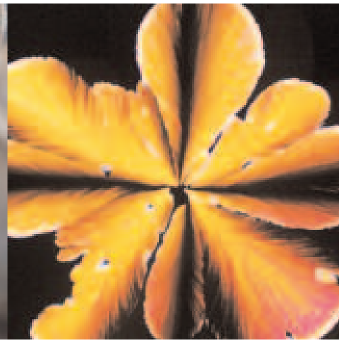
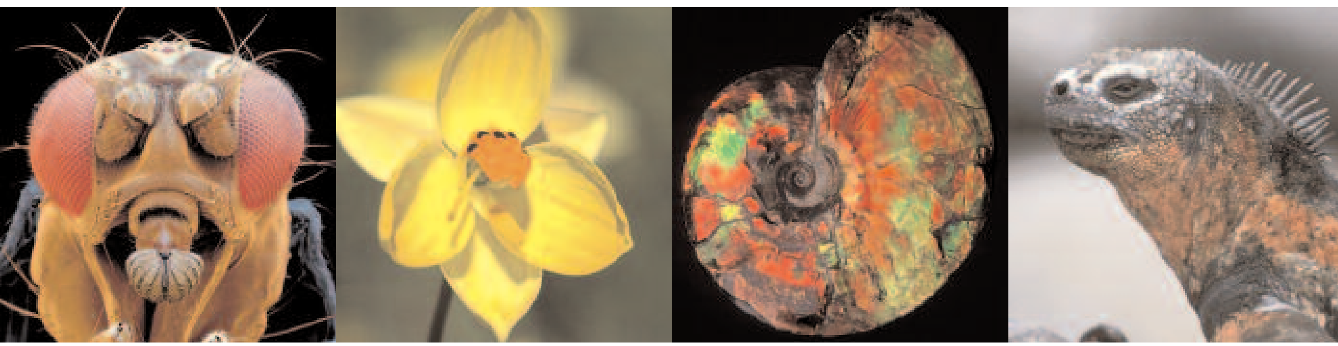
The Analysis of Biological Data






The Analysis of Biological Data

Michael C. Whitlock and Dolph Schluter



ROBERTS AND COMPANY PUBLISHERS
Greenwood Village, Colorado

Roberts and Company Publishers
4950 South Yosemite Street, F2 #197
Greenwood Village, Colorado 80111 USA
Internet: www.roberts-publishers.com
Telephone: (303) 221-3325
Facsimile: (303) 221-3326

ORDER INFORMATION
Telephone: (800) 351-1161 or (516) 422-4050
Facsimile: (516) 422-4097
Internet: www.roberts-publishers.com

Publisher: Ben Roberts
Development Editor and Copyeditor: John Murdzek
Proofreader: Gunder Hefta
Art Studio: Lineworks, Inc.
Text Designer: Mark Ong, Side by Side Studios
Cover Photographer: Christopher Marley
Cover Designer: Mark Ong, Side by Side Studios
Permissions Coordinator: Laura Gabbard Roberts
Compositor: Side by Side Studios

ISBN: 978-0-9815194-0-1

Library of Congress Cataloging-in-Publication Data

10 9 8 7 6 5 4 3 2 1

To Sally and Wilson, Andrea and Maggie

Contents in brief

Preface
Acknowledgments
About the authors

PART 1 INTRODUCTION TO STATISTICS

1. Statistics and samples

INTERLEAF 1 Biology and the history of statistics

2. Displaying data

3. Describing data

4. Estimating with uncertainty

INTERLEAF 2 Pseudoreplication

5. Probability

6. Hypothesis testing

INTERLEAF 3 Why statistical significance is not the same as biological importance

PART 2 PROPORTIONS AND FREQUENCIES

7. Analyzing proportions

INTERLEAF 4 Correlation does not require causation

8. Fitting probability models to frequency data

INTERLEAF 5 Making a plan

9. Contingency analysis: associations between categorical variables

PART 3 COMPARING NUMERICAL VALUES

10. The normal distribution

INTERLEAF 6 Controls in medical studies

11. Inference for a normal population

12. Comparing two means

INTERLEAF 7 Which test should I use?

