
A SECOND COURSE IN STATISTICS

REGRESSION ANALYSIS

Eighth Edition

WILLIAM MENDENHALL
University of Florida

TERRY SINCICH
University of South Florida

 **Pearson**

Director, Portfolio Management: Deirdre Lynch
Courseware Portfolio Manager: Patrick Barbera
Courseware Portfolio Management Assistant:
Morgan Danna
Content Producer: Angela Montoya
Managing Producer: Karen Wernholm
Producer: Shana Siegmund
Product Marketing Manager: Yvonne Vannatta
Product Marketing Assistant: Jon Bryant
Field Marketing Manager: Evan St. Cyr

Senior Author Support/Technology
Specialist: Joe Vetere
Manager, Rights and Permissions: Gina Cheselka
Text and Cover Design: Jerilyn Bockorick
Production Coordination, Composition,
and Illustrations: Pearson CSC
Manufacturing Buyer: Carol Melville,
LSC Communications
Cover Image: enot-poloskun/Getty Images

Copyright © 2020, 2012, 2003 by **Pearson Education, Inc.** All Rights Reserved. Printed in the United States of America. This publication is protected by copyright, and permission should be obtained from the publisher prior to any prohibited reproduction, storage in a retrieval system, or transmission in any form or by any means, electronic, mechanical, photocopying, recording, or otherwise. For information regarding permissions, request forms and the appropriate contacts within the Pearson Education Global Rights & Permissions department, please visit www.pearsoned.com/permissions/.

PEARSON, ALWAYS LEARNING, and MYLAB are exclusive trademarks owned by Pearson Education, Inc. or its affiliates in the U.S. and/or other countries.

Unless otherwise indicated herein, any third-party trademarks that may appear in this work are the property of their respective owners and any references to third-party trademarks, logos or other trade dress are for demonstrative or descriptive purposes only. Such references are not intended to imply any sponsorship, endorsement, authorization, or promotion of Pearson's products by the owners of such marks, or any relationship between the owner and Pearson Education, Inc. or its affiliates, authors, licensees or distributors.

[For instructor editions: This work is solely for the use of instructors and administrators for the purpose of teaching courses and assessing student learning. Unauthorized dissemination, publication or sale of the work, in whole or in part (including posting on the internet) will destroy the integrity of the work and is strictly prohibited.]

MICROSOFT AND/OR ITS RESPECTIVE SUPPLIERS MAKE NO REPRESENTATIONS ABOUT THE SUITABILITY OF THE INFORMATION CONTAINED IN THE DOCUMENTS AND RELATED GRAPHICS PUBLISHED AS PART OF THE SERVICES FOR ANY PURPOSE. ALL SUCH DOCUMENTS AND RELATED GRAPHICS ARE PROVIDED "AS IS" WITHOUT WARRANTY OF ANY KIND. MICROSOFT AND/OR ITS RESPECTIVE SUPPLIERS HEREBY DISCLAIM ALL WARRANTIES AND CONDITIONS WITH REGARD TO THIS INFORMATION, INCLUDING ALL WARRANTIES AND CONDITIONS OF MERCHANTABILITY, WHETHER EXPRESS, IMPLIED OR STATUTORY, FITNESS FOR A PARTICULAR PURPOSE, TITLE AND NON-INFRINGEMENT. IN NO EVENT SHALL MICROSOFT AND/OR ITS RESPECTIVE SUPPLIERS BE LIABLE FOR ANY SPECIAL, INDIRECT OR CONSEQUENTIAL DAMAGES OR ANY DAMAGES WHATSOEVER RESULTING FROM LOSS OF USE, DATA OR PROFITS, WHETHER IN AN ACTION OF CONTRACT, NEGLIGENCE OR OTHER TORTIOUS ACTION, ARISING OUT OF OR IN CONNECTION WITH THE USE OR PERFORMANCE OF INFORMATION AVAILABLE FROM THE SERVICES. THE DOCUMENTS AND RELATED GRAPHICS CONTAINED HEREIN COULD INCLUDE TECHNICAL INACCURACIES OR TYPOGRAPHICAL ERRORS. CHANGES ARE PERIODICALLY ADDED TO THE INFORMATION HEREIN. MICROSOFT AND/OR ITS RESPECTIVE SUPPLIERS MAY MAKE IMPROVEMENTS AND/OR CHANGES IN THE PRODUCT(S) AND/OR THE PROGRAM(S) DESCRIBED HEREIN AT ANY TIME. PARTIAL SCREEN SHOTS MAY BE VIEWED IN FULL WITHIN THE SOFTWARE VERSION SPECIFIED.

