Syllabus
EES 4891/5891: Probability and Statistics for Geosciences

Jonathan Gilligan
Vanderbilt University

Spring 2025

1 Nuts and Bolts

1.1 Class Meetings

TR 8:00–9:15 Stevenson 5722

1.2 Professor

Jonathan Gilligan (they/them/theirs)
Professor of Earth & Environmental Sciences
Professor of Civil & Environmental Engineering
Office: Stevenson 5735 (Stevenson #5, 7th floor)
jonathan.gilligan@vanderbilt.edu
Office Hours: Tue. 9:25–10:30, Wed. 1:30–3:30, or by appointment.

1.3 Email

If you want to communicate with me, be sure to begin the subject line of your email with "EES 4670" or "EES 5670". This helps assure that I will see your message quickly and respond to it.

I have set my email reader to flag all messages like this as important, so I will read them first. This also assures that I do not mistake your email for spam. I typically receive over 100 emails per day, so if you do not follow these instructions I may not notice your email.

2 Course Description

Probability and Statistics for Geosciences. Fundamentals of probability and statistics for the Earth & Enviromental Sciences, with applications in R. Probability distributions, descriptive statistics, statistical testing, regression analysis, elements of time-series analysis and multivariate statistics, principal components analysis, reproducible research methods, principles of statistical computing using R. [3](MNS)

2.1 Prerequisites

You should be familiar with differential calculus and linear algebra. I will review basic concepts, but the course will be difficult for you if you are completely unfamiliar with these areas of math.

This course will be mathematical and will make extensive use of the R software system, but **I do not assume** that you already know R or advanced mathematics beyond calculus and linear algebra.

3 Goals for the Course

My goals for this course are that at the end of the semester:

- You will understand theories of probability and be familiar with the properties of discrete and continuous probability distributions.
- You will understand what the Normal probability distribution is, and why it plays a central role in probability and statistics.
- You will be able to perform descriptive statistical analyses of data.
- You will be able to use statistical tests to test propositions about data, such as identifying differences between data collected from different sources or under different conditions.
- You will be able to estimate the parameters of probability distributions from observed data.
- You will be able to analyze time-series data.
- You will be able to analyze sets of multiple variables, such as measurements of multiple elements or isotopes from each of a number of samples.
- You will be skilled at organizing and managing your data analysis projects using Reproducible Research methods.

  Increasingly, journals and funding agencies require researchers to use reproducible research methods, so that other people can easily review and understand how the analysis was conducted, and so that if questions arise—even years later—it will be straightforward to go back and understand exactly how the analysis was conducted.

3.1 Reading Material

There are two required textbooks:

From time to time during the term, I will assign supplementary reading on the Internet or in handouts that I will post to the class web site.

TEXTBOOKS

- Julien Emile-Geay, *Data Analysis in the Earth & Environmental Sciences*, 5th Edition (2023) `https://doi.org/10.6084/m9.figshare.1014336`.

  This will be the main textbook for most of the semester. The book is published open-access and you can download the PDF for free from `https://doi.org/10.6084/m9.figshare.1014336`, Brightspace, or the course web site, at `https://ees5891.jgilligan.org/files/course_files/ProbStatEarthEnvironSci_2023.pdf`.

- Hadley Wickham, Mine Çetinkaya-Rundel, and Garrett Grolemund, *R for Data Science*, 2nd Edition, (O'Reilly, 2023; ISBN 978-1492097402) `https://r4ds.hadley.nz`.

  This book is the best practical introduction I have found for getting started with R and getting things done in data analysis. The author is the chief data scientist as the Posit

company and wrote a huge number of widely used free packages to extend and enrich R. This book follows his philosophy of how to organize data sensibly for analyzing and presenting it.

You can buy a paper copy, if you wish, but the full text is available for free online at `https://r4ds.hadley.nz`

COURSE WEB SITE

In addition to Brightspace, I have set up a companion web site for this course at `https://ees5891.jgilligan.org`, where I post the reading and homework assignments, my slides from class, and other useful material. That web site will be the central place to keep up with material for the course during the semester. This web site will direct you to Brightspace if there is anything you need to find there.

3.2 Computer Software

For this class, we will work in R, and I strongly recommend that you install the free version of RStudio Desktop for working with R. All the software we will use this semester is free and can be downloaded and installed on Windows, Mac, and Linux systems. You can find details at the "Tools" page of the course web site at `https://ees5891.jgilligan.org`.

We will also use the `git` revision control software as part of our Reproducible Research practice. You will use this to manage files for assignments and the semester research project. I will spend a class explaining why we use `git`, and how to use it effectively for your homework and other projects.

4 Assignments

4.1 Overview of reading assignments

I will post detailed reading assignments to the course website `https://ees5891.jgilligan.org` that give specific pages to read for each class and notes on important things you should understand. **I expect you to complete the reading before you come to class on the day for which the reading is assigned**, so you can participate in discussions of the assigned material and ask questions if there are things you don't understand.

4.2 Graded Work

HOMEWORK

Homework assignments will be posted on the course web site, and must be submitted by the beginning of class on the day it's due.

You will turn in homework to Brightspace or GITHUB as indicated on the assignment.

PROJECT

In the second half of the semester, you will do a research project, in which you will choose a data set that's interesting to you and apply statistical methods to analyze it. You will present

the results of your project in class during the last week of the semester and turn in a written report about your project.