13. Handling violations of assumptions

14. Designing experiments

INTERLEAF 8 Data dredging

15. Comparing means of more than two groups

INTERLEAF 9 Experimental and statistical mistakes

PART 4 REGRESSION AND CORRELATION

16. Correlation between numerical variables

INTERLEAF 10 Publication bias

17. Regression

INTERLEAF 11 Using species as data points

PART 5 MODERN STATISTICAL METHODS

18. Multiple explanatory variables

19. Computer-intensive methods

20. Likelihood

21. Meta-analysis: combining information from multiple studies

Answers to practice problems

Literature cited

Statistical tables

Photo credits

Index

Contents

Preface

Acknowledgments

About the authors

PART 1 **INTRODUCTION TO STATISTICS**

1. **Statistics and samples**

1.1 **What is statistics?**

1.2 **Sampling populations**

Example 1.2: Raining cats

Populations and samples
Properties of good samples
Random sampling
How to take a random sample
The sample of convenience
Volunteer bias
Real data in biology

1.3 **Types of data and variables**

Categorical and numerical variables
Explanatory and response variables

1.4 **Frequency distributions and probability distributions**

1.5 **Types of studies**

1.6 **Summary**

Practice problems

Assignment Problems

INTERLEAF 1 **Biology and the history of statistics**

2. **Displaying data**

2.1 **Displaying frequency distributions**

Frequency tables and bar graphs for categorical data

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

Rounding means, standard deviations, and other quantities
Coefficient of variation
Calculating mean and standard deviation from a frequency table

3.2 Median and interquartile range

Example 3.2: I'd give my right arm for a female
The median
The interquartile range
The box plot

3.3 How measures of location and spread compare

Example 3.3: Disarming fish
Mean versus median
Standard deviation versus interquartile range

3.4 Proportions
Calculating a proportion
The proportion is like a sample mean

3.5 Summary

3.6 Quick formula summary

Practice problems

Assignment problems

4. Estimating with uncertainty

4.1 The sampling distribution of an estimate

Example 4.1: The length of human genes
Estimating mean gene length with a random sample
The sampling distribution of $\bar{Y}$

4.2 Measuring the uncertainty of an estimate
Standard error
The standard error of $\bar{Y}$
The standard error of $\bar{Y}$ from data

4.3 Confidence intervals
The 2SE rule of thumb

4.4 Summary

4.5 Quick Formula Summary

Practice problems

Assignment problems

INTERLEAF 2 Pseudoreplication

5. Probability
5.1 The probability of an event
5.2 Venn diagrams
5.3 Mutually exclusive events
5.4 Probability distributions
Discrete probability distributions
Continuous probability distributions
5.5 Either this or that: adding probabilities
The addition rule
The probabilities of all possible mutually exclusive events add to one
The general addition rule
5.6 Independence and the multiplication rule
Multiplication rule 1
Example 5.6A: Smoking and high blood pressure
And versus or
Independence of more than two events
Example 5.6B: This thing ate my money!
Example 5.6C: Mendel's peas
5.7 Probability trees
Example 5.7: Sex and birth order
5.8 Dependent events
Example 5.8: Is this meat taken?
5.9 Conditional probability and Bayes' theorem
Conditional probability
The general multiplication rule
Bayes' theorem
Example 5.9: Detection of Down syndrome
5.10 Summary
Practice problems
Assignment problems
6. Hypothesis testing
6.1 Making and using hypotheses
Null hypothesis
Alternative hypothesis
To reject or not to reject
6.2 Hypothesis testing: an example

Example 6.2: The right hand of toad
Stating the hypotheses
The test statistic
The null distribution
Quantifying uncertainty: the *P*-value
Statistical significance
Reporting the results
6.3 Errors in hypothesis testing
Type I and Type II errors
6.4 When the null hypothesis is not rejected
Example 6.4: The genetics of mirror-image flowers 138
The test
Interpreting a non-significant result
6.5 One-sided tests
6.6 Hypothesis testing versus confidence intervals
6.7 Summary
Practice problems
Assignment problems

INTERLEAF 3 Why statistical significance is not the same as biological importance 148

