

Contents in brief

Preface	xxv
Acknowledgments	xxix
About the authors	xxxiii

PART 1 INTRODUCTION TO STATISTICS

	1. Statistics and samples	1
INTERLEAF 1	Biology and the history of statistics	20
	2. Displaying data	23
	3. Describing data	59
	4. Estimating with uncertainty	83
INTERLEAF 2	Pseudoreplication	97
	5. Probability	99
	6. Hypothesis testing	127
INTERLEAF 3	Why statistical significance is not the same as biological importance	148

PART 2 PROPORTIONS AND FREQUENCIES

	7. Analyzing proportions	151
INTERLEAF 4	Correlation does not require causation	173
	8. Fitting probability models to frequency data	175
INTERLEAF 5	Making a plan	204
	9. Contingency analysis: associations between categorical variables	207

PART 3 COMPARING NUMERICAL VALUES**10. The normal distribution 231****INTERLEAF 6 Controls in medical studies 256****11. Inference for a normal population 259****12. Comparing two means 279****INTERLEAF 7 Which test should I use? 315****13. Handling violations of assumptions 319****14. Designing experiments 357****INTERLEAF 8 Data dredging****15. Comparing means of more than two groups 393****INTERLEAF 9 Experimental and statistical mistakes 429****PART 4 REGRESSION AND CORRELATION****16. Correlation between numerical variables 431****INTERLEAF 10 Publication bias 460****17. Regression 463****INTERLEAF 11 Using species as data points 509****PART 5 MODERN STATISTICAL METHODS****18. Multiple explanatory variables 513****19. Computer-intensive methods 539****20. Likelihood 567****21. Meta-analysis: combining information from multiple studies 593**

Answers to practice problems 613

Literature cited 651

Statistical tables 667

Photo credits 689

Index 691

Contents

Preface	xxvii
Acknowledgments	xxxi
About the authors	xxxiii

PART 1 INTRODUCTION TO STATISTICS

1. Statistics and samples 1

1.1	What is statistics?	1
1.2	Sampling populations	3
	Example 1.2: Raining cats	3
	Populations and samples	4
	Properties of good samples	5
	Random sampling	7
	How to take a random sample	7
	The sample of convenience	8
	Volunteer bias	10
	Real data in biology	11
1.3	Types of data and variables	11
	Categorical and numerical variables	11
	Explanatory and response variables	13
1.4	Frequency distributions and probability distributions	13
1.5	Types of studies	15
1.6	Summary	16
	Practice problems	17
	Assignment Problems	18

INTERLEAF 1 Biology and the history of statistics 20

2. Displaying data 23

2.1	Displaying frequency distributions	24
	Frequency tables and bar graphs for categorical data	25

	Example 2.1A: Causes of teenage deaths	25
	Construction rules for bar graphs	26
	Frequency tables and histograms for numerical data	27
	Example 2.1B: Abundance of desert bird species	27
	Describing the shape of a histogram	30
	Interval width can affect histogram shape	31
	Example 2.1C: How many peaks?	31
	Construction rules for histograms	32
2.2	Quantiles of a frequency distribution	33
	Percentiles and quantiles	33
	Cumulative frequency distribution	33
2.3	Associations between categorical variables	34
	Example 2.3: Reproductive effort and avian malaria	35
	Contingency tables	35
	Grouped bar graph	36
	Mosaic plot	37
2.4	Comparing numerical variables between groups	38
	Comparing histograms between groups	38
	Example 2.4: Blood responses to high elevation	38
	Comparing cumulative frequencies	40
2.5	Displaying relationships between a pair of numerical variables	40
	Scatter plot	40
	Example 2.5A: Sins of the father	41
	Line graph	42
	Example 2.5B: Cyclic fluctuations in lynx numbers	42
	Maps	43
	Example 2.5C: The Antarctic ozone hole	43
2.6	Principles of effective display	44
	Principles of graphical display	44
	Follow similar principles in display tables	47
2.7	Summary	48
	Practice problems	49
	Assignment problems	53
3.	Describing data	59
3.1	Arithmetic mean and standard deviation	60
	Example 3.1: Gliding snakes	60
	The sample mean	61
	Variance and standard deviation	62

