

# Contents

About the Authors	iv
To Teachers: About This Book	xi
Preface	xvi
To Students: What Is Statistics?	xxiii
Applications	xxvi
Data Table Index	xxix
Beyond the Basics Index	xxx

## PART I Looking at Data

---

### CHAPTER 1 Looking at Data—Distributions 1

#### Introduction 1

#### 1.1 Data 2

Key characteristics of a data set 4

#### Section 1.1 Summary 7

#### Section 1.1 Exercises 7

#### 1.2 Displaying Distributions with Graphs 9

Categorical variables: Bar graphs and pie charts 9

Quantitative variables: Stemplots and histograms 12

Histograms 14

Examining distributions 17

Dealing with outliers 18

Time plots 20

#### Section 1.2 Summary 21

#### Section 1.2 Exercises 21

#### 1.3 Describing Distributions with Numbers 25

Measuring center: The mean 26

Measuring center: The median 28

Comparing the mean and the median 29

Measuring spread: The quartiles 30

The five-number summary and boxplots 31

The  $1.5 \times IQR$  rule for suspected outliers 33

Measuring spread: The standard deviation 35

Properties of the standard deviation 37

Choosing measures of center and spread 38

Changing the unit of measurement 39

#### Section 1.3 Summary 42

#### Section 1.3 Exercises 43

#### 1.4 Density Curves and Normal Distributions 46

Density curves 47

Measuring center and spread for density curves 48

Normal distributions 49

The 68–95–99.7 rule 51

Standardizing observations 52

Normal distribution calculations 54

Using the standard Normal table 56

Inverse Normal calculations 58

Normal quantile plots 59

#### Beyond the Basics: Density estimation 61

#### Section 1.4 Summary 62

#### Section 1.4 Exercises 63

#### Chapter 1 Exercises 67

---

### CHAPTER 2 Looking at Data—

#### Relationships 71

#### Introduction 71

#### 2.1 Relationships 72

Examining relationships 73

#### Section 2.1 Summary 76

#### Section 2.1 Exercises 76

#### 2.2 Scatterplots 77

Interpreting scatterplots 80

The log transformation 83

Adding categorical variables to scatterplots 84

Scatterplot smoothers 85

Categorical explanatory variables 87

#### Section 2.2 Summary 87

#### Section 2.2 Exercises 88

#### 2.3 Correlation 91

The correlation  $r$  92

Properties of correlation 94

#### Section 2.3 Summary 96

#### Section 2.3 Exercises 96

#### 2.4 Least-Squares Regression 98

Fitting a line to data 99

Prediction 101

The least-squares regression line 101

- Facts about least-squares regression 105
- Correlation and regression 107
- Interpretation of  $r^2$  107
- Section 2.4 Summary 108**
- Section 2.4 Exercises 109**
- 2.5 Cautions about Correlation and Regression 113**
  - Extrapolation 113
  - Residuals 115
  - The distribution of the residuals 118
  - Outliers and influential observations 119
  - Beware of the lurking variable 122
  - Beware of correlations based on averaged data 123
  - Beware of restricted ranges 123
- Beyond the Basics: Data mining 124**
- Section 2.5 Summary 125**
- Section 2.5 Exercises 125**
- 2.6 Data Analysis for Two-Way Tables 128**
  - The two-way table 128
  - Joint distribution 130
  - Marginal distributions 131
  - Describing relations in two-way tables 132
  - Conditional distributions 132
  - Simpson's paradox 135
- Section 2.6 Summary 136**
- Section 2.6 Exercises 137**
- 2.7 The Question of Causation 140**
  - Explaining association 140
  - Establishing causation 142
- Section 2.7 Summary 143**
- Section 2.7 Exercises 144**
- Chapter 2 Exercises 145**

---

## CHAPTER 3 Producing Data 153

Introduction 153

### 3.1 Sources of Data 154

- Anecdotal data 154
- Available data 155
- Sample surveys and experiments 157

#### Section 3.1 Summary 160

#### Section 3.1 Exercises 160

### 3.2 Design of Experiments 161

- Comparative experiments 163
- Randomization 165
- Randomized comparative experiments 167
- How to randomize 168
- Randomization using software 168
- Randomization using random digits 169
- Cautions about experimentation 171