PART 2 PROPORTIONS AND FREQUENCIES

7. Analyzing proportions
7.1 The binomial distribution
Formula for the binomial distribution
Number of successes in a random sample
Sampling distribution of the proportion
7.2 Testing a proportion: the binomial test
EXAMPLE 7.2: Sex and the X
Approximations for the binomial test
7.3 Estimating proportions
Example 7.3: Radiologists' missing sons
Estimating the standard error of a proportion
Confidence intervals for proportions—the Agresti–Coull method
Confidence intervals for proportions—the Wald method
7.4 Deriving the binomial distribution
7.5 Summary

7.6 Quick Formula Summary
Practice problems
Assignment problems

INTERLEAF 4 Correlation does not require causation

8. Fitting probability models to frequency data
8.1 Example of a random model: the proportional model
Example 8.1: No weekend getaway
8.2 χ^2 goodness-of-fit test
Null and alternative hypotheses
Observed and expected frequencies
The χ^2 test statistic
The sampling distribution of χ^2 under the null hypothesis
Calculating the *P*-value
Critical values for the χ^2 distribution
8.3 Assumptions of the χ^2 goodness-of-fit test
8.4 Goodness-of-fit tests when there are only two categories
Example 8.4: Gene content of the human X chromosome
8.5 Fitting the binomial distribution
Example 8.5: Designer two-child families?
8.6 Random in space or time: the Poisson distribution
Formula for the Poisson distribution
Testing randomness with the Poisson distribution
Example 8.6: Mass extinctions
Comparing the variance with the mean
8.7 Summary
8.8 Quick Formula Summary
Practice problems
Assignment problems

INTERLEAF 5 Making a plan

9. Contingency analysis: associations between categorical variables
9.1 Associating two categorical variables
9.2 Estimating association in 2 × 2 tables: odds ratio
Odds
Example 9.2: Take two aspirin and call me in the morning?

Odds ratio
Standard error and confidence interval for odds ratio
9.3 The χ^2 contingency test
Example 9.3: The gnarly worm gets the bird
Hypotheses
Expected frequencies assuming independence
The χ^2 statistic
Degrees of freedom
P-value and conclusion
A shortcut for calculating the expected frequencies
The χ^2 contingency test is a special case of the χ^2 goodness-of-fit test
Assumptions of the χ^2 contingency test
Correction for continuity
9.4 Fisher's exact test
Example 9.4: The feeding habits of vampire bats
9.5 *G*-tests
9.6 Summary
9.7 Quick Formula Summary
Practice problems
Assignment problems

PART 3: COMPARING NUMERICAL VALUES

10. The normal distribution

10.1 Bell-shaped curves and the normal distribution
10.2 The formula for the normal distribution
10.3 Properties of the normal distribution
10.4 The standard normal distribution and statistical tables
Using the standard normal table
Using the standard normal to describe any normal distribution
Example 10.4: One small step for man?
10.5 The normal distribution of sample means
Calculating probabilities of sample means
10.6 Central limit theorem
Example 10.6: Pushing your buttons
10.7 Normal approximation for the binomial distribution
Example 10.7: The only good bug is a dead bug

10.8 Summary

10.9 Quick Formula Summary

Practice problems

Assignment problems

INTERLEAF 6 **Controls in medical studies**

11. Inference for a normal population

11.1 The *t*-distribution for sample means

Student's *t*-distribution

Finding critical values of the *t*-distribution

11.2 The confidence interval for the mean of a normal distribution

Example 11.2: Eye to eye

The 95% confidence interval for the mean

The 99% confidence interval for the mean

11.3 The one-sample *t*-test

Example 11.3: Human body temperature

The effects of larger sample size—body temperature revisited

11.4 The assumptions of the one-sample *t*-test

11.5 Estimating the standard deviation and variance of a normal population

Confidence limits for the variance

Confidence limits for the standard deviation

Assumptions

11.6 Summary

11.7 Quick Formula Summary

Practice problems

Assignment problems

12. Comparing two means

12.1 Paired sample versus two independent samples

12.2 Paired comparison of means

Estimating mean difference from paired data

Example 12.2: So macho it makes you sick?

