

## Provides an easy-to-understand guide to statistical linear models and its uses in data analysis

This book defines a broad spectrum of statistical linear models that is useful in the analysis of data. Considerable rewriting was done to make the book more reader friendly than the first edition. *Linear Models, Second Edition* is written in such a way as to be self-contained for a person with a background in basic statistics, calculus and linear algebra. The text includes numerous applied illustrations, numerical examples, and exercises, now augmented with computer outputs in SAS and R. Also new to this edition is:

- A greatly improved internal design and format
- A short introductory chapter to ease understanding of the order in which topics are taken up
- Discussion of additional topics including multiple comparisons and shrinkage estimators
- Enhanced discussions of generalized inverses, the MINQUE, Bayes and Maximum Likelihood estimators for estimating variance components

Furthermore, in this edition, the second author adds many pedagogical elements throughout the book. These include numbered examples, end-of-example and end-of-proof symbols, selected hints and solutions to exercises available on the book's website, and references to "big data" in everyday life. Featuring a thorough update, *Linear Models, Second Edition* includes:

- A new internal format, additional instructional pedagogy, selected hints and solutions to exercises, and several more real-life applications
- Many examples using SAS and R with timely data sets
- Over 400 examples and exercises throughout the book to reinforce understanding

*Linear Models, Second Edition* is a textbook and a reference for upper-level undergraduate and beginning graduate-level courses on linear models, statisticians, engineers, and scientists who use multiple regression or analysis of variance in their work.

**The late SHAYLE R. SEARLE, PhD**, was Professor Emeritus of Biometry at Cornell University. He was the author of the first edition of *Linear Models*, *Linear Models for Unbalanced Data*, and *Generalized, Linear, and Mixed Models* (with Charles E. McCulloch), all from Wiley. The first edition of *Linear Models* appears in the Wiley Classics Library.

**MARVIN H. J. GRUBER, PhD**, is Professor Emeritus at Rochester Institute of Technology, School of Mathematical Sciences. Dr. Gruber has written a number of papers and has given numerous presentations at professional meetings during his tenure as a professor at RIT. His fields of interest include regression estimators and the improvement of their efficiency using shrinkage estimators. He has written and published two books on this topic. Another of his books, *Matrix Algebra for Linear Models*, also published by Wiley, provides good preparation for studying Linear Models. He is a member of the American Mathematical Society, the Institute of Mathematical Statistics and the American Statistical Association.



Subscribe to our free Statistics eNewsletter at [wiley.com/enewsletters](http://wiley.com/enewsletters)  
[www.wiley.com/statistics](http://www.wiley.com/statistics)

# WILEY



Also available  
as an e-book

Cover image:  
GettyImages/Caspar Benson

ISBN 978-1-118-95283-2



9 781118 952832



<b>Preface</b>	<b>xvii</b>
<b>Preface to First Edition</b>	<b>xxi</b>
<b>About the Companion Website</b>	<b>xxv</b>
<b>Introduction and Overview</b>	<b>1</b>
<b>1. Generalized Inverse Matrices</b>	<b>7</b>
1. Introduction, 7	
a. Definition and Existence of a Generalized Inverse, 8	
b. An Algorithm for Obtaining a Generalized Inverse, 11	
c. Obtaining Generalized Inverses Using the Singular Value Decomposition (SVD), 14	
2. Solving Linear Equations, 17	
a. Consistent Equations, 17	
b. Obtaining Solutions, 18	
c. Properties of Solutions, 20	
3. The Penrose Inverse, 26	
4. Other Definitions, 30	
5. Symmetric Matrices, 32	
a. Properties of a Generalized Inverse, 32	
b. Two More Generalized Inverses of $\mathbf{X}'\mathbf{X}$ , 35	
6. Arbitrariness in a Generalized Inverse, 37	
7. Other Results, 42	
8. Exercises, 44	



