AESC 6100: Applied Agricultural Data Science – Fall 2024

Dr. Paul Severns
Paul.Severns@uga.edu
2315 Miller Plant Sciences
Office Hours: By appointment - email to schedule

Lecture: Tu & Th 2:20-3:35, Miller Plant Sciences (MPS) 2102
Also available through Zoom for remote participation only for Griffin and Tifton Campuses

AESC 6100 Course Description (CAPA)**:** Covering a range of modern approaches for analyzing and interpreting structured and unstructured data commonly encountered in agricultural systems, this course serves as a foundation of descriptive and predictive analytics in the agri- food sciences and provides context for more specialized courses in data analytics.

Pre-requisites: Department Approval, highly recommended that you take a graduate level statistics course prior to this class

The course syllabus is a general plan for the course; deviations announced to the class by the instructor may be necessary.

Course Objectives and Outcomes

1). Introduction to data analyses that are not typically taught in graduate level statistics service classes.

2). Learn how these analyses are often used and how they probably shouldn't be used (but frequently are).

3). Learn the basic underpinnings of presented statistical analyses, their assumptions, strengths and weaknesses.

4). Learn how statistical and biological/economic/behavioral inferences are related by philosophy and data analysis.

5). Develop criteria for deciding when to interpolate, extrapolate, and identify when inferences are supported and when they are not.

5). Develop and enhance skills that enable you to critically evaluate statistical analyses as a reviewer (e.g. Did the researchers use the analysis in an acceptable manner?) and a consumer (e.g. Is this R package reliable?).

6). Apply your acquired knowledge to determine whether statistical analyses in published papers and statistical analysis packages are reliable and defensible.

****7). *Gain critical experience in understanding and explaining to others why you chose to use a specific analysis over others (e.g. assumptions/data types are better fit to the analysis you chose over the alternatives).*

Course grades (3 components) from 115 total points (A = 100%-93%, A- < 93% to 90%, B+ = 89%-87%, B =86%-83%, B- 82%-80%, C = 79%-70%, D = 69%-60%, F < 60%)

Assignment 1). Print, sign (digital or hand signature is acceptable) and return a hard (or email electronic copy) of this syllabus end page to me **before Sept 2, 2023**. This signed copy documents that you have read and understand the course policies, expectations, and evaluation (5 pts).

Assignment 2). Literature critique of a published paper of your choice (25 pts each x highest 2 scores = 50 pts). Takes place of midterms. I have 5 opportunities to hand in these critiques as a "life happens" accommodation for this course. You may turn in up to three critiques but no less than two. I just take the top two grades if you turn in three critiques.

Assignment 3). In place of a final exam you will critically review a statistical analysis package of your choice. Ideally, this statistical package is one that you might use for your thesis research (45 pts).

Assignment 4). Class reading assignment short answers (prior to in class paper discussion). 15 pts total (3pts x 5 assignments) = participation grade.

A very important policy on Zoom recordings. I will record the classes on Zoom (if there are students from Tifton/Griffin campuses in class), but I will not make the recordings available unless I receive an email request from the student requesting the recording for a specific date. The request(s) will need to detail why you were absent from class. Unfortunately, past students have taken advantage of the Zoom recordings and elected not attended class. We learn from each other, and by hearing the answers to and posing questions in class. This approach is more constructive, focused on student interests, and works well when students attend the class and are actively engaged. If students do not attend the class, we all lose out on potentially important questions and contributions.

Attendance is mandatory. However, I do realize that there are instances where attendance is not possible. Please contact me through email if you either anticipate being absent or are absent due to an illness, life-event, etc., so I can help you to makeup the missed material.

Important: All assignments are to be written in your own words and are individual. If I find evidence of plagiarism and/or group work, you will receive a zero for the course and I will contact your major advisor.

Dates	Topic	Assignment Due
Week 1 (Aug 15)	Introduction, Class Structure, Policies	
Week 2	Statistics and Inferences	
Week 3	Agricultural Data Types and Challenges, Non-parametric tests and randomization	Signed Syllabus Due Before 31-Aug-24
Week 4	Model Selection Methods and Approaches	
Week 5	Function/Curve Fitting – paper*	In class reading
Week 6	Decision Trees – paper*	In class reading
Week 7	Random Forest, Boosted Regression*	**1st Critique Sept. 27** (only strictly enforced due date other than the final critique)**
Week 8	Maximum Entropy – paper*	In class reading
Week 9	Cluster and Dendrogram Analyses - paper*	In class reading
Week 10	Ordinations-PCA, PCoA, CCA*	**Suggested Lit Critique**
Week 11	Ordination NMDS – paper*	In class reading
Week 12	Text Mining, Sentiment Analysis*	**Suggested Lit Critique**
Week 13	Networks – paper*	In class reading
Week 14	Neural Networks*	**Suggested Lit Critique** A Neural Network Playground
Week 15	Thanksgiving Week –	
Week 16	Machine Learning, Additional Analyses*	**(Note: Dec. 3 last day to turn in a critique)**
Dec 5-11	Finals Week	*Analysis Package Review Due before Dec 10*