Paired *t*-test

Assumptions

12.3 Two-sample comparison of means

Example 12.3: Spike or be spiked
Confidence interval for the difference between two means
Two-sample *t*-test
Assumptions
A two-sample test when standard deviations are unequal
12.4 Using the correct sampling units
Example 12.4: So long; thanks to all the fish
12.5 The fallacy of indirect comparison
Example 12.5: Mommy's baby, Daddy's maybe
12.6 Interpreting overlap of confidence intervals
12.7 Comparing variances
The *F*-test of equal variances
Levene's test for homogeneity of variances
12.8 Summary
12.9 Quick Formula Summary
Practice problems
Assignment problems

INTERLEAF 7 **Which test should I use?**

13. Handling violations of assumptions
13.1 Detecting deviations from normality
Graphical methods
Example 13.1: The benefits of marine reserves
Formal test of normality
13.2 When to ignore violations of assumptions
Violations of normality
Unequal standard deviations
13.3 Data transformations
Log transformation
Arcsine transformation
The square-root transformation
Other transformations
Confidence intervals with transformations
A caveat: avoid multiple testing with transformations
13.4 Nonparametric alternatives to one-sample and paired *t*-tests
Sign test
Example 13.4: Sexual conflict and the origin of new species
The Wilcoxon signed-rank test

13.5 **Comparing two groups: the Mann–Whitney *U*-test**
Example 13.5: Sexual cannibalism in sagebrush crickets
Tied ranks
Large samples and the normal approximation
13.6 **Assumptions of nonparametric tests**
13.7 **Type I and Type II error rates of nonparametric methods**
13.8 **Summary**
13.9 **Quick Formula Summary**
Practice problems
Assignment problems

14. Designing experiments

14.1 **Why do experiments?**
Confounding variables
Experimental artifacts
14.2 **Lessons from clinical trials**
Example 14.2: Reducing HIV transmission
Design components
14.3 **How to reduce bias**
Simultaneous control group
Randomization
Blinding
14.4 **How to reduce the influence of sampling error**
Replication
Balance
Blocking
Example 14.4A: Holey waters
Extreme treatments
Example 14.4B: Plastic hormones
14.5 **Experiments with more than one factor**
Example 14.5: Lethal combination
14.6 **What if you can't do experiments?**
Match and adjust
14.7 **Choosing a sample size**
Plan for precision
Plan for power
Plan for data loss
14.8 **Summary**

14.9 **Quick Formula Summary**
Practice problems
Assignment problems

INTERLEAF 8 **Data dredging**

15. Comparing means of more than two groups

15.1 **The analysis of variance**
Example 15.1: The knees who say night
Hypotheses
ANOVA in a nutshell
Calculating the mean squares
The variance ratio, *F*
ANOVA tables
Variability explained: R^2
ANOVA with two groups

15.2 **Assumptions and alternatives**
The robustness of ANOVA
Data transformations
Nonparametric alternatives to ANOVA

15.3 **Planned comparisons**
Planned comparison between two means

15.4 **Unplanned comparisons**
Example 15.4: Wood wide web
Testing all pairs of means using the Tukey–Kramer method
Assumptions

15.5 **Fixed and random effects**

15.6 **ANOVA with randomly chosen groups**
Example 15.6: Walking stick limbs
ANOVA table
Variance components
Repeatability
Assumptions

15.7 **Summary**

15.8 **Quick Formula Summary**
Practice problems
Assignment problems

INTERLEAF 9 **Experimental and statistical mistakes**

PART 4 REGRESSION AND CORRELATION

16. Correlation between numerical variables

16.1 Estimating a linear correlation coefficient
The correlation coefficient
Example 16.1: Manly digits
Standard error
Approximate confidence interval

16.2 Testing the null hypothesis of zero correlation
Example 16.2: What big inbreeding coefficients you have

16.3 Assumptions

16.4 The correlation coefficient depends on the range

16.5 Spearman's rank correlation
Example 16.5: The miracles of memory
Procedure for large *n*
Assumptions of Spearman's correlation