You may examine data from a research project you're working on, or it can be data from a public data source that you are interested in understanding better.

TESTS AND EXAMINATIONS

There will not be any tests or examinations in this course. Your grade will be based on class participation, homework, modeling projects, and in-class presentations.

4.3 Basis for Grading

Class participation	5%
Homework	45%
Research Project	50%

5 Final Note

This is the first time I have taught this course, and during the term, I will assess how things are going, and may change the assignments and sequence of readings to help you get the most out of it.

6 Meet Your Professor

Jonathan Gilligan has worked in many areas of science and public policy. Their past research includes work on laser physics, quantum optics, laser surgery, electrical properties of the heart, using modified spy planes to study the ozone layer in the stratosphere, and connections between religion and care for the environment. They are a Professor of Earth & Environmental Sciences, Professor of Civil & Environmental Engineering, and the director of the Vanderbilt Climate and Society Grand Challenge Initiative and serves on the steering committee for the Vanderbilt Center for Sustainability, Energy, and Climate (VSEC). They serve on the technical advisory board of the Southwestern Urban Corridor Integrated Field Laboratory, which is helping cities in the Southwestern U.S. plan for climate change, and on the steering committee for the Community Surface Dynamics Modeling System, a large interdisciplinary modeling center that studies environmental changes to landscapes and coasts, and they are a Voices of Science fellow of the American Geophysical Union.

Professor Gilligan's current research investigates the impacts of climate change on socity; the impacts of drinking water contamination on people's health, and developing new directions for climate policy in the US.

In 2017, Professor Gilligan and Professor Michael Vandenbergh shared the Morrison Prize for the highest-impact paper on sustainability law and policy published in the previous year. Professors Gilligan and Vandenbergh have developed this work into a book, *Beyond Politics: The Private Governance Approach to Climate Change* (Cambridge University Press, 2017).

In addition to their academic work, Professor Gilligan dabbles in writing for the theater. Their stage adaptation of Nathaniel Hawthorne's *The Scarlet Letter*, co-written with their mother Carol Gilligan, has been staged at The Culture Project in New York City, starring Marisa Tomei, Ron Cephas Jones, and Bobby Cannavale, and was later performed at Prime Stage Theatre, Pittsburgh and in a touring production by The National Players. Prof. Gilligan and Carol Gilligan also wrote the libretto for an opera, *Pearl*, in collaboration composer Amy Scurria, and producer/conductor Sara Jobin, which was performed at Shakespeare & Company in Lenox MA, starring Maureen O'Flynn, John Bellemer, Marnie Breckenridge, John Cheek, and Michael Corvino, and in Shanghai China, starring Li Xin, Wang Yang, John Bellemer, and Lin Shu.

Schedule of Classes (Subject to Change)

IMPORTANT NOTE: This schedule gives a rough indication of the reading for each day. See the assignment sheets posted on Brightspace and the course web site for the detailed daily assignments.

Date	Topic	Reading
Tue., Jan. 7	Introduction	No reading
Thu., Jan. 9	What is Probability?	*Emile-Geay* Ch. 1, *Emile-Geay* Ch. 2 pp. 15–22
Tue., Jan. 14	The R Statistical Programming Language	*Wickham* Intro & Ch. 1
Thu., Jan. 16	Mathematics of Probability	*Emile-Geay* Ch. 2 pp. 22–30
Tue., Jan. 21	Reproducible Research in R	No reading
Thu., Jan. 23	Probability Distributions	*Emile-Geay* Ch. 3
Tue., Jan. 28	Transforming Data in R	*Wickham* Ch. 2–3
Thu., Jan. 30	Working with Data in R	No reading
Tue., Feb. 4	Importing Data into R	*Wickham* Ch. 7–8
Thu., Feb. 6	The Normal Distribution	*Emile-Geay* Ch. 4 focus on pp. 51–59
Tue., Feb. 11	Error Theory	*Emile-Geay* Ch. 4 focus on pp, 60–62
Thu., Feb. 13	Exercises with Normality	No reading
Tue., Feb. 18	Statistical Estimation	*Emile-Geay* Ch. 5 pp. 63–76, *Emile-Geay* Ch. 5 pp. 76–80
Thu., Feb. 20	Exercises with Estimation	No reading
Tue., Feb. 25	Confirmatory Data Analysis	*Emile-Geay* Ch. 6 sections I–IV
Thu., Feb. 27	Statistical Testing	*Emile-Geay* Ch. 6 sections V–IX
Tue., Mar. 4	Review	No reading
Thu., Mar. 6	Exercises with Hypothesis Testing	No reading
Tue., Mar. 11 Thu., Mar. 13	**SPRING BREAK**	

Date	Topic	Reading
Tue., Mar. 18	Multivariate Analysis	*Emile-Geay* Ch. 11
Thu., Mar. 20	Least Squares Analysis	*Emile-Geay* Ch. 13
Tue., Mar. 25	Linear Regression	No reading
Thu., Mar. 27	Regression with Discrete vs. Continuous Variables	No reading
Tue., Apr. 1	Principal Components Analysis	No reading
Thu., Apr. 3	Exercises with Regression	No reading
Tue., Apr. 8	Time Series Data: Fourier Analysis	*Emile-Geay* Ch. 7
Thu., Apr. 10	Modeling Time Series	*Emile-Geay* Ch. 8
Tue., Apr. 15	Exercises in Time Series Analysis	No reading
Thu., Apr. 17	Review of the Semester	No reading