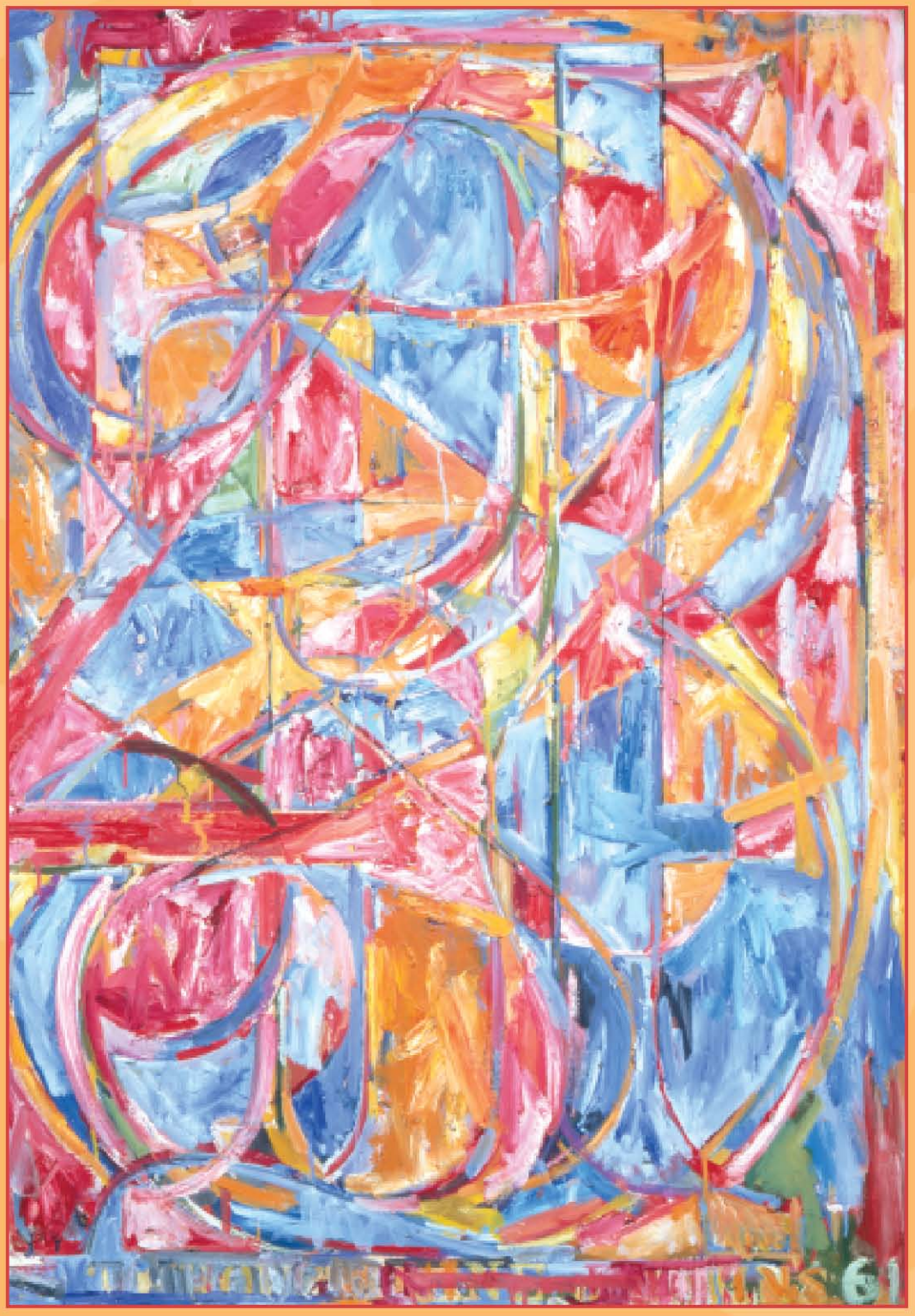
INTRODUCTION TO THE PRACTICE OF STATISTICS



SIXTH EDITION

MOORE
McCABE
CRAIG

Authors' note about the cover

Introduction to the Practice of Statistics emphasizes the use of graphical and numerical summaries to understand data. The front cover shows a painting entitled *0 to 9,* by the American artist Jasper Johns in 1961. In this work, the structure of the painting is determined by number sequence, just as our graphical summaries are determined by the numerical calculations that we perform when we analyze data. Can you find all of the digits in the painting?

Introduction to the Practice of Statistics

Senior Publisher:	CRAIG BLEYER
Publisher:	RUTH BARUTH
Development Editors:	SHONA BURKE, ANNE SCANLAN-ROHRER
Senior Media Editor:	ROLAND CHEYNEY
Assistant Editor:	BRIAN TEDESCO
Editorial Assistant:	KATRINA WILHELM
Marketing Coordinator:	DAVE QUINN
Photo Editor:	CECILIA VARAS
Photo Researcher:	ELYSE RIEDER
Cover and Text Designer:	VICKI TOMASELLI
Senior Project Editor:	MARY LOUISE BYRD
Illustrations:	INTEGRE TECHNICAL PUBLISHING CO.
Production Manager:	JULIA DE ROSA
Composition:	INTEGRE TECHNICAL PUBLISHING CO.
Printing and Binding:	RR DONNELLEY

TI-83™ screen shots are used with permission of the publisher: ©1996, Texas Instruments Incorporated. TI-83™ Graphic Calculator is a registered trademark of Texas Instruments Incorporated. Minitab is a registered trademark of Minitab, Inc. Microsoft © and Windows © are registered trademarks of the Microsoft Corporation in the United States and other countries. Excel screen shots are reprinted with permission from the Microsoft Corporation. S-PLUS is a registered trademark of the Insightful Corporation. SAS© is a registered trademark of SAS Institute, Inc. *CrunchIt!* is a trademark of Integrated Analytics LLC.

Library of Congress Control Number: 2007938575

ISBN-13: 978-1-4292-1623-4
ISBN-10: 1-4292-1623-9 (Extended Version, Casebound)

ISBN-13: 978-1-4292-1622-7
ISBN-10: 1-4292-1622-0 (Casebound)

ISBN-13: 978-1-4292-1621-0
ISBN-10: 1-4292-1621-2 (Paperback)

© 2009 by W. H. Freeman and Company. All rights reserved.

Printed in the United States of America

First printing

W. H. Freeman and Company
41 Madison Avenue
New York, NY 10010
Houndmills, Basingstoke RG21 6XS, England
www.whfreeman.com

Introduction to the Practice of Statistics

SIXTH EDITION

DAVID S. MOORE

GEORGE P. McCABE

BRUCE A. CRAIG

Purdue University

W. H. Freeman and Company
New York

This page intentionally left blank

Brief Contents

To Teachers: *About This Book* xv
To Students: *What Is Statistics?* xxv
About the Authors xxix
Data Table Index xxxi
Beyond the Basics Index xxxiii

PART I Looking at Data

CHAPTER 1: Looking at Data—Distributions 1

CHAPTER 2: Looking at Data—Relationships 83

CHAPTER 3: Producing Data 171

PART II Probability and Inference

CHAPTER 4: Probability: The Study of Randomness 237

CHAPTER 5: Sampling Distributions 311

CHAPTER 6: Introduction to Inference 353

CHAPTER 7: Inference for Distributions 417

CHAPTER 8: Inference for Proportions 487

PART III Topics in Inference

CHAPTER 9: Analysis of Two-Way Tables 525

CHAPTER 10: Inference for Regression 559

CHAPTER 11: Multiple Regression 607

CHAPTER 12: One-Way Analysis of Variance 637

CHAPTER 13: Two-Way Analysis of Variance 683

Companion Chapters (on the IPS Web site **www.whfreeman.com/ips6e** and CD-ROM)

