******Special Price******
Non-Returnable
No Refund/ No Return

6 01 08
USED BOOK
$ 68.25

CD-ROM Included

SIXTH EDITION

PROBABILITY AND STATISTICAL INFERENCE

HOGG AND TANIS

W / 1 CD

Sixth Edition

Probability and Statistical Inference

Robert V. Hogg
University of Iowa

Elliot A. Tanis
Hope College

Prentice Hall

Prentice Hall, Upper Saddle River, NJ 07458

Library of Congress Cataloging-in-Publication Data

Hogg Robert V.
Probability and statistical inference / Robert V. Hogg,
Elliot A. Tanis— 6th ed.
p. cm.
ISBN 0–13–027294–9
1. Probabilities. 2. Mathematical statistics. I. Tanis, Elliot A.
II. Title.

QA273.H694 2001 00-025352
519.2–dc21 CIP

Executive Project Manager: *Ann Heath*
Acquisitions Editor: *Kathleen Boothby Sestak*
Editor in Chief: *Sally Yagan*
Production Editor: *Lynn Savino Wendel*
Assistant Vice President of Production and Manufacturing: *David W. Riccardi*
Executive Managing Editor: *Kathleen Schiaparelli*
Senior Managing Editor: *Linda Mihatov Behrens*
Manufacturing Buyer: *Alan Fischer*
Manufacturing Manager: *Trudy Pisciotti*
Director of Marketing: *John Tweeddale*
Marketing Manager: *Angela Battle*
Marketing Assistant: *Vince Jansen*
Art Director: *Jayne Conte*
Editorial Assistant: *Joanne Wendelken*
Cover Designer: *Joseph Sengotta*
Composition: *PreTEX, Inc.*

Prentice Hall

© 2001, 1997 by Prentice-Hall, Inc.
Upper Saddle River, New Jersey 07458

Previous editions copyright 1993, 1988, 1983, 1977 by Macmillan Publishing Company, a division of Macmillan, Inc.

All rights reserved. No part of this book may be reproduced, in any form or by any means, without permission in writing from the publisher.

Maple and Maple V are registered trademarks of Waterloo Maple Inc.

Printed in the United States of America

10 9 8 7 6 5 4 3

ISBN 0-13-027294-9

Prentice-Hall International (UK) Limited, *London*
Prentice-Hall of Australia Pty. Limited, *Sydney*
Prentice-Hall Canada Inc., *Toronto*
Prentice-Hall Hispanoamericana, S.A., *Mexico*
Prentice-Hall of India Private Limited, *New Delhi*
Prentice-Hall of Japan, Inc., *Tokyo*
Pearson Education Asia Pte. Ltd.
Editora Prentice-Hall do Brasil, Ltda., *Rio de Janeiro*

Preface

We are pleased with the reception that was given to the first five editions of *Probability and Statistical Inference*. The sixth edition is still designed for use in a course having from three to six semester hours of credit. No previous study of statistics is assumed, and a standard course in calculus provides an adequate mathematical background. Certain sections have been starred and are not needed in subsequent sections. This, however, does not mean that these starred sections are unimportant, and we hope many of you will study them.

We still view this book as the basis of a junior or senior level course in the mathematics of probability and statistics that is taught by many departments of mathematics and/or statistics. We have tried to make it more "user friendly"; yet we do want to reinforce certain basic concepts of mathematics, particularly calculus. To help the student with methods of algebra of sets and calculus, we include a *Review of Selected Mathematical Techniques* in Appendix A. This review includes a method that makes integration by parts easier. Also we derive the important *Rule of 72* that provides an approximation to the number of years necessary for money to double.

MAJOR CHANGES IN THIS EDITION

Chapter 1 still provides an excellent introduction to good descriptive statistics and exploratory data analysis and the corresponding empirical distributions. However, probability models are also introduced in Chapter 1 so that the student recognizes from the beginning that the characteristics of the empirical distributions are estimates of those of probability distributions. Hopefully, this creates some interest among students in checking to see if a probability model is appropriate for the situation under consideration throughout the text.

