STAT 2593 - Problem Set 1
Due: Monday, October 16 on Crowdmark
Resubmit: Wednesday, November 1 on Crowdmark

Reminder that you are permitted to discuss to these problems with classmates, but every student must submit their own solutions which are their own work. **Please indicate any students that you discussed solutions with on your submission**. Please ensure that your submissions on Crowdmark are legible, and separated based on the problems included at the submission link. Submissions can be handwritten or typeset.

Part 1: True or False (20 Marks; 2 Marks Each)

For each of the following problems indicate whether the statement is true or false, and give a short justification for your answer. Correct answers without justification will receive only partial credit.

Problem 1: It is not possible for the sample standard deviation to exceed the sample mean, assuming that **all observations are positive**.

Problem 2: Suppose you have a list of numbers. Originally the smallest number in the list is 21.4. A modification is made which changes the smallest number to 17.3 without otherwise changing the list.
It is possible to determine the whether the median changes, and if so, by how much.

Problem 3: Two samples of manufactured rods have been taken. The first has 125 observations, with an average length of 120.3cm and a standard deviation of 10cm. The second batch is for 75 observations, with an average length of 104.2cm and a standard deviation of 10cm.
When the two batches are combined, the standard deviation of the 200 samples will exceed 10cm.

Problem 4: When there is an even number of observations, the sample median will *not* equal a value observed in the sample.

Problem 5: Suppose that my partner and I only eat one dinner per day. Consider the events A: I make dinner at home, B: my partner makes dinner at home, C: we order takeout for dinner, or D: we go out to eat. These events cover all possible actions for dinner. There is a pair of events from A, B, C, and D which are independent.

Problem 6: A chemical engineer is designing an experiment to determine the effect of various factors on the yield of a particular reaction. They are considering the effects of temperature, stirring rates, and the types of catalyst. Each experiment consists of selecting a different combination of the three factors. They consider five different temperature settings and four different stirring rates.
If a total of 87 experiments are run, there must be at least 5 catalysts considered.

Problem 7: Charles and Sadie are searching for mistakes in a series of math proofs. They each have a constant probability of detecting any of the given errors, denoted p_C and p_S respectively. They search the proofs independently from one another. In the document Charles finds 20 errors and Sadie finds 15 errors, 10 of which were also found by Charles.
It is expected that the total number of errors remaining in the document, after their proof-reading, is 5.
Hint: using the question setup, what is the expected number of total errors in the document? What is the expected number each will find? How many have they found?

Problem 8: A particular manufacturing company is forming several committees which are to be formed from members of management, labour, and the public. There are two representatives from management, three from labour, and four from the public. The advisory committee requires one representative from each of the groups. The audit committee requires one representative from management, one from the public, and two from labour.
There are more valid audit committees than there are advisory committees.

