APSTA-GE-2044
Generalized Linear Models and Extensions

Klint Kanopka, Ph.D. (he/him)
Assistant Professor of Applied Statistics

Spring 2024-2025

1 Course Description

A second year course in advanced statistical techniques that covers useful quantitative tools in health & policy research. Assuming a strong foundation in regression & the general linear model, this course focuses on data analysis that utilizes models for categorical, discrete or limited outcomes that are commonly seen in health & policy studies. Examples include health status, number of clinic visits, etc. In this course students will also learn the principles of likelihood-based inference, which will assist them in some of the more advanced statistics courses.

2 Student Learning Outcomes

During the course of this class:

1. Students will identify categorical and discrete outcomes and articulate the challenges they present in data analysis.
2. Students will implement and evaluate statistical models that deal with categorical, discrete, and limited outcomes.
3. Students will apply generalized linear models to analyze real data.
4. Students will identify exponential family distributions and articulate how they give rise to generalized linear models.

3 Course Logistics

3.1 Teaching Team

- **Klint Kanopka** - Instructor

  - klint.kanopka@nyu.edu
  - Office Hours: M 2p-3p @ Kimball, Rm 205W
- **Sophia Deng** - Grader
  - sd5718@nyu.edu

3.2 Meeting Times and Locations

In-person lecture attendance is required, but lectures will be recorded for reference.

- **Lecture:** T 1.45a-4.45p @ 194 Mercer, Rm 208

3.3 Required and Recommended Reading

- *Categorical Data Analysis* [1]
- *Generalized Linear Model for Categorical and Continuous Limited Dependent Variables* [4]
- *Generalized Linear Models: An Applied Approach* [2]

3.4 Prerequisites

Students should be experienced with the `R` Programming Language [3], regression, probability, and data visualization.

4 Grading

This is a two-credit course. Table 1 contains the proportion of the final grade coming from each assignment category. Table 2 contains the transformation from the weighted proportion of earned points to letter grades.

Category	p
Problem Sets	0.9
Participation (Lecture)	0.1

Table 1: Final grade weighting scheme

4.1 Problem Sets

The course contains four equally weighted problem sets (PS0-PS3).

	G^-	G	G^+
A	[.895, .945)	[.945, 1]	—
B	[.795, .825)	[.825, .865)	[.865, .895)
C	[.695, .825)	[.725, .765)	[.765, .795)
D	[.600, .640)	[.640, .670)	[.670, .695)
F	—	[0, .600)	—

Table 2: Grading Scale

4.1.1 Submission Guidelines

All assignments are to be submitted on Brightspace using Quarto (preferred) or Rmarkdown and adhere to the following conventions:

1. **Submit assignments by the deadline.** Assignment deadlines are typically Fridays at 11.59p. Late assignments will receive a 10% penalty per day.

2. **Name your files correctly.** For each homework assignment, name your files using the convention: `LastName_FirstInitial_PS#.ext`

3. **Put your name and the assignment in text of the file.** These should occupy the author and title fields, respectively.

4. **Submit both a source file and a knitted file.** Source files have the extension `.qmd` or `.Rmd`. Files should be knitted to `.pdf`. As an example for the first homework, I would submit two files:

   - `Kanopka_K_PS1.qmd`
   - `Kanopka_K_PS1.pdf`

5. **Your source file should run without modification.** If you load a file, please use relative paths. Do not load libraries unless explicitly requested in an assignment.

6. **Do not use `install.packages()` calls in your assignments.** Please install packages locally and then load them in your assignment source code.

4.2 Participation

Each unexcused absence from lecture will result in a deduction of 20% of available participation points. To receive an excused absence, please contact me in advance.

5 Course Policies

5.1 Academic Integrity

I take academic integrity incredibly seriously. Please review NYU Steinhardt's Academic Policies and Procedures for more information on specific policies, the disciplinary process, and sanctions.

5.2 Collaboration

I **strongly** encourage students to form study groups. Students may discuss and work on problem sets in groups. Each student must (1) report at the top of each question what other students they consulted with, (2) write their code and solution independently, and (3) understand their work well enough in order to reconstruct it entirely on their own.

5.3 AI Tool Policy

All assignments should be your own original work. The use of generative AI tools in this course is explicitly not allowed. Not all AI tools are generative (and thus banned), however. Here are some guiding examples to consider when using various AI-based tools. If you have a question about a specific tool or use case, please reach out to me.

Examples of AI that are okay to use in the course include:

- Grammar and spelling checkers (e.g., Grammarly)
- Transcription or translation tools (e.g., OtterAI)
- Generative AI tools that help you troubleshoot and understand code you have written

Examples of AI that are **not** okay to use in this course:

- Generative AI tools that analyze data for you (e.g., Chartify, Rows.ai)
- Generative AI tools that write your code for you (e.g., Copilot)
- Generative AI tools that construct text responses to questions from assignments for you (e.g., ChatGPT)

5.4 Students with Disabilities

New York University is committed to providing equal educational opportunity and participation for all students, and academic accommodations are available for qualified students who disclose their disability to the Moses Center. Students requesting academic accommodations are advised to reach out to the Moses Center for Student Accessibility as early as possible in the semester for assistance

- Teleophone: 212-998-4980

- Email: mosescsa@nyu.edu
- Web: `http://www.nyu.edu/csd`

5.5 Mental Health Statement

If you are experiencing undue personal and/or academic stress during the semester that may be interfering with your ability to perform academically, the NYU Wellness Exchange (212-443-9999) offers a range of services to assist and support you. I am available to speak with you about stresses related to your work in my course, and I can assist you in connecting with the Wellness Exchange. Additionally, if you anticipate any challenges with completing the assignments, readings, exams and other work required in this course, I encourage you to register with the Moses Center (212 998 4980) in advance so that you may be granted the proper academic accommodations.

5.6 Inclusion

NYU values an inclusive and equitable environment for all our students. I hope to foster a sense of community in this class and consider it a place where individuals of all backgrounds, beliefs, ethnicities, national origins, gender identities, sexual orientations, religious and political affiliations, and abilities will be treated with respect. I intend that all students' learning needs be addressed in and out of class and that the diversity students bring to this class be viewed as a resource and strength. Please feel free to speak with me if this standard is not being upheld.

6 Course Calendar

Table 3 contains topics, assignments, and deadlines. Note that this is subject to change.

Week	Lecture	Assignments
1	Topics: • GLM overview • Binary Outcomes, Logistic Regression Preparation: • TBD	Released: • HW0, HW1
2	Topics: • The Exponential Family • Probit Regression Preparation: • TBD	Due: • HW0
3	Topics: • Classical Test Theory • Item Response Theory • Rasch Models for Dichotomous Responses Preparation: • TBD	Released: • HW2 Due: • HW1
4	Topics: • Models for counts: Poisson models and extensions Preparation: • TBD	

5	Topics: • Analysis of Contingency Tables Preparation: • TBD	Released: • HW3 Due: • HW2
6	Topics: • GLM for clustered and longitudinal data I Preparation: • TBD	
7	Topics: • GLM for clustered and longitudinal data II Preparation: • TBD	Due: • HW3

Table 3: Course Calendar

References

[1] A. Agresti. *Categorical data analysis*, volume 792. John Wiley & Sons, 2012.

[2] J. P. Hoffmann. *Generalized linear models: An applied approach*. Pearson, 2004.

[3] R Core Team. *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria, 2021. URL `https://www.R-project.org/`.

[4] M. Smithson and E. C. Merkle. *Generalized linear models for categorical and continuous limited dependent variables*. CRC Press, 2013.