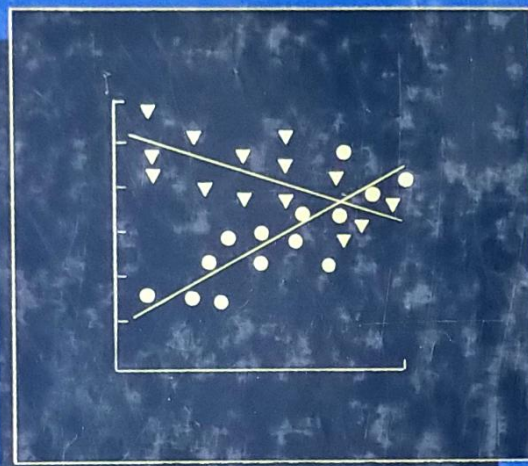


# APPLIED LINEAR STATISTICAL MODELS

F O U R T H      E D I T I O N



NETER   KUTNER   NACHTSHEIM   WASSERMAN



FOURTH  
EDITION

# Applied Linear Statistical Models



**John Neter**

*University of Georgia*

**Michael H. Kutner**

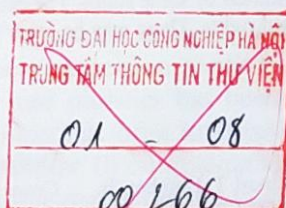
*The Cleveland Clinic Foundation*

**Christopher J. Nachtsheim**

*University of Minnesota*

**William Wasserman**

*Syracuse University*



GIFT OF THE ASIA FOUNDATION  
NOT FOR RE-SALE

QUÀ TẶNG CỦA QUỸ CHÂU Á  
KHÔNG ĐƯỢC BÁN LẠI

**McGraw-Hill WCB  
McGraw-Hill**

Boston, Massachusetts Burr Ridge, Illinois Dubuque, Iowa  
Madison, Wisconsin New York, New York San Francisco, California St. Louis, Missouri

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



**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.

## WCB/McGraw-Hill

*A Division of The McGraw-Hill Companies*

© The McGraw-Hill Companies, Inc., 1974, 1983, 1990, 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 statistical models / John Neter . . . [et al.]. — 3rd ed.

p. cm.

Includes index.

ISBN 0-256-11736-5

1. Regression analysis. 2. Analysis of variance. 3. Experimental design. 4. Linear models (Statistics)

QA278.2.A66 1996

519.5'35—dc20

95-37447

*Printed in the United States of America*

12 13 14 15 16 17 DOW /DOW 0 9 8 7 6 5 4 3



# Preface

Linear statistical models for regression, analysis of variance, and experimental designs 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 Statistical Models*, Fourth 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 fourth edition differs from the third in a number of important respects.

1. In the area of regression analysis, 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.

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.

2. We have retained the three chapters on nonlinear regression, logistic regression, and correlation analysis that were dropped from the third edition. 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.

3. In the area of analysis of variance and experimental designs, we have added new Chapters 31 and 32 on fractional factorial designs and response surface method-



# 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  $\sigma^2$  27
- 1.8 Normal Error Regression Model 29

#### 2 Inferences in Regression Analysis 44

- 2.1 Inferences concerning  $\beta_1$  44
- 2.2 Inferences concerning  $\beta_0$  53
- 2.3 Some Considerations on Making Inferences concerning  $\beta_0$  and  $\beta_1$  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
- 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  $\beta_0$  and  $\beta_1$  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—*DFBETAS*, 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

### PART V

## Analysis of Variance: I

### 16 Single-Factor ANOVA Model and Tests 663

- 16.1 Relation between Regression and Analysis of Variance 663
- 16.2 Experimental and Observational Studies, Factors, and Treatments 666
- 16.3 Design of Analysis of Variance Studies 668
- 16.4 Uses of Analysis of Variance Models 670
- 16.5 Use of Computers 671
- 16.6 ANOVA Model I—Fixed Factor Levels 671
- 16.7 Fitting of ANOVA Model 676
- 16.8 Analysis of Variance 681
- 16.9  $F$  Test for Equality of Factor Level Means 690
- 16.10 Alternative Formulation of Model I 693
- 16.11 Regression Approach to Single-Factor Analysis of Variance 696