Matched pairs designs 172

Block designs 173

#### Section 3.2 Summary 175

#### Section 3.2 Exercises 175

### 3.3 Sampling Design 178

- Simple random samples 180
- How to select a simple random sample 181
- Stratified random samples 184
- Multistage random samples 184
- Cautions about sample surveys 185

#### Beyond the Basics: Capture-recapture sampling 187

#### Section 3.3 Summary 188

#### Section 3.3 Exercises 189

### 3.4 Ethics 192

- Institutional review boards 193
- Informed consent 194
- Confidentiality 195
- Clinical trials 196
- Behavioral and social science experiments 198

#### Section 3.4 Summary 200

#### Section 3.4 Exercises 200

#### Chapter 3 Exercises 201

## PART II Probability and Inference

---

### CHAPTER 4 Probability: The Study of Randomness 204

Introduction 204

#### 4.1 Randomness 205

- The language of probability 206
- Thinking about randomness 207
- The uses of probability 208

#### Section 4.1 Summary 208

#### Section 4.1 Exercises 209

#### 4.2 Probability Models 210

- Sample spaces 210
- Probability rules 213
- Assigning probabilities: Finite number of outcomes 215
- Assigning probabilities: Equally likely outcomes 216
- Independence and the multiplication rule 217
- Applying the probability rules 220

#### Section 4.2 Summary 221

#### Section 4.2 Exercises 221

#### 4.3 Random Variables 224

- Discrete random variables 225
- Continuous random variables 228
- Normal distributions as probability distributions 231

Section 4.3 Summary	232	Binomial mean and standard deviation	304
Section 4.3 Exercises	233	Sample proportions	306
<b>4.4 Means and Variances of Random Variables 235</b>		Normal approximation for counts and proportions	308
The mean of a random variable	236	The continuity correction	311
Statistical estimation and the law of large numbers	239	Binomial formula	312
Thinking about the law of large numbers	240	The Poisson distributions for sample counts	314
<b>Beyond the Basics: More laws of large numbers 241</b>		<b>Section 5.3 Summary</b>	<b>318</b>
Rules for means	242	<b>Section 5.3 Exercises</b>	<b>319</b>
The variance of a random variable	244	<b>Chapter 5 Exercises</b>	<b>322</b>
Rules for variances and standard deviations	246		
<b>Section 4.4 Summary</b>	<b>250</b>	<hr/>	
<b>Section 4.4 Exercises</b>	<b>251</b>	<b>CHAPTER 6 Introduction to Inference 328</b>	
<b>4.5 General Probability Rules 253</b>		<b>Introduction 328</b>	
General addition rules	253	Overview of inference	329
Conditional probability	256	<b>6.1 Estimating with Confidence 330</b>	
General multiplication rules	259	Statistical confidence	331
Tree diagrams	260	Confidence intervals	333
Bayes's rule	262	Confidence interval for a population mean	335
Independence again	263	How confidence intervals behave	338
<b>Section 4.5 Summary</b>	<b>264</b>	Choosing the sample size	340
<b>Section 4.5 Exercises</b>	<b>265</b>	Some cautions	341
<b>Chapter 4 Exercises</b>	<b>267</b>	<b>Section 6.1 Summary</b>	<b>343</b>
		<b>Section 6.1 Exercises</b>	<b>343</b>
		<b>6.2 Tests of Significance 346</b>	
		The reasoning of significance tests	347
		Stating hypotheses	348
		Test statistics	350
		<i>P</i> -values	351
		Statistical significance	353
		Tests for a population mean	355
		Two-sided significance tests and confidence intervals	360
		The <i>P</i> -value versus a statement of significance	362
		<b>Section 6.2 Summary</b>	<b>363</b>
		<b>Section 6.2 Exercises</b>	<b>364</b>
		<b>6.3 Use and Abuse of Tests 367</b>	
		Choosing a level of significance	368
		What statistical significance does not mean	369
		Don't ignore lack of significance	370
		Statistical inference is not valid for all sets of data	371
		Beware of searching for significance	372
		<b>Section 6.3 Summary</b>	<b>372</b>
		<b>Section 6.3 Exercises</b>	<b>373</b>
		<b>6.4 Inference as a Decision 374</b>	
		Two types of error	375
		Error probabilities	376
		The common practice of testing hypotheses	378
		<b>Section 6.4 Summary</b>	<b>378</b>
		<b>Section 6.4 Exercises</b>	<b>379</b>
		<b>Chapter 6 Exercises</b>	<b>380</b>
<b>CHAPTER 5 Sampling Distributions 270</b>			
<b>Introduction 270</b>			
<b>5.1 Toward Statistical Inference 271</b>			
Sampling variability	272		
Sampling distributions	274		
Bias and variability	276		
Sampling from large populations	277		
Why randomize?	278		
<b>Section 5.1 Summary</b>	<b>279</b>		
<b>Section 5.1 Exercises</b>	<b>279</b>		
<b>5.2 The Sampling Distribution of a Sample Mean 281</b>			
The mean and standard deviation of $\bar{x}$	284		
The central limit theorem	286		
A few more facts related to the sampling distribution of $\bar{x}$	292		
<b>Beyond the Basics: Weibull distributions 293</b>			
<b>Section 5.2 Summary</b>	<b>295</b>		
<b>Section 5.2 Exercises</b>	<b>295</b>		
<b>5.3 Sampling Distributions for Counts and Proportions 297</b>			
The binomial distributions for sample counts	299		
Binomial distributions in statistical sampling	300		
Finding binomial probabilities	301		

