

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.

© 1995 Springer-Verlag New York, Inc.

All rights reserved. This work may not be translated or copied in whole or in part without the written permission of the publisher (Springer-Verlag New York, Inc., 175 Fifth Avenue, New York, NY 10010, USA), except for brief excerpts in connection with reviews or scholarly analysis. Use in connection with any form of information storage and retrieval, electronic adaptation, computer software, or by similar or dissimilar methodology now known or hereafter developed is forbidden. The use of general descriptive names, trade names, trademarks, etc., in this publication, even if the former are not especially identified, is not to be taken as a sign that such names, as understood by the Trade Marks and Merchandise Marks Act, may accordingly be used freely by anyone.

This reprint has been authorized by Springer-Verlag (Berlin/Heidelberg/New York) for sale in the People's Republic of China only and not for export therefrom.
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

世界图书出版公司北京公司已获得 Springer-Verlag 授权在中国
境内独家重印发行.

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.
- Bolifarine/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

Preface	v
1 Introduction	1
2 Linear Models	3
2.1 Regression Models in Econometrics	3
2.2 Econometric Models	6
2.3 The Reduced Form	10
2.4 The Multivariate Regression Model	12
2.5 The Classical Multivariate Linear Regression Model	15
2.6 The Generalized Linear Regression Model	16
3 The Linear Regression Model	19
3.1 The Linear Model	19
3.2 The Principle of Ordinary Least Squares (OLS)	20
3.3 Geometric Properties of OLS	21
3.4 Best Linear Unbiased Estimation	23
3.4.1 Basic Theorems	23
3.4.2 Linear Estimators	28
3.4.3 Mean Dispersion Error	29
3.5 Estimation (Prediction) of the Error Term e and σ^2	31
3.6 Classical Regression under Normal Errors	32
3.7 Testing Linear Hypotheses	34
3.8 Analysis of Variance and Goodness of Fit	41

3.8.1	Bivariate Regression	41
3.8.2	Multiple Regression	46
3.8.3	A Complex Example	50
3.8.4	Graphical Presentation	54
3.9	The Canonical Form	60
3.10	Methods for Dealing with Multicollinearity	61
3.10.1	Principal Components Regression	62
3.10.2	Ridge Estimation	63
3.10.3	Shrinkage Estimates	67
3.10.4	Partial Least Squares	67
3.11	Projection Pursuit Regression	71
3.12	Total Least Squares	72
3.13	Minimax Estimation	75
3.13.1	Inequality Restrictions	75
3.13.2	The Minimax Principle	78
3.14	Censored Regression	83
3.14.1	Introduction	83
3.14.2	LAD Estimators and Asymptotic Normality	84
3.14.3	Tests of Linear Hypotheses	85
4	The Generalized Linear Regression Model	89
4.1	Optimal Linear Estimation of β	89
4.2	The Aitken Estimator	96
4.3	Misspecification of the Dispersion Matrix	98
4.4	Heteroscedasticity and Autoregression	101
5	Exact and Stochastic Linear Restrictions	111
5.1	Use of Prior Information	111
5.2	The Restricted Least-Squares Estimator	112
5.3	Stepwise Inclusion of Exact Linear Restrictions	115
5.4	Biased Linear Restrictions and MDE Comparison with the OLSE	120
5.5	MDE Matrix Comparisons of Two Biased Estimators	123
5.6	MDE Matrix Comparison of Two Linear Biased Estimators	129
5.7	MDE Comparison of Two (Biased) Restricted Estimators	130
5.7.1	Special Case: Stepwise Biased Restrictions	134
5.8	Stochastic Linear Restrictions	138
5.8.1	Mixed Estimator	138
5.8.2	Assumptions about the Dispersion Matrix	140
5.8.3	Biased Stochastic Restrictions	143
5.9	Weakened Linear Restrictions	147
5.9.1	Weakly (R, r) -Unbiasedness	147
5.9.2	Optimal Weakly (R, r) -Unbiased Estimators	148
5.9.3	Feasible Estimators—Optimal Substitution of β in $\hat{\beta}_1(\beta, A)$	152