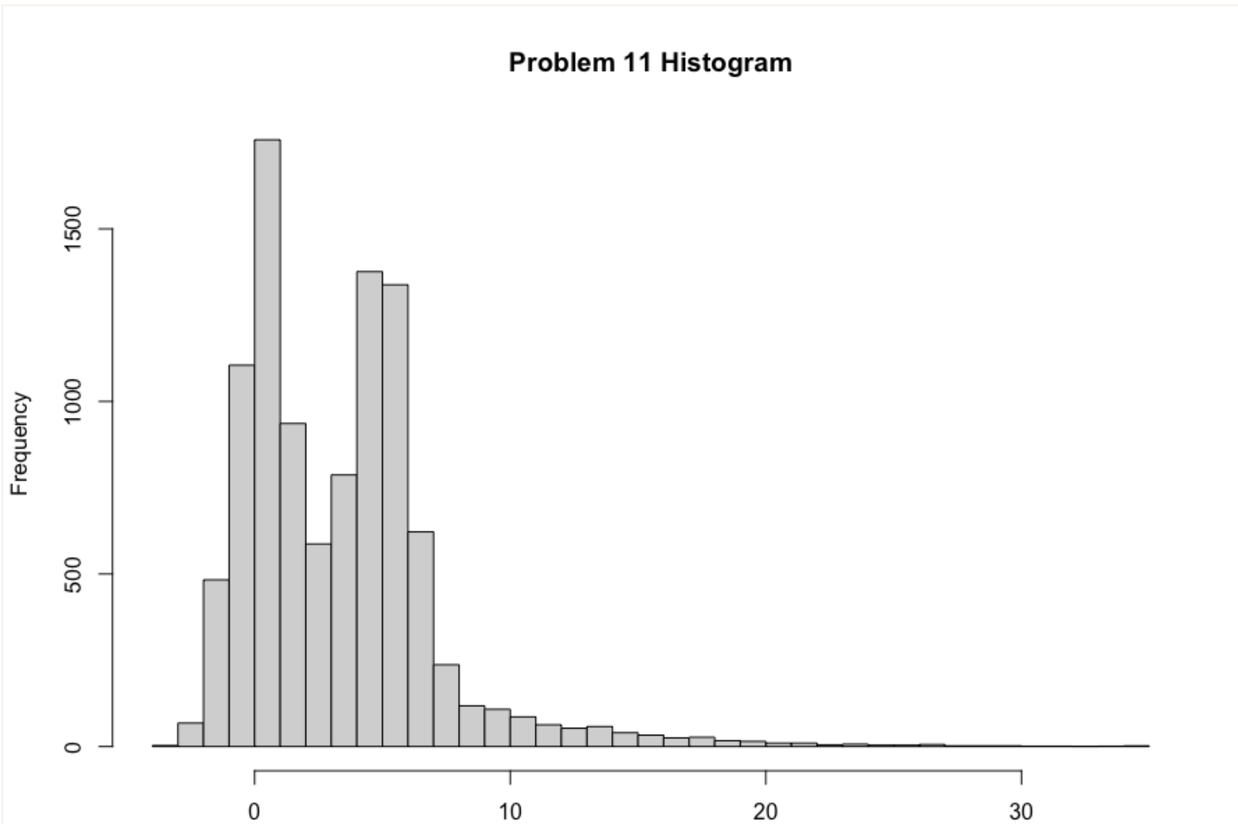
Problem 9: The number of days in a week that are rainy is modeled well by a binomial random variable with $n = 7$.

Problem 10: A particular venue hosts 52 concerts in a year, one every single weekend. Of the 52 different acts, 15 are bands that I would like to see. Suppose that I am available 7 weekends of the year to possibly attend concerts.
If we assume that the acts that I would like to see, and my availability, are random throughout the year, then I should expect to be able to see 2 bands that I would like to see.

Part 2: Conceptual Question (30 Marks)

For the following questions, provide your answers with justification and clear communication. The answers do not need to be long, but correct responses without complete justification will receive only partial credit.

Problem 11: (5 Marks) Describe the distribution displayed in the following histogram.



Problem 12: (5 Marks) A certain process for manufacturing CPUs has been in use for a period of time, and it is known that 12% of the CPUs it produces are defective. A new process that is supposed to reduce the proportion of defectives is being tested. In a simple random sample of 100 CPUs produced by the new process, 12 were defective.

(a) One of the engineers suggests that the test proves that the new process is no better than the old process, since the proportion of defectives in the sample is the same. Is this conclusion justified? Explain.

(b) Assume that there had been only 11 defective CPUs in the sample of 100. Would this have proven that the new process is better? Explain.

(c) Which outcome represents strong evidence that the new process is better: finding 11 defective CPUs or finding 2 defective CPUs?

Problem 13: (4 Marks) Prove that $S_{XX} = \sum_{i=1}^{n}(x_i - \overline{x})^2$ can be written as $\sum_{i=1}^{n} x_i^2 - n\overline{x}^2$.

Problem 14: (3 Marks) Jaclyn runs an Etsy store selling an assortment of handmade products. She keeps track of the weights of all of her items in a dataset. A new shipment company she begins to use wants weights recorded in kilograms, rather than the pounds that she has been using. As a result, Jaclyn converts the weights of her data from pounds to kilograms (1kg = 2.2lbs).

(a) How does this impact the mean weight, the variance of the weights, and the standard deviation of the weights?

(b) She is happy with the new company and has started to save some money from the better shipment rates. As a result, she is considering moving from envelopes to boxes to increase the perceived value of her goods. This packaging option is heavier, increasing the weight of each package by 50g.

If she chooses to do this, how would it impact the mean weight, the variance of the weights, and the standard deviation of the weights in her data?

Problem 15: (3 Marks) Consider an event from a statistical experiment, A. Is it possible for $A \perp A$? If so, explain when this can happen. If not, explain why it is not possible.

