Syllabus

STAT 3302

Statistical Modeling for Discovery II

Spring 2026

Credit Hours: 3

Course times and location: Synchronous class on Tuesdays 5.30-6.25 pm. Asynchronous classes are posted on Carmen Canvas. Midterms and final will be in-person.

Mode of delivery: Distance Enhanced (DE)

Course overview

Instructor

- **Name**: Sally Paganin
- **Email Address**: paganin.1@osu.edu
- **Office location**: 229 Cockins Hall
- **Course Zoom Link:** link on Carmen Canvas
- **Office Hours:** **Mondays 3:30 pm - 4:30 pm on Zoom** or **by appointment**

Note: My preferred method of contact is e-mail

Graduate Teaching Assistant

Name: Wentao Jiang
Email Address: Jiang.3122@buckeyemail.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 multicategory 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 expected learning outcomes

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

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

Prerequisites

STAT 3301 (Statistical Modeling for Discovery I) and MATH 2568 (Linear Algebra); or permission of instructor.

How this online course works

Mode of delivery

This is a distance-enhanced course, meaning it includes both **recorded content** and **live online meetings**. Recorded lectures will be posted each week for you to watch on your own schedule. In addition, we will meet for a **55-minute live online class every Tuesday from 5:30–6:25 pm**.

Pace of online activities

This course is divided into weekly modules. Recorded lectures will be posted every Wednesday. Students are expected to watch the lectures before the live meeting on the Tuesday of the following week. A discussion board will be provided to collect questions for the live meeting. The weekly

live lecture is intended to review questions and exercises from the week's material. These sessions are meant to be interactive and provide an opportunity to connect as a whole class. Live Tuesday sessions will not be recorded.

For expectation and studying tips for distance-enhanced courses see https://teaching.resources.osu.edu/HybridDistanceEnchancedCourses

Credit hours and work expectations

This is a **3-credit-hour course**. According to Ohio State policy (go.osu.edu/credithours), students should expect around 9 hours of engagement with the class each week to receive a grade of (C) average. Actual hours spent will vary by student learning habits and the assignments each week.

Participation requirements

Because this is an online course, attendance is based on your online activity and participation. The following is a summary of students' expected participation:

Participating in online activities

at least once per week

You are expected to log in to the course in Carmen every week. During most weeks you will probably log in many times. If you have a situation that might missing an entire week of class, discuss it with me as soon as possible.

Participating in a group discussion

at least once per week

Most weeks will include a **group discussion activity**, often based on a dataset or a data-analysis question.

Office hours and live sessions

optional

All scheduled live course events, including Tuesday live sessions and office hours, are optional. However, attending live sessions can be very

helpful for asking questions and staying on track. Live Tuesday classes will not be recorded.

Course communication guidelines

The following are my expectations for how we should communicate as a class. Above all, please remember to be respectful and thoughtful

Writing style

While there is no need to participate in class discussions as if you were writing a research paper, you should remember to write using good grammar, spelling, and punctuation. A more conversational tone is fine for non-academic topics.

Tone and civility

Let's maintain a supportive learning community where everyone feels safe and where people can disagree amicably. Remember that sarcasm doesn't always come across online. I will provide specific guidance for discussions on controversial or personal topics.

Citing your sources

When we have academic discussions, please cite your sources to back up what you say. For the textbook or other course materials, list at least the title and page numbers. For online sources, include a link.

Protecting and saving your work

Consider composing your academic posts in a word processor, where you can save your work, and then copying into the Carmen discussion.

Course materials and technologies

Textbooks

Required

1. Dobson, A. J., & Barnett, A. (2018). *An introduction to generalized linear models* (4th ed.). Chapman & Hall/CRC.
   Available via **Ohio State University Libraries**.

2. Rencher, A. C., & Christensen, W. F. (2012). *Methods of multivariate analysis* (3rd ed.). Wiley.
   Available via **Ohio State University Libraries**.

Recommended (optional)

The following book contains many examples that we will also discuss in class.

- Agresti, A. (2006). *Introduction to categorical data analysis* (2nd ed.). Wiley.
  Available via **Ohio State University Libraries**.

