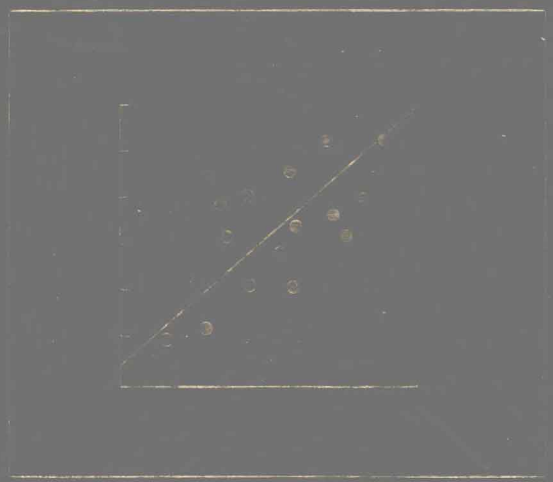
APPLIED LINEAR REGRESSION MODELS

THIRD EDITION



NETER KUTNER NACHTSHEIM WASSERMAN

THIRD EDITION

Applied Linear Regression Models

John Neter
University of Georgia

Michael H. Kutner
The Cleveland Clinic Foundation

Christopher J. Nachtsheim
University of Minnesota

William Wasserman
Syracuse University

IRWIN
Chicago • Bogatá • Boston • Buenos Aires • Caracas
London • Madrid • Mexico City • Sydney • Toronto

IRWIN **Concerned about Our Environment**

In recognition of the fact that our company is a large end-user of fragile yet replenishable resources, we at IRWIN can assure you that every effort is made to meet or exceed Environmental Protection Agency (EPA) recommendations and requirements for a "greener" workplace.

To preserve these natural assets, a number of environmental policies, both companywide and department-specific, have been implemented. From the use of 50% recycled paper in our textbooks to the printing of promotional materials with recycled stock and soy inks to our office paper recycling program, we are committed to reducing waste and replacing environmentally unsafe products with safer alternatives.

© The McGraw-Hill Companies, Inc., 1983, 1989, and 1996

All rights reserved. No part of this publication may be reproduced, stored in a retrieval system, or transmitted, in any form or by any means, electronic, mechanical, photocopying, recording, or otherwise, without the prior written permission of the publisher.

Trademark Acknowledgments:

IBM and IBM PC are registered trademarks of International Business Machines Corporation.

Disclaimer

Richard D. Irwin, Inc. makes no warranties, either expressed or implied, regarding the enclosed computer software package, its merchantability or its fitness for any particular purpose. The exclusion of implied warranties is not permitted by some states. The exclusion may not apply to you. This warranty provides you with specific legal rights. There may be other rights that you may have which may vary from state to state.

Irwin Book Team

Publisher: Tom Casson
Senior sponsoring editor: Richard T. Hercher, Jr.
Senior developmental editor: Gail Korosa
Senior marketing manager: Brian Kibby
Project editor: Lynne Basler
Production supervisor: Laurie Kersch
Designer: Matthew Baldwin
Cover designer: Stuart Paterson
Assistant Manager, Graphics: Charlene R. Perez
Compositor/Art Studio: Graphic Composition, Inc.
Typeface: 10/12 Times Roman
Printer: R. R. Donnelley & Sons Company

Library of Congress Cataloging-in-Publication Data

Applied linear regression models / John Neter . . . [et al.]. — 3rd ed.
p. cm.
Includes index.
ISBN 0-256-08601-X
1. Regression analysis. I. Neter, John
QA278.2.A65 1996
519.5'36—dc20 95–37447

Printed in the United States of America
6 7 8 9 0 DO 2

Applied Linear Regression Models

THE IRWIN SERIES IN PRODUCTION OPERATIONS MANAGEMENT

Aquilano, Chase, and Davis
Fundamentals of Operations Management
Second Edition

Chase and Aquilano
Production and Operations Management: *Manufacturing and Services*
Seventh Edition