CHAPTER 14: Logistic Regression 14-1

CHAPTER 15: Nonparametric Tests 15-1

CHAPTER 16: Bootstrap Methods and Permutation Tests 16-1

CHAPTER 17: Statistics for Quality: Control and Capability 17-1

Data Appendix D-1
Tables T-1
Answers to Odd-Numbered Exercises A-1
Notes and Data Sources N-1
Photo Credits C-1
Index I-1

This page intentionally left blank

Contents

To Teachers: *About This Book* xv
To Students: *What Is Statistics?* xxv
About the Authors xxix
Data Table Index xxxi
Beyond the Basics Index xxxiii

PART I Looking at Data

CHAPTER 1
Looking at Data—Distributions 1

Introduction 1
Variables 2
Measurement: know your variables 5

1.1 Displaying Distributions with Graphs 6
Graphs for categorical variables 6
Data analysis in action: don't hang up on me 7
Stemplots 9
Histograms 12
Examining distributions 15
Dealing with outliers 17
Time plots 18
Beyond the basics: decomposing time series* 19
Section 1.1 Summary 22
Section 1.1 Exercises 22

1.2 Describing Distributions with Numbers 30
Measuring center: the mean 30
Measuring center: the median 32
Mean versus median 34
Measuring spread: the quartiles 34
The five-number summary and boxplots 36
The 1.5 × *IQR* rule for suspected outliers 38
Measuring spread: the standard deviation 40
Properties of the standard deviation 42
Choosing measures of center and spread 43
Changing the unit of measurement 45
Section 1.2 Summary 47
Section 1.2 Exercises 48

1.3 Density Curves and Normal Distributions 53
Density curves 55
Measuring center and spread for density curves 56
Normal distributions 58
The 68–95–99.7 rule 59
Standardizing observations 61
Normal distribution calculations 62
Using the standard Normal table 64
Inverse Normal calculations 66
Normal quantile plots 68
Beyond the basics: density estimation 71
Section 1.3 Summary 71
Section 1.3 Exercises 72
Chapter 1 Exercises 78

CHAPTER 2
Looking at Data—Relationships 83

Introduction 83
Examining relationships 84

2.1 Scatterplots 86
Interpreting scatterplots 88
Adding categorical variables to scatterplots 89
More examples of scatterplots 90
Beyond the basics: scatterplot smoothers 92
Categorical explanatory variables 93
Section 2.1 Summary 94
Section 2.1 Exercises 95

2.2 Correlation 101
The correlation *r* 102
Properties of correlation 102
Section 2.2 Summary 105
Section 2.2 Exercises 105

2.3 Least-Squares Regression 108
Fitting a line to data 110
Prediction 111
Least-squares regression 112
Interpreting the regression line 115

Sections marked with an asterisk are optional.

Correlation and regression 115
*Understanding r^2 118
Beyond the basics: transforming relationships 119
Section 2.3 Summary 121
Section 2.3 Exercises 122

2.4 Cautions about Correlation and Regression 125
Residuals 126
Outliers and influential observations 129
Beware the lurking variable 132
Beware correlations based on averaged data 135
The restricted-range problem 135
Beyond the basics: data mining 136
Section 2.4 Summary 137
Section 2.4 Exercises 137

2.5 Data Analysis for Two-Way Tables 142
The two-way table 142
Joint distribution 144
Marginal distributions 145
Describing relations in two-way tables 146
Conditional distributions 146
Simpson's paradox 148
The perils of aggregation 151
Section 2.5 Summary 151
Section 2.5 Exercises 152

2.6 The Question of Causation 154
Explaining association: causation 154
Explaining association: common response 155
Explaining association: confounding 156
Establishing causation 157
Section 2.6 Summary 159
Section 2.6 Exercises 159
Chapter 2 Exercises 161

CHAPTER 3
Producing Data 171

Introduction 171
Anecdotal data 171
Available data 172
Sample surveys and experiments 173

3.1 Design of Experiments 178
Comparative experiments 180
Randomization 181
Randomized comparative experiments 183
How to randomize 184
Cautions about experimentation 188
Matched pairs designs 189
Block designs 190
Section 3.1 Summary 191
Section 3.1 Exercises 192

3.2 Sampling Design 197
Simple random samples 200
Stratified samples 202
Multistage samples 203
Cautions about sample surveys 204
Section 3.2 Summary 207
Section 3.2 Exercises 208

3.3 Toward Statistical Inference 212
Sampling variability 213
Sampling distributions 214
Bias and variability 217
Sampling from large populations 219
Why randomize? 219
Beyond the basics: capture-recapture sampling 220
Section 3.3 Summary 221
Section 3.3 Exercises 221

3.4 Ethics 224
Institutional review boards 225
Informed consent 226
Confidentiality 227
Clinical trials 228
Behavioral and social science experiments 230
Section 3.4 Summary 232
Section 3.4 Exercises 232
Chapter 3 Exercises 234

PART II Probability and Inference

CHAPTER 4
Probability: The Study of Randomness 237

Introduction 237

4.1 Randomness 237
The language of probability 239
Thinking about randomness 240

The uses of probability 241
Section 4.1 Summary 241
Section 4.1 Exercises 241

4.2 Probability Models 242
Sample spaces 243
Probability rules 245
Assigning probabilities: finite number of outcomes 248
Assigning probabilities: equally likely outcomes 249
Independence and the multiplication rule 251
Applying the probability rules 254
Section 4.2 Summary 255
Section 4.2 Exercises 255

4.3 Random Variables 258
Discrete random variables 259
Continuous random variables 263
Normal distributions as probability distributions 265
Section 4.3 Summary 267
Section 4.3 Exercises 267

4.4 Means and Variances of Random Variables 270
The mean of a random variable 270
Statistical estimation and the law of large numbers 273
Thinking about the law of large numbers 275
Beyond the basics: more laws of large numbers 277
Rules for means 277
The variance of a random variable 279
Rules for variances and standard deviations 281
Section 4.4 Summary 285
Section 4.4 Exercises 286

4.5 General Probability Rules* 289
General addition rules 290
Conditional probability 293
General multiplication rules 298
Tree diagrams 299
Bayes's rule 301
Independence again 302
Section 4.5 Summary 302
Section 4.5 Exercises 303
Chapter 4 Exercises 307

CHAPTER 5
Sampling Distributions 311

Introduction 311

5.1 Sampling Distributions for Counts and Proportions 313
The binomial distributions for sample counts 314
Binomial distributions in statistical sampling 315
Finding binomial probabilities: software and tables 316
Binomial mean and standard deviation 319
Sample proportions 321
Normal approximation for counts and proportions 322
The continuity correction* 326
Binomial formula* 327
Section 5.1 Summary 330
Section 5.1 Exercises 331

5.2 The Sampling Distribution of a Sample Mean 335
The mean and standard deviation of $\bar{x}$ 337
The central limit theorem 339
A few more facts 343
Beyond the basics: Weibull distributions 344
Section 5.2 Summary 346
Section 5.2 Exercises 346
Chapter 5 Exercises 350

CHAPTER 6
Introduction to Inference 353

Introduction 353
Overview of inference 354

6.1 Estimating with Confidence 356
Statistical confidence 356
Confidence intervals 358
Confidence interval for a population mean 360
How confidence intervals behave 363
Choosing the sample size 364
Some cautions 366
Beyond the basics: the bootstrap 368
Section 6.1 Summary 368
Section 6.1 Exercises 369

6.2 Tests of Significance 372
The reasoning of significance tests 372

Stating hypotheses 374
Test statistics 376
P-values 377
Statistical significance 379
Tests for a population mean 382
Two-sided significance tests and confidence intervals 386
P-values versus fixed α 388
Section 6.2 Summary 390
Section 6.2 Exercises 390

6.3 Use and Abuse of Tests 394

Choosing a level of significance 395
What statistical significance does not mean 396
Don't ignore lack of significance 397
Statistical inference is not valid for all sets of data 398
Beware of searching for significance 398
Section 6.3 Summary 399
Section 6.3 Exercises 399

