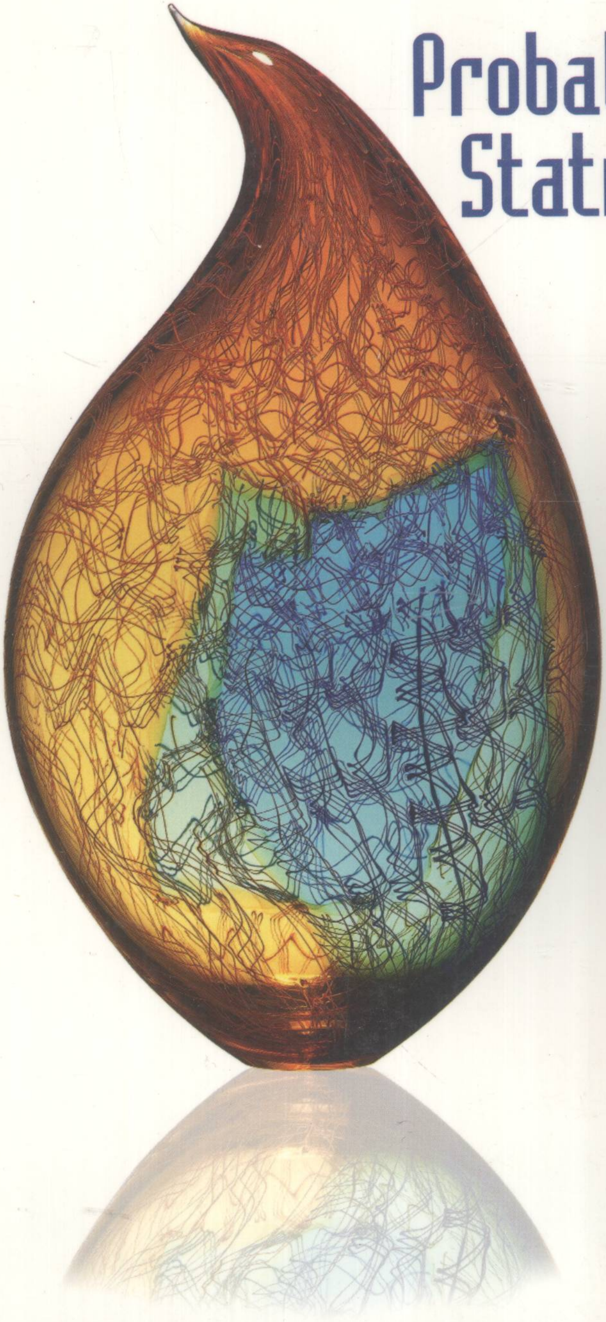
INTERNATIONAL EDITION

CD-ROM INCLUDED



Probability and Statistical Inference

7e

Hogg • Tanis

PROBABILITY AND STATISTICAL INFERENCE

SEVENTH EDITION

Robert V. Hogg
University of Iowa

Elliot A. Tanis
Hope College

PEARSON
Prentice
Hall

PEARSON EDUCATION INTERNATIONAL

Executive Acquisitions Editor: *George Lobell*
Editor-in-Chief: *Sally Yagan*
Production Editor: *Jeanne Audino/Bayani Mendoza de Leon*
Assistant Managing Editor: *Bayani Mendoza de Leon*
Senior Managing Editor: *Linda Mihatov Behrens*
Executive Managing Editor: *Kathleen Schiaparelli*
Assistant Manufacturing Manager/Buyer: *Michael Bell*
Marketing Manager: *Halee Dinsey*
Marketing Assistant: *Joon Moon*
Managing Editor, Digital Supplements: *Nicole M. Jackson*
Media Production Editor: *Donna Crilly*
Director of Creative Services: *Paul Belfanti*
Art Director: *Geoffrey Cassar*
Interior Designer: *Kristine Carney*
Cover Designer: *DaSa Design*
Art Editor: *Thomas Benfatti*
Editorial Assistant: *Jennifer Urban*
Manager, Cover Visual Research & Permissions: *Karen Sanatar*
Cover Image: *Lino Tagliapietra, Silea, 2000. Hand-blown colored glass with inserts of multiple colored filigree zanirico canes, h. 50 × 28 × 18 cm.*
Photo: Russell Johnson.

