Math 2820L: Statistics Laboratory

**Instructor:** Jonathan Ashbrock

**Email:** jonathan.ashbrock@vanderbilt.edu

**Location:** Wilson Hall, Room 120

**Meeting Times:** Wednesday, 3:10-4 pm

**Office Hours:** Monday, Friday 11:00-12pm in SC 1227

**Course Description:**

Applications of the theory developed in 2820. Emphasis on data analysis and interpretation. Topics include the one- and two-sample problems, paired data, correlation and regression, chi-square, and model building. Pre- or corequisite: 2810 or 2820.

**Topics Covered:** This semester will focus primarily on the following topics:

* Programming in **R**
* Interpreting and visualizing data
* Modeling data using regression techniques
* Common Data Analysis Algorithms
* Hypothesis Testing

The main focus of the course is the first three topics, and appropriate time will be devoted to each. However, given extra time I think it would be of great interest to the students to see some algorithms that are used in answering more general questions that an applied mathematician may be asked to answer in industry. As the last topic will be determined based on class time remaining, the topic is undecided as of yet. However, possible topics include clustering algorithms, principal component analysis, random forests, or support vector machines.

**Grading Policy:**  The class grade will be based upon homework assignments and a final project. The grading will be broken down as such:

* Homework - **70%**
* Final Project - **30%**

**Homework:** Homework assignments will typically either be to turn in **R** code for a project (with sufficiently many comments!) or a paper detailing the results obtained from your **R** code. Unless otherwise stated, each person is expected to turn in their own homework assignment but students are allowed to collaborate with others on IDEAS ONLY. This means you can discuss in general how to write a program but you cannot share code with one another. Any copied or shared code will be given a grade of 0 for the assignment.

Homework will be due on the date given at the beginning of the class period. Any homework not turned in at this time will lose 25% and will lose an additional 25% at 12:00 midnight each following day.

**Final Project:** The final project will consist of an analysis of data which uses all the techniques learned in this lab. It is likely that the final project will be done in groups and there will be both a paper and a presentation aspect to the final project. The presentations will take place during our exam time and perhaps the last class beforehand. More details will follow.

**Computer Policy:** Students are expected to bring their computer to each and every class. The course is designed to teach students to program in **R** and you cannot write a computer program without a computer. The first class will be devoted to helping the students get **R** installed on their laptop computer.

**Attendance Policy:** Attendance will be taken each class. A student who attends each class period (except for absences deemed excusable by myself) will have their lowest homework grade replaced by their highest.

In addition to replacing a homework assignment, attendance has the following benefit. The work done in class will be directly related to the homework assignments and so the homework assignments can be begun in class time. My experience with any coding project is that BY FAR the most difficult part is getting started. Once the project is begun and the relevant coding techniques are understood, the project is much simpler.