Course technology

Required Equipment

- **Computer**: A current Mac (macOS) or PC (Windows or Linux) with a high-speed internet connection, capable of running R and RStudio (described below).
- **Webcam**: built-in or external webcam, fully installed and tested
- **Microphone**: built-in laptop or tablet mic or external microphone
- **Other**: a mobile device (smartphone or tablet) to use for BuckeyePass authentication

Required software

R (The R Project for Statistical Computing):

- This course requires the use of the statistical software **R**, which is free and open-source.
- Please install the **latest stable version of R** available at the time you set up your system.
- R is available for Windows, macOS, and Linux from the **Comprehensive R Archive Network (CRAN)**.
- An in-depth introduction to R is available through the **R Introduction Manual**.

RStudio (optional, recommended):

RStudio is an easy-to-use interface for working with R. It is available for Windows, macOS, and Linux and can be downloaded for free from **Posit**.

Note: R must be installed before installing RStudio.

Grading and instructor response

How your grade is calculated

Assignment Category		Percentage
Participation	Weekly discussion	5%
Homework	Almost biweekly	15%
Midterm 1	Tue Feb 24, 2026 (in person)	20%
Midterm 2	Thu March 26, 2026 (in person)	20%
Project	Thu April 23, 2026 (last day of classes)	10%
Final Exam	Thu April 30, 2026 6:00-7:45pm	30%
Total		**100%**

Description of major course assignments

Participation

- **Description**

Participation is based on regular engagement with weekly online discussions, with submissions typically due on Mondays. Activities will include individual reflections early in the semester and small-group discussions later in the term, often centered on assigned datasets or the course project. For group activities, one written discussion summary per group is required

- **Academic integrity and collaboration guidelines**

Collaboration is expected for group-based participation activities. For individual participation assignments, all submitted work must be your own. For group participation assignments, all group members are responsible for the content of the submission. **Do not copy or plagiarize material from other students, online sources, or external resources**; university policies on academic integrity apply. Generative AI tools (e.g., Gemini, ChatGPT) may be used to help brainstorm ideas or clarify concepts, **but AI-generated text may not be copied directly into submissions. Any use of generative AI tools must be disclosed, including how the tools were used.**

Homework

- **Description**

Some homework assignments will **require handwritten solutions** (either completed electronically or written on paper and scanned). Other assignments will require **a combination of R code and written explanations**, completed using the **R Quarto template** provided with the course materials. The required format for each assignment will be clearly stated when it is posted.

Regardless of format, **all homework must be submitted on Carmen/Gradescope as a single PDF file**. If an assignment includes both handwritten and computer-generated components, all materials must be **combined into one PDF** before submission. You may use **Adobe Acrobat** to merge PDF files (available through Ohio State).

- **Academic integrity and collaboration guidelines**

You are encouraged to work collaboratively on homework assignments; however, **you must not copy any part of another student's work**. Your submitted answers—including code, written explanations, and solutions—must be your own. You may discuss problem-solving strategies with others, but the final solutions you submit must be written independently.

University policies on plagiarism apply. Do not copy or plagiarize material from other students, online sources, or any external resources.

You **may use Generative AI tools** (e.g., Gemini, ChatGPT, etc.) to help you understand concepts or clarify questions. However, **you may not copy and paste code, solutions, or written text** from the internet or from any AI tool into your submission.

If you collaborate with others and/or use generative AI tools, you **must include a disclosure statement at the beginning of your submission** specifying i) with whom you collaborated, and ii) which generative AI tools you used, and how you used them.

Exams

○ Description

There will be **two midterm exams** and **one cumulative final exam**. All exams will be **closed-book and closed-notes**, and **no calculator will be needed**. The use of computers, smartphones, or other electronic devices is not permitted during exams.

If an exceptional circumstance arises (e.g., a medical or family emergency), please contact the instructor as soon as possible **before the exam date** to discuss possible arrangements.

○ Academic integrity and collaboration guidelines

You may **not collaborate with anyone** on exams. You may **not access or view previous exams or solutions**, except for practice materials explicitly provided by the instructor.

Exams are **closed-book and closed-notes**. **No electronic devices** are permitted. A **handwritten cheat sheet** may be permitted only if explicitly announced in advance, along with clear guidelines on its content.

You may **not discuss or share any exam content** with others until **all students have completed the exam**, including those who have not yet taken it.

Project

- **Description**

Students will complete a group project in teams of up to four students. Group assignments, including the designation of a group leader, will be made automatically.

The project will involve identifying a dataset, formulating research questions that can be addressed using the data, and conducting an appropriate statistical analysis to answer those questions.

Project proposals will be due shortly before **spring break**, and the final project will be due **before the final exam**. Additional details and specific deadlines will be provided during the semester and announced on Carmen.

