Course introduction and objectives
This is an introductory course in linear regression analysis, with a focus on the theory and application of the method. At the end of the course, students should be able to understand and critically evaluate research that uses regression, design and complete a study using regression, and have an understanding of regression that is a foundation for learning more advanced, related methods. The emphasis will be on successful application of linear regression and understanding the uses of the method for social science research.

The course assumes that students are familiar with basic statistical concepts, such as levels of measurement, descriptive statistics, sampling distributions, statistical inference, and hypothesis testing. After briefly reviewing these foundations, the course will focus first on bivariate regression (one independent variable and one dependent variable) before moving on to multivariate regression (multiple independent variables and one dependent variable). Once we have covered the basics of the method, we will discuss issues related to model specification or model building and assessing overall fit. Then, we will examine in more detail the assumptions of linear regression with an emphasis on how these assumptions affect our applications of these methods in practice. We will discuss problems that are most often associated with different types of datasets, and strategies for addressing those problems. Finally, the last class will explore how linear regression is related to more advanced statistical methods, most of which build on linear regression by relaxing some of the assumptions or allowing for more complex data or error structures.

To be successful in the course, students will not need to be mathematicians or statisticians, but they will need a desire to learn, to think analytically, to solve problems, and be open to new ways of thinking. Students will also need basic algebra skills and basic computer skills.

Students will complete a series of homework assignments that apply these methods using statistical software. The course is designed to teach students about linear regression, not how to use any particular software for regression. Indeed, students who master course material should be able to pick up new software (particularly software with a menu-driven interface, though we will emphasize the value of writing command syntax files) quite quickly in the future. Consequently, though students will use statistical software to complete assignments, assignments will include the required software commands, and class time will not emphasize particular software features. Lecture materials will include instructions and examples using SPSS and Stata, and support will be available for students who are already familiar with basic data analysis functions in R. NOTE: SP participants can install a free copy of Stata on their personal computers for use this summer.

Readings The latter two texts may be available electronically through your home academic library

Background reference texts for basic inferential statistics (prior to this course):


Other texts that cover material in this course (for reference or future use)


Texts that cover linear regression and related methods (usually in more depth or at a level more advanced than this course):


Evaluation
There will be three homework assignments for the course. They will be posted to the CTools site (https://ctools.umich.edu/gateway/).
• Homework 1 (applied bivariate regression)
• Homework 2 (applied multivariate regression)
• Homework 3 (multivariate regression diagnostic)

Course Outline
1. Introductions, review & where we are headed (Week 1: Monday)
Overview & background material:

2. Bivariate regression: Introduction, estimation, and measuring fit (Week 1: Tuesday)
Lewis-Beck, pp. 9-26
Schroeder, Sjoquist, and Stephan, pp. 11-29
Gujarati and Porter (Essentials of Econometrics), pp. 19-38, 38-52

Lewis-Beck, pp. 26-54
Schroeder, Sjoquist, and Stephan, pp. 23-32, 36-53
Gujarati and Porter (Essentials of Econometrics), pp., 53-102, 104-107
Applications:

4. Multiple regression: Measuring fit, model assumptions (Week 1: Thursday)
Lewis-Beck, pp. 63-66
Schroeder, Sjoquist, and Stephan, pp. 32-36

5. Nominal independent variables, nonlinear models & model specification (Week 1: Friday)
Gujarati and Porter (Essentials of Econometrics), pp. 112-113; 132-215; 219-244
Lewis-Beck, pp. 30-45
Schroeder, Sjoquist, and Stephan, pp. 56-61, pp. 67-70
6. Problems: Multicollinearity, nonnormal and nonconstant errors, measurement error (Week 2: Monday)
Lewis-Beck, pp. 58-63
Schroeder, Sjoquist, and Stephan, pp. 70-72, 75-77
Gujarati and Porter (Essentials of Econometrics), pp. 229-230, 245-311

7. Problems: Residual analysis, outliers, and influential observations (Week 2: Tuesday)
Gujarati and Porter (Essentials of Econometrics), pp. 234-235; 283-285

8. July 4th holiday (Week 2: Wednesday)

9: Review of regression assumptions, model specification, and diagnostics (Week 2: Thursday)

10: Variations on linear regression (Week 2: Friday)
Schroeder, Sjoquist, and Stephan, pp. 72-80
Gujarati and Porter (Essentials of Econometrics), pp. 201-205, 347-401