---

**CHAPTER 7 Inference for Means 383**

Introduction 383

**7.1 Inference for the Mean of a Population 384**

- The  $t$  distributions 384
- One-sample  $t$  confidence interval 386
- The one-sample  $t$  test 388
- Using software 391
- Matched pairs  $t$  procedures 393
- Robustness of the  $t$  procedures 397
- Inference for non-Normal populations 398

**Beyond the Basics: The bootstrap 403**

Section 7.1 Summary 404

Section 7.1 Exercises 405

**7.2 Comparing Two Means 410**

- The two-sample  $z$  statistic 411
- The two-sample  $t$  procedures 413
- The two-sample  $t$  confidence interval 414
- The two-sample  $t$  significance test 416
- Robustness of the two-sample procedures 418
- Inference for small samples 420
- The pooled two-sample  $t$  procedures 423

Section 7.2 Summary 427

Section 7.2 Exercises 428

**7.3 Sample Size Calculations 433**

- Sample size for confidence intervals 433
- Power of a significance test 438

Section 7.3 Summary 442

Section 7.3 Exercises 442

Chapter 7 Exercises 444

---

**CHAPTER 8 Inference for Proportions 450**

Introduction 450

**8.1 Inference for a Single Proportion 451**

- Large-sample confidence interval for a single proportion 452

**Beyond the Basics: Plus four confidence interval for a single proportion 455**

- Significance test for a single proportion 456
- Choosing a sample size for a confidence interval 459
- Choosing a sample size for a significance test 462

Section 8.1 Summary 464

Section 8.1 Exercises 464

**8.2 Comparing Two Proportions 468**

- Large-sample confidence interval for a difference in proportions 470

**Beyond the Basics: Plus four confidence interval for a difference in proportions 472**

- Significance test for a difference in proportions 473

- Choosing a sample size for two sample proportions 476

**Beyond the Basics: Relative risk 479**

Section 8.2 Summary 480

Section 8.2 Exercises 481

Chapter 8 Exercises 483

---

**PART III Topics in Inference**

---

**CHAPTER 9 Inference for Categorical Data 486**

Introduction 486

**9.1 Inference for Two-Way Tables 487**

- The hypothesis: No association 492
- Expected cell counts 492
- The chi-square test 493
- Computations 496
- Computing conditional distributions 497
- The chi-square test and the  $z$  test 500