Chapters 2–4 provide concepts in probability and basic distributions. These have been simplified somewhat from the previous edition by introducing a few of the easiest distributions through examples. Also the probability generating function has been dropped in this edition, although we note that the moment-generating function can serve in that capacity. Of course, the latter can also be used to compute the moments of a distribution.

By request of many statisticians, we have introduced multivariate distributions much earlier, but in such a way that only the first section of that chapter is needed for the chapters that follow on techniques used in statistical inference. Hence, if an instructor so desires, he or she can start statistical methods sooner without conditional distributions.

While this book is written primarily as a mathematical introduction to probability and statistics, there are a great many examples and exercises concerned with applications. For illustrations, the reader will find applications in the areas of biology, education, economics, engineering, environmental studies, exercise science, health science, manufacturing, opinion polls, psychology, sociology, and sports. That is, there are many exercises in the text, some illustrating the mathematics of probability and statistics but a great number are concerned with applications. We are certainly more concerned with model checking in this edition than in the previous editions.

In addition, there is major effort to emphasize confidence intervals more than previously, and we clearly spell out the relationship between confidence intervals and tests of hypotheses. In that regard, we have increased the emphasis on one-sided confidence intervals somewhat because often a practitioner wants a lower or upper bound for the parameter in question, and these have a natural relationship with one-sided tests of hypotheses.

The chapter on confidence intervals is so organized that the instructor can introduce early basic concepts of regression and distribution-free techniques, if he or she chooses to do so. That is, the separate chapter on nonparametric methods has been dropped, and those basic concepts are included among the other methods of statistical inference. After the considerations of the first few basic sections in the chapters on estimation and tests, the instructor may select among some of the later sections in each of those chapters. It is also noted that, at the suggestion of two of the reviewers, Section 8.11 on resampling methods is new to this edition, and we hope some instructors will choose to consider this new technique of statistical inference.

While the text is written primarily for a two-semester sequence of a probability course (selections from Chapters 1–6) followed by a statistical inference course (selections from Chapters 7–10), it can be used for a 4-credit hour semester course for those students with a good calculus background. In such a course, Chapters 1–4 (except the starred sections) are considered followed by Section 5.1 and most of Chapter 6; these make up about 70 percent of the course. The remaining 30 percent consists of selections from Chapters 7 and 8, unless the instructor wants to consider some of the theory given in the early part of Chapter 9. Those interested in the statistical methods used in quality improvement would want to consider at least part of Chapter 10.

Different from most textbooks, we have included a prologue, a centerpiece, and an epilogue. The main emphasis in these is that variation occurs in almost every process, and the study of probability and statistics helps us understand this variability. Accordingly, the study of statistics is extremely useful in many fields of endeavor and can lead students to interesting positions in the future.

试读结束：需要全本请在线购买： www.ertongbook.com

FEATURES

Throughout the book, figures and real applications will help the student understand statistics and what statistical methods can accomplish. For some exercises, it is assumed that calculators or computers are available; thus the solutions will not always involve "nice" numbers. Solutions using a computer are given if a complicated data set is involved. The data sets for all of the exercises are available on a data disk and on the Web.

ANCILLARIES

A **Solutions Manual** containing worked-out solutions to the even-numbered exercises in the text is available to instructors from the publisher. Many of the numerical exercises were solved using *Maple*. For additional exercises that involve simulations, a separate manual *Probability & Statistics: Explorations with MAPLE*, second edition, by Zaven Karian and Elliot Tanis is available for purchase. Several exercises in that manual also make use of the power of *Maple* as a computer algebra system.

A **CD-ROM** is new for this edition. There are two folders on the CD. The folder titled "Data" includes the data files for most of the exercises and examples in the text. Each chapter for which data files are given has it's own folder and most include separate folders for pertinent sections. The files are saved in a number of different formats to provide maximum flexibility to the user. The folder labeled figures contains two types of figures. Several of the figures in the text are stored as Minitab projects (`.mpj`) within chapter folders. They can be opened from Minitab or it may be possible to double click on one of these files and Minitab will open the selected file. The other figures included in this folder were generated using *Maple*. These include animated versions of several figures in the text that allow the user to view the figures in a dynamic environment that will help clarify and reinforce the concepts being presented. To view the animated figures, as well as others that were generated using *Maple*, the user must have access to a web browser such as Microsoft Explorer or Netscape Navigator. If you have a web browser, the most efficient approach to viewing the figures is to open the Figures Folder and double click on the `directory.htm` file. This file should open in your web browser and present you with a list of figures from which to select.