Problem 16: (4 Marks) Suppose two, six-sided dice are rolled. The values are looked at, and the larger is kept. If the values are equal, one of the two is kept at random. If X represent the value of the die that remains, show that its probability mass function is

$$P(X = x) = \frac{(2x-1)}{36}, \quad x = 1, 2, \dots, 6.$$

Problem 17: (6 Marks) A multiple choice test contains 50 questions, and is to be given as a pass-fail exam. For each question, there are three answers and only one is correct. The two incorrect answers are equally plausible, so that a student guessing will have a $\frac{1}{3}$ chance of correctly answering.

(a) Suppose that each correct answer is worth one mark, and you lose no marks for guessing. If the instructor wanted to ensure that a student who was guessing randomly had no more than a 1% chance to pass the exam. What is the smallest value that the pass-fail threshold can be set at?

(b) Instead of one mark per correct answer and no penalties for an incorrect answer, suppose that the students receive two marks for a correct question and one mark for an incorrect answer. Find the expected score, and the variance of that score, for a student who guesses on every question.

(c) Suppose that a score of 28 is used as the pass-fail threshold. Moreover, assume that students receive one mark for correct answers and lose one mark for incorrect answers. Garth only studied half the material, and as a result, answered the first 25 questions correct. For the other half, they can do no better than random guessing. Garth is considering guessing on anywhere between 1 and 6 more questions: how many should they guess on to have the highest probability of passing?

Part 3: Knowledge Extension (30 Marks)

Please note, the following questions are designed to expand on the concepts covered in the course. While you have all of the tools required to solve these questions, they require deeper thinking than the previous conceptual questions.

Problem 18: (10 Marks) Dan and Isa play a game with a single, fair, six-sided die as follows. Dan keeps throwing the die until she obtains the sequence $5-5$ on any two successive throws. Then Isa takes the die, and proceeds exactly as Dan did, except she stops whenever the sequence $5-6$ appears on any two successive throws. The winner of the game is the person who tossed the die the fewest times. For instance, consider the following:

```
Dan: 5 6 4 1 5 5
Isa: 3 5 3 5 4 6 5 6
```

Dan would win this example, since she saw a $5-5$ in only 6 tosses, where it took Isa 8 to see a $5-6$.

(a) On average, will both players require the same number of throws to finish their objectives? Explain without doing any explicit calculations.

(b) How many throws will the full game last, on average?

Problem 19: (10 Marks) Three friends, Lindsey, Toby, and Islay, take turns playing a game where a random number is drawn between 1 and 10 without any of them knowing. In the order Lindsey then Toby then Islay, the friends take turns guessing the value of the number. Lindsey first guesses, then looks to see whether the guess was correct. If not, Toby guesses one of the 9 remaining values, and looks to see whether it was correct or not. If not, Islay guesses one of the 8 remaining values, and checks to see whether it was correct.
If no one correctly guessed, a new random number is selected and a new round starts. If at any point anyone correctly guesses, they are the winner.

(a) Is this a fair game (in that each friend has an equal chance to win)? Demonstrate your reasoning.

(b) How many total guesses are expected in the game?

Problem 20: (10 Marks) Suppose that there is a spinner wheel, with six different colours on it. If each of the six colours occupies the same size region, then any a single spin of the spinner will land on any specific colour with probability of $\frac{1}{6}$. Suppose that cyan is one of these colours.

If the wheel were to be spun 144 times, we could represent the total number of times that cyan is observed as a random variable X, such that $X \sim \text{Bin}(144, \frac{1}{6})$. As a result, we would have $E[X] = 24$ and $\text{var}(X) = 144 \times \frac{1}{6} \times \frac{5}{6} = 20$.

(a) Suppose we are given two different spinning wheels, which are not equally distributed. Wheel A shows cyan with probability $\frac{1}{4}$ and wheel B shows cyan with probability $\frac{1}{12}$.

Consider performing 144 spins of the wheel again, this time where A is spun 72 times then B is spun 72 times.

How does the expected number of times landing on cyan, and the variance of this value, compare to the case with a fair wheel?

(b) Consider the same two wheels, A and B previously described. Once again, 144 spins of a wheel will be performed. This time, for every spin of the wheel, one is randomly selected and then spun.

How does the expected number of times landing on cyan, and the variance of this value, compare to the case with a fair wheel?

(c) Consider the same two wheels, A and B previously described. Once again, 144 spins of a wheel will be performed, however, this time one of the two wheels is selected at random, and then is used for all of the 144 spins.

Explain how the mean and the variance of this procedure will compare to the others. Note: you do not need to calculate definitive values.