THE OHIO STATE UNIVERSITY

COLLEGE OF ARTS AND SCIENCES

STAT 6450 – APPLIED REGRESSION ANALYSIS SUMMER 2021

Instructor

Instructor: R. Scott Linder
Email address: linder.5@osu.edu
Lectures: On CarmenZoom, Tuesdays and Thursdays, 12:40 – 2:40 pm
Office hours: On Carmen Zoom (Tuesdays and Thursdays 3 – 4 pm)

Prerequisite or corequisite

Statistics 6201 or equivalent.

Grader or Teaching Assistant

Lun Li (li.9515@buckeyemail.osu.edu)

Course delivery

The course will meet **synchronously (live) over CarmenZoom during class meeting times**.
CarmenZoom Link: https://osu.zoom.us/j/3875850653?pwd=Q00rbTZzOTA2amNucVJnTTlRNHEvZz09

Any asynchronous content will be delivered by recordings or documents posted on the class CarmenCanvas site:

Please feel free to ask questions or comment during class meetings. Please have your camera turned and your microphone muted. If you need to miss a class, you can watch a recording of the class meeting missed, posted with a delay.

Office hours will be held via CarmenZoom (using the link above) during times listed above.

Course description

Stat 6450 introduces statistical regression analysis techniques with focus on basic theory of linear models and (especially) on the application of regression models in practice. Topics include:

- Simple Linear Regression (SLR) model**, including model fitting, inference about model parameters, prediction, diagnostics for checking conditions and remedial measures for addressing problems, transformations, visualization.**
- Multiple Linear Regression (MLR) model, including model fitting, matrix algebra notation and theory, inference about model parameters, categorical predictors, prediction, the ANOVA sums of squares and degrees of freedom, piecewise models, diagnostics and condition verification, remediation for addressing problems with assumptions

- Variable selection and model building, stepwise regression methods, LASSO regression, ridge regression, polynomial regression and cross-validation.
- Generalized linear models, including logistic regression and Poisson regression models, inference about model parameters, prediction and verifying conditions.

Course learning outcomes

By the end of this course, students should:

- Understand how applications motivate regression analysis
- Understand theoretical assumptions underpinning the linear model and their importance in properly conducting a regression analysis
- Know how to estimate parameters in regression models
- Be able to validate the modeling assumptions with formal tests and visual diagnostic tools
- Know how to make inferences regarding the linear model
- Be able to build and validate regression models in a principled manner
- Be able to apply the above knowledge and techniques in on your own data or problems

Course materials

Required textbook

Applied Linear Regression, 4th edition, by Sanford Weisberg.

Available here: https://library.ohio-state.edu/record=b8665795~S7

Necessary equipment

- Computer with high-speed internet connection, webcam and microphone.

Necessary software

- The statistical software package called ***R*** (free http://www.r-project.org/).
    - Download *R* for Windows, Mac, and Linux, from the CRAN archive at https://cran.r-project.org.
    - An in-depth introduction to *R* is available at http://cran.r-project.org/doc/manuals/R-intro.pdf
    - Hands-on tutorials are available in the Swirl system, which you can learn about at http://swirlstats.com/. In particular, "*R* Programming: The basics of programming in *R*" is an appropriate first tutorial for students who have never used *R*.
    - An easier to use interface to *R* is available in the software package *RStudio*. This package is available for Windows, Mac, and Linux and can be downloaded for free from http://rstudio.org. **Note that RStudio requires *R* to be installed.**
- Microsoft Office 365 ProPlus Ohio State students are eligible for free through Microsoft's Student Advantage program.
    - Provides access to Word, Excel, PowerPoint, Outlook and other programs, depending on platform. Users also receive 1 TB of OneDrive for Business storage.
    - Office 365 is installed within your BuckeyeMail account. Full instructions for downloading and installation can be found https://ocio.osu.edu/kb04733.

Course technology

For help with technology issues, questions, or requests, contact the OSU IT Service Desk. Standard support hours are available at https://ocio.osu.edu/help/hours.

- **Self-Service and Chat support:** http://ocio.osu.edu/selfservice
- **Phone:** 614-688-HELP (4357)
- **Email:** 8help@osu.edu
- **TDD:** 614-688-8743

Course Grades

Your course average will be computed using the following weights. Assignment, project and exam grades will be posted to the CarmenCanvas page.

