STAT 404
Regression for Social and Behavioral Research
Spring 2017

Prerequisite: STAT 301 or STAT 326 or STAT 401.

Lecture: Tu Th 11:00–11:50 pm, Town 0280
Lab: F 2:10–4:00pm, Gilman 2272

Instructor: Professor Olga Chyzh, ochyzh@iastate.edu
Office Hours: Th 3:00–5:00 pm and by appointment, Snedecor Hall 1413

Teaching Assistant: TBA
Office Hours: TBA

Overview and Objectives

This course starts with an in-depth treatment of OLS regression, and then proceeds to extend standard normal linear to generalized linear models, such as logistic regression, probit, multinomial logit and probit, poisson regression, and beyond. It builds on the students' previous knowledge and understanding of basic concepts in statistics, including moments, distributions, expected values, hypothesis testing, and multiple regression analysis. The objective of the class is to build a solid foundation in regression analysis, as it applies to social and behavioral research. Besides regular homework assignments, there will also be computer assignments intended to highlight the basic theoretical concepts in the context of both real-world and simulated data.

Learning Outcomes

Upon successful completion of this course, students should be able to:

1. Translate political phenomena into mathematical notation.
2. Understand the value and limitations of generalized linear modeling.
3. Given a data generating process, select an appropriate statistical model and method.
4. Test substantive hypotheses using generalized linear models.
5. Interpret a variety of types of model estimates.
6. Describe the assumptions of generalized linear models and address violations of them where possible.
7. Use `R` to import, manipulate and describe data, implement models, conduct diagnostics and sensitivity analysis, and produce publication-quality figures.

Requirements

Grades will be based on class attendance and participation (10%), lab assignments (30%), and three examinations (60%).

Attendance and Participation: 10%
Attendance is required. Students' comprehension of the lecture material is monitored using clicker questions. The purpose of the clicker questions is to monitor student effort and progress, *not* to penalize students who need extra time to learn complicated content. Therefore, each clicker question are awarded 90% of the points for attendance and 10% for giving the correct answer. All clicker questions are administered using Top Hat classroom response system (see section on required materials below). Each clicker question makes up only a small portion of the final grade, which means that missing a class or two will not make a meaningful difference on a student grade. The only excused absences are those due to university-approved activities (conditional on **advance** written warning from the relevant department) or a major life-altering event, such as a serious illness or a death in the family. Missed clicker questions, due to any other reasons, will result in a score of 0.

Lab Assignments: 30%
There is a total of 10 lab assignments. These assignments become available on Blackboard every Friday at 4 pm (after lab) and are due the following Friday at 2:00 pm (before lab) via Blackboard. Lab assignments consist of questions from both lectures and labs. The answer key to each assignment will become available at 11:59 pm on the day that the assignment is due. The students are expected to review the key and seek clarification should any questions arise. After the key is posted, no late assignments will be accepted for any reason. The only exceptions/arrangements will be made for cases of extreme adversities, e.g. hospitalization. Job interviews and/or academic conferences do not warrant an extension, as it is the student's responsibility to account for planned events in their schedule.

Examinations: 60%
The course includes three examinations: two midterms and a final. Each examination is worth 20% of the final grade. The exams are administered during lab time. Exams are technically non-cumulative (questions on later exams are not directly testing the knowledge of the material from previous exams), yet may assume reasonable mastery of the preceding material. For example, students must know how to calculate standard errors (exam 1 material) in order to interpret the results of a Poisson regression (exam 3 material).

Exam Dates:

- Friday, September 29, 2:10-4:00 pm;
- Friday, October 27, 2:10-4:00 pm;
- Monday, December 11, 9:45-11:45 am.

Labs:
During weekly labs, the instructor and the teaching assistant will go over additional problems and examples, as well as introduce students to the basics of programming in R statistical

software. Mastery of R programming will be regularly assessed as part of the lab assignments and exams.

