

Contents in brief

Preface	xxvii
Acknowledgments	xxxi
About the authors	xxxiii

PART 1 INTRODUCTION TO STATISTICS

	1. Statistics and samples	1
INTERLEAF 1	Biology and the history of statistics	23
	2. Displaying data	25
	3. Describing data	65
	4. Estimating with uncertainty	95
INTERLEAF 2	Pseudoreplication	115
	5. Probability	117
	6. Hypothesis testing	149
INTERLEAF 3	Why statistical significance is not the same as biological importance	176

PART 2 PROPORTIONS AND FREQUENCIES

	7. Analyzing proportions	179
INTERLEAF 4	Correlation does not require causation	201
	8. Fitting probability models to frequency data	203
INTERLEAF 5	Making a plan	233
	9. Contingency analysis: associations between categorical variables	235
	Review Problems 1	269

PART 3 COMPARING NUMERICAL VALUES

10. The normal distribution 273

INTERLEAF 6 Controls in medical studies 301

11. Inference for a normal population 303

12. Comparing two means 327

INTERLEAF 7 Which test should I use? 366

13. Handling violations of assumptions 369

Review Problems 2 417

14. Designing experiments 423

INTERLEAF 8 Data dredging 456

15. Comparing means of more than two groups 459

INTERLEAF 9 Experimental and statistical mistakes 500

PART 4 REGRESSION AND CORRELATION

16. Correlation between numerical variables 503

INTERLEAF 10 Publication bias 535

17. Regression 539

INTERLEAF 11 Using species as data points 593

Review Problems 3 597

PART 5 MODERN STATISTICAL METHODS

18. Multiple explanatory variables 605

19. Computer-intensive methods 635

20. Likelihood 655

21. Meta-analysis: combining information from multiple studies 681

Statistical tables	701
Literature cited	725
Answers to practice problems	747
Photo credits	803
Index	807

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	8
	The sample of convenience	9
	Volunteer bias	10
	Data in the real world	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	19
INTERLEAF 1	Biology and the history of statistics	23
2.	Displaying data	25
2.1	Guidelines for effective graphs	26
	How to draw a bad graph	26
	How to draw a good graph	27
2.2	Showing data for one variable	30
	Showing categorical data: frequency table and bar graph	30

- Example 2.2A: Crouching tiger 31
 - Making a good bar graph 32
 - A bar graph is usually better than a pie chart 33
 - Showing numerical data: frequency table and histogram 33
- Example 2.2B: Abundance of desert bird species 34
 - Describing the shape of a histogram 36
 - How to draw a good histogram 37
 - Other graphs for numerical data 38
- 2.3 Showing association between two variables 38
 - Showing association between categorical variables 38
 - Example 2.3A: Reproductive effort and avian malaria 39
 - Showing association between numerical variables: scatter plot 42
 - Example 2.3B: Sins of the father 42
 - Showing association between a numerical and a categorical variable 43
 - Example 2.3C: Blood responses to high elevation 43
- 2.4 Showing trends in time and space 46
 - Line graph 46
 - Example 2.4A: Bad science can be deadly 46
 - Maps 47
 - Example 2.4B: Biodiversity hotspots 48
- 2.5 How to make good tables 48
 - Follow similar principles for display tables 49
- 2.6 Summary 51
 - Practice problems 52
 - Assignment problems 58
- 3. Describing data 65
 - 3.1 Arithmetic mean and standard deviation 66
 - Example 3.1: Gliding snakes 66
 - The sample mean 67
 - Variance and standard deviation 68
 - Rounding means, standard deviations, and other quantities 70
 - Coefficient of variation 70
 - Calculating mean and standard deviation from a frequency table 71
 - Effect of changing measurement scale 72
 - 3.2 Median and interquartile range 73
 - Example 3.2: I'd give my right arm for a female 74
 - The median 74

	The interquartile range	75
	The box plot	76
3.3	How measures of location and spread compare	77
	Example 3.3: Disarming fish	77
	Mean versus median	78
	Standard deviation versus interquartile range	80
3.4	Cumulative frequency distribution	80
	Percentiles and quantiles	80
	Displaying cumulative relative frequencies	81
3.5	Proportions	82
	Calculating a proportion	82
	The proportion is like a sample mean	83
3.6	Summary	83
3.7	Quick Formula Summary	84
	Practice problems	85
	Assignment problems	89
4.	Estimating with uncertainty	95
4.1	The sampling distribution of an estimate	96
	Example 4.1: The length of human genes	96
	Estimating mean gene length with a random sample	98
	The sampling distribution of \bar{Y}	99
4.2	Measuring the uncertainty of an estimate	100
	Standard error	101
	The standard error of \bar{Y}	101
	The standard error of \bar{Y} from data	101
4.3	Confidence intervals	102
	The 2SE rule of thumb	104
4.4	Error bars	104
4.5	Summary	106
4.6	Quick Formula Summary	107
	Practice problems	107
	Assignment problems	110
INTERLEAF 2	Pseudoreplication	115
5.	Probability	117
5.1	The probability of an event	117
5.2	Venn diagrams	119
5.3	Mutually exclusive events	120

