Advanced Regression

Fall 2024 Block 1

Instructor

Instructor Dr. Tyler George

Cornell College, West 311

tgeorge@cornellcollege.edu

Class Meetings

August 28th - September 18th

M-F,9am-11am and 1pm-3pm

West 201

Course Calendar

Office Hours

MWTh 3:05pm-4:05pm and by appt.

West 311

Optional Appointment

I am available far beyond these times listed. Please email me and we can set up a time to chat about class material or whatever you prefer! I will generally announce changes to office hours in class but I still suggest checking the Course Calendar to verify availability.

You Are A Priority

My goal this block is to help you learn the material. I want to first and foremost recognize that you are an individual and thus are unique and may learn uniquely. Additionally, your health and wellbeing are priority one. Learning cannot happen effectively if you don't meet your other personal needs. That all being said, I have structured the class in a way that I, from experience teaching and learning myself, think will be most beneficial for the majority of students. I promise you that I will do my best to create an inclusive and engaging learning environment. I ask that you keep an open line of communication between us for when you may need help and/or flexibility. You and your learning are why I am here.

Course Description

Following a second regression course, this class will begin with a review of multiple linear regression, but now using R. New topics will include probability distributions, likelihoods, differentiating binary vs binomial logistic regression, and poisson regression including its variants. The class of generalized linear models will then be presented which unifies all past modeling approaches. All methods are presented using realistic case studies. Conducting and communicating the modeling process including exploratory data analysis, model exploration and selection, and inferences are all emphasized. Prerequisite(s): STA 202 and DSC 223. Alternate years.

Learning Objectives

This course supports the Educational Priorities and Outcomes of Cornell College with emphases on knowledge, communication, and reasoning. Specifically, the learning objectives of this course are: - Learn statistical distributions, likelihoods functions, and types of models including poisson regression, quasi-logistic regression, multilevel models, and polynomial regression. - Ability to communicate statistical ideas clearly and accurately (communication). - Understand the class of generalized linear models and the models contained (knowledge, reasoning). - Ability to ascertain which type of analysis is appropriate for a particular data set and question (reasoning).

Prerequisite

To be successful in this class, you should have completed STA 201, STA 202, and DSC 223.

Open Access Books – Free!

All of materials for this class are free.

The main text book is: Beyond Multiple Linear Regression by Paul Roback and Julie Legler – it is freely available online. Chapters 1-9.

The secondary text is: An Introduction to Statistical Learning with Applications in R by Gareth James, Daniela Witten, Trevor Hastie, and Robert Tibshirani – it is freely available online. Chapter 7.

Course Site and Moodle

Our course will run from a combination of Moodle and the course website at https://stats-tgeorge.github.io/STA363_AdvReg/.

Software – No need to install

We will use a combination of technologies in this course including R, and RStudio (server). Luckily for you I have put lots of effort into setting all of this on a machine we have on campus that we will all access with a web browser! You don't need to install any – in fact for a while I prefer you don't. More on this in class. If you are an off campus student, please let me know right away, as you may need to checkout a laptop (free) from IT to work on homework from home.

You can access the RStudio Server at: http://turing.cornellcollege.edu:8787/auth-sign-in?appUri=%2F

If you have any technical problems you should contact IT as soon as possible. Submit a Work Order!

Group Work

In this class, I would like you to work in groups for a variety of reasons. A large part of this class is communicating analysis. At the beginning of the block, groups will be formed. You should expect to work with this group every day. When we work in groups in class we may decide on roles, specifically who is controlling the one screen will rotate). Group members will rotate roles between tasks to help make sure everybody is sharing work. You won't be working in a group for everything; any quizzes, and exams may be individual.

Evaluations and grades

Grade Category Descriptions

Homework:

Homeworks will be graded for correctness. I will generally evaluate a random subset of the assigned questions. The goal is the practice the application of the method and then be able to interpret the result.