A **web page** for this text is available at `www.prenhall.com/hogg`. Check out this web page to see what is available.

ACKNOWLEDGMENTS

We are indebted to the *Biometrika* Trustees for permission to include Tables IV and VII, which are abridgements and adaptations of tables published in *Biometrika Tables for Statisticians*. We are also grateful to the Literary Executor of the late Sir Ronald A. Fisher, F.R.S., to Dr. Frank Yates, F.R.S., and to Longman Group, Ltd, London, for permission to use Table III from their book *Statistical Tables for Biological, Agricultural, and Medical Research* (6th ed., 1974), reproduced as our Table VI.

We wish to thank our colleagues, students, and friends for many suggestions and for their generosity in supplying data for exercises and examples. In particular, we would like to acknowledge the helpful reviews of Bhaskar Bhattacharya, YongHee Kim, Ditlev Monrad, and A.G. Warrack. We thank Priscilla Gathoni for checking our answers. Moreover, a number of the faculty members at the University of Iowa made very useful suggestions; these include: Tim Robertson, Joe Lang, Ralph Russo, Steve Hillis, Dick Dykstra, and Jon Cryer. Cathy Mader from Hope College was extremely generous with providing help with LaTeX commands. We also thank our students, Paul Scheet and Eric Goodman. We acknowledge the Curtis A. Jacobs Memorial Fund of Hope College for providing financial support for Eric. We also thank the University of Iowa and Hope College for providing time and encouragement. Finally, our families, through six editions, have been most understanding during the preparation of all of this material; we truly appreciate their patience and needed their love.

R.V.H.
E.A.T.

Prologue

The discipline of statistics deals with the collection and the analysis of data. The advances in computing technology, particularly in relationship to changes in science and business, have increased the need for more statistical scientists to examine the huge amount of data being collected. We know that data are not equivalent to information. Once data, hopefully of high quality, are collected, there is an almost unlimited need for statisticians to make sense of these data. That is, data must be analyzed, providing information upon which decisions can be made. In light of this great demand, opportunities for the discipline of statistics have never been greater, and there is a special need for more, bright young persons going into statistical science.

If we think of fields in which data play a major part, the list is almost endless: accounting, actuarial science, atmospheric science, biological science, economics, educational measurement, environmental science, epidemiology, finance, genetics, manufacturing, marketing, medicine, pharmaceutical industries, psychology, sociology, sports, and on and on. Due to all of these areas in which statistics is useful, it really should be taught as an applied science. Nevertheless, to go very far in such an applied science, it is necessary to understand the importance of creating models for each situation under study. Now no model is ever exactly right, but some are extremely useful as an approximation to the real situation. Most appropriate models in statistics require a certain mathematical background in probability. Accordingly, this textbook, while alluding to applications in the examples and the exercises, is really about the mathematics needed for the appreciation of probabilistic models necessary for statistical inferences.

In a sense, statistical techniques are really the heart of the scientific method. Observations are made that suggest conjectures. These conjectures are tested and data collected and analyzed, providing information about the truth of the conjectures. Sometimes these are supported by the data, but often the conjectures need to be modified and more data collected to test the modifications and so on. Clearly, in this iterative process, statistics plays a major role with its emphasis on proper design and analysis of experiments and the resulting inferences upon which decisions can be made. Through statistics, information is provided for taking certain actions: for illustrations, improving manufactured products, providing better services, marketing new products or services, forecasting energy needs, classifying diseases better, and so on.

