INTRODUCTION TO PROBABILITY AND STATISTICS FOR ENGINEERS AND SCIENTISTS

Fourth Edition
LIMITED WARRANTY AND DISCLAIMER OF LIABILITY

Academic Press, ("AP") and anyone else who has been involved in the creation or production of the accompanying code ("the product") cannot and do not warrant the performance or results that may be obtained by using the product. The product is sold "as is" without warranty of merchantability or fitness for any particular purpose. AP warrants only that the magnetic diskette(s) on which the code is recorded is free from defects in material and faulty workmanship under the normal use and service for a period of ninety (90) days from the date the product is delivered. The purchaser's sole and exclusive remedy in the event of a defect is expressly limited to either replacement of the diskette(s) or refund of the purchase price, at AP's sole discretion.

In no event, whether as a result of breach of contract, warranty, or tort (including negligence), will AP or anyone who has been involved in the creation or production of the product be liable to purchaser for any damages, including any lost profits, lost savings, or other incidental or consequential damages arising out of the use or inability to use the product or any modifications thereof, or due to the contents of the code, even if AP has been advised on the possibility of such damages, or for any claim by any other party.

Any request for replacement of a defective diskette must be postage prepaid and must be accompanied by the original defective diskette, your mailing address and telephone number, and proof of date of purchase and purchase price. Send such requests, stating the nature of the problem, to Academic Press Customer Service, 6277 Sea Harbor Drive, Orlando, FL 32887, 1-800-321-5068. AP shall have no obligation to refund the purchase price or to replace a diskette based on claims of defects in the nature or operation of the product.

Some states do not allow limitation on how long an implied warranty lasts, nor exclusions or limitations of incidental or consequential damage, so the above limitations and exclusions may not apply to you. This warranty gives you specific legal rights, and you may also have other rights, which vary from jurisdiction to jurisdiction.

The re-export of United States original software is subject to the United States laws under the Export Administration Act of 1969 as amended. Any further sale of the product shall be in compliance with the United States Department of Commerce Administration regulations. Compliance with such regulations is your responsibility and not the responsibility of AP.
For
Elise
Contents

Preface ........................................................................................................... xiii

Chapter 1  Introduction to Statistics ......................................................... 1
  1.1  Introduction ............................................................................. 1
  1.2  Data Collection and Descriptive Statistics .................................. 1
  1.3  Inferential Statistics and Probability Models .............................. 2
  1.4  Populations and Samples ....................................................... 3
  1.5  A Brief History of Statistics .................................................. 3
      Problems ....................................................................................... 7

Chapter 2  Descriptive Statistics ............................................................ 9
  2.1  Introduction ............................................................................. 9
  2.2  Describing Data Sets ............................................................ 9
      2.2.1  Frequency Tables and Graphs .......................................... 10
      2.2.2  Relative Frequency Tables and Graphs ............................ 10
      2.2.3  Grouped Data, Histograms, Ogives, and Stem and Leaf Plots ........................................ 14
  2.3  Summarizing Data Sets ......................................................... 17
      2.3.1  Sample Mean, Sample Median, and Sample Mode .............. 17
      2.3.2  Sample Variance and Sample Standard Deviation ............ 22
      2.3.3  Sample Percentiles and Box Plots .................................. 24
  2.4  Chebyshev’s Inequality ......................................................... 27
  2.5  Normal Data Sets ................................................................. 31
  2.6  Paired Data Sets and the Sample Correlation Coefficient .......... 33
      Problems ....................................................................................... 41

Chapter 3  Elements of Probability ......................................................... 55
  3.1  Introduction ............................................................................. 55
  3.2  Sample Space and Events ...................................................... 56
  3.3  Venn Diagrams and the Algebra of Events .............................. 58
  3.4  Axioms of Probability .......................................................... 59
Contents

3.5 Sample Spaces Having Equally Likely Outcomes ................. 61
3.6 Conditional Probability ........................................ 67
3.7 Bayes' Formula ............................................... 70
3.8 Independent Events ........................................... 75
Problems .................................................... 80