Required Materials:

- Monogan, James E. III. Political Analysis Using R. Available online from http://link.springer.com/book/10.1007/978-3-319-23446-5 (free to download when connected to ISU network).
- We will be using the Top Hat (www.tophat.com) classroom response system in class. You will be able to submit answers to in-class questions using smartphones, tablets, laptops, or through text messaging. You can visit the Top Hat Overview (support.tophat.com/hc/en-us/articles/200019034-Top-Hat-Overview-Getting-Started) within the Top Hat Success Center which outlines how you can register for a Top Hat account, as well as provides a brief overview to get you up and running with the system. An email invitation to join your Top Hat space will be sent to you, but if you dont receive it, you can still create your student account at tophat.com and join the class using join code 865320. You will be required to purchase a Top Hat license from the ISU Book Store or online in order to access any quizzes or questions your instructor creates in the Top Hat system. Should you require assistance with Top Hat at any time please contact the IT Solution Center at 515-294-4000 or solution@iastate.edu.

Recommended Materials:

- Moore, Will H. and David A. Siegel. A Mathematical Course for Political and Social Research.
- Long, Scott J. Regression Models for Categorical and Limited Dependent Variables.
- Greene, William H. Econometric Analysis
- Tilman M. Davies. The Book of R: A First Course in Programming and Statistics.

Grading Scale:

Final grades are assigned in accordance with the following scale:

A	$\geq 93\%$
A-	$\geq 90\%$
B+	$\geq 87\%$
B	$\geq 80\%$
see below	$< 80\%$

Students that score $< 80\%$, but demonstrate reasonable effort to complete **every** part of the course, including attendance/participation, will be awarded a grade of B-.

Students who score $< 80\%$ and *fail* to demonstrate reasonable effort to complete **every** part of the course, including attendance/participation, will be awarded a grade of C or below.

Students that fail to complete 2 or more assignments will be assigned a grade of F in the course. Lab assignments are intended to help the instructor monitor students' continuous effort to internalize the material, which is the key to passing the course.

Administrative Issues

Missing an Exam: I expect at least **two weeks** advance notification that you will miss an exam (i.e. participating in university sponsored activity). You must arrange with me to make up the work as soon as possible. If you are physically unable to contact me in advance (e.g., if you wake up extremely ill), you must contact me (via email) as soon as possible and provide written documentation in order to make up the exam (e.g., a note from a physician).

Grade Complaints: If for some reason a student is unhappy with their grade on an exam, they may submit their complaint in writing, explain the particular discrepancy, and recommend an appropriate recourse. The instructor will read the memo, re-read the disputed answer, and then assign a new grade. Note that the instructor reserves the right to assign a lower grade after re-reading the answer. Approximately 50% of grade complaints in my Fall 2016 section of this course resulted in the same or lower grade.

Academic Integrity: All students in attendance at Iowa State University are expected to be honorable and to observe standards of conduct appropriate to a community of scholars. Academic misconduct includes all acts of dishonesty in any academically related matter and any knowing or intentional help or attempt to help, or conspiracy to help, another student. The Academic Misconduct Disciplinary Policy will be followed in the event of academic misconduct.

Examples of academic dishonesty include-but are not limited to-the following:

- Plagiarism—that is, using another's ideas or writings without proper attribution, including sources from the internet;
- Copying from another person during an examination;
- Assisting another person with an individual assignment or an exam.

Disability Statement: If you are registered with the Office of Disability Services, please make an appointment with me as soon as possible to discuss any course accommodations that may be necessary. *It is your responsibility to do this in a timely manner.*

I reserve the right to modify the syllabus to reflect the pace of the course.