Printed in the United States of America

10 9 8 7 6 5 4 3 2 1

ISBN: 0-13-129382-6

Pearson Education, Ltd., *London*
Pearson Education Australia PTY. Limited, *Sydney*
Pearson Education Singapore, Pte., Ltd
Pearson Education North Asia Ltd, *Hong Kong*
Pearson Education Canada, Ltd., *Toronto*
Pearson Education de Mexico, S.A. de C.V.
Pearson Education—Japan, *Tokyo*
Pearson Education Malaysia, Pte. Ltd
Pearson Education, Upper Saddle River, *New Jersey*

Discrete Distributions

Bernoulli

$0 < p < 1$

$$f(x) = p^x(1-p)^{1-x}, \qquad x = 0, 1$$

$$M(t) = 1 - p + pe^t$$

$$\mu = p, \qquad \sigma^2 = p(1-p)$$

Binomial

$b(n, p)$

$0 < p < 1$

$$f(x) = \frac{n!}{x!(n-x)!}\, p^x(1-p)^{n-x}, \qquad x = 0, 1, 2, \ldots, n$$

$$M(t) = (1 - p + pe^t)^n$$

$$\mu = np, \qquad \sigma^2 = np(1-p)$$

Geometric

$0 < p < 1$

$$f(x) = (1-p)^{x-1}p, \qquad x = 1, 2, 3, \ldots$$

$$M(t) = \frac{pe^t}{1-(1-p)e^t}, \qquad t < -\ln(1-p)$$

$$\mu = \frac{1}{p}, \qquad \sigma^2 = \frac{1-p}{p^2}$$

Hypergeometric

$N_1 > 0, \quad N_2 > 0$

$N = N_1 + N_2$

$$f(x) = \frac{\binom{N_1}{x}\binom{N_2}{n-x}}{\binom{N}{n}}, \qquad x \le n, x \le N_1, n - x \le N_2$$

$$\mu = n\left(\frac{N_1}{N}\right), \qquad \sigma^2 = n\left(\frac{N_1}{N}\right)\left(\frac{N_2}{N}\right)\left(\frac{N-n}{N-1}\right)$$

Negative Binomial

$0 < p < 1$

$r = 1, 2, 3, \ldots$

$$f(x) = \binom{x-1}{r-1} p^r(1-p)^{x-r}, \qquad x = r, r+1, r+2, \ldots$$

$$M(t) = \frac{(pe^t)^r}{[1-(1-p)e^t]^r}, \qquad t < -\ln(1-p)$$

$$\mu = r\left(\frac{1}{p}\right), \qquad \sigma^2 = \frac{r(1-p)}{p^2}$$

Poisson

$0 < \lambda$

$$f(x) = \frac{\lambda^x e^{-\lambda}}{x!}, \qquad x = 0, 1, 2, \ldots$$

$$M(t) = e^{\lambda(e^t - 1)}$$

$$\mu = \lambda, \qquad \sigma^2 = \lambda$$

Uniform

$m > 0$

$$f(x) = \frac{1}{m}, \qquad x = 1, 2, \ldots, m$$

$$\mu = \frac{m+1}{2}, \qquad \sigma^2 = \frac{m^2-1}{12}$$

Continuous Distributions

Beta
$0 < \alpha$
$0 < \beta$

$$f(x) = \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} x^{\alpha-1}(1-x)^{\beta-1}, \qquad 0 < x < 1$$

$$\mu = \frac{\alpha}{\alpha+\beta}, \qquad \sigma^2 = \frac{\alpha\beta}{(\alpha+\beta+1)(\alpha+\beta)^2}$$

Chi-square
$\chi^2(r)$
$r = 1, 2, \ldots$

$$f(x) = \frac{1}{\Gamma(r/2)2^{r/2}} x^{r/2-1}e^{-x/2}, \qquad 0 \le x < \infty$$

$$M(t) = \frac{1}{(1-2t)^{r/2}}, \qquad t < \frac{1}{2}$$

$$\mu = r, \qquad \sigma^2 = 2r$$

Exponential
$0 < \theta$

$$f(x) = \frac{1}{\theta} e^{-x/\theta}, \qquad 0 \le x < \infty$$

$$M(t) = \frac{1}{1-\theta t}, \qquad t < \frac{1}{\theta}$$

$$\mu = \theta, \qquad \sigma^2 = \theta^2$$

Gamma
$0 < \alpha$
$0 < \theta$

$$f(x) = \frac{1}{\Gamma(\alpha)\theta^\alpha} x^{\alpha-1}e^{-x/\theta}, \qquad 0 \le x < \infty$$