- 5.4 **Probability distributions** 120
 - Discrete probability distributions 121
 - Continuous probability distributions 121
- 5.5 **Either this or that: adding probabilities** 123
 - The addition rule 123
 - The probabilities of all possible mutually exclusive outcomes add to one 124
 - The general addition rule 125
- 5.6 **Independence and the multiplication rule** 126
 - Multiplication rule 127
 - Example 5.6A: Smoking and high blood pressure 127
 - “And” versus “or” 128
 - Independence of more than two events 128
 - Example 5.6B: Mendel’s peas 128
- 5.7 **Probability trees** 129
 - Example 5.7: Sex and birth order 129
- 5.8 **Dependent events** 131
 - Example 5.8: Is this meat taken? 131
- 5.9 **Conditional probability and Bayes’ theorem** 133
 - Conditional probability 134
 - The general multiplication rule 135
 - Sampling without replacement 136
 - Bayes’ theorem 136
 - Example 5.9: Detection of Down syndrome 137
- 5.10 **Summary** 138
 - Practice problems 139
 - Assignment problems 144
- 6. **Hypothesis testing** 149
 - 6.1 **Making and using hypotheses** 150
 - Null hypothesis 151
 - Alternative hypothesis 152
 - To reject or not to reject 152
 - 6.2 **Hypothesis testing: an example** 153
 - Example 6.2: The right hand of toad 153
 - Stating the hypotheses 153
 - The test statistic 154
 - The null distribution 154
 - Quantifying uncertainty: the P -value 156
 - Draw the appropriate conclusion 158
 - Reporting the results 159

- 6.3 Errors in hypothesis testing 159
 - Type I and Type II errors 160
- 6.4 When the null hypothesis is not rejected 161
 - Example 6.4: The genetics of mirror-image flowers 161
 - The test 162
 - Interpreting a non-significant result 163
- 6.5 One-sided tests 164
- 6.6 Hypothesis testing versus confidence intervals 166
- 6.7 Summary 167
 - Practice problems 168
 - Assignment problems 170
- INTERLEAF 3 Why statistical significance is not the same as biological importance 176

- PART 2 PROPORTIONS AND FREQUENCIES**
- 7. Analyzing proportions 179
 - 7.1 The binomial distribution 180
 - Formula for the binomial distribution 180
 - Number of successes in a random sample 182
 - Sampling distribution of the proportion 183
 - 7.2 Testing a proportion: the binomial test 185
 - Example 7.2: Sex and the X 185
 - Approximations for the binomial test 188
 - 7.3 Estimating proportions 188
 - Example 7.3: Radiologists' missing sons 189
 - Estimating the standard error of a proportion 189
 - Confidence intervals for proportions—
the Agresti–Coull method 189
 - Confidence intervals for proportions—the Wald method 190
 - 7.4 Deriving the binomial distribution 191
 - 7.5 Summary 192
 - 7.6 Quick Formula Summary 192
 - Practice problems 193
 - Assignment problems 198
- INTERLEAF 4 Correlation does not require causation 201

8.	Fitting probability models to frequency data	203
8.1	Example of a probability model: the proportional model	204
	Example 8.1: No weekend getaway	204
8.2	χ^2 goodness-of-fit test	205
	Null and alternative hypotheses	205
	Observed and expected frequencies	206
	The χ^2 test statistic	207
	The sampling distribution of χ^2 under the null hypothesis	208
	Calculating the P -value	209
	Critical values for the χ^2 distribution	210
8.3	Assumptions of the χ^2 goodness-of-fit test	212
8.4	Goodness-of-fit tests when there are only two categories	213
	Example 8.4: Gene content of the human X chromosome	213
8.5	Fitting the binomial distribution	214
	Example 8.5: Designer two-child families?	215
8.6	Random in space or time: the Poisson distribution	217
	Formula for the Poisson distribution	219
	Testing randomness with the Poisson distribution	219
	Example 8.6: Mass extinctions	219
	Comparing the variance to the mean	223
8.7	Summary	223
8.8	Quick Formula Summary	224
	Practice problems	225
	Assignment problems	228
INTERLEAF 5	Making a plan	233
9.	Contingency analysis: associations between categorical variables	235
9.1	Associating two categorical variables	236
9.2	Estimating association in 2×2 tables: odds ratio	236
	Odds	237
	Example 9.2: Take two aspirin and call me in the morning?	238
	Odds ratio	239
	Standard error and confidence interval for odds ratio	240
9.3	Estimating association in 2×2 table: relative risk	242
	Odds ratio vs. relative risk	242
	Example 9.3: Your litter box and your brain	243
9.4	The χ^2 contingency test	245
	Example 9.4: The gnarly worm gets the bird	246