**Beyond the Basics: Meta-analysis 501**

Section 9.1 Summary 503

Section 9.1 Exercises 503

**9.2 Goodness of Fit 505**

Section 9.2 Summary 510

Section 9.2 Exercises 510

Chapter 9 Exercises 511

---

**CHAPTER 10 Inference for Regression 515**

Introduction 515

**10.1 Simple Linear Regression 516**

- Statistical model for linear regression 517
- Preliminary data analysis and inference considerations 518
- Revisiting the simple linear regressions model 519
- Estimating the regression parameters 520
- Checking conditions for regression inference 524
- Confidence intervals and significance tests 525
- Confidence intervals for mean response 528
- Prediction intervals 530
- Transforming variables 531

**Beyond the Basics: Nonlinear regression 533**

Section 10.1 Summary 534

Section 10.1 Exercises 536

**10.2 More Detail about Simple Linear Regression 539**

- Analysis of variance for regression 539
- The ANOVA  $F$  test 542

Calculations for regression inference 544  
Inference for correlation 550

**Section 10.2 Summary 553**

**Section 10.2 Exercises 554**

**Chapter 10 Exercises 555**

## **CHAPTER 11 Multiple Regression 563**

**Introduction 563**

### **11.1 Inference for Multiple Regression 564**

Population multiple regression  
equation 564  
Data for multiple regression 565  
Multiple linear regression model 566  
Estimation of the multiple regression  
parameters 568  
Confidence intervals and significance tests  
for regression coefficients 569  
ANOVA table for multiple regression 570  
Squared multiple correlation  $R^2$  571

**Section 11.1 Summary 572**

**Section 11.1 Exercises 573**

### **11.2 A Case Study 576**

Preliminary analysis 577  
Relationships between pairs  
of variables 578  
Fitting a multiple regression model 580  
Interpretation of results 582  
Examining the residuals 582  
Refining the model 583  
Considering other sets of explanatory  
variables 585  
Test for a collection of regression  
coefficients 587

**Beyond the Basics: Regression trees 588**

**Section 11.2 Summary 590**

**Section 11.2 Exercises 590**

**Chapter 11 Exercises 593**

## **CHAPTER 12 One-Way Analysis of Variance 598**

**Introduction 598**

### **12.1 Inference for One-Way Analysis of Variance 599**

The one-way ANOVA setting 600  
Comparing means 601  
The two-sample  $t$  statistic 602  
An overview of ANOVA 603  
The ANOVA model 606  
Estimates of population parameters 608  
Testing hypotheses in one-way ANOVA 610  
The ANOVA table 613  
The  $F$  test 616  
Software 618

**Beyond the Basics: Testing the equality of spread 619**

**Section 12.1 Summary 620**

**Section 12.1 Exercises 621**

### **12.2 Comparing the Means 624**

Contrasts 624  
Multiple comparisons 630  
Simultaneous confidence intervals 634  
Power of the one-way ANOVA  $F$  test 635

**Section 12.2 Summary 637**

**Section 12.2 Exercises 638**

**Chapter 12 Exercises 640**

## **CHAPTER 13 Two-Way Analysis of Variance 650**

**Introduction 650**

### **13.1 The Two-Way ANOVA Model 651**

Advantages of two-way ANOVA 651  
The two-way ANOVA model 655  
Main effects and interactions 656

**Section 13.1 Summary 661**

**Section 13.1 Exercises 661**

### **13.2 Inference for Two-Way ANOVA 663**

The two-way ANOVA table 664  
Carrying out a two-way ANOVA 665

**Section 13.2 Summary 669**

**Section 13.2 Exercises 669**

**Chapter 13 Exercises 671**

## **Companion Chapters**

(available online through Achieve)

## **CHAPTER 14 Logistic Regression 14-1**

**Introduction 14-1**

### **14.1 The Logistic Regression Model 14-2**

Binomial distributions and odds 14-2  
Odds for two groups 14-3  
Model for logistic regression 14-5  
Fitting and interpreting the logistic regression  
model 14-7

