CLIM762: Statistical Methods in Climate Research, Spring 2024.
Quick Overview

Instructor: Timothy DelSole (**Email**: tdelsole@gmu.edu)
Office Locations and Phone Number: Research Hall Rm112 (703-993-5715)
Office Hours: By appointment (request made by email).
Catalogue Description: Introduction to a core set of statistical methods that have proven useful to modern climate and predictability research. Topics include hypothesis testing, autoregression models, analysis of variance, statistical forecasting, linear regression, cross validation, principal component analysis, detection and attribution of climate change.
Prerequisites: linear algebra, programming experience, STAT 344; or permission of instructor.
Class Meeting: In-person. Mondays, 4:30-7:10pm, Research Hall, Rm 121.
Grading: Homework: 35%; Class presentation: 10%; Midterm: 20%; Final: 20%; Project: 15%
Policies: Late assignments will be penalized one letter grade per day late (exceptions for medical reasons). Lowest homework grade is dropped before calculating total homework score. Each student is expected to present at least one homework problem per class session. Absences in class will be excused for medical reasons or attendance at professional conferences, with prior notification. Outside of class, students are expected to (1) read a chapter or view a lecture video, (2) work on current homework, (3) write up previous homework for submission.
Honor System: About 90% of the homeworks were given in previous years. Students are expected to refrain from consulting answers for previous years' homeworks.
Class Web Site: Blackboard
Course Text *Statistical Methods for Climate Scientists* by DelSole and Tippett. You do not need to purchase this book. I will distribute PDFs (with permission from the publisher). *These PDFs are protected by copyright laws. They should not be shared with anyone outside this course.*
Software Homework computations are expected to be done using the package **R**.

Class Schedule

01/22 Probability concepts
01/29 Hypothesis tests
02/05 Confidence Intervals
02/12 Hypothesis Tests based on Ranks
02/19 Stochastic Processes
02/26 Power spectra
03/04 MID-TERM EXAM
03/11 **Spring Break**
03/18 Multivariate Concepts
03/25 Linear regression: estimation
04/01 Linear regression: inferences
04/08 Principal Component Analysis
04/15 Model selection
04/22 END-OF-SEMESTER EXAM
04/29 Pitfalls of Statistical Inference
05/06 Class Presentations

1 Class Format

This class will follow a "flipped classroom" model. This model is designed to maximize our classroom interaction and your understanding of the course material. Here is how this approach will work:

- Pre-class preparation:
  - Reading and Homework: Before each class, please read the assigned chapter and/or view a lecture video, and complete all the homework problems to the best of your abilities.
  - Presentation Preparation: Be prepared to present your solutions or approaches for each homework problem, focusing on the strategies you employed.
- In-Class Activities:
  - Student Presentations: You will present your homework solutions during class.
  - Random Selection: I will randomly select students to present specific problems, so be ready to discuss any of the assigned homework problems.
  - Peer Collaboration: Classmates are encouraged to engage in discussions, offer insights, and help each other understand different approaches.
- Grading and Homework Submission:
  - In-class presentations will be graded. However, the focus will be on the effort and reasoning demonstrated, not solely on finding the correct answer.
  - All homework problems, including those discussed in class, should be submitted.

As the semester evolves, you generally will be doing three things each week: (1) reading a chapter, (2) working on a new homework set, (3) writing up a homework set, as outlined in the table below.

week	class	in-class topic	in-class HW	HW due on class day	reading	out-of-class HW
0	**01/15**	–	–	–	ch1 & R tutorial	ch1
1	**01/22**	Probability concepts	ch1	–	ch2 & R tutorial	ch2
2	**01/29**	Hypothesis tests	ch2	ch1	ch3	ch3
3	**02/05**	Confidence Intervals	ch3	ch2	ch4	ch4
4	**02/12**	Tests based on Ranks	ch4	ch3	ch5	ch5
5	**02/19**	Stochastic Processes	ch5	ch4	ch6	ch6
6	**02/26**	Power spectra	ch6	ch5	–	–
7	**03/04**	MID-TERM EXAM	–	ch6	ch7	ch7
8	**03/11**	**Spring Break**	–	–	–	–
9	**03/18**	Multivariate Concepts	ch7	–	ch8	ch8
10	**03/25**	Linear regression: estimation	ch8	ch7	ch9	ch9
11	**04/01**	Linear regression: inferences	ch9	ch8	ch10	ch10
12	**04/08**	Principal Component Analysis	ch10	ch9	ch12	ch12
13	**04/15**	Model selection	ch12	ch10	–	–
14	**04/22**	END-OF-SEMESTER EXAM	–	ch12	ch11	ch11
15	**04/29**	Pitfalls of Statistical Inference	ch11	–	–	–
16	**05/06**	Class Presentations				

2 Policy on AI tools

ChatGPT can correctly answer most homework problems in this course (I checked). However, if you simply submit answers from ChatGPT without understanding them, you will learn very little (plus you will violate university ethical policies). Effective use of AI tools requires students to take more control of their own education to make wise decisions about how to use these tools to enhance their educational experience. As a teacher, I was initially tempted to ban AI tools completely, but on the other hand, I have a responsibility to train students for future jobs, and use of these tools will likely be very commonplace in your next job. In this course, AI tools are allowed with the following guidelines.