** This first literature critique must be turned in by Sept. 29.
*indicates that any of the analyses covered in these lectures are appropriate for a lit critique.
Attendance: You are expected to attend class unless illness and/or other acceptable reasons for absence occurs. Please email me if you will be absent and we can make arrangements for missing class(es).

In depth description of the 3 major assignment types in this class

In class reading assignment (15 pts total, 3 pts each, must turn in 5 assignments for the full points)

Often, I will introduce you to an analysis, explain how it works, where its weaknesses and strengths are, and some key assumptions the analysis relies upon. Because many machine learning analyses will provide a solution, it is critically important to evaluate whether the solution is reliable and the interpretation of the results justified. One good method to reinforce the initial exposure to a new analysis is to observe how that analysis was applied in a published study and discuss whether the author(s)' use and interpretation of the analysis is justified. For the weeks where I have indicated "**in class reading**" on the syllabus we will all read a paper that I select and discuss it as a group. This should help prepare you for the literature critique assignment (below) and also give you practice at picking out important assumptions of a study/statistical analyses and understanding how those assumptions may

impact the analysis and its interpretation. For each class that we have an in-class reading assignment, you will turn in a short, one-page document for which there is a template on the course eLC website. I simply ask you to identify at least 2 assumptions that you think are important in the paper we a reading as a class and give me a one to two sentence <u>explanation</u> of why those assumptions are important. **This assignment is due BEFORE 12:00 pm on the day when we discuss the paper as a class (typically Thursdays).** This assignment helps me understand what you have understood and perhaps not understood and serves as a very important guide for a productive in-class discussion.

<u>Literature Critique Instructions – Examples of high scoring critiques posted on the class eLC site</u>

For this portion of your course grade I will take your 2 highest scores from 3 assignments and those will be your points (50 points total). You may turn in 3 assignments or only 2 (no questions asked if you turn in just the 2 assignments). ***I will not accept late assignments for the first critique due by August 31, 2024 or after the final assignment due date of December 3, 2024.***

Directions:

1) More than 2 pages in length but no more than 6 pages, 1.5 spacing, 11 or 12 point font.
2) If your critique is not sufficiently proofread and poorly written (typos, grammatically incorrect sentences, and unclear writing) your critique grade will drop by at least 5 pts even if all the required components are present. **Proofread!**
3) Find a published manuscript that used one of the analyses we covered in the time period over which the critique is due (the paper can be in your field of study or not, it is your choice).
4) For the critique you must (**in your own words**):
   a. Provide the citation for the study you are critiquing and write a one paragraph summary of the study.
   b. <u>Explain</u> why the authors used the analysis and for what purpose. What were they hoping to learn or which hypothesis/(es) were they testing?
   c. Did the authors sufficiently describe the analysis and why they were using that particular analysis? Did they provide a reason for the analytical approach? <u>Explain</u>.
   d. What are the key assumptions of the analysis/study and did the authors do anything to address those assumptions?
   e. Did the authors provide enough information that if you had a similar data set you could run that same analysis? You may also need to check the citations for the analysis, as this is part of the methods. If this is the case, give the citation, read the cited paper, and report whether this citation provided the pertinent information. <u>Explain</u> what the authors included (or excluded) for their analysis description that was either useful or necessary but lacking.
   f. Do you think the author(s)' interpretation of the analysis results were reasonable and defensible? <u>Explain</u> why. Hint: Think about analysis assumptions/limitations and how the authors addressed those.

For this review you will select an analysis package, hopefully one that you will use for your thesis research, and perform a review focusing on the package's reliability, documentation, and accuracy/veracity. I am asking you to vet the statistical package you select. You may also review a statistical package featuring any of the analyses we discussed in class. For this assignment, the statistical package for review cannot be an environment (e.g. R, Python, SASS, S-PLUS) but rather an analysis package that is either stand-alone or runs within a specific programming environment. This could be an analysis package that runs in R (or any other platform) or a self-contained analysis package (e.g. STRUCTURE). Certain types of statistical packages are not appropriate for review, such as GLMM, Regression, ANOVA, MANOVA, t-tests, etc.. You need to pick an analysis package that is outside of what you would typically encounter in a graduate level, general service, statistics course. **I strongly recommend you run your choice by me before you begin your review as doing so will keep you from either doing too much work or not enough work.** The review should be at least 6 pages (excluding citations) but no more than 10 pages, 1.5-spaced, with 11-12 pt font.