Course Outline

Introduction, OLS Regression and Its Interpretation (Week 1: Aug. 20)

- This week is an overview of OLS regression, which includes calculating and interpreting its coefficients and standard errors, testing hypothesis related to the coefficients, and assessing model fit
- *Central Concepts:* OLS regression, sum of squares, regression coefficients, standard errors, coefficient of determination, predicted values.
- *Lab 1: Visualizing Data (Monogan Ch. 1–4).*

Mathematical Derivation of Least Squares in Scalar Form (Week 2: Aug. 27)

- Where do the formulas for the regression coefficient and intercept come from? The students will be introduced to the theory behind finding the least squared errors.
- *Central Concepts:* least squared errors, optimization.
- *Lab 2: Simple Regression (Monogan Ch 5.3).*

The Mathematical Derivation of Least Squares in Matrix Form (Weeks 3-4: Sept. 3, 10)

- The students will be introduced to a number of concepts in matrix algebra, which will culminate in a derivation of regression coefficients in matrix form.
- *Central Concepts:* matrix manipulations, eigenvalues, eigenvectors, Cholesky decomposition, identity matrix.
- *Lab 3: Multiple Regression I.*
- *Lab 4: Multiple Regression II.*

OLS via the Maximum Likelihood Approach (Weeks 5-6: Sept 17, 24)

- Students will be introduced to the idea of maximum likelihood, which will be applied to OLS.
- *Central Concepts:* maximum likelihood, the joint probability distribution, the logarithm of the likelihood.
- *Lab 5: Maximum Likelihood.*
- **Exam 1 on Friday, September 29, 2:10-4:00 pm.**

Logistic Regression (Weeks 7-8: Oct. 1, 8)

- Many social science applications involve binary dependent variables, such as winning an election, initiating an international conflict, being diagnosed with depression. As the OLS assumptions require a continuous dependent variable, we can not use OLS to model such outcomes. Instead, we model binary/dichotomous outcomes using a logistic regression.
- *Central Concepts:* logistic transformation, odds vs. probabilities, predicted probabilities, first differences.
- *Lab 6: Logistic Regression.*

Probit Regression (Weeks 9-10: Oct. 15, 22)

- Probit regression is an alternative to logistic regression that relies on a normal transformation rather than a logistic transformation.
- *Central Concepts:* the normal transformation.
- *Lab 7: Probit.*
- **Exam 2 on Friday, October 27, 2:10-4:00 pm.**

Multinomial Logit (Week 11: Oct. 29)

- Some outcome variables are measured on a categorical scale with more than two outcomes. For example, an voter may cast a vote for the Democratic, Republican, or a third-party candidate. We will learn how to treat these types of outcomes within a regression context.
- *Central Concepts:* multinomial regression, independence of irrelevant alternatives.
- *Lab 8: Multinomial Logit.*

Regression with an Ordinal Dependent Variable (Week 12: Nov. 5)

- Some outcome variables are measured on a categorical scale with more than two outcomes. For example, an survey respondent may "strongly disagree," "disagree," "agree," or "strongly agree" with a statement. We will learn how to treat these types of outcomes within a regression context.
- *Central Concepts:* parallel slopes assumption.
- *Lab 9: Regression with Ordered Outcomes.*

Regression for Count Data (Weeks 13-14: Nov. 12, 19)

- Students will learn how to model count data, such as the number of attacks, the number of endorsements, the number of patients diagnozed with a disease, etc.
- *Central Concepts:* count data, poisson regression, negative binomial regression.

Special Topics: Regression with Interactions and Squared Terms (Week 15: Nov 26)

- Effects of an explanatory variables may be conditional on the values of another explanatory variable. For example, the effect of exercise on weight may be conditional on gender: i.e. males may be able to lose weight as a result of exercise faster than females. We can model such effects using interaction terms. This section of the course focuses on the interpretation of regression coefficients for interactive variables.
- *Central Concepts:* interaction term, constitutive terms, non-linear terms, marginal effects, tables and graphs of marginal effects, confidence intervals around marginal effects.
- *Lab 10: Interactions (Monogan Ch. 6.2).*

Wrap-up and Review (Week 16: Dec. 3)

- The instructor will hold a review session for the final exam.
- *Lab: Additional review session held by the TA.*

Final Examination: Monday, December 11, 9:45 - 11:45 am.