Preface

The theory of linear models provides the foundation for two of the most important tools in applied statistics: regression and analysis of designed experiments. Thus it is no surprise that there are many books written on this topic. These books adopt a variety of approaches in providing an understanding of linear models. Recent books generally embrace the coordinate-free approach invoking vector spaces. In this book, *Linear Models: An Integrated Approach*, we also use vector spaces — but the emphasis is on statistical ideas and interpretations.

Why another book? Because we firmly believe that a comprehensive story of linear models can be told in simple language by appealing to the reader's statistical thinking. We develop the basic theory using essentially two simple statistical concepts, the linear zero function and the principle of covariance adjustment. Although these two ideas go back a few decades, their potential in developing the theory of linear models has not been fully exploited till recently. In this context, linear zero functions correspond to ancillary statistics. It is fascinating how they can be used to make complex expressions look obvious, particularly when the error dispersion matrix is singular. We believe that there is no easier or better way for the exposition of the general linear model. We also review the other 'unified theories' of linear unbiased estimation in the general linear model, and provide their statistical motivation which may not be available elsewhere.

The syllabus of a graduate-level course on 'Linear Models' taught in many universities includes a good bit of application in regression or analysis of designed experiments. This leaves less space for the underlying theory of linear models in most contemporary books on the subject. An important secondary objective of this book is to provide a more complete and up-to-date treatment of this relatively neglected area.

Linear Models: An Integrated Approach aims to achieve 'integration' in more ways than one. Establishing basic principles which link the general linear model with the special case of homoscedastic errors is one aspect of it. This linkage continues well beyond the derivation of the main results. The approach based on linear zero functions is used to derive and interpret results for many modifications and extensions. These include change in the residual sum of squares due to restrictions, effects of nuisance parameters and inclusion or exclusion of observations or parameters, as well as the multivariate linear model. Results on the decomposition of sum of squares in designed experiments, recursive inference, Kalman filter, missing plot techniques, deletion diagnostics and so on for the general linear model follow in a natural way. Simultaneous tests, confidence and tolerance intervals for parameters of the general linear model are also provided.

Another aspect of 'integration' is the unified treatment of linear models with partially specified dispersion matrix, which include such special cases as variance components or mixed effects models and linear models with serially or spatially correlated errors.

The third aspect of 'integration' is a comprehensive discussion of the foundations of linear inference, which is developed by carefully sequencing material that had been scattered in a number of articles. This theory runs parallel to the general theory of statistical inference, and shows how the approach based on linear zero functions emanates naturally from the basic principles of inference — without any distributional assumption.

After giving a brief introduction to the linear model in Chapter 1, we provide in Chapters 2 and 3 a summary of the algebraic and statistical results that would be used in the later chapters. In the next two chapters, we consider the linear model with uncorrelated errors of equal variance. We develop the theory of linear estimation in Chapter 4 and proceed to discuss confidence regions, testing and prediction in Chapter 5. Analysis of variance and covariance in designed experiments is considered in Chapter 6. The results of Chapters 4 and 5 are then extended to the case of the general linear model with an arbitrary dispersion matrix, in Chapter 7. In Chapter 8 we consider the case when the error dispersion matrix is misspecified or partially specified. We deal with updates in the general linear model in Chapter 9 and with multivariate response in Chapter 10. In Chapter 11 we discuss the statistical foundation for linear inference, alternative linear estimators, a geometric perspective and asymptotic results.

The book contains over 300 end-of-the-chapter exercises. These are meant to (a) illustrate the material covered, (b) supplement some results with proofs, interpretations and extensions, (c) introduce or expand ideas that are related to the text of the chapter, and (d) give glimpses of interesting research issues. These are arranged, more or less, in the order of the corresponding sections. Solutions to selected exercises are given in the appendix, while solutions to almost all the other exercises are given in the URL `http://www.isical.ac.in/~sdebasis/linmodel`. The URL also contains soft copies of the data sets.

Linear inference in the linear model often serves as a benchmark for other methods and models. While providing a comprehensive account of the state of the art in this area, we were unable to cover other inference procedures or other related models — for which several books are already available. Topics that are left out include nonlinear methods of inference such as Bayesian, robust and rank-based methods, resampling techniques, inference in the generalized linear model and data-analytic methods such as transformation of variables and use of missing data.

The book is meant primarily for students and researchers in statistics. Engineers and scientists who need a thorough understanding of linear models for their work may also find it useful. Even though practitioners may not find structured instructions, they will find in this book a reference for many of the tools they need and use — such as leverages, residuals, deletion diagnostics, indicators of collinearity and various plots. We also hope that researchers will find stimulus for further work from the perspective of current research discussed in the later chapters and the suggestions and leads provided in some of the exercises.

Familiarity with statistical inference and linear algebra at the upper division or first-year graduate level is a prerequisite for reading this book. Essential topics in these areas are briefly reviewed in Chapters 2 and 3. Mastery of algebra is *not* a prerequisite; we have simplified the proofs to the extent that algebra does not obscure the statistical content.

A one-semester graduate-level introductory course in linear models can be taught by using Chapters 1–6 and selected topics from the other chapters. A follow-up course can be taught from Chapters 7–11. Some sections in Chapters 4–9 are marked with an asterisk; these may be omitted during the first reading. For students who have already had a first course in linear models/regression elsewhere, a second course may be taught by rushing through Chapters 4–6, covering Chapter 7 in detail and then teaching selected topics from Chapters 8–11.

We do not make a distinction among lemmas, theorems and corollaries in this book. All of these are called propositions. The propositions, definitions, remarks and examples are numbered consecutively within a section, in a common sequence. Equations are also numbered consecutively within a section. Throughout the book, vectors are represented by lowercase and boldface letters, while uppercase and boldface letters are used to denote matrices. No notational distinction is made between a random vector and a particular realization of it.

Errata for the book appear at the URL mentioned earlier. We would appreciate receiving comments and suggestions on the book sent by e-mail to `sdebasis@isical.ac.in`.

The approach adopted in this book has its roots in the lecture notes of R.C. Bose (1949). It was conceived in its present form by P. Bhimasankaram of Indian Statistical Institute, who further indebted us by providing extensive suggestions on several versions of the manuscript. Comments from Professors Bikas K. Sinha and Anis C. Mukhopadhyay of Indian Statistical Institute and Professor Thomas Mathew of University of Maryland, Baltimore County were also very useful. The first author thanks his family and friends for coping with several long periods of absence and specially his son Shairik for putting up a brave face even as he missed his father's company.

Debasis Sengupta
Sreenivasa Rao Jammalamadaka