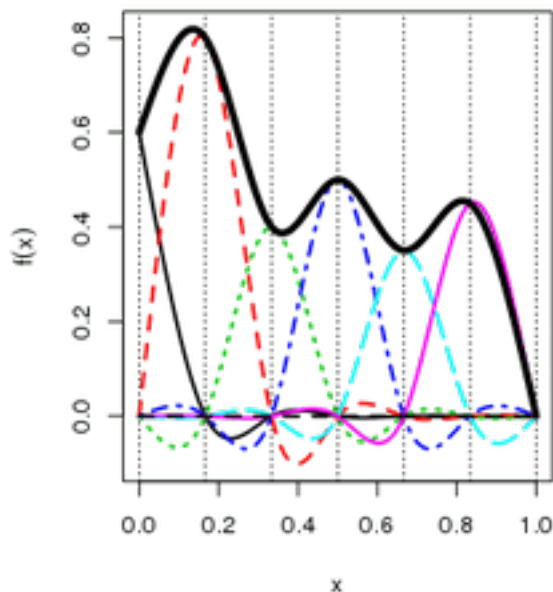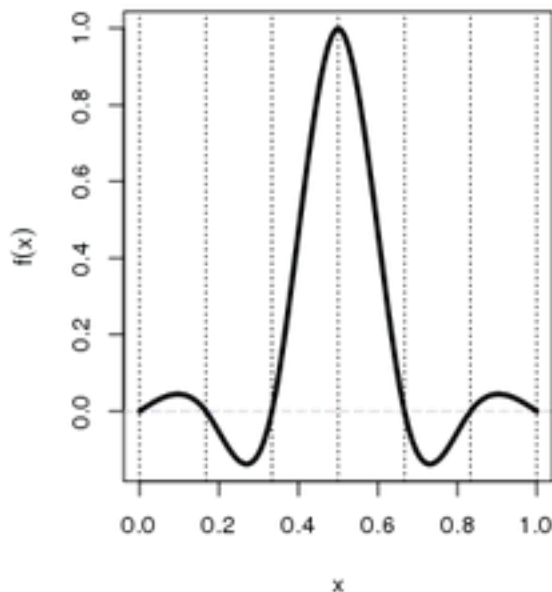# STAT 4010/5010
## Statistical Methods and Applications II



# Course Description

Probability and statistics are essential tools for engineers, data scientists, and many other professionals. Statistics can be understood as the study and application of (1) methods for reliably gathering and presenting information (descriptive statistics), and (2) methods for drawing conclusions about the world from limited information (inferential statistics). In this course, we will be primarily concerned with (2). In the prerequisite to this course—either STAT/MATH 4520/5520, or STAT 4000/5000—you studied probability theory and basic statistical inference, such as confidence intervals and hypothesis tests. In this course, we will use those tools to develop, analyze, and apply some more complicated statistical methods and models. Included in our study will be linear models, generalized linear models, nonparametric regression models, generalized additive models, tools for experimental design and causal inference, and bootstrapping. We will also learn some basic (but important) data science skills in the Jupyter/R environment, and we will consider important ethical issues that arise in the analysis of data. Our ultimate goal is to use tools from mathematics to analyze sample data and try to make predictions or draw conclusions about law-like relationships that hold in more general populations. My aim in this course is to help students become proficient in the technical aspects of statistics and data analysis; but, I also aim to help students identify when certain techniques are justified, and when certain techniques might lead them astray.

**Instructor:** Brian Zaharatos

**Office:** ECOT 318

**Office Hrs:** TBA (see Canvas)

**Email:** brian.zaharatos@colorado.edu

# A Note about Jupyter and R

On many assignments in this course, we will use the R programming language; we will run R within the web application Jupyter. Both Jupyter and R are great (free!) tools for data analysis/science, and I think they will be beneficial to you beyond this course. We will spend some time in class downloading these programs and becoming comfortable using them. Learning new languages and applications can be difficult and frustrating, but also rewarding. I'm here to help you on this journey! Taking the first in-class Jupyter assignment seriously will make all future assignments much easier.

