INTRODUCTORY COLLEGE MATHEMATICS

HACKWORTH
and
HOWLAND

SAUNDERS SERIES IN MODULAR MATHEMATICS

Statistics

INTRODUCTORY COLLEGE MATHEMATICS

Statistics

ROBERT D. HACKWORTH, Ed.D.
Department of Mathematics
St. Petersburg Junior College at Clearwater
Clearwater, Florida

and

JOSEPH HOWLAND, M.A.T.
Department of Mathematics
St. Petersburg Junior College at Clearwater
Clearwater, Florida

SAUNDERS SERIES IN MODULAR MATHEMATICS

W. B. Saunders Company: West Washington Square
Philadelphia, PA 19105

12 Dyott Street
London, WC1A 1DB

833 Oxford Street
Toronto, Ontario M8Z 5T9, Canada

INTRODUCTORY COLLEGE MATHEMATICS 0-7216-4418-X
Statistics

 Library of Congress catalog card number 75-23625.

Last digit is the print number: 9 8 7 6 5 4 3 2 1

PREFACE

Statistics

This book is one of the sixteen content modules in the Saunders Series in Modular Mathematics. The modules can be divided into three levels, the first of which requires only a working knowledge of arithmetic. The second level needs some elementary skills of algebra and the third level, knowledge comparable to the first two levels. *Statistics* is in level 2. The groupings according to difficulty are shown below.

Level 1	Level 2	Level 3
Tables and Graphs	*Numeration*	*Real Number System*
Consumer Mathematics	*Metric Measure*	*History of Real Numbers*
Algebra 1	*Probability*	*Indirect Measurement*
Sets and Logic	*Statistics*	*Algebra 2*
Geometry	*Geometric Measures*	*Computers*
		Linear Programming

The modules have been class tested in a variety of situations: large and small discussion groups, lecture classes, and in individualized study programs. The emphasis of all modules is upon ideas and concepts.

Because every citizen today faces a barrage of numerical information *Statistics* is appropriate for all students regardless of their major. The level of difficulty in *Statistics* is appropriate for freshmen and sophomore students.

Statistics begins by explaining the concepts of samples, populations and descriptive versus inferential statistics. After emphasizing skill in constructing bar graphs, histograms, and line graphs, the concepts of mean, median, mode, and the range of grouped and ungrouped data are presented. The module ends by discussing standard deviation and using the normal curve to find probabilities of data occurrence.

In preparing each module, we have been greatly aided by the valuable suggestions of the following excellent reviewers: William Andrews, Triton College, Ken Goldstein, Miami-Dade Community College, Don Hostetler, Mesa Community College, Karl Klee, Queensboro Community College, Pamela Matthews, Chabot College, Robert Nowlan, Southern Connecticut State College, Ken Seydel, Skyline College, Ara Sullenberger, Tarrant County Junior College, and Ruth Wing, Palm Beach Junior College. We thank them and the staff at W. B. Saunders Company for their support.

Robert D. Hackworth
Joseph W. Howland

NOTE TO THE STUDENT

Objectives:

Upon completing this unit the reader is expected to be able to demonstrate the following skills and concepts:

1. Knowledge of samples, populations, descriptive and inferential statistics.

2. Construct bar graphs, histograms, line graphs for sets of grouped or ungrouped data.

3. Find the mean, median, mode, and range of both grouped and ungrouped data.

4. Compute the standard deviation of a set of data.

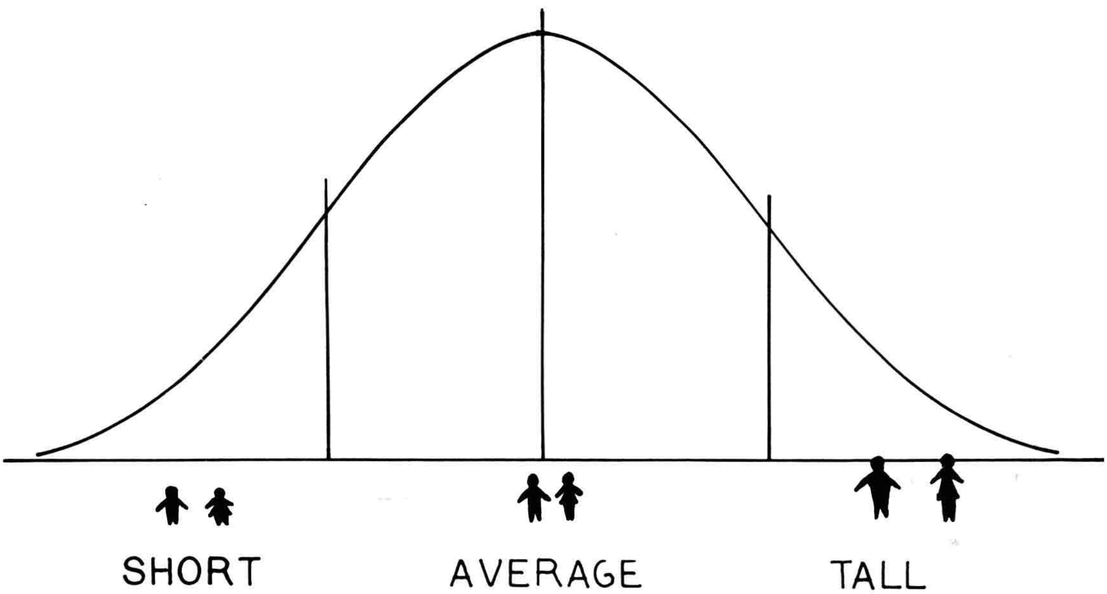
5. Use normal distributions and normal curves in finding probabilities of data occurrence.

Three types of problem sets, with answers, are included in this module. Progress Tests appear at the end of each section. These Progress Tests are always short. The questions asked in Progress Tests always come directly from the material in the section immediately preceding the test.

Exercise Sets appear less frequently in the module. More problems appear in an Exercise Set than in a Progress Test. These problems arise from all sections of the module preceding the Exercise Set. The problems in Part I of the Exercise Sets are specifically chosen to match the objectives of the module. Part II of each Exercise Set contains Challenge Problems.

A Self-Test is found at the end of the module. The Self-Test contains problems representative of the entire module.

In learning the material, the student is encouraged to try each problem set as it is encountered, check all answers, and restudy those sections where difficulties are discovered. This procedure is guaranteed to be both efficient and effective.


SHORT
AVERAGE
TALL

CONTENTS

STATISTICS

INTRODUCTION

For most people, a blood pressure reading (140 over 80), a metric temperature (25°Celsius), and a financial report entry ($1.37 earnings per share) are examples of numbers that have little meaning. Is a blood pressure reading of 140 over 80 a sign of health or sickness? Is a temperature of 25° Celsius warm or cool? Is an earnings per share of $1.37 good or bad? The numbers themselves are of little value unless the reader knows the context in which they are used, but each of us is constantly bombarded by information, given in numbers, for situations that are novel or strange to our own experience.

Statistics is a field of mathematics for presenting quantified (number) information in a meaningful, understandable manner. Statistics is sometimes viewed as a problem-solving mathematics because it can be used to convert raw, unrelated data to numbers whose meaning can be the basis for well-reasoned, wise decision-making. In science, industry, education, and government there is a constant need to process the number information that is produced, discarding the irrelevant, and carefully analyzing and interpreting the important. Statistics is the study of the best processes for using and explaining quantified information.

Beneficial as statistics can be when properly used, one of the most important outcomes of a study of statistics should be an understanding of the misleading and dangerous information that often results from "statistics" that have been misused. Two types of misuses of statistics are common. First, some statistics are the result of good intentions and poor procedure. Second, some statistics are the result of poor intentions with procedures selected to provide misleading information.

A quotation sometimes attributed to Mark Twain and sometimes attributed to an English nobleman says, "There are lies, damn lies, and statistics." That quotation, for people who generally accept numerical information without question, should serve as a strong warning against such ready acceptance in the future. Numerical information, regardless of its source, deserves close attention and study before being accepted as fact.

Two examples of well-intentioned, but poor use, of statistics can be cited. The first example illustrates how conclusions are made from statistics that are not justified by the data. The parents of two boys obtained IQ test scores for their offspring. One son had a score of 110 and the other scored 108. The first son was thereafter considered the smartest and treated accordingly. For a variety of reasons, there is no statistical justification for interpreting the 110 score as showing, much less proving, greater intelligence than the 108 score. This same type of misuse of statistics accompanies the return of tests in almost every classroom. Students conclude that they, not their tests, are better or worse than others.

