Stat 5100: Modern Regression Methods (QI/CI)

Section MW1, Spring Semester 2021
Monday, Wednesday, and Friday, 1:30-2:20 p.m.

Instructor

Name: Dr. Brennan Bean
Email: brennan.bean@usu.edu (please use Canvas for email communication)
Office: Animal Science 203 (only available via zoom this semester)
Office Hours: Mondays and Wednesdays (not Fridays): 2:30 – 4:00pm (or by appointment)

Prerequisites

From the catalog: *STAT 2000 or STAT 3000 with a C- or better.* Similar courses are also acceptable.

Introduction

Statistics is "the science that solves data problems."[1] Foundational to this science are models that appropriately predict quantities/probabilities given available information. This course is designed to help students appropriately **create, evaluate, and implement** modern regression models using statistical software. Students will also learn to **communicate** the details of their model construction and evaluation and **recommend decisions** based on their model results. Experience with these regression models will **prepare** students for further studies in statistics and data science.

Teaching Assistant (TA)/Undergraduate Teaching Fellow (UTF)

TA: Ethan Ancell
UTF: Ruth Taylor
Please refer to the syllabus tab on the course canvas page for the days/times of office hours for both the TA and the UTF. Both students are qualified to help graduate and undergraduate students on any assignment, though R programming questions should be addressed to Ethan.

Writing Fellows

This course includes three projects (two individual, one group) which all require written reports. You will be required to have each individual report reviewed by a writing fellow prior to final submission. The writing fellows will not grade the quality of your work, but successfully completing meetings with writing fellows is part of your grade.
(Alternative for Graduate Students): Graduate students are welcome to have their papers reviewed by the writing fellows. Graduate students may alternatively choose to have their paper reviewed by a USU faculty member (besides the instructor). To receive full points for the assignment, the faculty member must email the teaching assistant directly with annotated comments on your paper draft.

Instructor Commitments

As a research faculty member, I am expected to devote about 50% of my efforts on research, 45% to teaching, and 5% to service activities. The intent of this course design is to make ensure I provide quality instruction and timely feedback without sacrificing the other aspects of my faculty responsibilities or personal life. **For this reason, I will only respond to student emails and Canvas discussion questions (see following section) once every weekday between noon and 6pm (excluding holidays).** Please expect a reply/answer from me during the full window of time that occurs after you send your message/question.

[1] Dr. Bin Yu, University of California at Berkeley

Asking Questions

Students are strongly encouraged to ask questions during class and office hours. Students are also encouraged to ask questions related to course assignments using the "Discussions" tab on the course canvas page. Posting assignment-related questions on the discussion board provides opportunities for other members of the class to answer questions. **Keep in mind, however, that you should NOT post entire blocks of code to answer questions from other students.** The discussion boards work best when all of us actively engage in both asking questions and providing answers. Note that personal/sensitive questions should not be posted to the discussion board but rather sent as a canvas email message. Conversely, homework related questions should not be sent as an email.

Grades

Final scores in this course will be determined by the number of points earned divided by 1,000. Thus, each 10 points on an assignment, regardless of category, represents 1% of the final score. The following grade scale is guaranteed, but the instructor reserves the right to adjust this grade scale in favor of the students if necessary.

Percent	Grade		Percent	Grade
93-100	A		77-79.9	C+
90-92.9	A-		73-76.9	C
87-89.9	B+		70-72.9	C-
83-86.9	B		65-69.9	D+
80-82.9	B-		60-64.9	D

Assignments

All assignments are due by 8:00pm on the listed due date (NOT midnight). Excluding the policy outlined for the first six homework assignments, late assignments will not be accepted. This is a communications intensive and quantitative intensive course. As such, points in this course will be awarded based on students' ability to **appropriately analyze and evaluate data** in a regression modeling framework and **effectively communicate** results both orally and in writing. Assignments and projects are designed to be writing intensive and programming intensive. All programming-related assignments are expected to be completed using SAS or R statistical software. For those who prefer different statistical software platforms, please remember that having a knowledge of multiple programming languages is an essential skill in both academia and industry. For those surprised by the writing-intensive nature of this quantitative course, please remember that the intensity is required to meet the general education requirement for undergraduates and a crucial experience in scientific writing for graduate students.