MICROSOFT® AND WINDOWS® ARE REGISTERED TRADEMARKS OF THE MICROSOFT CORPORATION IN THE U.S.A. AND OTHER COUNTRIES. SCREEN SHOTS AND ICONS REPRINTED WITH PERMISSION FROM THE MICROSOFT CORPORATION. THIS BOOK IS NOT SPONSORED OR ENDORSED BY OR AFFILIATED WITH THE MICROSOFT CORPORATION.

Library of Congress Cataloging-in-Publication Data

Names: Mendenhall, William, author. | Sincich, Terry, author.

Title: A second course in statistics : regression analysis / William Mendenhall, Terry Sincich.

Other titles: Corrected title: Second course in statistics : regression analysis

Description: Eighth edition. | [Hoboken, New Jersey] : Pearson Education, Inc., [2020] | Includes bibliographical references and index.

Identifiers: LCCN 2018040176 | ISBN 9780135163795 | ISBN 013516379X

Subjects: LCSH: Commercial statistics. | Statistics. | Regression analysis.

Classification: LCC HF1017 .M46 2020 | DDC 519.5/36--dc23 LC record available at <https://lcn.loc.gov/2018040176>



ISBN 13: 978-0-13-516379-5
ISBN 10: 0-13-516379-X

CONTENTS

Preface ix

1 A REVIEW OF BASIC CONCEPTS (OPTIONAL) 1

- 1.1 Statistics and Data 1
- 1.2 Populations, Samples, and Random Sampling 4
- 1.3 Describing Qualitative Data 8
- 1.4 Describing Quantitative Data Graphically 13
- 1.5 Describing Quantitative Data Numerically 20
- 1.6 The Normal Probability Distribution 26
- 1.7 Sampling Distributions and the Central Limit Theorem 30
- 1.8 Estimating a Population Mean 34
- 1.9 Testing a Hypothesis About a Population Mean 43
- 1.10 Inferences About the Difference Between Two Population Means 52
- 1.11 Comparing Two Population Variances 67

2 INTRODUCTION TO REGRESSION ANALYSIS 83

- 2.1 Modeling a Response 83
- 2.2 Overview of Regression Analysis 85
- 2.3 Regression Applications 87
- 2.4 Collecting the Data for Regression 90

3 SIMPLE LINEAR REGRESSION 93

- 3.1 Introduction 93
- 3.2 The Straight-Line Probabilistic Model 94
- 3.3 Fitting the Model: The Method of Least Squares 96
- 3.4 Model Assumptions 107
- 3.5 An Estimator of σ^2 108
- 3.6 Assessing the Utility of the Model: Making Inferences About the Slope β_1 112
- 3.7 The Coefficient of Correlation 118
- 3.8 The Coefficient of Determination 124
- 3.9 Using the Model for Estimation and Prediction 130

- 3.10 A Complete Example 137
- 3.11 Regression Through the Origin (Optional) 144

CASE STUDY 1 LEGAL ADVERTISING — DOES IT PAY? 161

4 MULTIPLE REGRESSION MODELS 168

- 4.1 General Form of a Multiple Regression Model 168
- 4.2 Model Assumptions 170
- 4.3 A First-Order Model with Quantitative Predictors 171
- 4.4 Fitting the Model: The Method of Least Squares 172
- 4.5 Estimation of σ^2 , the Variance of ε 175
- 4.6 Testing the Utility of a Model: The Analysis of Variance F -Test 177
- 4.7 Inferences About the Individual β Parameters 180
- 4.8 Multiple Coefficients of Determination: R^2 and R_a^2 182
- 4.9 Using the Model for Estimation and Prediction 191
- 4.10 An Interaction Model with Quantitative Predictors 196
- 4.11 A Quadratic (Second-Order) Model with a Quantitative Predictor 203
- 4.12 More Complex Multiple Regression Models (Optional) 211
- 4.13 A Test for Comparing Nested Models 230
- 4.14 A Complete Example 240

