Contents

1 Introduction 1
1.1 The challenge of teaching introductory statistics 1
1.2 Fitting demonstrations, examples, and projects into a course 1
1.3 What makes a good example? 3
1.4 Why is statistics important? 3
1.5 The best of the best 4
1.6 Our motivation for writing this book 4

PART I INTRODUCTORY PROBABILITY AND STATISTICS

2 First week of class 11
2.1 Guessing ages 11
2.2 Where are the cancers? 13
2.3 Estimating a big number 14
2.4 What's in the news? 15
2.5 Collecting data from students 17

3 Descriptive statistics 19
3.1 Displaying graphs on the blackboard 19
3.2 Time series 19
3.2.1 World record times for the mile run 20
3.3 Numerical variables, distributions, and histograms 20
3.3.1 Categorical and continuous variables 20
3.3.2 Handedness 21
3.3.3 Soft drink consumption 22
3.4 Numerical summaries 22
3.4.1 Average soft drink consumption 22
3.4.2 The average student 24
3.5 Data in more than one dimension 24
3.5.1 Guessing exam scores 25
3.5.2 Who opposed the Vietnam War? 27
3.6 The normal distribution in one and two dimensions 28
3.6.1 Heights of men and women 29
3.6.2 Heights of conscripts 30
3.6.3 Scores on two exams 30
3.7 Linear transformations and linear combinations 31
3.7.1 College admissions 31
CONTENTS

3.7.2 Social and economic indexes
3.7.3 Age adjustment
3.8 Logarithmic transformations
  3.8.1 Simple examples: amoebas, squares, and cubes
  3.8.2 Log-linear transformation: world population
  3.8.3 Log-log transformation: metabolic rates

4 Linear regression and correlation
  4.1 Fitting linear regressions
    4.1.1 Simple examples of least squares
    4.1.2 Tall people have higher incomes
    4.1.3 Logarithm of world population
  4.2 Correlation
    4.2.1 Correlations of body measurements
    4.2.2 Correlation and causation in observational data
  4.3 Regression to the mean
    4.3.1 Mini-quizzes
    4.3.2 Exam scores, heights, and the general principle

5 Data collection
  5.1 Sample surveys
    5.1.1 Sampling from the telephone book
    5.1.2 First digits and Benford’s law
    5.1.3 Wacky surveys
    5.1.4 An election exit poll
    5.1.5 Simple examples of bias
    5.1.6 How large is your family?
  5.2 Class projects in survey sampling
    5.2.1 The steps of the project
    5.2.2 Topics for student surveys
  5.3 Experiments
    5.3.1 An experiment that looks like a survey
    5.3.2 Randomizing the order of exam questions
    5.3.3 Taste tests
  5.4 Observational studies
    5.4.1 The Surgeon General’s report on smoking
    5.4.2 Large population studies
    5.4.3 Coaching for the SAT

6 Statistical literacy and the news media
  6.1 Introduction
  6.2 Assignment based on instructional packets
  6.3 Assignment where students find their own articles
  6.4 Guidelines for finding and evaluating sources
  6.5 Discussion and student reactions
  6.6 Examples of course packets

7 Probability
  7.1 Constructing probability models
  7.2 Random variables
    7.2.1 Random processes
    7.2.2 Random experiments
    7.2.3 Random variables
    7.2.4 Probability distributions
  7.3 Probability models
    7.3.1 Binomial models
    7.3.2 Poisson models
    7.3.3 Lognormal models
  7.4 Probability distributions
    7.4.1 Law of Large Numbers
    7.4.2 Variability
    7.4.3 Spreads
  7.5 Conditional probabilities
    7.5.1 Bayes’ Theorem
    7.5.2 Likelihood
  7.6 You can lie with statistics
    7.6.1 Deception
    7.6.2 Spreads
    7.6.3 Facts

