CONTENTS

Preface xi

Chapter 1. Statistics in the modern day 1

PART I COMPUTING 15

Chapter 2. C 17

2.1 Lines 18
2.2 Variables and their declarations 28
2.3 Functions 34
2.4 The debugger 43
2.5 Compiling and running 48
2.6 Pointers 53
2.7 Arrays and other pointer tricks 59
2.8 Strings 65
2.9 ※ Errors 69

Chapter 3. Databases 74

3.1 Basic queries 76
3.2 ※ Doing more with queries 80
3.3 Joins and subqueries 87
3.4 On database design 94
3.5 Folding queries into C code 98

3.6 Maddening details 103
3.7 Some examples 108

Chapter 4. Matrices and models 113
4.1 The GSL's matrices and vectors 114
4.2 `apop_data` 120
4.3 Shunting data 123
4.4 Linear algebra 129
4.5 Numbers 135
4.6 ※ `gsl_matrix` and `gsl_vector` internals 140
4.7 Models 143

Chapter 5. Graphics 157
5.1 `plot` 160
5.2 ※ Some common settings 163
5.3 From arrays to plots 166
5.4 A sampling of special plots 171
5.5 Animation 177
5.6 On producing good plots 180
5.7 ※ Graphs—nodes and flowcharts 182
5.8 ※ Printing and LaTeX 185

Chapter 6. ※ More coding tools 189
6.1 Function pointers 190
6.2 Data structures 193
6.3 Parameters 203
6.4 ※ Syntactic sugar 210
6.5 More tools 214

PART II STATISTICS 217

Chapter 7. Distributions for description 219
7.1 Moments 219
7.2 Sample distributions 235
7.3 Using the sample distributions 252
7.4 Non-parametric description 261

Chapter 8. Linear projections 264
8.1 ※ Principal component analysis 265
8.2 OLS and friends 270
8.3 Discrete variables 280
8.4 Multilevel modeling 288

Chapter 9. Hypothesis testing with the CLT 295
9.1 The Central Limit Theorem 297
9.2 Meet the Gaussian family 301
9.3 Testing a hypothesis 307
9.4 ANOVA 312
9.5 Regression 315
9.6 Goodness of fit 319

Chapter 10. Maximum likelihood estimation 325
10.1 Log likelihood and friends 326
10.2 Description: Maximum likelihood estimators 337
10.3 Missing data 345
10.4 Testing with likelihoods 348

Chapter 11. Monte Carlo 356
11.1 Random number generation 357
11.2 Description: Finding statistics for a distribution 364
11.3 Inference: Finding statistics for a parameter 367
11.4 Drawing a distribution 371
11.5 Non-parametric testing 375

Appendix A: Environments and makefiles 381
A.1 Environment variables 381
A.2 Paths 385
A.3 Make 387

Appendix B: Text processing 392
B.1 Shell scripts 393
B.2 Some tools for scripting 398
B.3 Regular expressions 403
B.4 Adding and deleting 413
B.5 More examples 415

Appendix C: Glossary 419

Bibliography 435

Index 443