

See discussions, stats, and author profiles for this publication at: <https://www.researchgate.net/publication/266497792>

A Second Course in Statistics: Regression Analysis

Article in *Journal of the American Statistical Association* · June 1997
DOI: 10.2307/2965740

CITATIONS
427

READS
19,561

2 authors:



William M Mendenhall
University of Florida
900 PUBLICATIONS **25,569** CITATIONS

SEE PROFILE



Terry Sincich
University of South Florida
25 PUBLICATIONS **1,750** CITATIONS

SEE PROFILE

Some of the authors of this publication are also working on these related projects:



Proton Therapy for Pancreas Cancer [View project](#)

FIFTH EDITION

A Second Course in Statistics: Regression Analysis

WILLIAM MENDENHALL

University of Florida

TERRY SINCICH

University of South Florida



PRENTICE HALL

Upper Saddle River, New Jersey 07458

Contents

Preface xii

CHAPTER 1 A Review of Basic Concepts (Optional) 1

- 1.1 Statistics and Data 2
- 1.2 Populations, Samples, and Random Sampling 6
- 1.3 Describing Data Sets Graphically 10
- 1.4 Describing Data Sets Numerically 19
- 1.5 The Normal Probability Distribution 29
- 1.6 Sampling Distributions and the Central Limit Theorem 35
- 1.7 Estimating a Population Mean 39
- 1.8 Testing a Hypothesis About a Population Mean 51
- 1.9 Inferences About the Difference Between Two Population Means 59
- 1.10 Comparing Two Population Variances 72

CHAPTER 2 Introduction to Regression Analysis 90

- 2.1 Modeling a Response 91
- 2.2 Overview of Regression Analysis 93
- 2.3 Regression Applications 96
- 2.4 Collecting the Data for Regression 97

CHAPTER 3 Simple Linear Regression 101

- 3.1 Introduction 102
- 3.2 The Straight-Line Probabilistic Model 102
- 3.3 Fitting the Model: The Method of Least Squares 105
- 3.4 Model Assumptions 115
- 3.5 An Estimator of σ^2 117
- 3.6 Assessing the Utility of the Model: Making Inferences About the Slope β_1 120
- 3.7 The Coefficient of Correlation 127
- 3.8 The Coefficient of Determination 133
- 3.9 Using the Model for Estimation and Prediction 139
- 3.10 Simple Linear Regression: An Example Using the Computer 146
- 3.11 Regression Through the Origin (Optional) 153
- 3.12 A Summary of the Steps to Follow in a Simple Linear Regression Analysis 162

CHAPTER 4 Multiple Regression 172

- 4.1 The General Linear Model 173
- 4.2 Model Assumptions 174
- 4.3 Fitting the Model: The Method of Least Squares 175
- 4.4 Estimation of σ^2 , the Variance of ϵ 178
- 4.5 Inferences About the β Parameters 180
- 4.6 The Multiple Coefficient of Determination, R^2 191
- 4.7 Testing the Utility of a Model: The Analysis of Variance F Test 193
- 4.8 Using the Model for Estimation and Prediction 204
- 4.9 Other Linear Models 211
- 4.10 A Test for Comparing Nested Models 233
- 4.11 Stepwise Regression 242
- 4.12 Other Variable Selection Techniques (Optional) 252
- 4.13 Multiple Regression: A Complete Example 257
- 4.14 A Summary of the Steps to Follow in a Multiple Regression Analysis 262

CHAPTER 5 Model Building 273

- 5.1 Introduction: Why Model Building Is Important 274
- 5.2 The Two Types of Independent Variables: Quantitative and Qualitative 275
- 5.3 Models with a Single Quantitative Independent Variable 277
- 5.4 First-Order Models with Two or More Quantitative Independent Variables 285
- 5.5 Second-Order Models with Two or More Quantitative Independent Variables 288
- 5.6 Coding Quantitative Independent Variables (Optional) 296
- 5.7 Models with One Qualitative Independent Variable 302
- 5.8 Models with Two Qualitative Independent Variables 306
- 5.9 Models with Three or More Qualitative Independent Variables 319
- 5.10 Models with Both Quantitative and Qualitative Independent Variables 322
- 5.11 Model Building: An Example 335

