

# **STAT 325/625: Statistical Case Studies**

**Syllabus, Fall 2016 (draft 8/29/2016)**

**John W. Emerson (Jay) and Xiaofei Wang (Susan)**  
**Yale University**

## **1 Practical Information**

- From Jay: Although my name is John, I would prefer that you call me Jay. If that makes you uncomfortable, Professor Emerson is just fine.
- From Susan: Professor Wang or Susan, whatever you decide you are comfortable with.
- E-mail is the best way to reach us: john.emerson@yale.edu, xiaofei.wang@yale.edu.
- Offices: Jay is in 24 Hillhouse Room B06, Susan is in Room 206.
- Office hours: We're generally around – feel free to drop by. If we're busy, we'll work something out.
- MEETING TIME/PLACE: Tuesday/Thursday 9:00-10:15. Hopefully (subject to confirmation), STAT 325 will meet primarily in 17 Hillhouse Room 111; STAT 625 will meet in the TEAL classroom of 17 Hillhouse (room 101). Occasionally, the classes will merge.
- Plus 1-hour-to-be-arranged (1 HTBA, see below).

## **2 Who may take this course?**

This course may be taken only with permission. The course is limited enrollment with the following priorities, although this process should be easier this year given the separation of 325 and 625:

- 2<sup>nd</sup>-year PhD students in Statistics
- 1<sup>st</sup>-year PhD and MA students in Statistics
- Undergraduate majors in Statistics and Applied Mathematics\*, in order of seniority if it comes down to a difficult choice.
- Others (undergraduate and graduate students from other programs)\*

Here's the \*: prior coursework in probability, statistics, and data analysis is required for undergraduates and the broad "other" category. It is not an alternative to STAT 230/530 (offered both this and next semester for the first time) or STAT 361/661 (next semester only). Undergraduate majors and PhD students should have STAT 361/661 or the equivalent data analysis experience (possibly STAT 230/530 if supported by other coursework) prior to STAT 325/625.

### 3 Objectives, Prerequisites, and Philosophy

Statistical Case Studies involves the statistical analysis of a variety of problems which vary each time the course is taught. Examples? In previous years we have studied nationalistic biases in the judging of Olympic diving; a job hiring discrimination lawsuit; the prediction of college basketball games; a study of global warming; and local real estate prices and patterns. We will emphasize methods of choosing data, acquiring data, and assessing data quality.

Graduate, professional, and undergraduate students from any department may be welcome if space is available (see the previous section), but must seek permission (discussing their background in statistics and goals for the semester) at or before the first class meeting. This is not a class to be shopped casually during the second week of the semester. Our expectation is that you are self-motivated and will push yourself. The most important prerequisite is a willingness to get your hands dirty working with real data sets. This course will entail a substantial amount of programming, which is best learned through examples, trial and error.

We will assume a basic familiarity with the **R** statistical programming environment. If you haven't used **R** before, please talk with us before the course starts so the expectations are clear.

### 4 Computing

We expect all of you will use your own laptops and will bring them to class every day. The ability to connect to a projector or display via HDMI or VGA is required; if you don't have these outputs (some Macs, for example, do not), please buy the appropriate adapter and bring it with you every day.

This is a computationally intensive course, and you will be required to experiment beyond the examples given in class. Unedited computer output will not be accepted; you must carefully document and explain your work. Much of what you will learn is "computer programming." Examples in class should serve as a basis for any required analysis, but you will be expected to use other sources for help: online help, web searches, and other references as the need arises.

In terms of preparation, the most important thing is the installation of the newest versions of **R** and **RStudio IDE**. Although **RStudio** isn't required, we strongly recommend using it. You will also want  $\text{\LaTeX}$  ( $\text{\MacTeX}$  on the Mac and  $\text{\MiKTeX}$  on Windows). For Windows, make sure you get the complete distribution of  $\text{\MiKTeX}$  – the basic version will not work!

## 5 Grades and homework

The homework (which can be pretty open-ended and may sometimes be done in small groups), presentations and participation (both in class and in the extra meetings), and the final individual case study constitute the entire grade for this class. There are no exams, and this is not a standard lecture course. Correctness of the submitted solution (in the traditional sense) is not necessarily more important than the preparation and process leading to the solution. A fair bit of emphasis will be placed on writing short reports and presentations. Students will be expected to take turns giving short presentations or demonstrations of their analyses without prior warning during any class. There will be one evening-long required exercise/challenge/adventure, tentatively on Sunday, November 13, 6 PM onwards, with food & beverages provided.

We tentatively plan on giving 50% Honors (H), 50% High Pass (HP) to graduate students. The undergraduate grade of A will not be automatic, but if you're taking this course we assume that you are motivated, hardworking, and likely not at risk of walking away with a B. The work required week-to-week may be substantial. But the class will not dribble into exam period – you'll be done during the last week of classes (unless you are required to take the Ph.D. Practical Exam).

## 6 Shopping this class and homework

All homework assignments (including the first one) are required of everyone, submitted on time. This will not be an easy course to slip into after the first class or two. Auditing is not permitted.

## 7 The schedule, and 1 HTBA?

The class will usually (but not always) meet at the scheduled time (9-10:15, Tuesday/Thursday). We will announce unusual dates (i.e. no class but lengthy case study challenges) in advance. For example, there will probably be no class on Tuesday, October 18.

Once again, we will be requiring – of all students – an extra small-group or individual session approximately once every other week. This will allow us to give more detailed feedback and guidance, and to better assess your progress. The times should offer enough flexibility to accommodate your schedules, but will tentatively be Tuesday & Thursday immediately after class, and Friday mornings. We will discuss this in more detail in the first class and will conduct a Doodle poll.

A tentative plan follows. For STAT 325, this 1 HTBA would likely be on Wednesday, although Tuesday 4-5 PM might be a good time slot for Susan. For STAT 625, Jay tentatively plans on Tuesday/Thursday 10:30-11:30, and Friday morning 9-10 AM (though other times Friday morning might also work).

## 8 Academic Integrity

Academic integrity is a core institutional value at Yale. It means, among other things, truth in presentation, diligence and precision in citing works and ideas we have used, and acknowledging our collaborations with others. In view of our commitment to maintaining the highest standards of academic integrity, the Graduate School Code of Conduct specifically prohibits the following forms of behavior: cheating on examinations, problem sets and all other forms of assessment; falsification and/or fabrication of data; plagiarism, that is, the failure in a dissertation, essay or other written exercise to acknowledge ideas, research, or language taken from others; and multiple submission of the same work without obtaining explicit written permission from both instructors before the material is submitted.

The statement above was provided by the Yale Graduate School. In this class, we have the same expectations of undergraduate students. We encourage students to work together on most everything unless noted otherwise. But “working together” is difficult to define when code (programming) is involved. Many students can benefit from constructive collaborations. However, if you benefit from “working together” then you had better be able to explain and discuss your solution should questions arise. “I don’t remember how I did this” would leave you on thin ice. Enough said.

Specific requirement: any collaboration on homework must be acknowledged up front. Something like “Lastname, Firstname – Homework 1. Worked with John S. and Jane D.” would be fine with us. This is very simple and encourages constructive collaboration if it aids the learning process. But not acknowledging such a collaboration is a considered a violation of academic integrity which we are obligated to report to Yale College or the Graduate School rather than arbitrate ourselves.