STA 538: Statistical Programming Using R

Instructor:
Professor Andrew (Drew) Crossett, Ph.D.
Office: 25 University Ave. 111
Phone: x2613
email: acrosset@wcupa.edu

Office Hours:
Monday 4:30-6:30pm
Tuesday 5:30-6:30pm
Wednesday 5:30-6:30pm
Thursday 5:30-6:30pm

Course Web Page:
This course has a D2L website. I will be using it quite heavily
so please make sure to check it often!

Lectures:
Section 1: MoTuWeTh 6:30-9:00pm, 25 University Ave. 103

Text:
An Introduction to R (recommended),
Venables, William N.

A Beginner's Guide to R (recommended),
Zuur, Alain and Ieno, Elena and Meesters, Erik

A First Course in Statistical Programming with R (recommended),
Braun, John W.

OVERVIEW & COURSE DESCRIPTION

In recent years, data sets have grown in size and complexity to the point where it is nearly impossible to do modern statistics without a computer. This problem has led to the advent of many statistical languages and software packages - each with advantages and disadvantages. One of the most widely used systems is called R, which was developed by the Comprehensive R Archive Network (CRAN). There are many reasons why this statistical computing language has become so popular. Most notably, R is freely available for common operating systems, such as Windows, MacOS X and Linux. Also, since R is a programming language, it is an extremely powerful tool in the sense that it is very customizable. As such, it is growing more and more necessary to be able to read, write and interpret R code in many statistical applications.

After learning the basics, this course will focus on performing many rigorous statistical analyses and simulating data in R.

Student Learning Outcomes for the Masters of Science in Applied Statistics:

1. Demonstrate an understanding of probability and statistical inference, including the fundamental laws of classical probability, discrete and continuous random variables, expectation theory, maximum likelihood methods, hypothesis testing, power, and bivariate and multivariate distribution theory.

2. Demonstrated the ability to apply the elementary methods of statistical analysis, namely those based on classical linear models, categorical methods, and non-parametric ideas to perform data analysis for the purposes of statistical inference.

3. Demonstrate proficiency in the effective use of computers for research data management and for analysis of data with standard statistical software packages, particularly SAS.

4. Learn to develop and critically assess design of experimental studies and the collection of data.

5. Apply one or more methods of statistical inference to a particular area of interest, particularly the program in the elective concentration.

6. Gain practical experience in statistical consulting and communicating with non- statisticians, culminating with interaction with research workers at a local company as part of the internship practicum.

OBJECTIVES

1. To learn basic programming skills in R. Including, but not limited to defining variables, for loops, if statements, function construction and using internal packages [SLO 3].

2. To develop methods for simulating data in a variety of situations [SLO 5].

3. To learn how to read/write data files and plot figures used for simulation studies [SLO 4].

4. To develop skills for analyzing and interpreting R output in applications of statistical methods to problems in many fields of science [SLO 6].

LECTURES

Class time will consist of a combination of lecture, discussion, question and answer, and lab time. Depending on the topic, I will typically start class with a lecture on the theory behind the methods we are discussing. Following this will be some time for us to discuss what was just talked about. After that, I will display R code to the entire class to get everyone started. Finally, the class will be broken up into 2-3 students to work together on in-class projects. I will walk around to answer questions. There will be a couple of 15-20 minute breaks throughout the class.

CLASS CANCELLATION POLICY

In the event that I am unable to meet a class, I will: (a) notify you in person at a prior meeting to that effect or (b) an official class cancellation notification on the *stationary of the Department of Mathematics and signed and date stamped by the Department's Secretary Barbara Maleno* will be posted on the classroom door. All other postings announcing the cancellation of this class are to be considered unofficial and are to be ignored.

HOMEWORK

- There will be regular homework assignments for this class. Homework will be assigned at the end of each class and will (usually) be **due the following week**. Homework are due in class **BEFORE** class starts. **No late homework will be accepted unless it's previously cleared by me**. Homework and their solutions will be available on the course website. Please submit neat, *stapled* paper copies or *well organized* electronic versions of your homework.

- You **MUST** send me an electronic copy of your R code for each assignment. You may send it separate or at the end of the typed up solutions. If the computer code does not run or I do not get the same results that you get, you will receive *at most* 50% credit for that problem. It is in your best interest to comment your code as much as possible. This will remind you what you did when you are studying and also allows me to give you partial credit.

