Preface

This textbook is related to a course that I have taught for many years at University of California, Berkeley. The course was intended for graduate students in the biological and health sciences. But it attracted students from other departments on the campus at well. Since the course had a general appeal, I decided to prepare a textbook based on the outline of the course but with some extensions in order to reach a wider audience.

Now the book consists of seventeen chapters in three parts. The first four chapters are devoted to descriptive statistics, probability, random variables, and probability distributions, as an introduction. The second two chapters cover the basic concept of statistical inference: interval estimation and hypothesis testing. The remaining eleven chapters are devoted to specific methods of analysis: from the "Student's" t test, the chi-square test, through regression and correlation, to the analysis of variances and covariances, and to experimental designs and non-parametric statistics. In spite of rapid growth in the field, these topics are still the foundation of statistical analysis.

I learned from my teaching that not everyone in the class likes derivations or proofs of formulas or theorems. Some will not take a formula for granted without a formal proof, others regards proofs as a nuisance. As a compromise I have taken almost all the proofs out of the main body of the text and placed them in a separate section entitled "Proof of the Results

in This Chapter" near the end of each chapter. The reader may skip this section altogether without losing the continuity of the text, or may study this section intensely to increase one's understanding of the material. Some may even want to work out the solutions themselves and check them with those in the book.

Although this is a book of applied statistics, I have tried to provide theoretical justification for the methods of analysis introduced. Whenever needed, I have placed theorems in a separate section just before a method of analysis or a test statistic is introduced in order to make the analysis of the test more meaningful. For example, in Chapter 7 there are theorems regarding the "Students's" t distribution before the t statistic is used for hypothesis testing. In Chapter 8, Sec. 2, there are theorems concerning the chi-square distribution before the chi-square test is introduced in the following sections.

The reader will find three quantities in the discussion of the analysis of variance and analysis of covariance quite helpful in the understanding the material. The first is the linear model, which makes very clear the meaning of the underlying assumptions and the null hypotheses. The second is the expectations of mean squares. The explicit formulas for these expectations lead directly to the formulas of the estimates and the test statistics. And the third is the table of analysis of variance that summarizes the main points in the analysis and helps the reader's understanding. We all should be grateful to those who devised such an elegant way of summarization.

In Chapter 16 on experimental designs, other than description and analysis in each model, the efficiency of blocking in random blocks and in Latin-squares designs is discussed in some detail to justify the blocking.

One can often explain a point better with pictures than words. This is true also in this volume. In preparing the material, I have plotted many graphs to help convey the message to the reader. Many of these graphs serve the purpose well. In addition, I have prepared problems and exercises to accompany the methods of analysis and placed them in the last section of each chapter. These problems and exercises are directly related to the material in the chapter. They are indeed an integral past of the book. Instructors of the course may want to use them as a source of homework assignments.

Some of the illustrative examples and problems are hypothetical while some are real. I also have used simple numbers in real problems for easy computations. I urge the students to work through the examples and problems with a hand calculator rather computer programs. Using a hand

calculator, one can see every step in the calculations, the understanding of the methods of analysis is deep. Using computer programs, one can have the final numerical results almost immediately but will miss the detail of the computations and the understanding of the material.

Finally, I wish to mention that this book can be used as textbook for both an introductory course and an intermediate course, but using different chapters.

(1) For an introductory course: Chapters 1 to 13 and Chapters 15 and 17. Chapters 14 and 16 may be reserved for further study.
(2) For an intermediate course: Chapters 5 to 17. Students may study Chapters 1 to 4 on their own as a review.

I am most grateful to many people who have contributed to the development of statistical methods of analysis. Without their work, this book could not have been written. While it is impossible to acknowledge every contributor individually, I would like to thank J. Neyman and E. S. Pearson, who were among the founders of basic concept of interval estimation and hypothesis testing; and to R. A. Fisher who, more than anyone else, was responsible for developing the theory and methods of experimental designs and analysis of variance.

Many thanks are due to my colleagues, S. Selvin and C. Langhouser, who provided me with examples for illustration, and to B. J. van den Berg for allowing me to use the data from the Child Health and Development Studies. I very much appreciate their generosity. I also want to express my thanks to T. Porco who worked out solutions to the problems and helped me in computer matter, to C. Y. Liu for two graphs, to M. C. White for editorial help, to S. Tiedemann for proof reading, and to B. Hutchings for her expert typing and administration related to the preparation of the manuscript.

C. L. Chiang, Berkeley
March 30, 2003