Biostatistics *for Medical and Biomedical Practitioners*

Julien I.E. Hoffman

Statistics

BIOSTATISTICS FOR MEDICAL AND BIOMEDICAL PRACTITIONERS

JULIEN I. E. HOFFMAN
Tiburon, California, USA

ELSEVIER

Amsterdam • Boston • ~~Heidelberg~~ • London • New York • Oxford
Paris • San Diego • San Francisco • Singapore • Sydney • Tokyo
Academic Press is an imprint of Elsevier

Academic Press is an imprint of Elsevier
125 London Wall, London EC2Y 5AS, UK
525 B Street, Suite 1800, San Diego, CA 92101-4495, USA
225 Wyman Street, Waltham, MA 02451, USA
The Boulevard, Langford Lane, Kidlington, Oxford OX5 1GB, UK

ISBN: 978-0-12-802387-7

British Library Cataloguing-in-Publication Data
A catalogue record for this book is available from the British Library

Library of Congress Cataloging-in-Publication Data
A catalog record for this book is available from the Library of Congress

For information on all Academic Press publications
visit our website at http://store.elsevier.com/

Publisher: Mica Haley
Acquisition Editor: Rafael Teixeira
Editorial Project Manager: Mariana Kuhl
Production Project Manager: Julia Haynes
Designer: Victoria Pearson

Typeset by TNQ Books and Journals
www.tnq.co.in

Transferred to Digital Printing 2015

BIOSTATISTICS FOR MEDICAL AND BIOMEDICAL PRACTITIONERS

ABOUT THE AUTHOR

Julien I. E. Hoffman was born and educated in Salisbury (now named Harare) in Southern Rhodesia (now named Zimbabwe) in 1925. He entered the Faculty of Medicine at the University of the Witwatersrand in Johannesburg, South Africa, and obtained a BSc Hons in Anatomy and Physiology in 1945 and a medical degree (MB, BCh) in 1949. After working for almost 3 years at the Central Middlesex Hospital in London, England, as a house officer and registrar in Medicine, he returned to Johannesburg as a senior registrar in the Department of Medicine for 18 months, and then returned to England to work for the Medical Research Council. In 1957, he went to Boston Children's Hospital to study congenital heart disease, and this was followed by 15 months as a Fellow in the Cardiovascular Research Institute at the University of California in San Francisco.

In 1962, he joined the faculty at the Albert Einstein College of Medicine in New York City as an Assistant Professor of Pediatrics and Internal Medicine, and in 1966, returned to the University of California in San Francisco as Associate Professor of Pediatrics and member of the Cardiovascular Research Institute. He was a clinical pediatric cardiologist, taking care of children with heart disease, but spent about 50% time running a research laboratory to study the pathophysiology of the coronary circulation.

His interest in statistics began while taking his Science degree. After going to England, he took short courses in Biostatistics from Bradford Hill and his colleagues. On returning to South Africa, as the only department member to know anything about statistics, he was assigned to perform statistical analyses for other members of the department. This was a period of learning by trial and error, helped by Dr J Kerrich, head of the University's Division of Statistics.

When he went to San Francisco as a Fellow, he was assigned to give statistics lectures to the Fellows and Residents, and when returned to San Francisco in 1966, he gave an officially sanctioned course in Biostatistics to research Fellows and Residents. These lectures were given annually for over 30 years. He was a member of the Biostatistics group, a semiformal group that supervised statistical teaching and consultation. He also was a statistical consultant to the journal *Circulation Research*, and was often assigned manuscripts by other journals, mostly from the *American Heart Association*, to check on the statistical procedures used.

PREFACE

In my 40 years experience of advising medical research workers about how to analyze their studies, certain problems arose frequently. For examples, many investigators wanted to compare two Poisson distributions, yet some introductory books on Biostatistics give little attention to the Poisson distribution, even though it is an important distribution that answers the question about how often a rare event occurs, such as the number of deliveries per hour in a delivery room. Few people can navigate the minefield of multiple comparisons, involved when several different groups are compared, often done incorrectly by performing multiple *t*-tests, yet most elementary texts do not deal with this problem adequately. Problems of repeated measures analysis in which several measurements are made in each member of the group and thus are not independent occur frequently in medical research but are not often discussed. Tolerance tests are often needed to set ranges of normal values so that a single measurement can be assessed as likely to be normal or abnormal (such as a single fasting blood glucose concentration). Because most basic books do not discuss this problem, most people incorrectly set confidence limits that should apply only to mean values. In fact, one of the incentives for this book was the lack of books in introductory Biostatistics that could be understood relatively easily, but nevertheless were advanced enough that most investigators would not need to buy additional books or hunt through unfamiliar journals for appropriate tests.

