Midterm 1 Review   **Math 222**

These are problems similar to the ones that might be on midterm 1. Be sure to also review class workshops and problems from the textbook.

1. The 2010 General Social Survey asked the question, "After an average work day, about how many hours do you have to relax or pursue activities that you enjoy?" to a random sample of 1,155 Americans. The average relaxing time was found to be 1.65 hours.

   (a) In this study, identify each of the following:

       Observational units: Variable:

       Parameter: Statistic:

2. In order 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 significant differences were observed in any measure of cold duration or severity between the four medication 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 or not 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.

3. Below are the final exam scores of twenty introductory statistics students.

```
47, 56, 59, 67, 71, 72, 73, 74, 78, 78,
79, 79, 81, 81, 82, 83, 83, 88, 89, 94
```

   (a) Draw a histogram for the scores.

   (b) Describe the distribution of the data (skewed left, skewed right, or roughly symmetric).

   (c) Without calculating anything, which is larger, the mean or the median of this data?

4. Facebook data indicate that 50% of Facebook users have 100 or more friends, and that the average friend count of users is 190. What do these findings suggest about the shape of the distribution of number of friends of Facebook users?

5. A food truck sells burgers and fries. A burger costs $10, fries cost $5. They estimate that 40% of customers just buy a burger, 50% buy a burger and fries, and the remaining 10% just buy fries.

   (a) Let X represent the amount of money that a random customer will spend at this food truck. Make a table showing the probability model for X (outcomes and probabilities).

   (b) Find the expected revenue per customer.

   (c) Find the variance and standard deviation in X.

   (d) If the food truck has 200 customers in one night, what is the mean and standard deviation in their total revenue for that night? Describe any assumptions you made in order to answer the question.

6. Men in the United States have an average height of 70.0 inches with a standard deviation 3.0 inches. Women in the USA have an average height of 64.3 inches with a standard deviation of 2.7 inches. Assume that both groups have a normal distribution.

   (a) How tall is a woman in the 90th percentile? Use the `qnorm(p, mean, sd)` function to express your answer.

   (b) What percent of men are between 68 and 72 inches tall? Use the `pnorm(x, mean, sd)` function to express your answer.

   (c) What is the probability that a man is over 6 feet tall (72 inches)? What is the probability that a woman is over 6 feet tall?

   (d) It turns out that 91% of women are shorter than 5'8" (68 inches) while only 25% of men are. Given that someone is shorter than 5'8", what is the conditional probability that they are a man? Hint: I recommend drawing a tree diagram.

7. When polling companies conduct political polls, they usually only choose samples of between 800 to 1000 people. What would happen if they choose larger samples?

   A. The random error would get larger, but bias would not be a problem.

   B. The random error would get smaller, but bias would not be a problem.

   C. The random error would get larger, but bias might still a problem.

   D. The random error would get smaller, but bias might still be a problem.

8. Probability models are based on assumptions. Sometimes, we use models even when we know that the assumptions are not all 100% true. Here are two situations where it is not a good idea to use a given probability model.

   (a) In one elementary school, there are 56 boys and 44 girls in the 5th grade. Suppose we randomly assign 25 fifth graders to one teacher's class. Why can't you use the Binom(25, 0.56) distribution to model the possible number of boys who will end up in that class?

   (b) The National AIDS Behavioral Surveys found that 0.2% (that's 0.002 as a decimal fraction) of adult heterosexuals had both received a blood transfusion and had a sexual partner from a group at high risk of AIDS. Suppose that this national proportion holds for your region. Explain why you cannot safely use a Normal approximation to model the distribution of the number of people in this group when you interview an random sample of 1000 adults.

9. Suppose X is the result when you roll a six-sided die and Y is the result when you roll a 20-sided die. The mean and standard deviation for X are $\mu_X = 3.5$, $\sigma_X = 1.7078$. The mean and standard deviation for Y are $\mu_Y = 10.5$ and $\sigma_Y = 5.7663$. Compute the following.

   (a) $E(X + Y)$

   (b) $\text{Var}(X + Y)$

   (c) The standard deviation of $X + Y$.

10. Data collected at elementary schools in DeKalb County, GA suggest that each year roughly 25% of students miss exactly one day of school, 15% miss 2 days, and 28% miss 3 or more days due to sickness.

   (a) What is the probability that a student chosen at random doesn't miss any days of school due to sickness this year?

   (b) If a parent has two kids at a DeKalb County elementary school, what is the probability that both of her kids will miss some school? Note any assumption you make.

   (c) If a parent has two kids at a DeKalb County elementary school, what is the probability that at least one of her kids will miss some school? Note any assumption you make.

   (d) Is the assumption that you need to make in parts (b) and (c) reasonable? Explain.

11. In the game roulette, if you bet on a color, like black or red, there is an 18/38 chance that you win. If you bet on a number, like 7, there is a 1/38 chance that you win.

   (a) If you play roulette 100 times, and bet on 7 every time, what is the probability that you win at least 3 times? Use the `pbinom(x, n, p)` function to express your answer.

   (b) If you play roulette 100 times, and bet on black every time, what is the probability that you win at least 50 times using the normal approximation? Use the `pnorm(x, mean, sd)` function to express your answer.