CHAPTER 6 Some Regression Pitfalls 347

- 6.1 Introduction 348
- 6.2 Observational Data Versus Designed Experiments 348
- 6.3 Deviating from the Assumptions 350
- 6.4 Parameter Estimability and Interpretation 351
- 6.5 Multicollinearity 355
- 6.6 Extrapolation: Predicting Outside the Experimental Region 361
- 6.7 Data Transformations 363

CHAPTER 7	Residual Analysis	377
	7.1 Introduction	378
	7.2 Plotting Residuals and Detecting Lack of Fit	378
	7.3 Detecting Unequal Variances	394
	7.4 Checking the Normality Assumption	409
	7.5 Detecting Outliers and Identifying Influential Observations	414
	7.6 Detecting Residual Correlation: The Durbin–Watson Test	430
CHAPTER 8	Special Topics in Regression (Optional)	451
	8.1 Introduction	452
	8.2 Piecewise Linear Regression	452
	8.3 Inverse Prediction	457
	8.4 Weighted Least Squares	465
	8.5 Modeling Qualitative Dependent Variables	472
	8.6 Logistic Regression	476
	8.7 Ridge Regression	485
	8.8 Robust Regression	487
	8.9 Model Validation	489
CHAPTER 9	Time Series Modeling and Forecasting	494
	9.1 What Is a Time Series?	495
	9.2 Time Series Components	495
	9.3 Forecasting Using Smoothing Techniques (Optional)	497
	9.4 Forecasting: The Regression Approach	513
	9.5 Autocorrelation and Autoregressive Error Models	522
	9.6 Other Models for Autocorrelated Errors (Optional)	527
	9.7 Constructing Time Series Models	528
	9.8 Fitting Time Series Models with Autoregressive Errors	533
	9.9 Forecasting with Time Series Autoregressive Models	541
	9.10 Seasonal Time Series Models: An Example	548
	9.11 Forecasting Using Lagged Values of the Dependent Variable (Optional)	551
CHAPTER 10	Principles of Experimental Design	558
	10.1 Introduction	559
	10.2 Experimental Design Terminology	559
	10.3 Controlling the Information in an Experiment	562

- 10.4 Noise-Reducing Designs 563
- 10.5 Volume-Increasing Designs 570
- 10.6 Selecting the Sample Size 576
- 10.7 The Importance of Randomization 578

CHAPTER 11 **The Analysis of Variance for Designed Experiments** 581

- 11.1 Introduction 582
- 11.2 The Logic Behind an Analysis of Variance 582
- 11.3 Completely Randomized Designs 584
- 11.4 Randomized Block Designs 603
- 11.5 Two-Factor Factorial Experiments 621
- 11.6 More Complex Factorial Designs (Optional) 644
- 11.7 Follow-Up Analysis: Tukey's Multiple Comparisons of Means 654
- 11.8 Other Multiple Comparisons Methods (Optional) 664
- 11.9 Checking ANOVA Assumptions 673

CASE STUDY 12 **Modeling the Sale Prices of Residential Properties in Four Neighborhoods** 696

- 12.1 The Problem 697
- 12.2 The Data 697
- 12.3 The Models 697
- 12.4 Model Comparisons 700
- 12.5 Interpreting the Prediction Equation 704
- 12.6 Predicting the Sale Price of a Property 708
- 12.7 Conclusions 711

CASE STUDY 13 **An Analysis of Rain Levels in California** 712

- 13.1 The Problem 713
- 13.2 The Data 713
- 13.3 A Model for Average Annual Precipitation 713
- 13.4 A Residual Analysis of the Model 715
- 13.5 Adjustments to the Model 718
- 13.6 Conclusions 721

CASE STUDY 14 **Reluctance to Transmit Bad News: The MUM Effect** 722

- 14.1 The Problem 723

- 14.2 The Design 723
- 14.3 Analysis of Variance Models and Results 724
- 14.4 Follow-Up Analysis 725
- 14.5 Conclusions 727

CASE STUDY 15 **An Investigation of Factors Affecting the Sale Price of Condominium Units Sold at Public Auction** 728

- 15.1 The Problem 729
- 15.2 The Data 730
- 15.3 The Models 731
- 15.4 The Regression Analyses 733
- 15.5 An Analysis of the Residuals from Model 3 733
- 15.6 What the Model 3 Regression Analysis Tells Us 741
- 15.7 Comparing the Mean Sale Price for Two Types of Units (Optional) 749
- 15.8 Conclusions 750

