to make sound conclusions. The objective of this course is then to make the student aware of the fundamental ideas and concepts and hence apply it or use it as a tool while dealing with statistical data and data analysis.

As every scientific research area commencing the preparation of simple project needs the application of the subject, we focused on introducing statistics from the fundamental uses of data collection towards solving some problems involving common statistical distributions in hypothesis testing.

Block one is concerned with introducing the subject. Block two is about classification and presentation of data. The third and the fourth blocks cover data analysis, particularly measures of averages and dispersion. In the fifth block, the concept of elementary probability will be discussed. Block six is about probability distribution and block seven is concerned with sampling and sampling distribution. The last two blocks deal with estimation and hypothesis testing, and linear correlation and regression analysis respectively.

In each block you will find solved problems (examples) and check your progress questions whose answers are given at the end of each block. Students are advised not to refer the answers before trying them. We have also included model examination questions.

We welcome suggestions and comments about this book through the University College.

The Producers.

TABLE OF CONTENTS

<u>CONTENTS</u>	<u>Page</u>
BLOCK 1: INTRODUCTION TO STATISTICS	
UNIT 1: Introduction to Statistics	1
UNIT2: Collection of Data	21
BLOCK 2: CLASSIFICATION AND PRESENTATION OF STATISTICAL DATA	
UNIT 3: Classification of Data.....	33
UNIT 4: Graphical Presentation of Data	50
BLOCK 3: MEASURES OF CENTRAL TENDENCY	
UNIT 5: Definition and Purpose of Averages	66
UNIT 6: Mathematical Measures of Central Tendency.....	68
UNIT 7: Positional Measures of Central Tendency.....	94
UNIT 8: Relationship between Mean, Median, and Mode	109
BLOCK 4: MEASURES OF DISPERSION (VARIATION)	

UNIT 9: Meaning and Types of Dispersion	
.....	119
UNIT 10: Range and Quartile Deviation	
.....	135
UNIT 11: Mean Deviation	
	148

BLOCK 5: ELEMENTARY PROBABILITY THEORY

UNIT 12: Counting Methods	
	190
UNIT 13: Probability	
	212

BLOCK 6: INTRODUCTION TO PROBABILITY DISTRIBUTIONS

UNIT 14: Discrete Probability Distribution	
.....	235
UNIT 15: Continuous Probability Distribution	
.....	257

BLOCK 7: SAMPLING AND SAMPLING DISTRIBUTION OF THE MEAN

UNIT 16: Concepts And Reasons For Sampling.....	272
UNIT 17: Types Of Sampling Techniques	
.....	276
UNIT 18: Sampling Distribution Of The Sample Mea	
.....	288

BLOCK 8: ESTIMATION AND TEST OF HYPOTHESIS

UNIT 19: Estimation.....	
300	
UNIT 20: Test Of Hypothesis.....	
317	

BLOCK 9: CORRELATION AND REGRESSION

UNIT 21: Correlation	
342	
UNIT 22: Regression	
364	

APPENDIXES	
387	

BLOCK 1: INTRODUCTION

UNIT 1: INTRODUCTION TO STATISTICS

UNIT 2: COLLECTION OF DATA

INTRODUCTION

When did the practice of Statistics start? It is not as such possible to tell the exact point at which Statistics was started. However, it is not done consciously, the practice of Statistics in one of the things human beings perform in their day-to-day activities. All the estimations, forecasts, comparisons, averaging, and so on are part of Statistics. Statistics is something that is part of our lives. We encounter it in all walks of life, but most of the time we don't recognize it.

It was around the 16th century that Statistics considered as a formal discipline. After passing through continuous developmental stages, Statistics currently reached at a point more than ever in its history.

Nevertheless, Statistical results are not ends by themselves. They are used to develop and strengthen ideas or theories that we have in other fields like economics, social and natural science, business, medicine, etc. Rober W. Bugess, as cotted by C.B Gubta, put the aim of Statistics as:

“ The fundamental gospel of Statistics is to push back the domain of ignorance, prejudice, rule of thumb, arbitrary or premature decisions, traditions and dogmatism, and to increase the domain in which decisions are made and principles are formulated on the basis of analyzed facts.”

BLOCK 2: CLASSIFICATION AND PRESENTATION OF STATISTICAL DATA

UNIT 3: CLASSIFICATION OF DATA

UNIT 4: GRAPHICAL PRESENTATION OF DATA

INTRODUCTION

In the preceding block, you have learnt about the collection of data. When conducting a statistical study, you must gather data for the particular variable under study.

In this block, you will learn about classification and presentation of data. The first part deals with 'classification of data' and the following unit deals with 'presentation of data'.

The purpose of this block is to explain how to organize data by constructing 'frequency distribution' and how to present the data by constructing graphs and charts. The graphs and charts illustrated in this block include histograms, frequency polygons, ogives, pie charts, bar charts, time series graphs, and pictographs (pictograms).

