Springer Series in Statistics

C. Radhakrishna Rao
Helge Toutenburg

Linear Models

Least Squares and Alternatives

线性模型

Springer-Verlag
世界图书出版公司

C. Radhakrishna Rao
Helge Toutenburg

Linear Models

Least Squares and Alternatives

With 33 Illustrations

Springer-Verlag
世界图书出版公司
北京·广州·上海·西安

C. Radhakrishna Rao
The Pennsylvania State University
Department of Statistics
University Park, PA 16802
USA

Helge Toutenburg
Institut für Statistik
Universität München
80799 München
Germany

Library of Congress Cataloging-in-Publication Data
Rao, C. Radhakrishna (Calyampudi Radhakrishna), 1920–
Linear models: least squares and alternatives / C. Radhakrishna Rao, Helge Toutenburg.
p. cm. – (Springer series in statistics)
Includes bibliographical references (p. -) and index.
ISBN 0-387-94562-8 (hardcover : alk. paper)
1. Linear models (Statistics) I. Toutenburg, Helge. II. Title. III. Series.
QA279.R3615 1995
519.5'36—dc20 95-23947

Printed on acid-free paper.

Reprinted in China by Beijing World Publishing Corporation, 1998

9 8 7 6 5 4 3 2 1

ISBN 0-387-94562-8 Springer-Verlag New York Berlin Heidelberg

书　　名：Linear Models
作　　者：C.R.Rao , H.Toutenburg
中 译 名：线性模型
出 版 者：世界图书出版公司北京公司
印 刷 者：北京中西印刷厂
发　　行：世界图书出版公司北京公司 (北京朝阳门内大街 137 号　100010)
开　　本：大 32 开 850 × 1168　印　张：11.5
版　　次：1998 年 8 月第 1 版　1998 年 8 月第 1 次印刷
书　　号：7-5062-3818-7/ O•228
版权登记：图字 01-98-0432
定　　价：56.00 元

Springer Series in Statistics

Advisors:
P. Diggle, S. Fienberg, K. Krickeberg,
I. Olkin, N. Wermuth

Springer
New York
Berlin
Heidelberg
Barcelona
Budapest
Hong Kong
London
Milan
Paris
Santa Clara
Singapore
Tokyo

Springer Series in Statistics

Andersen/Borgan/Gill/Keiding: Statistical Models Based on Counting Processes.
Anderson: Continuous-Time Markov Chains: An Applications-Oriented Approach.
Andrews/Herzberg: Data: A Collection of Problems from Many Fields for the Student and Research Worker.
Anscombe: Computing in Statistical Science through APL.
Berger: Statistical Decision Theory and Bayesian Analysis, 2nd edition.
Bolfarine/Zacks: Prediction Theory for Finite Populations.
Brémaud: Point Processes and Queues: Martingale Dynamics.
Brockwell/Davis: Time Series: Theory and Methods, 2nd edition.
Choi: ARMA Model Identification.
Daley/Vere-Jones: An Introduction to the Theory of Point Processes.
Dzhaparidze: Parameter Estimation and Hypothesis Testing in Spectral Analysis of Stationary Time Series.
Fahrmeir/Tutz: Multivariate Statistical Modelling Based on Generalized Linear Models.
Farrell: Multivariate Calculation.
Federer: Statistical Design and Analysis for Intercropping Experiments.
Fienberg/Hoaglin/Kruskal/Tanur (Eds.): A Statistical Model: Frederick Mosteller's Contributions to Statistics, Science and Public Policy.
Fisher/Sen: The Collected Works of Wassily Hoeffding
Good: Permutation Tests A Practical Guide to Resampling Methods for Testing Hypotheses
Goodman/Kruskal Measures of Association for Cross Classifications
Grandell Aspects of Risk Theory.
Hall: The Bootstrap and Edgeworth Expansion.
Härdle: Smoothing Techniques With Implementation in S
Hartigan Bayes Theory
Heyer: Theory of Statistical Experiments
Jolliffe. Principal Component Analysis
Kotz/Johnson (Eds) Breakthroughs in Statistics Volume I.
Kotz/Johnson (Eds.): Breakthroughs in Statistics Volume II.
Kres: Statistical Tables for Multivariate Analysis.
Leadbetter/Lindgren/Rootzén: Extremes and Related Properties of Random Sequences and Processes.
Le Cam: Asymptotic Methods in Statistical Decision Theory.
Le Cam/Yang: Asymptotics in Statistics: Some Basic Concepts.
Manoukian: Modern Concepts and Theorems of Mathematical Statistics.
Manton/Singer/Suzman (Eds.): Forecasting the Health of Elderly Populations.
Miller, Jr.: Simultaneous Statistical Inference, 2nd edition.
Mosteller/Wallace: Applied Bayesian and Classical Inference: The Case of *The Federalist Papers.*

