STA 3180: Statistical Modeling
Spring 2022

UF Course Catalog: Overview of modern statistical modeling. Topics include linear regression, binary regression and classification, cross-validation, nonlinear regression and smoothing, tree-based methods, the bootstrap, and causal inference. Approaches will be illustrated in R. **Prerequisite:** STA 3100

Instructor: Beth Johnson
Email: ejohnson5@ufl.edu
Office: 115-B Griffin-Floyd Hall
Telephone: 352-273-1897

Instructor Office Hours: TBD, zoom link will be posted on Canvas course page
Graduate Teaching Assistants: TBD
GTA Office Hours/Location: TBD

Days: MWF
Time: 1:45 pm – 2:45 pm
Place: ROL 0115

Course Description and Objectives

Course Description: This is the third core course in the data analytics certificate* and is designed for students who are not statistics or data science majors and will introduce students to modern statistical methods essential for understanding large and complex data that arise in fields from biology to astronomy to the social sciences.

*Pending UF approval

Course Objectives:

Upon completion of this course,

Students will be able to use the statistical program R to:

Analyze data using methods for linear regression

Analyze data using methods for binary regression and classification

Analyze data using methods for nonlinear regression and smoothing

Analyze data using tree-based methods

Students will be able to explain what is necessary to make proper causal inferences.

Students will build collaborative skills by working with a group to complete a project.

Course Materials

Required Textbook: *An Introduction to Statistical Learning with Applications in R*, second edition, by James, Witten, Hastie, and Tibshirani (Springer, 2013), which can be downloaded at no cost from the website for the book (which is maintained by the authors).

Recommended Textbooks:

Hadley Wickham and Garrett Grolemund , 2017. R for Data Science, O'Reilly, Addison Wesley, download free pdf.

Jared P. Lander, 2017. T for Everyone: Advanced Analytics and Graphics, Second Editions, Addison Wesley Data and Analytics Series

Scientific Calculator (around $10 to $15): You will need a calculator capable of basic arithmetic operations and taking square roots will be needed for in-class exams. Internet-enabled electronic devices, such as cell phones or tablets, cannot be used as calculators during exams.

Web-enabled device: You will need some type of web-enabled device such as a laptop, smartphone, or tablet to use in-class to access Canvas as needed.

Course Resources

The Canvas, https://lss.at.ufl.edu/, course website will be used extensively throughout the semester to post notes and make course announcements. You must log on using your gatorlink username and password and access the course webpages from there. Important information about the course will be posted here including this syllabus, announcements, notes, assignments and your grades throughout the semester and computer output to supplement the examples done in class. Please check this site often.

Course Computer Software

Some assignments will require you to use the statistical software package, R, to analyze and visualize data. R is free and used around the world. There are now over 13,000 R packages.

The Comprehensive R Archive Network (CRAN) is the primary place to download R. The Lander and Wickham texts above describe obtaining R and RStudio. The free desktop version of RStudio is fine.

Course Approach

In this third core course of data analytics certificate, we will focus on the developing the following skills: using statistical software to analyze data, interpreting results from a statistical analysis, and stating clear conclusions in context for a lay audience.

These skills will all be assessed through various modes such as computer assignments, a class project and in-class exams.

Help

Remember to ask for help! You can come by during my scheduled office hours or make an appointment to see me. I can also answer some questions via email. *Emails received during the working week will be answered within 24 hours however emails received over the weekend may not be answered until Monday morning.*

Course Assignments

Your final course grade will be based on a combination of assessment types including computer homework assignments, in-class exams, class activities/discussion posts and a final group project. Due dates will be posted on the course schedule, on the Canvas course page and announced in class.

Class activities and participation: Participation is an important component of this course. You are expected to participate in in-class exercises, activities, and discussion posts.

Computer Homework Assignments: There will be at least five computer homework assignments referred to as CA's. Each computer assignment is like mini project and should be taken very seriously. You are required to type all solutions (unless otherwise specified). That is, you must integrate in Microsoft Word (or another similar program) the statistical results for any software package or applet (including any graphs) used to solve the problems with your typed answers to the questions. You must upload your solutions document as a pdf file electronically onto Canvas to get your assignment graded. It is your responsibility to make sure your posted file is readable. Unreadable files cannot be graded.

Due dates for all computer assignments will be posted on Canvas. There will be a 5-point penalty per day for late work and no assignments will be accepted more than two days after the due date.

