Regression Analysis and Linear Models

Concepts, Applications, and Implementation

Richard B. Darlington
Andrew F. Hayes

Series Editor's Note by Todd D. Little

THE GUILFORD PRESS
New York London

Contents

List of Symbols and Abbreviations xvii

1 • Statistical Control and Linear Models 1

1.1 Statistical Control / 1
1.1.1 The Need for Control /1
1.1.2 *Five Methods of Control /2*
1.1.3 *Examples of Statistical Control / 4*
1.2 An Overview of Linear Models / 8
1.2.1 *What You Should Know Already /12*
1.2.2 *Statistical Software for Linear Modeling and Statistical Control /12*
1.2.3 *About Formulas /14*
1.2.4 *On Symbolic Representations /15*
1.3 Chapter Summary / 16

2 • The Simple Regression Model 17

2.1 Scatterplots and Conditional Distributions / 17
2.1.1 *Scatterplots /17*
2.1.2 *A Line through Conditional Means /18*
2.1.3 *Errors of Estimate / 21*
2.2 The Simple Regression Model / 23
2.2.1 *The Regression Line* / 23
2.2.2 *Variance, Covariance, and Correlation* / 24
2.2.3 *Finding the Regression Line* / 25
2.2.4 *Example Computations / 26*
2.2.5 *Linear Regression Analysis by Computer / 28*
2.3 The Regression Coefficient versus the Correlation Coefficient / 31
2.3.1 *Properties of the Regression and Correlation Coefficients / 32*
2.3.2 *Uses of the Regression and Correlation Coefficients / 34*
2.4 Residuals / 35
2.4.1 *The Three Components of Y / 35*
2.4.2 *Algebraic Properties of Residuals* / 36
2.4.3 *Residuals as Y Adjusted for Differences in X / 37*
2.4.4 *Residual Analysis 137*
2.5 Chapter Summary / 41

3 • Partial Relationship and the Multiple Regression Model 43

3.1 Regression Analysis with More Than One Predictor Variable / 43
3.1.1 *An Example/43*
3.1.2 *Regressors / 46*

3.1.3 *Models/47*
3.1.4 *Representing a Model Geometrically 149*
3.1.5 *Model Errors / 50*
3.1.6 *An Alternative View of the Model* / 52
3.2 The Best-Fitting Model / 55
3.2.1 *Model Estimation with Computer Software / 55*
3.2.2 *Partial Regression Coefficients / 58*
3.2.3 *The Regression Constant* / 63
3.2.4 *Problems with Three or More* Repressors / 64
3.2.5 *The Multiple Correlation R / 68*
3.3 Scale-Free Measures of Partial Association / 70
3.3.1 *Semipartial Correlation / 70*
3.3.2 *Partial Correlation / 71*
3.3.3 *The Standardized Regression Coefficient* / 73
3.4 Some Relations among Statistics / 75
3.4.1 *Relations among Simple, Multiple, Partial, and Semipartial Correlations* / 75
3.4.2 *Venn Diagrams / 78*
3.4.3 *Partial Relationships and Simple Relationships May Have Different Signs / 80*
3.4.4 *How Covariates Affect Regression Coefficients / 81*
3.4.5 *Formulas for bj, $prsr_p$ and R/82*
3.5 Chapter Summary / 83

4 • Statistical Inference in Regression **85**

4.1 Concepts in Statistical Inference / 85
4.1.1 *Statistics and Parameters / 85*
4.1.2 *Assumptions for Proper Inference / 88*
4.1.3 *Expected Values and Unbiased Estimation 191*
4.2 The ANOVA Summary Table / 92
4.2.1 *Data = Model + Error* / 95
4.2.2 *Total and Regression Sums of Squares / 97*
4.2.3 *Degrees of Freedom / 99*
4.2.4 *Mean Squares /100*
4.3 Inference about the Multiple Correlation / 102
4.3.1 *Biased and Less Biased Estimation of $_TR^2$* /102
4.3.2 *Testing a Hypothesis about $_TR$/104*
4.4 The Distribution of and Inference about a Partial Regression Coefficient / 105
4.4.1 *Testing a Null Hypothesis about $_Tbj$ /105*
4.4.2 *Interval Estimates for $_Tbj$ /106*
4.4.3 *Factors Affecting the Standard Error of bj /107*
4.4.4 *Tolerance /109*
4.5 Inferences about Partial Correlations / 112
4.5.1 *Testing a Null Hypothesis about $_T$pr$_j$ and $_Tsrj$* / 222
4.5.2 *Other Inferences about Partial Correlations /113*
4.6 Inferences about Conditional Means / 116
4.7 Miscellaneous Issues in Inference / 118
4.7.1 *How Great a Drawback Is Collinearity? /118*
4.7.2 *Contradicting Inferences /119*
4.7.3 *Sample Size and Nonsignificant Covariates /121*
4.7.4 *Inference in Simple Regression (When k = 1) /121*
4.8 Chapter Summary / 122