(continued after index)

Preface

The book is based on both authors' several years of experience in teaching linear models at various levels. It gives an up-to-date account of the theory and applications of linear models. The book can be used as a text for courses in statistics at the graduate level and as an accompanying text for courses in other areas. Some of the highlights in this book are as follows.

A relatively extensive chapter on matrix theory (Appendix A) provides the necessary tools for proving theorems discussed in the text and offers a selection of classical and modern algebraic results that are useful in research work in econometrics, engineering, and optimization theory. The matrix theory of the last ten years has produced a series of fundamental results about the definiteness of matrices, especially for the differences of matrices, which enable superiority comparisons of two biased estimates to be made for the first time.

We have attempted to provide a unified theory of inference from linear models with minimal assumptions. Besides the usual least-squares theory, alternative methods of estimation and testing based on convex loss functions and general estimating equations are discussed. Special emphasis is given to sensitivity analysis and model selection.

A special chapter is devoted to the analysis of categorical data based on logit, loglinear, and logistic regression models.

The material covered, theoretical discussion, and its practical applications will be useful not only to students but also to researchers and consultants in statistics.

We would like to thank our colleagues Dr. G. Trenkler and Dr. V. K. Srivastava for their valuable advice during the preparation of the book. We wish to acknowledge our appreciation of the generous help received from Andrea Schöpp, Andreas Fieger, and Christian Kastner for preparing a fair copy. Finally, we would like to thank Dr. Martin Gilchrist of Springer-Verlag for his cooperation in drafting and finalizing the book.

We request that readers bring to our attention any errors they may find in the book and also give suggestions for adding new material and/or improving the presentation of the existing material.

C. Radhakrishna Rao
University Park, Pennsylvania
USA

Helge Toutenburg
Munich,
Germany

July 1995

Springer Series in Statistics

(continued from p. ii)

Pollard: Convergence of Stochastic Processes.
Pratt/Gibbons: Concepts of Nonparametric Theory.
Rao/Toutenburg: Linear Models: Least Squares and Alternatives.
Read/Cressie: Goodness-of-Fit Statistics for Discrete Multivariate Data.
Reinsel: Elements of Multivariate Time Series Analysis.
Reiss: A Course on Point Processes.
Reiss: Approximate Distributions of Order Statistics: With Applications to Non-parametric Statistics.
Rieder: Robust Asymptotic Statistics.
Ross: Nonlinear Estimation.
Sachs: Applied Statistics: A Handbook of Techniques, 2nd edition.
Salsburg: The Use of Restricted Significance Tests in Clinical Trials.
Särndal/Swensson/Wretman: Model Assisted Survey Sampling.
Schervish: Theory of Statistics.
Seneta: Non-Negative Matrices and Markov Chains, 2nd edition.
Shedler: Regeneration and Networks of Queues.
Siegmund: Sequential Analysis: Tests and Confidence Intervals.
Tanner: Tools for Statistical Inference: Methods for the Exploration of Posterior Distributions and Likelihood Functions, 2nd edition.
Todorovic: An Introduction to Stochastic Processes and Their Applications.
Tong: The Multivariate Normal Distribution.
Vapnik: Estimation of Dependences Based on Empirical Data.
Weerahandi: Exact Statistical Methods for Data Analysis.
West/Harrison: Bayesian Forecasting and Dynamic Models.
Wolter: Introduction to Variance Estimation.
Yaglom: Correlation Theory of Stationary and Related Random Functions I: Basic Results.
Yaglom: Correlation Theory of Stationary and Related Random Functions II: Supplementary Notes and References.