The book is intended to help physicians and biologists who might have had a short course on Statistics several years ago, but have forgotten all but a few of the terms and concepts and have not used their knowledge of statistics for reading the literature critically or for designing experiments. The general aim is to extend their knowledge of statistics, to indicate when various tests are applicable, what their requirements are, and what can happen when they are used inappropriately.

This book has four components.

1. It covers the standard statistical approaches for making descriptions and inferences—for example, mean and standard deviation, confidence limits, hypothesis testing, *t*-tests, chi-square tests, binomial, Poisson, and normal distributions, analysis of variance, linear regression and correlation, logistic regression, and life tables—to help readers understand how the tests are constructed, how to look for and avoid using inappropriate tests, and how to interpret the results. Examples of injudicious use of these tests are given. Although some basic formulas are presented, these are not essential for understanding what the tests do and how they should be interpreted.
2. Some chapters include a section on advanced methods that should be ignored on a first reading but provide information when needed, and others have an appendix where

some simple algebraic proofs are given. As this is not intended to be a mathematically rigorous book, most mathematical proofs are omitted, but a few are important teaching tools in their own right and should be studied. However, knowledge of mathematics (and differential calculus in particular) beyond elementary algebra is not required to use the material provided in this book. The equations indicate what the tests are doing.

3. Scattered throughout the chapters are variations on tests that are often needed but not frequently found in basic texts. These sections are often labeled "Alternative Methods," and they should be read and understood because they often provide simpler and more effective ways of approaching statistical inference. These include:
   a. Robust statistics for dealing with grossly abnormal distributions, both univariate and bivariate.
   b. Extending McNemar's test, a test for comparing paired counts, to more than two categories; for example, if matched pairs of patients are given one of two treatments and the results recorded as improved, the same, or worse, how should these be analyzed?
   c. Equivalence or noninferiority testing, to determine if a new drug or vaccine is equivalent to or not inferior to those in standard use.
   d. Finding the break point between two regression lines. For example, if the lactate:pyruvate ratio remains unchanged when systemic oxygen delivery is reduced below normal until some critical point is reached when the ratio starts to rise, how do we determine the critical oxygen delivery value?
   e. Competing risks analysis used when following the survival of a group of patients after some treatment, say replacement of the mitral valve, and allowing for deaths from noncardiac causes.
   f. Tolerance testing to determine if a single new measurement is compatible with a normative group.
   g. Crossover tests, in which for a group of subjects each person receives two treatments, thus acting as his or her own control.
   h. Use of weighted kappa statistics for evaluating how much two observers agree on a diagnosis.

   Some of these analyses can be found only in journals or advanced texts, and collecting them here may save investigators from having to search in unfamiliar sources to find them.

4. Some chapters describe more complex inferences and their associated tests. The average investigator is not likely to use any of these tests without consultation with a statistician, but does need to know that these techniques exist and, if even vaguely, what to look for and how to interpret the results of these tests when they appear in publications. These subjects include:
   a. Poisson regression (Chapter 34), in which a predicted count, for example, the number of carious teeth, is determined by how many subjects have zero, 1, 2, etc., carious teeth.

b. Resampling methods (Chapter 37), in which computer-intensive calculations allow the determination of the distributions and confidence limits for mean, median, standard deviations, correlation coefficients, and many other parameters without needing to assume a particular distribution.

c. The negative binomial distribution (Chapter 19) that allows investigation of distributions that are not random but in which the data are aggregated. If we took samples of seawater and counted the plankton in each sample, a random distribution of plankton would allow us to fit a standard distribution such as a binomial or Poisson distribution. If, however, some samples had excessive numbers of plankton and others had very few, a negative binomial distribution may be the way to evaluate the distribution.

d. Meta-analysis (Chapter 36), in which the results of several small studies are aggregated to provide a larger sample, for example, combining several small studies of the effects of beta-adrenergic blockers on the incidence of a second myocardial infarction, is often used. The pitfalls of doing such an analysis are seldom made clear in basic statistics texts.

e. Every investigator should be aware of multiple and nonlinear regression techniques (Chapter 30) because they may be important in planning experiments. They are also used frequently in publications, but usually without mentioning their drawbacks.

With the general availability of personal computers and statistical software, it is no longer necessary to detail computations that should be done by computer programs. There are many simple free online programs that calculate most of the commonly used statistical descriptions (mean, median, standard deviation, skewness, interquartile distance, slope, correlation, etc.) as well as commonly used inferential tests (*t*-test, chi-square, ANOVA, Poisson probabilities, binomial probabilities, life tables, etc.), along with their associated graphics. More complex tests require commercial programs. There are free online programs for almost all the tests described in this book, and hyperlinks are provided for these.