	Rounding means, standard deviations, and other quantities	64
	Coefficient of variation	64
	Calculating mean and standard deviation from a frequency table	65
3.2	Median and interquartile range	66
	Example 3.2: I'd give my right arm for a female	66
	The median	67
	The interquartile range	68
	The box plot	69
3.3	How measures of location and spread compare	70
	Example 3.3: Disarming fish	70
	Mean versus median	71
	Standard deviation versus interquartile range	72
3.4	Proportions	73
	Calculating a proportion	73
	The proportion is like a sample mean	74
3.5	Summary	74
3.6	Quick Formula Summary	75
	Practice problems	76
	Assignment problems	79
4.	Estimating with uncertainty	83
4.1	The sampling distribution of an estimate	84
	Example 4.1: The length of human genes	84
	Estimating mean gene length with a random sample	86
	The sampling distribution of \bar{Y}	87
4.2	Measuring the uncertainty of an estimate	89
	Standard error	89
	The standard error of \bar{Y}	89
	The standard error of \bar{Y} from data	90
4.3	Confidence intervals	90
	The 2SE rule of thumb	92
4.4	Summary	93
4.5	Quick Formula Summary	93
	Practice problems	94
	Assignment problems	95
INTERLEAF 2	Pseudoreplication	97

5. Probability 99

- 5.1 **The probability of an event 99**
- 5.2 **Venn diagrams 101**
- 5.3 **Mutually exclusive events 101**
- 5.4 **Probability distributions 102**
 - Discrete probability distributions 102
 - Continuous probability distributions 103
- 5.5 **Either this or that: adding probabilities 104**
 - The addition rule 104
 - The probabilities of all possible mutually exclusive events add to one 106
 - The general addition rule 106
- 5.6 **Independence and the multiplication rule 107**
 - Multiplication rule 108
 - Example 5.6A: Smoking and high blood pressure 108**
 - And versus or 109
 - Independence of more than two events 109
 - Example 5.6B: This thing ate my money! 110**
 - Example 5.6C: Mendel's peas 110**
- 5.7 **Probability trees 111**
 - Example 5.7: Sex and birth order 111**
- 5.8 **Dependent events 112**
 - Example 5.8: Is this meat taken? 112**
- 5.9 **Conditional probability and Bayes' theorem 115**
 - Conditional probability 115
 - The general multiplication rule 117
 - Bayes' theorem 117
 - Example 5.9: Detection of Down syndrome 118**
- 5.10 **Summary 120**
 - Practice problems 121**
 - Assignment problems 124**

6. Hypothesis testing 127

- 6.1 **Making and using hypotheses 128**
 - Null hypothesis 128
 - Alternative hypothesis 129
 - To reject or not to reject 130
- 6.2 **Hypothesis testing: an example 130**

	Example 6.2: The right hand of toad	130
	Stating the hypotheses	131
	The test statistic	131
	The null distribution	132
	Quantifying uncertainty: the P -value	134
	Statistical significance	135
	Reporting the results	136
6.3	Errors in hypothesis testing	137
	Type I and Type II errors	137
6.4	When the null hypothesis is not rejected	138
	Example 6.4: The genetics of mirror-image flowers	138
	The test	139
	Interpreting a non-significant result	140
6.5	One-sided tests	141
6.6	Hypothesis testing versus confidence intervals	143
6.7	Summary	144
	Practice problems	145
	Assignment problems	146
INTERLEAF 3	Why statistical significance is not the same as biological importance	148
PART 2	PROPORTIONS AND FREQUENCIES	
7.	Analyzing proportions	151
7.1	The binomial distribution	152
	Formula for the binomial distribution	152
	Number of successes in a random sample	153
	Sampling distribution of the proportion	156
7.2	Testing a proportion: the binomial test	157
	EXAMPLE 7.2: Sex and the X	158
	Approximations for the binomial test	160
7.3	Estimating proportions	161
	Example 7.3: Radiologists' missing sons	161
	Estimating the standard error of a proportion	161
	Confidence intervals for proportions— the Agresti–Coull method	162
	Confidence intervals for proportions—the Wald method	163
7.4	Deriving the binomial distribution	163

- 7.5 Summary 165
- 7.6 Quick Formula Summary 165
 - Practice problems 166
 - Assignment problems 170

INTERLEAF 4 Correlation does not require causation 173

8. Fitting probability models to frequency data 175

- 8.1 Example of a random model: the proportional model 176
 - Example 8.1: No weekend getaway 176**
- 8.2 χ^2 goodness-of-fit test 177
 - Null and alternative hypotheses 177
 - Observed and expected frequencies 178
 - The χ^2 test statistic 178
 - The sampling distribution of χ^2 under the null hypothesis 180
 - Calculating the P -value 182
 - Critical values for the χ^2 distribution 182
- 8.3 Assumptions of the χ^2 goodness-of-fit test 184
- 8.4 Goodness-of-fit tests when there are only two categories 185
 - Example 8.4: Gene content of the human X chromosome 185**
- 8.5 Fitting the binomial distribution 187
 - Example 8.5: Designer two-child families? 187**
- 8.6 Random in space or time: the Poisson distribution 190
 - Formula for the Poisson distribution 192
 - Testing randomness with the Poisson distribution 192
 - Example 8.6: Mass extinctions 193**
 - Comparing the variance with the mean 196
- 8.7 Summary 197
- 8.8 Quick Formula Summary 198
 - Practice problems 198
 - Assignment problems 201