16.6 The effects of measurement error on correlation

16.7 Summary

16.8 Quick Formula Summary

Practice problems

Assignment problems

INTERLEAF 10 Publication bias

17. Regression

17.1 Linear regression
Example 17.1: The lion's nose
The method of least squares
Formula for the line
Calculating the slope and intercept
Populations and samples
Predicted values
Residuals
Standard error of slope
Confidence interval for the slope

17.2 Confidence in predictions
Confidence intervals for predictions
Extrapolation

17.3 Testing hypotheses about a slope

Example 17.3: Chickadee alarms
The *t*-test of regression slope
The ANOVA approach
Using R^2 to measure the fit of the line to data
17.4 Regression toward the mean
17.5 Assumptions of regression
Outliers
Detecting non-linearity
Detecting non-normality and unequal variance
17.6 Transformations
17.7 The effects of measurement error on regression
17.8 Nonlinear regression
A curve with an asymptote
Quadratic curves
Formula-free curve fitting
Example 17.8: The incredible shrinking seal
Fitting a binary response variable
17.9 Summary
17.10 Quick Formula Summary
Practice problems
Assignment problems

INTERLEAF 11 **Using species as data points**

PART 5 **MODERN STATISTICAL METHODS**

18. Multiple explanatory variables
18.1 From linear regression to general linear models
Modeling with linear regression
Generalizing linear regression
Analyzing a categorical treatment variable
Example 18.1: I feel your pain
18.2 Analyzing experiments with blocking
Analyzing data from a randomized block design
Example 18.2: Zooplankton depredation
Model formula
Fitting the model to data
18.3 Analyzing factorial designs
Analysis of two fixed factors

Example 18.3: Interaction zone
Model formula
Fitting the model to data
The importance of distinguishing fixed and random factors
18.4 Adjusting for the effects of a covariate
Example 18.4: Mole-rat layabouts
Testing interaction
Dropping the interaction term
18.5 Assumptions of general linear models
18.6. Summary
Practice problems
Assignment problems

19. Computer-intensive methods
19.1 Hypothesis testing using simulation
Example 19.1: How did he know? The nonrandomness of haphazard choice
19.2 Randomization test
Example 19.2: Girls just wanna have genetic diversity
Assumptions of randomization tests
19.3 Bootstrap standard errors and confidence intervals
Example 19.3: The language center in chimps' brains
Bootstrap standard error
Confidence intervals by bootstrapping
Bootstrapping data sets with multiple samples
Assumptions and limitations of the bootstrap
19.4 Summary
Practice problems
Assignment problems

20. Likelihood
20.1 What is likelihood?
20.2 Two uses of likelihood in biology
Phylogeny estimation
Gene mapping
20.3 Maximum likelihood estimation
Example 20.3: Unruly passengers
Probability model
The likelihood formula

The maximum likelihood estimate
Likelihood-based confidence intervals

20.4 Versatility of maximum likelihood estimation

Example 20.4: Conservation scoop

Probability model
The likelihood formula
The maximum likelihood estimate
Bias

20.5 Log-likelihood ratio test

Likelihood ratio test statistic
Testing a population proportion

20.6 Summary

20.7 Quick Formula Summary

Practice problems

Assignment problems

21. Meta-analysis: combining information from multiple studies

21.1 What is meta-analysis?

Why repeat a study?

21.2 The power of meta-analysis

Example 21.2: Aspirin and myocardial infarction

21.3 Meta-analysis can give a balanced view

Example 21.3: The Transylvania effect.

21.4 The steps of a meta-analysis

Define the question

Example 21.4: Testosterone and aggression

Review the literature
Compute effect sizes
Determine the average effect size
Calculate confidence intervals and make hypothesis tests
Look for effects of study quality
Look for associations

21.5 File-drawer problem

21.6 How to make your paper accessible to meta-analysis

21.7 Summary

21.8 Quick Formula Summary

Practice problems

Assignment problems

Answers to practice problems

Literature cited

Statistical Tables

Using statistical tables

Statistical Table A: The χ^2 distribution

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

Statistical Table C: The Student *t*-distribution

Statistical Table D: The *F*-distribution

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

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

Statistical Table G: Critical values for the Spearman's correlation coefficient

Photo credits

Index

Preface

Modern biologists need the powerful tools of data analysis. As a result, an increasing number of universities offer, or even require, a basic data analysis course for all their biology students. We have been teaching such a course at the University of British Columbia for the last two decades. Over this period, we have sought a textbook that covered the material we needed in a first course at just the right level. We found that most texts were too technical and encyclopedic, or else they didn't go far enough, missing methods that were crucial to the practice of modern biology. We wanted a book that had a strong emphasis on intuitive understanding to convey meaning, rather than an over-reliance on formulas. We wanted to teach by example, and the examples needed to be interesting. Most importantly, we needed a biology book, addressing topics important to biologists handling real data.

