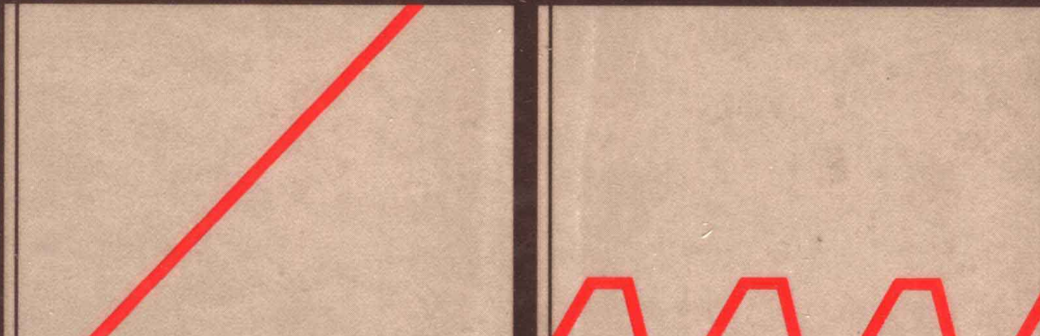


Richard W. Madsen/Melvin L. Moeschberger

INTRODUCTORY STATISTICS FOR BUSINESS AND ECONOMICS

INTRODUCTORY STATISTICS

For Business and Economics

Richard W. Madsen
University of Missouri—Columbia

Melvin L. Moeschberger
The Ohio State University

PRENTICE-HALL, INC., ENGLEWOOD CLIFFS, NEW JERSEY 07632

Library of Congress Cataloging in Publication Data
Madsen, Richard W.
Introductory statistics for business and economics.
Includes index.
1. Social sciences—Statistical methods.
2. Statistics. 3. Commercial statistics. 4. Economics
—Statistical methods. I. Moeschberger, Melvin L.
II. Title.
HA29.M232 1983 519.5 82-25005
ISBN 0-13-501577-4

Printed in the United States of America
10 9 8 7 6 5 4 3 2 1

ISBN 0-13-501577-4

Prentice-Hall International, Inc., London
Prentice-Hall of Australia Pty. Limited, Sydney
Editora Prentice-Hall do Brasil Ltda. Rio de Janeiro
Prentice-Hall Canada Inc., Toronto
Prentice-Hall of India Private Limited, New Delhi
Prentice-Hall of Japan, Inc., Tokyo
Prentice-Hall of Southeast Asia Pte. Ltd., Singapore
Whitehall Books Limited, Wellington, New Zealand

PREFACE

Probability and statistics play an important role in our present society, both in our professional and our daily lives. This book was written to help its readers learn statistical concepts that will be useful to them in both roles. It contains sufficient material to be used as a text in a two-semester sequence, or it can be used as a one-semester course by including only selected sections of the first eight or nine chapters. (Some sections suggested for omission are denoted by a star.) A possible outline for such a course is given at the end of this preface.

We believe that readers with a good background in algebra will be mathematically comfortable with the book. While we have included proofs for some theorems, generally we have left the more difficult proofs to more advanced courses. Such a mathematical prerequisite should pose no unnecessary obstacle to this book's intended audience: namely, undergraduate students in business and economics.

Not *all* students take the time to read the preface of a book (in fact, some do not even read the book!), but for those who do, we would like to make some suggestions. First, before each lecture, read over the material to be covered by your instructor. In this first reading, you may wish to skim the examples. After the lecture, read the material a second time, this time with pencil, paper, and calculator at hand. Try to follow and work out equations as necessary. When you come to examples, try to work them out on your own. If you have difficulty, read the solution to the example given in the text. Also try to sharpen your problem-solving skills by working the odd-numbered exercises at the end of each chapter. Check your answers with those given in the appendix. If you have difficulty, go back and reread the relevant section of the text. If you still have trouble solving a problem, get outside help.

We would like to point out to the instructor some of the features of this book. One feature is flexibility. As we mentioned earlier, the book can be used as a text in either a one- or two-semester course. For the benefit of students who are somewhat weaker mathematically, it would be possible to omit some of the proofs to theorems. Most of the later chapters (those on multiple regression, analysis of variance, nonparametric statistics, Bayesian statistics, decision theory, and time series) can be covered in most any order once the basics of statistical inference have been covered. As another example of flexibility, an instructor so inclined could cover most of the material on descriptive statistics (Chapter 6) at the beginning of the course rather than later on. We have so positioned the chapter on descriptive statistics, however, because we want the student to see the relationship of data collection to the sampling theory developed earlier.

