Unit 3 Review

<table>
<thead>
<tr>
<th>Sampling Distribution</th>
<th>Categorical Variable</th>
<th>Quantitative Variable</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sample Statistic</td>
<td>( \hat{p} ) (sample proportion)</td>
<td>( \bar{x} ) (sample mean)</td>
</tr>
<tr>
<td>Center (Mean)</td>
<td>( \pi ) (population proportion)</td>
<td>( \mu ) (population mean)</td>
</tr>
<tr>
<td>Standard Deviation</td>
<td>( \sqrt{n \left(1 - \pi\right)/n} )</td>
<td>( s / \sqrt{n} )</td>
</tr>
<tr>
<td>Shape for Large</td>
<td>Approximately normal</td>
<td>Normal or approximately normal</td>
</tr>
<tr>
<td>Sample Size ((n))</td>
<td>When ( n \geq 10 ) and</td>
<td>When population is</td>
</tr>
<tr>
<td></td>
<td>( n(1 - \pi) \geq 10 )</td>
<td>normal or ( n \geq 30 )</td>
</tr>
</tbody>
</table>

1. Suppose that two employees have to be fired from a group of seven:
   - Amy, female, age 35
   - Bob, male, age 41
   - Chad, male, age 43
   - Diana, female, age 44
   - Elvis, male, age 49
   - Frank, male, age 61
   - Ginger, female, age 62

Suppose for now that the company selects the two employees to be fired at random, so all possible pairs of people have the same probability of being fired.

a. List all outcomes in the sample space. In other words, list all possible pairs of people that can be chosen from this group of seven people. (To save time and space, just use initials. For example, write AB to represent the (Amy, Bob) pair.)

b. Determine the probability (both as a fraction and a decimal) that both employees selected are female.

c. Determine the probability (both as a fraction and a decimal) that the average age of the employees selected is older than 50.

d. Determine the probability that both employees selected are older than 60. Suppose that Frank and Ginger are the employees selected to be fired. They proceed to file a lawsuit, alleging that the company is guilty of age discrimination for choosing the two oldest employees to fire. The company responds that it did not discriminate; it simply happened to choose these two by random chance.

e. Does your probability analysis provide evidence to support the allegation of age discrimination? Explain (with 1–3 sentences) the reasoning process behind your answer.

2. Suppose that the drying time for a certain type of paint under specified test conditions is known to be normally distributed with mean 75 minutes and standard deviation 5 minutes. Suppose also that chemists have devised a new additive that they hope will reduce the mean drying time (without changing the standard deviation). A test is then conducted to measure the drying time for a test specimen, and the company executives decide that they will be convinced that the additive is effective only if the drying time on this specimen is less than 70 minutes.

a. If the additive actually has no effect at all on the drying time, what is the probability the company executives will mistakenly conclude that it is effective? Include a sketch with your calculation.

b. If you want to alter the cutoff value from 70 in order to reduce the error probability in part a to .05, what cutoff value should you choose?
c. Now suppose that the standard deviation of the drying times to 65 minutes is 2 rather than 5 minutes. Without doing any new calculations, describe how this change would affect your answers to parts a and b. Give an intuitive explanation for your reasoning in both cases.

3. Suppose that 80% of all the incoming email messages for a college’s computer system are spam.
   a. Is the 80% value a parameter or a statistic for the population of incoming email messages at this college? Explain.
   b. What symbol would be used to represent the .8 value?

Consider taking a random sample of 200 incoming email messages at this college and calculating the sample proportion of these messages that are spam.
   c. Determine the mean and standard deviation of the sampling distribution of this sample proportion.
   d. Based on this mean and standard deviation, would it be surprising if the sample turned out to have less than 50% spam? Explain.

4. Suppose 20% of all heart transplant patients do not survive the operation.
   a. Think about taking repeated random samples of 371 patients from this population. Describe how the sample proportion who die would vary from sample to sample. 
      (Hint: Be sure to refer to the shape, center, and spread of its sampling distribution.)
      Also include a well-labeled sketch to represent this distribution.
   b. Suppose you take a random sample of 371 heart transplant patients. Determine the probability that the sample proportion who die would be .213 or higher.

5. Suppose I tell you that I flipped a coin multiple times and got 75% heads. Would you be reasonably convinced that this was not a fair coin (where “fair” means that the coin has a 50/50 chance of landing heads or tails)? If so, explain why. If not, describe what additional information you would ask for and explain why that information is necessary.

6. The distribution of house prices is skewed to the right because most houses cost a modest amount but a few cost a very large amount. If you take a random sample of 1000 houses, can you reasonably expect the distribution of the house prices to be approximately normal? Explain your answer.

7. Reconsider the activities involving the sampling of Reese’s Pieces candies. Continue to suppose that 45% of all Reese’s Pieces are orange.
   a. Is .45 a parameter or a statistic? What symbol would you use to represent it?
   b. Does sampling variability refer to a statistic, a parameter, or both?
   c. Are you more likely to get between 35% and 55% orange candies if you take a random sample of 40 candies or a random sample of 400 candies? Explain briefly.
   d. Are you more likely to get more than 60% orange candies if you take a random sample of 40 candies or a random sample of 400 candies? Explain briefly.

Now suppose that 20 people each take a random sample of 40 candies, calculate the sample proportion of orange candies, and create a dotplot of their results.
   e. Identify the observational units and variable of this dotplot.
8. Suppose that a phone company reports that the average duration of a cell phone call is 1.7 minutes, with a standard deviation of 1.4 minutes.
   a. Would it be reasonable to use a normal distribution to model the duration of cell phone calls? Explain, based primarily on the values reported above.
   b. Suppose you want to examine a random sample of 60 cell phone calls. Do you think it would be reasonable to use the Central Limit Theorem to describe the sampling distribution of the sample mean call duration? Explain.
   c. What does the CLT say about the sampling distribution of the sample mean call duration in a random sample of 60 calls?
   d. Draw a well-labeled sketch to accompany your answer to question 3.
   e. Describe how the sketch would change if the sample size were 160 calls rather than 60 calls.

9. If you bet one dollar on a color (red or black) in a spin of a roulette wheel, then you will either win one dollar or lose one dollar. Since there are 18 red, 18 black, and 2 green (house wins if green comes up), the probability is 20/38 (about .526) that you will lose one dollar and 18/38 (about .474) that you will win one dollar. Explain what it means to say that the probability of winning one dollar is .474, as if to someone with no formal knowledge of statistics.