



THE OHIO STATE UNIVERSITY

COLLEGE OF ARTS AND SCIENCES

STAT 3302 (28525): Statistical Modeling for Discovery II

Spring 2023

Instructor:	Nasser Sadeghkhan	Time:	M W F 1:50 PM–2:45 PM
Email:	Sadeghkhan.1@osu.edu	Place:	Pomerene Hall (PO) 0250

Office Hours: W F 4:00-5:00 pm or by appointment. My office is located at 205B Cockins Hall.

TA: Daryl Swartzentruber (swartzentruber.20@osu.edu)

Course Description: This course continues to investigate statistical models for data analysis and discovery in big-data settings. The regression methods developed in Stat 3301 are extended to data settings with binary and multi-category outcomes. An introduction to some of the most commonly used statistical methods for exploring and analyzing multivariate data is provided. Interpretation and communication of the results of analyses is emphasized.

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

- Build, fit and interpret statistical models for binary outcomes.
- Understand the difference between nominal and ordinal outcomes and build regression models that are appropriate for each.
- Recognize the types of questions that can be answered by regression models for multi-category data and structure models to answer those questions.
- Comprehend the statistical principles that underlie basic methods for multivariate data analysis.

Course delivery:

- This course is the **three credit hours** which meets **Mondays, Wednesdays and Fridays from 12:40 pm to 1:35 pm at Pomerene Hall 250**. Lectures will be delivered **in-person class** during the scheduled class meeting times.
- Students are expected to attend and participate in these **in-person class** meetings. **Please arrive on-time**. Class meetings will be used to provide in-depth investigation of the topics for the week using a lecture format. Students will participate in these class sessions by engaging in discussions prompted by the instructor and by asking and answering questions. Students should plan to take notes during class. **No recorded version** of lectures are provided.

Main References:

- A. J. Dobson and A. Barnett (2008), An Introduction to Generalized Linear Models, Forth Edition, Chapman & Hall/CRC Texts in Statistical Science. Available online at <https://library.ohio-state.edu/record=b8615141~S7>
- A. C. Rencher and W. F. Christensen (2012), Methods of Multivariate Analysis, Third Edition, Wiley. Available online at <http://onlinelibrary.wiley.com.proxy.lib.ohio-state.edu/book/10.1002/9781118391686>.

I will highlight other useful references as the course progresses.

Prerequisites: Linear algebra (Math 2568), STAT 3301 (Statistical Modeling for Discovery I).

Necessary software

- This class requires you to use the statistical software package called R (The R Project for Statistical Computing; <http://www.r-project.org/>) to illustrate certain aspects. Here is the information for obtaining R.
 - 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>
 - 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.
- This class requires the use of the (free) R Markdown authoring framework to complete assignments. Information about R Markdown will be provided in class; an online guide with overview information can be found at <https://rmarkdown.rstudio.com>.
- 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>.

Evaluation: The evaluation will be determined based on **six** assignments, **two** midterm exams, a project and **one** final exam.

Assignments 20%, First Midterm 20%, Second Midterm 20%, Project 10%, Final 30%.

Grades will be recorded on [Carmen](#)

Important Dates:

Assignments	Almost biweekly (on Carmen)
Midterm #1	M–Feb 13 (in class)
Midterm #2	M–Mar 27 (in class)
Proposals of project	by F Mar 31
Project Deadline	by M Apr 24
Final Exam	Tu May 2, 12:00- 1:45 (in class)

Grading Scale:

93–100	A
90–92.9	A-
87–89.9	B+
83–86.9	B
80–82.9	B-
77–79.9	C+
73–76.9	C
70 –72.9	C-
67 –69.9	D+
60 –66.9	D
Below 60	E

Course Policy:

- There will be no online streaming nor recording.
- Lecture notes are partial, and you need to add some material during classes. So try not to miss classes!
- The lecture notes will be released one week in advance of every session on [Carmen](#).