- Definition of AI Tools: AI tools include, but are not limited to, software and online resources capable of generating content, solving problems, or providing data analysis. Examples include language models, data analysis programs, and algorithmic problem solvers.
- Permitted Use of AI Tools:
  - Supplemental Learning: You may use AI tools to enhance your understanding of course material, explore concepts, and engage in self-directed learning.
  - Clarification and Research: AI tools can be used for clarifying complex topics, conducting research, or gathering supplementary information.
- Restrictions on Use:
  - Directly submitting work generated by an AI tool as your own is strictly prohibited. This includes, but is not limited to, written assignments, code, and statistical analysis.
  - Understanding Over Automation: The focus of assignments is your understanding and application of the material. Using AI tools to bypass this learning process undermines the educational objectives of the course. The flip classroom model has been adopted to encourage you to understand solutions well enough to explain them to your peers.
- Disclosure and Transparency:
  - Mandatory Disclosure: If you use an AI tool to assist with any part of your homework, you must disclose this use. Clearly indicate which portions were aided by AI and provide a brief explanation of how it contributed to your work. If you found errors or could improve over AI tools, mention that too.
  - Ethical Considerations: Consider the ethical implications of using AI in your work. Honesty and integrity are paramount in academic pursuits.
- Academic Integrity Violations: Failure to adhere to this policy will be considered a violation of academic integrity. Consequences will be in line with the universitys academic integrity policy and may include a grade penalty or disciplinary action.
- Encouragement of Skill Development:
  - Focus on Personal Skill Enhancement: While AI tools can be beneficial, the primary goal is to develop your own skills and understanding. The use of these tools should not replace personal effort and intellectual engagement with the course material.
  - For a thoughtful discussion of when to use AI tools, I recommend this article: A Sports Analogy for Understanding Different Ways to Use AI [clickable link]

3 Coding Assignments Should be Done in R

All homework assignments must be completed using the 'R' software. This decision is based on several factors, including the suitability of 'R' for our coursework and its widespread use in statistics.

- Downloading 'R': 'R' is a free software, and you can download it from http://www.r-project.org/. Please download and install it on your computer the first week of the semester.
- Tutorial on 'R': Blackboard contains a tutorial on 'R' that covers everything you need to know for the homework assignments. Past students have found it useful, and indicate that it takes about 2-3 hours to complete. Your feedback on this tutorial is also very welcome, as it helps me improve the material for future classes.
- Why not Python or Other Languages?
  - Using a single programming language, in this case, 'R', simplifies both the grading process and in-class discussions, and ensures consistency in evaluating student work.
  - I provide solutions to all homework assignments. Due to time constraints, it is not feasible to offer equivalent codes in multiple programming languages like Python or Matlab.
  - Relevance of 'R': R is the most commonly used software among statisticians and has a vast array of statistical packages. Its relevance extends beyond academia; for instance, my last graduate student secured a job at a research center where R is exclusively used.

4 Format for Returning R codes

- Some homework assignments ask you to write R code and submit it. You should submit it using the following naming conventions:

  ```
  hw02.[last name].responses.pdf
  hw02.[last name].functions.R
  hw02.[last name].main.R
  ```

  For example, if I were submitting, the name would be `hw02.delsole.responses.pdf`.

- The file `hw02.[last name].responses.pdf` contains the following (in this order):
  1. output from running your main code
  2. written response to each question *directly after each relevant output*.
  3. after all questions have been answered, start a new page and copy/paste the text of your R code `hw02.[last name].main.R`.
  4. After the main code, start a new page and copy/paste the text of `hw02.[last name].functions.R`.

  More details: When printing the output from running your R code, copy the *exact* output. Do not hand type the output and report just the first few significant figures. I want to see the exact output. After the output, insert short-answer response to questions. For instance, some questions ask you to run your R code and then interpret the results. Please insert your interpretations *directly after the output*. After all that, include the text of `*.main.R` and `*.functions.R`.

- Some questions contain math-type questions, such as "derive..." or "show that..." questions. You can write your responses out and then scan it. *I would like you to insert your math in the file* `hw02.[last name].responses.pdf`.
- The file `hw02.[last name].functions.R` contains *all* your R functions in a single file. Unless otherwise instructed, the output of your function should be a `list` with *lower case* variable names. The variable names should be spelled *exactly* as in the assignment directions. Each variable in the list should be a scalar (i.e., no vectors, unless specifically instructed).
- The file `hw02.[last name].main.R` contains the R code that loads the data and calls the function. Please submit the entire code. This allows me to give partial credit for minor errors.

5 End of Semester Project

At the end of the semester, students present an original research project. The project must involve a statistical technique that was not covered in the course. The goal is to independently learn a new technique and to present it to the class, using the material in this course as a foundation. If you are working on a thesis, then I highly recommend choosing a topic that contributes directly to your thesis. I strongly encourage you to discuss your project idea with me before finalizing your proposal.

Project Proposal

- Deadline: Submit a proposal describing your chosen research question by April 26, 2024.
- Proposal Content: Your proposal should clearly outline the research question and the proposed method of investigation, with as much detail as possible.
- Length: no more than a paragraph
- Feedback: I will review each proposal and notify you of its appropriateness by April 29, 2024. This step is intended to ensure the feasibility and appropriateness of your chosen project.

Project Presentation

- Time: presentations will be given during the class exam day May 6, 2024.
- Oral Presentation: Present your findings in a 15-minute oral presentation (strictly enforced!).
- Please submit your presentation slides to me immediately after all presentations are concluded.

Presentations will be Evaluated based on the Following Criteria

- Student clearly explains the motivating question of the talk, without using jargon.
- Student clearly explains the statistical procedure that addresses the question.
- Student presents results in a logical, interesting sequence that the audience can follow.
- Student demonstrates full knowledge of the procedure by answering questions during class.
- Student concludes with a definite statement about the outcome of the project.
- Presentation (excluding questions) lasts no more than 15min (strictly enforced!).
- Student submits presentations slides immediately after the presentation period.