We couldn't find the book that we needed, so we decided to write this one to fill the gap. We include several unusual features that we have discovered to be helpful for effectively reaching our audience:

Interesting biology examples. Our teaching has shown us that biology students learn data analysis best in the context of interesting examples drawn from the medical and biological literature. Statistics is a means to an end, a tool to learn about nature. By emphasizing what we can learn about biology, the power and value of statistics becomes plain. Plus, it's just more fun for everyone concerned.

Every chapter has several biological examples of key concepts, and each example is prefaced by a substantial description of the biological setting. The examples are illustrated with photos of the real organisms, so that students can look at what they're learning about. The emphasis on real and interesting examples carries into the problem sets; for each chapter, there are dozens of questions based on real data about biological issues.

Intuitive explanations of key concepts. Statistical reasoning requires a lot of new ways of thinking. Students can get lost in the barrage of new jargon and multitudinous tests. We have found that starting from an intuitive foundation, away from all the details, is extremely valuable. We take an intuitive approach to basic questions: What's a good sample? What's a confidence interval? Why do an experiment? The first several chapters establish this basic knowledge, and the rest of the book builds on it.

Practical data analysis. As its title suggests, this book focuses on data rather than the mathematical foundations of statistics. We teach how to make good graphical displays, and we emphasize that a good graph is the beginning point of any good data analysis. We give equal time to estimation and hypothesis testing, and we avoid treating the *P*-value as an end in itself. The book does not demand a knowledge of mathematics beyond simple algebra. We focus on practicality over nuance, on biological usefulness over theoretical hand-wringing. We teach not only the "right" way of doing something, but also highlight some of the pitfalls that might be encountered.

We know that a computer will be available for most calculations, so we focus on the concepts of biological data analysis and how statistics can help extract scientific insight from data. With the power of modern computers at hand, the challenge in analyzing data becomes knowing what method to use and why.[1] We imagine and hope that every course using this book will have a component encouraging students to use computer statistical packages. We are also aware that the diversity of such packages is immense, and so we have not tied the book to any particular program.

Practical experimental design. A biologist cannot do good statistics—or good science—without a practical understanding of experimental design. Unlike most books, we discuss basic topics in experimental design, such as controls, randomization, pseudoreplication, and blocking, and we do it in a practical, intuitive way.

Up-to-date on the basics. Believe it or not, the best confidence interval for the proportion is not the one you probably learned as an undergraduate. Nonparametric statistics do not effectively test for differences in means (or medians, for that matter) without some fairly strong assumptions that we normally hear little about. With these and many other topics, we have brought the coverage of basic, everyday topics in statistics up to date.

Coverage of modern topics. Modern biology uses a larger toolkit than a generation ago. In this book, we go beyond most introductory books by establishing the conceptual principles of important topics, such as likelihood, nonlinear regression, randomization, meta-analysis, and the bootstrap.

Useful summaries. Near the end of each chapter is a short, clear summary of the key concepts, and most chapters end with Quick Formula Summaries that put most equations in one easy-to-find place.

Interleaves. Between chapters are short essays that we call interleaves. These interleaves cover a variety of conceptual and common-sense topics that are crucial for the interpretation of statistical results in scientific research. Several of them focus on

---

1 "A computer lets you make more mistakes faster than any invention in human history—with the possible exceptions of handguns and tequila." —Mitch Ratcliffe, in *Technology Review,* 1992

ways that science can go wrong when concepts are misapplied—and how to account for such mistakes. Although the interleaves are set outside the boundaries of the chapters, they complement the material in the core chapters, and we strongly recommend that they not be skipped.

