Regression Analysis I: An Introduction

Saundra K. Schneider
Michigan State University

This course provides an introduction to the theory, methods, and practice of regression analysis. The goals are to provide students with the skills that are necessary to: (1) read, understand, and evaluate the professional literature that uses regression analysis; (2) design and carry out studies that employ regression techniques for testing substantive theories; and (3) prepare to learn about more advanced statistical procedures.

Any course of this type must assume a working knowledge of elementary statistical concepts and techniques. We will conduct a *brief* review at the beginning of the course, but students must be familiar with such ideas as descriptive statistics, sampling distributions, statistical inference, and hypothesis testing, before moving on to the more complicated matters that will comprise the majority of the course material. The course will not dwell on statistical theory. But, we will focus on the nature of the basic regression model, and the development of the regression estimators. We will see that this model depends very heavily on several assumptions. Therefore, we will examine these assumptions in detail, considering why they are necessary, whether they are valid in practical research situations, and the consequences of violating them in particular applications of the regression techniques. These formal, analytic treatments will be counterbalanced by the use of frequent substantive examples and class exercises. Again, the overall course objective is *not* to turn you into a statistician-- instead, we are trying to maximize your research skills as a social scientist.

Formal course requirements are as follows: (1) Class attendance and active participation. This is mandatory. Statistical knowledge is cumulative, and gaps in the early material will always have detrimental consequences later on. (2) Completion of class assignments. Most of these are computer exercises, designed to familiarize you with the application of various concepts and techniques introduced in class. Each of these assignments will focus on a specific set of topics. However, the latter assignments are cumulative in the sense that they build upon earlier material in the class.

The following are the ***recommended*** texts for the course:

Michael S. Lewis-Beck. *Applied Regression: An Introduction*.

Larry D. Schroeder, David L. Sjoquist, Paula E. Stephan. *Understanding Regression Analysis: An Introductory Guide*.

Damodar N. Gujarati and Dawn C. Porter. *Basic Econometrics*, 5th Edition.

The following books are useful ***reference*** books:

George W. Bohrnstedt, David Knoke, and Alissa Potter Mee. *Statistics for Social Data Analysis* (4th Edition).

Lawrence C. Hamilton. *Modern Data Analysis: A First Course in Applied Statistics*.

Thomas H. Wonnacott and Ronald J. Wonnacott. *Introductory Statistics.*

Neil A. Weiss. *Introductory Statistics* (9th Edition).

The following books are ***supplemental***:

William D. Berry. *Understanding Regression Assumptions.*

William D. Berry and Stanley Feldman. *Multiple Regression in Practice*.

Peter Kennedy. *A Guide to Econometrics* (6th Edition).

John Fox. *Regression Diagnostics.*

Jeffrey M. Wooldridge. *Introductory Econometrics: A Modern Approach* (3rd Edition).

Students should pay special attention to the readings in the **recommended** texts identified with asterisks in the syllabus. This material is critical for the course. It would be wise to read *all* the material assigned in the recommended texts and to purchase these texts for our own library. You should also have access to a basic reference book, such as Bohrnstedt/Knoke/Mee, Hamilton, Weiss, or Wonnacott and Wonnacott. Although these reference books are not required texts, they will prove useful for reviewing basic concepts and introductory material. And they will also provide reasonable alternative discussions of the bivariate and multiple regression models. Most of the supplemental books are either too specialized or advanced to be used as central texts in a course of this type. However, several of them are very good and would be extremely useful books to add to your own library.

After you have selected your texts, use the readings listed on the following pages to follow along with the material. You do NOT need to read all of the material in all the texts. But, it is wise to keep up with the readings in the recommended texts you have chosen.

