Data Analysis in the Environmental Sciences

Course website: https://canvas.ucsc.edu/courses/23091

Instructor: Claudie Beaulieu, E&MS A443, clbeauli@ucsc.edu

Lectures: TTh 13.30-15.05, E&MS D236

Discussion: W 16.00-18.00, Crown Mac Lab room 201

Office Hours: TTh 15.05-16.05, other times are also possible by appointment.

Accommodations: If you have documented/undocumented disabilities, please register with the Disability Resource Center (https://drc.ucsc.edu/index.html) and inform me about specific accommodation needs at the start of the course.

Course description: The main aim of this course is to introduce data analysis methods regularly encountered within environmental sciences. You will learn how to think about data, its uncertainty, how models and data are related and depend on underlying assumptions, and how to synthesize information contained in data. The class will focus on two main areas of study: 1) environmental sampling and risk analysis and 2) climate and environmental change detection. **Prerequisite(s):** MATH 11B or 19B or equivalent.

Learning outcomes:

1. Summarize, synthesize and visualize an environmental dataset
2. Describe and quantify data and sampling uncertainties
3. Utilize statistical distributions to assess environmental risk
4. Apply statistical tests to detect and quantify climate and environmental change
5. Express and verify underlying assumptions in statistical tests and models

Problem sets: 7 problem sets due on Tuesdays, altogether count towards 35% of the final grade.

Take-home exams: There will be a midterm and a final take-home exam due on 11/05/2019 (Midterm) and 12/11/2019 (Final). The take-home exams will count towards 25% and 40% of the final grade, respectively.

Textbook: There is no required textbook for this class. However, if you would like to do some additional reading, I would recommend "*Statistical Methods for the Atmospheric Sciences*" by Wilks. The textbook is in reserve at the library.

Course slides: I will post slides on Canvas one day before each lecture. Since my lectures typically contain problems to solve in class, each lecture will be updated afterwards with the solutions.

Schedule of lectures

Topic 1: Environmental sampling and risk

Week 1:
L1 (September 26th): Intro to environmental data analysis

Week 2:
L2 (October 1st): Intro to probabilities
L3 (October 3rd): Making sense of a new data set

Week 3:
L4 (October 8th): Probability distributions
L5 (October 10th): Probability distributions continued

Week 4:
L6 (October 15th): Normal distribution & the central limit theorem
L7 (October 17th): Statistical expectations and variances

Week 5:
L8 (October 22nd): Sampling the environment
L9 (October 24th): Frequency analysis

Topic 2: Climate and environmental change detection

Week 6:
L10 (October 29th): Revision for the midterm
L11 (October 31st): Intro to hypothesis testing

Week 7:
L12 (November 5th): Tests on mean & variance differences
L13 (November 7th): Correlation

Week 8:
L14 (November 12th): How to lie with statistics
L15 (November 14th): Simple linear regression

Week 9:
L16 (November 19th): Trend detection
L17 (November 21st): Multiple linear regression

Week 10:
L18 (November 26th): Multiple linear regression continued
Note: November 28th is thanksgiving – no class

Week 11:
L19 (December 3rd): Revision for the final
L20 (December 5th): Q&A for the final

Schedule of problem sets & exams:

	To be published on Canvas	Due
PS1	October 1st	October 8th
PS2	October 8th	October 15th
PS3	October 15th	October 22nd
PS4	October 22nd	October 29th
Midterm exam	October 29th	November 5th
PS5	November 5th	November 12th
PS6	November 12th	November 19th
PS7	November 19th	November 26th
Final exam	December 3rd	December 11th

Programming language: We will use R to analyse data in this class. R is a free software environment for statistical computing and graphics. It compiles and run on UNIX platforms, Windows and Macs. You can download it here: http://www.r-project.org/. I recommend also downloading RStudio here: https://www.rstudio.com/products/rstudio/download/, which is a more user-friendly interface of R including a console, syntax-highlighting editor that supports direct code execution, and a variety of robust tools for plotting, viewing history, debugging and managing your workspace. If you have no previous experience with a programming language, this might sound intimidating. Please note that you are not expected to have any previous experience with programming in R at the beginning of this course, we will take it step-by-step. You can think about R as an additional tool in your toolbox! R is available on all Mac Labs on campus, as well as McHenry and Science &Engineering library public Macs. See ITS website to find out where Mac Labs are: https://its.ucsc.edu/computer-labs/index.html

Problem sets: There will be seven weekly assignments that together count for 35% of your final grade. The assignments will be released on Canvas every Tuesdays, we will discuss them in class on each Tuesday afternoon lecture, and Wednesday discussion sections are meant to provide help with the problem sets, that are due on the following Tuesday. Each problem set typically contains some data analysis calculations and interpretation, and feedback. Copy of the R code generated to answer questions (where applicable) will have to be submitted as part of the problem set.

Exams: The two exams in this class are take-home. They will be given to you one week before their due date. They will consist in data analysis problems to be done in R, and your R code will have to be submitted as part of the exam. The midterm exam counts for 25% and the final exam counts for 40% of your final grade.

Group work: I strongly encourage you to work in groups in this class, as students can achieve deeper learning through talking with other students. Actually, not only you are encouraged, but you will earn points on the problem sets for doing so. While group work is strongly encouraged, you must turn in your own work using your own words and your own code. Copying someone else's words/code is not allowed and can lead to serious consequences. Please see the following for a description of the UCSC Academic Integrity policy: https://registrar.ucsc.edu/navigator/section1/academicintegrity. I trust you to be honest and turn in your own work that reflects your own understanding.

Grading scale:
As: 100-85%
Bs: 85-70%
Cs: 70-60%
Ds: 60-50%
F: <50%
You are not competing for grades with other students. It is possible for everyone taking this class to get an A. If the above scale results in too few As, Bs, and Cs, then a curve will be used to assign at least 10% As, 20% Bs, and 30% Cs.

Late Work: Problem sets are due as listed in the table above and on each assignment, generally a week after it is handed out. Late problem sets will be eligible for 80% of original points after one day, 60% after two days, and so on, after 5 days passed the due date a problem set is not eligible for points anymore. Missed take-home exam deadlines are not eligible for points. However, medical emergencies and traumatic events may happen. If you contact me within 24h and provide documentation, I will be happy to give you an extension.

Tips for success:
1- Showing up: class and discussion sections attendance and participation
2- Planning ahead: Start the weekly problem set BEFORE discussion section, so that you can get help where needed early on
3- Working in groups: I guarantee you that you will achieve deeper learning by discussing with your peers (and you will earn points)
4- Providing feedback: I will ask you for feedback weekly at the end of each problem sets. If some concepts were less well understood, I will make adjustments to lectures to revisit these concepts.
5- Asking questions: I encourage you to ask questions as soon as you get confused. Because each week's material is built on the previous ones, it is important not to drag concepts less well understood.
6- Practicing: I appreciate that learning a programming language can be frustrating. There is no better way to learn than by doing it.

Withdrawal: By November 6th. Other key dates for registration and enrolment can be found here: https://registrar.ucsc.edu/soc/key-dates-enrollment.html

Disclaimer: This syllabus is intended to provide guidance on topics to be covered in the class, and I will follow as closely as possible. However, I reserve the right to modify, supplement and make changes as needed if I judge it will benefit your learning.