

Module 0: Welcome to STA 310

Rebecca C. Steorts (slide and course adaptation from Maria Tackett)

Welcome!

Teaching Team

Instructor:

Professor Rebecca Steorts
Old Chem 216
rebecca.steorts@duke.edu

Teaching assistants

Suchismita Roy
Wenxin Guo

Course logistics

Lectures

Mondays and Wednesdays, 11:45 - 1:00 pm, Gray 228

Labs

Lab 01: Friday, 11:45 - 1:00 pm, Social Sciences 105

Generalized Linear Models

In statistics, a generalized linear model (GLM) is a flexible generalization of ordinary linear regression. The GLM generalizes linear regression by allowing the linear model to be related to the response variable via a link function and by allowing the magnitude of the variance of each measurement to be a function of its predicted value.¹

¹Source: Generalized linear model

Generalized Linear Models

In statistics, a generalized linear model (GLM) is a flexible generalization of ordinary linear regression. The GLM generalizes linear regression by allowing the linear model to be related to the response variable via a link function and by allowing the magnitude of the variance of each measurement to be a function of its predicted value.¹

Example: Logistic regression

$$\begin{aligned}\pi = P(y = 1|x) &\Rightarrow \text{Link function: } \log\left(\frac{\pi}{1-\pi}\right) \\ &\Rightarrow \log\left(\frac{\pi}{1-\pi}\right) = \beta_0 + \beta_1 x\end{aligned}$$

¹Source: Generalized linear model

Course learning objectives

By the end of the semester, you will be able to . . .

- ▶ describe generalized linear models (GLMs) as a unified framework.

Course learning objectives

By the end of the semester, you will be able to . . .

- ▶ describe generalized linear models (GLMs) as a unified framework.
- ▶ explain how specific models fit into the GLM framework

Course learning objectives

By the end of the semester, you will be able to . . .

- ▶ describe generalized linear models (GLMs) as a unified framework.
- ▶ explain how specific models fit into the GLM framework
- ▶ identify the appropriate model given the data and analysis objective.

Course learning objectives

By the end of the semester, you will be able to . . .

- ▶ describe generalized linear models (GLMs) as a unified framework.
- ▶ explain how specific models fit into the GLM framework
- ▶ identify the appropriate model given the data and analysis objective.
- ▶ analyze real-world data by fitting and interpreting GLMs.

Course learning objectives

By the end of the semester, you will be able to . . .

- ▶ describe generalized linear models (GLMs) as a unified framework.
- ▶ explain how specific models fit into the GLM framework
- ▶ identify the appropriate model given the data and analysis objective.
- ▶ analyze real-world data by fitting and interpreting GLMs.
- ▶ use R for analysis and write reports

Course learning objectives

By the end of the semester, you will be able to . . .

- ▶ describe generalized linear models (GLMs) as a unified framework.
- ▶ explain how specific models fit into the GLM framework
- ▶ identify the appropriate model given the data and analysis objective.
- ▶ analyze real-world data by fitting and interpreting GLMs.
- ▶ use R for analysis and write reports
- ▶ effectively communicate results from statistical analyses to a general audience in writing.

Course topics

Generalized Linear Models

- ▶ Review of distributions, likelihoods, and regression
- ▶ Introduce models for non-normal response variables
- ▶ Estimation, interpretation, and inference
- ▶ Mathematical details of GLMs as a unified framework

Research Article

Intersectionality of Race and Question-Asking in Women After Right Hemisphere Brain Damage

Danai Kasambira Fannin,^a Jada Elleby,^a Maria Tackett,^b and Jamila Minga^c

^aDepartment of Communication Sciences and Disorders, North Carolina Central University, Durham ^bDepartment of Statistical Science, Duke University, Durham, NC ^cDepartment of Head and Neck Surgery & Communication Sciences and Department of Neurology, Vascular and Stroke Division, Duke University School of Medicine, Durham, NC

*“... we used **negative binomial regression** to model the association between the number of questions produced, race, and group after adjusting for the additional covariates age and years of education. **Poisson and zero-inflated Poisson regression models** were also considered. . . the negative binomial model was a good fit for the data given the **overdispersion** in the distribution of number of questions asked.”²*

²Fannin, D. K., Elleby, J., Tackett, M., & Minga, J. (2023). Intersectionality of Race and Question-Asking in Women After Right Hemisphere Brain Damage. *Journal of Speech, Language, and Hearing Research*, 66(1), 314-324.

Papers

Officiating bias: The effect of foul differential on foul calls in NCAA basketball

Kyle J. Anderson  & David A. Pierce

Pages 687-694 | Accepted 07 Jan 2009, Published online: 20 May 2009

 Cite this article  <https://doi.org/10.1080/02640410902729733>

 Full Article

 Figures & data

 References

 Citations

 Metrics

 Reprints & Permissions

[Read this article](#)

"... a **logistic regression model** is used to test how the likelihood of a foul is affected by which team is the home team, the foul differential, and the score differential... The logistic regression was run under several specifications... using **clustered observation standard errors**, with each game as a cluster. This is done as an attempt to adjust for the fact that **observations may not be independent** as required under the logistic specification.³

³Anderson, K. J., & Pierce, D. A. (2009). Officiating bias: The effect of foul differential on foul calls in NCAA basketball. *Journal of sports sciences*, 27(7), 687-694.

