Contents

1 Introduction 1
- 1.1 The challenge of teaching introductory statistics 1
- 1.2 Fitting demonstrations and examples 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 29
  - 3.6.3 Scores on two exams 29
- 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 Statistical graphics 38
4.1 Guiding principles 39
4.2 Lecture topics 40
4.3 Assignments 41
4.4 Deconstruct and reconstruct a plot 43
4.5 One-minute revelation 45
4.6 Turning tables 46

5 Linear regression and correlation 48
5.1 Fitting linear regressions 48
5.1.1 Simple examples of least squares 48
5.1.2 Tall people have higher incomes 49
5.1.3 Logarithm of world population 51
5.2 Correlation 53
5.2.1 Correlations of body measurements 53
5.2.2 Correlation and causation in observational data 54
5.3 Regression to the mean 55
5.3.1 Mini-quizzes 55
5.3.2 Exam scores, heights, and the general principle 56

6 Data collection 58
6.1 Sample surveys 58
6.1.1 Sampling from the telephone book 58
6.1.2 First digits and Benford's law 62
6.1.3 Wacky surveys 64
6.1.4 An election exit poll 65
6.1.5 Simple examples of bias 66
6.1.6 How large is your family? 66
6.2 Class projects in survey sampling 67
6.2.1 The steps of the project 68
6.2.2 Topics for student surveys 73
6.3 How big was the crowd? 76
6.4 Experiments 78
6.4.1 An experiment that looks like a survey 78
6.4.2 Randomizing the order of exam questions 81
6.4.3 Taste tests 82
6.4.4 Can they taste the difference? 85
6.5 Observational studies 85
6.5.1 The Surgeon General's report on smoking 86

6.5.2 Large population studies 87
6.5.3 Coaching for the SAT 88

7 Statistical literacy and the news media 90
7.1 Introduction 90
7.2 Assignment based on instructional packets 91
7.3 Assignment where students find their own articles 93
7.4 Guidelines for finding and evaluating sources 96
7.5 Discussion and student reactions 98
7.6 Examples of course packets 98
7.6.1 A controlled experiment: Fluids for trauma victims 99
7.6.2 A sample survey: 1 in 4 youths abused, survey finds 104
7.6.3 An observational study: Monster in the crib 108
7.6.4 A model-based analysis: Illegal aliens 112

8 Probability 117
8.1 Constructing probability examples 117
8.2 Random numbers via dice or handouts 117
8.2.1 Random digits via dice 117
8.2.2 Random digits via handouts 117
8.2.3 Normal distribution 118
8.2.4 Poisson distribution 118
8.3 Probabilities of compound events 118
8.3.1 Babies 118
8.3.2 Real vs. fake coin flips 119
8.3.3 Lotteries 121
8.4 Probability modeling 122
8.4.1 Lengths of baseball World Series 122
8.4.2 Voting and coalitions 124
8.4.3 Space shuttle failure and other rare events 124
8.5 Conditional probability 125
8.5.1 What's the color on the other side of the card? 125
8.5.2 Lie detectors and false positives 127
8.6 You can load a die but you can't bias a coin flip 128
8.6.1 Demonstration using wooden dice 129
8.6.2 Sporting events and quantitative literacy 131
8.6.3 Physical explanation 132

9 Statistical inference 134
9.1 Weighing a "random" sample 134
9.2 From probability to inference: totals and averages 135
9.2.1 Where are the missing girls? 135
9.2.2 Real-time gambler's ruin 136
9.3 Confidence intervals: examples 137
9.3.1 Biases in age guessing 137
9.3.2 Comparing two groups 138

9.3.3 Land or water? 138
9.3.4 Poll differentials: a discrete distribution 139
9.3.5 Golf: can you putt like the pros? 140
9.4 Confidence intervals: theory 140
9.4.1 Coverage of confidence intervals 140
9.4.2 Noncoverage of confidence intervals 142
9.5 Hypothesis testing: z, t, and χ^2 tests 142
9.5.1 Hypothesis tests from confidence intervals 143
9.5.2 Binomial model: sampling from the phone book 144
9.5.3 Hypergeometric model: taste testing 145
9.5.4 Benford's law of first digits 145
9.5.5 Length of baseball World Series 145
9.6 Simple examples of applied inference 146
9.6.1 How good is your memory? 146
9.6.2 How common is your name? 147
9.7 Advanced concepts of inference 148
9.7.1 Shooting baskets and statistical power 148
9.7.2 Do-it-yourself data dredging 148
9.7.3 Praying for your health 149

10 Multiple regression and nonlinear models 151
10.1 Regression of income on height and sex 151
10.1.1 Inference for regression coefficients 151
10.1.2 Multiple regression 151
10.1.3 Regression with interactions 153
10.1.4 Transformations 154
10.2 Exam scores 155
10.2.1 Studying the fairness of random exams 155
10.2.2 Measuring the reliability of exam questions 155
10.3 A nonlinear model for golf putting 156
10.3.1 Looking at data 157
10.3.2 Constructing a probability model 157
10.3.3 Checking the fit of the model to the data 158
10.4 Pythagoras goes linear 160

11 Lying with statistics 162
11.1 Examples of misleading presentations of numbers 162
11.1.1 Fabricated or meaningless numbers 162
11.1.2 Misinformation 162
11.1.3 Ignoring the baseline 164
11.1.4 Arbitrary comparisons or data dredging 164
11.2 Selection bias 166
11.2.1 Distinguishing from other sorts of bias 166
11.2.2 Some examples presented as puzzles 168
11.2.3 Avoiding over-skepticism 168
11.3 Reviewing the semester's material 169