The given exercises and examples are of two types: routine problems to give the student practice in evaluating formulas and using tables, and word problems to help the student see the practical application of the theory. Most of the word problems are business oriented, although we have included problems with other orientations for variety and to show the broad applicability of statistics. Several problems are based on real data. These problems are denoted by an asterisk (*).

As with all works of this type, this book is not simply the product of our own work. We are grateful to Rita Merola for supplying us with the real estate data in Appendix A2. We acknowledge the helpful comments from reviewers of the manuscript. We are also grateful for comments from students and instructors who have used our earlier book *Statistical Concepts*, since much of this book is based on that work. We appreciate the support we received from our wives and children during writing. In addition to her role of support, Carole Madsen contributed greatly by typing the manuscript. (In return for this personal acknowledgment, Carole has agreed to accept all responsibility for typographical errors in this book.)

OUTLINE FOR A ONE-SEMESTER COURSE

The following is a suggested outline giving sections that might be covered in a one-semester course. Sections could be added or deleted at the instructor's discretion. A more detailed outline is given in the Teacher's Manual.

Chapter 1	Sections 1.1-1.9
Chapter 2	Sections 2.1-2.8
Chapter 3	Sections 3.1-3.5
Chapter 4	Sections 4.1-4.4
Chapter 5	Sections 5.1-5.2, 5.4, 5.6-5.7
Chapter 6	Sections 6.1-6.6
Chapter 7	Sections 7.1-7.2, 7.4-7.5, 7.10
Chapter 8	Sections 8.1-8.3, 8.5
Chapter 9	Sections 9.1, 9.2, 9.4

CONTENTS

1

PROBABILITY

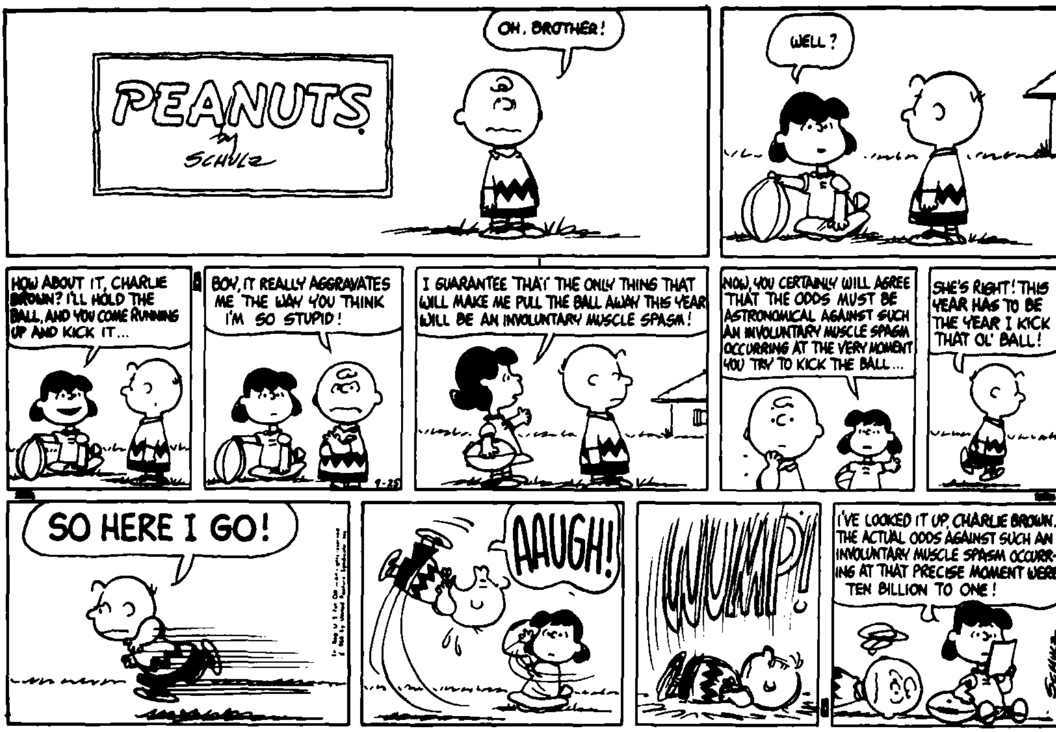
1.1 INTRODUCTION

A "Time Essay" in *Time* magazine, entitled "Getting Dizzy by the Numbers," began,

> 'The very hairs of your head,' says Matthew 10:30, 'are all numbered.' There is little reason to doubt it. Increasingly, everything tends to get numbered one way or another, everything that can be counted, measured, averaged, estimated or quantified. Intelligence is gauged by a quotient, the humidity by a ratio, the pollen by its count, and the trends of birth, death, marriage and divorce by rates. In this epoch of runaway demographics, society is as often described and analyzed with statistics as with words. . . .