## Textbooks

- *Linear Models with R*, Second Edition, by Julian Faraway (ISBN: 9781439887332)
- *Extending the Linear Model with R,* Second Edition, by Julian Faraway (ISBN: 9781498720960)
- Advanced Data Analysis from an Elementary Point of View, by Cosma Rohilla Shalizi (**optional**)
- Probability and Statistics with R, 2nd Edition by M.D. Ugarte, A.F. Militino, A.T. Arnholt. (**optional**)
- Other notes to be posted to Canvas.

## Learning Objectives

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

1. load a dataset into R, clean the data, and perform exploratory data analysis;
2. define, articulate important properties of, and construct appropriate linear regression models (including ANOVA/ANCOVA), generalized linear models, nonparametric regression models, and generalized additive models for real and simulated data;
3. make predictions and inferences from data based on the above mentioned models;
4. articulate the difference between an observational study and a designed experiment, and apply experimental design techniques in the correct context;
5. describe some important ethical issues that arise in the analysis of data, and develop a framework for making ethical data-driven decisions.

## Course Webpage

Course materials such as this syllabus, a course schedule, assignments, announcements, and your grades will be uploaded to Canvas. Check Canvas frequently!

# Assignments

## Exams (15% each)

There will be two midterm exams and one final exam. The date of these exams are on our course schedule (on Canvas).

## Homework (30%)

Homework will be due (roughly) once every other week (due dates and times will appear on the Canvas assignments). Late homework will not be accepted or graded, except in extraordinary circumstances. Homework assignments will have a theoretical section and a computational section. You are asked to electronically submit a single Jupyter file to Canvas; this file will contain your answers to both sections. Ideally, your answers to theoretical section will be typeset in Markdown cells (below each question) using LaTeX.  You may also embed a properly scanned pdf of handwritten answers to theoretical questions into the Jupyter file. The computational portion of the homework should include all relevant R code and output, and a write-up and interpretation of your results. You can collaborate with your classmates on the homework assignments, but you must write up the results independently of each other. Graduate students (5010) will often have additional homework questions to differentiate them from undergraduate students (4010).

## Final Project (15%)

Instructions for the final project will be given out in class. This project provides an opportunity for students to answer important scientific or business related questions by conducting a thorough data analysis using the tools from this course (or extensions of those tools). There will be a written and short presentation component to the project.

## Participation/Classwork (10%)

Participation in this course is essential for doing well. We will frequently have opportunities for class participation. The majority of this portion of your grade will come from "in-class" assignments. In addition, active participation in class, online discussion, attending office hours, and, potentially, pop quizzes can also help your participation grade.

# A Note about Linear Algebra

Linear algebra provides much of the mathematical language for a rigorous formulation of statistical models.  Consequently, some proficiency with linear algebra will be important for a strong conceptual understanding of the models used in this course, especially for STAT 5010 students. The beginning of the semester will provide an opportunity for students learn some essential tools of linear algebra, if necessary.  Previous coursework in linear algebra (e.g., APPM 3310 Matrix Methods) will be sufficient.

# Policies

## Trigger Warning

It is possible that discussions in this course, especially those pertaining to real-world data, could be potentially disturbing or traumatizing. If you feel the need to leave class during a discussion that you find disturbing or traumatizing, for however long, you may do so without academic penalty. You will, however, be responsible for any material you miss. If you do leave class for a significant time, please make arrangements to get notes from another student (or see me). If a topic is disturbing to you to the extent that you do not feel comfortable working on it, I am happy to try to make reasonable accommodations, e.g., work with you on a different topic that demonstrates the same (or similar) learning objectives.

## Classroom Behavior

Both students and faculty are responsible for maintaining an appropriate learning environment in all instructional settings, whether in person, remote or online. Those who fail to adhere to such behavioral standards may be subject to discipline. Professional courtesy and sensitivity are especially important with respect to individuals and topics dealing with race, color, national origin, sex, pregnancy, age, disability, creed, religion, sexual orientation, gender identity, gender expression, veteran status, political affiliation or political philosophy. For more information, see the policies on classroom behavior and the Student Conduct & Conflict Resolution policies.