After five years of writing, you hold the result in your hands. We think *The Analysis of Biological Data* provides a good background in data analysis for biologists, covering a broad range of topics in a practical and intuitive way. It works for our classes; we hope that it works for yours, too.

Organization of the book

The Analysis of Biological Data is divided into five blocks, each with a handful of chapters. We recommend starting with the first block, because it introduces many basic concepts that are used throughout the book. These early chapters are meant to be read in their entirety.

After the first block, most chapters progress from the most general topics at the start to more specialized topics by the end. Each chapter is structured so that a basic understanding of the topic may be obtained from the earliest sections. For example, in the chapter on analysis of variance (Chapter 15), the basics are taught in the first two sections; reading Sections 15.1 and 15.2 gives roughly the same material that most introductory statistics texts provide about this method. Sections 15.3–15.6 explain additional twists and other interesting applications.

The last block of chapters (Chapters 18–21) is mainly for the adventurous and the curious. These chapters introduce several topics, such as likelihood, bootstrapping, and meta-analysis, that are commonly encountered in the biological and medical literature but that are not often mentioned in an introductory course. These chapters introduce the basic principles of each topic, how the methods work, and point to where you might look to find out more.

A basic course could be taught by using only Chapters 1–17 and, within this subset of chapters, by stopping after Sections 5.6, 7.3, 8.4, 9.3, 12.6, 13.6, 15.2, 16.4, and 17.5 in their respective chapters. We suggest that all courses highlight the topics covered in the interleaves.

Each chapter ends with a series of problems that are designed to test students' understanding of the concepts and the practical application of statistics. The problems are divided into Practice Problems and Assignment Problems. Short answers to all Practice Problems are provided in the back of the book; answers to the Assignment Problems are available to instructors only from the publisher. For a copy, contact Ben Roberts at bwr@roberts-publishers.com or (303) 221-3325. Other teaching resources for the book are available online at http://www.roberts-publishers.com/whitlock/teaching.

A word about the data

The data used in this book are real, with a few well-marked exceptions. For the most part, these data were obtained directly from published papers. In some cases, we contacted the authors of articles who generously provided the raw data for our use. Often, when raw data were not provided in the original paper, we resorted to obtaining data points by digitizing graphical depictions, such as scatter plots and histograms. Inevitably, the numbers we extracted differ slightly from the original numbers because of measurement error. In rare cases, we generated data by computer that matched the statistical summaries in the paper. In all cases, the results we present are consistent with the conclusions of the original papers.

Acknowledgments

This book would not have been possible for the two of us alone. Many other people contributed to it in substantial ways. The clarity and accuracy of its contents was improved by the careful attention of a lot of generous readers, including Arianne Albert, Brad Anholt, Cecile Ane, Eric Baack, James Bryant, Martin Buntinas, C. Ray Chandler, Christiana Drake, Jonathan Dushoff, Steven George, Aleeza Gerstein, George Gilchrist, Brett Goodwin, Mike Hickerson, Darren Irwin, Nusrat Jahan, Philip Johns, Istvan Karsai, Robert Keen, John Kelly, Rex Kenner, Ben Kerr, Joseph G. Kunkel, Todd Livdahl, Brian C. McCarthy, Eli Minkoff, Robert Montgomerie, Spencer Muse, Courtney Murren, Claudia Neuhauser, Patrick C. Phillips, Jay Pitocchelli, James Robinson, Simon Robson, Michael Rosenberg, Noah Rosenberg, Nathan Rank, Bruce Rannala, Mark Rizzardi, Michael Russell, Ronald W. Russell, Andrew Schaffner, Andrea Schluter, William Thomas, Michael Travisano, Thomas Valone, Bruce Walsh, Grace A. Wyngaard, and Sam Yeaman. Many of these people read multiple chapters, and they all improved the clarity and accuracy of the book. Sally Otto and Allan Stewart-Oaten earned our undying gratitude by reading and commenting on the entire book. Of course, any errors that remain are our own fault; we didn't always take everyone's advice, even perhaps when we should have. If we have forgotten anyone, you have our thanks even if our memories are poor.