Directions:

1) Provide the name, url, and citation for the statistical analysis package under review. Use Google Scholar and provide the number of times the analysis package has been cited.

2) Give a brief – 1 paragraph summary of what analysis the package is supposed to perform and how you might use it for your thesis research (or if not for your thesis research an application in your field of study).

3) Explain the analysis approach, in general terms, describing the type of data you (personally) would analyze with this statistical package.

4) Critically evaluate the statistical package and address the following questions:

a. Do the authors provide a user's manual and an example data set for analysis? Is this example data set useful to you? Explain why or why not.

b. Is the statistical analysis sufficiently described by the authors? What is the evidence of this? Explain. Consider whether the authors explain their analytical approach and if they compare it with other methods. Are there assumptions? Are those assumptions stated/addressed?

c. Do the authors provide citations for their analysis? If so, read up to three of these cited manuscripts (include the ones you read in your response) and explain why each citation did or did not provide you with enough information to evaluate the analysis itself and/or its performance/reliability.

d. If there are no citations for the analytical approach, did the authors provide a proof or direct explanation of the analysis in a way that would provide all the necessary information to reconstruct the analysis in a different platform?

5) Explain why you think the statistical analysis package you reviewed is trustworthy or not. If you found the analysis package to meet your expectations after your critical review (or to fall

shore of your expectations), reflect on the number of citations for this statistical package. How would you advise someone to use the analysis package? What are the strengths and weakness of the package?

Required Course Material

No text or course material is required by the student. All course materials will be supplied via eLC.

Policy on the use of LLMs (Large Language Models – Generative AI – ChatGPT – etc.) for written assignments

A) Submitting work and assessments created by someone or something else, as if it was your own, is plagiarism and is a form of cheating and this includes AI-generated content. Please refer to UGA's University Honor Code and Academic Honesty Policy (below).
B) Natural language processing models work by predicting what text is most likely to follow previous text based on the information it has ingested. Therefore, it can often return incorrect or false information. For example, it may return non-existent academic references.
C) The assignments in this class are specific to both the articles that you select and the statistical packages that are approved for the final project. The questions and topics you are to address are specific and will vary from paper to paper, package to package. Since I read all of the papers you select for literature critiques and the statistical package documentation you submit for your final project, LLMs are almost assuredly unlikely to be of any help to you (most likely a hinderance even).

UGA Student Honor Code: *"I will be academically honest in all of my academic work and will not tolerate academic dishonesty of others." A Culture of Honesty, the University's policy and procedures for handling cases of suspected dishonesty, can be found at* honesty.uga.edu*. Every course syllabus should include the instructor's expectations related to academic honesty.*

UGA Well-Being Resources:

UGA Well-being Resources promote student success by cultivating a culture that supports a more active, healthy, and engaged student community.

Anyone needing assistance is encouraged to contact Student Care & Outreach (SCO) in the Division of Student Affairs at 706-542-8479 or visit [sco.uga.edu](sco.uga.edu). Student Care & Outreach helps students navigate difficult circumstances by connecting them with the most appropriate resources or services. They also administer the Embark@UGA program which supports students experiencing, or who have experienced, homelessness, foster care, or housing insecurity.

UGA provides both clinical and non-clinical options to support student well-being and mental health, any time, any place. Whether on campus, or studying from home or abroad, UGA Well-being Resources are here to help.

- *Well-being Resources: well-being.uga.edu*
- *Student Care and Outreach: sco.uga.edu*
- *University Health Center: healthcenter.uga.edu*
- *Counseling and Psychiatric Services: caps.uga.edu or CAPS 24/7 crisis support at 706-542-2273*
- *Health Promotion/ Fontaine Center: healthpromotion.uga.edu*
- *Disability Resource Center and Testing Services drc.uga.edu*

Additional information, including free digital well-being resources, can be accessed through the UGA app or by visiting https://well-being.uga.edu.

Name   **5 pts (due before Aug 31)**

Signature

Date

What department are you in and what is your research project (just a couple of sentences)?

Why are you taking this course (just a couple of sentences will suffice)?

Have you taken any graduate level statistics courses prior to this class?

What year of your grad program are you into and are you a MS or PhD student?

I will not share this information with anyone, this is just for me to get to know you better and direct the class so that it is helpful for the majority of students who are taking it. These questions are just a small way that I can make this effort and let you know that I am trying to see you as person rather than an "810- number".