## 2. Distributions and Quadratic Forms

1. Introduction, 49
2. Symmetric Matrices, 52
3. Positive Definiteness, 53
4. Distributions, 58
  - a. Multivariate Density Functions, 58
  - b. Moments, 59
  - c. Linear Transformations, 60
  - d. Moment and Cumulative Generating Functions, 62
  - e. Univariate Normal, 64
  - f. Multivariate Normal, 64
    - (i) Density Function, 64
    - (ii) Aitken's Integral, 64
    - (iii) Moment Generating Function, 65
    - (iv) Marginal Distributions, 66
    - (v) Conditional Distributions, 67
    - (vi) Independence of Normal Random Variables, 68
  - g. Central  $\chi^2$ ,  $F$ , and  $t$ , 69
  - h. Non-central  $\chi^2$ , 71
  - i. Non-central  $F$ , 73
  - j. The Non-central  $t$  Distribution, 73
5. Distribution of Quadratic Forms, 74
  - a. Cumulants, 75
  - b. Distributions, 78
  - c. Independence, 80
6. Bilinear Forms, 87
7. Exercises, 89

## 3. Regression for the Full-Rank Model

1. Introduction, 95
  - a. The Model, 95
  - b. Observations, 97
  - c. Estimation, 98
  - d. The General Case of  $k \times$  Variables, 100
  - e. Intercept and No-Intercept Models, 104
2. Deviations From Means, 105
3. Some Methods of Estimation, 109
  - a. Ordinary Least Squares, 109
  - b. Generalized Least Squares, 109
  - c. Maximum Likelihood, 110
  - d. The Best Linear Unbiased Estimator (b.l.u.e.)(Gauss–Markov Theorem), 110
  - e. Least-squares Theory When The Parameters are Random Variables, 112



4. Consequences of Estimation, 115
  - a. Unbiasedness, 115
  - b. Variances, 115
  - c. Estimating  $E(y)$ , 116
  - d. Residual Error Sum of Squares, 119
  - e. Estimating the Residual Error Variance, 120
  - f. Partitioning the Total Sum of Squares, 121
  - g. Multiple Correlation, 122
5. Distributional Properties, 126
  - a. The Vector of Observations  $y$  is Normal, 126
  - b. The Least-square Estimator  $\hat{\mathbf{b}}$  is Normal, 127
  - c. The Least-square Estimator  $\hat{\mathbf{b}}$  and the Estimator of the Variance  $\hat{\sigma}^2$  are Independent, 127
  - d. The Distribution of  $SSE/\sigma^2$  is a  $\chi^2$  Distribution, 128
  - e. Non-central  $\chi^2$ 's, 128
  - f.  $F$ -distributions, 129
  - g. Analyses of Variance, 129
  - h. Tests of Hypotheses, 131
  - i. Confidence Intervals, 133
  - j. More Examples, 136
  - k. Pure Error, 139
6. The General Linear Hypothesis, 141
  - a. Testing Linear Hypothesis, 141
  - b. Estimation Under the Null Hypothesis, 143
  - c. Four Common Hypotheses, 145
  - d. Reduced Models, 148
    - (i) The Hypothesis  $\mathbf{K}'\mathbf{b} = \mathbf{m}$ , 148
    - (ii) The Hypothesis  $\mathbf{K}'\mathbf{b} = \mathbf{0}$ , 150
    - (iii) The Hypothesis  $b_q = 0$ , 152
  - e. Stochastic Constraints, 158
  - f. Exact Quadratic Constraints (Ridge Regression), 160
7. Related Topics, 162
  - a. The Likelihood Ratio Test, 163
  - b. Type I and Type II Errors, 164
  - c. The Power of a Test, 165
  - d. Estimating Residuals, 166
8. Summary of Regression Calculations, 168
9. Exercises, 169

#### 4. Introducing Linear Models: Regression on Dummy Variables

1. Regression on Allocated Codes, 175
  - a. Allocated Codes, 175
  - b. Difficulties and Criticism, 176
  - c. Grouped Variables, 177
  - d. Unbalanced Data, 178



2. Regression on Dummy (0, 1) Variables, 180
  - a. Factors and Levels, 180
  - b. The Regression, 181
3. Describing Linear Models, 184
  - a. A One-Way Classification, 184
  - b. A Two-Way Classification, 186
  - c. A Three-Way Classification, 188
  - d. Main Effects and Interactions, 188
    - (i) Main Effects, 188
    - (ii) Interactions, 190
  - e. Nested and Crossed Classifications, 194
4. The Normal Equations, 198
5. Exercises, 201