CASE STUDY 2 MODELING THE SALE PRICES OF RESIDENTIAL PROPERTIES IN FOUR NEIGHBORHOODS 256

5 PRINCIPLES OF MODEL BUILDING 269

- 5.1 Introduction: Why Model Building Is Important 269
- 5.2 The Two Types of Independent Variables: Quantitative and Qualitative 271
- 5.3 Models with a Single Quantitative Independent Variable 273
- 5.4 First-Order Models with Two or More Quantitative Independent Variables 282
- 5.5 Second-Order Models with Two or More Quantitative Independent Variables 284
- 5.6 Coding Quantitative Independent Variables (Optional) 292
- 5.7 Models with One Qualitative Independent Variable 298
- 5.8 Models with Two Qualitative Independent Variables 303

- 5.9 Models with Three or More Qualitative Independent Variables 313
- 5.10 Models with Both Quantitative and Qualitative Independent Variables 316
- 5.11 External Model Validation (Optional) 326

6 VARIABLE SCREENING METHODS 337

- 6.1 Introduction: Why Use a Variable Screening Method? 337
- 6.2 Stepwise Regression 338
- 6.3 All-Possible-Regressions Selection Procedure 344
- 6.4 Caveats 348

CASE STUDY 3 DEREGULATION OF THE INTRASTATE TRUCKING INDUSTRY 355

7 SOME REGRESSION PITFALLS 365

- 7.1 Introduction 365
- 7.2 Observational Data versus Designed Experiments 365
- 7.3 Parameter Estimability and Interpretation 368
- 7.4 Multicollinearity 373
- 7.5 Extrapolation: Predicting Outside the Experimental Region 379
- 7.6 Variable Transformations 381

8 RESIDUAL ANALYSIS 393

- 8.1 Introduction 393
- 8.2 Regression Residuals 394
- 8.3 Detecting Lack of Fit 399
- 8.4 Detecting Unequal Variances 408
- 8.5 Checking the Normality Assumption 419
- 8.6 Detecting Outliers and Identifying Influential Observations 423
- 8.7 Detecting Residual Correlation: The Durbin–Watson Test 435

CASE STUDY 4 AN ANALYSIS OF RAIN LEVELS IN CALIFORNIA 449

CASE STUDY 5 AN INVESTIGATION OF FACTORS AFFECTING THE SALE PRICE OF CONDOMINIUM UNITS SOLD AT PUBLIC AUCTION 457

9 SPECIAL TOPICS IN REGRESSION (OPTIONAL) 475

- 9.1 Introduction 475**
- 9.2 Piecewise Linear Regression 475**
- 9.3 Inverse Prediction 485**
- 9.4 Weighted Least Squares 493**
- 9.5 Modeling Qualitative Dependent Variables 501**
- 9.6 Logistic Regression 504**
- 9.7 Poisson Regression 517**
- 9.8 Ridge and LASSO Regression 525**
- 9.9 Robust Regression 530**
- 9.10 Nonparametric Regression Models 533**

10 INTRODUCTION TO TIME SERIES MODELING AND FORECASTING 539

- 10.1 What Is a Time Series? 539**
- 10.2 Time Series Components 540**
- 10.3 Forecasting Using Smoothing Techniques (Optional) 542**
- 10.4 Forecasting: The Regression Approach 557**
- 10.5 Autocorrelation and Autoregressive Error Models 565**
- 10.6 Other Models for Autocorrelated Errors (Optional) 568**
- 10.7 Constructing Time Series Models 570**
- 10.8 Fitting Time Series Models with Autoregressive Errors 574**
- 10.9 Forecasting with Time Series Autoregressive Models 581**
- 10.10 Seasonal Time Series Models: An Example 587**
- 10.11 Forecasting Using Lagged Values of the Dependent Variable (Optional) 590**

CASE STUDY 6 MODELING DAILY PEAK ELECTRICITY DEMANDS 596

11 PRINCIPLES OF EXPERIMENTAL DESIGN 608

- 11.1 Introduction 608**
- 11.2 Experimental Design Terminology 608**
- 11.3 Controlling the Information in an Experiment 611**
- 11.4 Noise-Reducing Designs 612**