INTERLEAF 5 Making a plan 204

9. Contingency analysis: associations between categorical variables 207

- 9.1 Associating two categorical variables 208
- 9.2 Estimating association in 2×2 tables: odds ratio 209
 - Odds 209

- Example 9.2: Take two aspirin and call me in the morning?** 210
 - Odds ratio 211
 - Standard error and confidence interval for odds ratio 212
- 9.3 The χ^2 contingency test** 213
 - Example 9.3: The gnarly worm gets the bird** 213
 - Hypotheses 215
 - Expected frequencies assuming independence 215
 - The χ^2 statistic 216
 - Degrees of freedom 216
 - P-value and conclusion 217
 - A shortcut for calculating the expected frequencies 217
 - The χ^2 contingency test is a special case of the χ^2 goodness-of-fit test 218
 - Assumptions of the χ^2 contingency test 218
 - Correction for continuity 219
- 9.4 Fisher's exact test** 219
 - Example 9.4: The feeding habits of vampire bats** 220
- 9.5 G-tests** 221
- 9.6 Summary** 222
- 9.7 Quick Formula Summary** 223
 - Practice problems 224
 - Assignment problems 227

PART 3: COMPARING NUMERICAL VALUES

- 10. The normal distribution** 231
 - 10.1 Bell-shaped curves and the normal distribution** 232
 - 10.2 The formula for the normal distribution** 234
 - 10.3 Properties of the normal distribution** 235
 - 10.4 The standard normal distribution and statistical tables** 236
 - Using the standard normal table 237
 - Using the standard normal to describe any normal distribution 239
 - Example 10.4: One small step for man?** 240
 - 10.5 The normal distribution of sample means** 242
 - Calculating probabilities of sample means 244
 - 10.6 Central limit theorem** 245
 - Example 10.6: Pushing your buttons** 245

10.7 Normal approximation for the binomial distribution 247**Example 10.7: The only good bug is a dead bug 248****10.8 Summary 250****10.9 Quick Formula Summary 250****Practice problems 251****Assignment problems 254****INTERLEAF 6 Controls in medical studies 256****11. Inference for a normal population 259****11.1 The t -distribution for sample means 259**Student's t -distribution 260Finding critical values of the t -distribution 262**11.2 The confidence interval for the mean of a normal distribution 263****Example 11.2: Eye to eye 263**

The 95% confidence interval for the mean 264

The 99% confidence interval for the mean 265

11.3 The one-sample t -test 266**Example 11.3: Human body temperature 266**

The effects of larger sample size—body temperature revisited 270

11.4 The assumptions of the one-sample t -test 270**11.5 Estimating the standard deviation and variance of a normal population 271**

Confidence limits for the variance 271

Confidence limits for the standard deviation 273

Assumptions 273

11.6 Summary 273**11.7 Quick Formula Summary 274****Practice problems 275****Assignment problems 277****12. Comparing two means 279****12.1 Paired sample versus two independent samples 280****12.2 Paired comparison of means 281**

Estimating mean difference from paired data 282

Example 12.2: So macho it makes you sick? 282Paired t -test 285

Assumptions 287

12.3 Two-sample comparison of means 287

	Example 12.3: Spike or be spiked	287
	Confidence interval for the difference between two means	288
	Two-sample t -test	291
	Assumptions	292
	A two-sample test when standard deviations are unequal	293
12.4	Using the correct sampling units	293
	Example 12.4: So long; thanks to all the fish	294
12.5	The fallacy of indirect comparison	297
	Example 12.5: Mommy's baby, Daddy's maybe	297
12.6	Interpreting overlap of confidence intervals	299
12.7	Comparing variances	300
	The F -test of equal variances	300
	Levene's test for homogeneity of variances	301
12.8	Summary	301
12.9	Quick Formula Summary	302
	Practice problems	306
	Assignment problems	311
INTERLEAF 7	Which test should I use?	315
13.	Handling violations of assumptions	319
13.1	Detecting deviations from normality	320
	Graphical methods	320
	Example 13.1: The benefits of marine reserves	323
	Formal test of normality	324
13.2	When to ignore violations of assumptions	325
	Violations of normality	325
	Unequal standard deviations	326
13.3	Data transformations	326
	Log transformation	327
	Arcsine transformation	330
	The square-root transformation	331
	Other transformations	331
	Confidence intervals with transformations	331
	A caveat: avoid multiple testing with transformations	332
13.4	Nonparametric alternatives to one-sample and paired t-tests	333
	Sign test	333
	Example 13.4: Sexual conflict and the origin of new species	334
	The Wilcoxon signed-rank test	337