Chapter 4 Random Variables and Expectation ......................... 89
4.1 Random Variables ............................................. 89
4.2 Types of Random Variables ................................... 92
4.3 Jointly Distributed Random Variables ........................... 96
  4.3.1 Independent Random Variables .............................. 101
  *4.3.2 Conditional Distributions ................................ 105
4.4 Expectation .................................................. 107
4.5 Properties of the Expected Value ................................ 111
  4.5.1 Expected Value of Sums of Random Variables .......... 115
4.6 Variance ..................................................... 118
4.7 Covariance and Variance of Sums of Random Variables .......... 121
4.8 Moment Generating Functions .................................. 125
4.9 Chebyshev's Inequality and the Weak Law of Large Numbers .... 127
Problems .................................................... 130

Chapter 5 Special Random Variables ................................ 141
5.1 The Bernoulli and Binomial Random Variables ................... 141
  5.1.1 Computing the Binomial Distribution Function .......... 147
5.2 The Poisson Random Variable .................................. 148
  5.2.1 Computing the Poisson Distribution Function .......... 155
5.3 The Hypergeometric Random Variable ........................ 156
5.4 The Uniform Random Variable .................................. 160
5.5 Normal Random Variables ...................................... 168
5.6 Exponential Random Variables .................................. 176
  *5.6.1 The Poisson Process .................................... 180
*5.7 The Gamma Distribution ........................................ 183
5.8 Distributions Arising from the Normal .......................... 186
  5.8.1 The Chi-Square Distribution .............................. 186
  5.8.2 The t-Distribution ...................................... 190
  5.8.3 The F-Distribution ...................................... 192
*5.9 The Logistics Distribution .................................... 193
Problems .................................................... 195

Chapter 6 Distributions of Sampling Statistics ...................... 203
6.1 Introduction ............................................... 203
6.2 The Sample Mean ............................................. 204
6.3 The Central Limit Theorem ..................................... 206
  6.3.1 Approximate Distribution of the Sample Mean ............... 212
  6.3.2 How Large a Sample Is Needed? ................................ 214
6.4 The Sample Variance ........................................ 215
6.5 Sampling Distributions from a Normal Population ............... 216
  6.5.1 Distribution of the Sample Mean .......................... 217
  6.5.2 Joint Distribution of $X$ and $S^2$ ........................... 217
6.6 Sampling from a Finite Population ............................. 219
Problems ...................................................................... 223

Chapter 7 Parameter Estimation ..................................... 231
7.1 Introduction ..................................................... 231
7.2 Maximum Likelihood Estimators ................................ 232
  *7.2.1 Estimating Life Distributions ................................ 240
7.3 Interval Estimates ............................................. 242
  7.3.1 Confidence Interval for a Normal Mean When the Variance
       Is Unknown .................................................. 248
  7.3.2 Confidence Intervals for the Variance of a Normal Distribution 253
7.4 Estimating the Difference in Means of Two Normal Populations 255
7.5 Approximate Confidence Interval for the Mean of a Bernoulli Random
       Variable ...................................................... 262
  *7.6 Confidence Interval of the Mean of the Exponential Distribution 267
  *7.7 Evaluating a Point Estimator .................................. 268
  *7.8 The Bayes Estimator .......................................... 274
Problems ...................................................................... 279