- 11.5 Volume-Increasing Designs 619
- 11.6 Selecting the Sample Size 625
- 11.7 The Importance of Randomization 627

12 THE ANALYSIS OF VARIANCE FOR DESIGNED EXPERIMENTS 630

- 12.1 Introduction 630
- 12.2 The Logic Behind an Analysis of Variance 631
- 12.3 One-Factor Completely Randomized Designs 632
- 12.4 Randomized Block Designs 648
- 12.5 Two-Factor Factorial Experiments 664
- 12.6 More Complex Factorial Designs (Optional) 687
- 12.7 Follow-Up Analysis: Tukey's Multiple Comparisons of Means 697
- 12.8 Other Multiple Comparisons Methods (Optional) 709
- 12.9 Checking ANOVA Assumptions 718

CASE STUDY 7 VOICE VERSUS FACE RECOGNITION — DOES ONE FOLLOW THE OTHER? 738

APPENDIX A DERIVATION OF THE LEAST SQUARES ESTIMATES OF β_0 AND β_1 IN SIMPLE LINEAR REGRESSION 744

APPENDIX B THE MECHANICS OF A MULTIPLE REGRESSION ANALYSIS 746

- B.1 Introduction 746
- B.2 Matrices and Matrix Multiplication 747
- B.3 Identity Matrices and Matrix Inversion 751
- B.4 Solving Systems of Simultaneous Linear Equations 754
- B.5 The Least Squares Equations and Their Solutions 756
- B.6 Calculating SSE and s^2 762
- B.7 Standard Errors of Estimators, Test Statistics, and Confidence Intervals for $\beta_0, \beta_1, \dots, \beta_k$ 762
- B.8 A Confidence Interval for a Linear Function of the β Parameters; a Confidence Interval for $E(y)$ 765

B.9	A Prediction Interval for Some Value of y to be Observed in the Future	771
------------	--	------------

APPENDIX **C** A PROCEDURE FOR INVERTING A MATRIX **775**

APPENDIX **D** USEFUL STATISTICAL TABLES **780**

Table D.1	Normal Curve Areas	781
Table D.2	Critical Values for Student's t	782
Table D.3	Critical Values for the F Statistic: $F_{.10}$	783
Table D.4	Critical Values for the F Statistic: $F_{.05}$	785
Table D.5	Critical Values for the F Statistic: $F_{.025}$	787
Table D.6	Critical Values for the F Statistic: $F_{.01}$	789
Table D.7	Critical Values for the Durbin–Watson d Statistic ($\alpha = .05$)	791
Table D.8	Critical Values for the Durbin–Watson d Statistic ($\alpha = .01$)	792
Table D.9	Critical Values for the χ^2 Statistic	793
Table D.10	Percentage Points of the Studentized Range $q(p, v)$, Upper 5%	795
Table D.11	Percentage Points of the Studentized Range $q(p, v)$, Upper 1%	797

APPENDIX **E** FILE LAYOUTS FOR CASE STUDY DATA SETS **799**

Answers to Selected Exercises **802**

Credits **812**

Index **819**

Online

SAS Tutorial

SPSS Tutorial

MINITAB Tutorial

R Tutorial

PREFACE

Overview

This text is designed for two types of statistics courses. The early chapters, combined with a selection of the case studies, are designed for use in the second half of a two-semester (two-quarter) introductory statistics sequence for undergraduates with statistics or non-statistics majors. Or, the text can be used for a course in applied regression analysis for masters or Ph.D. students in other fields.

At first glance, these two uses for the text may seem inconsistent. How could a text be appropriate for both undergraduate and graduate students? The answer lies in the content. In contrast to a course in statistical theory, the level of mathematical knowledge required for an applied regression analysis course is minimal. Consequently, the difficulty encountered in learning the mechanics is much the same for both undergraduate and graduate students. The challenge is in the application – diagnosing practical problems, deciding on the appropriate linear model for a given situation, and knowing which inferential technique will answer the researcher’s practical question. This *takes experience*, and it explains why a student with a non-statistics major can take an undergraduate course in applied regression analysis and still benefit from covering the same ground in a graduate course.

