# CONTENTS

**Preface**  v  
**Prologue**  ix  

## 1 PROBABILITY  1

1.1 Properties of Probability  1  
1.2 Methods of Enumeration  11  
1.3 Conditional Probability  20  
1.4 Independent Events  29  
1.5 Bayes' Theorem  35  

## 2 DISCRETE DISTRIBUTIONS  41

2.1 Random Variables of the Discrete Type  41  
2.2 Mathematical Expectation  47  
2.3 Special Mathematical Expectations  53  
2.4 The Binomial Distribution  63  
2.5 The Hypergeometric Distribution  71  
2.6 The Negative Binomial Distribution  76  
2.7 The Poisson Distribution  81  

## 3 CONTINUOUS DISTRIBUTIONS  91

3.1 Random Variables of the Continuous Type  91  
3.2 The Exponential, Gamma, and Chi-Square Distributions  100  
3.3 The Normal Distribution  110  
3.4 Additional Models  119  

## 4 BIVARIATE DISTRIBUTIONS  129

4.1 Bivariate Distributions of the Discrete Type  129  
4.2 The Correlation Coefficient  139  
4.3 Conditional Distributions  145  
4.4 Bivariate Distributions of the Continuous Type  153  
4.5 The Bivariate Normal Distribution  162  

## 5 DISTRIBUTIONS OF FUNCTIONS OF RANDOM VARIABLES  169

5.1 Functions of One Random Variable  169  
5.2 Transformations of Two Random Variables  178  
5.3 Several Independent Random Variables  187  
5.4 The Moment-Generating Function Technique  194  
5.5 Random Functions Associated with Normal Distributions  199  
5.6 The Central Limit Theorem  207  
5.7 Approximations for Discrete Distributions  213  
5.8 Chebyshev's Inequality and Convergence in Probability  220  
5.9 Limiting Moment-Generating Functions  224  

## 6 POINT ESTIMATION  233

6.1 Descriptive Statistics  233  
6.2 Exploratory Data Analysis  245  
6.3 Order Statistics  256  
6.4 Maximum Likelihood and Method of Moments Estimation  264  
6.5 A Simple Regression Problem  277  
6.6 Asymptotic Distributions of Maximum Likelihood Estimators  285  
6.7 Sufficient Statistics  290  
6.8 Bayesian Estimation  298
Contents

7 INTERVAL ESTIMATION 307
  7.1 Confidence Intervals for Means 307
  7.2 Confidence Intervals for the Difference of Two Means 314
  7.3 Confidence Intervals for Proportions 323
  7.4 Sample Size 329
  7.5 Distribution-Free Confidence Intervals for Percentiles 337
  7.6 More Regression 344
  7.7 Resampling Methods 353

8 TESTS OF STATISTICAL HYPOTHESES 361
  8.1 Tests About One Mean 361
  8.2 Tests of the Equality of Two Means 369
  8.3 Tests for Variances 378
  8.4 Tests About Proportions 385
  8.5 Some Distribution-Free Tests 392
  8.6 Power of a Statistical Test 403
  8.7 Best Critical Regions 410
  8.8 Likelihood Ratio Tests 418

9 MORE TESTS 425
  9.1 Chi-Square Goodness-of-Fit Tests 425
  9.2 Contingency Tables 435
  9.3 One-Factor Analysis of Variance 446
  9.4 Two-Way Analysis of Variance 456
  9.5 General Factorial and $2^k$ Factorial Designs 465
  9.6 Tests Concerning Regression and Correlation 471
  9.7 Statistical Quality Control 477

APPENDICES

A REFERENCES 489

B TABLES 491

C ANSWERS TO ODD-NUMBERED EXERCISES 513

D REVIEW OF SELECTED MATHEMATICAL TECHNIQUES 525
  D.1 Algebra of Sets 525
  D.2 Mathematical Tools for the Hypergeometric Distribution 529
  D.3 Limits 532
  D.4 Infinite Series 533
  D.5 Integration 537
  D.6 Multivariate Calculus 539

Index 545
In this Tenth Edition of *Probability and Statistical Inference*, Elliot Tanis and Dale Zimmerman would like to acknowledge the many contributions that Robert Hogg made to the first nine editions. Dr. Hogg died on December 23, 2014, but his insights continue on in this tenth edition. We are indebted to his influence on our lives and work.

**CONTENT AND COURSE PLANNING**

This text is designed for a two-semester course, but it can be adapted for a one-semester course. A good calculus background is needed, but no previous study of probability or statistics is required.

This new edition has more than 25 new examples and more than 75 new exercises. However, its chapters are organized in much the same manner as in the ninth edition. The first five again focus on probability, including the following topics: conditional probability, independence, Bayes’ theorem, discrete and continuous distributions, certain mathematical expectations including moment-generating functions, bivariate distributions along with marginal and conditional distributions, correlation, functions of random variables and their distributions, the central limit theorem, and Chebyshev’s inequality. We added a section on the hypergeometric distribution, adding to material that had previously been scattered throughout the first and second chapters. Also, to this portion of the book we added material on new topics, including the index of skewness and the laws of total probability for expectations and the variance. While the strong probability coverage of the first five chapters is important for all students, feedback we have received indicates that it has been particularly helpful to actuarial students who are studying for Exam P in the Society of Actuaries’ series (or Exam 1 of the Casualty Actuarial Society).

