Regression Analysis I
ICPSR Summer Program: Session I

Instructor: Shahryar Minhas
Email: minhassh@msu.edu
Instructor Office Hours: W F: 1pm-2pm EST

Lecture: 8:30-12:00 PM EST
Location: LSA 3254
Zoom: https://umich.zoom.us/j/99576553842

Course Description

This is a course on quantitative methodology, by which we mean a course on the application of statistical methods to problems in political science, with a focus on linear regression. Specifically, the course will provide an introduction to how regression works, practical applications of the method, including inference and writing about the results. Focus of the course will be on ordinary least squares regression, which describes the relationship between one or more independent variables and a dependent variable.

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. Course also assumes familiarity with basic algebra and have had some exposure to courses or work that have provided experience in connecting data to research questions.

Inclusion and diversity are core values of my classroom. I am committed to creating a safe, supportive, and welcoming environment where all students can pursue academic and personal success. We all deserve each other's respect, support, recognition, and protection.

Maintaining a respectful and inclusive community requires vigilance. We must, therefore, all stand up against derogatory and discriminatory language or actions whenever we see them. It is essential that we all work together to foster an inclusive community.

Asynchronous Lectures

Class will be taught in a hybrid format where we will have students attending in-person, via zoom, and then others who might just rely on the recordings. Zoom link for all lectures is: https://umich.zoom.us/j/99576553842. Lectures are automatically recorded to the cloud and will be made available on Canvas by the end of the day of that lecture.

Class Communications

I will try to be very responsive to any questions that you have over email. In addition, I've set up a Slack channel for the course. Here is an invite link: https://join.slack.com/t/slack-5wv3910/shared_invite/zt-2kdbpckuf-fF3PSL6_ZHSF1bAi1zEnWA.

Learning Objectives

This course is designed as a series of modules that build upon each other, by the end of the course you should be able to do the following:

- Estimate a linear regression model and interpret its output
- Use linear regression to test various hypotheses regarding a substantive claim
- Begin to write about matters related to linear regression, using accessible, non-technical, but precise language, particularly with regards to hypothesis testing

Acknowledgments

The lectures and homeworks in this course make use of modified materials that I have prepared from previous courses, but some of the homework materials and structures benefited from conversations with Shawna Metzger and materials from Edmund Malesky.

Required Materials

Books

This course has two suggested books (if you need help getting access to them just let me know):

- Introductory Econometrics: A Modern Approach 6 Edition by Jeffrey Wooldridge (abbreviated IE)
- Regression Analysis: A Practical Introduction 1st Edition by Jeremy Arkes (abbreviated RA)

Statistical Software

We'll be using the R programming language to learn linear regression. You will need to install R (the language) and RStudio (a program for writing, testing, and publishing R code). The course does expect that you have some introductory R knowledge.

To download R, go to https://cran.r-project.org/ and click the link for your operating system under the section titled "Download and Install R". Then follow the instructions to install the version appropriate for your specific system version.

To download RStudio, go to https://www.rstudio.com/products/rstudio/download/#download and click the link for your operating system. This will download an installer file. Once the installer is downloaded, double click it to run, and then follow the instructions to complete the installation. This will be covered in the first video lecture, as well.

Grading

Students will have a weekly homework (after the first week) typically assigned on Monday and due on Friday. Students can compile their homeworks using RMarkdown, TeX, or Word, but the submitted file must be a in a .pdf format.

Grades will be assigned based on performance on the homework assignments. Each assignment will be graded (0–no submission, 1–does not demonstrate comprehension of the methodology, 2–adequate comprehension demonstrated, 3–excellent comprehension demonstrated). The final grade will be calculated using the following grading scheme to the sum of homework grades.

A+	≥ 9
A	≥ 8
A-	≥ 6
B+	≥ 4
B-	< 4

Late assignments will not be accepted.

Course Schedule

Find below an overview of the course, including the topics of lecture and assigned readings. Based on how fast or slow we are going we might cover additional materials (based on class interest and what I can prepare) or less. The below is just to give you a sense of what we're hoping to accomplish, I will always try and tailor the pace of a class to the needs of the students.

Note that links have been provided to all readings that are not from Wooldridge or Arkes.

Course Introduction, Bit on measurement, Setting up R

To start to get familiar with R, you should take a look at a programming camp I host for incoming MSU graduate students, site is available here: http://s7minhas.com/msuProgCamp/timeline.html. You should be able to download all the materials associated for each topic. I will be discussing R in class as well.