Introductory Statistics Course

It is difficult to identify the amount of material that should be included in the second semester of a two-semester sequence in introductory statistics. Optionally, a few lectures should be devoted to Chapter 1 (A Review of Basic Concepts) to make certain that all students possess a common background knowledge of the basic concepts covered in a first-semester (first-quarter) course. Chapter 2 (Introduction to Regression Analysis), Chapter 3 (Simple Linear Regression), Chapter 4 (Multiple Regression Models), Chapter 5 (Model Building), Chapter 6 (Variable Screening Methods), Chapter 7 (Some Regression Pitfalls), and Chapter 8 (Residual Analysis) provide the core for an applied regression analysis course. These chapters could be supplemented by the addition of Chapter 10 (Time Series Modeling and Forecasting), Chapter 11 (Principles of Experimental Design), and Chapter 12 (The Analysis of Variance for Designed Experiments).

Applied Regression for Graduates

In our opinion, the quality of an applied graduate course is not measured by the number of topics covered or the amount of material memorized by the students. The measure is how well they can apply the techniques covered in the course to the solution of real problems encountered in their field of study. Consequently, we advocate moving on to new topics only after the students have demonstrated ability (through testing) to apply the techniques under discussion. In-class consulting sessions, where a case study is presented and the students have the opportunity to diagnose the problem and recommend an appropriate method of analysis, are very

helpful in teaching applied regression analysis. This approach is particularly useful in helping students master the difficult topic of model selection and model building (Chapters 4-8) and relating questions about the model to real-world questions. The seven case studies (which follow relevant chapters) illustrate the type of material that might be useful for this purpose.

A course in applied regression analysis for graduate students would start in the same manner as the undergraduate course, but would move more rapidly over the review material and would more than likely be supplemented by Appendix A (Derivation of the Least Squares Estimates), Appendix B (The Mechanics of a Multiple Regression Analysis), and/or Appendix C (A Procedure for Inverting a Matrix), one of the statistical software Windows tutorials available at the course website (SAS, SPSS, MINITAB, or R), Chapter 9 (Special Topics in Regression), and other chapters selected by the instructor. As in the undergraduate course, we recommend the use of case studies and in-class consulting sessions to help students develop an ability to formulate appropriate statistical models and to interpret the results of their analyses.

Features

1. **Readability.** We have purposely tried to make this a teaching (rather than a reference) text. Concepts are explained in a logical intuitive manner using worked examples.
2. **Emphasis on model building.** The formulation of an appropriate statistical model is fundamental to any regression analysis. This topic is treated in Chapters 4-8 and is emphasized throughout the text.
3. **Emphasis on developing regression skills.** In addition to teaching the basic concepts and methodology of regression analysis, this text stresses its use, as a tool, in solving applied problems. Consequently, a major objective of the text is to develop a skill in applying regression analysis to appropriate real-life situations.
4. **Real data-based examples and exercises.** The text contains many worked examples that illustrate important aspects of model construction, data analysis, and the interpretation of results. Nearly every exercise is based on data and research extracted from a news article, magazine, or journal. Exercises are located at the ends of key sections and at the ends of chapters.
5. **Case studies.** The text contains seven case studies, each of which addresses a real-life research problem. The student can see how regression analysis was used to answer the practical questions posed by the problem, proceeding with the formulation of appropriate statistical models to the analysis and interpretation of sample data.
6. **Data sets.** The online resource provides complete data sets that are associated with the case studies, exercises and examples. These can be used by instructors and students to practice model-building and data analyses.
7. **Extensive use of statistical software.** Tutorials on how to use any of four popular statistical software packages – SAS, SPSS, MINITAB, and R – are provided online. Printouts associated with the respective software packages are presented and discussed throughout the text.
8. **End-of-Chapter Summaries.** Important points are reinforced through flow graphs (which aid in selecting the appropriate statistical method) and boxed notes with key words, formulas, definitions, lists, and key concepts.

