

# Preface

This book owes its existence to the lecture “Statistics for Systems Biology”, which I taught in the fall semester 2010 at the Department for Biosystems Science and Engineering of the Swiss Federal Institute of Technology (ETH Zurich). To a large part, the audience consisted of students with backgrounds in biological sciences, which explains the large proportion of biological examples in this text.

Nevertheless, I hope that this text will be helpful for readers with very different backgrounds who need to quantify and analyze data to answer interesting questions. This book is not intended to be a manual, nor can it provide the answer to all questions and problems that one will encounter when analyzing data. Both the book title and the title of the book series indicate that space is limited and this book therefore concentrates more on the ideas and concepts rather than on presenting a vast array of different methods and applications. While all the standard material for an introductory course is covered, this text is very much inspired by Larry Wasserman’s excellent book *All of Statistics* [1] and consequently discusses several topics usually not found in introductory texts, such as the bootstrap, robust estimators, and multiple testing, which are all found in modern statistics software. Due to the space constraints, this book does not cover methods from Bayesian statistics and does not provide any exercises. Frequent reference is made to the software R (freely available from <http://www.r-project.org>), but the text itself is largely independent from a particular software.

Should this book provide the reader with enough understanding of the fundamental concepts of statistics and thereby enable her or him to avoid some pitfalls in the analysis of data and interpretation of the results, such as by providing proper confidence intervals, not “accepting” a null hypothesis, or correcting for multiple testing where it is due, I shall be contented.

The book is organized in four chapters: **Chapter 1** introduces the basics of probability theory, which allows to describe non-deterministic processes and is thus essential for statistics. **Chapter 2** covers the inference of parameters and properties from given data, and introduces various types of estimators, their properties, and the computation of confidence intervals to quantify how good a given estimate is. Robust alternatives to important estimators are also provided.

[Chapter 3](#) is devoted to hypothesis testing, with a main focus on the fundamental ideas and the interpretation of results. This chapter also contains sections on robust methods and correction for multiple testing, which become more and more important, especially in biology. Finally, [Chap. 4](#) presents linear regression with one and several covariates and one-way analysis-of-variance. This chapter uses R more intensively to avoid tedious manual calculations, which the reader hopefully appreciates.

There surely is no shortage in statistics books. For further reading, I suggest to have a look at the two books by Wasserman: *All of Statistics* [1] and *All of Nonparametric Statistics* [2], which contain a much broader range of topics. The two books by Lehmann, *Theory of Point Estimation* [3] and *Testing Statistical Hypotheses* [4] contain almost everything one ever wanted to know about the material in [Chaps. 2](#) and [3](#). For statistics using R, *Statistics—An Introduction using R* [5] by Crawley and *Introductory Statistics with R* [6] by Dalgaard are good choices, and *The R Book* [7] by Crawley offers a monumental reference. The *Tiny R Handbook* [8], published in the same series by Springer, might be a good companion to this book. For statistics related to bioinformatics, *Statistical Methods in Bioinformatics* [9] by Ewens and Grant provides lots of relevant information; the DNA sequence example is partly adapted from that book. Finally, for the german-speaking audience, I would recommend the two books by Pruscha *Statistisches Methodenbuch* [10], focusing on practical methods, and *Vorlesungen über mathematische Statistik* [11], its theory counterpart.

This script was typeset in LATEX, with all except the first two figures and all numerical data directly generated in R and included using Sweave [12].

I am indebted to many people that allowed this book to enter existence: I thank Jörg Stelling for his constant encouragement and support and for enabling me to work on this book. Elmar Hulliger, Ellis Whitehead, Markus Beat Dürr, Fabian Rudolf, and Robert Gnügge helped correcting various errors and provided many helpful suggestions. I thank my fiancée Elke Schlechter for her love and support. Financial support by the EU FP7 project UNICELLSYS is gratefully acknowledged. For all errors and flaws still lurking in the text, the figures, and the examples, I will nevertheless need to take full responsibility.

Basel, July 2011

Hans-Michael Kaltenbach

## References

1. Wasserman, L.: *All of Statistics*. Springer, Heidelberg (2004)
2. Wasserman, L.: *All of Nonparametric Statistics*. Springer, Heidelberg (2006)
3. Lehmann, E.L., Casella, G.: *Theory of Point Estimation*, 2nd edn. Springer, Heidelberg (1998)
4. Lehmann, E.L., Romano, J.P.: *Testing Statistical Hypotheses*, 3rd edn. Springer, Heidelberg (2005)
5. Crawley, M.J.: *Statistics—An Introduction using R*. Wiley, New York (2005)

6. Dalgaard, R.: *Introductory Statistics with R*, 2nd edn. Springer, Heidelberg (2008)
7. Crawley, M.J.: *The R Book*. Wiley, New York (2007)
8. Allerhand, M.: *A Tiny Handbook of R*. Springer, Heidelberg (2011)
9. Ewens, W.J., Grant, G.R.: *Statistical Methods in Bioinformatics*. Springer, Heidelberg (2001)
10. Pruscha, H.: *Statistisches Methodenbuch*. Springer, Heidelberg (2006)
11. Pruscha, H.: *Vorlesungen über Mathematische Statistik*. Springer, Heidelberg (2000)
12. Leisch, F.: *Sweave: Dynamic generation of statistical reports*. In: Härdle, W., Rönz, B. (eds.) *Compstat 2002—Proceedings in Computational Statistics*, pp 575–580 (2002)

A Concise Guide to Statistics

Kaltenbach, H.-M.

2012, XIII, 111 p. 33 illus., Softcover

ISBN: 978-3-642-23501-6