6.4 Power and Inference as a Decision* 401

Power 401
Increasing the power 405
Inference as decision* 406
Two types of error 406
Error probabilities 407
The common practice of testing hypotheses 409
Section 6.4 Summary 410
Section 6.4 Exercises 410
Chapter 6 Exercises 412

CHAPTER 7
Inference for Distributions 417

Introduction 417

7.1 Inference for the Mean of a Population 418

The *t* distributions 418
The one-sample *t* confidence interval 420
The one-sample *t* test 422
Matched pairs *t* procedures 428
Robustness of the *t* procedures 432
The power of the *t* test* 433
Inference for non-Normal populations* 435
Section 7.1 Summary 440
Section 7.1 Exercises 441

7.2 Comparing Two Means 447

The two-sample *z* statistic 448
The two-sample *t* procedures 450
The two-sample *t* significance test 451
The two-sample *t* confidence interval 454
Robustness of the two-sample procedures 456
Inference for small samples 457
Software approximation for the degrees of freedom* 460
The pooled two-sample *t* procedures* 461
Section 7.2 Summary 466
Section 7.2 Exercises 467

7.3 Optional Topics in Comparing Distributions* 473

Inference for population spread 473
The *F* test for equality of spread 474
Robustness of Normal inference procedures 476
The power of the two-sample *t* test 477
Section 7.3 Summary 479
Section 7.3 Exercises 479
Chapter 7 Exercises 481

CHAPTER 8
Inference for Proportions 487

Introduction 487

8.1 Inference for a Single Proportion 488

Large-sample confidence interval for a single proportion 488
Beyond the basics: the plus four confidence interval for a single proportion 491
Significance test for a single proportion 493
Confidence intervals provide additional information 496
Choosing a sample size 498
Section 8.1 Summary 501
Section 8.1 Exercises 502

8.2 Comparing Two Proportions 505

Large-sample confidence interval for a difference in proportions 506
Beyond the basics: plus four confidence interval for a difference in proportions 509
Significance test for a difference in proportions 511
Beyond the basics: relative risk 515
Section 8.2 Summary 516
Section 8.2 Exercises 517
Chapter 8 Exercises 519

PART III Topics in Inference

CHAPTER 9
Analysis of Two-Way Tables 525

Introduction 525

9.1 Inference for Two-Way Tables 526
- The hypothesis: no association 529
- Expected cell counts 529
- The chi-square test 530
- The chi-square test and the z test 533
- Beyond the basics: meta-analysis 534
- Section 9.1 Summary 536

9.2 Formulas and Models for Two-Way Tables* 536
- Computations 536
- Computing conditional distributions 537
- Computing expected cell counts 540
- The X^2 statistic and its P-value 540
- Models for two-way tables 541
- Concluding remarks 544
- Section 9.2 Summary 545

9.3 Goodness of Fit* 545
- Section 9.3 Summary 548
- Chapter 9 Exercises 548

CHAPTER 10
Inference for Regression 559

Introduction 559

10.1 Simple Linear Regression 560
- Statistical model for linear regression 560
- Data for simple linear regression 561
- Estimating the regression parameters 565
- Confidence intervals and significance tests 570
- Confidence intervals for mean response 572
- Prediction intervals 574
- Beyond the basics: nonlinear regression 576
- Section 10.1 Summary 578

10.2 More Detail about Simple Linear Regression* 579
- Analysis of variance for regression 579
- The ANOVA F test 581
- Calculations for regression inference 583
- Inference for correlation 590
- Section 10.2 Summary 593
- Chapter 10 Exercises 594

CHAPTER 11
Multiple Regression 607

Introduction 607

11.1 Inference for Multiple Regression 607
- Population multiple regression equation 607
- Data for multiple regression 608
- Multiple linear regression model 609
- Estimation of the multiple regression parameters 610
- Confidence intervals and significance tests for regression coefficients 611
- ANOVA table for multiple regression 612
- Squared multiple correlation R^2 613

11.2 A Case Study 615
- Preliminary analysis 615
- Relationships between pairs of variables 616
- Regression on high school grades 618
- Interpretation of results 619
- Residuals 620
- Refining the model 621
- Regression on SAT scores 622
- Regression using all variables 623
- Test for a collection of regression coefficients 623
- Beyond the basics: multiple logistic regression 625
- Chapter 11 Summary 627
- Chapter 11 Exercises 628

CHAPTER 12
One-Way Analysis of Variance 637

Introduction 637

12.1 Inference for One-Way Analysis of Variance 638
- Data for one-way ANOVA 638
- Comparing means 639
- The two-sample t statistic 640
- An overview of ANOVA 641
- The ANOVA model 644
- Estimates of population parameters 646
- Testing hypotheses in one-way ANOVA 648
- The ANOVA table 649
- The F test 652

12.2 Comparing the Means 655
Contrasts 655
Multiple comparisons 661
Software 665
Power* 666
Section 12.2 Summary 669
Chapter 12 Exercises 670

CHAPTER 13
Two-Way Analysis of Variance 683

Introduction 683

13.1 The Two-Way ANOVA Model 684
Advantages of two-way ANOVA 684
The two-way ANOVA model 688
Main effects and interactions 689

13.2 Inference for Two-Way ANOVA 694
The ANOVA table for two-way ANOVA 694
Section 13.2 Summary 698
Chapter 13 Exercises 699

Companion Chapters (on the IPS Web site **www.whfreeman.com/ips6e** and CD-ROM)

CHAPTER 14
Logistic Regression 14-1

Introduction 14-1

14.1 The Logistic Regression Model 14-1
Binomial distributions and odds 14-2
Odds for two samples 14-3
Model for logistic regression 14-4
Fitting and interpreting the logistic regression model 14-6

14.2 Inference for Logistic Regression 14-8
Confidence intervals and significance tests 14-8
Multiple logistic regression 14-14
Section 14.2 Summary 14-16
Chapter 14 Exercises 14-17
Chapter 14 Notes 14-23

CHAPTER 15
Nonparametric 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? 15-8
Ties 15-10
Rank, *t*, and permutation tests 15-12
Section 15.1 Summary 15-14
Section 15.1 Exercises 15-15

15.2 The Wilcoxon Signed Rank Test 15-17
The Normal approximation 15-20
Ties 15-21
Section 15.2 Summary 15-23
Section 15.2 Exercises 15-23

15.3 The Kruskal-Wallis Test* 15-26
Hypotheses and assumptions 15-28
The Kruskal-Wallis test 15-28
Section 15.3 Summary 15-30
Section 15.3 Exercises 15-31
Chapter 15 Exercises 15-33
Chapter 15 Notes 15-35

CHAPTER 16
Bootstrap Methods and Permutation Tests 16-1

Introduction 16-1
Software 16-2

16.1 The Bootstrap Idea 16-3
The big idea: resampling and the bootstrap distribution 16-4
Thinking about the bootstrap idea 16-9
Using software 16-10
Section 16.1 Summary 16-11
Section 16.1 Exercises 16-11

16.2 First Steps in Using the Bootstrap 16-13
Bootstrap *t* confidence intervals 16-13
Bootstrapping to compare two groups 16-17
Beyond the basics: the bootstrap for a scatterplot smoother 16-20
Section 16.2 Summary 16-21
Section 16.2 Exercises 16-22

16.3 How Accurate Is a Bootstrap Distribution?* 16-24
Bootstrapping small samples 16-26
Bootstrapping a sample median 16-28
Section 16.3 Summary 16-28
Section 16.3 Exercises 16-30

16.4 Bootstrap Confidence Intervals 16-30
Bootstrap percentile confidence intervals 16-31
More accurate bootstrap confidence intervals: BCa and tilting 16-32
Confidence intervals for the correlation 16-35
Section 16.4 Summary 16-38
Section 16.4 Exercises 16-38