5 • Extending Regression Analysis Principles **125**

5.1 Dichotomous Regressors / 125
5.1.1 *Indicator or Dummy Variables /125*
5.1.2 *Estimates of Y Are Group Means* / 226
5.1.3 *The Regression Coefficient for an Indicator Is a Difference /128*

5.1.4 *A Graphic Representation /129*
5.1.5 *A Caution about Standardized Regression Coefficients for Dichotomous Regressors /130*
5.1.6 *Artificial Categorization of Numerical Variables* / 232
5.2 Regression to the Mean / 135
5.2.1 *How Regression Got Its Name /135*
5.2.2 *The Phenomenon /135*
5.2.3 *Versions of the Phenomenon /138*
5.2.4 *Misconceptions and Mistakes Fostered by Regression to the Mean /140*
5.2.5 *Accounting for Regression to the Mean Using Linear Models /141*
5.3 Multidimensional Sets / 144
5.3.1 *The Partial and Semipartial Multiple Correlation /145*
5.3.2 *What It Means If PR = 0 or SR = 0 /148*
5.3.3 *Inference Concerning Sets of Variables /148*
5.4 A Glance at the Big Picture / 152
5.4.1 *Further Extensions of Regression /153*
5.4.2 *Some Difficulties and Limitations /153*
5.5 Chapter Summary / 155

6* Statistical versus Experimental Control 157

6.1 Why Random Assignment? / 158
6.1.1 *Limitations of Statistical Control /158*
6.1.2 *The Advantage of Random Assignment 1159*
6.1.3 *The Meaning of Random Assignment /160*
6.2 Limitations of Random Assignment / 162
6.2.1 *Limitations Common to Statistical Control and Random Assignment /162*
6.2.2 *Limitations Specific to Random Assignment /165*
6.2.3 *Correlation and Causation /166*
6.3 Supplementing Random Assignment with Statistical Control / 169
6.3.1 *Increased Precision and Power /169*
6.3.2 *Invulnerability to Chance Differences between Groups 1174*
6.3.3 *Quantifying and Assessing Indirect Effects 1175*
6.4 Chapter Summary / 176

7 • Regression for Prediction 177

7.1 Mechanical Prediction and Regression / 177
7.1.1 *The Advantages of Mechanical Prediction 1177*
7.1.2 Regression as a Mechanical Prediction Method /178
7.1.3 *A Focus on R Rather Than on the Regression Weights /180*
7.2 Estimating True Validity / 181
7.2.1 *Shrunken versus Adjusted R /181*
7.2.2 *Estimating $_{T}RS$ /183*
7.2.3 *Shrunken R Using Statistical Software 1186*
7.3 Selecting Predictor Variables / 188
7.3.1 *Stepwise Regression /189*
7.3.2 All Subsets Regression 1192
7.3.3 *How Do Variable Selection Methods Perform? /192*
7A Predictor Variable Configurations / 195
7.4.1 *Partial Redundancy (the Standard Configuration) /196*
7.4.2 *Complete Redundancy /198*
7.4.3 *Independence /199*
7.4.4 *Complementarity /199*
7.4.5 *Suppression / 200*
7.4.6 *How These Configurations Relate to the Correlation between Predictors / 201*
7AH Configurations of Three or More Predictors / 205
7.5 Revisiting the Value of Human Judgment / 205
7.6 Chapter Summary / 207