Hill
Manufacturing Strategy: *Text & Cases*
Second Edition

Hopp and Spearman
Factory Physics
First Edition

Lambert and Stock
Strategic Logistics Management
Third Edition

Leenders, Fearon, and England
Purchasing and Materials Management
Ninth Edition

Lotfi and Pegels
Decision Support Systems For Operations and Management Science
Third Edition

Nahmias
Production and Operations Analysis
Second Edition

Niebel
Motion and Time Study
Ninth Edition

Schonberger and Knod
Operations Management: *Continuous Improvement*
Fifth Edition

Stevenson
Production/Operations Management
Fifth Edition

Vollmann, Berry, and Whybark
Manufacturing Planning & Control Systems
Third Edition

Zimmerman
ABC Operations Software
Beta Edition

THE IRWIN SERIES IN STATISTICS

Aczel
Statistics: *Concepts and Applications*
First Edition

Aczel
Complete Business Statistics
Second Edition

Bryant and Smith
Practical Data Analysis: *Case Studies in Business Statistics, Volumes I and II*
First Edition

Duncan
Quality Control & Industrial Statistics
Fifth Edition

Cooper and Emory
Business Research Methods
Fifth Edition

Gitlow, Oppenheim, and Oppenheim
Quality Management: *Tools and Methods For Improvement*
Second Edition

Hall
Computerized Business Statistics
Fourth Edition

Hanke and Reitsch
Understanding Business Statistics
Second Edition

Lind and Mason
Basic Statistics for Business and Economics
First Edition

Mason and Lind
Statistical Techniques in Business and Economics
Ninth Edition

Neter, Kutner, Nachtsheim, and Wasserman
Applied Linear Statistical Models
Fourth Edition

Neter, Kutner, Nachtsheim, and Wasserman
Applied Linear Regression Models
Third Edition

Shin
The Irwin Statistical Software Series: *Minitab, SAS, SPSS Guides*
First Edition

Siegel
Practical Business Statistics
Second Edition

Webster
Applied Statistics For Business and Economics
Second Edition

Wilson and Keating
Business Forecasting
Second Edition

THE IRWIN SERIES IN QUANTITATIVE METHODS AND MANAGEMENT SCIENCE

Bierman, Bonini, and Hausman
Quantitative Analysis For Business Decisions
Eighth Edition

Knowles
Management Science: *Building and Using Models*
First Edition

Lotfi and Pegels
Decision Support Systems For Operations and Management Science
Third Edition

Stevenson
Introduction to Management Science
Second Edition

Turban and Meredith
Fundamentals of Management Science
Sixth Edition

此为试读,需要完整PDF请访问: www.ertongbook.com

To
Dorothy, Ron, David,
Nancy, Michelle, Allison,
Maureen, Abigael, Andrew,
Cathy, Christopher, Timothy, Randall, Erin, Fiona

Preface

Linear regression models are widely used today in business administration, economics, engineering, and the social, health, and biological sciences. Successful applications of these models require a sound understanding of both the underlying theory and the practical problems that are encountered in using the models in real-life situations. While *Applied Linear Regression Models,* Third Edition, is basically an applied book, it seeks to blend theory and applications effectively, avoiding the extremes of presenting theory in isolation and of giving elements of applications without the needed understanding of the theoretical foundations.

The third edition differs from the second in a number of important respects.

1. We have reorganized the chapters in order to get to multiple linear regression analysis more quickly. Old Chapter 1 on basic results in probability and statistics has been placed in Appendix A, old Chapter 9 on polynomial regression has been interwoven in the discussion of multiple linear regression, and old Chapter 10 on qualitative predictor variables now comes after a full discussion of multiple regression model building and diagnostics.