Engagement/Preparation (50 points)

There will be a series of at least six assignments (10 points each) designed to encourage discussion on course topics and prepare students to succeed on homework, quizzes, and papers. These assignments will be due via electronic submission on Canvas and points will be awarded based on completeness. Submissions that are too hard to read (in the case of a scanned document) or are clearly an inadequate attempt will receive half or no points. **Only your five highest engagement/preparation scores will be retained.**

Quizzes (200 points)
There will be six canvas quizzes (40 points each) in this course that will cover content from the previous two weeks of lecture as well as the previous homework assignment. Each quiz (except the last one) will have a time limit of one hour and can be taken only once. You can use any course resources to complete this quiz, **but you must work on your own** (excluding the last quiz, which can be completed in groups). Any attempts to consult with other people, in or out of the class, on your quiz is considered cheating. Discussing the content of the quiz with any member of the class before the official quiz close date is also considered cheating. **Only your five highest quiz scores will be retained.**

Homework (350 points)
There will be 7 homework assignments (and a syllabus quiz) this semester assigned approximately every other week. Late homework assignments will receive a 20% deduction from the original score for each day they are late, except for the final homework assignment for which late submissions will not be accepted. All assignments must be typed (no scanned, handwritten work). While students *should* work together on homework assignments, each student should write the answers and their code on their own. **Directly copied code or answers will be considered cheating**. Note also that there are **no dropped scores for homework assignments**.

Project 1 (75 Points)
Students will conduct a simple linear regression analysis using data provided by the instructor and write a report based on their results. Papers will be drafted using a template provided by the instructor. Points will be awarded for adequate, well-written responses to the questions asked in the provided template. Students will also be required to have their papers reviewed by the writing fellows.

Project 2 (125 Points)
Students will conduct a multiple linear regression analysis using data provided by the instructor and write a report based on their results. Points will be awarded for well written reports that adequately demonstrate the appropriateness, predictions, and implications of the final model. Students will be required to have their papers reviewed by the writing fellows and will also be required to receive and provide peer review to 2-3 other students in the course.

Final Project (200 points)
Students will work remotely in groups of 3-4 to create a multiple linear regression model on a dataset of their choice. Students will also compare their regression model results to the results from at least one other "modern" regression method discussed in class. Points for the project will be divided amongst a project proposal, written paper, oral presentation, and response to review. Points will also be awarded based upon a group member evaluation provided near the end of the project. Note that part of your final project grade will involve oral presentations that will occur during our **scheduled final time on Friday, April 30th from 1:30-3:20pm. Virtual class attendance is required on this day to receive full final project presentation points.**

Experience Cache Valley (10 points – extra credit)
Students have the option to participate in an activity in or immediately surrounding Cache Valley and write about their experience. Activities in clear violation of COVID-19 related public health recommendations will not receive credit. Adaptations for students not living in or near Cache Valley will be made on a case by case basis. See Canvas for assignment details.

Materials

There are two recommended, but not required, textbooks for this course:

- *Applied Linear Regression Models (4th Edition),* Kutner et al. (ISBN: 978-0073014661)
- *Forecasting and Time Series: An Applied Approach (3rd Edition),* Bowerman and O'Connel (1993)

All necessary materials from these references will be provided to students via Canvas.

Academic Honesty

In this course, cheating is defined as **any attempt made by a student to deceive** the instructor in the representation of their work. This definition includes all forms of plagiarism on homework and group projects, as well as any type of consultation with other students on course quizzes unless expressly authorized. Any observed instances of cheating will be reported to the University and warrants automatic failure from the course.

Student Expectations

This course has not official attendance requirement. However, students are expected to attend each class period and respectfully engage in the classroom discussion. Respect is demonstrated by being on time, keeping your webcam on to facilitate small group discussions, and muting your microphone when you are not asking a question. Note that students that do not display appropriate zoom visuals or fail to mute their microphone may be removed from the zoom meetings by the instructor. Students are responsible for all information presented in class and are **responsible to regularly check Canvas** for announcements, due dates, and other communications from the instructor. Because *all* assignments in this course are submitted electronically, no accommodations will be made for students requesting to submit late assignments, even for university excused absences. With one notable exception (see next section).

The "Life Happens" Card

All students inevitably face extenuating circumstances during their academic career. To accommodate this, each student is provided with an imaginary "life happens" card, which can be used once and only once during the semester. **The life happens card will extend the due date by one week on one (and only one) homework assignment or *final* draft of an individual writing assignment** (see exception for homework 7 in the assignment description)**.** The life happens card cannot be used to extend the due date on any other assignment, including group assignments. Note that homework assignments will not be accepted after the extended due date for partial credit. This card is intended to accommodate serious extenuating circumstances, including medical emergencies. Additional accommodations will not be made should the life happens card be used early in the semester for non-serious reasons, only to miss another assignment for more serious reasons later in the semester.