Homework Assignment Average	30%
Midterm Exam	25%
Final Project	20%
Final Exam	25%
	100%

Your course grade will be determined from your course average. Exact cutoffs for determining final course grades will not be determined until final course averages have been computed. However, here are some guidelines:

If your course average is at least 92%, your course grade will be A.
If your course average is at least 90%, your course grade will be at least A–.
If your course average is at least 82%, your course grade will be at least B
If your course average is at least 80%, your course grade will be at least B–

Note: If your course average is slightly less than a bound given above, you may still earn the higher letter grade. The bounds above are viewed sufficient and maybe not necessary.

Assignment information

Homework: Homework will be assigned regularly. Expect approximately 5 assignments consisting of some written problems and some computer programming/data analysis problems. You are encouraged to work together on the problems, but each student must hand in their own work, written in their own words. Your lowest homework assignment score will be dropped**. Late submission of assignments will not be accepted without prior agreement.** If you are unable to complete an assignment on time, please communicate with me.

All assignment submissions must be submitted online as a single PDF file through the CarmenCanvas page. Please be sure that the questions are clearly labeled, all supporting work (including edited and easy-to-read computer code and output) can be easily identified, and that all figures/tables are referenced and interpreted in the text. Numbers from the output you refer to in your discussion should be highlighted or circled.

Exams: There will be two exams -- a midterm exam and a final exam. Coverage includes lecture material, assigned reading, and homework. All exams are closed book/closed notes and will be delivered through carmen and proctored online through CarmenZoom. Further details will be given in advance of each exam. See the calendar included with this syllabus for tentative exam dates.

Formulas will be provided, and **s**tatistical tables will be provided as needed. You will need a calculator, but no communication devices are allowed (e.g. mobile phones). Makeup exams will be possible if you contact me early (preferably before the exam), provided you can document a valid reason for missing.

Project: A data-analysis group class project will consist of an oral presentation of results obtained from analyzing a dataset, and will require use of the *R* software. Students are expected to work in groups of 2-3 and each group can select a dataset of their choice to analyze (preferably from the research/educational domain of the student). No written report is required. **Groups will present their results to the class on the last class meeting date prior to the final exam date.** Presentations may be pre-recorded. Information and guidelines about projects will be provided several weeks into the semester.

Health and safety

The Ohio State University Wexner Medical Center's Cornavirus Outbreak site (https://wexnermedical.osu.edu/features/coronavirus) includes the latest information about COVID-19 as well as guidance for students, faculty and staff. Guidelines and requirements for campus safety from the University's COVID-19 Transition Task Force were published on July 1 on the Safe and Healthy website (https://safeandhealthy.osu.edu).

Student academic services

Student academic services offered on the OSU main campus http://advising.osu.edu/welcome.shtml.

Student support services

Student support services offered on the OSU main campus http://ssc.osu.edu.

Academic integrity policy

Policies for this online course

- **Exams**: You must complete the midterm and final exams yourself, without any external help or communication, and using only materials explicitly authorized by the instructor.
- **Written assignments**: Written assignments, including discussion posts, should be your own original work.

Ohio State's academic integrity policy

It is the responsibility of the Committee on Academic Misconduct to investigate all reported cases of student academic misconduct. The term "academic misconduct" includes all forms of student academic misconduct wherever committed; illustrated by, but not limited to, cases of plagiarism and dishonest practices in connection with examinations. Instructors shall report all instances of alleged academic misconduct to the committee (Faculty Rule 3335-5-487). For additional information, see the Code of Student Conduct http://studentlife.osu.edu/csc/.

Copyright disclaimer

The materials used in connection with this course may be subject to copyright protection and are only for the use of students officially enrolled in the course for the educational purposes associated with the course. Copyright law must be considered before copying, retaining, or disseminating materials outside of the course.

Statement on title IX

Title IX makes it clear that violence and harassment based on sex and gender are Civil Rights offenses subject to the same kinds of accountability and the same kinds of support applied to offenses against other protected categories. If you or someone you know has been sexually harassed or assaulted, you may find the appropriate resources at http://titleix.osu.edu or by contacting the Ohio State Title IX Coordinator at titleix@osu.edu

Accessibility accommodations for students with disabilities

The university strives to make all learning experiences as accessible as possible. If you anticipate or experience academic barriers based on your disability (including mental health, chronic, or temporary medical conditions), please let me know immediately so that we can privately discuss options. To establish reasonable accommodations, I may request that you register with Student Life Disability Services. After registration, make arrangements with me as soon as possible to discuss your accommodations so that they may be implemented in a timely fashion. SLDS contact information: slds@osu.edu; 614-292-3307; http://slds.osu.edu; 098 Baker Hall, 113 W. 12th Avenue.