Statisticians recognize that there are often small errors in their inferences, and they attempt to make the probabilities of those mistakes as small as possible. The reason that these uncertainties even exist is due to the fact that there is variation in the data. Even though experiments are repeated under seemingly the same conditions, the results vary from trial to trial. We try to improve the quality of the data by making them as reliable as possible, but the data simply do not fall on given patterns. There is variation in almost all processes. In light of this uncertainty, the statistician tries to determine the pattern in the best possible way, but always explaining the error structures of the statistical estimates.

This is an important lesson to be learned: Variation is almost everywhere. It is the statistician's job to understand the variation. Often, like in manufacturing, the desire is to reduce variation because the products will be more consistent. In other words, the "car doors will fit better" in the manufacturing of automobiles if the variation is decreased by making each door closer to its target values.

Many statisticians in industry have stressed the need of "statistical thinking" in everyday operations. It is based upon three points, two of which have been mentioned in the preceding paragraph: (1) Variation exists in all processes; (2) Understanding and reducing undesirable variation is a key to success; and (3) All work occurs in a system of interconnected processes. W. Edwards Deming, an esteemed statistician and quality improvement guru, stressed these three points, particularly the third one. He would carefully note that you could not maximize the total operation by maximizing the individual components unless they are independent of each other. However, in most instances, they are highly dependent; and persons in different departments must work together in creating the best products and services. If not, what one unit does to better itself could very well hurt others. He often sighted an orchestra as an illustration of the need for the members to work together to create an outcome that is consistent and desirable.

Any student of statistics should understand the nature of variability and the necessity of creating probabilistic models for that variability. We can not avoid making inferences and decisions in the face of this uncertainty; however, these results are greatly influenced by the probabilistic models selected. Some persons are better model builders than others and accordingly will make better inferences and decisions. The assumptions needed for each statistical model are carefully examined and hopefully the reader will become a better model builder.

Finally, we must mention how dependent modern statistical analyses are upon the computer. The statisticians and computer scientists really should work together in areas of exploratory data analysis and "data mining." Statistical software development is critical today, for the best of it is needed in complicated data analyses. In light of this growing relationship between these two fields, it is good advice for bright students to take substantial offerings in statistics and in computer science.

Students majoring in statistics, computer science, or in some joint program are in great demand in the workplace or for future graduate studies. Clearly, they can earn advanced degrees in statistics or computer science or both. But, more important, they are highly desirable candidates for graduate work in other areas: actuarial science, industrial engineering, finance, marketing, accounting, management science, psychology, economics, law, sociology, medicine, health sciences, and so on. So many of these fields have been "mathematized" to the degree that those programs are begging for majors in statistics and/or computer science. Often these students become "stars" in these other areas. We truly hope that we can interest students enough so that they want to study more statistics. If they do, the opportunities for extremely successful careers are almost endless.

Contents

试读结束：需要全本请在线购买： www.ertongbook.com

CHAPTER 1

Empirical and Probability Distributions

1.1 BASIC CONCEPTS

It is usually difficult to explain to the general public what statisticians do. Many think of us as "math nerds" who seem to enjoy dealing with numbers. And there is some truth to that concept. But if we consider the bigger picture, many recognize that statisticians can be extremely helpful in many investigations.

Consider the following.

1. There is some problem or situation that needs to be considered; so statisticians are often asked to work with investigators or research scientists.
2. Suppose that some measure (or measures) are needed to help us understand the situation better. The measurement problem is often extremely difficult, and creating good measures is a valuable skill. For illustration, in higher education, how do we measure good teaching? This is a question to which we have not found a satisfactory answer, although several measures, such as student evaluations, have been used in the past.
3. After the measuring instrument has been developed, we must collect data through observation, possibly the results of a survey or an experiment.
4. Using these data, statisticians summarize the results, often using descriptive statistics and graphical methods.
5. These summaries are then used to analyze the situation. Here it is possible that statisticians make what are called statistical inferences.
6. Finally a report is presented, along with some recommendations that are based upon the data and the analysis of them. Frequently such a recommendation might be to perform the survey or experiment again, possibly changing some of the questions or factors involved. This is how statistics is used in what is referred to as the scientific method, because often the analysis of the data suggests other experiments. Accordingly, the scientist must consider different possibilities in his or her search for an answer and thus performs similar experiments over and over again.