Topics and Reading Assignments

I. Introduction to Regression Analysis

Reading: ***Gujarati and Porter, pp. 15-32

II. Preliminary Material and Statistical Review

A. Frequency Distributions, Univariate Summary Statistics, Probability Distributions

Reading: ***Gujarati and Porter, pp. 801-823

Hamilton, pp. 3-110
Bohrnstedt, Knoke, and Mee, pp. 27-92, 135-154
Wonnacott and Wonnacott, pp. 25-60, 109-116, 124-141
Weiss, pp. 2-231

B. Statistical Inference and the Properties of Statistical Estimators

Reading: *Gujarati and Porter, pp. 823- 837

Hamilton, pp. 241-259

1. Confidence Intervals & Hypothesis Tests

Reading:

Hamilton, pp. 260-354
Bohrnstedt, Knoke, and Mee, pp. 154-179
Wonnacott and Wonnacott, pp. 254-264, 287-297, 300-310, 314-317
Weiss, pp. 280-485

2. Differences Between Two Means, Two Variances, Etc.

Reading:

Hamilton, pp. 397-456
Bohrnstedt, Knoke, and Mee, pp. 187-212
Wonnacott and Wonnacott, pp. 265-273
Weiss, pp. 486-647

C. Linear Combinations

Reading:

Wooldridge, pp. 707-802

III. The Bivariate Regression Model

A. Introduction: Basic Ideas and Concepts

Reading:

*****Lewis-Beck, pp. 9-26
*****Schroeder, Sjoquist, and Stephan, pp. 11-23
*Gujarati and Porter, pp. 34-54

Hamilton, pp. 457-476
Berry, pp. 1-22
Bohrnstedt, Knoke, and Mee, pp. 253-266
Wonnacott and Wonnacott, pp. 357-370
Weiss, pp. 694-741

B. The Least Squares Criterion and Estimation in the Bivariate Regression Model

Reading:

*****Gujarati and Porter, pp. 55-61

Berry and Feldman, pp. 31-41
Hamilton, pp. 468-477
Bohrnstedt, Knoke, and Mee, pp. 266-274, 284-286
Wonnacott and Wonnacott, pp. 474-496
Kennedy, pp. 11-59
Wooldridge, pp. 50-66, 89-95, 106-109, 123-126, 176-181, 187-190

C. Goodness of fit, the Correlation Coefficient and R^2

Reading:

*Schroeder, Sjoquist, and Stephan, pp. 23-29
*Gujarati and Porter, pp. 73-94

Hamilton, pp. 477-483

D. Assumptions Underlying the Bivariate Linear Regression Model

Reading:

Berry and Feldman, pp. 9-12
Kennedy, pp. 11-59

E. Statistical Inference, Confidence Intervals, and Hypothesis Tests

Reading: ***Lewis-Beck, pp. 26-47
***Schroeder, Sjoquist, and Stephan, pp. 36-53
*Gujarati and Porter, pp. 107-147

Hamilton, pp. 503-525
Bohrnsted, Knoke, and Mee, pp. 277-284
Wonnacott and Wonnacott, pp. 372-395
Kennedy, pp. 51-90
Wooldridge, pp. 126-147
Weiss, pp. 742-797

F. Summary, Extensions, and a Preliminary Look at Residuals, Outliers, and Influential Cases

Reading: ***Gujarati and Porter, pp. 147-188

Hamilton, pp. 492-495, 535-551
Berry, pp. 22-88

IV. The Multiple Regression Model

A. Introduction: Notation, Assumptions, and Interpretation

Reading: ***Lewis-Beck, pp. 47-54
*Schroeder, Sjoquist, and Stephan, pp. 29-32
*Gujarati and Porter, pp. 188-195

Hamilton (*MDA*), pp. 563-566
Bohrnstedt, Knoke, and Mee, pp. 381-390
Wonnacott and Wonnacott, pp. 396-406
Berry and Feldman, pp. 9-18
Wooldridge, pp. 73-88

B **Measures of Goodness of Fit**

Reading: ***Schroeder, Sjoquist, and Stephan, pp. 32-36
***Gujarati and Porter, pp. 196-206