Learning goals:

- How do we go from a concept to a variable
- What are common types of variables that arise
- How do we perform descriptive analysis
- Where does probability fit in

Readings:

- IE:
  - Chapter 1

Recommended:

- RA:
  - Chapter 1
- MSU Progamming Camp:
  - Day 1
  - Day 2

Univariate Hypothesis Testing

Learning goals:

- How do we determine whether there is a statistical difference between groups in our data
- How do we structure a hypothesis test
- How do we interpret p-values

Readings:

- IE:
  - Appendix B
  - Appendix C

Recommended:

- MSU Progamming Camp:
  - Day 3

Glimpsing at OLS

Learning goals:

- What is the problem that we are trying to solve when we estimate an OLS model
- What goes on the left and what goes on the right hand side of a regression equation (i.e., what is an independent variable, what is a dependent variable)
- OLS is a way to draw fancy lines through scatters of points

Readings:

- IE:
  - Start Chapter 2

Recommended:

- RA:
  - Chapter 2
- MSU Progamming Camp:
  - Day 4-7

Deriving OLS

Learning goals:

- How exactly are we estimating the "effect" that an independent variable has on a dependent variable
- What is the minimization problem in OLS
- Brief foray into how we can solve minimization problems via some calculus
- How do we accomplish this minimization
- Interpret in words the OLS equation for $\hat{\beta}$

Readings:

- IE:
  - Continue Chapter 2

Recommended:

- RA:
  - Chapter 2
- MSU Progamming Camp:
  - Day 4-7

OLS Assumptions and Standard Errors

Learning goals:

- Now that we have estimated the effect of our indepent variable on a dependent variable, what is the uncertainty around our estimated effect
- What extra assumptions do we need to be able to estimate the uncertainty of our $\hat{\beta}$ estimates
- What's the procedure for estimating the uncertainty
- Interpret in words the OLS equation for $se(\hat{\beta})$

Readings:

- IE:
  - Finish Chapter 2

Recommended:

- RA:
  - Chapter 5
- MSU Progamming Camp:
  - Day 4-7

Hypothesis Testing in OLS

Learning goals:

- How do we do hypothesis testing in the context of OLS

Readings:

- IE:
  - Start Chapter 5
  - Start Appendix D
  - Start Appendix E

Recommended:

- RA:
  - Chapter 4
- Eggers et al. (2021), "No evidence for systematic voter fraud". PNAS.
- Schrodt (2014), "Seven Deadly Sins". Journal of Peace Research.

Multiple Regression

Learning goals:

- Rarely will it be the case that we just have one independent variable, how do we generalize OLS to multiple independent variables
- Brief foray into linear algebra
- How does the minimization problem change?
- Interpret again in words the OLS equation for $\hat{\beta}$

Readings:

- IE:
  - Continue Chapter 5
  - Continue Appendix D
  - Continue Appendix E

Recommended:

- RA:
  - Chapter 4
- Eggers et al. (2021), "No evidence for systematic voter fraud". PNAS.
- Schrodt (2014), "Seven Deadly Sins". Journal of Peace Research.

Omitted Variables

Learning goals:

- What happens if the true data generating process for our dependent variable is: $democracy = \beta_0 + \beta_1 * income + \beta_2 * literacy$ but we leave out literacy?
- Are we able to obtain an accurate estimate of β_1

Readings:

- IE:
  - Finish Chapter 5
  - Finish Appendix D
  - Finish Appendix E

Recommended:

- RA:
  - Chapter 4
- Eggers et al. (2021), "No evidence for systematic voter fraud". PNAS.
- Schrodt (2014), "Seven Deadly Sins". Journal of Peace Research.

Dummy variables and interactions

Learning goals:

- What is the role of dummy variables in regression?
- What if the effect that an independent variable has on the dependent variable is actually contingent on a third variable?

Readings:

- IE:
  - Start Section 6.2
  - Start Sections 7.1-7.7
  - Start Section 14.1

Recommended:

- Brambor et al. (2006), "Understanding Interaction Models". Political Analysis.
- Hainmueller at al (2019), "How Much Should We Trust Estimates from Multiplicative Interaction Models¿'. Political Analysis.

Dummy variables and interactions (cont'd)

Learning goals:

- What is the role of dummy variables in regression?
- What if the effect that an independent variable has on the dependent variable is actually contingent on a third variable?

Readings:

- IE:
  - Finish Section 6.2
  - Finish Sections 7.1-7.7
  - Finish Section 14.1

Recommended:

- Brambor et al. (2006), "Understanding Interaction Models". Political Analysis.
- Hainmueller at al (2019), "How Much Should We Trust Estimates from Multiplicative Interaction Models¿'. Political Analysis.

Model evaluation

Learning goals:

- How do we assess the predictive accuracy of our models
- When should we care about the predictive accuracy of our models

Readings:

- RA:
  - Section 7.1 - 7.2

Recommended:

- Ward et al. (2010), "The perils of policy by p-value". Journal of Peace Research.
- Colaresi & Mahmood (2017), "Do the robot". Journal of Peace Research.

Measurement error

Learning goals:

- Rarely will it be the case in the social sciences that we can operationalize our variables without some error, what is measurement error
- What are the consequences of measurement error
- When do we need to worry about measurement error

Readings:

- IE:
  - Chapter 9

Heteroskedasticity & Endogeneity

Learning goals:

- What are these fancy sounding words (heteroskedasticity and endogeneity), how do they manifest in OLS, how can we diagnose them, and can we fix them

Readings:

- IE:
  - Chapter 8
- RA:
  - Start Chapter 6