We owe a debt to the students of BIOL 300 at the University of British Columbia, who class-tested this book over the last several years. The book also benefited by class testing at several colleges and universities, in courses by Brad Anholt (University of Victoria), Eric Baack (Luther College), George Gilchrist (College of William and Mary), Mike Hickerson (Queens College, City University of New York), Nusrat Jahan (James Madison University), Susan Lehman (Brock University), Jean Richardson (Brock University), Simon Robson (James Cook University), and Grace A. Wyngaard (James Madison University). George Gilchrist and his students gave us a very detailed and extremely helpful set of comments at a crucial stage of the book. The following students from UBC and other institutions uncovered significant errors in draft versions of the book: Jessica Beaubier, Edward Cheung, Lorena Cheung, Stephanie Cheung, Denise Choi, Samrad Ghavimi, Inderjit Grewal, Sarah Hamanishi, Gurpreet Khaira, Jung Min Kim, Arleigh Lambert, Alexander Leung, Mira Li, Flora Liu, Dianna Louie, Johnston Mak, Sarah Neumann, Ruth Ogbamichael, Marion Pearson, Trevor Schofield, Meredith Soon, Erin Stacey, Michelle Uzelac, Hillary Ward, Chris Wong, Irene Yu, Anush Zakaryan, Paul Zhou, and Jon-Paul Zacharias.

A number of researchers kindly sent us their original data, including Matt Arnegard, Audrey Barker-Plotkin, Butch Brodie, Pamela Colosimo, Kevin Fowler, Chris Harley, Luke Harmon, Andrew Hendry, Peter Keightley, Fredrik Liljeros, Jean Thierry-Mieg, Jeffrey S. Mogil, Patrik Nosil, Margarita Ramos, Rick Relyea, Jake Socha, Brian Starzomski, Richard Svanback, Andrew Trites, Jason Weir, Jack Werren, and Martin Wikelski.

The book was edited and copyedited by editor extraordinaire John Murdzek: his humor may be warped, but his editorial direction is straight as can be. John worked overtime getting this book out on a tight schedule, even finding the typo in "*Photinus ignites.*" Tom Webster turned our graphs and illustrations into beautiful art, and Laura Roberts spent many an hour tracking down the wonderful photos, and Jeff Whitlock very generously provided many of the beautiful photos in this book. Mark Ong and his team at Side by Side Studios did a fantastic job of turning the manuscript into a book. Eric Baack has our special appreciation for slaving over the problem sets to create the answer keys. Gunder Hefta and Aleeza Gerstein corected numerous errrors with their careful proofreading. Finally, Ben Roberts deserves our greatest thanks, for all of his support and vision in making this book happen, and especially for Clause 24.

The book was started while MCW was supported by the Peter Wall Institute for Advanced Studies at UBC as a Distinguished-Scholar-in-Residence, and the majority of the final stages of the book were written while he was a Sabbatical Scholar at the National Evolutionary Synthesis Center in North Carolina (NSF #EF-0423641). DS began working on the book while a visiting professor in Developmental Biology at Stanford University. The scholarly support and environment provided by each of these institutions was exceptional—and greatly appreciated.

Finally, we would like to give great thanks to all of the people that have taught us the most over the years. MCW would like to thank Dave McCauley, Mike Wade, Nick Barton, Ben Pierce, Kevin Fowler, Patrick Phillips, Sally Otto, and Betty Whitlock. DS would like to thank Trevor Price, Don Ludwig, Andrea Schluter, Peter and Rosemary Grant, Jamie Smith, David Kingsley, Tom Schoener and Ron Brooks. We dedicate this book to these, our teachers.

About the authors

Michael Whitlock is an evolutionary biologist and population geneticist. He is a Professor of Zoology at the University of British Columbia, where he has taught statistics to biology students since 1995. He is currently Editor-in-Chief of *The American Naturalist.*

Dolph Schluter is Professor and Canada Research Chair in the Zoology Department and Biodiversity Research Center at the University of British Columbia. He is known for his research on the ecology and evolution of Galapagos finches and threespine stickleback. He is a fellow of the Royal Societies of Canada and London.