So the discipline of statistics deals with the *collection* and *analysis of data*. When measurements are taken, even seemingly under the same conditions, the results usually vary. Despite this variability, a statistician tries to find a pattern; yet due to the "noise," not all of the points lie on the pattern. In the face of this variability, he or she must still determine the best way to describe the pattern. Accordingly, statisticians know that mistakes will be made in data analysis, and they try to minimize those errors as much as possible and then give bounds on the possible errors. By considering these bounds, decision makers can decide how much confidence they want to place on these data and the analysis of them. If the bounds are wide, perhaps more data should be collected. If they are small, however, the person involved in the study might want to make a decision and proceed accordingly.

Variability is a fact of life, and proper statistical methods can help us understand data collected under inherent variability. Because of this variability, many decisions have to be made that involve uncertainties. In medical research, interest may center on the effectiveness of a new vaccine for mumps; an agronomist must decide if an increase in yield can be attributed to a new strain of wheat; a meteorologist is interested in predicting the probability of rain; the state legislature must decide whether increasing speed limits will result in more accidents; the admissions officer of a college must predict the college performance of an incoming freshman; a biologist is interested in estimating the clutch size for a particular type of bird; an economist desires to estimate the unemployment rate; an environmentalist tests whether new controls have resulted in a reduction in pollution.

In reviewing the preceding, relatively short list of possible areas of applications of statistics, the reader should recognize that good statistics is closely associated with careful thinking in many investigations. For illustration, students should appreciate how statistics is used in the endless cycle of the scientific method. We observe nature and ask questions; we run experiments and collect data that shed light on these questions; we analyze the data and compare the results of the analysis to what we previously thought; we raise new questions; and on and on. Or if you like, statistics is clearly part of the important "plan—do—study—act" cycle: Questions are raised and investigations planned and carried out. The resulting data are studied and analyzed and then acted upon, often raising new questions.

There are many aspects of statistics. Some people get interested in the subject by collecting data and trying to make sense out of these observations. In some cases the answers are obvious and little training in statistical methods is necessary. But if a person goes very far in many investigations, he or she soon realizes that there is a need for some theory to help describe the error structure associated with the various estimates of the patterns. That is, at some point, appropriate probability and mathematical models are required to make sense out of complicated data sets. Statistics and the probabilistic foundation on which statistical methods are based can provide the models to help people make decisions such as these. So in this book, we are more concerned about the mathematical, rather than the applied, aspects of statistics, although we give enough real examples so that the reader can get a good sense of a number of important applications of statistical methods.

In the study of statistics we consider experiments for which the outcome cannot be predicted with certainty. Such experiments are called **random experiments**. Each experiment ends in an outcome that cannot be determined with certainty before the experiment is performed. However, the experiment is such that the collection of every possible outcome can be described and perhaps listed. This collection of all outcomes is called the **outcome space**, the **sample space**, or, more simply, the **space** S. The following examples will help illustrate what we mean by random experiments, outcomes, and their associated spaces.

Example 1.1-1 Two dice are cast and the total number of spots on the sides that are "up" are counted. The outcome space is $S = \{2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12\}$. ▲

Example 1.1-2 Each of six students selects an integer at random from the first 52 positive integers. We are interested in whether at least two of these six integers match (M) or whether all are different (D). Thus $S = \{M, D\}$. ▲

Example 1.1-3 A fair coin is flipped successively at random until the first head is observed. If we let x denote the number of flips of the coin that are required, then $S = \{x : x = 1, 2, 3, 4, \ldots\}$, which consists of an infinite, but countable, number of outcomes. ▲

Example 1.1-4 A box of breakfast cereal contains one of four different prizes. The purchase of one box of cereal yields one of the prizes as the outcome, and the sample space is the set of four different prizes. ▲

Example 1.1-5 In Example 1.1-4, assume that the prizes are put into the boxes randomly. A family continues to buy this cereal until they obtain a complete set of the four different prizes. The number of boxes of cereal that must be purchased is one of the outcomes in $S = \{b : b = 4, 5, 6, \ldots\}$. ▲