8 Statistical inference
  8.1 Weighing evidence
  8.2 From probability
    8.2.1 Weighting
    8.2.2 Reliability
  8.3 Confidence intervals
    8.3.1 Bias
    8.3.2 Confidence intervals
    8.3.3 Large population
    8.3.4 Small population
    8.3.5 Goldilocks
  8.4 Confidence limits
    8.4.1 Coverage
    8.4.2 Noncoverage
  8.5 Hypothesis testing
    8.5.1 Hypothesis testing
    8.5.2 Binomial testing
<table>
<thead>
<tr>
<th>Page</th>
<th>Section</th>
</tr>
</thead>
<tbody>
<tr>
<td>31</td>
<td>6.6.1</td>
</tr>
<tr>
<td>32</td>
<td>6.6.2</td>
</tr>
<tr>
<td>33</td>
<td>6.6.3</td>
</tr>
<tr>
<td>33</td>
<td>6.6.4</td>
</tr>
<tr>
<td>35</td>
<td>7</td>
</tr>
<tr>
<td>38</td>
<td>7.1</td>
</tr>
<tr>
<td>38</td>
<td>7.2</td>
</tr>
<tr>
<td>39</td>
<td>7.2.1</td>
</tr>
<tr>
<td>41</td>
<td>7.2.2</td>
</tr>
<tr>
<td>43</td>
<td>7.2.3</td>
</tr>
<tr>
<td>43</td>
<td>7.2.4</td>
</tr>
<tr>
<td>44</td>
<td>7.3</td>
</tr>
<tr>
<td>45</td>
<td>7.3.1</td>
</tr>
<tr>
<td>45</td>
<td>7.3.2</td>
</tr>
<tr>
<td>46</td>
<td>7.3.3</td>
</tr>
<tr>
<td>48</td>
<td>7.4</td>
</tr>
<tr>
<td>48</td>
<td>7.4.1</td>
</tr>
<tr>
<td>51</td>
<td>7.4.2</td>
</tr>
<tr>
<td>51</td>
<td>7.4.3</td>
</tr>
<tr>
<td>52</td>
<td>7.5</td>
</tr>
<tr>
<td>54</td>
<td>7.5.1</td>
</tr>
<tr>
<td>54</td>
<td>7.5.2</td>
</tr>
<tr>
<td>56</td>
<td>7.6</td>
</tr>
<tr>
<td>56</td>
<td>7.6.1</td>
</tr>
<tr>
<td>57</td>
<td>7.6.2</td>
</tr>
<tr>
<td>57</td>
<td>7.6.3</td>
</tr>
<tr>
<td>58</td>
<td>8</td>
</tr>
<tr>
<td>63</td>
<td>8.1</td>
</tr>
<tr>
<td>66</td>
<td>8.2</td>
</tr>
<tr>
<td>69</td>
<td>8.2.1</td>
</tr>
<tr>
<td>72</td>
<td>8.2.2</td>
</tr>
<tr>
<td>73</td>
<td>8.3</td>
</tr>
<tr>
<td>73</td>
<td>8.3.1</td>
</tr>
<tr>
<td>75</td>
<td>8.3.2</td>
</tr>
<tr>
<td>76</td>
<td>8.3.3</td>
</tr>
<tr>
<td>76</td>
<td>8.3.4</td>
</tr>
<tr>
<td>77</td>
<td>8.3.5</td>
</tr>
<tr>
<td>79</td>
<td>8.4</td>
</tr>
<tr>
<td>82</td>
<td>8.4.1</td>
</tr>
<tr>
<td>84</td>
<td>8.4.2</td>
</tr>
<tr>
<td>84</td>
<td>8.5</td>
</tr>
<tr>
<td>84</td>
<td>8.5.1</td>
</tr>
<tr>
<td>84</td>
<td>8.5.2</td>
</tr>
</tbody>
</table>
CONTENTS

8.5.3 Hypergeometric model: taste testing 131
8.5.4 Benford's law of first digits 131
8.5.5 Length of baseball World Series 131
8.6 Simple examples of applied inference 132
8.6.1 How good is your memory? 132
8.6.2 How common is your name? 133
8.7 Advanced concepts of inference 134
8.7.1 Shooting baskets and statistical power 134
8.7.2 Do-it-yourself data dredging 134
8.7.3 Praying for your health 135

9 Multiple regression and nonlinear models 137
9.1 Regression of income on height and sex 137
9.1.1 Inference for regression coefficients 137
9.1.2 Multiple regression 137
9.1.3 Regression with interactions 139
9.1.4 Transformations 140
9.2 Exam scores 141
9.2.1 Studying the fairness of random exams 141
9.2.2 Measuring the reliability of exam questions 141
9.3 A nonlinear model for golf putting 142
9.3.1 Looking at data 143
9.3.2 Constructing a probability model 143
9.3.3 Checking the fit of the model to the data 144
9.4 Pythagoras goes linear 145

10 Lying with statistics 147
10.1 Examples of misleading presentations of numbers 147
10.1.1 Fabricated or meaningless numbers 147
10.1.2 Misinformation 147
10.1.3 Ignoring the baseline 149
10.1.4 Arbitrary comparisons or data dredging 149
10.1.5 Misleading comparisons 151
10.2 Selection bias 153
10.2.1 Distinguishing from other sorts of bias 153
10.2.2 Some examples presented as puzzles 154
10.2.3 Avoiding over-skepticism 155
10.3 Reviewing the semester's material 155
10.3.1 Classroom discussion 155
10.3.2 Assignments: find the lie or create the lie 156
10.4 1 in 2 marriages end in divorce? 156
10.5 Ethics and statistics 158
10.5.1 Cutting corners in a medical study 158
10.5.2 Searching for statistical significance 159
10.5.3 Controversies about randomized experiments 159
10.5.4 How important is blindness? 160