Chapter 8 Hypothesis Testing ........................................ 293
8.1 Introduction ...................................................... 294
8.2 Significance Levels .............................................. 294
8.3 Tests Concerning the Mean of a Normal Population .......... 295
  8.3.1 Case of Known Variance ..................................... 295
  8.3.2 Case of Unknown Variance: The $t$-Test ..................... 307
8.4 Testing the Equality of Means of Two Normal Populations ... 314
  8.4.1 Case of Known Variances .................................... 314
  8.4.2 Case of Unknown Variances .................................. 316
  8.4.3 Case of Unknown and Unequal Variances .................. 320
  8.4.4 The Paired $t$-Test .......................................... 321
8.5 Hypothesis Tests Concerning the Variance of a Normal Population 323
  8.5.1 Testing for the Equality of Variances of Two Normal Populations 324
8.6 Hypothesis Tests in Bernoulli Populations ........................ 325
  8.6.1 Testing the Equality of Parameters in Two Bernoulli Populations 329
8.7 Tests Concerning the Mean of a Poisson Distribution .................. 332
8.7.1 Testing the Relationship Between Two Poisson Parameters .......... 333
Problems .................................................... 336

Chapter 9 Regression ............................................... 353
9.1 Introduction ................................................. 353
9.2 Least Squares Estimators of the Regression Parameters ............... 355
9.3 Distribution of the Estimators ................................... 357
9.4 Statistical Inferences About the Regression Parameters ............... 363
  9.4.1 Inferences Concerning β .................................... 364
  9.4.2 Inferences Concerning α ..................................... 372
  9.4.3 Inferences Concerning the Mean Response \( \alpha + \beta x_0 \) ................. 373
  9.4.4 Prediction Interval of a Future Response ........................ 375
  9.4.5 Summary of Distributional Results ............................. 377
9.5 The Coefficient of Determination and the Sample Correlation
  Coefficient .................................................. 378
9.6 Analysis of Residuals: Assessing the Model ........................... 380
9.7 Transforming to Linearity ........................................ 383
9.8 Weighted Least Squares ........................................ 386
9.9 Polynomial Regression ......................................... 393
9.10 Multiple Linear Regression ..................................... 396
  9.10.1 Predicting Future Responses .................................. 407
9.11 Logistic Regression Models for Binary Output Data ................. 412
Problems .................................................... 415

Chapter 10 Analysis of Variance ....................................... 441
10.1 Introduction ................................................. 441
10.2 An Overview ................................................. 442
10.3 One-Way Analysis of Variance ................................... 444
  10.3.1 Multiple Comparisons of Sample Means ......................... 452
  10.3.2 One-Way Analysis of Variance with Unequal Sample Sizes .......... 454
10.4 Two-Factor Analysis of Variance: Introduction and Parameter
  Estimation ................................................... 456
10.5 Two-Factor Analysis of Variance: Testing Hypotheses ................. 460
10.6 Two-Way Analysis of Variance with Interaction ..................... 465
Problems .................................................... 473

Chapter 11 Goodness of Fit Tests and Categorical Data Analysis .......... 485
11.1 Introduction .................................................. 485
11.2 Goodness of Fit Tests When All Parameters Are Specified .......... 486
  11.2.1 Determining the Critical Region by Simulation .................... 492
11.3 Goodness of Fit Tests When Some Parameters Are Unspecified ........ 495
## Contents

### 11.4 Tests of Independence in Contingency Tables
- Page 497

### 11.5 Tests of Independence in Contingency Tables Having Fixed Marginal Totals
- Page 501

*### 11.6 The Kolmogorov–Smirnov Goodness of Fit Test for Continuous Data
- Page 506

### Problems
- Page 510

### Chapter 12 Nonparametric Hypothesis Tests
- Page 517
  - 12.1 Introduction
  - Page 517
  - 12.2 The Sign Test
  - Page 517
  - 12.3 The Signed Rank Test
  - Page 521
  - 12.4 The Two-Sample Problem
    * 12.4.1 The Classical Approximation and Simulation
    - Page 531
  - 12.5 The Runs Test for Randomness
  - Page 535
  - Problems
  - Page 539