Exams: There will be two exams which will require students to read and analyze statistical output and answer some conceptual questions. In case of conflict or illness, if a student is unable to take an exam at the scheduled time, they must get in touch with the instructor prior to the exam time for any arrangements to be made for a makeup. Each case will be reviewed individually. Valid and detailed documentation is a prerequisite under such extenuating circumstances. A grade of zero is the minimum punishment of any type of dishonesty on an exam.

Research Project: Students will work in teams, of no more than three people, to complete a research project based on their approved topic. Each team will write a statistical paper, of approximately 10 pages, prepare a poster, and give an oral presentation to an audience of faculty and peers. See grading rubrics.

Grading Scheme:

Exams	30%
Computer Homework Assignments	40%
Class Activities	5%
Research Project	25%

The instructor reserves the right to adjust the percentages if needed.

Your final overall numeric score is rounded to the nearest integer.

So, for example, if your average is 76.4 your grade will be 76. If your grade is 76.5, your grade will be 77.

Numeric Grade	Letter
92.5-100	A
90-92.4	A-
87.5-89.9	B+
82.5-87.4	B
80-82.4	B-
77.5-79.9	C+
72.5-77.4	C
70-72.4	C-
62.5-69.9	D
60-62.4	D-
0-59.9	E

Letter grades will be assigned according to the table shown.

Course Policies

Grading Policies:

*You have at most seven days after a graded assignment is posted to question the grade. If you do not notify me in writing of any issues with your grade within that time, then the posted score stands.

*Requirements for class attendance and make-up exams, assignments, and other work in this course as well as policies regarding absences, religious holidays, illness, and student athletes are consistent with UF Attendance Policies.

Additional make-up policy requirements:

- Every effort should be made to complete the assignment/exam during the open period. Only extreme situations will warrant a makeup. Contact the instructor prior to the exam - as soon as you realize you will be unable to take the assignment/exam at the scheduled time. Each case will be reviewed individually. Valid and detailed documentation is a prerequisite for scheduling a makeup under such extenuating circumstances.
- If you have an emergency on the day of the assignment/exam, the instructor must be contacted by midnight of the day of the assignment/exam.
- Make-ups need to be scheduled within a week from the assignment deadline. Student is responsible for scheduling.
- Additional Note: Being on vacation or booking a trip prior to the completion of the semester is not a valid reason to request a makeup. Please reference the most recent Academic Calendar, https://catalog.ufl.edu/UGRD/dates-deadlines/pdfs/.

*If you have a disability that requires academic accommodation, contact the Disability Resource Center (DRC). The DRC will provide documentation to the students who must then provide this documentation to the instructor when requesting information. You must submit this documentation prior to submitting any assignments for which you are requesting accommodation.

* Incompletes are only assigned when extraordinary circumstances (such as an accident, or extended hospitalization), arising after the date for dropping the course, prevent the student from completing the course requirements. Having a failing grade in the course is not a valid reason for requesting an Incomplete. Information on Medical Withdrawal can be found at https://umatter.ufl.edu/ . Information on how to Drop a class can be found in UF's Academic Catalog https://catalog.ufl.edu/ and https://catalog.ufl.edu/UGRD/academic-regulations/dropping-courses-withdrawals/

*There is no "extra credit" or forgiven grades – you are responsible for all your work done (or left undone).

Honor Code on Exams: You are required to abide by the University of Florida Student Honor Code. Any violation of the academic integrity expected of you will result in a minimum academic sanction of **a failing grade on the assignment or assessment.** Any alleged violations of the Student Honor Code will result in a referral to Student Conduct and Conflict Resolution. Please review the Student Honor Code and Student Conduct Code at sccr.dso.ufl.edu/policies/student-honor-code-student-conduct-code/

The Honor Code will be enforced for all exams.

Classroom Behavior: During class students should turn off their cellular phones and refrain from eating, drinking, reading newspapers, doing homework, listening to music, excessive talking and all other behaviors that are distracting and disrespectful to the instructor and their fellow students.

Privacy Policy: Student records are confidential. Only information designated "UF directory information" may be released without your written consent. This applies to parents or anyone else who contacts me about your grades.