Participation

This will be measure this by your class attendance, and your work on labs and class examples. With class work, please save your script files in the folder that I can view on the RStudio server. I will look at these files to see if you are following along.

Labs and Mini-Projects

This class includes a variety of related techniques where generally only 1 of them is the correct technique for particular research question. For this reason, we will do labs following a single method and mini "projects" at the end of each major new method or collection of methods. With labs, the data will be provided. With mini-projects, you will likely need to find, and clean data, and then apply the new technique.

Exams

You will have two exams this block, September 6th and 18th. Each will have two components. Component 1 will be on these dates and you will get a choice of oral or written format. Component 2 will be a take-home, open-book, open note, exam. You will be 12 hours or more to complete this component.

Assignment	Points
Homework	200
Participation	100
Labs & Mini Projects	300
Exams, two 200pts exams	400
Total	1000

Grade	Range	Grade	Range
A	93-100%	C	73-76%

Grade	Range	Grade	Range
A-	90–92%	C-	70-72%
B+	87–89%	D+	67-69%
B	83-86%	D	63-66%
B-	80-82%	D-	60-62%
C+	77-79%	F	<60%

Use of AI

I expect you to generate your own work in this class. When you submit any kind of work (including projects, exams, homeworks), you are asserting that you have generated and written the text, and **code**, unless you indicate otherwise by the use of quotation marks and proper attribution for the source. Submitting content as your own that has been generated by someone other than you, or was created or assisted by a computer application or tool, including artificial intelligence (AI) tools such as ChatGPT is cheating and constitutes a violation of our Academic Honesty policy. You may use simple word processing tools to update spelling and grammar in your assignments, but unless given permission otherwise, you may not use AI tools to draft your work, even if you edit, revise, or paraphrase it. There may be opportunities for you to use AI tools in this class. Where they exist, I will clearly specify when and in what capacity it is permissible for you to use these tools.

DISABILITIES AND ACCOMODATIONS POLICY

Cornell College makes reasonable accommodations for persons with disabilities. Students should notify the Office of Academic Support and Advising and their course instructor of any disability related accommodations within the first three days of the term for which the accommodations are required, due to the fast pace of the block format. For more information on the documentation required to establish the need for accommodations and the process of requesting the accommodations.

ACADEMIC HONESTY POLICY

Cornell College expects all members of the Cornell community to act with academic integrity. An important aspect of academic integrity is respecting the work of others. A student is expected to explicitly acknowledge ideas, claims, observations, or data of others, unless generally known. When a piece of work is submitted for credit, a student is asserting that the submission is her or his work unless there is a citation of a specific source. If there is no appropriate acknowledgment of sources, whether intended or not, this may constitute a violation of the College's requirement for honesty in academic work and may be treated as a case of academic

dishonesty. The procedures regarding how the College deals with cases of academic dishonesty appear in The Catalog, under the heading "Academic Honesty."

Illness Policy

If you are experiencing COVID-19 symptoms, do not attend class. Perform a home test or contact Director of Student Health Services Lynn O'Brien at **student_health@cornellcollege.edu** immediately to arrange a COVID-19 test at the Health Center. If you need to isolate due to COVID-19, or if you become unable to attend class for any other health reason, contact me as soon as possible to determine if you are able to continue in the class. A Withdrawal for Health Reasons may be required.

Mandatory Reporter Reminder

It is my goal that you feel supported and able to share information related to your life experiences during classroom discussions, in your written work, and in any one-on-one meetings with me. You should also know that all Cornell College faculty and staff are mandatory reporters. This means that I will keep information you share with me private to the greatest extent possible. However, I am required to share information regarding sexual assault, abuse, criminal behavior, or about a student who may be a danger to themselves or to others. If you wish to speak to someone confidentially who is not a mandatory reporter, you can schedule an appointment with one of the counselors in the Ebersole Health and Wellbeing Center or contact the College Chaplain, Rev. Melea White, at **mwhite@cornelllcollege.edu**.