

STAT 3302: Statistical Modeling for Discovery II

Spring 2024

Section 28840

Instructor

Colleen Sunderland-Ober (sunderland.46@osu.edu)

Office Hours (via Zoom): Tuesday 5:30-7:00 pm and Thursday 5:30-7:00 pm

Grader

Kiljae Lee (lee.10428@osu.edu)

Course Description

This course continues to investigate statistical models for data analysis and discovery in big-data settings. The regression methods you learned about in STAT 3301 will be extended to data that have binary or multi-category outcomes. The most commonly used statistical methods for exploring and analyzing multivariate data will be introduced, with a focus on interpreting and communicating the results of analysis.

Course Learning Outcomes

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

- Construct, fit, and interpret statistical models for binary outcomes;
- Recognize the difference between nominal and ordinal outcomes, and be able to fit regression models that are appropriate for each;
- Identify the types of questions that can be answered by regression models for multiple-category data, and build models that can answer those questions;
- Understand the statistical principles underlying methods for multivariate data analysis.

Course delivery

The course content will be delivered asynchronously. You will receive a weekly email with details about what you are expected to do that week, and what deadlines are approaching. Lecture videos equivalent to a total three 55-minute classroom hours per week will be uploaded to the Carmen site on Mondays. These may be a series of shorter videos, each focusing on a particular topic. Slides for each week's lectures will be available on Carmen. It is your responsibility to watch the videos and take notes or annotate the slides. I will be available twice a week during office hours on Zoom to answer your questions about the lecture content or the assignments. It is also possible to meet on Zoom by appointment.

Course materials

- 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>
- 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>

Course technology

Necessary technical skills

Students need to be able to use Carmen, CarmenZoom, and email. Students should be able to open and read pdf files, and use a web browser.

Necessary computing equipment

Students need a computer with a web browser, microphone and camera.

Necessary computer software

This class requires you to use the statistical software package R (the R Project for Statistical Computing (<http://www.r-project.org/>)

- You can download R for Windows, Mac, and Linux for free at the CRAN archive: <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.
- You are strongly encouraged to use 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 homework assignments, two midterm exams, a project and one final exam. Assignments 20%, First Midterm 20% (Feb. 20), Second Midterm 20% (Mar. 26), Project 10%, Final 30% (Apr. 26). Grades will be recorded on Carmen.

Grading

| | |
|---------------|----|
| 93–100 | A |
| 90–92 | A- |
| 87–89 | B+ |
| 83–86 | B |
| 80–82 | B- |
| 77–79 | C+ |
| 73–76 | C |
| 70 –72..... | C- |
| 67 –69..... | D+ |
| 60 –66..... | D |
| Below 60..... | E |

Course policies

Homework will be assigned on an approximately biweekly basis and made available on Carmen. Your answers must be submitted on Carmen as a **single pdf**. You are encouraged to use RMarkdown. Homework problems that do not require R may be handwritten (electronically, or on paper and scanned) but they must be combined together with the rest of your submission into a **single PDF**. You can use Adobe Acrobat (see <https://it.osu.edu/adobeLinks to an external site.>) to combine PDF files.

You may collaborate with classmates on your homework, but ultimately your submitted answers (code, prose, solutions) must be your own work. For example, I encourage you to discuss strategies for solving problems, but the actual solution that you write must be your own. Keep in mind the university policies on plagiarism. Do not copy or plagiarize anything from others or that you may find on the Internet.

There will be two midterm exams and a cumulative final exam. The exams will be available for 24 hours, but once you open an exam, there is a fixed amount of time you have to complete it. For midterm exams, this will be 80 minutes (70 minutes to work on the exam, and 10 minutes to get it loaded on Carmen). For the final exam, this will be 1 hour 55 minutes (1 hour 45 minutes to work on the exam, and 10 minutes to get your work loaded onto Carmen). The exams are open book, open notes. If there is an exceptional circumstance, e.g. medical or family emergency, then please contact me as soon as possible before the exam date to make potential arrangements.

There will be a project that you will complete in groups of 4 students. The group leader and members will be assigned automatically. 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.

Proposals for the project will be due just before spring break. The complete project will be due before the final exam. Further details, including deadlines, will be given as the semester progresses and announced on Carmen. There are limited options to change your group, under some circumstances if you want to. Email me so we can discuss and find a way.

If you absolutely need to turn in an assignment late and have a valid excuse, please see me for the necessary arrangements. However, you must notify me in advance in such a situation. Exceptions to this policy will be permitted only in extreme situations such as serious injury immediately prior to an assignment being due or severe illness requiring hospitalization.

Religious Accommodations

It is Ohio State's policy to reasonably accommodate the sincerely held religious beliefs and practices of all students. The policy permits a student to be absent for up to three days each academic semester for reasons of faith or religious or spiritual belief.

Students planning to use religious beliefs or practices accommodations for course requirements must inform the instructor in writing no later than 14 days after the course begins. The instructor is then responsible for scheduling an alternative time and date for the course requirement, which may be before or after the original time and date of the course requirement. These alternative accommodations will remain confidential. It is the student's responsibility to ensure that all course assignments are completed.

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.

Academic Integrity

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.

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.

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.

Course Calendar (may be subject to change)

Week 1 (Jan. 8 – 12)

Introduction, review of binomial model and MLE

Week 2 (Jan. 15 – 19)

Simple logistic regression (model formulation)

Week 3 (Jan. 22 – 26)

Estimating parameters, multivariable logistic regression

Week 4 (Jan. 29 – Feb. 2)

Model evaluation and diagnostics in logistic regression

Week 5 (Feb. 5 – 9)

Model building in logistic regression

Week 6 (Feb. 12 – 16)

Binomial regression and Poisson regression

Week 7 (Feb. 19 – 23)

Poisson regression, continued

Midterm Exam 1 (February 20-21)

Week 8 (Feb. 26 – Mar. 1)

Models for ordinal and multicategory data

Week 9 (Mar. 4 – 8)

Introduction to multivariate data

03/11-03/15: spring break (no classes)

Week 10 (Mar. 18 – 22)

Multivariate numerical summaries

Week 11 (Mar. 25 – 29)

Multivariate normal distribution

Midterm Exam 2 (Mar. 26-27)

Week 12 (Apr. 1 – 5)

Multivariate normal distribution, continued

Week 13 (Apr. 8 – 12)

Principal components analysis

Week 14 (Apr. 15 – 19)

Linear discriminant analysis

Project due

Week 15 (Apr. 22 – 26)

TBA (catch-up, review, etc.)

Final Exam (Apr. 26)