Bohrnstedt, Knoke, and Mee, pp. 392-396
Wonnacott and Wonnacott, pp. 496-501

C. Statistical Inference and the Role of Hypothesis Testing

Reading: *Gujarati and Porter, pp. 233-243

Hamilton, pp. 566-568
Bohrnstedt, Knoke, and Mee, pp. 396-409
Wonnacott and Wonnacott, pp. 406-408
Berry and Feldman, pp. 9-18
Kennedy, pp. 60-80
Wooldridge, pp. 147-167, 214-218

D. Summary and a Brief Look at Extensions

Reading: *Gujarati and Porter, pp. 243-277

Hamilton (*RWG*), pp. 83-101

V. Model Building in Multiple Regression Analysis

A. Models of Substantive Phenomena and the Importance of Model Assumptions

Reading: *Lewis-Beck, pp. 63-66

Hamilton, pp. 574-576
Wonnacott and Wonnacott, pp. 410-424
Berry, pp. 1-24

B. Model Specification

Reading: *Lewis-Beck, pp. 30-45
*Schroeder, Sjoquist, and Stephan, pp. 67-70
*Gujarati and Porter, pp. 467-522

Berry, pp. 30-45
Berry and Feldman, pp. 18-26
Kennedy, pp. 71-92

C. Nominal Independent Variables

Reading: *Schroeder, Sjoquist, and Stephan, pp. 56-58
*Gujarati and Porter, pp. 277-314

Hamilton, pp. 576-580
Bohrnstedt, Knoke, and Mee, pp. 409-419
Kennedy, pp. 248-258
Wooldridge, pp. 230-252

D. Functional Forms and Nonlinear Models

Reading: *Schroeder, Sjoguist, and Stephan, pp. 58-61
*Gujarati and Porter, pp. 523-540

Berry, pp. 60-66
Hamilton, pp. 583-584
Berry and Feldman, pp. 51-72
Kennedy, pp. 93-111
Wooldridge, pp. 304-390

VI. Potential Problems in Multiple Regression Analysis

A. Interpretation of Results

Reading:

Hamilton, pp. 568-573
Bohrnstedt, Knoke, and Mee, pp. 274-275, 390-392
Fox, pp.3-5

B. Multicollinearity and Its Effects

Reading: *Lewis-Beck, pp. 58-63
*****Schroeder, Sjoquist, and Stephan, pp. 71-72
*Gujarati and Porter, pp. 320-364

Wonnacott and Wonnacott, pp. 501-506
Hamilton, pp. 580-581
Berry, pp. 24-27
Berry and Feldman, pp. 37-50
Kennedy, pp. 192-202
Fox, pp. 10-21
Wooldridge, pp. 101-105

C. Nonnormal and Nonconstant (Heteroscedastic) Errors

Reading: *Schroeder, Sjoquist, and Stephan, pp. 75-77
*Gujarati and Porter, pp. 365-411

Berry and Feldman, pp. 73-88
Berry, pp. 67, 72-81
Fox, pp. 40-53
Kennedy, pp. 133-139
Wooldridge, pp. 181-185

D. Measurement Error

Reading: *Schroeder, Sjoquist, and Stephan, pp. 70-71
*Gujarati and Porter, pp. 524-528

Berry and Feldman, pp. 26-37
Berry, pp. 45-60
Kennedy, pp. 157-163
Wooldridge, pp. 318-325

E. Residual Analysis, Outliers, and Influential Observations

Reading: *Gujarati and Porter, pp. 496-497

Berry, pp. 27-29
Fox, pp. 21-40
Kennedy, pp. 372-388

VII. Additional Topics

A. Dichotomous Dependent Variables

Reading: *****Schroeder, Sjoquist, and Stephan, pp. 79-80
*****Gujarati and Porter, pp. 541-591

Wooldridge, pp. 252-258

B **Simultaneous Equation Models**

Reading: *****Schroeder, Sjoquist, and Stephan, pp. 77-79
*****Gujarati and Porter, pp. 671-688

C. A Brief Introduction to Time Series Models

Reading: ***Schroeder, Sjoquist, and Stephan, pp. 72-75
***Gujarati and Porter, pp. 737-772

Berry, pp. 67-72
Kennedy, pp. 139-156, 163-179