**Stat 322/522: Study Design**
**Project 1**

For this project assignment, you design, collect, and analyze data from a survey that you run. You work in pairs on the project, so that you can split the effort from data collection. Each member of a team is from the same course, e.g., students in STA 322 should work with other students in STA 322. If you cannot find someone else to work with, contact Prof. Reiter as soon as possible.

Your work should comprise a written report that describes the survey design and analysis. Please use R Markdown or some other word processor. Turn in one report per team, with all team members' names on the front page of the report. Include your code at the end of the report. Additionally, please upload to the Sakai dropbox a csv file that includes the data that you used for the report. Reports that do not include the code or data will receive a lower grade. Please refer to the course website for the Duke Community Standard requirements for the project.

**Due Date:** Turn in your project by noon on March 8. You can turn in the report at the beginning of any class up to March 8, or upload the report to the Sakai Dropbox by the due date. If you do the latter, please send me an email indicating whose Dropbox it is in.

**Grading:** Maximum of 10 points. Grade will be based on appropriateness of survey design and analysis for the problem, clear justification of your design and analysis choices, and clear explanation of results.

**Topic of the Survey**: How much should Duke students who take courses in natural science departments expect to spend on required textbooks if they buy them new at the Duke textbook store? You will run a survey to answer this question. As a sampling frame, we will use the online site of the Duke textbook store, `https://eposweb-320.sequoiars.com/ePOS?form=shared3/gm/main.html\&this_category=17\&store=320\&design=duke_textbooks`. At that site, select the choice "New & Used Click Here." Then click on the "Spring 2019" button. This should pop up the names of departments. If you click on a department, you will see a list of courses for that department. When you select a course, you can see what books are required in the course. Filter on "required" books by selecting the appropriate option once you have selected the courses.

Answer the following questions in your report, using point estimates and 95% confidence intervals.

1. What is the total cost of required books in the natural sciences at Duke?
2. What is the average number of required books per course in the natural sciences at Duke?
3. What is the average of the (total) cost of required books per course in the natural sciences at Duke?

When specifying the design, consider ways to make the estimates as accurate as possible, e.g., stratified sampling, and the data collection as convenient as possible, e.g., cluster sampling. You don't get extra points for being overly complicated in the design. Rather, use what we have learned to help you collect data with an efficient probability sample that allows you to get unbiased estimates. You should use a probability sample, even if you could figure out how to scrape all the data to get the actual population quantities. I leave sample size up to you depending on your resources (in this case, your time) and desire for accuracy, but I suggest sampling at least 50 courses. It's not too hard to get the data once you have the sample, so you should be able to take a decent sample size without spending too much time on data collection.

The natural science departments at Duke are listed at `https://trinity.duke.edu/natural-sciences`. To determine what qualifies as courses, include only courses numbered 699 and lower. Exclude course numbers for independent studies. If a course has multiple lecture sections, count only the first one that appears on the textbook store website (e.g., STA101.001 counts but STA 101.002 does not). Count only the lectures in courses with labs.