STA 221: Regression Analysis

Syllabus - Fall 2025

Teaching team & office hours

	Contact	Office hours	Location
Dr. Alexander Fisher	aaf29@duke.edu	Tu/Fr: 2-3pm	Old Chem 223B
Cathy Lee	pin.chian.lee@duke.edu	Th: 1-3pm	Zoom (see Canvas for link)
Xukun Zhu	xukun.zhu@duke.edu	We 4:30-6:30pm	Old Chem 025
Krish Bansal	krish.bansal@duke.edu	Mo: 1:30-2:30pm / We: 12-1pm	Old Chem 203A

Meetings

Lecture	Tu/Th 11:45am - 1:00pm	Old Chemistry 116
Lab 01	We 1:25pm - 2:40pm	Perkins LINK 087 (Classroom 3)
Lab 02	We 3:05pm - 4:20pm	Perkins LINK 087 (Classroom 3)

Course website: sta221-fa25.github.io

Course description

In STA 221, students will learn how linear and logistic regression models are used to explore multivariable relationships, apply these methods to real data and learn the mathematical underpinnings of the models. Students will develop computing skills to implement a reproducible data analysis workflow and gain experience communicating statistical results. Throughout the semester, students will work on a team project where they will develop a research question, answer it using methods learned in the course, and share results through a written report and presentation.

Topics include applications of linear and logistic regression, least squares estimation, maximum likelihood estimation, analysis of variance, model diagnostics, and model selection. Students will gain experience using the computing tools R and GitHub to analyze real-world data from a variety of fields.

Prerequisites

Either any STA 100-level course or STA 230, 231, or 240L and MATH 216, 218, or 221. The recommended co-requisite is STA 230, 231, or 240L.

Course material

There is no official textbook for the course; readings will be made available as they are assigned. We will use the statistical software package R both in-class, and on take-home assignments in this course. R is freely available at http://www.r-project.org/. RStudio, the popular IDE for R, is freely available at https://posit.co/downloads/. Additionally, students may access R and RStudio through Docker containers provided by Duke Office of Information Technology. Containers can be accessed at https://cmgr.oit.duke.edu/containers.

Course learning objectives

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

- analyze data to explore real-world multivariable relationships.
- fit, interpret, and draw conclusions from linear and logistic regression models.
- implement a reproducible analysis workflow using R for analysis, Quarto to write reports and GitHub for version control and collaboration.
- explain the mathematical foundations of linear and logistic regression.
- effectively communicate statistical results to a general audience.
- assess the ethical considerations and implications of analysis decisions.

Evaluation

Assignment	Description
Homework (25%)	Individual take-home assignments, submitted to Gradescope.
Midterms (45%)	Two exams with an in-class and take-home component.
Final project (15%)	Team-based final project.
Quizzes (5%)	In-class pop-quizzes.
Labs (10%)	Exercises assigned in lab, submitted to Gradescope.

A $>= 93$, A- < 93, B+ < 90, B < 87, B- < 83, C+ < 80, C < 77, C- < 73, D+ < 70, D < 67, D- < 63, F < 60

> **i** A note on quizzes
>
> On pseudo-random class days, there will be a brief quiz on the previous lectures. If you score $> 60\%$ cumulatively on your final quiz grade, you will receive full participation credit. Your lowest **two** quizzes will also be dropped.

Policies

Academic integrity

By enrolling in this course, you commit to upholding Duke's community standard reproduced as follows:

> I will not lie, cheat, or steal in my academic endeavors;
>
> I will conduct myself honorably in all my endeavors; and
>
> I will act if the Standard is compromised.

Any violations of academic integrity will automatically result in a 0 for the assignment and will be reported to the Office of Student Conduct for further action. For the Exams and Quizzes, students are required to work alone. For the Homework assignments, students may work with a study group but each student must write up and submit their own answers.

Late work

Late homework may be submitted within 48 hours of the assignment deadline. Late homework submitted within 24 hours (even 1 minute late) will receive a 5% late penalty. Late work submitted between 24 to 48 hours of the deadline will receive a 10% late penalty. Work submitted after 48 hours will not be accepted. Exams cannot be turned in late and can only be excused under exceptional circumstances. The Duke policy for illness requires a short-term illness report or a letter from the Dean; except in emergencies, all other absenteeism must be approved in advance (e.g., an athlete who must miss class may be excused by prior arrangement for specific days). For emergencies, email notification is needed at the first reasonable time.

Outside resources and generative AI statement

The use of online resources (including generative AI, as well as static webpages like StackOverflow, etc.) is strictly prohibited on in-class assignments. For take home assignments, you may make use of online resources for coding portions on assignments. If you directly use code

from a source (or use it as inspiration), you must explicitly cite where you obtained the code. If you used generative AI to create the code, you should include your prompt(s) in your citation as well. Any code that is discovered to be recycled or created by AI and is not explicitly cited will be treated as plagiarism.

> ⚠ Warning
>
> Extensive use of AI on take-home assessments will likely set you up for poor performance on graded in-class assignments.

Errors in grading

Errors in grading must be brought to the attention of the TA or instructor during office hours within 1 week of receiving the grade.