CASE STUDY 16 **Modeling Daily Peak Electricity Demands** 753

- 16.1 The Problem 754
- 16.2 The Data 754
- 16.3 The Models 755
- 16.4 The Regression and Autoregression Analyses 759
- 16.5 Forecasting Daily Peak Electricity Demand 763
- 16.6 Conclusions 765

APPENDIX A **The Mechanics of a Multiple Regression Analysis** 767

- A.1 Introduction 768
- A.2 Matrices and Matrix Multiplication 769
- A.3 Identity Matrices and Matrix Inversion 774
- A.4 Solving Systems of Simultaneous Linear Equations 779
- A.5 The Least Squares Equations and Their Solution 782
- A.6 Calculating SSE and s^2 788
- A.7 Standard Errors of Estimators, Test Statistics, and Confidence Intervals for $\beta_0, \beta_1, \dots, \beta_k$ 789
- A.8 A Confidence Interval for a Linear Function of the β Parameters; A Confidence Interval for $E(y)$ 793
- A.9 A Prediction Interval for Some Value of y to Be Observed in the Future 799

APPENDIX B **A Procedure for Inverting a Matrix** 805

APPENDIX C Useful Statistical Tables 811

Table 1	Normal Curve Areas	812
Table 2	Critical Values for Student's t	813
Table 3	Critical Values for the F Statistic: $F_{.10}$	814
Table 4	Critical Values for the F Statistic: $F_{.05}$	816
Table 5	Critical Values for the F Statistic: $F_{.025}$	818
Table 6	Critical Values for the F Statistic: $F_{.01}$	820
Table 7	Random Numbers	822
Table 8	Critical Values for the Durbin-Watson d Statistic ($\alpha = .05$)	825
Table 9	Critical Values for the Durbin-Watson d Statistic ($\alpha = .01$)	826
Table 10	Critical Values for the χ^2 Statistic	827
Table 11	Percentage Points of the Studentized Range, $q(p, \nu)$, Upper 5%	829
Table 12	Percentage Points of the Studentized Range, $q(p, \nu)$, Upper 1%	831

APPENDIX D SAS Tutorial 833

D.1	Introduction	834
D.2	Creating a SAS Data Set	834
D.3	Accessing an External Data File	835
D.4	Relative Frequency Distributions, Descriptive Statistics, Correlations, and Plots	836
D.5	Simple Linear Regression	837
D.6	Multiple Regression	838
D.7	Stepwise Regression	839
D.8	Residual Analysis and Regression Diagnostics	840
D.9	Logistic Regression	841
D.10	Time Series Forecasting Models and Durbin-Watson Test	842
D.11	Analysis of Variance	842

APPENDIX E SPSS Tutorial 844

E.1	Introduction	845
E.2	Creating an SPSS Data File	845
E.3	Accessing an External Data File	846
E.4	Relative Frequency Distributions, Descriptive Statistics, Correlations, and Plots	847
E.5	Simple Linear Regression	848
E.6	Multiple Regression	849
E.7	Stepwise Regression	850
E.8	Residual Analysis and Regression Diagnostics	850

- E.9 Logistic Regression 851
- E.10 Analysis of Variance 852

APPENDIX F MINITAB Tutorial 854

- F.1 Introduction 855
- F.2 Creating a MINITAB Data Worksheet 855
- F.3 Accessing an External Data File 856
- F.4 Relative Frequency Distributions, Descriptive Statistics, Correlations, and Plots 857
- F.5 Simple Linear Regression 858
- F.6 Multiple Regression 859
- F.7 Stepwise Regression 859
- F.8 Residual Analysis and Regression Diagnostics 860
- F.9 Time Series Smoothing Methods and Durbin–Watson Test 861
- F.10 Analysis of Variance 862

APPENDIX G ASP Tutorial 864

- G.1 Introduction 865
- G.2 Hardware Requirements 865
- G.3 Getting Started 865
- G.4 The Main Menu 866
- G.5 Alternative Commands Menu 866
- G.6 Creating a Data Matrix 868
- G.7 Accessing an External Data File 869
- G.8 Analyzing a Data Matrix 869
- G.9 Available Documentation 869

APPENDIX H Data Set

Sealed Bid Data for Fixed and Competitive Highway Construction Contracts 870

APPENDIX I Data Set

Real Estate Appraisals and Sales Data for Seven Neighborhoods in Tampa, Florida 874

APPENDIX J Data Set

Condominium Sales Data 880

Answers to Odd-Numbered Exercises 884

Index 895