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

3.7.2 Social and economic indexes 31
3.7.3 Age adjustment 32
3.8 Logarithmic transformations 32
3.8.1 Simple examples: amoebas, squares, and cubes 33
3.8.2 Log-linear transformation: world population 33
3.8.3 Log-log transformation: metabolic rates 35

4 Linear regression and correlation 38
4.1 Fitting linear regressions 38
4.1.1 Simple examples of least squares 38
4.1.2 Tall people have higher incomes 39
4.1.3 Logarithm of world population 41
4.2 Correlation 43
4.2.1 Correlations of body measurements 43
4.2.2 Correlation and causation in observational data 44
4.3 Regression to the mean 45
4.3.1 Mini-quizzes 45
4.3.2 Exam scores, heights, and the general principle 46

5 Data collection 48
5.1 Sample surveys 48
5.1.1 Sampling from the telephone book 48
5.1.2 First digits and Benford's law 52
5.1.3 Wacky surveys 54
5.1.4 An election exit poll 55
5.1.5 Simple examples of bias 56
5.1.6 How large is your family? 56
5.2 Class projects in survey sampling 57
5.2.1 The steps of the project 58
5.2.2 Topics for student surveys 63
5.3 Experiments 66
5.3.1 An experiment that looks like a survey 66
5.3.2 Randomizing the order of exam questions 68
5.3.3 Taste tests 69
5.4 Observational studies 72
5.4.1 The Surgeon General's report on smoking 73
5.4.2 Large population studies 73
5.4.3 Coaching for the SAT 75

6 Statistical literacy and the news media 76
6.1 Introduction 76
6.2 Assignment based on instructional packets 77
6.3 Assignment where students find their own articles 79
6.4 Guidelines for finding and evaluating sources 82
6.5 Discussion and student reactions 84
6.6 Examples of course packets 84

6.6.1 A controlled experiment: IV fluids for trauma victims 85
6.6.2 A sample survey: 1 in 4 youths abused, survey finds 90
6.6.3 An observational study: Monster in the crib 93
6.6.4 A model-based analysis: Illegal aliens put uneven load 98

7 Probability 103
7.1 Constructing probability examples 103
7.2 Random numbers via dice or handouts 103
7.2.1 Random digits via dice 103
7.2.2 Random digits via handouts 103
7.2.3 Normal distribution 104
7.2.4 Poisson distribution 104
7.3 Probabilities of compound events 104
7.3.1 Babies 104
7.3.2 Real vs. fake coin flips 105
7.3.3 Lotteries 107
7.4 Probability modeling 108
7.4.1 Lengths of baseball World Series 108
7.4.2 Voting and coalitions 110
7.4.3 Space shuttle failure and other rare events 110
7.5 Conditional probability 111
7.5.1 What's the color on the other side of the card? 111
7.5.2 Lie detectors and false positives 113
7.6 You can load a die but you can't bias a coin flip 114
7.6.1 Demonstration using plastic checkers and wooden dice 115
7.6.2 Sporting events and quantitative literacy 117
7.6.3 Physical explanation 118

8 Statistical inference 120
8.1 Weighing a "random" sample 120
8.2 From probability to inference: distributions of totals and averages 121
8.2.1 Where are the missing girls? 121
8.2.2 Real-time gambler's ruin 122
8.3 Confidence intervals: examples 123
8.3.1 Biases in age guessing 123
8.3.2 Comparing two groups 124
8.3.3 Land or water? 124
8.3.4 Poll differentials: a discrete distribution 125
8.3.5 Golf: can you putt like the pros? 126
8.4 Confidence intervals: theory 126
8.4.1 Coverage of confidence intervals 126
8.4.2 Noncoverage of confidence intervals 128
8.5 Hypothesis testing: z, t, and χ^2 tests 128
8.5.1 Hypothesis tests from examples of confidence intervals 129
8.5.2 Binomial model: sampling from the phone book 130

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

10.5.5 Use of information in statistical inferences 161

PART II PUTTING IT ALL TOGETHER

11 How to do it 167
11.1 Getting started 167
11.1.1 Multitasking 167
11.1.2 Advance planning 167
11.1.3 Fitting an activity to your class 168
11.1.4 Common mistakes 168
11.2 In-class activities 171
11.2.1 Setting up effective demonstrations 171
11.2.2 Promoting discussion 172
11.2.3 Getting to know the students 173
11.2.4 Fostering group work 173
11.3 Using exams to teach statistical concepts 175
11.4 Projects 175
11.4.1 Monitoring progress 177
11.4.2 Organizing independent projects 178
11.4.3 Topics for projects 181
11.4.4 Statistical design and analysis 183
11.5 Resources 185
11.5.1 What's in a spaghetti box? Cooking up activities from scratch 185
11.5.2 Books 186
11.5.3 Periodicals 187
11.5.4 Web sites 187
11.5.5 People 188

12 Structuring an introductory statistics course 189
12.1 Before the semester begins 189
12.2 Finding time for student activities in class 190
12.3 A detailed schedule for a semester-long course 190
12.4 Outline for an alternative schedule of activities 198

PART III MORE ADVANCED COURSES

13 Decision theory and Bayesian statistics 203
13.1 Decision analysis 204
13.1.1 How many quarters are in the jar? 204
13.1.2 Utility of money 207
13.1.3 Risk aversion 209
13.1.4 What is the value of a life? 210
13.1.5 Probabilistic answers to true–false questions 211
13.1.6 Homework project: evaluating real-life forecasts 212
13.1.7 Real decision problems 213

13.2 Bayesian statistics 215
13.2.1 Where are the cancers? 215
13.2.2 Subjective probability intervals and calibration 216
13.2.3 Drawing parameters out of a hat 219
13.2.4 Where are the cancers? A simulation 219
13.2.5 Hierarchical modeling and shrinkage 220

14 Student activities in survey sampling 222
14.1 First week of class 222
14.1.1 News clippings 222
14.1.2 Class survey 223
14.2 Random number generation 224
14.2.1 What do random numbers look like? 224
14.2.2 Random numbers from coin flips 224
14.3 Estimation and confidence intervals 225
14.4 A visit to Clusterville 226
14.5 Statistical literacy and discussion topics 228
14.6 Projects 230
14.6.1 Research papers on complex surveys 231
14.6.2 Sampling and inference in StatCity 232
14.6.3 A special topic in sampling 236

15 Problems and projects in probability 237
15.1 Setting up a probability course as a seminar 237
15.2 Introductory problems 238
15.2.1 Probabilities of compound events 239
15.2.2 Introducing the concept of expectation 240
15.3 Challenging problems 241
15.4 Does the Poisson distribution fit real data? 243
15.5 Organizing student projects 244
15.6 Examples of structured projects 244
15.6.1 Fluctuations in coin tossing—arcsine laws 245
15.6.2 Recurrence and transience in Markov chains 247
15.7 Examples of unstructured projects 249
15.7.1 Martingales 249
15.7.2 Generating functions and branching processes 250
15.7.3 Limit distributions of Markov chains 250
15.7.4 Permutations 251
15.8 Research papers as projects 252

16 Directed projects in a mathematical statistics course 254
16.1 Organization of a case study 255
16.2 Fitting the cases into a course 255
16.2.1 Covering the cases in lectures 256
16.2.2 Group work in class 256
16.2.3 Cases as reports 257

16.2.4 Independent projects in a seminar course 257
16.3 A case study: quality control 258
16.4 A directed project: helicopter design 259
16.4.1 General instructions 259
16.4.2 Designing the study and fitting a response surface 261

Notes 265

References 277

Author Index 288

Subject Index 292