We certainly agree with the essay writer that in this day and age statistics is used in virtually all areas of life. We believe that understanding statistical concepts can be very helpful to you both in and out of your professional life. Since fundamental statistical concepts are based on the study of probability, we begin our study there.

Most of you are no doubt familiar with the saying that there is nothing certain in this life except death and taxes—a rather cynical way of conveying the idea that life is filled with uncertainty and variability. It also illustrates the way people make inferences about the future by considering their past experience. "Uncertainty and variability" and "inference" are two concepts with which we will be concerned throughout this text. Probability is a mathematical means of studying uncertainty and variability, and statistics is primarily concerned with making inferences based on partial information and past experience.



Readers familiar with the Peanuts comic strip are aware of Lucy's habit of pulling away the football when Charlie Brown is about to kick it. This has happened so often that when you see the first panel, you make an inference about what the final outcome will be—that is, you make an inference about the future based on past experience. (We could speculate about what would happen if Lucy actually let him kick the football, but we will leave such speculations to you, the reader.)

As an illustration of uncertainty, variability, and inference, we might consider a typical intercollegiate football game between, say, the University of Missouri and Kansas University to be played at Columbia, Missouri. On the day before the game, could you give precise answers to any of the following questions? What will the paid attendance be at the game? Will the toss of the coin at the beginning of the game be heads or tails? How many field goals will be attempted? How many hot dogs will be sold? What will be the final score? Which team will win? At what time will the game be over? How many car accidents will there be within the city limits in the hour after the game?

Of course, you can't give precise answers to these questions because of the uncertainty involved. Furthermore, while some of the answers are of interest only to a sports fan, the other answers are important to various officials: the stadium official who must assign ticket takers to various gates, the concessionaire who must order the hot dogs, the police chief who assigns traffic control officers, and so on. How do these officials make their decisions? They must make inferences based on partial information and past experience.

Probability and statistics can be helpful in answering questions like these. You will see from the examples used throughout the text that, in addition to applications to business and economics, probability and statistics have applications to fields such as agronomy, biology, forestry, genetics, medical technology, political science, quality control, social science, and zoology. We should point out that some people try to use statistics dishonestly. Of course, this is no surprise to you. You've often heard phrases such as, "Figures don't lie, but liars figure." Such phrases have arisen because of the way people have misused statistics. We will point out in this text some ways in which people misuse and abuse statistics, not so that you will do the same, but so that you can recognize such abuses and not be misled by them.

1.2 SAMPLE SPACES

What do you think of when someone mentions the word "experiment?" For some people the word conjures up visions of a mad scientist and his helper Igor working in a dimly lit lab surrounded by bubbly test tubes. For others, it brings memories of a high school or college physics or chemistry lab. In any case, we might think of an *experiment* as being a well-defined act or procedure leading to some outcome. Frequently, experiments, under very well-controlled conditions, have the property of reproducibility; that is, the very same sequence of steps done at different points in time lead to the exact same outcome. However, sometimes experimental outcomes differ because of changes in hard-to-control conditions (for instance, gas experiments are sensitive to changes in pressure and temperature), while for other experiments the outcomes differ because of an inherent randomness in the experiment (counting the number of cosmic rays which strike a satellite in a one-minute time interval). We will be primarily interested in experiments whose outcomes differ from one experimental trial to the next due to an inherent randomness. We will refer to such experiments as *random experiments.*

Definition 1.1: The *sample space* S of a random experiment is the set of all possible outcomes of that experiment.

Example 1.1: A referee flips a coin at the start of a football game. Find the sample space for this "experiment."

Since the two possible outcomes are "heads" (H) and "tails" (T), it follows that $S = \{H, T\}$.

Example 1.2: Mrs. Hubbard has 6 children: Alan, Betty, Carl, Diane, Eric, and Frieda, ages 1—6 respectively. As in many households, they frequently argue over whose turn it is to do a chore. Mrs. Hubbard solves such arguments by tossing a die that has a child's name on each side. The child whose name comes up is assigned to do the chore. Find the sample space for the experiment of choosing a child by tossing this die.