$$M(t) = \frac{1}{(1-\theta t)^\alpha}, \qquad t < \frac{1}{\theta}$$

$$\mu = \alpha\theta, \qquad \sigma^2 = \alpha\theta^2$$

Normal
$N(\mu, \sigma^2)$
$-\infty < \mu < \infty$
$0 < \sigma$

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-(x-\mu)^2/2\sigma^2}, \qquad -\infty < x < \infty$$

$$M(t) = e^{\mu t + \sigma^2 t^2/2}$$

$$E(X) = \mu, \qquad \mathrm{Var}(X) = \sigma^2$$

Uniform
$U(a, b)$
$-\infty < a < b < \infty$

$$f(x) = \frac{1}{b-a}, \qquad a \le x \le b$$

$$M(t) = \frac{e^{tb}-e^{ta}}{t(b-a)}, \qquad t \ne 0; \qquad M(0) = 1$$

$$\mu = \frac{a+b}{2}, \qquad \sigma^2 = \frac{(b-a)^2}{12}$$

RELATED TITLES OF INTEREST FROM PRENTICE HALL

Mathematical Statistics, Vol. I
Peter Bickel and Kjell Doksum
0-13-850363-X

John E. Freund's Mathematical Statistics with Applications 7e
Irwin Miller and Maryless Miller
0-13-142706-7

An Introduction to Mathematical Statistics and Its Applications 4e
Richard Larsen and Morris Marx
0-13-186793-8

Statistics and Data Analysis
Ajit Tamhane and Dorothy Dunlop
0-13-744426-5

Introduction to Mathematical Statistics, 6e
Hogg/Craig/McKean
0-13-008507-3

Miller and Freund's Probability and Statistics for Engineers 7e
Richard Johnson
0-13-143745-3

Probability and Statistics for Engineers and Scientists 7e
Ron Walpole, Ray Myers, Sharon Myers, and Keying Ye
0-13-041529-4

Applied Statistics for Engineers and Scientists
Joe Petruccelli, B. Nandram, and M. Chen
0-13-565953-1

Applied Statistics for Engineers and Scientists: Using Excel and Minitab
David Levine, Patricia Ramsey, and Robert Smidt
0-13-488801-4

Time Series Analysis: Forecasting & Control 3e
George Box, Gwilyn Jenkins, and Greg Reinsel
0-13-060774-6

Contact **www.prenhall.com** to order any of these titles

PREFACE

We are pleased with the reception that was given to the first six editions of *Probability and Statistical Inference*. The seventh edition is still designed for use in a course having from three to six semester hours of credit. No previous study of probability or statistics is assumed, and a standard course in calculus provides an adequate mathematical background. Certain sections have been starred and they 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. In particular, the first five chapters provide a substantial one-semester course in probability and probability distributions. While many statisticians are teaching a course at this level by minimizing probability and concentrating on statistics, we have found that those studying statistics, actuarial science, electrical engineering, economics, finance, genetics, and so on need probability as much as statistics. Thus we choose to place about equal emphasis on these two topics. Chapters 6–10 consist of the second semester of a two-semester sequence as they cover topics in statistics and statistical inference. We have discovered that a fairly good four-semester-hour course can be constructed by an instructor by selecting topics from the first five chapters and Chapters 6 and 8.

We have tried to make the seventh edition 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*, which provides an approximation to the number of years necessary for money to double.

ENHANCEMENTS IN THIS EDITION

- There is better and more logical organization, resulting in a major chapter on the normal distribution.
- A new section gives a brief history of probability, indicating how the normal distribution was discovered.
- More real examples and exercises concerning probability were added that will appeal to students of actuarial science, finance, economics, and so on.

- There is a short but excellent Bayesian chapter, including real examples and an indication of how Bayesians prove theorems by establishing "Dutch books."
- In the section on bootstrapping there is an explanation of the origin of this word.
- There are examples of Simpson's paradox.
- Some different statistical techniques, including ordered restricted estimates, have been added.
- There is somewhat more emphasis on the importance of sufficient statistics, noting that such statistics, if they exist, are always used in statistical inferences.
- Tests of hypotheses and confidence intervals are tied together.
- An explanation of the Six Sigma program is in the Epilogue.
- The figures are improved with the use of color.
- The CD-ROM includes not only all of the data sets in different formats, but also many more new applications of Minitab and *Maple*.
- Illustrations of *Maple* as a computer algebra system are given in the text and on the CD-ROM.