BLOCK 3: MEASURES OF CENTRAL TENDENCY

UNIT 5: DEFINITION AND PURPOSE OF AVERAGES

UNIT 6: MATHEMATICAL MEASURES OF CENTRAL TENDENCY

UNIT 7: POSITIONAL MEASURES OF CENTRAL TENDENCY

UNIT 8: MEANING AND TYPES OF DISPERSION

INTRODUCTION

Visual presentation of data would disclose some characteristic features of a mass of data. And further summarization of data is so essential to show the relationship between variables and to correlate one variable with another. To describe the characteristic features of the entire mass of data with single quotient, the more obvious measure that helps to make quicker and better decision is the measure of Central Tendency, also called the Averages. An average gives a bird's eye view of huge mass of data, which are not easily intelligible, since it refers to a numerical value that is a central point about which other values in a series get dispersed.

BLOCK 4: MEASURES OF DISPERSION (VARIATION)

UNIT 9: RANGE AND QUARTILE DEVIATION

UNIT 10: MEAN DEVIATION

UNIT 11: STANDARD DEVIATION

INTRODUCTION

Although, averages serve the purpose of describing the characteristics of a distribution, they cannot give a comprehensive idea and as such hide many important facts about the distribution. One of the important facts concealed (hidden) by the measures of central tendency is regarding the variability in the values. Averages fail to explain the extent of deviation from the average value in the distribution. In the absence of such information, the averages among different distributions cannot be meaningfully compared.

Comparison can be made only in cases where all the values are the same as the average or when there are no significant deviations in the

values from the average. However, this situation is very rare especially, in case of the data pertaining to various socio-economic problems. Hence, any average by itself fails to give if it is not identified and accompanied information describing the range of things and their deviations from the central value. Some other values are required for describing the characteristic of distribution and making comparisons with other distributions. As Simpson and Kafka rightly pointed out, “An average doesn’t tell the full story. It is hardly fully representative, of a mass unless we know the manner individual items scatter around it. A further description of the series is necessary if we are to gauge how representative the average is “. To support and supplement measures of central tendency, other measures like dispersion, and skew ness are devised. While the measures of dispersion explain the degree of variation in the individual items from the central value, skew ness measures the degree of symmetry or asymmetry of the distribution.

BLOCK 5: ELEMENTARY PROBABILITY THEORY

UNIT 12: COUNTING METHODS

UNIT 13: PROBABILITY

INTRODUCTION

The theory of probability is a far-reaching branch of statistics, which helps to obtain numerical information about the occurrence or non-occurrence of an event. It is a measure of the likely hood that an event will happen in the future which plays a major role in our daily life

where we continually faced with decisions leading to uncertainty. The use of information from descriptive data requires knowledge of probability as one of the major tools in drawing valid generalization and wise and reasonable decisions about the large size population from which the sample data is taken.

Dear students, this block consists of two units; the first of which is counting methods which gives you clear idea of how to determine the possible number of elements in an event as well as in a sample space of an experiment, while the second unit give you the techniques of determining the probability of an event. You also need to realize that in ordinary language, probability chance, likely hood, and odds are interchangeable.

BLOCK 6: INTRODUCTION TO PROBABILITY DISTRIBUTIONS

UNIT 14: DISCRETE PROBABILITY DISTRIBUTION

UNIT 15: CONTINUOUS PROBABILITY DISTRIBUTION

INTRODUCTION

In the preceding block, you have learnt about probability, one of the most important tools in statistics and which is important in decision making as it provides a mechanism for measuring, expressing and analyzing the uncertainties associated with future events.

You also saw that the probability of an event could be computed either by summing the probabilities of the experimental outcomes (sample points) comprising the event or by using the relationships established by the addition, conditional probability and multiplication laws of probability.

In this block, you will learn what a 'probability distribution' is. The first part deals with discrete probability distribution and the in the following unit, an introduction to continuous probability distribution will be dealt with, focusing on the normal probability distribution.

**BLOCK 7: SAMPLING AND SAMPLING
DISTRIBUTION OF THE MEAN**

UNIT 16: CONCEPTS AND REASONS FOR SAMPLING

UNIT 17: TYPES OF SAMPLING TECHNIQUES

UNIT 18: SAMPLING DISTRIBUTION OF THE SAMPLE MEAN

BLOCK 9: CORRELATION AND REGRESSION

UNIT 21 CORRELATION

UNIT 22 REGRESSION

INTRODUCTION

In this block, we shall look at methods, which investigate whether two quantitative variables are related. In our practical life, we come across different sets of data that deal with more than one variable which are interrelated and interdependent. For example, the instructor wonders whether the Mathematics mark and the Statistics mark are related: did a good performance in one subject go with a good performance in the other?

She/he decides that s/he can most easily discover this by plotting the marks for all the students on a sheet of graph paper. In addition, she/he tries to see the relationship mathematically. These mathematical methods are Karl Pearson's coefficient of correlation and Spearman's Rank coefficient of correlation.