## 5. Models Not of Full Rank

205

1. The Normal Equations, 205
  - a. The Normal Equations, 206
  - b. Solutions to the Normal Equations, 209
2. Consequences of a Solution, 210
  - a. Expected Value of  $\mathbf{b}^\circ$ , 210
  - b. Variance Covariance Matrices of  $\mathbf{b}^\circ$  (Variance Covariance Matrices), 211
  - c. Estimating  $E(y)$ , 212
  - d. Residual Error Sum of Squares, 212
  - e. Estimating the Residual Error Variance, 213
  - f. Partitioning the Total Sum of Squares, 214
  - g. Coefficient of Determination, 215
3. Distributional Properties, 217
  - a. The Observation Vector  $\mathbf{y}$  is Normal, 217
  - b. The Solution to the Normal Equations  $\mathbf{b}^\circ$  is Normally Distributed, 217
  - c. The Solution to the Normal Equations  $\mathbf{b}^\circ$  and the Estimator of the Residual Error Variance  $\hat{\sigma}^2$  are Independent, 217
  - d. The Error Sum of Squares Divided by the Population Variance  $SSE/\sigma^2$  is Chi-square  $\chi^2$ , 217
  - e. Non-central  $\chi^2$ 's, 218
  - f. Non-central  $F$ -distributions, 219
  - g. Analyses of Variance, 220
  - h. Tests of Hypotheses, 221
4. Estimable Functions, 223
  - a. Definition, 223
  - b. Properties of Estimable Functions, 224
    - (i) The Expected Value of Any Observation is Estimable, 224
    - (ii) Linear Combinations of Estimable Functions are Estimable, 224



- (iii) The Forms of an Estimable Function, 225
- (iv) Invariance to the Solution  $\mathbf{b}^\circ$ , 225
- (v) The Best Linear Unbiased Estimator (b.l.u.e.)  
Gauss–Markov Theorem, 225

- c. Confidence Intervals, 227
- d. What Functions Are Estimable?, 228
- e. Linearly Independent Estimable Functions, 229
- f. Testing for Estimability, 229
- g. General Expressions, 233
- 5. The General Linear Hypothesis, 236
  - a. Testable Hypotheses, 236
  - b. Testing Testable Hypothesis, 237
  - c. The Hypothesis  $\mathbf{K}'\mathbf{b} = \mathbf{0}$ , 240
  - d. Non-testable Hypothesis, 241
  - e. Checking for Testability, 243
  - f. Some Examples of Testing Hypothesis, 245
  - g. Independent and Orthogonal Contrasts, 248
  - h. Examples of Orthogonal Contrasts, 250
- 6. Restricted Models, 255
  - a. Restrictions Involving Estimable Functions, 257
  - b. Restrictions Involving Non-estimable Functions, 259
  - c. Stochastic Constraints, 260
- 7. The “Usual Constraints”, 264
  - a. Limitations on Constraints, 266
  - b. Constraints of the Form  $b_i^\circ = 0$ , 266
  - c. Procedure for Deriving  $\mathbf{b}^\circ$  and  $\mathbf{G}$ , 269
  - d. Restrictions on the Model, 270
  - e. Illustrative Examples of Results in Subsections a–d, 272
- 8. Generalizations, 276
  - a. Non-singular  $\mathbf{V}$ , 277
  - b. Singular  $\mathbf{V}$ , 277
- 9. An Example, 280
- 10. Summary, 283
- 11. Exercises, 283

## 6. Two Elementary Models

- 1. Summary of the General Results, 288
- 2. The One-Way Classification, 291
  - a. The Model, 291
  - b. The Normal Equations, 294
  - c. Solving the Normal Equations, 294
  - d. Analysis of Variance, 296
  - e. Estimable Functions, 299
  - f. Tests of Linear Hypotheses, 304
    - (i) General Hypotheses, 304



