CLIM762: Statistical Methods in Climate Research, Spring 2025.

Catalog 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.

Course	CLIM 762: Statistical Methods for Climate Researh (3 credits)
Instructor	Timothy DelSole [web profile]
Contact	email: tdelsole@gmu.edu; office: Research Hall Rm112
Prerequisites	linear algebra, programming experience, STAT 344; or permission of instructor
Meeting Times	Mondays, 4:30-7:10pm
Location	Research Hall, Rm 121
Grading	Homework: 45%; weekly quizzes: 15%; Midterm: 20%; Final: 20%
Office Hours	By appointment (request made by email).
Software	Homeworks are expected to be done using the package **RStudio**
Required Texts	*Statistical Methods for Climate Scientists* by DelSole and Tippett. You do not need to purchase this book

Class Schedule

week	date	topic	reading
1	27-Jan	Probability concepts	chapter 1
2	03-Feb	Introduction to R and R Studio	Tutorial
3	10-Feb	Hypothesis Tests	chapter 2
4	17-Feb	Confidence Intervals	chapter 3
5	24-Feb	Hypothesis Tests based on Ranks	chapter 4
6	03-Mar	Stochastic Processes	chapter 5
7	10-Mar	SPRING BREAK	
8	17-Mar	MID-TERM EXAM	
9	24-Mar	Power Spectra	chapter 6
10	31-Mar	Multivariate Concepts	chapter 7
11	07-Apr	Linear Regression: Estimation	chapter 8
12	14-Apr	Linear Regression: Inference	chapter 9
13	21-Apr	Principal Component Analysis	chapter 12
14	28-Apr	Model Selection	chapter 10
15	05-May	Pitfalls of Statistical Inference	chapter 11
	12-May	FINAL EXAM	

Learning Outcomes

By the end of this course, students will be able to:

1. Apply core statistical methods such as hypothesis testing and regression analysis to analyze climate data and address research questions.
2. Interpret statistical quantities like confidence intervals, power spectra, principal components.
3. Understand best practices and recognize and avoid common pitfalls in statistical inference.
4. Understand the potential applications of machine learning in climate research.

Technical Requirements

- This course will be hosted on Canvas. Please ensure you are familiar with accessing and navigating this platform. Resources and support are available at: https://lms.gmu.edu/getting-started-students/ to help you get started. If you have any questions, do not hesitate to reach out to me or contact the ITS Support Center for assistance.
- Use RStudio on your laptop and use it to make computations with data and generate plots.

Course Format

The course is structured to provide both qualitative and quantitative understanding of physical climatology through a combination of readings, lectures, student presentations, and assignments.

- **Lectures:** The instructor will deliver lectures every Monday.
- **Reading Assignments:** Students are expected to complete one chapter of reading per week before the corresponding lecture. Although these readings are not graded, they are essential for providing a preview of the material to be discussed in class.
- **Homework Assignments:** Weekly homework assignments, distributed each Monday, are designed to reinforce concepts from the lectures and readings. These assignments are due the following Monday by the end of the day.
- **Quizzes:** Most Mondays, there will be a closed-book, 10-minute quiz covering the material from the previous week's lectures, readings, and homework. These quizzes are designed to ensure that students have understood and can apply the concepts covered in the assignments.

Policies

- **Attendance:** Regular attendance is expected. Please notify the instructor in advance if you need to miss a class.
- **Homework assignments:** No late assignments will be accepted. The lowest individual homework grade will be dropped before calculating the total homework score.
- **Quizzes:** There are no makeup quizzes. Lowest quiz score will be dropped.
- **Collaboration with others:** Permitted as long as you (1) try the problem yourself first, (2) write solutions/code yourself, (3) acknowledge all sources (textbooks, collaborators, ChatGPT).
- **Academic Integrity:** Students are expected to adhere to the university's academic integrity policy. Plagiarism and cheating will not be tolerated.
- **Honor System**: About 90% of the homeworks were given in previous years. Students are expected to refrain from consulting answers for previous years' homeworks.
- Common University Policies (Including Academic Integrity, Disability Accommodations, Diversity, Title IX, and Student Privacy) [clickable link]

Grades

Grading	
homework assignments	45%
weekly quizzes	15%
mid-term	20%
Final	20%

Grade Schema	
grade	**percentages**
A+	97.0 - 100
A	93.0 - 96.9
A-	90.0 - 92.9
B+	87.0 - 89.0
B	83.0 - 86.9
B-	80.0 - 82.9
C+	77.0 - 79.9
C	73.0 - 76.9
C-	70.0 - 72.9
D+	67.1 - 69.9
D-	60.0 - 67.0
F	less than 60.0

1 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. Quizzes has been adopted to encourage you to understand the material.
- 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 university's 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]

2 Coding Assignments Should be Done in RStudio

All homework assignments must be completed using RStudio. RStudio is an interface for a package called R. The decision to use RStudio is based on several factors, including the suitability of RStudio for our coursework and its widespread use in statistics.

- Downloading RStudio: RStudio is a free software, and you can download it from https://posit.co/download/rstudio-desktop/. Please download and install it on your computer the first week of the semester.
- Tutorial on RStudio: Canvas contains a tutorial on RStudio that covers everything you need to know for the first few 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, RStudio, 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 RStudio: R is the most commonly used software among statisticians and has a vast array of statistical packages. Its relevance extends beyond academia; for instance, a recent graduate student secured a job at a research center where R is exclusively used.

3 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`.

- 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.