Generalized Linear and Mixed Models
APSTA-GE 2044

Meeting Time and Location

Runs for seven weeks (03/25-05/06)
Time: Friday 9:15-12:15 (first two-thirds lecture; last one-third lab/Q&A; there will be a short break between).
Location: Silver Room: 411 (in-person)

Instructor

Marc Scott, PhD
Email: marc.scott@nyu.edu,
Office hours: Friday 12:15-1pm in Kimball 2nd Floor Conference Room

Course Assistant

Kwan Bo (Joseph) Shim
Email: ks6017@nyu.edu
Office hours: Thursday 11:30-12:30 via Zoom.

Course Description:

This is a course in advanced statistical techniques that covers generalized linear models and extensions that are commonly used in health and policy research. Assuming a strong foundation in the general linear model (linear regression and ANOVA) and exposure to the linear mixed model (a.k.a. multilevel models), this course focuses on data analysis that utilizes models for categorical, discrete or limited outcomes, some of them may come with hierarchical structures. Examples are drawn from broad areas of health and policy research including determinants of self-reported health status or factors influencing number of clinic visits. 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.

Course Prerequisites:

- REQUIRED: RESCH-GE 2003 Intermediate Statistical Methods or an equivalent course that covers linear regression analysis at intermediate level.
- PREFERRED: a course in multilevel data analysis or longitudinal data analysis
- Students are assumed to be able to perform basic data management and basic statistical analysis such as linear regression in R.

Learning Objectives:

By the end of the course, students will be able to:

1. Learn about the different characteristics of categorical and discrete outcomes and their inherent challenges when part of a data analysis.
2. Be introduced to and master the use of a wide range of statistical models that deal with categorical, discrete and limited outcomes, such as logistic/probit regressions, Poisson regression models and their variations through examples in social and behavioral sciences.
3. Build practical skills to apply appropriate statistical models and analyze data using statistical software.

4. Gain deeper understanding of the statistical theory, specifically likelihood-based inference that underlies statistical practices.

Course Format:
Discussions + hands-on problem solving exercises + question & answer sessions

Course Outline (approximate):

	Topic
Weeks 1-3	GLM overview, Logistic Regression: logistic and probit models
	Extensions to logistic regression: ordinal and multinomial models (hwk1)
Weeks 4-5	Models for counts: Poisson models and extensions
	Analysis of Contingency Tables (hwk2)
Weeks 6-7	GLM for clustered and longitudinal data I
	GLM for clustered and longitudinal data II (hwk3)

Recommended Texts:

- **[SM]**Michael Smithson and Edgar C. Merkle. 2014. *Generalized Linear Model for Categorical and Continuous Limited Dependent Variables, CRC Press* (Intermediate math level, also contains Stata and R code.)
- **[HM]** Hardin and Joseph Hilbe. 2007. *Generalized Linear Models and Extensions using Stata.* Second edition, Stata Press. (Good Stata reference book)
- **[RS]** Sophia Rabe-Hesketh, Anders Skrondal, (2012) Multilevel and Longitudinal Modeling Using Stata, Volume II: Categorical Responses, Counts, and Survival, Third Edition, Stata Press (Good Stata reference book)
- John P. Hoffmann. 2004 *Generalized Linear Models: An Applied Approach*. Pearson. (Minimal mathl)
- J Scott Long and Jeremy Freese. 2005. *Regression Models for Categorical Dependent Variables Using Stata*, 2nd Edition. Stata Press. (Has a good introduction to Stata)
- Alan Agresti. 2002 *Categorical Data Analysis*. Wiley. (The "bible" book in this topic. Discussions on log-linear model)

Labs

The "lab" in this class occurs in the last third of the 3 hour meeting. The primary purpose of this meeting is to do exercises related to the class material. These may be individual or group activities (attendance is required). Question and answer sessions are part of the lab.

Participation

You will receive participation credit based on the quality and quantity of your participation in the Discussion Forums. This is 19% of your grade, and each Module is tracked separately, so participation in each is necessary to receive non-zero credit for a module.

Class ***attendance is mandatory*** with these exceptions: if you are ill, stay home; if you have a Moses Center accommodation associated with in-class attendance.

Recordings

I will record lecture, but students who Zoom into the class cannot participate directly.

Course Requirements:

The grade for this course will be determined as follows: 3 problem sets (27% each for a total of 81%), plus 19% class participation based on Discussion Forums (exercises posted on Brightspace). **The course and problem sets will be managed through NYU Brightspace** (using Quizzes).

The grading system is as follows:

A: 95% and above	A-: 90-94.99%	
B+: 86-89.99%	B: 83-85.99%	B-: 80-82.99%
C+: 76-79.99%	C: 73-75.99%	C-: 70-72.99%
D+: 67-69.99%	D: 64-66.99%	D-: 60-63.99%
F: less than 60%		

Academic Integrity:

All students are responsible for understanding and complying with the NYU Steinhardt Statement on Academic Integrity. A copy is available at: http://steinhardt.nyu.edu/policies/academic_integrity.

Students with Disabilities:

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