### 17 Analysis of Factor Level Effects 710

- 17.1 Introduction 710
- 17.2 Plots of Estimated Factor Level Means 712
- 17.3 Estimation and Testing of Factor Level Effects 716
- 17.4 Need for Simultaneous Inference Procedures 724
- 17.5 Tukey Multiple Comparison Procedure 725
- 17.6 Scheffé Multiple Comparison Procedure 732
- 17.7 Bonferroni Multiple Comparison Procedure 736



- 17.8 Holm Simultaneous Testing Procedure 739
- 17.9 Analysis of Factor Effects when Factor Quantitative 742

## **18 ANOVA Diagnostics and Remedial Measures 756**

- 18.1 Residual Analysis 756
- 18.2 Tests for Constancy of Error Variance 763
- 18.3 Overview of Remedial Measures 768
- 18.4 Weighted Least Squares 768
- 18.5 Transformations of Response Variable 772
- 18.6 Effects of Departures from Model 776
- 18.7 Nonparametric Rank  $F$  Test 777
- 18.8 Case Example—Heart Transplant 780

### **PART VI**

## **Analysis of Variance: II**

### **19 Two-Factor Analysis of Variance—Equal Sample Sizes 795**

- 19.1 Multifactor Studies 795
- 19.2 Meaning of ANOVA Model Elements 799
- 19.3 Model I (Fixed Factor Levels) for Two-Factor Studies 812
- 19.4 Analysis of Variance 817
- 19.5 Evaluation of Appropriateness of ANOVA Model 825
- 19.6  $F$  Tests 829
- 19.7 Regression Approach to Two-Factor Analysis of Variance 832
- 19.8 Pooling Sums of Squares in Two-Factor Analysis of Variance 837

### **20 Analysis of Factor Effects in Two-Factor Studies—Equal Sample Sizes 849**

- 20.1 Strategy for Analysis 849
- 20.2 Analysis of Factor Effects when Factors Do Not Interact 851
- 20.3 Analysis of Factor Effects when Interactions Important 859
- 20.4 Analysis when One or Both Factors Quantitative 865

### **21 Two-Factor Studies—One Case Per Treatment 875**

- 21.1 No-Interaction Model 875
- 21.2 Tukey Test for Additivity 882

### **22 Two-Factor Studies—Unequal Sample Sizes and Unequal Treatment Importance 889**

- 22.1 Unequal Sample Sizes 889
- 22.2 Use of Regression Approach for Testing Factor Effects when Sample Sizes are Unequal 891
- 22.3 Inferences about Factor Effects when Sample Sizes Are Unequal 897
- 22.4 Empty Cells in Two-Factor Studies 902
- 22.5 ANOVA Inferences when Treatment Means Are of Unequal Importance 905
- 22.6 Statistical Computing Packages 915

### **23 Multifactor Studies 924**

- 23.1 ANOVA Model I for Three-Factor Studies 924
- 23.2 Analysis of Variance 932
- 23.3 Evaluation of Appropriateness of ANOVA Model 938
- 23.4 Analysis of Factor Effects 938
- 23.5 Example of Three-Factor Study 942
- 23.6 Unequal Sample Sizes in Multifactor Studies 948

### **24 Random and Mixed Effects Models 958**

- 24.1 Single-Factor Studies—ANOVA Model II 958
- 24.2 Two-Factor Studies—ANOVA Models II and III 976
- 24.3 Two-Factor Studies—ANOVA Tests for Models II and III 981
- 24.4 Two-Factor Studies—Estimation of Factor Effects For Models II and III 984
- 24.5 Three-Factor-Studies—ANOVA Models II and III 989
- 24.6 ANOVA Models II and III with Unequal Sample Sizes 994

### **25 Analysis of Covariance 1010**

- 25.1 Basic Ideas 1010
- 25.2 Single-Factor Covariance Model 1014
- 25.3 Example of Single-Factor Covariance Analysis 1019
- 25.4 Multifactor Covariance Analysis 1028
- 25.5 Additional Considerations for the Use of Covariance Analysis 1032



