

Stat 3301: Statistical Modeling for Discovery I, Autumn 2026 (23086)

Course Details

Version 1: August 20, 2026

Course Overview

Instructor Information

Instructor: Thomas Metzger, PhD

Email: metzger.181@osu.edu; **please do not use my Buckeyemail address.** Messages sent through Carmen often go into my spam folder; for fastest communication, use **metzger.181@osu.edu**.

Office hours: MW 9:30-10:30 am Cockins 340E, or by appointment.

Grader/TA: Ms. Skylar (Danyang) Shi, shi.2154@osu.edu

Course Delivery

This course meets 1:50 - 2:45 pm every Monday, Wednesday, and Friday in Pomerene Hall 150. Attendance is required.

Course Description

Statistical models for data analysis in the linear regression framework. The challenges of developing meaningful models for data are explored, with emphasis on the model building process, the use of numerical and graphical diagnostics for assessing model fit, and interpretation and communication of results. Statistical foundations are introduced along with basic inferential techniques. Prereq: C- or above in 3202; or 4202 and 5732; or permission of instructor. Prereq or concur: Math 2568; or permission of instructor.

Course Learning Outcomes

By the end of this course, students should successfully be able to:

- Use graphical and numerical summaries of data to describe relationships between variables.
- Formulate, fit, evaluate and compare regression models that describe relationships between variables.
- Understand and be able to describe the statistical foundations of standard regression models.
- Identify common violations of the assumptions that underly standard regression models.
- Perform a complete regression analysis and communicate the results in both statistical and problem-specific terms.

Materials

Text

- Required text: Weisberg, Sanford. Applied Linear Regression. 4th Edition. John Wiley & Sons, 2005.

Hardware and Software

- Required hardware: a webcam and microphone are required. There may be occasional individual meetings between students and the TA or professor for oral exams and intermittent check-ins.
- Required software: we will extensively use the statistical software package called R (The R Project for Statistical Computing; <http://www.r-project.org/>). This software package is available for free. You can 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>. Tutorials are available in the Swirl system, which you can learn about at <http://swirlstats.com/>. “R Programming: The basics of programming in R” is an appropriate first tutorial for students who have never used R.
- Required software: we will also use the R interface 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.
- Required software: Microsoft Office 365 ProPlus. All Ohio State students are now eligible for free Microsoft Office 365 ProPlus through Microsoft’s Student Advantage program. Each student can install Office on five PCs or Macs, five tablets (Windows, iPad® and Android™) and five phones. Students are able to access Word, Excel, PowerPoint, Outlook and other programs, depending on platform. Users will 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>.

Grades and Assignments

Homework (10%)

Homework assignments will be submitted as a PDF in knit Markdown format. Code and work should be included and format should be neat and professional. Homework grades will be earned based on a brief oral defense with either the professor or TA conducted in person during office hours or another scheduled time. Students who cannot defend their work will receive no credit for the assignment and may be submitted to COAM. AI may be used as a tool, but all submitted content must represent a student’s work. Students who do not complete and understand the homework assignments with their own unique thought and understanding will have a difficult time passing this course. Homework assignments will often serve as inspiration for quiz and exam problems.

Lecture Quizzes (10%)

Short quizzes will occasionally be given in class to ensure you are keeping on track with the material. Quiz problems will often be inspired by examples done in class or on homework. The lowest quiz grade will be dropped. Absolutely no outside resources are permitted on quizzes.

Written Exams (60%)

Three written exams, each comprising 20% of your grade, will take place during the semester. Exam 1, approximately covering Lectures 1A-4C, will take place Week 5 TBD. Exam 2, approximately covering lectures 5A-9C, will take place Week 10 TBD. Exam 3, approximately covering Lectures 10A-14C, will take place Week 15 TBD. Absolutely no outside resources are permitted on exams.

Project (20%)

Students will complete a data analysis project due at the end of the semester. Details on the project to be posted in early November. Tentatively, there to be a 8-10 page write up that will be completed in groups and a short individual oral defense (5-10 min) to be completed during finals week. The oral defense will

take place at a mutually agreeable time during finals week, or during the scheduled final exam time for this course (Wednesday, December 16 from 12:00-1:45 pm).

Quiz and Exam Etiquette

Students must keep their hands and quiz/exam document visible and above the desktop at all times. All items such as coats and backpacks must be kept off the table and on the back of the chair or floor. Any food or drink must not be in packaging containing print. Phones should be silenced and kept out of reach. Earbuds, smart glasses, and smart watches must be removed. You may be asked to orally defend your answers after the exam if there is any suspicion of academic misconduct. Students will be asked to defend their answers later if there is any suspicion of having obtained illicit help.

Academic Integrity

Given that the learning goals of this class include developing and refining your coding abilities, articulating statistical concepts orally and in writing, and connecting complex statistical concepts with mathematical and computer science tools, the use of generative artificial intelligence tools such as ChatGPT or Copilot is not permitted in this course for purposes including, but not limited to, generating code, finding solutions, or solving problems. Ultimately, any use of GenAI tools for work in this class for labs and homeworks may be scrutinized to ensure it is in adherence with Ohio State's Academic Integrity policy and Code of Student Conduct. If I suspect that you have used GenAI inappropriately on an assignment for this course, I will ask you to explain your process for completing the assignment in question. The unauthorized or inappropriate use of GenAI tools will result in referral to the Committee on Academic Misconduct.

AI Usage

May be Permitted/Appropriate:

- Pasting in error messages generated by code you have written yourself
- Asking to be pointed in the right direction for resources - Paraphrasing a topic you have phrased yourself
- Correcting grammar - Translating unknown words and phrases - Organizing or paraphrasing thoughts or ideas you have already created

Not permitted/Inappropriate:

- Pasting in problems that are verbatim or minimally/superficially adjusted - Asking to have a problem started for you - Using AI to solve proofs and problems - Copying or minimally/superficially altering code written by AI - AI does the thinking for you

The bottom line: if the work you submit represents your own uniquely generated thoughts, it is appropriate. If it represents someone (or something) else's thoughts or work, it is not appropriate. In general, do not submit work you can't defend, because if I have suspicions, I will ask you to defend it.

Tentative Course Schedule

Week	Dates	Topics
Week 1	Aug 24 - Aug 28	Introduction, summarizing data in R
Week 2	Aug 31 - Sept 4	Relationships among variables
Week 3	Sept 7 - Sept 11	Intro to simple linear regression (SLR), parameter estimation
Week 4	Sept 14 - Sept 18	Inference and prediction under SLR models
Week 5	Sept 21 - Sept 25	Sources of variability, coefficient of determination, residuals
Week 6	Sept 28 - Oct 2	Transformations of regressors and response variables
Week 7	Oct 5 - Oct 9	Multiple linear regression estimation, inference and prediction
Week 8	Oct 12 - Oct 16	Interpreting coefficients, correlated predictors
Week 9	Oct 19 - Oct 23	Regression with categorical and continuous regressors
Week 10	Oct 26 - Oct 30	Multiple categorical regressors, comparing nested models
Week 11	Nov 2 - Nov 6	Nonlinear relationships and polynomial regression
Week 12	Nov 9 - Nov 13	Regression model diagnostics
Week 13	Nov 16 - Nov 20	Common violations of independence
Week 14	Nov 23 - Nov 27	Descriptive versus causal interpretations of regression
Week 15	Nov 30 - Dec 4	Approaches to model comparison
Week 16	Dec 7 - Dec 11	Approaches to model selection