Meet your classmates!

- ▶ Get in groups of 2 - 3
- ▶ Each person in the group...
 - ▶ Introduce yourself
 - ▶ Share a boring fact about yourself
- ▶ Everyone will introduce one person from your group to the class

Course details

Pre-reqs

Pre-reqs

STA 210 and STA 230 / STA 240

Background knowledge

Statistical methods

- ▶ Linear and logistic regression
- ▶ Statistical inference
- ▶ Basic understanding of random variables

Computing

- ▶ Using R for data analysis
- ▶ Writing reports using Rmd
- ▶ Understanding of github
- ▶ Understanding reproducibility

Course toolkit

GitHub: <https://github.com/resteorts/generalized-linear-models>

- ▶ Course information and course schedule

Canvas

- ▶ Changes to Schedule
- ▶ Ed Discussion
- ▶ Homework uploads

Gradescope (link on course webpage)

- ▶ Homework uploads (make sure to upload to Canvas as well).

Ed Discussion (link on course webpage)

- ▶ Course discussion

Class Meetings

Lectures

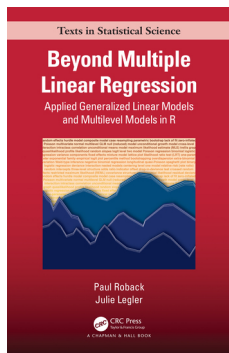
- ▶ Some traditional lecture
- ▶ Short individual and group activities
- ▶ Bring fully-charged laptop / tablet to use R

Labs (start January 17)

- ▶ Work on class assignments with TA support
- ▶ Time for clarifying questions regarding course material

Attendance is strongly expected (if you are healthy!)

Readings



- ▶ Primary textbook: *Beyond Multiple Linear Regression* by Roback and Legler
- ▶ Other texts:
 - ▶ *R for Data Science (2nd edition)* by Wickham, Çetinkaya-Rundel, and Grolemund
 - ▶ *Tidy Modeling with R* by Kuhn and Silge
- ▶ Articles and videos periodically assigned

Computing toolkit

R and RStudio

- ▶ Install R and RStudio on your laptop
- ▶ Click here for instructions to install RStudio and configure git

or

Access RStudio through Docker container provided by Duke OIT

- ▶ Reserve a generic **RStudio** container (there is no course specific container)

Canvas and Gradescope

- ▶ All homework assignments will be uploaded to Gradescope and Canvas.
- ▶ Gradescope allows more fair and balanced grading.
- ▶ Canvas allows us to check the reproducibility of your work.
- ▶ Unfortunately, there is no platform that does both (to my knowledge).
- ▶ Feedback will be given in Gradescope and is individual and private.

Ed Discussion

- ▶ Online discussion forum (like Piazza, etc.)
- ▶ Platform to ask questions about course content, logistics, assignments, etc.
- ▶ Content organized by channels. Before posting, please browse previous posts to see if your question has already been answered. If not, please post your question in the relevant channel.
- ▶ Questions about grades, absences, and other private matters should be emailed to me with “STA 310” in the subject line.

Activities & Assessment

Homework (40%)

- ▶ Individual assignments
- ▶ Combination of conceptual questions, guided analyses, and open-ended analyses
- ▶ Lowest homework grade is dropped

Quizzes (60%)

- ▶ Individual online quizzes
- ▶ Covers content since the previous quiz, including readings, lecture notes, in-class activities, and homework
- ▶ Lowest quiz grade is dropped

Grading

Final grades will be calculated as follows

Category	Percentage
Homework	40%
Quizzes	60%

See the course syllabus for letter grade thresholds.

Course community

Course community

- ▶ Uphold the Duke Community Standard:
 - ▶ *I will not lie, cheat, or steal in my academic endeavors;*
 - ▶ *I will conduct myself honorably in all my endeavors;*
 - ▶ *I will act if the Standard is compromised.*
- ▶ Commit to respect, honor, and celebrate our diverse community
- ▶ Commit to being part of a learning environment that is welcoming and accessible to everyone

Accessibility

- ▶ The Student Disability Access Office (SDAO) is available to ensure that students are able to engage with their courses and related assignments.
- ▶ If you have documented accommodations from SDAO, please send the documentation within the first week to make sure all accommodations can be put in place as quickly as possible!
- ▶ I am committed to making all course activities and materials accessible. If any course component is not accessible to you in any way, please don't hesitate to let me know.

Support

- ▶ **Office hours** to meet with a member of the teaching team.
 - ▶ Find the course schedule on the course webpage
 - ▶ Office hours begin January 16
 - ▶ Please see me after class if you have questions before then.
- ▶ **Ed Discussion** for questions about course logistics, content, and assignments
- ▶ **Email** for questions not appropriate for Ed Discussion, e.g., regarding personal matters or grades
 - ▶ Please put **STA 310** in the subject line

See the syllabus regarding additional academic and mental health and wellness resources.

Questions

Questions?