The sample space will consist of the children who can be chosen, hence,

$$S = \{\text{Alan, Betty, Carl, Diane, Eric, and Frieda}\}$$

Example 1.3: Amanda is the manager of an appliance store specializing in large appliances. The store's policy is to give away free every thousandth appliance of a given type. At present, 999 freezers have been sold since the last one was given away. If Amanda counts the number of customers purchasing appliances before one buys a freezer, what is the sample space?

A customer making a purchase may be buying a freezer, but of course may be buying another appliance, such as a refrigerator or washer. Any number of customers may purchase other appliances before one comes along wanting to purchase a freezer. Consequently, $S = \{1, 2, 3, \ldots\}$.

In the first two examples, we see that the sample space is finite; that is, there is a natural number that corresponds to the number of points or elements of S. In the first example, there are 2 points in S; in the second example, there are 6. However, in the third example, we see that, theoretically, any number of customers might have to be observed before one wants a freezer, so S must contain all of the integers. If a sample space is such that all elements could be written down in a list, but where the list could never be finished because it is infinitely long, then the number of elements in S is said to be *countably infinite*.

Definition 1.2: If a sample space contains a finite or countably infinite number of elements, it is said to be *countable*. Such sample spaces are also called *discrete*.

Example 1.4: Howard is the advertising manager for a television station which broadcasts college football games. In order to know how many advertising spots can be sold for a game starting at 1:30 P.M., Howard would like to know the time of the final gun. What is the sample space?

Since each quarter of a college football game is 15 minutes, and since the half-time intermission is 20 minutes, the earliest the game could be over is 2:50 P.M. Of course, the time is generally later than that due to timeouts, out-of-bounds plays, incompleted passes, and so on (when the official clock stops). Since the sample space is the set of all possible outcomes, we would have $S = \{t\colon t \geq 2\colon 50 \text{ P.M.}\}$.

If we were to assume further that all games would be automatically ended at 6:00 P.M. due to darkness, then we would have $S = \{t\colon 2\colon 50 \text{ P.M.} \leq t \leq 6\colon 00 \text{ P.M.}\}$.

Whenever we attempt to measure time, we are limited by our measuring devices. In the above example, we might expect the time of the final gun to be reported to the nearest minute or second, in which case we'd have to say that S is discrete. However, ignoring our inability to measure time to any number of decimal places, time is actually a continuous quantity, and, consequently, S consists of all numbers in some interval of the real line.

Definition 1.3: If a sample space contains the set of all numbers in some interval of the real line, that sample space is said to be *continuous*.

1.3 EVENTS

Let's assume that the most recent argument that needs to be resolved in the Hubbard household (Example 1.2) is whose turn it is to feed the dog Fang. Mrs. Hubbard is about to resolve the argument by rolling her special die. In this case, Fang is interested in the outcome of this experiment, since he has noticed that when one of the girls feeds him, he gets more food than when a boy feeds him. Fang is hoping that a girl is chosen. If we let $G = \{\text{Betty, Diane, Frieda}\}$ and $B = \{\text{Alan, Carl, Eric}\}$, then Fang would hope that one of the elements of G rather than of B would be selected.

This is an example of a situation where a person (using the term loosely) is interested in knowing whether a member of a certain subset of S was chosen without being interested in which particular member of that subset was chosen.

Definition 1.4: An *event* is a subset of a sample space S.

Events are generally denoted by capital letters near the beginning of the alphabet. B and G, as defined earlier, are examples of events. Events may be defined by a verbal description or by listing the elements of S that satisfy the description. For example,

$$A = \{\text{child chosen is at least 4 years old}\} = \{\text{Diane, Eric, Frieda}\}$$

Many other events could be defined using this sample space (some of which are given in Exercise 8). In fact, since there are 6 elements of S, there are $2^6 = 64$ different events that can be defined. (We will show in Section 1.4 that if a sample space contains n distinct points, then the number of events that can be defined using that sample space is 2^n.)

Example 1.5: The Sirloin Steakhouse has 30 tables. If the number of tables occupied at 5:00 P.M. on Friday night is to be counted, then

$$S = \{0, 1, 2, \ldots, 30\}$$

What do events A and B represent if $A = \{0, 1, 2, \ldots, 10\}$ and $B = \{11, 12, 13, \ldots, 30\}$?

In words, A represents the event that 10 or fewer tables are occupied, while B represents the event that more than 10 tables are occupied.