16.5 Significance Testing Using Permutation Tests 16-41
Using software 16-45
Permutation tests in practice 16-45
Permutation tests in other settings 16-49
Section 16.5 Summary 16-52
Section 16.5 Exercises 16-52
Chapter 16 Exercises 16-56
Chapter 16 Notes 16-59

CHAPTER 17
Statistics for Quality: Control and Capability 17-1

Introduction 17-1
Use of data to assess quality 17-2

17.1 Processes and Statistical Process Control 17-3
Describing processes 17-3
Statistical process control 17-6
$\bar{x}$ charts for process monitoring 17-8
s charts for process monitoring 17-12
Section 17.1 Summary 17-17
Section 17.1 Exercises 17-17

17.2 Using Control Charts 17-21
$\bar{x}$ and R charts 17-22
Additional out-of-control rules 17-23
Setting up control charts 17-25
Comments on statistical control 17-30
Don't confuse control with capability! 17-33
Section 17.2 Summary 17-34
Section 17.2 Exercises 17-35

17.3 Process Capability Indexes* 17-39
The capability indexes C_p and C_{pk} 17-41
Cautions about capability indexes 17-44
Section 17.3 Summary 17-45
Section 17.3 Exercises 17-46

17.4 Control Charts for Sample Proportions 17-49
Control limits for p charts 17-50
Section 17.4 Summary 17-54
Section 17.4 Exercises 17-54
Chapter 17 Exercises 17-56
Chapter 17 Notes 17-57

Data Appendix D-1
Tables T-1
Answers to Odd-Numbered Exercises A-1
Notes and Data Sources N-1
Photo Credits C-1
Index I-1

This page intentionally left blank

TO TEACHERS: *About This Book*

Statistics is the science of data. *Introduction to the Practice of Statistics* (*IPS*) is an introductory text based on this principle. We present the most-used methods of basic statistics in a way that emphasizes working with data and mastering statistical reasoning. *IPS* is elementary in mathematical level but conceptually rich in statistical ideas and serious in its aim to help students think about data and use statistical methods with understanding.

Some schematic history will help place *IPS* in the universe of texts for a first course in statistics for students from a variety of disciplines. Traditional texts were almost entirely devoted to methods of inference, with quick coverage of means, medians, and histograms as a preliminary. No doubt this reflected the fact that inference is the only part of statistics that has a mathematical theory behind it. Several innovative books aimed at nontraditional audiences pioneered a broader approach that paid more attention to design of samples and experiments, the messiness of real data, and discussion of real-world statistical studies and controversies. All were written by widely known statisticians whose main business was not writing textbooks. *The Nature of Statistics* (Wallis and Roberts) has passed away, but *Statistics* (Freedman and collaborators) and *Statistics: Concepts and Controversies* (Moore) remain alive and well. None of these books tried to meet the needs of a typical first course because their audiences did not need full coverage of standard statistical methods.

IPS was the first book to successfully combine attention to broader content and reasoning with comprehensive presentation of the most-used statistical methods. It reflects the consensus among statisticians—even stronger now than when the first edition appeared—concerning the content of an introduction to our discipline. This consensus is expressed in a report from the joint curriculum committee of the American Statistical Association and the Mathematical Association of America[1] and in discussions in leading journals.[2] *IPS* has been successful for several reasons:

1. *IPS* examines the nature of modern statistical practice at a level suitable for beginners. Attention to data analysis and data production as well as to probability and inference is "new" only in the world of textbooks. Users of statistical methods have always paid attention to all of these. Contemporary research in statistics, driven by advances in computing, puts more emphasis on sophisticated "looking at data" and on data-analytic ways of thinking. Formal inference remains important and receives careful treatment, but it appears as part of a larger picture.

2. *IPS* has a logical overall progression, so data analysis and data production strengthen the presentation of inference rather than stand apart from it. We stress that data analysis is an essential preliminary to inference because inference requires clean data. The most useful "goodness of fit" procedure, for example, is the normal quantile plot presented in Chapter 1 and used frequently in the inference chapters. We emphasize that when you do formal statistical inference, you are acting as if your data come from properly randomized data production. We use random samples and experimental randomization to motivate the need for probability as a language for inference.

3. *IPS* presents data analysis as more than a collection of techniques for exploring data. We integrate techniques with discussion of systematic ways of thinking about data. We also work hard to make data-analytic thinking accessible to beginners by presenting a series of simple principles: always plot your data; look for overall patterns and deviations from them; when looking at the overall pattern of a distribution for one variable, consider shape, center, and spread; for relations between two variables, consider form, direction, and strength; always ask whether a relationship between variables is influenced by other variables lurking in the background. Inference is similarly treated as more than a collection of methods. We warn students about pitfalls in clear cautionary discussions—about regression and correlation, experiments, sample surveys, confidence intervals, and significance tests. Our goal throughout *IPS* is to present principles and techniques together in a way that is accessible to beginners and lays a foundation for students who will go on to more advanced study.
4. *IPS* integrates discussion of techniques, reasoning, and practice using real examples to drive the exposition. Students learn the technique of least-squares regression and how to interpret the regression slope. But they also learn the conceptual ties between regression and correlation, the importance of looking for influential observations (always plot your data), and to beware of averaged data and the restricted-range effect.
5. *IPS* is aware of current developments both in statistical science and in teaching statistics. For example, the first edition already favored the version of the two-sample t procedures that does not assume equal population variances and discussed the great difference in robustness between standard tests for means and for variances. In the fourth edition, we introduced the modified ("plus four") confidence intervals for proportions that are shown by both computational studies[3] and theory[4] to be superior to the standard intervals for all but very large samples. Brief optional "Beyond the Basics" sections give quick overviews of topics such as density estimation, scatterplot smoothers, nonlinear regression, and data mining. Chapter 16 on resampling methods offers an extended introduction to one of the most important recent advances in statistical methodology.

The title of the book expresses our intent to introduce readers to statistics as it is used in practice. Statistics in practice is concerned with gaining understanding from data; it focuses on problem solving rather than on methods that may be useful in specific settings. A text cannot fully imitate practice because it must teach specific methods in a logical order and must use data that are not the reader's own. Nonetheless, our interest and experience in applying statistics have influenced the nature of *IPS* in several ways.

Statistical Thinking Statistics is interesting and useful because it provides strategies and tools for using data to gain insight into real problems. As the continuing revolution in computing automates the details of doing calculations and making graphs, an emphasis on statistical concepts and on insight from data becomes both more practical for students and teachers and more important for users who must supply what is not automated. No student should complete a first statistics course, for example, without a firm grasp of the distinction between observational studies and experiments and of why randomized comparative experiments are the gold standard for evidence of causation.

We have seen many statistical mistakes, but few have involved simply getting a calculation wrong. We therefore ask students to learn to explore data, always starting with plots, to think about the context of the data and the design of the study that produced the data, the possible influence of wild observations on conclusions, and the reasoning that lies behind standard methods of inference. Users of statistics who form these habits from the beginning are well prepared to learn and use more advanced methods.

Data Data are numbers with a context, as we say in "To Students: What Is Statistics?" A newborn who weighs 10.3 pounds is a big baby, and the birth weight could not plausibly be 10.3 ounces or 10.3 kilograms. Because context makes numbers meaningful, our examples and exercises use real data with real contexts that we briefly describe. Calculating the mean of five numbers is arithmetic, not statistics. We hope that the presence of background information, even in exercises intended for routine drill, will encourage students to always consider the meaning of their calculations as well as the calculations themselves. Note in this connection that a calculation or a graph or "reject H_0" is rarely a full answer to a statistical problem. We strongly encourage requiring students always to state a brief conclusion in the context of the problem. This helps build data sense as well as the communication skills that employers value.