- (ii) The Test Based on  $F(M)$ , 305
  - (iii) The Test Based on  $F(R_m)$ , 307
- g. Independent and Orthogonal Contrasts, 308
- h. Models that Include Restrictions, 310
- i. Balanced Data, 312
- 3. Reductions in Sums of Squares, 313
  - a. The  $R(\ )$  Notation, 313
  - b. Analyses of Variance, 314
  - c. Tests of Hypotheses, 315
- 4. Multiple Comparisons, 316
- 5. Robustness of Analysis of Variance to Assumptions, 321
  - a. Non-normality of the Error, 321
  - b. Unequal Variances, 325
    - (i) Bartlett's Test, 326
    - (ii) Levene's Test, 327
    - (iii) Welch's (1951)  $F$ -test, 328
    - (iv) Brown-Forsyth (1974b) Test, 329
  - c. Non-independent Observations, 330
- 6. The Two-Way Nested Classification, 331
  - a. Model, 332
  - b. Normal Equations, 332
  - c. Solving the Normal Equations, 333
  - d. Analysis of Variance, 334
  - e. Estimable Functions, 336
  - f. Tests of Hypothesis, 337
  - g. Models that Include Restrictions, 339
  - h. Balanced Data, 339
- 7. Normal Equations for Design Models, 340
- 8. A Few Computer Outputs, 341
- 9. Exercises, 343

## 7. The Two-Way Crossed Classification

- 1. The Two-Way Classification Without Interaction, 347
  - a. Model, 348
  - b. Normal Equations, 349
  - c. Solving the Normal Equations, 350
  - d. Absorbing Equations, 352
  - e. Analyses of Variance, 356
    - (i) Basic Calculations, 356
    - (ii) Fitting the Model, 357
    - (iii) Fitting Rows Before Columns, 357
    - (iv) Fitting Columns Before Rows, 359
    - (v) Ignoring and/or Adjusting for Effects, 362
    - (vi) Interpretation of Results, 363



- f. Estimable Functions, 368
- g. Tests of Hypothesis, 370
- h. Models that Include Restrictions, 373
- i. Balanced Data, 374
- 2. The Two-Way Classification with Interaction, 380
  - a. Model, 381
  - b. Normal Equations, 383
  - c. Solving the Normal Equations, 384
  - d. Analysis of Variance, 385
    - (i) Basic Calculations, 385
    - (ii) Fitting Different Models, 389
    - (iii) Computational Alternatives, 395
    - (iv) Interpretation of Results, 397
    - (v) Fitting Main Effects Before Interaction, 397
  - e. Estimable Functions, 398
  - f. Tests of Hypotheses, 403
    - (i) The General Hypothesis, 403
    - (ii) The Hypothesis for  $F(M)$ , 404
    - (iii) Hypotheses for  $F(\alpha|\mu)$  and  $F(\beta|\mu)$ , 405
    - (iv) Hypotheses for  $F(\alpha|\mu, \beta)$  and  $F(\beta|\mu, \alpha)$ , 407
    - (v) Hypotheses for  $F(\gamma|\mu, \alpha, \beta)$ , 410
    - (vi) Reduction to the No-Interaction Model, 412
    - (vii) Independence Properties, 413
  - g. Models that Include Restrictions, 413
  - h. All Cells Filled, 414
  - i. Balanced Data, 415
- 3. Interpretation of Hypotheses, 420
- 4. Connectedness, 422
- 5. The  $\mu_{ij}$  Models, 427
- 6. Exercises, 429

## 8. Some Other Analyses

- 1. Large-Scale Survey-Type Data, 437
  - a. Example, 438
  - b. Fitting a Linear Model, 438
  - c. Main-Effects-Only Models, 440
  - d. Stepwise Fitting, 442
  - e. Connectedness, 442
  - f. The  $\mu_{ij}$ -models, 443
- 2. Covariance, 445
  - a. A General Formulation, 446
    - (i) The Model, 446
    - (ii) Solving the Normal Equations, 446
    - (iii) Estimability, 447