New to the 8th Edition

Although the scope and coverage remain the same, the eighth edition contains several substantial changes, additions, and enhancements:

1. **New and Updated Case Studies.** *Case Study 2: Modeling Sale Prices of Residential Properties*, has been updated with current data. A new case study (*Case Study 7: Voice Versus Face Recognition – Does One Follow the Other?*) now follows the chapter on analysis of variance.
2. **Real Data-based Exercises.** Many new and updated exercises, based on contemporary studies and real data in a variety of fields, have been added. Most of these exercises foster and promote critical thinking skills.
3. **Statistical Software Output.** All statistical software printouts shown in the text have been updated to reflect the most recent version of the software: Minitab, SAS, and SPSS.
4. **Updated Statistical Software Tutorials.** They can be found at the following website: www.pearson.com/math-stats-resources. The text's online resource provides updated instructions on how to use the Windows versions of SAS, SPSS, MINITAB, and R. Step-by-step instructions and screen shots for each method presented in the text are shown.
5. **Updated and New Sections in Chapter 9: Special Topics in Regression.** The section on logistic regression (Section 9.6) has been expanded. A new section (Section 9.7) on Poisson regression has been added. And, in addition to ridge regression, Section 9.8 now includes a discussion of Lasso regression.

Numerous less obvious changes in details have been made throughout the text in response to suggestions by current users of the earlier editions.

Supplements

The text is also accompanied by the following supplementary material:

1. **Instructor's solutions manual.** The instructor's exercise solutions manual presents the full solutions to the other half (the even) exercises contained in the text. For adopters, the manual is complimentary from the publisher.
2. **Data Files.** They can be found at the book's resource website: www.pearson.com/math-stats-resources. The text's online resource provides data files for all data sets marked with a data (📄) icon in the text. These include data sets for text examples, exercises, and case studies. The data files are saved in “.csv” format for easy importing into statistical software such as R, as well as in SAS (“.sas7bdat”), SPSS (“.sav”), and Minitab (“.mtw”) format.

Acknowledgments

We want to thank the many people who contributed time, advice, and other assistance to this project. We owe particular thanks to the many reviewers who provided suggestions and recommendations at the onset of the project and for the succeeding editions (including the 8th):

Jack Miller (University of Michigan)
Scott Grimshaw (Brigham Young University)

Liam O'Brien (Colby College)
Subarna K Samanta (The College of New Jersey)
Wolde Woubneh (Kean University)
Alan Huebner (University of Notre Dame)
Jen-Wen Lin (University of Toronto)
Karen Keating (Kansas State University)
Seamus Freyne (Mississippi State University)
Martin Tanner (Northwestern University)
Rebecca L. Pierce (Ball State University)
Julius Esunge (University of Mary Washington)
Brant Deppa (Winona State University)
Ross Hosky (Appalachian State University)
David Holmes (College of New Jersey)
Patrick McKnight (George Mason University)
David Kidd (George Mason University)
W.R. Stephenson (Iowa State University)
Lingyun Ma (University of Georgia)
Pinyuen Chen (Syracuse University)
Gokarna Aryal (Purdue University, Calumet)
Monnie McGee (Southern Methodist University)
Ruben Zamar (University of British Columbia)
Tom O'Gorman (Northern Illinois University)
William Bridges, Jr. (Clemson University)
Paul Maiste (Johns Hopkins University)
Mohammed Askalani, Mankato State University (Minnesota)
Ken Boehm, Pacific Telesis (California)
Andrew C. Brod, University of North Carolina at Greensboro
James Daly, California State Polytechnic Institute at San Luis Obispo
Assane Djeto, University of Nevada - Las Vegas
Robert Elrod, Georgia State University
James Ford, University of Delaware
Carol Ghomi, University of Houston
James Holstein, University of Missouri at Columbia
Steve Hora, Texas Technological University
K. G. Janardan, Eastern Michigan University
Thomas Johnson, North Carolina State University
Ann Kittler, Ryerson College (Toronto)
James T. McClave, University of Florida
John Monahan, North Carolina State University
Kris Moore, Baylor University
Farrokh Nasri, Hofstra University
Robert Pavur, University of North Texas

P. V. Rao, University of Florida
Tom Rothrock, Info Tech, Inc.
Ray Twery, University of North Carolina at Charlotte
Joseph Van Matre, University of Alabama at Birmingham
William Weida, United States Air Force Academy
Dean Wichern, Texas A&M University
James Willis, Louisiana State University

We are particularly grateful to Charles Bond, Evan Anderson, Jim McClave, Herman Kelting, Rob Turner, P.J. Taylor and Mike Jacob, who provided data sets and/or background information used in the case studies.