- Hypotheses 247
- Expected frequencies assuming independence 247
- The χ^2 statistic 249
- Degrees of freedom 249
- P*-value and conclusion 249
- A shortcut for calculating the expected frequencies 250
- The χ^2 contingency test is a special case of the χ^2 goodness-of-fit test 251
- Assumptions of the χ^2 contingency test 251
- Correction for continuity 251
- 9.5 Fisher's exact test 252
 - Example 9.5: The feeding habits of vampire bats 252
- 9.6 G-tests 254
- 9.7 Summary 255
- 9.8 Quick Formula Summary 255
 - Practice problems 257
 - Assignment problems 262
- Review Problems 1 269

PART 3: COMPARING NUMERICAL VALUES

- 10. The normal distribution 273
 - 10.1 Bell-shaped curves and the normal distribution 274
 - 10.2 The formula for the normal distribution 276
 - 10.3 Properties of the normal distribution 277
 - 10.4 The standard normal distribution and statistical tables 278
 - Using the standard normal table 279
 - Using the standard normal to describe any normal distribution 280
 - Example 10.4: One small step for man? 281
 - 10.5 The normal distribution of sample means 283
 - Calculating probabilities of sample means 285
 - 10.6 Central limit theorem 286
 - Example 10.6: Young adults and the Spanish flu 286
 - 10.7 Normal approximation to the binomial distribution 289
 - Example 10.7: The only good bug is a dead bug 289
 - 10.8 Summary 292
 - 10.9 Quick Formula Summary 293

	Practice problems	293
	Assignment problems	297
INTERLEAF 6	Controls in medical studies	301
11.	Inference for a normal population	303
11.1	The t -distribution for sample means	304
	Student's t -distribution	304
	Finding critical values of the t -distribution	306
11.2	The confidence interval for the mean of a normal distribution	307
	Example 11.2: Eye to eye	307
	The 95% confidence interval for the mean	308
	The 99% confidence interval for the mean	309
11.3	The one-sample t -test	310
	Example 11.3: Human body temperature	310
	The effects of larger sample size—body temperature revisited	313
11.4	Assumptions of the one-sample t -test	314
11.5	Estimating the standard deviation and variance of a normal population	314
	Confidence limits for the variance	315
	Confidence limits for the standard deviation	316
	Assumptions	316
11.6	Summary	317
11.7	Quick Formula Summary	317
	Practice problems	318
	Assignment problems	321
12.	Comparing two means	327
12.1	Paired sample versus two independent samples	328
12.2	Paired comparison of means	329
	Estimating mean difference from paired data	330
	Example 12.2: So macho it makes you sick?	330
	Paired t -test	333
	Assumptions	334
12.3	Two-sample comparison of means	335
	Example 12.3: Spike or be spiked	335
	Confidence interval for the difference between two means	336
	Two-sample t -test	338
	Assumptions	340
	A two-sample t -test when standard deviations are unequal	341

- 12.4 Using the correct sampling units 341
 - Example 12.4: So long; thanks to all the fish 342
- 12.5 The fallacy of indirect comparison 344
 - Example 12.5: Mommy's baby, Daddy's maybe 345
- 12.6 Interpreting overlap of confidence intervals 346
- 12.7 Comparing variances 347
 - The *F*-test of equal variances 348
 - Levene's test for homogeneity of variances 348
- 12.8 Summary 349
- 12.9 Quick Formula Summary 350
 - Practice problems 353
 - Assignment problems 360
- INTERLEAF 7 Which test should I use? 366
- 13. Handling violations of assumptions 369
 - 13.1 Detecting deviations from normality 370
 - Graphical methods 370
 - Example 13.1: The benefits of marine reserves 373
 - Formal test of normality 374
 - 13.2 When to ignore violations of assumptions 375
 - Violations of normality 375
 - Unequal standard deviations 376
 - 13.3 Data transformations 377
 - Log transformation 377
 - Arcsine transformation 380
 - The square-root transformation 381
 - Other transformations 381
 - Confidence intervals with transformations 381
 - A caveat: Avoid multiple testing with transformations 382
 - 13.4 Nonparametric alternatives to one-sample and paired *t*-tests 383
 - Sign test 383
 - Example 13.4: Sexual conflict and the origin of new species 384
 - The Wilcoxon signed-rank test 387
 - 13.5 Comparing two groups: the Mann–Whitney *U*-test 387
 - Example 13.5: Sexual cannibalism in sagebrush crickets 387
 - Tied ranks 391
 - Large samples and the normal approximation 392
 - 13.6 Assumptions of nonparametric tests 392