The remaining four chapters of the book focus on statistical inference. Topics carried over from the previous edition include descriptive and order statistics, point estimation including maximum likelihood and method of moments estimation, sufficient statistics, Bayesian estimation, simple linear regression, interval estimation, and hypothesis testing. New material has been added on the topics of percentile matching and the invariance of maximum likelihood estimation, and we’ve added a new section on hypothesis testing for variances, which also includes confidence intervals for a variance and for the ratio of two variances. We present confidence intervals for means, variances, proportions, and regression coefficients; distribution-free confidence intervals for percentiles; and resampling methods (in particular, bootstrapping). Our coverage of hypothesis testing includes standard tests on means (including distribution-free tests), variances, proportions, and regression coefficients, power and sample size, best critical regions (Neyman-Pearson), and likelihood ratio tests. On the more applied side, we describe chi-square tests for goodness of fit and for association in contingency tables, analysis of variance including general factorial designs, and statistical quality control.
Preface

The first semester of the course should contain most of the topics in Chapters 1–5. The second semester includes some topics omitted there and many of those in Chapters 6–9. A more basic course might omit some of the starred sections, but we believe that the order of topics will give the instructor the flexibility needed in his or her course. The usual nonparametric and Bayesian techniques are placed at appropriate places in the text rather than in separate chapters. We find that many persons like the applications associated with statistical quality control in the last section.

The Prologue suggests many fields in which statistical methods can be used. At the end of each chapter, we give some interesting historical comments, which have proven to be very worthwhile in the past editions. The answers given in this text for exercises that involve the standard distributions were often calculated using our probability tables which, of course, are rounded off for printing. If you use a statistical package, your answers may differ slightly from those given.

ANCILLARIES

Data sets for this text are available on Pearson’s Student Resources website: https://www.pearson.com/math-stats-resources.


Some of the numerical exercises were solved with 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 download from Pearson’s Student Resources website. This is located at https://www.pearson.com/math-stats-resources. Several exercises in that manual also make use of the power of Maple as a computer algebra system.

If you find errors in this text, please send them to dale-zimmerman@uiowa.edu so that they can be corrected in a future printing. These errata will also be posted on http://homepage.divms.uiowa.edu/~dzimmer/.

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 thank the reviewers of the ninth edition who made suggestions for this edition. They are Maureen Cox from St. Bonaventure University, Lynne Seymour from the University of Georgia, Kevin Keen from the University of Northern British Columbia, Clifford Taylor from Concordia University Ann Arbor, Melanie Autin from Western Kentucky University, Aubie Anisef from Douglas College, Manohar Aggarwal from the University of Memphis, Joseph Huber from the University of Kansas, and Christopher Swanson from Ashland University. Mark Mills from Central College in Iowa, Matthew Bognar from the University of Iowa, and David Schweitzer from Liberty University also made many helpful comments, and Hongda Zhang of the University of Iowa wrote solutions to some of the new exercises. We also acknowledge the excellent suggestions from our copy editor Jody Callahan and the fine work of our accuracy checker Kyle Siegrist. We also thank the University of Iowa and Hope College for providing office space and encouragement. Finally, our families have been most understanding during the preparation of all of this material.
We would especially like to thank our wives, Elaine and Bridget. We truly appreciate their patience and needed their love.

Elliot A. Tanis
tanis@hope.edu

Dale L. Zimmerman
dale-zimmerman@uiowa.edu
The discipline of statistics deals with the collection and analysis of data. Advances in computing technology, particularly in relation 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 a strong need for statisticians to make sense of them. That is, data must be analyzed in order to provide 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 to go 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. Because statistics is useful in all these areas, 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. To be applied properly, most appropriate models in statistics require a certain mathematical background in probability. Accordingly, while alluding to applications in the examples and exercises, this textbook 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 are collected and analyzed, providing information about the truth of the conjectures. Sometimes the conjectures are supported by the data, but often the conjectures need to be modified and more data must be 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 that is relevant to taking certain actions, including 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 errors in their inferences, and they attempt to quantify the probabilities of those mistakes and make them as small as possible. 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. In light of this uncertainty, the statistician tries to summarize the data in the best possible way, 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 variation. Often, as in manufacturing, the desire is to reduce variation so that the products will be more consistent. In other words, car doors will fit better in the manufacturing of automobiles if the variation is decreased by making each door closer to its target values.
Any student of statistics should understand the nature of variability and the necessity for creating probabilistic models of that variability. We cannot avoid making inferences and decisions in the face of this uncertainty; however, these inferences and decisions 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; it is hoped that thereby the reader will become a better model builder.

Finally, we must mention how modern statistical analyses have become dependent upon the computer. Increasingly, statisticians and computer scientists are working 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 a program at their interface such as data science are in great demand in the workplace and 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 fields have been “mathematized” that their programs are begging for majors in statistics or computer science. Often, such students become “stars” in these other areas. We truly hope that we can interest students enough that they want to study more statistics. If they do, they will find that the opportunities for very successful careers are numerous.