11.3.1 Classroom discussion 169
11.3.2 Assignments: Find the lie or create the lie 171
11.4 1 in 2 marriages end in divorce? 171
11.5 Ethics and statistics 171
11.5.1 Cutting corners in a medical study 171
11.5.2 Searching for statistical significance 172
11.5.3 Controversies about randomized experiments 173
11.5.4 How important is blindness? 173
11.5.5 Use of information in statistical inferences 174

PART II PUTTING IT ALL TOGETHER

12 How to do it 179
12.1 Getting started 179
12.1.1 Multitasking 179
12.1.2 Advance planning 179
12.1.3 Fitting an activity to your class 180
12.1.4 Common mistakes 180
12.2 In-class activities 183
12.2.1 Setting up effective demonstrations 183
12.2.2 Promoting discussion 184
12.2.3 Getting to know the students 185
12.2.4 Fostering group work 185
12.3 Tricks for the large lecture 187
12.4 Using exams to teach statistical concepts 190
12.5 Projects 190
12.5.1 Monitoring progress 192
12.5.2 Organizing independent projects 198
12.5.3 Topics for projects 201
12.5.4 Statistical design and analysis 204
12.6 Resources 205
12.6.1 What's in a spaghetti box? 206
12.6.2 Books 207
12.6.3 Periodicals 207
12.6.4 Web sites 208
12.6.5 People 208

13 Structuring an introductory statistics course 209
13.1 Before the semester begins 209
13.2 Finding time for student activities in class 210
13.3 A detailed schedule for a semester-long course 210
13.4 Outline for an alternative schedule of activities 218

14 Teaching statistics to social scientists 221
14.1 Starting with predictions, graphs, and deterministic models 221
14.2 Teaching style 223

14.3 A case study: the sampling distribution of the sample mean 224
14.4 Starting an applied regression course 224
14.5 How is there time to cover all the material? 226

15 Statistics diaries 228
15.1 Examples of student diaries 228
15.2 Using diaries in statistics classes 250

16 A course in statistical communication and graphics 252
16.1 Background 252
16.2 Plan for a 13-week course 254

PART III MORE ADVANCED COURSES

17 Decision theory and Bayesian statistics 277
17.1 Decision analysis 278
17.1.1 How many quarters are in the jar? 278
17.1.2 Utility of money 281
17.1.3 Risk aversion 283
17.1.4 What is the value of a life? 284
17.1.5 Probabilistic answers to true–false questions 285
17.1.6 Homework project: evaluating real-life forecasts 286
17.1.7 Real decision problems 287
17.2 Bayesian statistics 289
17.2.1 Where are the cancers? 289
17.2.2 Subjective probability intervals and calibration 290
17.2.3 Drawing parameters out of a hat 293
17.2.4 Where are the cancers? A simulation 293
17.2.5 Hierarchical modeling and shrinkage 294

18 Student activities in survey sampling 296
18.1 First week of class 296
18.1.1 News clippings 296
18.1.2 Question bias 297
18.1.3 Class survey 297
18.2 Random number generation 299
18.2.1 What do random numbers look like? 299
18.2.2 Random numbers from coin flips 300
18.3 Estimation and confidence intervals 300
18.4 A visit to Clusterville 303
18.5 Statistical literacy and discussion topics 303
18.6 Projects 305
18.6.1 Analyzing data from a complex survey 305
18.6.2 Research papers on complex surveys 310
18.6.3 Sampling and inference in StatCity 311
18.6.4 A special topic in sampling 314

19 Problems and projects in probability 315
19.1 Setting up a probability course as a seminar 315
19.2 Introductory problems 316
19.2.1 Probabilities of compound events 317
19.2.2 Introducing the concept of expectation 318
19.3 Challenging problems 319
19.4 Does the Poisson distribution fit real data? 321
19.5 Organizing student projects 322
19.6 Examples of structured projects 322
19.6.1 Fluctuations in coin tossing—arcsine laws 323
19.6.2 Recurrence and transience in Markov chains 325
19.7 Examples of unstructured projects 327
19.7.1 Martingales 327
19.7.2 Generating functions and branching processes 328
19.7.3 Limit distributions of Markov chains 328
19.7.4 Permutations 329
19.8 Research papers as projects 330

20 Directed projects in a mathematical statistics course 332
20.1 Organization of a case study 333
20.2 Fitting the cases into a course 333
20.2.1 Covering the cases in lectures 334
20.2.2 Group work in class 334
20.2.3 Cases as reports 335
20.2.4 Independent projects in a seminar course 335
20.3 A case study: quality control 336
20.4 A directed project: helicopter design 337
20.4.1 General instructions 337
20.4.2 Designing the study and fitting a response surface 339

21 Statistical thinking in a data science course 342
21.1 Goals 343
21.1.1 Statistical thinking in a computational context 343
21.1.2 Core paradigms 343
21.1.3 Learn how to learn new technologies 344
21.1.4 Connect to real modern problems 344
21.2 Topics 344
21.2.1 Language basics 345
21.2.2 Graphics 345
21.2.3 Data structures 346
21.2.4 Programming concepts 346
21.2.5 Text manipulation 347
21.2.6 Information technologies 348
21.2.7 Statistical methods 349
21.3 Projects and student work 350
21.4 Copy the master 350

21.5 Spam filtering assignment 354
21.6 Interactive visualization assignment 357

Notes 361

References 375

Author Index 387

Subject Index 392