Mathematics Although statistics is a mathematical science, it is not a field of mathematics and should not be taught as if it were. A fruitful mathematical theory (based on probability, which *is* a field of mathematics) underlies some parts of basic statistics, but by no means all. The distinction between observation and experiment, for example, is a core statistical idea that is ignored by the theory.[5] Mathematically trained teachers, rightly resisting a formula-based approach, sometimes identify conceptual understanding with mathematical understanding. When teaching statistics, we must emphasize statistical ideas and recognize that mathematics is not the only vehicle for conceptual understanding. *IPS* requires only the ability to read and use equations without having each step parsed. We require no algebraic derivations, let alone calculus. Because this is a *statistics* text, it is richer in ideas and requires more thought than the low mathematical level suggests.

Calculators and Computers Statistical calculations and graphics are in practice automated by software. We encourage instructors to use software of their choice or a graphing calculator that includes functions for both data analysis and basic inference. *IPS* includes some topics that reflect the dominance of software in practice, such as normal quantile plots and the version of the two-sample t procedures that does not require equal variances. Several times we display the output of multiple software systems for the same problem. The point is that a student who knows the basics can interpret almost any output. Students like this reassurance, and it helps focus their attention on understanding rather than reading output.

Judgment Statistics in practice requires judgment. It is easy to list the mathematical assumptions that justify use of a particular procedure, but not so easy to decide when the procedure can be safely used in practice. Because judgment develops through experience, an introductory course should present clear guidelines and not make unreasonable demands on the judgment of

students. We have given guidelines—for example, on using the t procedures for comparing two means but avoiding the F procedures for comparing two variances—that we follow ourselves. Similarly, many exercises require students to use some judgment and (equally important) to explain their choices in words. Many students would prefer to stick to calculating, and many statistics texts allow them to. Requiring more will do them much good in the long run.

Teaching Experiences We have successfully used *IPS* in courses taught to quite diverse student audiences. For general undergraduates from mixed disciplines, we cover Chapters 1 to 8 and Chapter 9, 10, or 12, omitting all optional material. For a quantitatively strong audience—sophomores planning to major in actuarial science or statistics—we move more quickly. We add Chapters 10 and 11 to the core material in Chapters 1 to 8 and include most optional content. We de-emphasize Chapter 4 (probability) because these students will take a probability course later in their program, though we make intensive use of software for simulating probabilities as well as for statistical analysis. The third group we teach contains beginning graduate students in such fields as education, family studies, and retailing. These mature but sometimes quantitatively unprepared students read the entire text (Chapters 11 and 13 lightly), again with reduced emphasis on Chapter 4 and some parts of Chapter 5. In all cases, beginning with data analysis and data production (Part I) helps students overcome their fear of statistics and builds a sound base for studying inference. We find that *IPS* can be flexibly adapted to quite varied audiences by paying attention to our clear designation of some material as optional and by varying the chapters assigned.

The Sixth Edition: *What's New?*

- **Co-author** We are delighted to welcome Professor Bruce Craig to the *Introduction to the Practice of Statistics* author team. Bruce is currently Director of the Statistical Consulting Service at Purdue University and is an outstanding teacher. His vast experience consulting and collaborating with individuals who use statistical methods in their work provides him with perspective on the field of statistics that resonates with the approach of this text.
- **Ethics** Chapter 3 now contains a new section (3.4) on ethics. We believe that this topic is a very important part of the undergraduate curriculum and that a course in statistics is an ideal forum to stimulate thought and discussion about ethical issues.
- **Text Organization** Logistic Regression, previously treated in Chapter 16, now appears in Chapter 14. Similarly, Bootstrap Methods and Permutation Tests has moved to Chapter 16. This change is in line with the increasing importance of logistic regression in statistical practice. In response to suggestions from current *IPS* users, we have moved the material on data analysis for two-way tables from Chapter 9 back to Chapter 2 (Section 2.5). In addition, the large sample confidence procedures are now the featured methods for one and two proportions in Chapter 9, and the plus-four have been moved to Beyond the Basics sections, a more appropriate location. The table of contents follows what we consider to be the best ordering of the topics from a

pedagogical point of view. However, the text chapters are generally written to enable instructors to teach the material in the order they prefer.

- **Design** A new design incorporates colorful, revised figures throughout to aid students' understanding of text material. Photographs related to chapter examples and exercises make connections to real-life applications and provide a visual context for topics.
- **Exercises and Examples** Exercises and examples are labeled to help instructors and students easily identify key topics and application areas. The number of total exercises has increased by 15%. Approximately half the total exercises are new or revised to reflect current data and a variety of topics. *IPS* examples and exercises cover a wide range of application areas. An application index is provided for instructors to easily select and assign content related to specific fields.
- **Use Your Knowledge Exercises** Short exercises designed to reinforce key concepts now appear throughout each chapter. These exercises are listed, with page numbers, at the end of each section for easy reference.

- **Look Back** At key points in the text Look Back margin notes direct the reader to the first explanation of a topic, providing page numbers for easy reference.

In addition to the new Sixth Edition enhancements, *IPS* has retained the successful pedagogical features from previous editions:

- **Caution** Warnings in the text, signaled by a caution icon, help students avoid common errors and misconceptions.

- **Challenge Exercises** More challenging exercises are signaled with an icon. Challenge exercises are varied: some are mathematical, some require open-ended investigation, and so on.

- **Applets** Applet icons are used throughout the text to signal where related, interactive statistical applets can be found on the text Web site (www.whfreeman.com/ips6e) and CD-ROM.
- **Statistics in Practice** Formerly found at the opening of each chapter, these accounts by professionals who use statistics on the job are now located on the *IPS* Web site and CD-ROM.
- **CrunchIt! Statistical Software** Developed by Webster West of Texas A&M University, CrunchIt! is an easy-to-use program for students and offers capabilities well beyond those needed for a first course. CrunchIt! output, along with other statistical software output, is integrated throughout the text. Access to CrunchIt! is available online through an access-code–protected Web site. Access codes are available in every new copy of *IPS* 6e or can be purchased online.

Acknowledgments

We are pleased that the first five editions of *Introduction to the Practice of Statistics* have helped move the teaching of introductory statistics in a direction supported by most statisticians. We are grateful to the many colleagues and students who have provided helpful comments, and we hope that they will

find this new edition another step forward. In particular, we would like to thank the following colleagues who offered specific comments on the new edition:

Mary A. Bergs
Mercy College of Northwest Ohio
John F. Brewster
University of Manitoba
Karen Buro
Grant MacEwan College
Smiley W. Cheng
University of Manitoba
Gerarda Darlington
University of Guelph
Linda Dawson
University of Washington, Tacoma
Michael Evans
University of Toronto
Mary Gray
American University
Rick Gumina
Colorado State University
Patricia Humphrey
Georgia Southern University
Mohammad Kazemi
The University of North Carolina at Charlotte
Jeff Kollath
Oregon State University
William J. Kubik
Hanover College
Charles Liberty
Keene State College
Brian Macpherson
University of Manitoba
Henry Mesa
Portland Community College
Helen Noble
San Diego State University
Richard Numrich
College of Southern Nevada
Becky Parker
Montana State University
Maria Consuelo C. Pickle
St. Petersburg College
German J. Pliego
University of St. Thomas
Philip Protter
Cornell University
John G. Reid
Mount Saint Vincent University
Diann Reischman
Grand Valley State University
Shane Rollans
Thompson Rivers University
Nancy Roper
Portland Community College
Teri Rysz
University of Cincinnati
Engin Sungur
University of Minnesota, Morris
Todd Swanson
Hope College
Anthony L. Truog
University of Wisconsin, Whitewater
Augustin Vukov
University of Toronto
Erwin Walker
Clemson University
Nathan Wetzel
University of Wisconsin, Stevens Point

The professionals at W. H. Freeman and Company, in particular Mary Louise Byrd, Ruth Baruth, and Shona Burke, have contributed greatly to the success of *IPS*. Additionally, we would like to thank Anne Scanlan-Rohrer, Pam Bruton, Jackie Miller, and Darryl Nester for their valuable contributions to the Sixth Edition. Most of all, we are grateful to the many people in varied disciplines and occupations with whom we have worked to gain understanding from data. They have provided both material for this book and the experience that enabled us to write it. What the eminent statistician John Tukey called "the real problems experience and the real data experience" have shaped our view of statistics, convincing us of the need for beginning instruction to focus on data and concepts, building intellectual skills that transfer to more elaborate settings and remain essential when all details are automated. We hope that users and potential users of statistical techniques will find this emphasis helpful.