5.9.4	RLSE Instead of the Mixed Estimator	153
6	Prediction Problems in the Generalized Regression Model	155
6.1	Introduction	155
6.2	Some Simple Linear Models	155
6.3	The Prediction Model	158
6.4	Optimal Heterogeneous Prediction	159
6.5	Optimal Homogeneous Prediction	161
6.6	MDE Matrix Comparisons between Optimal and Classical Predictors	165
6.6.1	Comparison of Classical and Optimal Prediction with Respect to the y_* -Superiority	167
6.6.2	Comparison of Classical and Optimal Predictors with Respect to the $X_*\beta$ -Superiority	170
6.7	Prediction Regions	172
7	Sensitivity Analysis	177
7.1	Introduction	177
7.2	Prediction Matrix	177
7.3	The Effect of a Single Observation on the Estimation of Parameters	183
7.3.1	Measures Based on Residuals	184
7.3.2	Algebraic Consequences of Omitting an Observation	185
7.3.3	Detection of Outliers	186
7.4	Diagnostic Plots for Testing the Model Assumptions	191
7.5	Measures Based on the Confidence Ellipsoid	192
7.6	Partial Regression Plots	199
8	Analysis of Incomplete Data Sets	203
8.1	Statistical Analysis with Missing Data	203
8.2	Missing Data in the Response	209
8.2.1	Least-Squares Analysis for Complete Data	209
8.2.2	Least-Squares Analysis for Filled-up Data	210
8.2.3	Analysis of Covariance—Bartlett's Method	211
8.3	Missing Values in the X -Matrix	212
8.3.1	Missing Values and Loss in Efficiency	214
8.3.2	Standard Methods for Incomplete X -Matrices	216
8.4	Maximum Likelihood Estimates of Missing Values	219
8.5	Weighted Mixed Regression	221
8.5.1	Minimizing the MDEP	223
8.5.2	The Two-Stage WMRE	226
9	Robust Regression	229
9.1	Introduction	229
9.2	Least Absolute Deviation Estimators—Univariate Case	230

9.3	M-Estimates: Univariate Case	234
9.4	Asymptotic Distributions of LAD Estimators	237
9.4.1	Univariate Case	237
9.4.2	Multivariate Case	238
9.5	General M-Estimates	240
9.6	Test of Significance	243
10	Models for Binary Response Variables	247
10.1	Generalized Linear Models	247
10.2	Contingency Tables	249
10.2.1	Introduction	249
10.2.2	Ways of Comparing Proportions	250
10.2.3	Sampling in Two-Way Contingency Tables	253
10.2.4	Likelihood Function and Maximum Likelihood Estimates	254
10.2.5	Testing the Goodness of Fit	256
10.3	GLM for Binary Response	259
10.3.1	Logit Models	259
10.3.2	Loglinear Models	259
10.3.3	Logistic Regression	260
10.3.4	Testing the Model	263
10.3.5	Distribution Functions as Link Function	263
10.4	Logit Models for Categorical Data	264
10.5	Goodness of Fit—Likelihood-Ratio Test	265
10.6	Loglinear Models for Categorical Variables	267
10.6.1	Two-Way Contingency Tables	267
10.6.2	Three-Way Contingency Tables	270
10.7	The Special Case of Binary Response	274
10.8	Coding of Categorical Explanatory Variables	276
10.8.1	Dummy and Effect Coding	276
10.8.2	Coding of Response Models	279
10.8.3	Coding of Models for the Hazard Rate	281
A	Matrix Algebra	285
A.1	Introduction	285
A.2	Trace of a Matrix	287
A.3	Determinant of a Matrix	288
A.4	Inverse of a Matrix	290
A.5	Orthogonal Matrices	291
A.6	Rank of a Matrix	292
A.7	Range and Null Space	292
A.8	Eigenvalues and Eigenvectors	293
A.9	Decomposition of Matrices	295
A.10	Definite Matrices and Quadratic Forms	298
A.11	Idempotent Matrices	304

A.12 Generalized Inverse	305
A.13 Projectors	314
A.14 Functions of Normally Distributed Variables	315
A.15 Differentiation of Scalar Functions of Matrices	318
A.16 Miscellaneous Results, Stochastic Convergence	321
B Tables	325
References	329
Index	345

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.