APSTA-GE-2044
Generalized Linear Models and Extensions

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

Spring 2025-2026

1 Course Description

A second year course in advanced statistical techniques that covers useful quantitative tools in education, 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 3p-4p @ Kimball, Rm 205W

3.2 Meeting Times and Locations

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

- **Lecture:** T 1.45p-4.45p @ 194 Mercer, Rm 204

3.3 Required and Recommended Reading

- *Categorical Data Analysis* [1]
- *Generalized Linear Models 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	1.0

Table 1: Final grade weighting scheme

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

Table 2: Grading Scale

4.1 Problem Sets

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

4.1.1 Submission Guidelines

All assignments are to be submitted on Gradescope as a `.pdf` 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 up to a maximum of three days.

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 the text of the file.** These should occupy the author and title fields, respectively.

4. **Submit a compiled file on Gradescope.** Source files are Quarto documents and have the extension `.qmd`. Please retain these to submit upon request. Submitted files should be compiled to `.pdf` and generated from the `.qmd` template provided. As an example, for the first homework, I would create two files:

   - `Kanopka_K_PS0.pdf`
   - `Kanopka_K_PS0.qmd`

   I would submit `Kanopka_K_PS0.pdf` on Gradescope and retain a copy of `Kanopka_K_PS0.qmd` locally.

5. **If requested, 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 Extra Credit

This course allows extra credit to be earned in any of the following three ways:[1]

1. The first person to report typographical errors or mistakes in any of my course materials will receive one (1) extra credit point per error (I expect there to be a lot of them).

2. "Valuable suggestions" will also result in one (1) extra credit point and acknowledgment in a footnote.

[1] This policy is inspired by the Knuth Reward Check

3. Reporting a coding error with a solution or a valuable suggestion that requires additional labor (code snippets, a figure, a dataset, etc.) will be worth more, up to a maximum of 2^8 points. This amount will be scaled based on the nature and duration of required labor.

These should be claimed in the `#2044-glmm-errata` Slack channel.

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, but students should feel comfortable consulting AI tools to support their work. The expectation is that students will write their own software implementations, interpret their results, and develop their own narrative responses, but may use AI tools to help with tasks like debugging and refining their grammar. If you have a question about a specific tool or use case, please reach out to me.

5.4 Students with Disabilities

Students with physical or learning disabilities are required to register with the Moses Center for Student Accessibility, 726 Broadway, 2nd Floor, (212-998-4980 and online at `http://www.nyu.edu/csd`). They must present a letter from the Center to the instructor at the start of the semester to be considered for appropriate accommodation.

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: • Models for counts: Poisson models and extensions Preparation: • TBD	Released: • HW2 Due: • HW1
4	Topics: • Analysis of Contingency Tables Preparation: • TBD	
5	Topics: • GLM for clustered and longitudinal data I Preparation: • TBD	Released: • HW3 Due: • HW2

6	Topics: • GLM for clustered and longitudinal data II Preparation: • TBD	
7	Topics: • Generalized Additive Models 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.