8 • Assessing the Importance of Regressors **209**

8.1 What Does It Mean for a Variable to Be Important? / 210
8.1.1 *Variable Importance in Substantive or Applied Terms / 210*
8.1.2 *Variable Importance in Statistical Terms / 211*
8.2 Should Correlations Be Squared? / 212
8.2.1 *Decision Theory / 213*
8.2.2 *Small Squared Correlations Can Reflect Noteworthy Effects 1217*
8.2.3 *Pearson's r as the Ratio of a Regression Coefficient to Its Maximum Possible Value / 218*
8.2.4 *Proportional Reduction in Estimation Error / 220*
8.2.5 *When the Standard Is Perfection* / 222
8.2.6 *Summary/223*
8.3 Determining the Relative Importance of Regressors in a Single Regression Model / 223
8.3.1 *The Limitations of the Standardized Regression Coefficient / 224*
8.3.2 *The Advantage of the Semipartial Correlation* / 225
8.3.3 *Some Equivalences among Measures / 226*
8.3.4 *Eta-Squared, Partial Eta-Squared, and Cohen's f-Squared /227*
8.3.5 *Comparing Two Regression Coefficients in the Same* Model / 229
8.4 Dominance Analysis / 233
8.4.1 *Complete and Partial Dominance* /235
8.4.2 *Example Computations / 236*
8.4.3 *Dominance Analysis Using a Regression Program / 237*
8.5 Chapter Summary / 240

9 • Multicategorical Regressors **243**

9.1 Multicategorical Variables as Sets / 244
9.1.1 *Indicator (Dummy) Coding* /245
9.1.2 *Constructing Indicator Variables* / 249
9.1.3 *The Reference Category / 250*
9.1.4 *Testing the Equality of Several Means* / 252
9.1.5 *Parallels with Analysis of Variance / 254*
9.1.6 *Interpreting Estimated Y and the Regression Coefficients / 255*
9.2 Multicategorical Regressors as or with Covariates / 258
9.2.1 *Multicategorical Variables as Covariates / 258*
9.2.2 *Comparing Croups and Statistical Control 1260*
9.2.3 *Interpretation of Regression Coefficients / 264*
9.2.4 *Adjusted Means /266*
9.2.5 *Parallels with ANCOVA /268*
9.2.6 *More Than One Covariate / 271*
9.3 Chapter Summary / 273

10 • More on Multicategorical Regressors **275**

10.1 Alternative Coding Systems / 276
10.1.1 *Sequential (Adjacent or Repeated Categories) Coding / 277*
10.1.2 *Helmert Coding / 283*
10.1.3 *Effect Coding / 287*
10.2 Comparisons and Contrasts / 289
10.2.1 *Contrasts / 289*
10.2.2 *Computing the Standard Error of a Contrast / 291*
10.2.3 *Contrasts Using Statistical Software* /292
10.2.4 *Covariates and the Comparison of Adjusted Means 1294*
10.3 Weighted Group Coding and Contrasts / 298
10.3.1 *Weighted Effect Coding / 298*
10.3.2 *Weighted Helmert Coding / 300*

10.3.3 *Weighted Contrasts / 304*
10.3.4 *Application to Adjusted Means / 308*
10.4 Chapter Summary / 308

11 • Multiple Tests **311**

11.1 The Multiple Test Problem / 312
11.1.1 An Illustration through Simulation / 312
11.1.2 *The Problem Defined / 315*
11.1.3 *The Role of Sample Size / 316*
11.1.4 *The Generality of the Problem / 317*
11.1.5 *Do Omnibus Tests Offer "Protection"? /319*
11.1.6 *Should You Be Concerned about the Multiple Test Problem? /319*
11.2 The Bonferroni Method / 320
11.2.1 *Independent Tests/321*
11.2.2 *The Bonferroni Method for Nonindependent Tests / 322*
11.2.3 *Revisiting the Illustration / 324*
11.2.4 Bonferroni Layering / 324
11.2.5 *Finding an "Exact" p-Value /325*
11.2.6 *Nonsense Values / 327*
11.2.7 *Flexibility of the Bonferroni Method / 327*
11.2.8 *Power of the Bonferroni Method/328*
11.3 Some Basic Issues Surrounding Multiple Tests / 328
11.3.1 *Why Correct for Multiple Tests at All? / 329*
11.3.2 *Why Not Correct for the Whole History of Science? / 330*
11.3.3 *Plausibility and Logical Independence of Hypotheses / 331*
11.3.4 *Planned versus Unplanned Tests / 335*
11.3.5 *Summary of the Basic Issues / 338*
11.4 Chapter Summary / 338

12 • Nonlinear Relationships **341**

12.1 Linear Regression Can Model Nonlinear Relationships / 341
12.1.1 *When Must Curves Be Fitted? / 342*
12.1.2 The Graphical Display of Curvilinearity / 344
12.2 Polynomial Regression / 347
12.2.1 *Basic Principles / 347*
12.2.2 *An Example / 350*
12.2.3 *The Meaning of the Regression Coefficients for Lower-Order Regressors / 352*
12.2.4 *Centering Variables in Polynomial Regression / 354*
12.2.5 *Finding a Parabola's Maximum or Minimum / 356*
12.3 Spline Regression / 357
12.3.1 *Linear Spline Regression / 358*
12.3.2 *Implementation in Statistical Software / 363*
12.3.3 *Polynomial Spline Regression / 364*
12.3.4 *Covariates, Weak Curvilinearity, and Choosing Joints / 368*
12.4 Transformations of Dependent Variables or Regressors / 369
12.4.1 *Logarithmic Transformation / 370*
12.4.2 *The Box-Cox Transformation / 372*
12.5 Chapter Summary / 374