## PART VII

**Study Designs****26 Design of Experiments, Randomization, and Sample Size Planning 1045**

- 26.1 Design of Experiments 1045
- 26.2 Contributions of Statistics to Experimentation 1046
- 26.3 Randomization Tests 1050
- 26.4 Planning of Sample Sizes with Power Approach 1052
- 26.5 Planning of Sample Sizes with Estimation Approach 1060
- 26.6 Planning of Sample Sizes to Find "Best" Treatment 1063

**27 Randomized Block Designs 1072**

- 27.1 Elements of Randomized Block Designs 1072
- 27.2 Model for Randomized Block Designs 1075
- 27.3 Analysis of Variance and Tests 1078
- 27.4 Evaluation of Appropriateness of Randomized Block Model 1081
- 27.5 Analysis of Treatment Effects 1084
- 27.6 Factorial Treatments 1086
- 27.7 Planning Randomized Block Experiments 1088
- 27.8 Regression Approach to Randomized Block Designs 1091
- 27.9 Covariance Analysis for Randomized Block Designs 1092
- 27.10 Nonparametric Rank  $F$  Test 1094
- 27.11 Missing Observations 1097
- 27.12 Random Block Effects 1100
- 27.13 Generalized Randomized Block Designs 1106
- 27.14 Use of More Than One Blocking Variable 1108

**28 Nested Designs, Subsampling, and Partially Nested Designs 1121**

- 28.1 Distinction between Nested and Crossed Factors 1121
- 28.2 Two-Factor Nested Designs 1124
- 28.3 Analysis of Variance for Two-Factor Nested Designs 1127
- 28.4 Evaluation of Appropriateness of Nested Design Model 1133
- 28.5 Analysis of Factor Effects in Two-Factor Nested Designs 1134
- 28.6 Unbalanced Nested Two-Factor Designs 1138

- 28.7 Subsampling in Single-Factor Study with Completely Randomized Design 1141
- 28.8 Pure Subsampling in Three Stages 1148
- 28.9 Three-Factor Partially Nested Designs 1149

**29 Repeated Measures and Related Designs 1164**

- 29.1 Elements of Repeated Measures Designs 1164
- 29.2 Single-Factor Experiments with Repeated Measures on All Treatments 1166
- 29.3 Two-Factor Experiments with Repeated Measures on Both Factors 1176
- 29.4 Two-Factor Experiments with Repeated Measures on One Factor 1184
- 29.5 Regression Approach to Repeated Measures Designs 1194
- 29.6 Split-Plot Designs 1194

**30 Latin Square and Related Designs 1207**

- 30.1 Basic Elements 1207
- 30.2 Latin Square Model 1211
- 30.3 Analysis of Latin Square Experiments 1212
- 30.4 Planning Latin Square Experiments 1218
- 30.5 Regression Approach to Latin Square Designs 1220
- 30.6 Additional Replications with Latin Square Designs 1221
- 30.7 Replications in Repeated Measures Designs 1225

**31 Exploratory Experiments—Two-Level Factorial and Fractional Factorial Designs 1234**

- 31.1 Two-Level Full Factorial Designs 1235
- 31.2 Analysis of Unreplicated Two-Level Studies 1241
- 31.3 Two-Level Fractional Factorial Designs 1249
- 31.4 Screening Experiments 1264
- 31.5 Incomplete Block Designs for Two-Level Factorial Experiments 1266

**32 Response Surface Methodology 1280**

- 32.1 Response Surface Experiments 1280
- 32.2 Central Composite Response Surface Designs 1281
- 32.3 Optimal Response Surface Designs 1289
- 32.4 Analysis of Response Surface Experiments 1297



32.5	Sequential Search for Optimum Conditions—Method of Steepest Ascent	1303
------	--	------

Appendix A	Some Basic Results in Probability and Statistics	1313
Appendix B	Tables	1333
Appendix C	Data Sets	1365
Appendix D	Rules for Developing ANOVA Models and Tables for Balanced Designs	1373
Appendix E	Selected Bibliography	1391
Index		1400