Assignment Policy:

- The assignments will be posted on [Carmen](#).
- Your answers must be uploaded electronically through Carmen in a single PDF file.
- Pay attention to the due date, it is usually 9-10 days after the assignment is posted. No late assignment will be accepted. Under very exceptional circumstance if you are unable to complete an assignment on time, please get in touch with me asap, so we can discuss your situation.
- You are encouraged to work together on the homework, but **do not** copy any part of a assignment.
- Homework problems that require R software should be completed in R **Markdown** and html or pdf files must be provided. Raw computer output is not acceptable. Homework problems that do not require R may be handwritten (electronically, or on paper and scanned) and uploaded.

Project Policy:

- In groups of 4, students will be responsible for completing a project. This will be done in the "Study Group" on Carmen.
- Proposals for project ideas will be due by Friday March 31 including 1) title of the project, 2) data set and 3) tentative objectives. You can apply all the techniques you will learn in this course. Additionally, you can use other techniques you have already studied or any other method you wish. But please consider that it is advised to apply most of the techniques you have learned in this course.
- The project will consist of finding a dataset, formulating questions that can be answered with the data, and performing an appropriate analysis to answer the questions posed.

- The group leader and the group members will be assigned automatically. This is a group work and all the members are required to work equally. You are free to split the task between you. You are responsible to communicate and be in touch with your group members for proceeding the projects.
- Last week of the semester is dedicated for presenting your project in a Powerpoint or a Beamer.
- The group leader is responsible of emailing me in by the proposal deadline (March 31) the following materials and in the subject of the email "section 20943, and group number":
 - name of the group members
 - title of the project
 - data set
 - tentative objectives

Prior to the project deadline (April 24), the group leader must email me the following materials:

- name of the group members
 - The presentation (Powerpoint or Beamer) that you already presented in the class.
 - Dataset in CSV file.
 - R codes you have used. Must be runnable.
- There are limited options to change your group, under some circumstances if you want to. You can email me, and we will discuss and find a way.

Exam Policy:

All exams are closed book/closed notes. A basic calculator is allowed – tablets, laptops, and cellphones are not. Midterm 1 covers the material up to and including Wed 16 Feb. Midterm 2 covers the material up to and including Wed 23 Mar. The final will cover all the material for the course. There will be **no make-up** exams. If exceptional circumstances (sudden onset of illness, unexpected family situations, etc.) arise, contact me asap, so we can discuss your situation.

Class Policy:

- **Arrive to class on time.**
- Be courteous when using mobile devices. Make sure your cell phone is turned fully off, or silent. No texting, reading emails, playing games, or whatever else it is that people do with those wretched gizmos.
- If you must use a laptop in class, then turn off the sound and do not type on laptop keyboards which is really distracting.
- Missing one class could easily lead to a disastrous domino effect. If you have to miss a lecture, then I strongly recommend you study the material you missed before you return to class. Do not hesitate to come by and ask me. I require that you know all material covered in class. You are responsible for making up anything that was covered in lectures you missed.

Health and safety:

The Ohio State University Wexner Medical Center's Coronavirus Outbreak site (<https://wexnermedical.osu.edu/features/coronavirus>) includes the latest information about COVID-19 as well as guidance for students, faculty and staff.

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>.

Ohio State's academic integrity policy:

It is the responsibility of the Committee on Academic Misconduct to investigate or establish procedures for the investigation of 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.

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; <https://slds.osu.edu>; 098 Baker Hall, 113 W. 12th Avenue.

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 (e.g., race). 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.

Course schedule (tentative): Note that they are tentative and may be subject to change.

- Week 1: Introduction and Motivation, Review of Binomial Model and Maximum Likelihood Estimation
- Week 2: Simple Logistic Regression: Model Formulation
- Week 3: Estimation of Parameters, Multivariable Logistic Regression
- Week 4: Model Evaluation and Diagnostics in Logistic Regression
- Week 5: Model Building in Logistic Regression
- Week 6: Multinomial and Multicategory Regression
- Week 7: Ordinal Logistic Regression
- Week 8: Poisson Regression
- Week 9: Introduction to Multivariate Data
- Week 10: Multivariate Numerical Summaries
- Week 11: The Multivariate Normal Distribution;
- Week 12: More on The Multivariate Normal Distribution
- Week 13: Principal Components Analysis
- Week 14: Linear Discriminant Analysis
- Week 15: Project presentations