2. We have expanded substantially the discussion of diagnostics and remedial measures. In Chapters 3 and 10, we have added robust tests for constancy of the error variance, smoothing techniques to explore the shape of the regression function, robust regression and nonparametric regression techniques, bootstrapping methods for evaluating the precision of sample estimates for complex situations, and estimation of the variance and standard deviation functions to obtain weights for weighted least squares.

3. The chapter on logistic regression (Chapter 14) has been extensively revised and expanded to include an introduction to polytomous logistic regression, Poisson regression, and generalized linear models, as well as greater coverage of diagnostic procedures using the model deviance, deviance residuals, and simulated envelopes.

4. We now include several case studies at strategic places to assist with the integration of the statistical methods discussed.

5. We have given greater emphasis to maximum likelihood methods and have substantially expanded the use of graphic presentation, including the use of scatter

plot matrices, three-dimensional rotating plots, three-dimensional response surface and contour plots, and conditional effects plots.

6. Throughout the text, we have made extensive revisions in the exposition on the basis of classroom experience to improve the clarity of the presentation.

We have included in this book not only the more conventional topics in regression, but also topics that are frequently slighted, though important in practice. We devote three chapters (Chapter 8–10) to the model-building process for regression, including computer-assisted selection procedures for identifying "good" subsets of predictor variables and validation of the chosen regression model. Two chapters (Chapters 11 and 14) are devoted to indicator variables, covering both response and predictor indicator variables. The use of residual analysis and other diagnostics for examining the appropriateness of a regression model is a recurring theme throughout this book. So is the use of remedial measures that may be helpful when the model is not appropriate. In the analysis of the results of a study, we give greater emphasis to the use of estimation procedures than to significance tests, because estimation is often more meaningful in practice. Also, since practical problems seldom are concerned with a single inference, we stress the use of simultaneous inference procedures.

Theoretical ideas are presented to the degree needed for good understanding in making sound applications. Proofs are given in those instances where we feel they serve to demonstrate an important method of approach. Emphasis is placed on a thorough understanding of the regression models, particularly the meaning of the model parameters, since such understanding is basic to proper applications. A wide variety of examples and cases is presented to illustrate the use of the theoretical principles, to show the great diversity of applications of regression models, and to demonstrate how analyses are carried out for different problems.

We use "Notes" and "Comments" sections in each chapter to present additional discussion and matters related to the mainstream of development. In this way, the basic ideas in a chapter are presented concisely and without distraction.

Applications of regression models frequently require extensive computations. We take the position that a computer or programmable calculator is available in most applied work and that almost every computer user has access to program packages for regression analysis. Hence, we explain the basic mathematical steps in fitting a regression model but do not dwell on computational details. This approach permits us to avoid many complex formulas and enables us to focus on basic principles. We make extensive use in this text of computer capabilities for performing computations and preparing graphic plots, and we illustrate a variety of computer printouts and plots and explain how they are used for analysis.

A selection of problems is provided at the end of each chapter. Here readers can reinforce their understanding of the methodology and use the concepts learned to analyze data. We have been careful to supply data-analysis problems that typify genuine applications. In most problems the calculations are best handled on a programmable calculator or computer. To facilitate data entry, a diskette in ASCII format is provided with the text that includes the data sets for all examples, problems, exercises, and projects, as well as for the data sets in Appendix C. The README.TXT file on the diskette provides information about the identification of the data sets.

We assume that the reader of *Applied Linear Regression Models,* Third Edition, has had an introductory course in statistical inference, covering the material outlined in Appendix A. Should some gaps in the reader's background exist, the relevant portions of an introductory text can be studied, or the instructor of the class may use supplemental materials for covering the missing segments. Appendix A is primarily

intended as a reference of basic statistical results for continuing use as the reader progresses through the book.

Calculus is not required for reading *Applied Linear Regression Models,* Third Edition. In a number of instances, we use calculus to demonstrate how some important results are obtained, but these demonstrations are confined to supplementary comments or notes and can be omitted without any loss of continuity. Readers who do know calculus will find these comments and notes in natural sequence so that the benefits of the mathematical developments are obtained in their immediate context. Some basic elements of matrix algebra are needed for multiple regression. Chapter 5 introduces these elements of matrix algebra in the context of simple regression for easy learning.