Media and Supplements

For Students

NEW! STATS PORTAL

portals.bfwpub.com/ips6e (Access code required. Available packaged with *Introduction to the Practice of Statistics*, Sixth Edition, or for purchase online.) StatsPortal is the digital gateway to *IPS* 6e, designed to enrich the course and enhance students' study skills through a collection of Web-based tools. StatsPortal integrates a rich suite of diagnostic, assessment, tutorial, and enrichment features, enabling students to master statistics at their own pace. It is organized around three main teaching and learning components:

- **Interactive eBook** offers a complete and customizable online version of the text, fully integrated with all the media resources available with *IPS* 6e. The eBook allows students to quickly search the text, highlight key areas, and add notes about what they're reading. Similarly, instructors can customize the eBook to add, hide, and reorder content, add their own material, and highlight key text for students.
- **Resources** organizes all the resources for *IPS* 6e into one location for students' ease of use. These resources include the following:
  - **StatTutor Tutorials** offer over 150 audio-multimedia tutorials tied directly to the textbook, including videos, applets, and animations.
  - **Statistical Applets** are 16 interactive applets to help students master key statistical concepts.
  - **CrunchIt! Statistical Software** allows users to analyze data from any Internet location. Designed with the novice user in mind, the software is not only easily accessible but also easy to use. CrunchIt! offers all the basic statistical routines covered in the introductory statistics courses. **CrunchIt!** statistical software is available via an access-code protected Web site. Access codes are available in every new copy of *IPS* 6e or can be purchased online.
  - **Stats@Work Simulations** put students in the role of statistical consultants, helping them better understand statistics interactively within the context of real-life scenarios. Students are asked to interpret and analyze data presented to them in report form, as well as to interpret current event news stories. All tutorials are graded and offer helpful hints and feedback.
  - **EESEE Case Studies** developed by The Ohio State University Statistics Department provide students with a wide variety of timely, real examples with real data. Each case study is built around several thought-provoking questions that make students think carefully about the statistical issues raised by the stories. **EESEE** case studies are available via an access-code-protected Web site. Access codes are available in every new copy of *IPS* 6e or can be purchased online.

- **Podcast Chapter Summary** provides students with an audio version of chapter summaries to download and review on an mp3 player.
- **Data Sets** are available in ASCII, Excel, JMP, Minitab, TI, SPSS, and S-Plus formats.
- **Online Tutoring with SMARTHINKING** is available for homework help from specially trained, professional educators.
- **Student Study Guide with Selected Solutions** includes explanations of crucial concepts and detailed solutions to key text problems with step-by-step models of important statistical techniques.
- **Statistical Software Manuals** for TI-83/84, Minitab, Excel, JMP, and SPSS provide instruction, examples, and exercises using specific statistical software packages.
- **Interactive Table Reader** allows students to use statistical tables interactively to seek the information they need.

Resources (instructors only)

- **Instructor's Guide with Full Solutions** includes worked-out solutions to all exercises, teaching suggestions, and chapter comments.
- **Test Bank** contains complete solutions for textbook exercises.
- **Lecture PowerPoint slides** offer a detailed lecture presentation of statistical concepts covered in each chapter of *IPS*.

- **Assignments** organizes assignments and guides instructors through an easy-to-create assignment process providing access to questions from the Test Bank, Web Quizzes, and Exercises from *IPS* 6e. The Assignment Center enables instructors to create their own assignments from a variety of question types for self-graded assignments. This powerful assignment manager allows instructors to select their preferred policies in regard to scheduling, maximum attempts, time limitations, feedback, and more!

Online Study Center: www.whfreeman.com/osc/ips6e

(Access code required. Available for purchase online.) In addition to all the offerings available on the Companion Web site, the OSC offers:

- **StatTutor Tutorials**
- **Stats@Work Simulations**
- **Study Guide with Selected Solutions**
- **Statistical Software Manuals**

Companion Web site: www.whfreeman.com/ips6e

Seamlessly integrates topics from the text. On this open-access Web site, students can find the following:

- **Interactive Statistical Applets** that allow students to manipulate data and see the corresponding results graphically.

- **Data Sets** in ASCII, Excel, JMP, Minitab, TI, SPSS, and S-Plus formats.
- **Interactive Exercises and Self-Quizzes** to help students prepare for tests.
- **Optional Companion Chapters 14, 15, 16, and 17,** covering logistic regression, nonparametric tests, bootstrap methods and permutation tests, and statistics for quality control and capability.
- **Supplementary Exercises** for every chapter.

Interactive Student CD-ROM Included with every new copy of *IPS*, the CD contains access to the companion chapters, applets, and data sets also found on the Companion Web site.

Special Software Packages Student versions of JMP, Minitab, S-PLUS, and SPSS are available on a CD-ROM packaged with the textbook. This software is not sold separately and must be packaged with a text or a manual. Contact your W. H. Freeman representative for information or visit www.whfreeman.com.

NEW! SMARTHINKING Online Tutoring (access code required) W. H. Freeman and Company is partnering with SMARTHINKING to provide students with free online tutoring and homework help from specially trained, professional educators. Twelve-month subscriptions are available for packaging with *IPS*.

Printed Study Guide prepared by Michael A. Fligner of The Ohio State University offers students explanations of crucial concepts in each section of *IPS*, plus detailed solutions to key text problems and stepped-through models of important statistical techniques. ISBN 1-4292-1473-2

For Instructors

The **Instructor's Web site** www.whfreeman.com/ips6e requires user registration as an instructor and features all the student Web materials plus:

- **Instructor version of EESEE** (Electronic Encyclopedia of Statistical Examples and Exercises), with solutions to the exercises in the student version and **CrunchIt!** statistical software.
- **Instructor's Guide,** including full solutions to all exercises in .pdf format.
- **PowerPoint slides** containing all textbook figures and tables.
- **Lecture PowerPoint slides** offering a detailed lecture presentation of statistical concepts covered in each chapter of *IPS*.
- **Full answers to the Supplementary Exercises** on the student Web site.

Instructor's Guide with Solutions by Darryl Nester, Bluffton University. This printed guide includes full solutions to all exercises and provides video and Internet resources and sample examinations. It also contains brief discussions of the *IPS* approach for each chapter. ISBN 1-4292-1472-4

Test Bank by Brian Macpherson, University of Manitoba. The test bank contains hundreds of multiple-choice questions to generate quizzes and tests. Available in print as well as electronically on CD-ROM (for Windows and Mac), where questions can be downloaded, edited, and resequenced to suit the instructor's needs.
Printed Version, ISBN 1-4292-1471-6
Computerized (CD) Version, ISBN 1-4292-1859-2

Enhanced Instructor's Resource CD-ROM Allows instructors to search and export (by key term or chapter) all the material from the student CD, plus:

- All text images and tables
- Statistical applets and data sets
- Instructor's Guide with full solutions
- PowerPoint files and lecture slides
- Test bank files

ISBN 1-4292-1503-8

Course Management Systems W. H. Freeman and Company provides courses for Blackboard, WebCT (Campus Edition and Vista), and Angel course management systems. They are completely integrated courses that you can easily customize and adapt to meet your teaching goals and course objectives. On request, Freeman also provides courses for users of Desire2Learn and Moodle. Visit www.bfwpub.com/lmc for more information.

i-clicker is a new two-way radio-frequency classroom response solution developed by educators for educators. University of Illinois physicists Tim Stelzer, Gary Gladding, Mats Selen, and Benny Brown created the i-clicker system after using competing classroom response solutions and discovering they were neither classroom-appropriate nor student-friendly. Each step of i-clicker's development has been informed by teaching and learning. i-clicker is superior to other systems from both pedagogical and technical standpoints. To learn more about packaging i-clicker with this textbook, please contact your local sales rep or visit www.iclicker.com.

