1   Introduction

Biology is becoming increasingly computational. New technologies are producing massive amounts of data, particularly in molecular biology, and are opening entirely new avenues of research. However, with these new technologies come new challenges and needs. Large amounts of quantitative information need to be organized, displayed, and understood. Even for biologists who are not directly involved in collecting data, thinking quantitatively and interpreting quantitative results, both in research and in the literature, are becoming important aspects of daily scientific training and work.

Large-scale sequencing projects are good examples of how science, enabled by technology, has given rise to massive amounts of data (e.g., genome and protein sequences) that need to be summarized, investigated, and quantitatively assembled. Case in point: The simplest software program used for assessing sequencing data is BLAST (Basic Local Alignment Search Tool), in which every search conducted by the program returns statistical summaries that describe how well search results "fit" a query. Understanding what these results are actually telling us is essential to making well-informed scientific choices.

Having an appreciation for the computer results provided by software applications or quantitative results from the literature requires a basic understanding of both statistical thinking and statistical procedures. Likewise, quantitative reasoning as applied to the design of an experiment and data evaluation are as important as formulating and answering biological questions based on challenging and complex experimental situations. With this in mind, this Manual is intended as a resource that provides an overview of terminology and concepts that occur frequently in quantitative thinking and in the statistical analysis of biological data. This Manual also explains in the simplest of terms the underlying principles of the more complex analyses (e.g., microarray experiments and RNA-Seq).

If you are looking for a one- or two-semester textbook on statistical procedures, you have the wrong book in hand. This Manual is not intended as a textbook—It is a "bench-side" Manual that is designed and intended to be used by people who are in need of a quick refresher or a big picture overview of a statistical procedure. Maybe it has been a long time (if ever) since you last took a course in mathematics or statistics, or even considered thinking quantitatively about your data. Maybe the amount of quantitative training you received is lacking or less than desirable. This Manual is aimed at you! The comprehensive index allows you to quickly access the information you need to think about your data and interpret your results effectively.

As you will see and read in this Manual, new concepts are illustrated by simple examples selected from different biological applications. Many of the examples that we use are accompanied by detailed R Commander instructions and are the same examples that are used in the earlier Excel version of this book (Bremer and Doerge 2009). We chose R because it is a powerful and versatile software that is freely available and widely used by biologists and statisticians today. We chose R Commander because this easy-to-understand graphical user interface provides easy and intuitive access to R. For applications that exceed the scope of this text, R programs and packages usually exist, but they may not be available via R Commander. Of course, there are many other statistics programs that can also be used instead of R (for a discussion of available programs, see Section 2.7).

In R Commander: Complete R Commander examples can be found in boxes throughout the text. For complete instructions on how to install R and R Commander, refer to Section 2.8.

Today, computationally trained biologists are gaining a competitive edge in science. Not only can they read the literature, they can also think critically about it and engage in meaningful scientific conversations rich with quantitative reasoning. Our purpose in providing this Manual is to assist biologists in becoming fluent and comfortable in the language of quantitative reasoning and to facilitate open and informed communication between the biological and the quantitative sciences.