Preface

A significant portion of the subject called statistics that is used in applied problems is directly or peripherally related to the linear statistical model. There is perhaps no subject within the field of statistics that has been more widely investigated and published. This quantity of publications is ample proof of the usefulness of the linear model in real problems.

There is, of course, no complete agreement by all statisticians about what the content of linear model theory includes. Some would say that it is a part of traditional multivariate analysis, and in most cases it can be viewed this way. However, in this book I consider it to be what is usually referred to as analysis of variance, correlation and regression, and design of experiments.

Since the literature is so vast and the number of users so large, it is impossible to write a book that can cover all the topics or that will be appropriate for all people who are interested in the three areas of statistics referred to above. I have attempted to write a book that will be useful to the following people: (1) students who will obtain a master's degree in statistics and will work as statisticians in industry or government; (2) students who will obtain a Ph.D. degree in statistics and who plan to do teaching, consulting, or research in applied statistics; (3) students who will obtain a master's or Ph.D. degree in a field other than statistics but who plan to work in an area where analysis of variance, regression and correlation, or design of experiments will be required; (4) anyone who is working in a subject area who wants a beginning theory text in linear model theory for self-study or reference.

Background Needed

This is not a book on mathematics nor on theoretical statistics. It is intended to be an introductory book that covers some topics in linear models that will be useful to experimenters, statistical consultants, and those who are training to be statistical consultants. The mathematics courses required for this book include two or three semesters of college-level calculus and one semester of matrix algebra. The statistics requirements are two or three courses in applied statistics, including study of correlation and regression and the design of experiments, and one year of theoretical statistics from books such as [D-13], [F-2], [H-16], [L-10], and [M-7] listed in "References and Further Reading" at the end of this book. For the past several years I have taught a course at Colorado State University from notes which resulted in this book, and some of the students had fewer courses than those listed above, but many students had taken more courses. About twenty-five percent of the students who took this course during those years were graduate students not in statistics; about seventy-five percent of the students were graduate students in statistics. I cover all the material in the book in a two-semester course that meets three times a week.

Approach

In writing the book I have used a definition-theorem-proof procedure, but I have presented the proofs in what I call a "semirigorous" fashion: some proofs are merely sketched and a few are omitted. After Chapter 2, if a proof is omitted, a reference is generally given where a proof can be found. I have attempted to talk to the reader and in many places I have added notes and reminders that some readers may find helpful.

As I stated above, there are so many topics of potential interest in linear models that it is impossible to cover but a few of them. The approach that I take in this book is to discuss principally what is referred to as "normal theory" linear models and "least squares theory" linear models. While it is true that other "theories" of the linear model, often referred to as "distribution free" or "nonparametric" theories of the linear model, have received a great deal of attention, I feel that at the present time the normal and least squares theories are important for a first course. From a knowledge of these, one can study the so-called more robust procedures for the linear model that will certainly become a major part of statistical application in the future.

Emphases and Special Topics

Perhaps the most important chapter in the book is Chapter 6, "The General Linear Model." The chapters following Chapter 6 can be viewed as special cases or generalizations or applications of the material in that chapter. The chapters before Chapter 6 contain material that can be viewed as preparatory.

The word "application" in the title of this book may need some explanation. It refers to the following three methods of presentation: (1) some topics are discussed that are useful in applied problems, such as linear model prediction, tolerance intervals, inverse regression, polynomial and trigonometric models, testing whether models are equivalent, finding maximum (or minimum) of a quadratic model, finding the point of intersection of two models, selection of variables, growth curves, discrimination (classification); (2) in many cases topics are introduced by referring to real situations where they might be useful; (3) computing procedures are discussed for most of the topics covered.

Some of the special topics that are covered are discussed only briefly since a full discussion would require much more space than could possibly be devoted to them. In these cases I have given a setting for understanding the basics of the topic; then I present some elementary theory, followed by references where additional material that is current can be found. This will perhaps allow the reader to find those topics that may be useful and to work through the material in this book, then through some of the current literature on that particular subject.

Examples and Problems

The book contains about 150 examples, many of them involving data. In most cases "real" data are not used since the examples are used mainly to illustrate a point or to help fix an idea for the reader. There is something to be said for using real data, but if sufficient time and space are not devoted to a discussion of the situation that generated the real data, these arguments lose some of their force. I felt that it would not be a good use of space to use real data with all the attending explanations that would be essential. The classroom instructor can enrich the course by using real examples if they seem appropriate.