Accessibility of course technology

This online course requires use of Carmen (Ohio State's learning management system) and other online communication and multimedia tools. If you need additional services to use these technologies, please request accommodations with your instructor.

- Carmen (Canvas) accessibility
- Streaming audio and video
- Synchronous course tools

Your mental health

As a student you may experience a range of issues that can cause barriers to learning. The Ohio State University offers services to assist you with addressing these and other concerns you may be experiencing. If you or someone you know are suffering from possible mental health issues, you can learn more about the broad range of confidential mental health services available on campus via the Office of Student Life's Counseling and Consultation Service (CCS) by visiting ccs.osu.edu or calling 614- 292-5766. CCS is located on the 4th Floor of the Younkin Success Center and 10th Floor of Lincoln Tower. You can reach an on call counselor when CCS is closed at 614-292-5766 and 24 hour emergency help is also available through the 24/7 National Suicide Prevention Hotline at 1-800-273- TALK or at suicidepreventionlifeline.org

Disclaimer

This syllabus should be taken as a fairly reliable guide for the course content and policy. However, you cannot claim any rights from it and in particular the instructor reserves the right to change due dates or the methods of grading and/or assessment if necessary. Any changes will be communicated to you through official course announcements with plenty of time to react.

Course schedule (tentative)

Class	Date	Topics, Readings, Assignments, Deadlines
1	R 5.13	Introduction; Class overview; Motivation for studying regression; Basic terminology review (probability distributions; moments; covariance and correlation, probability models, statistics, estimation methods)
2	T 5.18	Observational studies vs. experiments; Statistical modeling; Simple Linear Regression (SLR) model; Assumptions for SLR model; Mean-centered SLR model; Expectation and Variance of *Y* under SLR.
3	R 5.20	Estimation of SLR parameters – least squares and maximum likelihood estimates; Properties of estimators.
4	T 5.25	Inferences about SLR model parameters; Confidence intervals for mean response and for predictions; Practical issues relating to inference.
5	R 5.27	Regression sums of squares, degrees of freedom, and the ANOVA F-test for the SLR model; Goodness of fit and R-squared coefficient.
6	T 6.1	Using visual diagnostic tools to detect model assumption violations or lack of fit; Decomposition of SSE into SSPE and SSLF, and analytic test for lack of fit.
7	R 6.3	Analytic diagnostic tests for model assumptions; Remedial measures for addressing violations to model assumptions.
8	T 6.8	Joint (simultaneous) confidence intervals for regression parameters; Confidence bands.
9	R 6.10	Linear algebra and matrix review; SLR model in matrix from; Review example.
10	T 6.15	Multiple linear regression (MLR) model – quadratic model and interaction model.
11	R 6.17	Midterm Exam (covers class material through R 6.10).
12	T 6.22	Inference for parameters in MLR model; Sums of squares, df and ANOVA in MLR model; Multicollinearity; Lack of fit test.
13	R 6.24	Extra sums of squares; General linear test; Added variable plot.
14	T 6.29	Weighted least squares; MLR diagnostic measures; Leverage and Cook's distance; Outliers and Studentized residuals.
15	R 7.1	Categorical predictor variables; Interpreting coefficients for qualitative predictors; Interactions revisited; Piecewise linear models.
16	T 7.6	Polynomial regression models; Variance inflation factors; Design principles; Bootstrap parameter estimation.
17	R 7.8	Model selection criteria – Adjusted R-squared, Mallow's Cp, AIC, BIC; Model selection strategies – best subset regression, forward selection, backward elimination and Stepwise methods.
18	T 7.13	Model validation with cross-validation; Example applying model selection and cross-validation methods.
19	R 7.15	Final MLR topics: Model selection for large p and small n; Ridge regression; Loss functions and penalties; Lasso regression.
20	T 7.20	Logistic regression for binary response variable.
21	R 7.22	Logistic regression inference and deviance; residuals and diagnostics.
22	T 7.27	Multiple Logistic regression – model selection and predictive performance.
23	R 7.29	Poisson regression (time permitting); Project presentations.
24	T 8.3	Final Exam (cumulative).