- Statistics is very conducive to collaboration. Therefore, unless otherwise instructed, I encourage you to discuss assignments with other students. The best way to work with others on homework is to do as much as you can on your own, and then get together. If you've done the problems independently, it's more likely that one of you will have the right answer. If you do them together from the outset, you might both make the same mistakes.

- **Academic Honesty:** While I encourage you to work together, **the written (or typed) solutions to homework problems must be your own and not copied from someone else**. If the distinction between cooperation and plagiarism seems unclear, please ask me in class or privately and I'll be happy to explain it at greater length. Also, "for questions regarding Academic Dishonesty, the No-Grade Policy, Sexual Harassment, or the Student Code of Conduct, students are encouraged to refer to their major department's handbook, the Undergraduate Course Catalogue, the Ram's Eye View, or the University Web Site." Improper conduct in any of these areas may result in a grade of zero points, a failing grade in the course and/or expulsion from the class.

EXAMS

- There will be TWO IN-CLASS EXAMINATIONS and a TAKE-HOME FINAL. The first exam is *tentatively* scheduled for WEDNESDAY, JANUARY 4TH. No books or notes will be allowed for this exam. The second in-class exam is *tentatively* scheduled for THURSDAY, JANUARY 12TH. This will be an open-notes exam. The take-home final exam will be cumulative and will be due by 8:30pm on FRIDAY, JANUARY 20TH.

• **No makeup examinations** will be given. A student who misses an examination because of a medical reason must provide **documented** evidence of serious medical incapacitation. Other reasons for missing an examination must be discussed with me **before** the day of the examination. Each case will be considered on an individual basis. The overall course grade for a student who misses an examination with a valid reason will be based on that student's remaining course work. A student who misses an examination without a valid excuse will receive a grade of zero on that examination. Sleeping through an exam is not a valid excuse and will result in a grade of zero.

COURSE GRADE

Your course grade consists of:

Exam 1	20%
Exam 2	20%
Homework	30%
Final	30%

ATTENDANCE POLICY:

I will not be taking attendance. However, I highly recommend that you attend class. Even though I will be posting some notes on the website, there will be a lot of supplemental material not in the notes. If you miss a class, your best bet to learn the material is to come to my office hours, meet with someone from class or, in extreme cases, contact me to meet at another time.

COMPUTING

Part of this course will take place in a computer lab so laptops are not necessary. However, if you feel more comfortable working from your own computer then you are more than welcome to use it. If you intend on using your own computer then you will need to download R. It is freely available at `http://cran.r-project.org/`. Also, you may need to download a data set before coming to class. If this is the case then I will have it posted to the course website the night before class. Make sure to check the website frequently!

OFFICE HOURS

I will hold office hours each week. I am also frequently available by appointment. Please make sure to provide ample advance notice for these appointments.

ADA

• If you have a disability that requires accommodations under the Americans with Disabilities Act (ADA), please present your letter of accommodations and meet with me as soon as possible so that I can support your success in an informed manner. Accommodations cannot be granted retroactively. If you would like to know more about West Chester University's Services for Students with Disabilities (OSSD), please contact the OSSD which is located 223 Lawrence Center. The OSSD hours of operation are Monday-Friday 8:30 a.m. - 4:30 p.m. Their phone number is 610-436-2564, their fax number is 610-426-2600, their email address is ossd@wcupa.edu, and their website is www.wcupa.edu/ussss/ossd.

EMAIL POLICY

• It is expected that faculty, staff, and students activate and maintain regular access to University provided e-mail accounts. Official university communications, including those from your instructor, will be sent through your university e-mail account. You are responsible for accessing that mail to be sure to obtain official University communications. Failure to access will not exempt individuals from the responsibilities associated with this course.

CAMPUS EMERGENCIES

• For campus emergencies call WCU Department of Public Safety at (610)436-3311.

TENTATIVE WEEKLY SCHEDULE

1. **December 19th:** Basic Programming Statements, Plotting and Introduction to R [CO 1]

2. **January 3rd:** Advanced Plotting and Programming (For Loops, Writing Functions, Missing Data, Tables) [CO 1]

3. **January 9th:** Distributions and Statistical Analyses (ANOVA, Regression, Categorical, Nonparametric Methods) [CO 2]

4. **January 16th:** Simulations, Markov Chains, Monte Carlo Methods [CO 3]

I am looking forward to teaching this class as I hope you are looking forward to taking it. Please do not hesitate to contact me if you have any questions or concerns whatsoever.

Welcome, and Good Luck!