13.7	Type I and Type II error rates of nonparametric methods	393
13.8	Permutation tests	394
	Assumptions of permutation tests	397
13.9	Summary	398
13.10	Quick Formula Summary	399
	Practice problems	400
	Assignment problems	408
	Review Problems 2	417
14.	Designing experiments	423
14.1	Why do experiments?	424
	Confounding variables	424
	Experimental artifacts	425
14.2	Lessons from clinical trials	426
	Example 14.2: Reducing HIV transmission	426
	Design components	426
14.3	How to reduce bias	428
	Simultaneous control group	428
	Randomization	429
	Blinding	430
14.4	How to reduce the influence of sampling error	431
	Replication	432
	Balance	434
	Blocking	434
	Example 14.4A: Holey waters	436
	Extreme treatments	437
	Example 14.4B: Plastic hormones	437
14.5	Experiments with more than one factor	438
	Example 14.5: Lethal combination	439
14.6	What if you can't do experiments?	440
	Match and adjust	440
14.7	Choosing a sample size	441
	Plan for precision	442
	Plan for power	444
	Plan for data loss	445
14.8	Summary	445
14.9	Quick Formula Summary	447
	Practice problems	450
	Assignment problems	453
	INTERLEAF 8 Data dredging	456

- 15. Comparing means of more than two groups 459
 - 15.1 The analysis of variance 460
 - Example 15.1: The knees who say night 460
 - Hypotheses 461
 - ANOVA in a nutshell 462
 - ANOVA tables 463
 - Partitioning the sum of squares 464
 - Calculating the mean squares 466
 - The variance ratio, F 467
 - Variation explained: R^2 469
 - ANOVA with two groups 470
 - 15.2 Assumptions and alternatives 470
 - The robustness of ANOVA 470
 - Data transformations 471
 - Nonparametric alternatives to ANOVA 471
 - 15.3 Planned comparisons 471
 - Planned comparison between two means 472
 - 15.4 Unplanned comparisons 473
 - Example 15.4: Wood wide web 474
 - Testing all pairs of means using the Tukey–Kramer method 475
 - Assumptions 477
 - 15.5 Fixed and random effects 477
 - 15.6 ANOVA with randomly chosen groups 478
 - Example 15.6: Walking stick limbs 478
 - ANOVA calculations 479
 - Variance components 480
 - Repeatability 481
 - Assumptions 482
 - 15.7 Summary 482
 - 15.8 Quick Formula Summary 483
 - Practice problems 486
 - Assignment problems 492
- INTERLEAF 9 Experimental and statistical mistakes 500

PART 4 REGRESSION AND CORRELATION

- 16. Correlation between numerical variables 503
 - 16.1 Estimating a linear correlation coefficient 504
 - The correlation coefficient 504
 - Example 16.1: Flipping the bird 506

	Standard error	508
	Approximate confidence interval	508
16.2	Testing the null hypothesis of zero correlation	510
	Example 16.2: What big inbreeding coefficients you have	510
16.3	Assumptions	512
16.4	The correlation coefficient depends on the range	514
16.5	Spearman's rank correlation	515
	Example 16.5: The miracles of memory	516
	Procedure for large n	518
	Assumptions of Spearman's correlation	519
16.6	The effects of measurement error on correlation	519
16.7	Summary	520
16.8	Quick Formula Summary	521
	Practice problems	524
	Assignment problems	529
INTERLEAF 10	Publication bias	535
17.	Regression	539
17.1	Linear regression	541
	Example 17.1: The lion's nose	541
	The method of least squares	542
	Formula for the line	543
	Calculating the slope and intercept	544
	Populations and samples	545
	Predicted values	546
	Residuals	546
	Standard error of slope	547
	Confidence interval for the slope	548
17.2	Confidence in predictions	548
	Confidence intervals for predictions	549
	Extrapolation	550
17.3	Testing hypotheses about a slope	551
	Example 17.3: Prairie home champion	551
	The t -test of regression slope	552
	The ANOVA approach	554
	Using R^2 to measure the fit of the line to data	555
17.4	Regression toward the mean	555
17.5	Assumptions of regression	557
	Outliers	558