Example 1.6: In Example 1.4, we described a continuous sample space based on the time of the final gun of a football game. If we assume that the game will end automatically at 6:00 P.M. due to darkness, then

$$S = \{t\colon 2\colon 50 \text{ P.M.} \leq t \leq 6\colon 00 \text{ P.M.}\}$$

The following would each correspond to subsets of S:

$A = \{\text{game is over after } 4\colon 00\}$
$B = \{\text{game is over no later than } 4\colon 00\}$
$C = \{\text{game is over between 3 and 4 o'clock}\}$
$D = \{\text{game is over by } 2\colon 30\}$

You will recall from your study of sets that a set that has no members (such as, the set of 4-footed ducks) is said to be empty. We will denote the empty set by the Greek letter phi (ϕ). The event D defined in Example 1.6 is actually the empty set. Why? Since S consists of times between 2:50 and 6:00, it is impossible for the game to be over by 2:30; hence, D contains no points of S and, consequently, must be empty.

In Examples 1.5 and 1.6, you can see that it is impossible for events A and B to occur simultaneously. In set terminology, the simultaneous occurrence of two sets is called the *intersection* of those sets. Since events are merely subsets of some sample space, it makes sense to talk about the intersection of two events, denoted by $\cap$. In set terminology, if two sets A and B have an empty intersection, they are said to be *disjoint*. When the sets happen to be events, a different term is used to convey the same notion.

Definition 1.5: Two events, A and B, are said to be *mutually exclusive* if $A \cap B = \phi$.

You must be careful here not to be confused by the notation. As you can see, the letters A and B denote quite different events in the different examples. In both Examples 1.5 and 1.6, the events A and B defined are mutually exclusive. However, in the illustration of Mrs. Hubbard's children given earlier in this section, we defined

$$A = \{\text{Diane, Eric, Frieda}\} \quad \text{and} \quad B = \{\text{Alan, Carl, Eric}\}$$

so $A \cap B = \{\text{Eric}\} \neq \phi$. In this case, A and B are *not* mutually exclusive.

We also wish to emphasize that the concept, mutually exclusive, relates to pairs of events. It is easy to define three different events, say A, B, and C, such that $A \cap B \cap C = \phi$, while $A \cap B \neq \phi$, $A \cap C \neq \phi$, and $B \cap C \neq \phi$. It would not be correct to say that A, B, and C are mutually exclusive. If we say that events $A_1, A_2, \ldots, A_n$ are pairwise mutually exclusive, we will mean $A_i \cap A_j = \phi$, $i \neq j$.

Since events are sets, we can consider other set operations in addition to intersections. For example, we can consider unions and complements of sets. In this book, we will use a "bar" to denote the complement of a set $\bar{A}$. (Other books may use A', $\tilde{A}$, A^c, and so on, to denote complements.)

Example 1.7: A gas station operator Tex has eight gas pumps at his station. Tex periodically counts the number of cars using the gas pumps. For this experiment, $S = \{0, 1, 2, \ldots, 8\}$. Defining

$$A = \{\text{an odd number of pumps are in use}\} = \{1, 3, 5, 7\}$$

$$B = \{\text{at least half of the pumps are in use}\} = \{4, 5, 6, 7, 8\}$$

$$C = \{\text{at most half of the pumps are in use}\} = \{0, 1, 2, 3, 4\}$$

find the events $\bar{A}$, $B \cup C$, and $B \cap C$.

Using the definitions of complements, unions, and intersections, we find

$$\bar{A} = \{0, 2, 4, 6, 8\}$$

$$B \cup C = \{0, 1, 2, \ldots, 8\} = S$$

$$B \cap C = \{4\}$$

In some situations, it is helpful to have a diagram to aid in finding the elements of a set formed by unions, intersections, or complements. One such diagram is called a *Venn diagram*. In these diagrams, the sample space S is represented by a rectangle, and events are represented by portions of S, usually, but not necessarily, circles. If events are known to be mutually exclusive, the corresponding circles are shown as nonoverlapping. Figure 1.1 gives a Venn diagram where A and C are mutually exclusive, but A and B are not mutually exclusive, nor are B and C. In Figure 1.2, the shaded regions correspond to $\bar{A}$, $A \cap B$, and $A \cup B$ respectively.

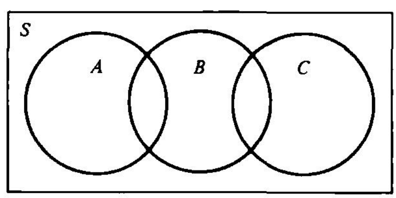


FIGURE 1.1
A Venn Diagram

FIGURE 1.2
Venn Diagrams Showing Complements, Intersections, and Unions