

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.

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.

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>.

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