Recommendations for Course Success

It is critical to remember that understanding the concepts discussed in this course will come through **repetition.** Rarely, if ever, is a complex concept understood after hearing it once in lecture. Rather, understanding will come by:

1. **Watching topic/programming videos before each class**, study other course materials before class as needed.
2. Engaging in class discussions by asking and answering questions.
3. Reviewing material between class sessions.
4. Starting homework assignments and projects soon after they are assigned.
5. Taking time to understand the provided example code before using it.

University Resources

Disability Resource Center

USU Welcomes students with disabilities. If students have, or suspect they may have, a physical, mental health, or learning disability that may require accommodations in this course, please contact the Disability Resource Center (DRC) as early in the semester as possible (University Inn #101, 435-797-2444, drc@usu.edu, usu.edu/drc). All disability-related accommodations must be approved by the DRC. Once approved, the DRC will coordinate with faculty to provide accommodations.

Office of Equity

USU strives to provide an environment for students and employees that is free from discrimination or harassment, including sexual misconduct. Should students experience harassment or discrimination at any point during the semester inside or outside of class, please reach out to the instructor or to USU's Office of Equity (Old Main #161, 435-797-1266, titleix@usu.edu, equity.usu.edu). As a responsible employee, the instructor is required to share information about any instances of sexual misconduct (sexual harassment, sexual assault, relationship violence (dating and domestic violence, or stalking) with the Office of Equity so that students can get connected to support and reporting resources. Students can learn more about the USU resources available for individuals who have experienced sexual misconduct at sexualassault.usu.edu.

CAPS

Mental health is critically important for the success of USU students. As a student, you may experience a range of issues that can cause barriers to learning, such as strained relationships, increased anxiety, alcohol/drug problems, feeling down, difficulty concentrating and/or lack of motivation. These mental health concerns or stressful events may lead to diminished academic performance or reduce your ability to participate in daily activities. Utah State University provides free services for students to assist them with addressing these and other concerns. You can learn more about the broad range of confidential mental health services available on campus at Counseling and Psychological Services (CAPS).

Students are also encouraged to download the "SafeUT App" to their smartphones. The SafeUT application is a 24/7 statewide crisis text and tip service that provides real-time crisis intervention to students through texting and a confidential tip program that can help anyone with emotional crises, bullying, relationship problems, mental health, or suicide related issues.

IDEA Objectives

The university using the IDEA learning objectives to measure the effectiveness of courses. Of the 13 pre-defined learning objectives, the ones most relevant to this course (in order of importance) include:

- (13) Learning appropriate methods for collecting, analyzing, and interpreting numerical information.
- (1) Gaining a basic understanding of the subject.
- (3) Learning to apply course material.
- (8) Developing skill in expressing myself orally and in writing.

Disclaimer and Acknowledgements

Many of the materials and assignments in this course are adapted from materials provided Dr. John Stevens and Dr. Richard Cutler. **The instructor reserves the right to adjust this syllabus as needed, most particularly due to the evolving situation surrounding COVID-19**. Any changes will be communicated to students via a Canvas announcement.

Topics

The following list summarizes the *anticipated* list of topics that we will cover in the course:

- **Introduction and data exploration**
  - Introduction to SAS, R, and LaTeX
  - Introduction to data exploration
  - Review of hypothesis testing
- **Single-predictor linear regression**
  - basic model (Ch. 1)
  - diagnostics and remedial measures (Ch. 3)
- **Simple model inference**
  - parameters and response (Ch. 2)
  - simultaneous and misc. (Ch. 4)
- **Multiple-predictor linear regression**
  - matrix approach (Ch. 5)
  - basic model and inference (Ch. 6)
  - advanced inference (Ch. 7)
  - multicollinearity (Ch. 7.6, Ch. 10.5)
- **Variables in multiple linear regression**
  - higher-order and qualitative predictors (Ch. 8)
  - model selection and validation (Ch. 9)
  - influential observations and outliers (Ch. 10)
- **Variations on linear regression**
  - nonlinear and robust regression (Ch. 11, 13)
  - penalized regression (ridge, LASSO, and elastic net; Ch. 11)
  - nonparametric regression (LOESS, Ch. 11)
  - generalized additive models (GAMs)
  - regression trees (Ch. 11) and random forests
- **Regression with Discrete Response**
  - logistic regression – binary and polytomous
  - Poisson regression
  - log-linear models
- **Time series**
  - simple first-order autocorrelation (Ch. 12)
  - Box-Jenkins / ARIMA(p, d, q) models (Borrowman and O'Connell)
- **Quantile Regression**