	Detecting nonlinearity	559
	Detecting non-normality and unequal variance	559
17.6	Transformations	560
17.7	The effects of measurement error on regression	562
17.8	Nonlinear regression	563
	A curve with an asymptote	564
	Quadratic curves	565
	Formula-free curve fitting	566
	Example 17.8: The incredible shrinking seal	566
17.9	Logistic regression: fitting a binary response variable	567
17.10	Summary	571
17.11	Quick Formula Summary	572
	Practice problems	576
	Assignment problems	582
INTERLEAF 11	Using species as data points	593
	Review Problems 3	597

PART 5 MODERN STATISTICAL METHODS

18.	Multiple explanatory variables	605
18.1	ANOVA and linear regression are linear models	606
	Modeling with linear regression	607
	Generalizing linear regression	608
	General linear models	610
18.2	Analyzing experiments with blocking	611
	Analyzing data from a randomized block design	611
	Example 18.2: Zooplankton depredation	612
	Model formula	612
	Fitting the model to data	613
18.3	Analyzing factorial designs	614
	Example 18.3: Interaction zone	615
	Model formula	616
	Testing the factors	617
	The importance of distinguishing fixed and random factors	619
18.4	Adjusting for the effects of a covariate	619
	Example 18.4: Mole-rat layabouts	620
	Testing interaction	621
	Fitting a model without an interaction term	622

- 18.5 Assumptions of general linear models 623
- 18.6 Summary 624
 - Practice problems 626
 - Assignment problems 629
- 19. Computer-intensive methods 635
 - 19.1 Hypothesis testing using simulation 636
 - Example 19.1: How did he know? The nonrandomness of haphazard choice 636
 - 19.2 Bootstrap standard errors and confidence intervals 639
 - Example 19.2: The language center in chimps' brains 640
 - Bootstrap standard error 642
 - Confidence intervals by bootstrapping 644
 - Bootstrapping with multiple groups 644
 - Assumptions and limitations of the bootstrap 646
 - 19.3 Summary 646
 - Practice problems 647
 - Assignment problems 651
- 20. Likelihood 655
 - 20.1 What is likelihood? 656
 - 20.2 Two uses of likelihood in biology 657
 - Phylogeny estimation 657
 - Gene mapping 658
 - 20.3 Maximum likelihood estimation 659
 - Example 20.3: Unruly passengers 659
 - Probability model 660
 - The likelihood formula 660
 - The maximum likelihood estimate 661
 - Likelihood-based confidence intervals 664
 - 20.4 Versatility of maximum likelihood estimation 665
 - Example 20.4: Conservation scoop 665
 - Probability model 666
 - The likelihood formula 666
 - The maximum likelihood estimate 667
 - Bias 668
 - 20.5 Log-likelihood ratio test 668
 - Likelihood ratio test statistic 669
 - Testing a population proportion 669
 - 20.6 Summary 670

20.7	Quick Formula Summary	671
	Practice problems	672
	Assignment problems	676
21.	Meta-analysis: combining information from multiple studies	681
21.1	What is meta-analysis?	683
	Why repeat a study?	683
21.2	The power of meta-analysis	684
	Example 21.2: Aspirin and myocardial infarction	684
21.3	Meta-analysis can give a balanced view	686
	Example 21.3: The Transylvania effect	686
21.4	The steps of a meta-analysis	687
	Define the question	687
	Example 21.4: Testosterone and aggression	688
	Review the literature	688
	Compute effect sizes	689
	Determine the average effect size	692
	Calculate confidence intervals and test hypotheses	692
	Look for effects of study quality	693
	Look for associations	693
21.5	File-drawer problem	694
21.6	How to make your paper accessible to meta-analysis	695
21.7	Summary	695
21.8	Quick Formula Summary	696
	Practice problems	697
	Assignment problems	698
	Statistical Tables	701
	Using statistical tables	701
	Statistical Table A: The χ^2 distribution	703
	Statistical Table B: The standard normal (Z) distribution	706
	Statistical Table C: The Student t -distribution	708
	Statistical Table D: The F -distribution	711
	Statistical Table E: Mann–Whitney U -distribution	718
	Statistical Table F: Tukey–Kramer q -distribution	720
	Statistical Table G: Critical values for the Spearman’s rank correlation	722

Literature cited 725

Answers to practice problems 747

Photo credits 803

Index 807