Faculty Course Evaluations: Student feedback is welcomed by the instructor and beneficial to future students in the course. Students are requested to provide feedback on the quality of instruction in this course by completing a brief confidential evaluation towards the end of the semester at https://evaluations.ufl.edu. Summaries of the evaluation results can be found at https://evaluations.ufl.edu/results.

Other University Services

U Matter, We Care: Your well-being is important to the University of Florida. The U Matter, We Care initiative is committed to creating a culture of care on our campus by encouraging members of our community to look out for one another and to reach out for help if a member of our community is in need. If you or a friend is in distress, please contact umatter@ufl.edu so that the U Matter, We Care Team can reach out to the student in distress. A nighttime and weekend crisis counselor is available by phone at 352-392-1575. The U Matter, We Care Team can help connect students to the many other helping resources available including, but not limited to, Victim Advocates, Housing staff, and the Counseling and Wellness Center. Please remember that asking for help is a sign of strength. In case of emergency, call 9-1-1.

***Sexual Assault Recovery Services (SARS): Student Health Center, 392-1161**
***University Police Department, 392-1111 (or 9-1-1 for emergencies), http://www.police.ufl.edu**

Covid Policy

In response to COVID-19, the following practices are in place to maintain your learning environment, to enhance the safety of our in-classroom interactions, and to further the health and safety of ourselves, our neighbors, and our loved ones.

If you are not vaccinated, get vaccinated. Vaccines are readily available at no cost and have been demonstrated to be safe and effective against the COVID-19 virus. Visit this link for details on where to get your shot, including options that do not require an appointment: https://coronavirus.ufhealth.org/vaccinations/vaccine-availability/. Students who receive the first dose of the vaccine somewhere off-campus and/or outside of Gainesville can still receive their second dose on campus.

You are expected to wear approved face coverings at all times during class and within buildings even if you are vaccinated. Please continue to follow healthy habits, including best practices like frequent hand washing. Following these practices is our responsibility as Gators.

- Sanitizing supplies are available in the classroom if you wish to wipe down your desks prior to sitting down and at the end of the class.
- Hand sanitizing stations will be located in every classroom.

If you sick, stay home and self-quarantine. Please visit the UF Health Screen, Test & Protect website about next steps, retake the questionnaire and schedule your test for no sooner than 24 hours after your symptoms began. Please call your primary care provider if you are ill and need immediate care or the UF Student Health Care Center at 352-392-1161 (or email covid@shcc.ufl.edu) to be evaluated for testing and to receive further instructions about returning to campus. UF Health Screen, Test & Protect offers guidance when you are sick, have been exposed to someone who has tested positive or have tested positive yourself. Visit the UF Health Screen, Test & Protect website for more information.

- Course materials will be provided to you with an excused absence, and you will be given a reasonable amount of time to make up work.
- If you are withheld from campus by the Department of Health through Screen, Test & Protect you are not permitted to use any on campus facilities. Students attempting to attend campus activities when withheld from campus will be referred to the Dean of Students Office.

- Continue to regularly visit coronavirus.UFHealth.org and coronavirus.ufl.edu for up-to-date information about COVID-19 and vaccination.

Tentative Course Schedule

Week	Topic
1	Review of Simple Linear Regression
2	Multiple Linear Regression
3	Model Selection
4	Residuals and Diagnostics
	Exam 1
5	Review of Binary Regression
6	Classification: Linear/Quadratic Discriminant Analysis
7	Cross Validation
8	Estimating Error Rates
9	Nonlinear Regression
10	Smoothing
	Exam 2
11	Decision Trees
12	Random Forests
13	Bootstrap
14	Causal Inference and A/B testing
15	Project Presentations