## Requirements for COVID-19

As a matter of public health and safety, all members of the CU Boulder community and all visitors to campus must follow university, department and building requirements and all public health orders in place to reduce the risk of spreading infectious disease. Students who fail to adhere to these requirements will be asked to leave class, and students who do not leave class when asked or who refuse to comply with these requirements will be referred to Student Conduct and Conflict Resolution. For more information, see the policy on classroom behavior and the Student Code of Conduct. If you require accommodation because a disability prevents you from fulfilling these safety measures, please follow the steps in the "Accommodation for Disabilities" statement on this syllabus.

CU Boulder currently requires masks in classrooms and laboratories regardless of vaccination status. This requirement is a precaution to supplement CU Boulder's COVID-19 vaccine requirement. Exemptions include individuals who cannot medically tolerate a face covering, as well as those who are hearing-impaired or otherwise disabled or who are communicating with someone who is hearing-impaired or otherwise disabled and where the ability to see the mouth is essential to communication. If you qualify for a mask-related accommodation, please follow the steps in the "Accommodation for Disabilities" statement on this syllabus. In addition, vaccinated instructional faculty who are engaged in an indoor instructional activity and are separated by at least 6 feet from the nearest person are exempt from wearing masks if they so choose.

If you feel ill and think you might have COVID-19, if you have tested positive for COVID-19, or if you are unvaccinated or partially vaccinated and have been in close contact with someone who has COVID-19, you should stay home and follow the further guidance of the Public Health Office (contacttracing@colorado.edu). If you are fully vaccinated and have been in close contact with someone who has COVID-19, you do not need to stay home; rather, you should self-monitor for symptoms and follow the further guidance of the Public Health Office (contacttracing@colorado.edu).

## Accommodation for Disabilities

If you qualify for accommodations because of a disability, please submit your accommodation letter from Disability Services to your faculty member in a timely manner so that your needs can be addressed. Disability Services determines accommodations based on documented disabilities in the academic environment. Information on requesting accommodations is located on the Disability Services website. Contact Disability Services at 303-492-8671 or dsinfo@colorado.edu for further assistance. If you have a temporary medical condition, see Temporary Medical Conditions on the Disability Services website.

## Preferred Student Names and Pronouns

CU Boulder recognizes that students' legal information doesn't always align with how they identify. Students may update their preferred names and pronouns via the student portal; those preferred names and pronouns are listed on instructors' class rosters. In the absence of such updates, the name that appears on the class roster is the student's legal name.

# Honor Code

All students enrolled in a University of Colorado Boulder course are responsible for knowing and adhering to the Honor Code academic integrity policy. Violations of the Honor Code may include, but are not limited to: plagiarism, cheating, fabrication, lying, bribery, threat, unauthorized access to academic materials, clicker fraud, submitting the same or similar work in more than one course without permission from all course instructors involved, and aiding academic dishonesty. All incidents of academic misconduct will be reported to the Honor Code (honor@colorado.edu); 303-492-5550). Students found responsible for violating the academic integrity policy will be subject to nonacademic sanctions from the Honor Code as well as academic sanctions from the faculty member. Additional information regarding the Honor Code academic integrity policy can be found on the Honor Code website.

# Sexual Misconduct, Discrimination, Harassment and/or Related Retaliation

CU Boulder is committed to fostering an inclusive and welcoming learning, working, and living environment. The university will not tolerate acts of sexual misconduct (harassment, exploitation, and assault), intimate partner violence (dating or domestic violence), stalking, or protected-class discrimination or harassment by or against members of our community. Individuals who believe they have been subject to misconduct or retaliatory actions for reporting a concern should contact the Office of Institutional Equity and Compliance (OIEC) at 303-492-2127 or email cureport@colorado.edu. Information about university policies, reporting options, and the support resources can be found on the OIEC website.

Please know that faculty and graduate instructors have a responsibility to inform OIEC when they are made aware of incidents of sexual misconduct, dating and domestic violence, stalking, discrimination, harassment and/or related retaliation, to ensure that individuals impacted receive information about their rights, support resources, and reporting options. To learn more about reporting and support options for a variety of concerns, visit Don't Ignore It.

# Religious Holidays

Campus policy regarding religious observances requires that faculty make every effort to deal reasonably and fairly with all students who, because of religious obligations, have conflicts with scheduled exams, assignments or required attendance.  In this class, please message me with any due dates that conflict with religious holidays. I am happy to make an accommodation.

See the campus policy regarding religious observances for full details.