13 • Linear Interaction **377**

13.1 Interaction Fundamentals / 377
13.1.1 *Interaction as a Difference in Slope /377*
13.1.2 *Interaction between Two Numerical Regressors / 378*
13.1.3 *Interaction versus Intercorrelation / 379*
13.1.4 *Simple Linear Interaction / 380*

13.1.5 *Representing Simple Linear Interaction with a Cross-Product / 381*
13.1.6 *The Symmetry of Interaction / 382*
13.1.7 *Interaction as a Warped Surface / 384*
13.1.8 *Covariates in a Regression Model with an Interaction* / 385
13.1.9 *The Meaning of the Regression Coefficients / 385*
13.1.10 *An Example with Estimation Using Statistical Software/386*
13.2 Interaction Involving a Categorical Regressor / 390
13.2.1 *Interaction between a Dichotomous and a Numerical Regressor / 390*
13.2.2 *The Meaning of the Regression Coefficients* / 392
13.2.3 *Interaction Involving a Multicategorical and a Numerical Regressor / 394*
13.2.4 *Inference When Interaction Requires More Than One Regression Coefficient / 397*
13.2.5 *A Substantive Example / 398*
13.2.6 *Interpretation of the Regression Coefficients* / 402
13.3 Interaction between Two Categorical Regressors / 404
13.3.1 *The 2x2 Design / 404*
13.3.2 *Interaction between a Dichotomous and q Multicategorical Regressor / 407*
13.3.3 *Interaction between Two Multicategorical Regressors / 408*
13.4 Chapter Summary / 408

14 * Probing Interactions and Various Complexities 411

14.1 Conditional Effects as Functions / 411
14.1.1 *When the Interaction Involves Dichotomous or Numerical Variables / 412*
14.1.2 *When the Interaction Involves a Multicategorical Variable / 414*
14.2 Inference about a Conditional Effect / 415
14.2.1 *When the Focal Predictor and Moderator Are Numerical or Dichotomous* / 435
14.2.2 *When the Focal Predictor or Moderator Is Multicategorical /419*
14.3 Probing an Interaction / 422
14.3.1 *Examining Conditional Effects at Various Values of the Moderator / 423*
14.3.2 *The Johnson-Neyman Technique / 425*
14.3.3 *Testing versus Probing an Interaction* / 427
14.3.4 *Comparing Conditional Effects / 428*
14.4 Complications and Confusions in the Study of Interactions / 429
14.4.1 *The Difficulty of Detecting Interactions / 429*
14.4.2 *Confusing Interaction with Curvilinearity* /430
14.4.3 *How the Scaling of Y Affects Interaction* / 432
14.4.4 *The Interpretation of Lower-Order Regression Coefficients When a Cross-Product Is Present /433*
14.4.5 *Some Myths about Testing Interaction* / 435
14.4.6 *Interaction and Nonsignificant Linear Terms / 437*
14.4.7 *Homogeneity of Regression in ANCOVA* /437
14.4.8 *Multiple, Higher-Order, and Curvilinear Interactions / 438*
14.4.9 *Artificial Categorization of Continua / 441*
14.5 Organizing Tests on Interaction / 441
14.5.1 *Three Approaches to Managing Complications / 442*
14.5.2 *Broad versus Narrow Tests* / 443
14.6 Chapter Summary / 445

15 • Mediation and Path Analysis 447

15.1 Path Analysis and Linear Regression / 448
15.1.1 *Direct, Indirect, and Total Effects / 448*
15.1.2 *The Regression Algebra of Path Analysis / 452*
15.1.3 *Covariates* / 454
15.1.4 *Inference about the Total and Direct Effects / 455*
15.1.5 *Inference about the Indirect Effect / 455*
15.1.6 *Implementation in Statistical Software / 458*

