Preface

Our first edition, *Introduction to Biostatistics: A Guide to Design, Analysis, and Discovery*, was published in 1995 and was well received both by reviewers and readers. That book broke new ground by expanding the range of methods covered beyond what typically was included in competing texts. It also emphasized the importance of understanding the context of a problem — the why and what — instead of considering only the how of analysis.

Although the past several years have seen much interest in a second edition, our involvement with numerous other projects prevented us from tackling a new edition. Now that the stars are in alignment (or whatever), we finally decided to create a second edition. We are excited that Mike Hernandez has agreed to collaborate with us on this edition.

This new edition builds on the strengths of the first effort while including several new topics reflecting changes in the practice of biostatistics. Although parts of the second edition still serve as an introduction to the world of biostatistics, other parts break new ground compared to competing texts. For some of these relatively more advanced topics, we strongly advise the reader to consult with experts in the field before setting out on the analyses.

This revised and expanded edition continues to encourage readers to consider the full context of the problem being examined. This context includes understanding what the goal of the study is, what the data actually represent, why and how the data were collected, how to choose appropriate analytic methods, whether or not one can generalize from the sample to the target population, and what problems occur when the data are incomplete due to people refusing to participate in the study or due to the researcher failing to obtain all the relevant data from some sample subjects. Although many biostatistics textbooks do a very good job in presenting statistical tests and estimators, they are limited in their presentations of the context. In addition, most textbooks do not emphasize the relevance of biostatistics to people's lives and well-being. We have written and revised this textbook to address these deficiencies and to provide a good guide to statistical methods.

This textbook also differs from some of the other texts in that it uses real data for most of the exercises and examples. For example, instead of using data resulting from tossing dice or dealing cards, real data on the relation between prenatal care and birth weight are used in the definition of probability and in the demonstration of the rules of

probability. We then show how these rules are applied to epidemiologic measures and the life table, major tools used by health analysts. Other major differences between this and other texts are found in Chapters 11, 14, and 15. In Chapter 11 we deal with the analysis of the follow-up life table; its use in survival analysis is considered in Chapter 14. In Chapter 15 we present strategies for analyzing survey data from complex sample designs. Survey data are used widely in public health and health services research, but most biostatistics texts do not deal with sample weights or methods for estimating the sample variance from complex surveys.

We also include material on tolerance and prediction intervals, topics generally ignored in other texts. We demonstrate in which situations these intervals should be used and how they provide different information than that provided by confidence intervals. In addition, we discuss the randomized response technique and the general linear model for analysis of data sets with an unequal number of observations in each cell, topics generally not covered in other texts. The randomized response technique is one way of dealing with response bias associated with sensitive questions, and it also illustrates the importance of statistical design in the data collection process.

Although we did not write this book with the assumption that readers have prior knowledge of statistical methods, we did assume that readers are not the type to be rendered unconscious by the sight of a formula. When presenting a formula, we first explain the concept that underlies the formula. We then show how the formula is a translation of the concept into something that can be measured. The emphasis is on when and how to apply the formula, not on its derivation. We also provide a review of some mathematical concepts that are used in our explanations in Appendix A. A website is provided that demonstrates the use of statistical software in carrying out the analyses shown in the text. As new versions of statistical packages become available, the website material will be updated.

The textbook is designed for a two-semester course for the first-year graduate student in health sciences. It is also intended to serve as a guide for the reader to discover and learn statistical concepts and methods more or less by oneself. If used for a one-semester course, possible deletions include the sections on the geometric mean, the Poisson distribution, the distribution-free approach to intervals, the confidence interval and test of hypothesis for the variance and coefficient of correlation, the Kruskal-Wallis and Friedman tests, the trend test for r by 2 contingency tables, the two-way ANOVA, ANOVA for unbalanced designs, the linear model representation of the ANOVA, the ordered and conditional logistic regression, the proportional hazards regression, and the analysis of survey data.

Several appendices are at the end of the book. Appendix A presents some basic mathematical concepts that are essential to understanding the statistical methods presented in this book. Appendix B contains several statistical tables referenced in the text. Appendix C is a listing of major governmental sources of health data, and Appendix D presents solutions to selected exercises.

Acknowledgments

Our special thanks go to Professor Anthony Schork of University of Michigan for his review of the first edition and his suggestions for the contents of the second edition.

There are several other anonymous reviewers of the contents of the second edition who made helpful comments and suggestions. We also wish to acknowledge those who provided with useful suggestions and comments for the first edition including Joel A. Harrison, Mary Forthofer, Herbert Gautschi, Irene Easling, Anna Baron, Mary Grace Kovar, and the students at the University of Texas School of Public Health Regional Program in El Paso, who reviewed parts or all of the first edition. Any problems in the text are the responsibility of the authors and not of the reviewers.