# Study design

Please work on the practice problems in your group. At least one of the following problems will be assigned to the weekly problem set.

1. In Spring 2024, Prof. Tang asked her STAT 311 students to fill out a mid-semester survey. She was particularly interested in the the amount of hours her STAT 311 students were spending per week on the course. The average time spent on the course was found to be 10.2 hours per week.

   a. Based on this information, identify each of the following: observation, variable, parameter, and statistic.
   b. Was this survey a census or (just) a sample? Why?

2. Time to implement the sampling methods we learned! **Keep track of your work on this problem; we will return to it in the next few classes.**

   We have a farmer who grows sunflowers for making sunflower oil. Her field is arranged in a grid pattern, with 12 rows and 12 columns as shown below. Water is important for crops, so irrigation ditches have been installed along the top and bottom of the field. It is expected that plants closer to a water source will perform better than those further away from water.

   The farmer would like to estimate the number of healthy plants in the field, along with a few other characteristics about the sunflowers. It would be unfeasible to conduct a census, so we should choose to sample a subset of the grid cells. Suppose we'd like to sample $n = 12$ grid cells total.
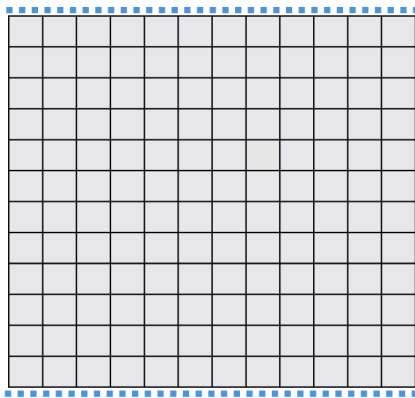
   a. Using words (sentences or bulleted list are fine) and perhaps labeling the figure below, describe exactly how you would obtain a sample of 12 grid cells using **simple random sampling**. I should be able to read your work and know what to do without any questions! Think about how you will perform the random sampling.

   Then implement the method that you've written down, and either using the figure below or drawing your own 12x12 field, shade in the squares that correspond to your sample.
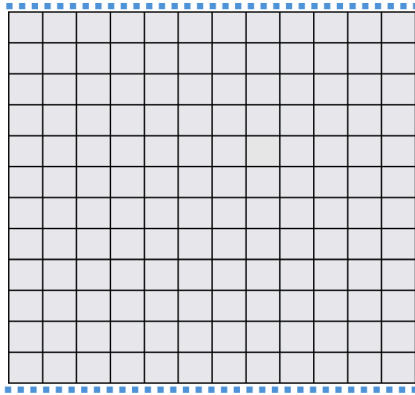
b. Using words (sentences or bulleted list are fine) and perhaps labeling the figure below, describe exactly how you would obtain a sample of 12 grid cells using **stratified sampling** where the strata are rows. I should be able to read your work and know what to do without any questions!

Then implement the method that you've written down, and either using the figure below or drawing your own 12x12 field, shade in the squares that correspond to your sample.
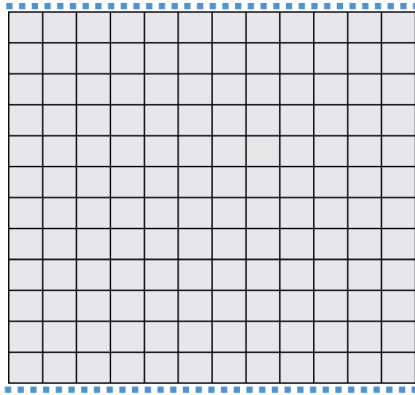


c. Using words (sentences or bulleted list are fine) and perhaps labeling the figure below, describe exactly how you would obtain a sample of 12 grid cells using **stratified sampling** where the strata are columns. I should be able to read your work and know what to do without any questions!

Then implement the method that you've written down, and either using the figure below or drawing your own 12x12 field, shade in the squares that correspond to your sample.
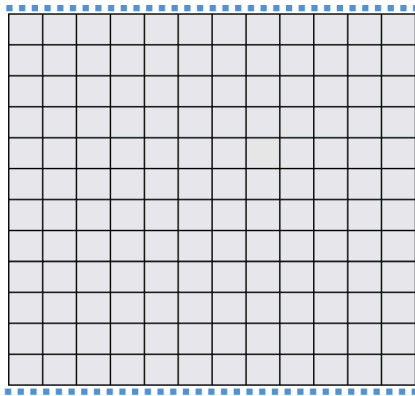
d. Using words (sentences or bulleted list are fine) and perhaps labeling the figure below, describe exactly how you would obtain a sample of 12 grid cells using **cluster sampling** where we have 24 clusters total. I should be able to read your work and know what to do without any questions!

Then implement the method that you've written down, and either using the figure below or drawing your own 12x12 field, shade in the squares that correspond to your sample.



e. Using words (sentences or bulleted list are fine) and perhaps labeling the figure below, describe exactly how you would obtain a sample of 12 grid cells using **multistage sampling**. I should be able to read your work and know what to do without any questions!

Then implement the method that you've written down, and either using the figure below or drawing your own 12x12 field, shade in the squares that correspond to your sample.
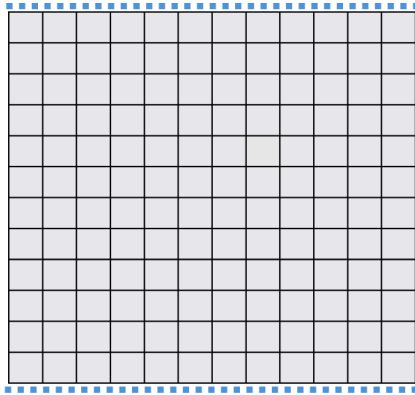
3. Suppose we want to estimate household size, where a *"household"* is defined as people living together in the same dwelling, and sharing living accommodations. If we select students at random at an elementary school and ask them what their family size is, will this be a good measure of household size? Or will our average be biased? If so, will it overestimate or underestimate the true value?

4. To assess the effectiveness of taking large doses of vitamin C in reducing the duration of the common cold, researchers recruited 400 healthy volunteers from staff and students at a university. A quarter of the patients were assigned a placebo, and the rest were evenly divided between 1g Vitamin C, 3g Vitamin C, or 3g Vitamin C plus additives to be taken at onset of a cold for the following two days. All tablets had identical appearance and packaging. The nurses who handed the prescribed pills to the patients knew which patient received which treatment, but the researchers assessing the patients when they were sick did not. No statistically discernible differences were observed in any measure of cold duration or severity between the four groups, and the placebo group had the shortest duration of symptoms.

   a. Was this an experiment or an observational study? Why?
   b. What are the explanatory and response variables in this study?
   c. Were the patients blinded to their treatment?
   d. Was this study double-blind?
   e. Participants are ultimately able to choose whether to use the pills prescribed to them. We might expect that not all of them will adhere and take their pills. Does this introduce a confounding variable to the study? Explain your reasoning.