15.2 Multiple Mediator Models / 464
15.2.1 *Path Analysis for a Parallel Multiple Mediator Model / 464*
15.2.2 *Path Analysis for a Serial Multiple Mediator Model / 467*
15.3 Extensions, Complications, and Miscellaneous Issues / 469
15.3.1 *Causality and Causal Order / 469*
15.3.2 *The Causal Steps Approach / 471*
15.3.3 *Mediation of a Nonsignificant Total Effect / 472*
15.3.4 *Multicategorical Independent Variables / 473*
15.3.5 *Fixing Direct Effects to Zero / 474*
15.3.6 *Nonlinear Effects / 475*
15.3.7 *Moderated Mediation / 475*
15.4 Chapter Summary / 476

16 • Detecting and Managing Irregularities 479

16.1 Regression Diagnostics / 480
16.1.1 *Shortcomings of Eyeballing the Data / 481*
16.1.2 *Types of Extreme Cases / 482*
16.1.3 *Quantifying Leverage, Distance, and Influence /484*
16.1.4 *Using Diagnostic Statistics / 490*
16.1.5 *Generating Regression Diagnostics with Computer Software / 494*
16.2 Detecting Assumption Violations / 495
16.2.1 *Detecting Nonlinearity / 496*
16.2.2 *Detecting Non-Normality / 498*
16.2.3 *Detecting Heteroscedasticity / 499*
16.2.4 *Testing Assumptions as a Set / 505*
16.2.5 *What about Nonindependence? / 506*
16.3 Dealing with Irregularities / 509
16.3.1 *Heteroscedasticity-Consistent Standard Errors / 511*
16.3.2 *The Jackknife / 512*
16.3.3 *Bootstrapping / 512*
16.3.4 *Permutation Tests / 513*
16.4 Inference without Random Sampling / 514
16.5 Keeping the Diagnostic Analysis Manageable / 516
16.6 Chapter Summary / 517

17 • Power, Measurement Error, and Various Miscellaneous Topics 519

17.1 Power and Precision of Estimation / 519
17.1.1 *Factors Determining Desirable Sample Size / 520*
17.1.2 *Revisiting the Standard Error of a Regression Coefficient / 521*
17.1.3 *On the Effect of Unnecessary Covariates / 524*
17.2 Measurement Error / 525
17.2.1 *What Is Measurement Error? / 525*
17.2.2 *Measurement Error in* Y / *526*
17.2.3 *Measurement Error in Independent Variables / 527*
17.2.4 *The Biggest Weakness of Regression: Measurement Error in Covariates / 527*
17.2.5 *Summary: The Effects of Measurement Error / 528*
17.2.6 *Managing Measurement Error / 530*
17.3 An Assortment of Problems / 532
17.3.1 *Violations of the Basic Assumptions / 532*
17.3.2 *Collinearity / 532*
17.3.3 *Singularity / 534*
17.3.4 *Specification Error and Overcontrol / 538*
17.3.5 *Noninterval Scaling / 541*
17.3.6 *Missing Data / 543*
17.3.7 *Rounding Error / 546*
17.4 Chapter Summary / 548

18 • Logistic Regression and Other Linear Models **551**

18.1 Logistic Regression / 551
- 18.1.1 *Measuring a Model's Fit to Data / 552*
- 18.1.2 *Odds and Logits / 554*
- 18.1.3 *The Logistic Regression Equation* / 556
- 18.1.4 *An Example with a Single Regressor / 557*
- 18.1.5 *Interpretation of and Inference about the Regression Coefficients / 560*
- 18.1.6 *Multiple Logistic Regression and Implementation in Computing Software / 562*
- 18.1.7 *Measuring and Testing the Fit of the Model / 565*
- 18.1.8 *Further Extensions / 568*
- 18.1.9 *Discriminant Function Analysis / 568*
- 18.1.10 *Using OLS Regression with a Dichotomous* Y / 569

18.2 Other Linear Modeling Methods / 570
- 18.2.1 *Ordered Logistic and Probit Regression / 570*
- 18.2.2 *Poisson Regression and Related Models of Count Outcomes* / 572
- 18.2.3 *Time Series Analysis / 573*
- 18.2.4 *Survival Analysis* / 573
- 18.2.5 *Structural Equation Modeling / 574*
- 18.2.6 *Multilevel Modeling / 575*
- 18.2.7 *Other Resources / 577*

18.3 Chapter Summary / 578

Appendices

A. The RLM Macro for SPSS and SAS **581**

B. Linear Regression Analysis Using R **603**

C. Statistical Tables **611**

D. The Matrix Algebra of Linear Regression Analysis **621**

References **627**

Author Index **637**

Subject Index **641**

About the Authors **661**

Data files for the examples used in the book and files containing the SPSS and SAS versions of RLM are available on the companion web page at *www.afhayes.com.*