Applied Linear Regression Models, Third Edition, is intended for use in undergraduate and graduate courses in regression analysis and in second courses in applied statistics. The extent to which material presented in this text is used in a particular course depends upon the amount of time available and the objectives of the course. The basic elements of regression are covered in Chapters 1, 2, 3, 4 (Sections 4.1–4.3), 5, 6, 7, 8, 9, and 10. Chapters 11, 12, 13, 14, and 15 can be covered as time permits and interests dictate.

This book can also be used for self-study by persons engaged in the fields of business administration, economics, engineering, and the social, health, and biological sciences who desire to obtain competence in the application of regression models.

An *Instructor Solutions Manual,* containing detailed solutions to all numerical problems and analytical exercises, is available from the publisher, Irwin, for use by instructors. To facilitate learning by students, instructors may order from the publisher, Irwin, copies of the *Student Solutions Manual* for use by students. The *Student Solutions Manual* provides intermediate and final numerical results for easy self-checking of solutions for selected problems. We use a check mark ($\surd$) in front of the problem number to designate the problems for which the solutions appear in the *Student Solutions Manual.*

A book such as this cannot be written without substantial assistance from numerous persons. We are indebted to the many contributors who have developed the theory and practice discussed in this book. We also would like to acknowledge appreciation to our students, who helped us in a variety of ways to fashion the method of presentation contained herein. We are grateful to the many users of *Applied Linear Statistical Models* and *Applied Linear Regression Models* who have provided us with comments and suggestions based on their teaching with these texts. We are also indebted to Professors James E. Holstein, University of Missouri, and David L. Sherry, University of West Florida, for their review of *Applied Linear Statistical Models,* First Edition; to Professors Samuel Kotz, University of Maryland at College Park, Ralph P. Russo, University of Iowa, and Peter F. Thall, The George Washington University, for their review of *Applied Linear Regression Models,* First Edition; to Professors John S. Y. Chiu, University of Washington, James A. Calvin, University of Iowa, and Michael F. Driscoll, Arizona State University, for their review of *Applied Linear Statistical Models,* Second Edition; to Professor Richard Anderson-Sprecher, University of Wyoming, for his review of *Applied Linear Regression Models,* Second Edition; and to Professors Alexander von Eye, The Pennsylvania State University, Samuel Kotz, University of Maryland at College Park, and John B. Willett, Harvard University, for their review of *Applied Linear Statistical Models,* Third Edition. These reviews provided many important suggestions, for which we are most grateful.

In addition, valuable assistance was provided by Professors Richard K. Burdick, Arizona State University, R. Dennis Cook, University of Minnesota, W. J. Conover,

Texas Tech University, and Mark E. Johnson, University of Central Florida, and by Drs. Richard J. Beckman, Los Alamos National Laboratory, and Ronald L. Iman, Sandia National Laboratories. We are most appreciative of their willing help.

We are also indebted to the 88 participants in a survey concerning *Applied Linear Regression Models,* Second Edition, and the 76 participants in a survey concerning *Applied Linear Statistical Models,* Third Edition. Helpful suggestions were received in these surveys, for which we are thankful.

George Cotsonis, A. Pedro Duarte Silva, and Manus Rungtusanatham assisted us diligently in the checking of the manuscript, in preparing computer-generated plots, and in other ways. The initial typing of the manuscript was ably done by Rebecca Brooks, Linda Keith, and Charlyene Harmon. Rebecca Brooks assisted us in outstanding fashion in the many difficult stages of the revision of a complex manuscript, aided by Sheila King. We are most grateful to all of these persons for their invaluable help and assistance.

Finally, our families bore patiently the pressures caused by our commitment to complete this revision. We are appreciative of their understanding.