Example 1.1-6 A fair coin is flipped successively at random until heads is observed on two successive flips. If we let y denote the number of flips of the coin that are required, then $S = \{y : y = 2, 3, 4, \ldots\}$. ▲

Example 1.1-7 To determine the percentage of body fat for a person, one measurement that is made is a person's weight under water. If w denotes this weight in kilograms, then the sample space could be $S = \{w : 0 < w < 7\}$, as we know from past experience that this weight does not exceed 7 kilograms. ▲

Example 1.1-8 An ornithologist is interested in the clutch size (number of eggs in a nest) for gallinules, a bird that lives in a marsh. If we let c equal the clutch size then a possible sample space would be $S = \{c : c = 0, 1, 2, \ldots, 15\}$, as 15 is the largest known clutch size. ▲

Note that the outcomes of a random experiment can be numerical, as in Examples 1.1-3, 1.1-6, 1.1-7, and 1.1-8, but they do not have to be, as shown by Examples 1.1-2 and 1.1-4. Often we "mathematize" those latter outcomes by assigning numbers to them. For instance, in Example 1.1-2, we could denote the outcome $\{D\}$ by the number zero and the outcome $\{M\}$ by the number one. In general, measurements on outcomes associated with random experiments are called **random variables** and these are usually denoted by some capital letter towards the end of the alphabet, like X, Y, and Z.

Note the numbers of outcomes in the sample spaces in these examples. In Examples 1.1-1, 1.1-4 and 1.1-8, the number of outcomes is finite. In Examples 1.1-3 and 1.1-6 the number of possible outcomes is infinite but countable. That is, there are as many outcomes as there are counting numbers (positive integers). The space for Example 1.1-7 is different from the other examples in that the set of possible outcomes is an interval of numbers. Theoretically, the weight could be any one of an infinite number of possible weights; here the number of possible outcomes is not countable. However, from a practical point of view, reported weights are selected from a finite number of possibilities because we can read and record the answer to only an accuracy determined by our scale. Many times, however, it is better to conceptualize the space as an interval of outcomes, and Example 1.1-7 is an example of a space of the continuous type.

If we consider a random experiment and its space, we note that under repeated performances of the experiment, some outcomes occur more frequently than others. For illustration, in Example 1.1-3, if this coin-flipping experiment is repeated over and over, the first head is observed on the first flip more often than on the second flip. If we can somehow, by theory or observations, determine the fractions of times a random experiment ends in the respective outcomes, we have described a *distribution of the random variable* (sometimes called a *population*). Often we cannot determine this distribution through theoretical reasoning but must actually perform the random experiment a number of times to obtain guesses or *estimates* of these fractions. The collection of the observations that are obtained from these repeated trials is often called a *sample*. The making of a conjecture about a distribution of a random variable based on the sample is called a *statistical inference*. That is, in statistics, we try to argue from the sample to the population. To understand the background behind statistical inferences that are made from the sample, we need a knowledge of some probability, basic distributions, and sampling distribution theory; these topics are considered in the early part of this book.

Given a sample or set of measurements, we would like to determine methods for describing the data. Suppose that we have some **counting** or **discrete data**. For example, you record the number of children in the family of each of your classmates. Or perhaps your state, on a regular basis, selects five integers out of the first 39 integers for the state lottery. You could count the number of those five integers that are odd.

In general, suppose we repeat a random experiment a number of times, say n. If a certain outcome has occurred f times in these n trials; the number f is called the **frequency** of the outcome. The ratio f/n is called the **relative frequency** of the outcome. A relative frequency is usually very unstable for small values of n, but it tends to stabilize about some number, say p, as n increases. The number p is called the **probability** of the outcome.

To develop an understanding of a particular set of discrete data, we can summarize the data in a frequency table and then construct a histogram of the data. A **frequency table** provides the number of occurrences of each possible outcome. A **histogram** presents graphically the tallied data as illustrated in the following example.

试读结束：需要全本请在线购买： www.ertongbook.com