This problem set covers material from Week 1, dates 9/08 – 9/11.

Instructions: Write or type complete solutions to the following problems and submit answers to the corresponding Gradescope assignment. Your solutions should be neatly-written, show all work and computations, include figures or graphs where appropriate, and include some written explanation of your method or process (enough that I can understand your reasoning without having to guess or make assumptions). A general rubric for homework problems appears on the final page of this assignment.

Monday 9/08

None!

Wednesday 9/10

1. We have a description of three research questions (a)-(c). For each, identify the explanatory variable and the response variable. Additionally, for each research question, write 1-2 sentences about how you might obtain a sample to answer the research question. You must use each of simple random, stratified, and cluster sampling once. Be sure to clearly identify your strata/clusters!

   (a) Researchers want to determine if the number of hours per week spent on social media influences anxiety levels among college students.

   (b) The College wants to evaluate whether the First Year Seminar (FYS) improve students' ability to write.

   (c) The College wants to learn if athletes or non-athletes use the fitness center more often during the week.

2. We would like to obtain a sample of Middlebury students in order to learn how many hours of class the average student has per week. The registrar keeps a list of all undergraduates in alphabetical order. Suppose there are 2,770 undergraduates at Middlebury College this semester. To obtain the sample, someone proposes that we choose a number at random from 1 to 50, count that far down the list. We'll take that name and every 50th name after it to create our sample.

   (a) Is this a probability sampling method? Why or why not?

   (b) Is this sampling method the same as simple random sampling? Why or why not?

   (c) For this research question, do you believe there is selection bias when using the proposed sampling method? Why or why not?

   (d) Suppose the list of all students is not in alphabetical order, but is first sorted by major and then within major, sorted in alphabetical order. Do you think this leads to selection bias? Why or why not?

Thursday 9/11

3. The University of Vermont is interested in understanding whether students who live on campus have higher GPAs than those who commute. The school uses a simple random sample of 500 students and observes their GPA and housing status.

   (a) Can this study be used to establish causality? Why or why not?

   (b) Is there any potential for bias to enter into the study, even if the sample is randomly selected? If so, what kind? If not, why not?

   (c) What are some potential confounding variables in this scenario?

   (d) How might a stratified sampling scheme improve the study?

4. We are interested in learning if listening to classical music improves memory performance in college students.

   (a) Design an experiment that follows the three principles of experimental design to answer the research question. Be sure to clearly define the population of interest, the response variable, and the different treatments.

   (b) What is one potential confounding variable in this experiment? How is this addressed in your study design above? If it is not already addressed, how would you address it?

General rubric

Points	Criteria
5	The solution is correct *and* well-written. The author leaves no doubt as to why the solution is valid.
4.5	The solution is well-written, and is correct except for some minor arithmetic or calculation mistake.
4	The solution is technically correct, but author has omitted some key justification for why the solution is valid. Alternatively, the solution is well-written, but is missing a small, but essential component.
3	The solution is well-written, but either overlooks a significant component of the problem or makes a significant mistake. Alternatively, in a multi-part problem, a majority of the solutions are correct and well-written, but one part is missing or is significantly incorrect.
2	The solution is either correct but not adequately written, or it is adequately written but overlooks a significant component of the problem or makes a significant mistake.
1	The solution is rudimentary, but contains some relevant ideas. Alternatively, the solution briefly indicates the correct answer, but provides no further justification.
0	Either the solution is missing entirely, or the author makes no non-trivial progress toward a solution (i.e. just writes the statement of the problem and/or restates given information).
Notes:	For problems with multiple parts, the score represents a holistic review of the entire problem. Additionally, half-points may be used if the solution falls between two point values above.
Notes:	For problems with code, well-written means only having lines of code that are necessary to solving the problem, as well as presenting the solution for the reader to easily see. It might also be worth adding comments to your code.