Homework 1 1/23 - 1/25

STA 336 Due 11:59pm Monday, January 30 Name:

Instructions: Write-up complete solutions to the following problems and submit answers on Gradescope. 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 rubric for homework problems appears on the final page of this assignment.

- Unless otherwise noted, problem numbers are taken from the 4th edition of DeGroot and Schervish's *Probability and Statistics*.

Monday 1/23

AP1. Suppose a coin with probability 0.4 of landing heads is flipped 100 times, and let X count the number of heads observed.

(a) What is the distribution of X (state both its name and parameters)?

(b) What is the probability that $X = 50$?

(c) Express $P(X \leq 30)$ as a sum. Then use R to find this probability.

(d) By the Central Limit Theorem, X is approximately Normally distributed. What is the mean and the variance of this Normal distribution?

(e) Use R and the previous Normal approximation to estimate $P(X \leq 30)$.

Wednesday 1/25

AP2. For each of the following situations, identify the *random variables* that are being measured, the *model* under consideration, the *parameter(s)* of this model, and the *parameter space*.

(a) A geologist measures the diameters of 10 pebbles in an old stream bed. Theoretical considerations lead the geologist to believe that the logarithm of pebble diameter is normally distributed with unknown mean μ and unknown variance σ^2. The geologist wishes to use the observations to obtain some information about μ and σ^2, but has in advance no knowledge of the magnitudes of the two parameters.

(b) A precision scale is being used to obtain 4 independent determinations of the mass of a very small object. Due to small fluctuations in environmental conditions, the scale gives different readings each time. Suppose that the true mass θ of the object is unknown, but that the variance in these readings is 0.01 gram. Moreover, suppose that the scale is known to be biased to the positive side by 0.1 grams (that is, the average value of errors is 0.1). Assume that the errors are otherwise identically distributed normal random variables.

(c) The number of eggs laid by an insect in one clutch is Poisson distributed with unknown mean λ. Once laid, each egg has an unknown chance p of hatching, and the hatching of one egg is independent of the hatching of the others. An entomologist studies a group of 100 such insects and counts the number of eggs laid, as well as eggs hatched, for each insect.

AP3. Suppose a jar contains 20 marbles, each of varying color and mass. Marbles are drawn uniformly from the jar one-by-one with replacement, and their masses are recorded. Let M_1, M_2, and M_3 be the masses of the first, second, and third marbles drawn. Suppose c is an unknown positive constant, and that the jar contains exactly one marble of each of the following masses: $\{c, 2c, 3c, \ldots, 19c, 20c\}$.

(a) Define a single sample space S that could be simultaneously used for all of M_1, M_2, M_3. That is, each of these variables should be functions with the same domain S, and so that for a particular outcome $\omega \in S$, it makes sense to evaluate $M_1(\omega), M_2(\omega), M_3(\omega)$.

(b) What is the support of the random variable M_1? (Recall, the support is the set of all values a variable can take).

(c) State the implied *model*, *parameter(s)* for this model, and the *parameter space*.

(d) Give examples of three different real-valued statistics (that is, statistics whose outputs are real numbers) which could be defined using just the data M_1, M_2, M_3. Which of these statistics do you think would be most useful for estimating the value of c? Why?

Homework 1 1/23 - 1/25

STA 336 Due 11:59pm Monday, January 30 Name:

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.