Statistical Learning from a Regression Perspective
R. A. Berk, 2008
New York, Springer
xviii + 360 pp., £55.99
ISBN 978-0-387-77500-5

There has been a good amount of development in the area of linear models and regression analysis. Advancements in computing facilities have led to the development of alternative techniques of modelling based on the philosophy of regression analysis. The emphasis of such approaches is more on computational aspects, concepts and algorithms than on developing only the hard core theory. Several such techniques have been developed and became popular in the last decade under the purview of topics of statistical learning. The present book compiles some of the recently developed statistical learning techniques. The emphasis in the book is more on the applications of such techniques through computations rather than on their mathematical and statistical properties. Nevertheless, the necessary mathematical and statistical details accompany the applications but they are not discussed in detail. This book should be of interest to statisticians for a few reasons. Firstly, it gathers together various techniques of statistical learning in one place and explains how to use them. Secondly, the author is well versed with the developments in the area of regression analysis; see Berk (2003). So he has successfully bridged the gaps between analytical and computational developments related to regression analysis. The book will serve as a base for such topics for a long time in spite of the rapid developments in the field of statistical learning. A key feature of this book is that the regression function is used in terms of conditional distributions.

The book is developed in eight chapters. The discussion in Chapter 1 presents the need and motivation for learning the tools that are alternative to traditional theory of regression analysis through several data-based examples and with various convincing arguments. Some underlying concepts of regression analysis that are needed for the development of theory in further chapters are also explained with minimal mathematical input. Next, Chapter 2 elaborates various aspects of regression splines and regression smoothers. It presents stepwise discussion on different types of splines and penalized smoothing, and addresses issues that are related to various methods of choosing smoothers with different type of variables along with illustrations. Chapter 3 deals with the issues, concepts and methodology of classification and regression trees. The trio of statistical learning procedures—bagging, random forests and boosting—are discussed in Chapters 4, 5 and 6 respectively. Each of these chapters presents motivation, reasoning for using the methodology, steps involved and related issues in detail. The topic of support vector machines is detailed in Chapter 7. Lastly, Chapter 8 discusses the limitations, advantages, disadvantages and practical suggestions on the topics that are presented in the earlier chapters. Every chapter contains a section which guides a reader on the use of software but the text is not tied to any specific software. The illustrations in all the chapters are data based and their analysis clarifies the application and usage of tools. Exercises are presented at the end of every chapter.

The reader should have an adequate statistical background to understand the topics of this book. Those researchers who hesitate to read too much mathematical and statistical content to understand the philosophy and issues in the topics of statistical learning and are more interested in the applied aspects will find the book interesting. The book is well suited for teaching a course at an advanced undergraduate or graduate level, even as a full semester course, and it also can be comprehensively used as a part of any course in statistics.

Reference

Berk, R. A. (2003) *Regression Analysis: a Constructive Critique*. Newbury Park: Sage.

Shalabh
Indian Institute of Technology Kanpur
E-mail: shalab@iitk.ac.in

Bayesian Methods for Data Analysis, 3rd edn
B. P. Carlin and T. A. Louis, 2009
Boca Raton, Chapman and Hall
xvi + 536 pp., $69.95
ISBN 978-1-584-88697-6

This is the third edition of a book that was originally aimed at advanced students of statistics. The authors widened the scope of the book in the second edition, and this has been widened further in this edition. Although it still shows signs of its more academic origins, the book contains some useful advice for practitioners. All the essential topics are covered: statistical inference, Bayesian computation, model criticism and selection, as well as Bayesian design. The authors argue that the Bayesian approach is to be recommended even if judged by the frequentist properties of posterior parameter estimates. The book includes a chapter on empirical Bayes which provides another approach that attempts to capture some of the benefits of Bayesian methods while sidestepping the more controversial aspects of the prior. Even when discussing the more complete Bayesian approach, they encourage users to think

of priors as 'tuning parameters' that can be used to produce a decision rule with broad validity. There is only the briefest discussion of how to obtain informative priors, which reinforces the feeling that the authors are keen to appeal to sceptics of the validity of formally including subjective prior beliefs in their statistical analyses.

Throughout the text one can find good practical advice on various implementational issues, and there is a whole chapter dedicated to case-studies. The chapter on Bayesian design provides very good coverage of some clinical trial design ideas that are receiving a considerable amount of interest in the pharmaceutical industry currently.

This book, by two very experienced and knowledgeable Bayesians, is a valuable contribution to the growing literature on the practical application of Bayesian methods. However, it is probably not the best starting place for a newcomer to the Bayesian approach.

Phil Woodward
Herne
E-mail: phil.woodward@pfizer.com

Statistical Design
G. Casella, 2008
New York, Springer
xxiv + 308 pp., £55.99
ISBN 978-0-837-75964-7

Many undergraduate and postgraduate programmes in statistics teach courses on design and analysis of experiments but they concentrate more on the analysis part than on the understanding of fundamental issues, which are generally taken as trivial to explain. Most of the available books on the topics in design of experiments aim at the standard analysis or deal with advanced topics whereas the basic fundamental underlying concepts are, in general, covered briefly. This book fills up such gaps and emphasizes the intricacies of various statistical concepts and principles related to the design of experiments which are well and clearly explained through various examples. The book is written with a refreshing style and the presentation is focused on the underlying fundamentals and principles which are essential for good design of experiments.

The book begins with a chapter on the basics of various underlying concepts and terminologies that are used in the design of experiments; for example, it starts even with answering 'what is an experimental unit'. All the topics in this chapter are well explained with several examples which make the reading interesting and concepts easy to understand in detail. The next Chapter 2 presents the details and analysis of completely randomized designs along with 2^2- and 2^4-factorial experiments. Chapter 3 deals with complete-block designs when block effects are fixed and discusses randomized complete-block design as well as Latin square design. The randomized complete-block design under the random-block effects is discussed in Chapter 4. The topics and analysis of split plot and crossover designs are presented in Chapter 5 and also compared with the whole plot analysis which helps a reader in better understanding the differences between the two cases. The last chapter, Chapter 6, addresses the confounding aspects in designs and discusses the balanced incomplete-block design, fractional factorial, balanced lattice design, reference design and loop design. The topics that are covered in this chapter are not discussed in detail and may not be sufficient for a first-time reader to understand the theory that is involved in these designs in detail. An appendix concisely presents the layout and corresponding analysis-of-variance tables of various designs that are discussed in the earlier chapters. There is a section at the end of each chapter which is divided into two subsections, i.e. technical notes and miscellanea. These provide the necessary proofs of the results and explain in more detail the terminologies which are used in the respective chapters. A reader who is more interested in the applications than in the mathematical details may skip these two subsections and can read the book without any problem. Every chapter is supplemented with many exercises which are divided into two categories, i.e. essential and accompaniment which will help the instructors teaching the course. All the topics and various steps that are involved in the analysis of designs are clearly detailed stepwise with examples in every chapter. It is difficult to find such a variety of examples on various designs in many available books. Some examples are based on microarray data, which are more recent developments and are not available in classical books.

The book is written for students in statistical courses and for practising statisticians. It can be useful for statisticians in different environments and with different levels of experience. The book is self-contained, well presented, well structured and well suited for teaching a course at an undergraduate level, even as a full semester course. As the author states in his preface, this book should not be taken as an encyclopaedia of the topics in the design of experiments. Most of the standard designs are covered and the way that they are presented makes the book excellent and different from many other books on these topics.

Shalabh
Indian Institute of Technology Kanpur
E-mail: shalab@iitk.ac.in