- (iv) A Model for Handling the Covariates, 447
  - (v) Analyses of Variance, 448
  - (vi) Tests of Hypotheses, 451
  - (vii) Summary, 453
- b. The One-Way Classification, 454
  - (i) A Single Regression, 454
  - (ii) Example, 459
  - (iii) The Intra-Class Regression Model, 464
  - (iv) Continuation of Example 1, 467
  - (v) Another Example, 470
- c. The Two-Way Classification (With Interaction), 470
- 3. Data Having All Cells Filled, 474
  - a. Estimating Missing Observations, 475
  - b. Setting Data Aside, 478
  - c. Analysis of Means, 479
    - (i) Unweighted Means Analysis, 479
    - (ii) Example, 482
    - (iii) Weighted Squares of Means, 484
    - (iv) Continuation of Example, 485
  - d. Separate Analyses, 487
- 4. Exercises, 487

## 9. Introduction to Variance Components

- 1. Fixed and Random Models, 493
  - a. A Fixed-Effects Model, 494
  - b. A Random-Effects Model, 494
  - c. Other Examples, 496
    - (i) Of Treatments and Varieties, 496
    - (ii) Of Mice and Men, 496
    - (iii) Of Cows and Bulls, 497
- 2. Mixed Models, 497
  - (i) Of Mice and Diets, 497
  - (ii) Of Treatments and Crosses, 498
  - (iii) On Measuring Shell Velocities, 498
  - (iv) Of Hospitals and Patients, 498
- 3. Fixed or Random, 499
- 4. Finite Populations, 500
- 5. Introduction to Estimation, 500
  - a. Variance Matrix Structures, 501
  - b. Analyses of Variance, 502
  - c. Estimation, 504
- 6. Rules for Balanced Data, 507
  - a. Establishing Analysis of Variance Tables, 507
    - (i) Factors and Levels, 507
    - (ii) Lines in the Analysis of Variance Table, 507
    - (iii) Interactions, 508



- b. Mixed Models, 565
  - c. Random-Effects Models, 566
  - d. Applications, 566
2. Analysis of Variance Method (Henderson's Method 1), 567
  - a. Model and Notation, 567
  - b. Analogous Sums of Squares, 568
    - (i) Empty Cells, 568
    - (ii) Balanced Data, 568
    - (iii) A Negative "Sum of Squares", 568
    - (iv) Uncorrected Sums of Squares, 569
  - c. Expectations, 569
    - (i) An Example of a Derivation of the Expectation of a Sum of Squares, 570
    - (ii) Mixed Models, 573
    - (iii) General Results, 574
    - (iv) Calculation by "Synthesis", 576
  - d. Sampling Variances of Estimators, 577
    - (i) Derivation, 578
    - (ii) Estimation, 581
    - (iii) Calculation by Synthesis, 585
3. Adjusting for Bias in Mixed Models, 588
  - a. General Method, 588
  - b. A Simplification, 588
  - c. A Special Case: Henderson's Method 2, 589
4. Fitting Constants Method (Henderson's Method 3), 590
  - a. General Properties, 590
  - b. The Two-Way Classification, 592
    - (i) Expected Values, 593
    - (ii) Estimation, 594
    - (iii) Calculation, 594
  - c. Too Many Equations, 595
  - d. Mixed Models, 597
  - e. Sampling Variances of Estimators, 597
5. Analysis of Means Methods, 598
6. Symmetric Sums Methods, 599
7. Infinitely Many Quadratics, 602
8. Maximum Likelihood for Mixed Models, 605
  - a. Estimating Fixed Effects, 606
  - b. Fixed Effects and Variance Components, 611
  - c. Large Sample Variances, 613
9. Mixed Models Having One Random Factor, 614
10. Best Quadratic Unbiased Estimation, 620
  - a. The Method of Townsend and Searle (1971) for a Zero Mean, 620
  - b. The Method of Swallow and Searle (1978) for a Non-Zero Mean, 622



- 11. Shrinkage Estimation of Regression Parameters and Variance Components, 626
  - a. Shrinkage Estimators, 626
  - b. The James–Stein Estimator, 627
  - c. Stein’s Estimator of the Variance, 627
  - d. A Shrinkage Estimator of Variance Components, 628
- 12. Exercises, 630

<b>References</b>	<b>633</b>
<b>Author Index</b>	<b>645</b>
<b>Subject Index</b>	<b>649</b>