IMPORTANT POINTS IN THIS EDITION

Chapter 1 is a basic chapter on probability after considering how discrete data can arise. This is followed by a chapter on discrete probability distributions. Chapter 3 starts with continuous type data, introducing stem-and-leaf diagrams, order statistics, and box plots and includes many standard continuous distribution. Chapter 4 introduces multivariate distributions, which are considered with more emphasis than in past editions, including mixtures of discrete and continuous types. Chapter 5 is the climax of the first semester as it is devoted to the important normal distribution, including a brief history of probability noting how the normal distribution was discovered. In these first five chapters over 100 new "real" probability examples and exercises are provided so that we are less dependent upon problems involving coins, cards, and dice. In particular, we hope the reader recognizes the importance of conditional distributions, correlation, and the use of distributions that are mixtures of the discrete and continuous cases.

We begin the second semester with a strong chapter (Chapter 6) on estimation which includes in a starred section the importance of sufficient statistics as shown by the theorem of Rao and Blackwell. This is followed by a new short but modern chapter (Chapter 7) on Bayesian estimation. George Woodworth and Kate Cowles made a number of suggestions concerning this new Bayesian chapter; in particular, Woodworth taught us about a "Dutch book" used in many Bayesian proofs. Chapter 8 is an excellent chapter on tests of statistical hypotheses, and the connection between tests and confidence intervals is given explicitly. Chapter 9 provides some theory of statistical tests. Chapter 10 is short, but interesting, as it considers some of the methods used to improve the quality of manufactured products, although it can be applied to the service industries too. The Six Sigma program is considered in the Epilogue.

We recognize the importance of the computer in modern statistics by including supplements that usually involve the use of the computer. Some of these illustrate the use of Minitab for calculating probabilities, analyzing data, and applications in statistical inference such as tests of hypotheses and confidence intervals. The power of a computer algebra system (CAS) for theoretical computations is illustrated using *Maple*. The importance of simulation is also demonstrated. Some of these illustrations are printed in the text. There are also many new illustrations on the enclosed CD-ROM and you are encouraged to check these out.

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.

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.

FEATURES

Throughout the book, figures and real applications will help the student understand probability and statistics and what they can accomplish. For some exercises, it is assumed that calculators or computers are available; thus the solutions will not always involve "nice" numbers. The data sets for all of the exercises are available on the enclosed CD-ROM. The data are provided in different formats so that they should be accessible to most computer programs. Finally, in the first part of the book concerning probability, there are supplementary comments inserted about statistics and computation.

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.

The **CD-ROM** contains all of the data sets in various formats. There are also applications of *Minitab* for drawing figures, calculating probabilities, calculating characteristics of a sample, and statistical inference. One folder contains some supplementary *Maple* programs that are useful in probability and statistics. It is these programs that were used for constructing the figures in this text and for other applications. *Maple* was also used to animate some of the figures in the text. All you need is a browser. Simply load in the directory and you can pull up these animations.

If you find any errors in this text, please send them to tanis@hope.edu so that they can be corrected in a future printing. The **errata** will also be posted on http://www.math.hope.edu/tanis/.

ACKNOWLEDGMENTS

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 and suggestions of Ching-Yuan Chiang, Richard Dudley, Mark Ghamsary, Paul Joyce, YongHee Kim-Park, Tom Rousseau, David Snyder, and Randall Swift. We also thank the University of Iowa and Hope College for providing office space and encouragement. Finally, our families, through seven editions, have been most understanding during the preparation of all of this material. We would especially like to thank our wives, Ann and Elaine. We truly appreciate their patience and needed their love.

Robert V. Hogg
robert-hogg@uiowa.edu

Elliot A. Tanis
tanis@hope.edu

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, has increased the need for more statistical scientists to examine the huge amount of data being collected. We know that data is not equivalent to information. Once data (hopefully of high quality) are collected, there is a strong 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, while 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, as 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 cited an orchestra as an illustration of the need of 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 cannot 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 a joint program are in great demand in the workplace or in graduate programs. 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, etc. So many of these fields have been "mathematized" that their 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 very successful careers are numerous.

CONTENTS