John Neter
Michael H. Kutner
Christopher J. Nachtsheim
William Wasserman

Contents

PART I

Simple Linear Regression

1 Linear Regression with One Predictor Variable 3

1.1 Relations between Variables 3
1.2 Regression Models and Their Uses 6
1.3 Simple Linear Regression Model with Distribution of Error Terms Unspecified 10
1.4 Data for Regression Analysis 14
1.5 Overview of Steps in Regression Analysis 15
1.6 Estimation of Regression Function 17
1.7 Estimation of Error Terms Variance σ^2 27
1.8 Normal Error Regression Model 29

2 Inferences in Regression Analysis 44

2.1 Inferences concerning β_1 44
2.2 Inferences concerning β_0 53
2.3 Some Considerations on Making Inferences concerning β_0 and β_1 54
2.4 Interval Estimation of $E\{Y_h\}$ 56
2.5 Prediction of New Observation 61
2.6 Confidence Band for Regression Line 67
2.7 Analysis of Variance Approach to Regression Analysis 69
2.8 General Linear Test Approach 78
2.9 Descriptive Measures of Association between X and Y in Regression Model 80
2.10 Considerations in Applying Regression Analysis 84
2.11 Case when X is Random 85

3 Diagnostics and Remedial Measures 95

3.1 Diagnostics for Predictor Variable 95
3.2 Residuals 97

此为试读,需要完整PDF请访问: www.ertongbook.com

3.3 Diagnostics for Residuals 98
3.4 Overview of Tests Involving Residuals 110
3.5 Correlation Test for Normality 111
3.6 Tests for Constancy of Error Variance 112
3.7 F Test for Lack of Fit 115
3.8 Overview of Remedial Measures 124
3.9 Transformations 126
3.10 Exploration of Shape of Regression Function 135
3.11 Case Example—Plutonium Measurement 138

4 Simultaneous Inferences and Other Topics in Regression Analysis 152

4.1 Joint Estimation of β_0 and β_1 152
4.2 Simultaneous Estimation of Mean Responses 155
4.3 Simultaneous Prediction Intervals for New Observations 158
4.4 Regression through Origin 159
4.5 Effects of Measurement Errors 164
4.6 Inverse Predictions 167
4.7 Choice of X Levels 169

5 Matrix Approach to Simple Linear Regression Analysis 176

5.1 Matrices 176
5.2 Matrix Addition and Subtraction 180
5.3 Matrix Multiplication 182
5.4 Special Types of Matrices 185
5.5 Linear Dependence and Rank of Matrix 188
5.6 Inverse of a Matrix 189
5.7 Some Basic Theorems for Matrices 194
5.8 Random Vectors and Matrices 194
5.9 Simple Linear Regression Model in Matrix Terms 198
5.10 Least Squares Estimation of Regression Parameters 200
5.11 Fitted Values and Residuals 202
5.12 Analysis of Variance Results 205
5.13 Inferences in Regression Analysis 208

PART II

Multiple Linear Regression

6 Multiple Regression—I 217

6.1 Multiple Regression Models 217
6.2 General Linear Regression Model in Matrix Terms 225
6.3 Estimation of Regression Coefficients 227
6.4 Fitted Values and Residuals 227
6.5 Analysis of Variance Results 228
6.6 Inferences about Regression Parameters 231
6.7 Estimation of Mean Response and Prediction of New Observation 233
6.8 Diagnostics and Remedial Measures 236

6.9 An Example—Multiple Regression with Two Predictor Variables 241

7 Multiple Regression—II 260

7.1 Extra Sums of Squares 260
7.2 Uses of Extra Sums of Squares in Tests for Regression Coefficients 268
7.3 Summary of Tests concerning Regression Coefficients 271
7.4 Coefficients of Partial Determination 274
7.5 Standardized Multiple Regression Model 277
7.6 Multicollinearity and Its Effects 285
7.7 Polynomial Regression Models 296
7.8 Interaction Regression Models 308
7.9 Constrained Regression 315