PART II Practical projects

11 How to do projects 161
11.1 Getting started 161
11.1.1 Planning 161
11.1.2 Doing 162
11.1.3 Writing up 163
11.1.4 Getting feedback 163
11.2 In-class projects 164
11.2.1 Spot the fake 164
11.2.2 I'm a celebrity! 164
11.2.3 The four horsemen 165
11.2.4 Do-it-yourself data dredging 165
11.3 Using each other's projects 166
11.4 Projects for fun 167
11.4.1 Making your own 167
11.4.2 Making a presentation 167
11.4.3 Testing statistical significance 167
11.4.4 Submitting a project 167
11.5 Resources for projects 168
11.5.1 Where to find data 168
11.5.2 Further reading 168
11.5.3 Further software 168
11.5.4 Further websites 168
11.5.5 Further exercises 168

12 Structuring a project 169
12.1 Before the final write-up 169
12.2 Finding a topic 169
12.3 A detailed outline 169
12.4 Outline for the final write-up 169

PART III More decision theory

13 Decision theory 171
13.1 Decision rules 171
13.1.1 Expected value 171
13.1.2 Utility 172
13.1.3 Risk 172
13.1.4 Value 172
13.1.5 Probabilities 172
13.1.6 Health 172
13.1.7 Risk 172
CONTENTS

10.5.5 Use of information in statistical inferences 161

PART II PUTTING IT ALL TOGETHER

11 How to do it

11.1 Getting started
    11.1.1 Multitasking
    11.1.2 Advance planning
    11.1.3 Fitting an activity to your class
    11.1.4 Common mistakes

11.2 In-class activities
    11.2.1 Setting up effective demonstrations
    11.2.2 Promoting discussion
    11.2.3 Getting to know the students
    11.2.4 Fostering group work

11.3 Using exams to teach statistical concepts

11.4 Projects
    11.4.1 Monitoring progress
    11.4.2 Organizing independent projects
    11.4.3 Topics for projects
    11.4.4 Statistical design and analysis

11.5 Resources
    11.5.1 What’s in a spaghetti box? Cooking up activities from scratch
    11.5.2 Books
    11.5.3 Periodicals
    11.5.4 Web sites
    11.5.5 People

12 Structuring an introductory statistics course

12.1 Before the semester begins
12.2 Finding time for student activities in class
12.3 A detailed schedule for a semester-long course
12.4 Outline for an alternative schedule of activities

PART III MORE ADVANCED COURSES

13 Decision theory and Bayesian statistics

13.1 Decision analysis
    13.1.1 How many quarters are in the jar?
    13.1.2 Utility of money
    13.1.3 Risk aversion
    13.1.4 What is the value of a life?
    13.1.5 Probabilistic answers to true-false questions
    13.1.6 Homework project: evaluating real-life forecasts
    13.1.7 Real decision problems
CONTENTS

13.2 Bayesian statistics
  13.2.1 Where are the cancers?
  13.2.2 Subjective probability intervals and calibration
  13.2.3 Drawing parameters out of a hat
  13.2.4 Where are the cancers? A simulation
  13.2.5 Hierarchical modeling and shrinkage

14 Student activities in survey sampling
  14.1 First week of class
    14.1.1 News clippings
    14.1.2 Class survey
  14.2 Random number generation
    14.2.1 What do random numbers look like?
    14.2.2 Random numbers from coin flips
  14.3 Estimation and confidence intervals
  14.4 A visit to Clusterville
  14.5 Statistical literacy and discussion topics
  14.6 Projects
    14.6.1 Research papers on complex surveys
    14.6.2 Sampling and inference in StatCity
    14.6.3 A special topic in sampling

15 Problems and projects in probability
  15.1 Setting up a probability course as a seminar
  15.2 Introductory problems
    15.2.1 Probabilities of compound events
    15.2.2 Introducing the concept of expectation
  15.3 Challenging problems
  15.4 Does the Poisson distribution fit real data?
  15.5 Organizing student projects
  15.6 Examples of structured projects
    15.6.1 Fluctuations in coin tossing—arc sine laws
    15.6.2 Recurrence and transience in Markov chains
  15.7 Examples of unstructured projects
    15.7.1 Martingales
    15.7.2 Generating functions and branching processes
    15.7.3 Limit distributions of Markov chains
    15.7.4 Permutations
  15.8 Research papers as projects

16 Directed projects in a mathematical statistics course
  16.1 Organization of a case study
  16.2 Fitting the cases into a course
    16.2.1 Covering the cases in lectures
    16.2.2 Group work in class
    16.2.3 Cases as reports
16.2.4 Independent projects in a seminar course
16.3 A case study: quality control
16.4 A directed project: helicopter design
    16.4.1 General instructions
    16.4.2 Designing the study and fitting a response surface

Notes
References
Author Index
Subject Index