**Section 14.1 Summary 14-9**

**Section 14.1 Exercises 14-10**

### **14.2 Inference for Logistic Regression 14-12**

Confidence intervals and significance  
tests 14-12

Inference for multiple logistic regression 14-18

**Section 14.2 Summary 14-19**

**Section 14.2 Exercises 14-20**

**Chapter 14 Exercises 14-21**

---

**CHAPTER 15 Nonparametric Rank Tests 15-1****Introduction 15-1****15.1 The Wilcoxon Rank Sum Test 15-3**

- The rank transformation 15-4
- The Wilcoxon rank sum test 15-5
- The Normal approximation 15-7
- What hypotheses does Wilcoxon test? 15-9
- Ties 15-10
- Nonparametric rank and  $t$  procedures 15-12

**Section 15.1 Summary 15-13****Section 15.1 Exercises 15-14****15.2 The Wilcoxon Signed Rank Test 15-15**

- The Normal approximation 15-20
- Ties 15-20
- Testing a hypothesis about the median of a distribution 15-22

**Section 15.2 Summary 15-22****Section 15.2 Exercises 15-23****15.3 The Kruskal-Wallis Test 15-25**

- Hypotheses and assumptions 15-26
- The Kruskal-Wallis test 15-26

**Section 15.3 Summary 15-28****Section 15.3 Exercises 15-29****Chapter 15 Exercises 15-31**

---

**CHAPTER 16 Bootstrap Methods and Permutation Tests 16-1****Introduction 16-1**

- Software 16-2

**16.1 The Bootstrap Idea 16-2**

- The big idea: Resampling and the bootstrap distribution 16-3
- Thinking about the bootstrap idea 16-7
- Using software 16-8

**Section 16.1 Summary 16-9****Section 16.1 Exercises 16-9****16.2 First Steps in Using the Bootstrap 16-12**

- Bootstrap  $t$  confidence intervals 16-12
- Bootstrapping to compare two groups 16-16

**Beyond the Basics: The bootstrap for a scatterplot smoother 16-19****Section 16.2 Summary 16-20****Section 16.2 Exercises 16-21****16.3 How Accurate Is a Bootstrap Distribution? 16-22**

- Bootstrapping small samples 16-24
- Bootstrapping a sample median 16-26

**Section 16.3 Summary 16-27****Section 16.3 Exercises 16-28****16.4 Bootstrap Confidence Intervals 16-28**

- Bootstrap percentile confidence intervals 16-29
- A more accurate bootstrap confidence interval: BCa 16-30
- Confidence intervals for the correlation 16-32

**Section 16.4 Summary 16-35****Section 16.4 Exercises 16-35****16.5 Significance Testing Using Permutation Tests 16-38**

- Using software 16-41
- Permutation tests in practice 16-42
- Permutation tests in other settings 16-44

**Section 16.5 Summary 16-47****Section 16.5 Exercises 16-47****Chapter 16 Exercises 16-50**

---

**CHAPTER 17 Statistics for Quality: Control and Capability 17-1****Introduction 17-1****17.1 Processes and Statistical Process Control 17-2**

- Describing processes 17-3
- Statistical process control 17-5
- $\bar{x}$  charts for process monitoring 17-7
- $s$  charts for process monitoring 17-11

**Section 17.1 Summary 17-15****Section 17.1 Exercises 17-15****17.2 Using Control Charts 17-19**

- $\bar{x}$  and  $R$  charts 17-19
- Additional out-of-control rules 17-20
- Setting up control charts 17-22
- Comments on statistical control 17-27
- Don't confuse control with capability! 17-29

**Section 17.2 Summary 17-30****Section 17.2 Exercises 17-31****17.3 Process Capability Indexes 17-35**

- The capability indexes  $C_p$  and  $C_{pk}$  17-37
- Cautions about capability indexes 17-39

**Section 17.3 Summary 17-41****Section 17.3 Exercises 17-41****17.4 Control Charts for Sample Proportions 17-44**

- Control limits for  $p$  charts 17-45

**Section 17.4 Summary 17-48****Section 17.4 Exercises 17-49****Chapter 17 Exercises 17-50**

Tables T-1

Answers to Odd-Numbered Exercises A-1

Notes and Data Sources N-1

Index I-1