Rubric for Oral Presentation

	1	2	3	4
Organization	Talk lacks any semblance of logical structure.	Relationship between portions of talk is unclear AND student understanding is compromised by this lack.	Occasionally the relationship between parts of the talk is unclear OR the organization of the talk, while existent, is not conducive to student understanding.	Logical development of talk is clear and appropriate. Organization of talk enables students to understand the material.
Clarity of communication	The talk is confusing throughout.	The theme of the talk is unclear OR many points in the talk are unclear. Incorrect statistical language is used.	Speaker generally communicates clearly, but some fine points are unclear, or some incorrect statistical language is used.	Speaker communicates coherently and clearly, and precise statistical language is used.
Representations (graphs, tables, equations, …) and notation	Speaker uses unclear or inappropriate representations, or speaker uses no representations whatsoever.	Speaker uses minimal representations or uses representations ineffectively.	Speaker effectively uses clear and appropriate representations, including technology, manipulatives, handouts, and/or writing on board.	Speaker effectively uses clear and appropriate representations, including technology, manipulatives, handouts, and/or writing on board, and addresses connections between different representations.
Knowledge of subject matter	Speaker shows errors in knowledge of statistical concepts.	Speaker explains concepts but in a rudimentary form.	Speaker clearly articulates statistical concepts.	Speaker clearly and coherently articulates statistical concepts and develops connections among concepts.
Mathematical proof	Speaker does not understand proof.	Speaker lacks appropriate reasoning and methods in explaining proof.	Speaker explains proof with acceptable reasoning and methods, and partially supplies detail.	Speaker thoroughly explains proof using sound reasoning, appropriate methods, and supplies pertinent detail.

Grading Expectations: Written Report Checklist

Your report will be graded according to how you meet the criteria detailed below:

Report Goal
Do you have a clearly defined goal or goals?
Intended Readers
What are the backgrounds of the intended readers?
What are they likely to know about the topic?

Title
Does the title completely describe the subject of the report?
Are there any unnecessary words?

Abstract
Is the abstract understandable without reading the rest of the report?
Does it summarize the objectives, results and implications of your study?

Introduction
Can the reader understand why the research was done?
Can the reader understand how the research fits into existing knowledge of the upper-level STA course topic?
Does the final paragraph outline the remainder of the report?

Methods and Results
Are the statistical methods and results clearly presented?
Are all tables and figures numbered, titled and clearly introduced in the prose?
Do the tables and figures tell their part of the story?

Discussion
Is your discussion consistent with the statistical results?
Have you clearly distinguished conjecture from fact?
Is your discussion free from statistical jargon?
Are the implications of your study clearly discussed?
Have you pointed out any shortcomings or limitations in your study?

Conclusion
Have you summarized the main points of your report?
References/Appendices
Have you given complete citations for all sources referenced in your report?
Are the citations consistent with the style you are using?
Are the appendices appropriately titled, identified by a letter or number and referred to in the body of the report?

Final Checks
Are the pages numbered?
Have you used a spell checker?
Have you carefully proofread your report after making all the changes?
Has a colleague critiqued your report?

Poster Presentation Evaluation Criteria

Presenter(s): _____ Reviewer: _____

Please rate the poster/presenter utilizing the 0 to 3 scale where
0=No attempt, 1=Developing, 2=Competent and 3=Exemplary

1. Statement of Research Problem/Rationale:

a) Clearly written questions or hypotheses being addressed	0	1	2	3
b) Well-written rationale/justification for the study	0	1	2	3
c) Information includes an explanation of key concepts/theories	0	1	2	3

2. Methods (Explanation/Appropriateness):

a) Clear description of data collection procedures if applicable	0	1	2	3
b) Clear description of statistical methods used	0	1	2	3
b) Methods are appropriate to address aim/question	0	1	2	3
c) Sufficient written information to suggest that statistics are calculated correctly	0	1	2	3

3. Analysis/Results:

a) Figures/Tables used to appropriately and clearly present the data	0	1	2	3
b) Findings are presented clearly and accurately	0	1	2	3
c) Analysis is well written and appropriately applied	0	1	2	3

4. Written Conclusion/Discussion:

a) Addressed study's problem/question	0	1	2	3
b) Conclusions are sufficiently supported by results	0	1	2	3
c) Results are placed into broader framework	0	1	2	3
d) Importance of findings is addressed	0	1	2	3

5. Verbal Interaction with Presenter(s)

a) Presenter(s) summarized study clearly	0	1	2	3
b) Length of poster summary by presenter(s) was appropriate	0	1	2	3
c) Presenter(s) demonstrated full knowledge of the material and can explain and elaborate on questions	0	1	2	3

6. Poster Design

a) Effective overall aesthetic /organization of poster	0	1	2	3
b) Flow of information is logical and facilitates understanding	0	1	2	3
c) Fonts are easy to read and vary appropriately for headings and text	0	1	2	3
d) Correct grammar and spelling	0	1	2	3

7. Overall

a) Question is interesting	0	1	2	3
b) Study is original	0	1	2	3
c) Potential to contribute to the field	0	1	2	3