The second example illustrates how the meaning of raw data must be understood before any arithmetical computations are made. A piece of music was written partly in three-quarter time and partly in two-quarter time. A student was asked what the tempo for the piece should be. The student found the average of $\frac{3}{4}$ and $\frac{2}{4}$ which was $\frac{5}{8}$. For any music student the absurdity of the answer, five-eighths time, must be apparent.

The intentional misuse of statistics in our society arises both out of outright dishonesty and an opinion that figures should be manipulated until the desired answer is obtained. Dishonest statistics frequently arise from surveys that are purposely biased in favor of a particular product or point-of-view. Whenever the announcer says, "Nine out of ten doctors...," "75% of the residents of Milwaukee...," or "Three out of four housewives..." the listener would be wise to be cautious.

It is difficult and sometimes time-consuming to differentiate between statistics properly used and those improperly used. Even an expert statistician may have difficulties because the type of information needed to verify or invalidate statistics is usually not readily available. However, there are four important questions that should be asked about any statistical study. If they are all answered affirmatively the statistics are likely to be factual. The questions are:

1. Were procedures employed to avoid bias or misrepresentation in the collection of the original (raw) data?

2. Were the purposes of the statistical study completely defined prior to any mathematical treatment of the data?

3. Were the statistical procedures employed decided upon prior to any mathematical treatment of the data?

4. Are the conclusions limited to the original purposes of the study?

Progress Test 1

Each of the following is a misuse of statistics. State why:

1. A sociologist conducted a study to determine if women were having fewer children today than 20 years ago. At the end of the study, it was announced that birth control pills were reducing the rates of birth.

2. A pharmaceutical company asked 10 friendly doctors to evaluate a new drug compared to other similar drugs. At the end of the study, the results favoring the new drug were widely publicized.

3. After using the Exono College entrance examination for five years, the staff announced the purpose of the test was correlating scores with vocational choices.

4. Three tries at statistical tests failed to show the strength of a political party in Guffaw County, but the fourth try was successful.

DESCRIPTIVE AND INFERENTIAL STATISTICS

There are two main types of statistics. One is called descriptive statistics and the other is called inferential statistics. Descriptive statistics seeks to accurately describe the situation from which it gathers its raw data. Descriptive statistics deals with the present time and limits itself to describing present conditions for the group of figures actually studied.

Inferential statistics deals with inferences, implications, and generalizations. Inferential statistics may be concerned with the future or might broaden its conclusions to include other groups of people than those actually studied.

Some of the mathematical manipulations of statistics are applied to both descriptive and inferential situations. Often, however, there are different mathematics involved. In either case, the end-products or conclusions are different. It is important to recognize that a descriptive statistical study and an inferential statistical study have these different objectives. A lack of understanding of the differences between descriptive and inferential statistics can easily lead to a misinterpretation of some statistical studies.

The data below shows the lengths of all babies born in a particular hospital in May. The lengths were measured to the nearest inch.

19, 21, 20, 21, 20, 23, 18, 21, 21, 24, 19,

21, 20, 19, 22, 21, 17, 22, 20, 20, 20, 21,

19, 23

For descriptive statistics the figures on baby lengths may be used in studies to:

1. Establish size measurements for later comparisons.

2. Correlate body length with body weight for these particular babies.

3. Make comparisons with the diets followed by the mothers of the infants.

For inferential statistics the figures on baby lengths might be used in studies to:

1. Predict the lengths of babies born in the future at that particular hospital.

2. Predict the length of babies born in other comparable hospitals that same month.

3. Establish manufacturing sizes for cribs or diapers for the following year.

A second example involves a larger set of raw data. The following are the mathematics test scores of 100 college freshmen majoring in Business at Exono College one year.