8 Building the Regression Model I: Selection of Predictor Variables 327

8.1 Overview of Model-Building Process 327
8.2 Surgical Unit Example 334
8.3 All-Possible-Regressions Procedure for Variables Reduction 336
8.4 Forward Stepwise Regression and Other Automatic Search Procedures for Variables Reduction 347
8.5 Some Final Comments on Model Building for Exploratory Observational Studies 353

9 Building the Regression Model II: Diagnostics 361

9.1 Model Adequacy for a Predictor Variable—Partial Regression Plots 361
9.2 Identifying Outlying Y Observations—Studentized Deleted Residuals 368
9.3 Identifying Outlying X Observations—Hat Matrix Leverage Values 375
9.4 Identifying Influential Cases—*DFFITS,* Cook's Distance, and *DFBETAS* Measures 378
9.5 Multicollinearity Diagnostics—Variance Inflation Factor 385
9.6 Surgical Unit Example—Continued 388

10 Building the Regression Model III: Remedial Measures and Validation 400

10.1 Unequal Error Variances Remedial Measures—Weighted Least Squares 400
10.2 Multicollinearity Remedial Measures—Ridge Regression 410
10.3 Remedial Measures for Influential Cases—Robust Regression 416
10.4 Remedial Measures for Unknown Response Function—Nonparametric Regression 425
10.5 Remedial Measures for Evaluating Precision in Nonstandard Situations—Bootstrapping 429
10.6 Model Validation 434
10.7 Case Example—Mathematics Proficiency 439

11 Qualitative Predictor Variables 455

11.1 One Qualitative Predictor Variable 455
11.2 Model Containing Interaction Effects 461
11.3 More Complex Models 464
11.4 Comparison of Two or More Regression Functions 468
11.5 Other Uses of Indicator Variables 474
11.6 Some Considerations in Using Indicator Variables 480
11.7 Case Example—MNDOT Traffic Estimation 483

12 Autocorrelation in Time Series Data 497

12.1 Problems of Autocorrelation 497
12.2 First-Order Autoregressive Error Model 501
12.3 Durbin-Watson Test for Autocorrelation 504
12.4 Remedial Measures for Autocorrelation 507
12.5 Forecasting with Autocorrelated Error Terms 517

PART III

Nonlinear Regression

13 Introduction to Nonlinear Regression 531

13.1 Linear and Nonlinear Regression Models 531
13.2 Example 535
13.3 Least Squares Estimation in Nonlinear Regression 536
13.4 Model Building and Diagnostics 547
13.5 Inferences about Nonlinear Regression Parameters 548
13.6 Learning Curve Example 555

14 Logistic Regression, Poisson Regression, and Generalized Linear Models 567

14.1 Regression Models with Binary Response Variable 567
14.2 Simple Logistic Response Function 570
14.3 Simple Logistic Regression 573
14.4 Multiple Logistic Regression 580
14.5 Model Building: Selection of Predictor Variables 585
14.6 Diagnostics 590
14.7 Inferences about Logistic Regression Parameters 599
14.8 Inferences about Mean Response 602
14.9 Prediction of a New Observation 605
14.10 Polytomous Logistic Regression 608
14.11 Poisson Regression 609
14.12 Generalized Linear Models 614

PART IV

Correlation Analysis

15 Normal Correlation Models 631

15.1 Distinction between Regression and Correlation Models 631

15.2 Bivariate Normal Distribution 632
15.3 Conditional Inferences 636
15.4 Inferences on Correlation Coefficients 640
15.5 Multivariate Normal Distribution 645
15.6 Spearman Rank Correlation Coefficient 651

Appendix A Some Basic Results in Probability and Statistics 663
Appendix B Tables 685
Appendix C Data Sets 702
Appendix D Selected Bibliography 708
Index 715

PART I

Simple Linear Regression

此为试读,需要完整PDF请访问: www.ertongbook.com