STAT 517: Advanced Statistical Models

Fall 2024

Instructor:	Ray Bai	**Time:**	MWF 1:10 PM – 2:00 PM
Email:	RBAI@mailbox.sc.edu	**Place:**	LeConte 103

Course Page:

https://blackboard.sc.edu/   (Check regularly for announcements and homework assignments)

Instructor Office Hours: Mondays and Wednesdays 3:30-4:30 pm in LeConte 207 or by appointment

Teaching Assistant: TBA

TA Office Hours: TBA

Course Description: The purpose of this course is to cover more advanced statistical models than those covered in STAT 515-516. Whereas the sequence STAT 515-516 thoroughly covered linear regression, STAT 517 will introduce models that go beyond the "traditional" linear regression model in several ways.

First, we will study *generalized linear models*, which will allow us to fit regression models where the response variable is not necessarily continuous (e.g. binary, multi-categorical, and count responses). Next, we will introduce *random effects* and *mixed effects models* for clustered and/or longitudinal data. Finally, we will relax the linearity assumption and study nonparametric and semiparametric models. The tentative schedule of topics is:

- **Week 1-2**:   review of R and linear regression, overview of generalized linear models

- **Week 3-5**:   models for binary responses

- **Week 6-7**:   models for multinomial and ordinal responses

- **Week 8-9**:   models for count responses

- **Week 10-12**: random effects and mixed effects models

- **Week 13-15**: nonparametric and semiparametric regression and classification

Learning Outcomes:

1. Be able to fit models to data when the classical linear regression model is inappropriate.

2. Check model assumptions of statistical models and conduct estimation and inference for model parameters.

3. Be able to analyze real datasets and draw meaningful conclusions from the data analysis.

Prerequisites: Grade of C or higher in STAT 516 or STAT 513

Main References: We will use typed handouts prepared by the instructor. Parts of these lecture notes are *not* complete and will be filled in during lecture. There is **no required** textbook in this class. However, the material covered in the class is predominantly based on the following textbook:

- Faraway, J. J. (2016). *Extending the Linear Model with R, Second Edition.* Chapman & Hall/CRC.

Computing: We will use the software R for computing and data analysis. You can download R for free from https://www.r-project.org/. The basics of R will be taught in class, and R will be used for data analysis. Extensive programming experience in R is not required to take this class.

Homework: There will be five homework assignments. Homework assignments require the use of R. Students are allowed to discuss the problems and work together with their classmates, but each student must write up and turn in their **own** solution. **You may <u>not</u> post any of the course material to a website like Chegg to solicit answers from people outside the class.** This will be monitored, and should this be discovered, the instructor will contact the website to obtain the poster ID, the answerer's ID, the viewers' IDs, and e-mails and IP addresses for all offenders. Doing so will result in disciplinary action.

Exams: There will be one midterm and one final exam. These exams are take-home exams and involve data analysis in R. **Collaboration and discussion with classmates is <u>not</u> allowed for the exams. Posting exam questions on a website like Chegg is also considered cheating and is <u>not</u> allowed.** Violations of this policy will result in a 0 on the exam and possibly other disciplinary action by the USC Office of Academic Integrity.

All assignments and exams will be released on Blackboard at a certain time, and students will have a fixed time window in which to submit their solutions through Blackboard. Late submissions will be penalized by a deduction of 1/3 of the points per day (i.e. a grade of 0 if more than two days late).

Graduate students: In accordance with Graduate School policy, graduate students who are enrolled in this course for credit must be given extra work that, if not completed, will cause the letter grade to drop by one whole letter. Graduate students taking this class for credit are **<u>required</u>** to write a report in the style of an academic article that either:

- discusses an article from the academic literature that employs methods discussed in the class, **OR**

- analyzes an original dataset not discussed in the class using methods covered in the class.

More details will be discussed individually with the graduate students in this class. Graduate students who do not complete this paper satisfactorily will have their grade decreased by 11 points (A $\to$ B, B+ $\to$ C+, etc.). Undergraduate students are **not** required to do this extra assignment.

Grading: Your grade will be determined according to the following distribution:

- Homework: 30% (6% for each assignment)

- Midterm: 30%

- Final Exam: 40%

The grading scale is as follows: 90-100 = A, 87-89 = B+, 80-86 = B, 77-79 = C+, 70-76 = C, 67-69 = D+, 60-66 = D, 59 and below = F.

Accommodation: If you need special accommodations for examinations or any other aspects of the course, please contact me before or during the first week of the semester. Note that reasonable accommodations are available for students with a documented disability. If you have a disability and may need accommodations to fully participate in this class, contact the Office of Student Disability Services by phone (803-777-6142) or e-mail sasds@mailbox.sc.edu. All accommodations must be approved through the Office of Student Disability Services.

Honor Code: See the Carolinian Creed in the *Carolina Community: Student Handbook and Policy Guide*. The *minimum* punishment for violations of the USC Honor Code is a grade of zero for the work in question. In accordance with university policy, there may be other punishments, including an automatic F in the class and/or expulsion from the university.