65,	73,	53,	75,	91,	49,	83,	87,	74,	85,
76,	71,	84,	65,	38,	88,	71,	46,	68,	45,
92,	84,	71,	69,	53,	81,	74,	56,	89,	92,
57,	74,	76,	85,	90,	96,	46,	59,	63,	56,
30,	59,	80,	60,	57,	65,	84,	68,	71,	74,
79,	72,	85,	68,	71,	25,	54,	82,	75,	72,
58,	91,	45,	21,	83,	82,	65,	72,	74,	90,
32,	46,	25,	49,	55,	70,	82,	84,	68,	74,
63,	72,	78,	76,	85,	71,	62,	86,	59,	78,
73,	71,	84,	52,	41,	93,	84,	79,	71,	75

For descriptive statistics the test scores might be used in studies to:

1. Assign each student to a particular mathematics course.

2. Compare this year's freshman class with that of last year.

3. Correlate test scores with achievement in the Introduction to Business course.

For inferential statistics the test scores might be used in studies to:

1. Predict student success in Exono College.

2. Forecast the number of accounting sections needed the following semester.

3. Estimate the mathematical background of the freshman business majors next year.

Progress Test 2

Label each of the following as either descriptive or inferential statistics:

1. A Gallup Poll was taken to determine what portion of Americans today favor the legalization of marijuana.

2. A medical research team determined the effect of a particular drug on an experimental group of rats.

3. A government economist used last month's unemployment data to forecast next year's economic growth.

4. A math teacher used a test on the first day of class to counsel students on the advisability of their remaining in the course.

SAMPLES AND POPULATIONS

A statistical study may attempt to support conclusions about people, medicines, tests, machine parts, economic systems, or any other grouping of people, objects, or things related in some manner. Regardless of the subject of the study, the statistics are always concerned with two groups. One group is called the sample and the other is called the population.

The sample of a statistical study consists of the group from which the raw data is obtained. The population of a statistical study is the group to which the conclusions apply. Sometimes the sample group and the population group are identical, but frequently, the sample group is a fractional part of the population.

A teacher who is studying the achievement levels of his class of 30 students has a population group of 30. With such a relatively small number in the population there is little, if any, need to study only a fractional part of the group as the sample.

A college administration which is studying the achievement levels of its student body consisting of 8,500 students has a population of 8,500. The paperwork alone in acquiring data on 8,500 students may prohibit the study, but a sample consisting of 30 to 50 students might be easily studied.

Not all populations are available for study and the researcher is forced to deal with samples. A study of the reaction of individuals faced with a danger situation (nuclear attack for instance) cannot possibly be developed on the entire population.

There is a tendency to rely only upon studies in which the entire population is used as the sample. One of the major areas of statistics, however, is the development of methods for assuring that conclusions reached upon the basis of samples are valid for their populations. It is beyond the scope of this module to show how the validity of samples is supported, but the reader should understand that statisticians do have justifiable faith in their ability to draw conclusions from a sample that are directly applicable to the population.

Nevertheless, one of the reasons that statisticians are comfortable when working with samples is the exceeding care that is taken in the selection process of a sample. Whenever a sample, different from the population, is used there is the possibility that it may or may not share the attributes of its population. A "good" sample must reflect the same relevant attributes as its population. A "poor" sample is one which may not reflect the same relevant attributes as its population. Notice that a sample is "poor" whenever there is reasonable suspicion to believe that is <u>may not</u> reflect the attributes of its population.

If the population were the set of all college students in the United States there would be a tremendous number of sub-groups of the population. Depending upon the particular statistical study, some of these sub-groups would be "good" samples and some would be "poor" samples. It is relatively easy to list a number of sub-groups that would be "poor" samples. Each of the following is a sub-group for the set of all college students in the United States:

1. The freshman class at Harvard is a sub-group that would be a "poor" sample for a population study of the economic costs of a college education.

2. A sub-group composed entirely or predominantly of Roman Catholics would be a "poor" sample for a study of the population's attitude toward legalized abortion.

3. A sub-group composed of students at colleges east of the Mississippi would be a "poor" sample for any study in which geographical location may have any effect whatsoever.

4. A sub-group composed entirely or predominantly of females would be a "poor" sample for a study of the population's attitudes toward women executives.

The preceding four examples are extreme cases of obviously "poor" samples. In actual situations only a researcher intending to misuse statistics would select such samples. However, real life statistical studies often involve samples that have less obvious, but equally important, drawbacks. The following Progress Test is an opportunity for the reader to discover some of the subtle difficulties involved in selecting a sample.