### Chapter 13 Quality Control
- Page 547
  - 13.1 Introduction
  - Page 547
  - 13.2 Control Charts for Average Values: The $\bar{X}$-Control Chart
    * 13.2.1 Case of Unknown $\mu$ and $\sigma$
    - Page 551
  - 13.3 $S$-Control Charts
  - Page 556
  - 13.4 Control Charts for the Fraction Defective
  - Page 559
  - 13.5 Control Charts for Number of Defects
  - Page 561
  - 13.6 Other Control Charts for Detecting Changes in the Population Mean
    * 13.6.1 Moving-Average Control Charts
    - Page 565
    * 13.6.2 Exponentially Weighted Moving-Average Control Charts
    - Page 567
    * 13.6.3 Cumulative Sum Control Charts
    - Page 573
  - Problems
  - Page 575

### Chapter 14* Life Testing
- Page 583
  - 14.1 Introduction
  - Page 583
  - 14.2 Hazard Rate Functions
  - Page 583
  - 14.3 The Exponential Distribution in Life Testing
    * 14.3.1 Simultaneous Testing — Stopping at the $r$th Failure
    - Page 586
    * 14.3.2 Sequential Testing
    - Page 592
    * 14.3.3 Simultaneous Testing — Stopping by a Fixed Time
    - Page 596
    * 14.3.4 The Bayesian Approach
    - Page 598
  - 14.4 A Two-Sample Problem
  - Page 600
  - 14.5 The Weibull Distribution in Life Testing
    * 14.5.1 Parameter Estimation by Least Squares
    - Page 604
  - Problems
  - Page 606
Chapter 15  Simulation, Bootstrap Statistical Methods, and Permutation Tests ............................................. 613

15.1  Introduction .......................................................... 613

15.2  Random Numbers ..................................................... 614

15.2.1  The Monte Carlo Simulation Approach ......................... 616

15.3  The Bootstrap Method ............................................... 617

15.4  Permutation Tests .................................................... 624

15.4.1  Normal Approximations in Permutation Tests .................... 627

15.4.2  Two-Sample Permutation Tests .................................... 631

15.5  Generating Discrete Random Variables ............................... 632

15.6  Generating Continuous Random Variables ........................ 634

15.6.1  Generating a Normal Random Variable ........................... 636

15.7  Determining the Number of Simulation Runs in a Monte Carlo Study . 637

Problems ................................................................. 638

Appendix of Tables .......................................................... 641

Index ........................................................................... 647

* Denotes optional material.
Preface

The fourth edition of this book continues to demonstrate how to apply probability theory to gain insight into real, everyday statistical problems and situations. As in the previous editions, carefully developed coverage of probability motivates probabilistic models of real phenomena and the statistical procedures that follow. This approach ultimately results in an intuitive understanding of statistical procedures and strategies most often used by practicing engineers and scientists.

This book has been written for an introductory course in statistics or in probability and statistics for students in engineering, computer science, mathematics, statistics, and the natural sciences. As such it assumes knowledge of elementary calculus.

ORGANIZATION AND COVERAGE

Chapter 1 presents a brief introduction to statistics, presenting its two branches of descriptive and inferential statistics, and a short history of the subject and some of the people whose early work provided a foundation for work done today.

The subject matter of descriptive statistics is then considered in Chapter 2. Graphs and tables that describe a data set are presented in this chapter, as are quantities that are used to summarize certain of the key properties of the data set.

To be able to draw conclusions from data, it is necessary to have an understanding of the data’s origination. For instance, it is often assumed that the data constitute a “random sample” from some population. To understand exactly what this means and what its consequences are for relating properties of the sample data to properties of the entire population, it is necessary to have some understanding of probability, and that is the subject of Chapter 3. This chapter introduces the idea of a probability experiment, explains the concept of the probability of an event, and presents the axioms of probability.

Our study of probability is continued in Chapter 4, which deals with the important concepts of random variables and expectation, and in Chapter 5, which considers some special types of random variables that often occur in applications. Such random variables as the binomial, Poisson, hypergeometric, normal, uniform, gamma, chi-square, $t$, and $F$ are presented.
In **Chapter 6**, we study the probability distribution of such sampling statistics as the sample mean and the sample variance. We show how to use a remarkable theoretical result of probability, known as the central limit theorem, to approximate the probability distribution of the sample mean. In addition, we present the joint probability distribution of the sample mean and the sample variance in the important special case in which the underlying data come from a normally distributed population.