Problems are given in appropriate chapters. They are placed after a procedure is described so that the reader can immediately practice what has been studied to make sure that the message is understood; the procedures are those that should be able to be performed by the average reader without statistical consultation. Although the problems are simple and could be done by hand, it is better to use one of the recommended online calculators because they save time and do not make arithmetic errors. This frees up time for the reader to consider what the results mean.

The simpler arithmetic techniques, however, are still described in this book because they lead to better understanding of statistical methods, and show the reader where various components of the calculation come from, and how the components are used and interpreted. In place of tedious instructions for doing the more complex arithmetic procedures, there is a greater concentration on the prerequisites for doing each test and for interpreting the results. It is easier than ever for the student to think about what the statistical tests are doing and how they contribute to solving the problem. On the other hand, we need to resist the temptation to give a cookbook approach to solving problems without giving some understanding of their bases, even though this may involve some elementary algebra. As Good and Hardin (2009) wrote: "Don't be too quick to turn on the computer. Bypassing the brain to compute by reflex is a sure recipe for disaster."

Many people who learn statistics devise and carry out their own experiments, and for them a good knowledge of statistics is essential. There are many excellent statistical consultants, but not enough of them to advise every investigator who is designing an experiment. An investigator should be able to develop efficient experiments in most fields, and should reserve consultation only for the more complex of these. But more numerous and important are those who do not intend to do their own research. Even if the person is not a research worker, he or she is still responsible for assessing the merit of the articles that they are reading to gain new knowledge. It is no longer good enough for such a reader to know only technical information about the drugs used, the techniques for measuring pressure and flow, or to understand the physiologic and biochemical changes that take place in the disease in question. It is as essential for the reader to learn to read critically, and to know how to evaluate the statistical tests used. Who has not been impressed by the argument that because Fisher's exact test showed a probability of 0.003, the two groups were different, or that because the probability was 0.08 the two groups were not different? Who has heard of Fisher's exact test? Of McNemar's test? Of the Kolmogorov–Smirnov two group test? Of Poisson regression? And if the reader has not heard of them, how can he or she know if they should have been used, or have been used and interpreted correctly? With the pace at which new knowledge is appearing and being incorporated into medical practice, everyone engaged in medical research or practice needs to be well grounded in Statistics. Statistical thinking is not an addendum to the scientific method, but is an integral component of it.

What can you expect to achieve after studying this book?

1. A better appreciation of the role of variation in determining average values, and how to take account of this variation in assessing the importance of measured values or the difference between these values in two or more groups.
2. A better understanding of the role of chance in producing unexpected results, and how to make decisions about what the next steps should be.
3. An appreciation of Bayes' theorem, that is, how to improve estimates of probability by adding in potential explanatory variables, and the importance of the population prevalence in making clinical predictions.
4. An ability to perform simple statistical tests (e.g., *t*-tests, chi-square, linear regression and correlation, McNemar's test, simple analysis of variance, calculate odds ratios and their confidence limits, and calculate simple life tables), as well as understanding their limitations.

5. Appreciate that there are many statistical techniques that can be used to address specific problems, such as meta-analysis, bioassay methods, nonlinear and multiple regression, negative binomial distributions, Poisson regression, and time series analysis. It is unlikely that after studying this book you will be able to perform these analyses on your own, but you should know that such methods exist and that a statistical consultant can help you to choose the right analytic method.

Statistical procedures are technologies to help investigators interpret their data, and are not ends in themselves. Like most technologies, Statistics is a good servant but a bad master. Francis Galton (1889) wrote: "Some people hate the very name of statistics, but I find them full of beauty and interest. Whenever they are not brutalized but delicately handled by the higher methods, and are warily interpreted, their power of dealing with complicated phenomena is extraordinary."

REFERENCES

Galton, F.I., 1889. Natural Inheritance. Macmillan and Company, London, p. 62.

Good, P.I., Hardin, J.W., 2009. Common Errors in Statistics (and How to Avoid Them). John Wiley & Sons, Hoboken, NJ, p. 274.

ACKNOWLEDGMENTS

Many people have helped me eradicate errors and clumsy phrasing. I owe a debt of gratitude to Joseph P. Archie, Jr, Raul Domenech, Stanton Glantz, Gus Vlahakes, and Calvin Zippin. My special thanks to my wife for help with editing.

I also wish to thank the staff of Elsevier for their hard work, with particular thanks to Rafael Teixeira, Mariana Kühl Leme, and Julia Haynes.

CONTENTS