Progress Test 3

A government teacher asked each member of his class to interview a representative sample of the Exono College student body to determine attitudes toward the campus police force. The sampling methods of four students are given below. In each case state a reason why the sample may not be representative of the population.

1. Student A took 50 questionnaires into the student coffee shop. Some students griped at first, but every student asked filled out the questionnaire.

2. Student B sent the questionnaire by mail to the official address of every student registered at Exono College. 15% of the students replied.

3. Student C asked 50 friends to complete the questionnaire and each one did.

4. Student D got an alphabetical list of every tenth student registered at Exono College. Most of them attended a meeting called for the purpose of filling out the questionnaire and, although there was a lot of conversation about some of the questions, almost every student at the meeting completed a questionnaire.

BAR GRAPHS, HISTOGRAMS, LINE GRAPHS AND NORMAL CURVE

The remainder of this module is concerned with organizing data and statistical methods for describing sets of data.

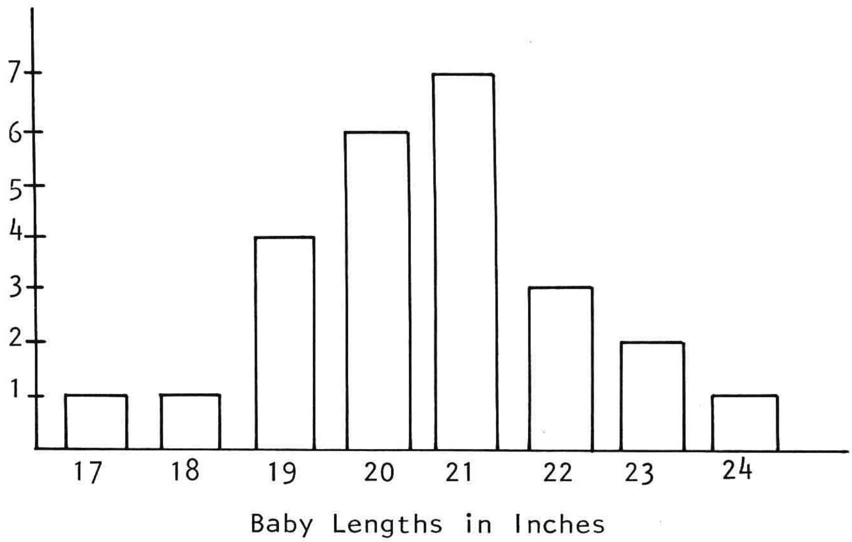
Earlier, two sets of data -- one involving lengths of babies and the other, scores of business majors on a mathematics test -- were given. At that time, the numbers were listed without any attempt at organization. Both sets of data are shown below, but the order in listing lengths and scores has been changed so that both lists proceed from the smallest item to the largest.

Baby lengths: 17, 18, 19, 19, 19, 19, 20, 20, 20, 20, 20, 20, 21, 21, 21, 21, 21, 21, 21, 22, 22, 22, 23, 23, 24.

Test scores: 21, 25, 25, 31, 32, 38, 41, 45, 45, 46, 46, 46, 49, 49, 52, 53, 53, 54, 55, 56, 56, 57, 57, 58, 59, 59, 59, 60, 62, 63, 63, 65, 65, 65, 65, 68, 68, 68, 68, 69, 70, 71, 71, 71, 71, 71, 71, 71, 71, 72, 72, 72, 72, 73, 73, 74, 74, 74, 74, 74, 74, 75, 75, 75, 76, 76, 78, 78, 79, 79, 80, 81, 82, 82, 82, 83, 83, 84, 84, 84, 84, 84, 84, 85, 85, 85, 85, 86, 86, 87, 88, 89, 90, 90, 91, 91, 92, 92, 93, 96.

The lists on page 9 are examples of simple statistics at work. The raw data of the original lists have been altered only by listing in order of size, but this alteration, in itself, simplifies the data and makes it easier to understand.

Another way of showing the information is through the use of bar graphs. The graph below shows the data on baby lengths.



The bar graph above has each data item from the list of baby lengths shown along the horizontal axis. The heights of the bars indicate the number of times the data items appeared in the list. The bar graph below shows the test score data, but differs from the graph on baby lengths in one very important aspect.