

This problem set covers material from Week 1, dates 2/09- 2/13. Textbook problems can be found at the end of the corresponding subsection.

**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 2/09

None!

## Wednesday 2/11

None!

## Friday 2/13

1. For the following scenario, identify the statistical model (be specific about all components, please)!

An ecologist would like to learn about the water temperature of Otter Creek. However, it is known that water temperatures can fluctuate and the thermometer used to obtain measurements is a bit faulty. For this reason, it is good practice to take repeated measurements, and the ecologist decides to obtain 5 independent measurements. Suppose that the true temperature  $\theta$  of the water is unknown, but the variance in the measurements obtained is 0.1 degrees. Further suppose that the thermometer used to obtain the measurements is known to be positively biased by 0.2 degrees (i.e. the average value of the errors is 0.2). Assume that the errors are otherwise identically distributed normal random variables.

2. A jar contains 20 marbles, each of varying color and mass (but otherwise same shape/size). Marbles are drawn uniformly at random with replacement, one-by-one, from the jar. Every time we draw a marble, we record its mass before putting it back into the jar. Suppose we draw three marbles. Let  $M_1$ ,  $M_2$ , and  $M_3$  be the masses of the three marbles drawn. Further suppose  $c$  is an unknown positive constant, and that the jar contains exactly one marble of each of the following masses:  $\{c, 2c, 3c, \dots, 19c, 20c\}$ .
  - (a) Identify the statistical model (be specific about all components, please)!
  - (b) Give examples of three different statistics which, by definition, could be defined using just the data  $M_1, M_2, M_3$ . Which of your statistics do you think would be most useful as an estimator for  $c$ ? Why?

**General rubric**

Points	Criteria
5	The solution is correct <i>and</i> 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.