- 13.5 **Comparing two groups: the Mann–Whitney *U*-test** 337
 - Example 13.5: Sexual cannibalism in sagebrush crickets** 338
 - Tied ranks 341
 - Large samples and the normal approximation 342
- 13.6 **Assumptions of nonparametric tests** 343
- 13.7 **Type I and Type II error rates of nonparametric methods** 343
- 13.8 **Summary** 344
- 13.9 **Quick Formula Summary** 345
 - Practice problems** 346
 - Assignment problems** 351
- 14. **Designing experiments** 357
 - 14.1 **Why do experiments?** 358
 - Confounding variables 358
 - Experimental artifacts 359
 - 14.2 **Lessons from clinical trials** 359
 - Example 14.2: Reducing HIV transmission** 360
 - Design components 360
 - 14.3 **How to reduce bias** 362
 - Simultaneous control group 362
 - Randomization 363
 - Blinding 364
 - 14.4 **How to reduce the influence of sampling error** 366
 - Replication 366
 - Balance 368
 - Blocking 369
 - Example 14.4A: Holey waters** 370
 - Extreme treatments 371
 - Example 14.4B: Plastic hormones** 371
 - 14.5 **Experiments with more than one factor** 372
 - Example 14.5: Lethal combination** 373
 - 14.6 **What if you can't do experiments?** 374
 - Match and adjust 375
 - 14.7 **Choosing a sample size** 375
 - Plan for precision 376
 - Plan for power 378
 - Plan for data loss 379
 - 14.8 **Summary** 380

	14.9 Quick Formula Summary	381
	Practice problems	384
	Assignment problems	386
INTERLEAF 8	Data dredging	390
	15. Comparing means of more than two groups	393
	15.1 The analysis of variance	394
	Example 15.1: The knees who say night	394
	Hypotheses	395
	ANOVA in a nutshell	396
	Calculating the mean squares	397
	The variance ratio, F	399
	ANOVA tables	401
	Variability explained: R^2	403
	ANOVA with two groups	403
	15.2 Assumptions and alternatives	403
	The robustness of ANOVA	403
	Data transformations	403
	Nonparametric alternatives to ANOVA	404
	15.3 Planned comparisons	404
	Planned comparison between two means	405
	15.4 Unplanned comparisons	406
	Example 15.4: Wood wide web	407
	Testing all pairs of means using the Tukey–Kramer method	408
	Assumptions	410
	15.5 Fixed and random effects	411
	15.6 ANOVA with randomly chosen groups	412
	Example 15.6: Walking stick limbs	412
	ANOVA table	413
	Variance components	414
	Repeatability	415
	Assumptions	416
	15.7 Summary	416
	15.8 Quick Formula Summary	417
	Practice problems	420
	Assignment problems	424
INTERLEAF 9	Experimental and statistical mistakes	429

PART 4 REGRESSION AND CORRELATION

16. Correlation between numerical variables 431

16.1 Estimating a linear correlation coefficient 432

The correlation coefficient 432

Example 16.1: Manly digits 434

Standard error 436

Approximate confidence interval 436

16.2 Testing the null hypothesis of zero correlation 438

Example 16.2: What big inbreeding coefficients you have 438

16.3 Assumptions 440

16.4 The correlation coefficient depends on the range 442

16.5 Spearman's rank correlation 443

Example 16.5: The miracles of memory 444

Procedure for large n 446

Assumptions of Spearman's correlation 447

16.6 The effects of measurement error on correlation 447

16.7 Summary 448

16.8 Quick Formula Summary 449

Practice problems 452

Assignment problems 455

INTERLEAF 10 Publication bias 460

17. Regression 463

17.1 Linear regression 465

Example 17.1: The lion's nose 465

The method of least squares 466

Formula for the line 467

Calculating the slope and intercept 468

Populations and samples 469

Predicted values 470

Residuals 470

Standard error of slope 472

Confidence interval for the slope 472

17.2 Confidence in predictions 472

Confidence intervals for predictions 473

Extrapolation 474

17.3 Testing hypotheses about a slope 475

	Example 17.3: Chickadee alarms	476
	The t -test of regression slope	477
	The ANOVA approach	478
	Using R^2 to measure the fit of the line to data	479
17.4	Regression toward the mean	480
17.5	Assumptions of regression	482
	Outliers	483
	Detecting non-linearity	483
	Detecting non-normality and unequal variance	484
17.6	Transformations	485
17.7	The effects of measurement error on regression	487
17.8	Nonlinear regression	488
	A curve with an asymptote	488
	Quadratic curves	489
	Formula-free curve fitting	490
	Example 17.8: The incredible shrinking seal	490
	Fitting a binary response variable	492
17.9	Summary	493
17.10	Quick Formula Summary	494
	Practice problems	498
	Assignment problems	503
INTERLEAF 11	Using species as data points	509