**Chapter 7** shows how to use data to estimate parameters of interest. For instance, a scientist might be interested in determining the proportion of Midwestern lakes that are afflicted by acid rain. Two types of estimators are studied. The first of these estimates the quantity of interest with a single number (for instance, it might estimate that 47 percent of Midwestern lakes suffer from acid rain), whereas the second provides an estimate in the form of an interval of values (for instance, it might estimate that between 45 and 49 percent of lakes suffer from acid rain). These latter estimators also tell us the “level of confidence” we can have in their validity. Thus, for instance, whereas we can be pretty certain that the exact percentage of afflicted lakes is not 47, it might very well be that we can be, say, 95 percent confident that the actual percentage is between 45 and 49.

**Chapter 8** introduces the important topic of statistical hypothesis testing, which is concerned with using data to test the plausibility of a specified hypothesis. For instance, such a test might reject the hypothesis that fewer than 44 percent of Midwestern lakes are afflicted by acid rain. The concept of the \( p \)-value, which measures the degree of plausibility of the hypothesis after the data have been observed, is introduced. A variety of hypothesis tests concerning the parameters of both one and two normal populations are considered. Hypothesis tests concerning Bernoulli and Poisson parameters are also presented.

**Chapter 9** deals with the important topic of regression. Both simple linear regression — including such subtopics as regression to the mean, residual analysis, and weighted least squares — and multiple linear regression are considered.

**Chapter 10** introduces the analysis of variance. Both one-way and two-way (with and without the possibility of interaction) problems are considered.

**Chapter 11** is concerned with goodness of fit tests, which can be used to test whether a proposed model is consistent with data. In it we present the classical chi-square goodness of fit test and apply it to test for independence in contingency tables. The final section of this chapter introduces the Kolmogorov–Smirnov procedure for testing whether data come from a specified continuous probability distribution.

**Chapter 12** deals with nonparametric hypothesis tests, which can be used when one is unable to suppose that the underlying distribution has some specified parametric form (such as normal).

**Chapter 13** considers the subject matter of quality control, a key statistical technique in manufacturing and production processes. A variety of control charts, including not only the Shewhart control charts but also more sophisticated ones based on moving averages and cumulative sums, are considered.

**Chapter 14** deals with problems related to life testing. In this chapter, the exponential, rather than the normal, distribution plays the key role.
In Chapter 15 (new to the fourth edition), we consider the statistical inference techniques of bootstrap statistical methods and permutation tests. We first show how probabilities can be obtained by simulation and then how to utilize simulation in these statistical inference approaches.

ABOUT THE CD
Packaged along with the text is a PC disk that can be used to solve most of the statistical problems in the text. For instance, the disk computes the $p$-values for most of the hypothesis tests, including those related to the analysis of variance and to regression. It can also be used to obtain probabilities for most of the common distributions. (For those students without access to a personal computer, tables that can be used to solve all of the problems in the text are provided.)

One program on the disk illustrates the central limit theorem. It considers random variables that take on one of the values 0, 1, 2, 3, 4, and allows the user to enter the probabilities for these values along with an integer $n$. The program then plots the probability mass function of the sum of $n$ independent random variables having this distribution. By increasing $n$, one can “see” the mass function converge to the shape of a normal density function.

ACKNOWLEDGMENTS
We thank the following people for their helpful comments on the Fourth Edition:

- Charles F. Dunkl, University of Virginia, Charlottesville
- Gabor Szekely, Bowling Green State University
- Krzysztof M. Ostaszewski, Illinois State University
- Michael Ratliff, Northern Arizona University
- Wei-Min Huang, Lehigh University
- Youngho Lee, Howard University
- Jacques Rioux, Drake University
- Lisa Gardner, Bradley University
- Murray Lieb, New Jersey Institute of Technology
- Philip Trotter, Cornell University