The book contains a large amount of material about methods for computing the quantities needed in problems where data are involved. Many computing methods are available and some of them are machine dependent. I chose to use the square root (Cholesky) procedure throughout the book. This decision was based on many factors, among which were the following: (1) the normal equations are very useful for the theoretical development of most topics in the linear model and they involve a nonnegative matrix; (2) in most cases (except possibly for high degree polynomial models) the normal equations can be compiled without error from the given data; (3) the square root procedure is one of the "best" all-around methods for obtaining the quantities needed when one starts with the normal equations, where the word "best" means a balance of storage requirements, programming ease, computing time, and accuracy of results. Of course, for very large data sets and poorly conditioned data matrices, a computing procedure that does not start with the normal equations would generally be expected to give more accurate results.

The book contains about 450 problems, and an instructor's solution manual is available that, in addition to answers, includes hints for the more difficult problems. The problems include some that are theoretical and some that require simple computing. Some are quite trivial and some are quite difficult.

Suggested Coverage

The book contains fifteen chapters. Chapters 1 and 2 contain review topics in mathematics (mainly matrix algebra) and statistics and can be read quickly by the reader who has adequate prerequisites for the course. Chapter 3 contains material on the multivariate normal, and Chapter 4 is devoted to quadratic forms of normal random variables. These two chapters can also be covered quite quickly. Chapter 5 discusses various models that are presented in the book and is a reading chapter. Chapter 6 is the first encounter with a detailed presentation of the linear model, and references to material in the first five chapters are made when that material is needed. Chapter 7 is a chapter on computing, and Chapter 8 is devoted to several applications of the general linear model. Chapter 9 contains a brief discussion of inference procedures when sampling from the multivariate normal distribution. Chapter 10 discusses the regression model and relates it to the general linear model, and Chapter 11 is devoted to inference procedures for simple, multiple, and partial correlation. Chapter 12 is devoted to applications of the regression model and includes regression prediction theory, selection of variables, growth curves, and discrimination. Chapter 13 discusses the general theory of the design model and the special case of the one-factor design model. In Chapter 14 there is a detailed discussion of the two-factor design model with and without interaction and with equal and unequal numbers. Chapter 15 is devoted to the subject of variance components.

As was stated previously, the material in this book can be covered in a two-semester course that meets three times per week. For a one-semester course the following is recommended. The material in Chapters 1, 2, 3, 4, and 5 can be covered quite rapidly but not in depth; cover Chapters 6 and 7 in depth; then select other chapters depending on one's interest. For applications of the linear model, cover Chapter 8; for correlation, cover Chapter 9 (but not in depth) and Chapter 11; for applications of the regression model, select Chapters 9, 10, 11, and 12; for design models, cover Chapters 13 and 14 (also Chapter 15 if one is interested in components of variance).

References

The book contains a "References and Further Reading" section where over 200 books and papers are listed. Since the literature on the linear model is so huge, one has to pick and choose to obtain a manageable set that adequately covers the main topics and includes the important publications. All items in this section are not referenced in the book. The intent is to include a list where one can start with a publication on a given topic, examine the referenced papers and others that are included, examine the references listed in these papers, and thereby obtain a set of several papers on the topic of interest. Very few publications dated after 1973 are included.

Acknowledgements

I would like to express appreciation to all who have contributed to this book. However, there have been so many contributors that I hesitate to name them for fear that some will be inadvertently omitted. I must thank my colleagues and the graduate students in the department of statistics at Colorado State University for their assistance which was provided in many ways. Thanks are also due to the administrative officers at Colorado State University for providing resources and an environment in which a project such as this can be accomplished. I am grateful to the Literary Executor of the late Sir Ronald A. Fisher, F.R.S., to Dr. Frank Yates, F.R.S. and to Longman Group Ltd., London, for permission to reprint Table III from their book *Statistical Tables for Biological, Agricultural and Medical Research*. (6th edition, 1974.) I want to thank the many authors and organizations for allowing me to reprint material and these acknowledgements are made in the book where the material appears. Finally, I want to thank my wife Jeanne who has helped me in so many ways, some of which I am sure I am not aware of.