PART 5 MODERN STATISTICAL METHODS

18. Multiple explanatory variables 513

18.1 From linear regression to general linear models 514

 Modeling with linear regression 514

 Generalizing linear regression 516

 Analyzing a categorical treatment variable 517

Example 18.1: I feel your pain 517

18.2 Analyzing experiments with blocking 519

 Analyzing data from a randomized block design 520

Example 18.2: Zooplankton depredation 520

 Model formula 520

 Fitting the model to data 521

18.3 Analyzing factorial designs 522

 Analysis of two fixed factors 523

- Example 18.3: Interaction zone 523**
 - Model formula 524
 - Fitting the model to data 524
 - The importance of distinguishing fixed and random factors 526
 - 18.4 Adjusting for the effects of a covariate 526**
 - Example 18.4: Mole-rat layabouts 527**
 - Testing interaction 528
 - Dropping the interaction term 529
 - 18.5 Assumptions of general linear models 530**
 - 18.6. Summary 532**
 - Practice problems 533**
 - Assignment problems 535**
- 19. Computer-intensive methods 539**
 - 19.1 Hypothesis testing using simulation 540**
 - Example 19.1: How did he know? The nonrandomness of haphazard choice 540**
 - 19.2 Randomization test 543**
 - Example 19.2: Girls just wanna have genetic diversity 545**
 - Assumptions of randomization tests 550
 - 19.3 Bootstrap standard errors and confidence intervals 550**
 - Example 19.3: The language center in chimps' brains 551**
 - Bootstrap standard error 553
 - Confidence intervals by bootstrapping 555
 - Bootstrapping data sets with multiple samples 556
 - Assumptions and limitations of the bootstrap 556
 - 19.4 Summary 556**
 - Practice problems 557**
 - Assignment problems 563**
- 20. Likelihood 567**
 - 20.1 What is likelihood? 568**
 - 20.2 Two uses of likelihood in biology 569**
 - Phylogeny estimation 569
 - Gene mapping 570
 - 20.3 Maximum likelihood estimation 571**
 - Example 20.3: Unruly passengers 571**
 - Probability model 571
 - The likelihood formula 572

	The maximum likelihood estimate	573
	Likelihood-based confidence intervals	574
20.4	Versatility of maximum likelihood estimation	577
	Example 20.4: Conservation scoop	577
	Probability model	577
	The likelihood formula	578
	The maximum likelihood estimate	579
	Bias	580
20.5	Log-likelihood ratio test	580
	Likelihood ratio test statistic	580
	Testing a population proportion	581
20.6	Summary	582
20.7	Quick Formula Summary	583
	Practice problems	584
	Assignment problems	588
21.	Meta-analysis: combining information from multiple studies	593
21.1	What is meta-analysis?	595
	Why repeat a study?	595
21.2	The power of meta-analysis	596
	Example 21.2: Aspirin and myocardial infarction	596
21.3	Meta-analysis can give a balanced view	598
	Example 21.3: The Transylvania effect.	598
21.4	The steps of a meta-analysis	599
	Define the question	599
	Example 21.4: Testosterone and aggression	600
	Review the literature	601
	Compute effect sizes	602
	Determine the average effect size	604
	Calculate confidence intervals and make hypothesis tests	605
	Look for effects of study quality	605
	Look for associations	606
21.5	File-drawer problem	606
21.6	How to make your paper accessible to meta-analysis	607
21.7	Summary	608
21.8	Quick Formula Summary	609
	Practice problems	610
	Assignment problems	611

Answers to practice problems 613

Literature cited 651

Statistical Tables 667

Using statistical tables 667

Statistical Table A: The χ^2 distribution 669

Statistical Table B: The standard normal (Z) distribution 672

Statistical Table C: The Student *t*-distribution 674

Statistical Table D: The *F*-distribution 677

Statistical Table E: Mann–Whitney *U*-distribution 682

Statistical Table F: Tukey–Kramer *q*-distribution 684

**Statistical Table G: Critical values for
the Spearman’s correlation coefficient 686**

Photo credits 689

Index 691