This problem set covers material from Week 2, dates 9/16 – 9/19.

Instructions: Write or type complete solutions to the following problems and submit answers to the corresponding Canvas 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 9/16

1. Calculate the average/sample mean in each of the following scenarios:

   (a) Suppose that we have some data where 20% of the data are 1's, 50% are 2's, and 30% are 3's. What is the sample mean?

   (b) A school has two classes, one with 10 students and one with 100 students. What is the average class size?

   (c) A school has two classes, one with 10 students and one with 100 students. What is the average size of the class that a student is enrolled in?

2. Consider the following two sets of data:

$$\mathbf{x} = (1,\ 3,\ 4,\ 5,\ 7) \qquad\qquad \mathbf{y} = (6,\ 8,\ 9,\ 10,\ 12)$$

   (a) For each set of data $\mathbf{x}$ and $\mathbf{y}$, find the average and the standard deviation. Please use the proper symbols/notation!

   (b) How is the set of data $\mathbf{y}$ related to $\mathbf{x}$? How does this relationship carry over/affect the average and the standard deviation of $\mathbf{y}$ in comparison to those of $\mathbf{x}$?

3. 21 people in a room have an average of 5 feet 6 inches. A 22nd person enters the room. How tall would this person have to be to raise the average height by 1 inch? *(1 foot = 12 inches)*

4. Returning to the sunflowers again! Recall that in Problem Set 1, we used five different sampling schemes to obtain a sample of $n = 12$ grid cells. On our course website, find the 'Sunflower field' link corresponding to this assignment. For each sampling method, use this field to "collect data" (i.e. find the number) on the number of healthy sunflowers in each of the 12 grid cells you sampled.

   (a) For each sampling scheme, calculate i) the average number of healthy sunflowers in a grid cell, as well as ii) the standard deviation of the number of healthy plants in a grid cell.

   - To show your work for this problem: it is sufficient to write-out the calculations by hand (using "..." notation is fine!) for just one of the sampling schemes. But be sure to report the statistics for all schemes.

- I highly encourage you to use R to help perform the actual calculations. If you do use R, write-out an example of the code you execute for just one of the sampling schemes.

(b) Briefly comment on how the statistics you calculated compare across the various survey methods, and briefly discuss why.

Wednesday 9/18

5. The statistic $\frac{\bar{x}}{m}$ can be used as a measure of skewness (recall that m is the median). Suppose we have a distribution where all observations are greater than 0 (i.e. $x_i > 0$ for all observations $i = 1, \ldots, n$). What is the expected shape of the distribution under the following conditions? Explain your reasoning.

   (a) $\frac{\bar{x}}{m} = 1$

   (b) $\frac{\bar{x}}{m} > 1$

   (c) $\frac{\bar{x}}{m} < 1$

6. Returning to the sunflowers!

   (a) Obtain the median number of healthy sunflowers in a grid cell for each of the sampling strategies. (You're welcome to use code, but it may be faster to do this by hand). Briefly comment on how the sample medians compare to the sample means you obtained on Monday's problems.

   (b) Pick one of the sampling methods. Draw out the boxplot for the data you obtained under that sampling scheme, providing the statistics necessary to create the boxplot. In your answer, please specify which sampling method you chose!

Thursday 9/19

7. Problems in the associated .Rmd file. For grading purposes, R problems 1-4 will be treated as one problem, as will 5-6, and 8-10.

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.