Contents

1
Introduction

Linear models play a central part in modern statistical methods. On the one hand, these models are able to approximate a large amount of metric data structures in their entire range of definition or at least piecewise. On the other hand, approaches such as the analysis of variance, which model effects as linear deviations from a total mean, have proved their flexibility. The theory of generalized models enables us, through appropriate link functions, to apprehend error structures that deviate from the normal distribution and hence ensuring, that a linear model is maintained in principle. Numerous iterative procedures for solving the normal equations were developed especially for those cases where no explicit solution is possible. For the derivation of explicit solutions in rank-deficient linear models, classical procedures are available: for example, ridge or principal component regression, partial least squares, as well as the methodology of the generalized inverse. The problem of missing data in the variables can be dealt with by appropriate imputation procedures.

Chapter 2 describes the hierarchy of the linear models, starting with the classical regression model, up to the structural model of econometrics.

Chapter 3 contains the standard procedures for estimating and testing in regression models with full or reduced rank of the design matrix, algebraic and geometric properties of the OLS estimate, as well as an introduction to minimax estimation when auxiliary information is available in the form of inequality restrictions. The concepts of partial and total least squares, projection pursuit regression, and censored regression are introduced.

Chapter 4 describes the theory of best linear estimates in the generalized regression model, effects of misspecified covariance matrices, as well as special covariance structures of heteroscedasticity and first order autoregression.

Chapter 5 is devoted to estimation under exact or stochastic linear restrictions. The comparison of two biased estimators according to the MDE criterion is based on recent theorems of matrix theory. The results are the outcome of intensive international research over the last ten years and appear here for the first time in a coherent form. This concerns the concept of the weak r-unbiasedness as well.

Chapter 6 contains the theory of the optimal linear prediction and gives, in addition to known results, an insight into recent studies about the MDE matrix comparison of optimal and classical predictions according to alternative superiority criteria.

Chapter 7 presents ideas and procedures for studying the effect of single data rows on the estimation of β. Here, different measures for revealing outliers or influential points, including graphical methods, are incorporated. Some examples illustrate this.

Chapter 8 deals with missing data in the design matrix X. After introducing the general problems and defining the various missing data mechanisms according to Rubin, we demonstrate "adjustment by follow-up interviews" for long-term studies with dropout. For the regression model the method of imputation is described, in addition to the analysis of the loss of efficiency in case of a reduction to the completely observed submodel. The method of weighted mixed estimates is presented for the first time in a textbook on linear models.

Chapter 9 contains recent contributions to robust statistical inference based on M-estimation.

Chapter 10 describes the model extensions for categorical response and explanatory variables. Here, the binary response and the loglinear model are of special interest. The model choice is demonstrated by means of examples. Categorical regression is integrated into the theory of generalized linear models.

An independent chapter (Appendix A) on matrix algebra summarizes standard theorems (including proofs) that are of interest for the book itself, but also for linear statistics in general. Of special interest are the theorems about decomposition of matrices (A.30–A.34), definite matrices (A.35–A.59), the generalized inverse, and especially about the definiteness of differences between matrices (Theorem A.71; cf. A.74–A.78).

The book offers an up-to-date and comprehensive account of the theory and applications of linear models.

Tables for the χ^2- and F-distributions are provided in Appendix B.

2

Linear Models

2.1 Regression Models in Econometrics

The methodology of regression analysis, one of the classical techniques of mathematical statistics, is an essential part of the modern econometric theory.

Econometrics combines elements of economics, mathematical economics, and mathematical statistics. The statistical methods used in econometrics are oriented toward specific econometric problems and hence are highly specialized. In economic laws stochastic variables play a distinctive role. Hence econometric models, adapted to the economic reality, have to be built on appropriate hypotheses about distribution properties of the random variables. The specification of such hypotheses is one of the main tasks of econometric modelling. For the modelling of an economic (or a scientific) relation, we assume that this relation has a relative constancy over a sufficiently long period of time (that is, over a sufficient length of observation period), since otherwise its general validity would not be ascertainable. We distinguish between two characteristics of a structural relationship, the *variables* and the *parameters*. The variables, which we will classify later on, are those characteristics whose values in the observation period can vary. Those characteristics that do not vary can be regarded as the structure of the relation. The *structure* consists of the functional form of the relation, including the relation between the main variables, the type of probability distribution of the random variables, and the parameters of the model equations.