

# STAT 517: ADVANCED STATISTICAL MODELS

Fall 2022

---

<b>Instructor:</b> Ray Bai	<b>Time:</b> MWF 2:20 PM – 3:10 PM
<b>Email:</b> <a href="mailto:RBAI@mailbox.sc.edu">RBAI@mailbox.sc.edu</a>	<b>Place:</b> LeConte 103

---

## Course Page:

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

**Instructor Office Hours:** Tuesdays and Thursdays 1:00-2:00 pm in LeConte 207

**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 repeated measures and longitudinal data. Then we will relax the linearity assumption and study nonparametric and semiparametric models, including additive models and tree-based models. If time permits, additional topics such as penalized regression will also be discussed. The tentative schedule of topics is:

- **Week 1:** review of R and linear models
- **Week 2-4:** models for binary responses
- **Week 5-7:** models for count responses and contingency tables
- **Week 8:** models for multinomial and ordinal responses
- **Week 9-10:** random effects and mixed effects models
- **Week 11-13:** nonparametric and semiparametric regression models
- **Week 14-15:** penalized regression models

## 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 513 or STAT 516

**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. Thus, it is in your best interest to attend lectures.

There is no required textbook in this class. However, the material covered in the class is predominantly based on the following textbooks:

1. Faraway, J. J. (2006). *Extending the Linear Model with R*. Chapman & Hall/CRC.
2. James, G., Witten, D., Hastie, T., and Tibshirani, R. (2013). *An Introduction to Statistical Learning*. Springer.

**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 mainly used for data analysis. Students do not need to write very complex functions or programs in R. Students interested in learning to program more complicated things in R should enroll in STAT 540: Computing in Statistics.

**Homework:** There will be six 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 not 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. Students may use their notes and books. **Collaboration and discussion with classmates is not allowed for the exams. Posting exam questions on a website like Chegg is also considered cheating and is not allowed.** To reiterate, the instructor is able to obtain the information of users who utilize an external website. Violations of this policy may result in a 0 on the exam, an F for the course, and/or punishment by the USC Office of Academic Integrity.

The exams and datasets 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).

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

- Homework: 30% (5% 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.

**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.

**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](mailto:sasds@mailbox.sc.edu). All accommodations must be approved through the Office of Student Disability Services.