TO STUDENTS: *What Is Statistics?*

Statistics is the science of collecting, organizing, and interpreting numerical facts, which we call *data*. We are bombarded by data in our everyday lives. The news mentions imported car sales, the latest poll of the president's popularity, and the average high temperature for today's date. Advertisements claim that data show the superiority of the advertiser's product. All sides in public debates about economics, education, and social policy argue from data. A knowledge of statistics helps separate sense from nonsense in the flood of data.

The study and collection of data are also important in the work of many professions, so training in the science of statistics is valuable preparation for a variety of careers. Each month, for example, government statistical offices release the latest numerical information on unemployment and inflation. Economists and financial advisors, as well as policymakers in government and business, study these data in order to make informed decisions. Doctors must understand the origin and trustworthiness of the data that appear in medical journals. Politicians rely on data from polls of public opinion. Business decisions are based on market research data that reveal consumer tastes. Engineers gather data on the quality and reliability of manufactured products. Most areas of academic study make use of numbers, and therefore also make use of the methods of statistics.

Understanding from Data

The goal of statistics is to gain understanding from data. To gain understanding, we often operate on a set of numbers—we average or graph them, for example. But we must do more, because data are not just numbers; they are numbers that have some context that helps us understand them.

You read that low birth weight is a major reason why infant mortality in the United States is higher than in most other advanced nations. The report goes on to say that 7.8% of children born in the United States have low birth weight, and that 13.4% of black infants have low birth weight.[1] To make sense of these numbers you must know what counts as low birth weight (less than 2500 grams, or 5.5 pounds) and have some feeling for the weights of babies. You probably recognize that 5.5 pounds is small, that 7.5 pounds (3400 grams) is about average, and that 10 pounds (4500 grams) is a big baby.

Another part of the context is the source of the data. How do we know that 7.8% of American babies have low birth weight or that the average weight of newborns is about 3400 grams? The data come from the National Center for Health Statistics, a government office to which the states report information from all birth certificates issued each month. These are the most complete data available about births in the United States.

When you do statistical problems—even straightforward textbook problems—don't just graph or calculate. Think about the context and state your conclusions in the specific setting of the problem. As you are learning how to do statistical calculations and graphs, remember that the goal of statistics is not calculation for its own sake but gaining understanding from numbers. The

calculations and graphs can be automated by a calculator or software, but you must supply the understanding. This book presents only the most common specific procedures for statistical analysis. A thorough grasp of the principles of statistics will enable you to quickly learn more advanced methods as needed. Always keep in mind, however, that a fancy computer analysis carried out without attention to basic principles will often produce elaborate nonsense. As you read, seek to understand the principles, as well as the necessary details of methods and recipes.

The Rise of Statistics

Historically, the ideas and methods of statistics developed gradually as society grew interested in collecting and using data for a variety of applications. The earliest origins of statistics lie in the desire of rulers to count the number of inhabitants or measure the value of taxable land in their domains. As the physical sciences developed in the seventeenth and eighteenth centuries, the importance of careful measurements of weights, distances, and other physical quantities grew. Astronomers and surveyors striving for exactness had to deal with variation in their measurements. Many measurements should be better than a single measurement, even though they vary among themselves. How can we best combine many varying observations? Statistical methods that are still important were invented to analyze scientific measurements.

By the nineteenth century, the agricultural, life, and behavioral sciences also began to rely on data to answer fundamental questions. How are the heights of parents and children related? Does a new variety of wheat produce higher yields than the old, and under what conditions of rainfall and fertilizer? Can a person's mental ability and behavior be measured just as we measure height and reaction time? Effective methods for dealing with such questions developed slowly and with much debate.[2]

As methods for producing and understanding data grew in number and sophistication, the new discipline of statistics took shape in the twentieth century. Ideas and techniques that originated in the collection of government data, in the study of astronomical or biological measurements, and in the attempt to understand heredity or intelligence came together to form a unified "science of data." That science of data—statistics—is the topic of this text.

The Organization of This Book

Part I of this book, called "Looking at Data," concerns data analysis and data production. The first two chapters deal with statistical methods for organizing and describing data. These chapters progress from simpler to more complex data. Chapter 1 examines data on a single variable; Chapter 2 is devoted to relationships among two or more variables. You will learn both how to examine data produced by others and how to organize and summarize your own data. These summaries will be first graphical, then numerical, and then, when appropriate, in the form of a mathematical model that gives a compact description of the overall pattern of the data. Chapter 3 outlines arrangements (called "designs") for producing data that answer specific questions. The principles presented in this chapter will help you to design proper samples and experiments and to evaluate such investigations in your field of study.

Part II, consisting of Chapters 4 to 8, introduces statistical inference—formal methods for drawing conclusions from properly produced data. Statistical inference uses the language of probability to describe how reliable its conclusions are, so some basic facts about probability are needed to understand inference. Probability is the subject of Chapters 4 and 5. Chapter 6, perhaps the most important chapter in the text, introduces the reasoning of statistical inference. Effective inference is based on good procedures for producing data (Chapter 3), careful examination of the data (Chapters 1 and 2), and an understanding of the nature of statistical inference as discussed in Chapter 6. Chapters 7 and 8 describe some of the most common specific methods of inference for drawing conclusions about means and proportions from one and two samples.

The five shorter chapters in Part III introduce somewhat more advanced methods of inference, dealing with relations in categorical data, regression and correlation, and analysis of variance. Supplement chapters, available on the book-companion CD and Web site, present additional statistical topics.

What Lies Ahead

Introduction to the Practice of Statistics is full of data from many different areas of life and study. Many exercises ask you to express briefly some understanding gained from the data. In practice, you would know much more about the background of the data you work with and about the questions you hope the data will answer. No textbook can be fully realistic. But it is important to form the habit of asking, "What do the data tell me?" rather than just concentrating on making graphs and doing calculations.

You should have some help in automating many of the graphs and calculations. You should certainly have a calculator with basic statistical functions. Look for key words such as "two-variable statistics" or "regression" when you shop for a calculator. More advanced (and more expensive) calculators will do much more, including some statistical graphs. You may be asked to use software as well. There are many kinds of statistical software, from spreadsheets to large programs for advanced users of statistics. The kind of computing available to learners varies a great deal from place to place—but the big ideas of statistics don't depend on any particular level of access to computing.

Because graphing and calculating are automated in statistical practice, the most important assets you can gain from the study of statistics are an understanding of the big ideas and the beginnings of good judgment in working with data. Ideas and judgment can't (at least yet) be automated. They guide you in telling the computer what to do and in interpreting its output. This book tries to explain the most important ideas of statistics, not just teach methods. Some examples of big ideas that you will meet are "always plot your data," "randomized comparative experiments," and "statistical significance."

You learn statistics by doing statistical problems. Practice, practice, practice. Be prepared to work problems. The basic principle of learning is persistence. Being organized and persistent is more helpful in reading this book than knowing lots of math. The main ideas of statistics, like the main ideas of any important subject, took a long time to discover and take some time to master. The gain will be worth the pain.