There are limited options for changing groups. If you believe a change is necessary, please email the instructor to discuss possible solutions.

- **Academic integrity and collaboration guidelines**

You are expected to **work collaboratively with your group members** on the project. As with homework assignments, **do not copy or plagiarize** material from online sources or elsewhere. You may use **Generative AI tools** (e.g., Gemini, ChatGPT) to assist with writing or coding; however, **you are fully responsible for all submitted work** and are expected to understand and be able to explain every part of your submission and presentation.

If you collaborate beyond your assigned group and/or use generative AI tools, you **must include a disclosure statement at the beginning of your submission** specifying: (i) with whom you collaborated, and (ii) which generative AI tools you used, and how you used them.

Late assignments

Assignments are expected to be submitted on time. If you anticipate needing to submit an assignment late due to a valid reason, please contact

the instructor **in advance** to discuss possible arrangements. Exceptions will be considered only in **exceptional circumstances**, such as serious injury or severe illness.

Late assignments will be accepted for up to **48 hours after the original deadline**, with a **2% deduction per hour** past the deadline. **No submissions will be accepted after 48 hours.**

Submitting the wrong file (e.g., a blank template, an incomplete draft, or a corrupted file), or submitting in an incorrect format, is **not a valid excuse**. It is your responsibility to ensure that the correct document is submitted in the proper format before the deadline. Students are strongly encouraged to begin assignments early; last-minute issues are not grounds for exceptions.

Each student is allowed **one late-submission waiver during the semester**, **no questions asked**. When using the waiver, the assignment must still be submitted **within the 48-hour late window**, and the late penalty will be waived for that assignment. To use the waiver, you must notify the instructor in writing and specify which assignment the waiver applies to.

Grading Scale

- 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
- Under 60: E

Instructor feedback and response time

I am providing the following list to give you an idea of my intended availability throughout the course. Remember that you can call 614-688-4357 (HELP) at any time if you have a technical problem.

Grading and feedback

For homework assignments, you can generally expect feedback and grades within **10 days.**

Preferred contact method

If you have a question, please contact me first through my Ohio State email address. I will reply to emails within **24 hours on days when class is in session at the university.**

Class announcements

I will send all important class-wide messages through the Announcements tool in CarmenCanvas. Please check your notification preferences (go.osu.edu/canvas-notifications) to ensure you receive these messages.

Academic policies

This course follows all required Ohio State University syllabus policies, including Academic Misconduct, Disability Services, Religious Accommodations, and Intellectual Diversity. Full statements are available at: https://ugeducation.osu.edu/academics/syllabus-policies-statements

Course Schedule

Refer to our Carmen course page for up-to-date assignment due dates.

Week	Date(s)	Topics / Readings / Assignments	Assessments Due
1	Jan 12–16	Introduction; review of the Bernoulli and Binomial model and inference	
2	Jan 19–23	Review of maximum likelihood estimation. Introduction to simple logistic regression	**Discussion**
3	Jan 26–30	Simple logistic regression. Estimation of parameters; interpretation	**Discussion HMW 1**
4	Feb 2–6	Model evaluation and diagnostics in logistic regression	**Discussion**
5	Feb 9–13	Interpretation of categorical covariates.	**Discussion HMW**
6	Feb 16–20	Model building in logistic regression	**Discussion**

Week	Date(s)	Topics / Readings / Assignments	Assessments Due
	Feb 24	**Midterm Exam 1**	**Discussion HMW 3**
7	Feb 25–27	Binomial regression	
8	Mar 2–6	Intro to GLM. Poisson regression	**Discussion**
9	Mar 9–13	Overdispersion in Poisson regression. Intro to nominal and ordinal models	**Discussion HMW 4**
	Mar 16–20	**No class (Spring Break)**	
10	Mar 23	Introduction to multivariate data	**Project proposal**
	Mar 26	**Midterm Exam 2**	

Week	Date(s)	Topics / Readings / Assignments	Assessments Due
11	Mar 30–Apr 3	Summaries of multivariate data. Random vectors.	**Discussion**
12	Apr 6–10	Multivariate Normal Distribution; modeling and inference.	**Discussion HMW 5**
13	Apr 13–17	Principal components analysis	**Discussion**
14	Apr 20–23	Review and further topics	**Discussion HMW 6**
15	Apr 27–29		**Project report and presentations**
Final	April 30 1	**Final Exam (6:00 pm – 7:45 pm)**	