This page intentionally left blank

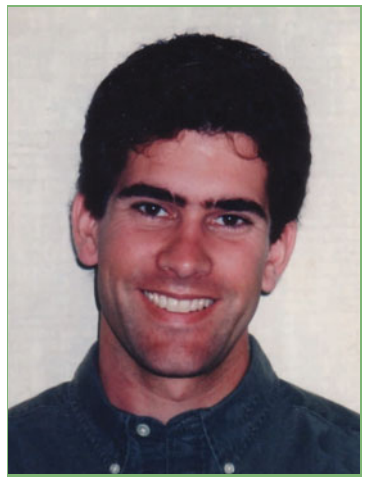
ABOUT THE AUTHORS

David S. Moore is Shanti S. Gupta Distinguished Professor of Statistics, Emeritus, at Purdue University and was 1998 president of the American Statistical Association. He received his A.B. from Princeton and his Ph.D. from Cornell, both in mathematics. He has written many research papers in statistical theory and served on the editorial boards of several major journals. Professor Moore is an elected fellow of the American Statistical Association and of the Institute of Mathematical Statistics and an elected member of the International Statistical Institute. He has served as program director for statistics and probability at the National Science Foundation.

In recent years, Professor Moore has devoted his attention to the teaching of statistics. He was the content developer for the Annenberg/Corporation for Public Broadcasting college-level telecourse *Against All Odds: Inside Statistics* and for the series of video modules *Statistics: Decisions through Data*, intended to aid the teaching of statistics in schools. He is the author of influential articles on statistics education and of several leading texts. Professor Moore has served as president of the International Association for Statistical Education and has received the Mathematical Association of America's national award for distinguished college or university teaching of mathematics.

George P. McCabe is the Associate Dean for Academic Affairs and a Professor of Statistics at Purdue University. In 1966 he received a B.S. degree in mathematics from Providence College and in 1970 a Ph.D. in mathematical statistics from Columbia University. His entire professional career has been spent at Purdue with sabbaticals at Princeton, the Commonwealth Scientific and Industrial Research Organization (CSIRO) in Melbourne, Australia, the University of Berne (Switzerland), the National Institute of Standards and Technology (NIST) in Boulder, Colorado, and the National University of Ireland in Galway. Professor McCabe is an elected fellow of the American Statistical Association and was 1998 chair of its section on Statistical Consulting. He has served on the editorial boards of several statistics journals. He has consulted with many major corporations and has testified as an expert witness on the use of statistics in several cases.

Professor McCabe's research interests have focused on applications of statistics. Much of his recent work has been focused on problems in nutrition, including nutrient requirements, calcium metabolism, and bone health. He is author or coauthor of over 150 publications in many different journals.



Bruce A. Craig is Professor of Statistics and Director of the Statistical Consulting Service at Purdue University. He received his B.S. in mathematics and economics from Washington University in St. Louis and his Ph.D. in statistics from the University of Wisconsin–Madison. He is an active member of the American Statistical Association and will be chair of its section on Statistical Consulting in 2009. He also is an active member of the Eastern North American Region of the International Biometrics Society and was elected by the voting membership to the Regional Committee between 2003 and 2006. Professor Craig serves on the editorial board of several statistical journals and serves on many data and safety monitoring boards, including Purdue's IRB.

Professor Craig's research interests focus on the development of novel statistical methodology to address research questions in the life sciences. Areas of current interest are protein structure determination, diagnostic testing, and animal abundance estimation. In 2005, he was named Purdue University Faculty Scholar.

DATA TABLE INDEX

TABLE 1.1	Service times (seconds) for calls to a customer service center	8
TABLE 1.2	Literacy rates (percent) in Islamic nations	10
TABLE 1.3	IQ test scores for 60 randomly chosen fifth-grade students	13
TABLE 1.4	Yearly discharge of the Mississippi River (in cubic kilometers of water)	18
TABLE 1.5	Average property damage per year due to tornadoes	25
TABLE 1.6	Carbon dioxide emissions (metric tons per person)	26
TABLE 1.7	Mean annual temperatures (°F) in two California cities	26
TABLE 1.8	Survival times (days) of guinea pigs in a medical experiment	29
TABLE 1.9	Educational data for 78 seventh-grade students	29
TABLE 1.10	Fuel economy (miles per gallon) for 2004 model vehicles	31
TABLE 1.11	Biological clock cycle lengths for a plant species in different locations	79
TABLE 2.1	NBA teams as businesses	98
TABLE 2.2	Two measurements of foot deformities	98
TABLE 2.3	World record times for the 10,000-meter run	99
TABLE 2.4	Growth of icicles over time	100
TABLE 2.5	Two measures of glucose level in diabetics	129
TABLE 2.6	Four data sets for exploring correlation and regression	142
TABLE 2.7	Fruit and vegetable consumption and smoking	162
TABLE 2.8	Percent growth in stock price and earnings for industry groups	163
TABLE 2.9	Neuron response to tones and monkey calls	164
TABLE 4.1	Grades awarded at a university, by school	294
TABLE 7.1	Monthly rates of return on a portfolio (percent)	425
TABLE 7.2	Aggressive behaviors of dementia patients	429
TABLE 7.3	Length (in seconds) of audio files sampled from an iPod	436

TABLE 7.4	DRP scores for third-graders	452
TABLE 7.5	Seated systolic blood pressure	463
TABLE 10.1	In-state tuition and fees (in dollars) for 32 public universities	595
TABLE 10.2	Watershed area and index of biotic integrity	596
TABLE 10.3	Percent forest and index of biotic integrity	597
TABLE 10.4	Annual number of tornadoes in the United States between 1953 and 2005	598
TABLE 10.5	C-reactive protein and serum retinol	598
TABLE 10.6	VO+ and osteocalcin	599
TABLE 10.7	VO− and TRAP	599
TABLE 10.8	Bank wages, length of service (LOS), and bank size	601
TABLE 10.9	SAT and ACT scores	604
TABLE 12.1	Quality ratings in three groups	665
TABLE 12.2	Age at death for women writers	676
TABLE 12.3	Price promotion data	677
TABLE 12.4	Piano lesson data	678
TABLE 13.1	Safety behaviors of abused women	701
TABLE 13.2	Tool diameter data	706
TABLE 13.3	Expected price data	707

BEYOND THE BASICS INDEX

CHAPTER 1	Decomposing time series	19
CHAPTER 1	Density estimation	71
CHAPTER 2	Scatterplot smoothers	92
CHAPTER 2	Transforming relationships	119
CHAPTER 2	Data mining	136
CHAPTER 3	Capture-recapture sampling	220
CHAPTER 4	More laws of large numbers	277
CHAPTER 5	Weibull distributions	344
CHAPTER 6	The bootstrap	368
CHAPTER 8	The plus four confidence interval for a single proportion	491
CHAPTER 8	Plus four confidence interval for a difference in proportions	509
CHAPTER 8	Relative risk	515
CHAPTER 9	Meta-analysis	534
CHAPTER 10	Nonlinear regression	576
CHAPTER 11	Multiple logistic regression	625
CHAPTER 16	The bootstrap for a scatterplot smoother	16-20

CHAPTER

1

Looking at Data—Distributions

Students planning a referendum on college fees. See Example 1.1.

Introduction

Statistics is the science of learning from data. Data are numerical facts. Here is an example of a situation where students used the results of a referendum to convince their university Board of Trustees to make a decision.

1.1 Displaying Distributions with Graphs

1.2 Describing Distributions with Numbers

1.3 Density Curves and Normal Distributions

EXAMPLE

1.1 Students vote for service learning scholarships. According to the National Service-Learning Clearinghouse: "Service-learning is a teaching and learning strategy that integrates meaningful community service with instruction and reflection to enrich the learning experience, teach civic responsibility, and strengthen communities."[1] University of Illinois at Urbana–Champaign students decided that they wanted to become involved in this national movement. They proposed a $15.00 per semester Legacy of Service and Learning Scholarship fee. Each year, $10.00 would be invested in an endowment and $5.00 would be used to fund current-use scholarships. In a referendum, students voted 3785 to 2977 in favor of the proposal. On April 11, 2006, the university Board of Trustees approved the proposal. Approximately $370,000 in current-use scholarship funds will be generated each year, and with the endowment, it is expected that in 20 years there will be more than a million dollars per year for these scholarships.