Name: ____________________________________ Date: __________________

SAMPLE PROPORTIONS
COMMON CORE ALGEBRA II

Many times we are interested in determining a confidence interval for the population mean, μ, based on a sample mean, $\overline{x}$. Sometimes, though, we want to simply know what proportion, *p*, of a population shares a certain characteristic. We again infer characteristics about *p* based on the proportion of a sample, $\hat{p}$ (p "hat" as it is often called).

***Exercise* #1:** A school is trying to determine the proportion of students who own cell phones. They do a survey of **all** juniors and find that 168 out of 236 of have cell phones. They then take a **sample** of freshmen and find that 30 out of 52 freshmen in the sample own cell phones.

(a) Calculate the population proportion, *p*, of juniors who own cell phones. Round to the nearest hundredth.

(b) Calculate the sample proportion, $\hat{p}$, of freshmen who own cell phones. Round to the nearest hundredth.

Clearly, in the last example, the sample proportion of freshmen who own cell phones is less than the population proportion of juniors who own cell phones. But, can we attribute that variability to the two "treatments", i.e. juniors versus freshmen, or could the variability be due to **sampling variability**, i.e. the random chance that we just picked a group of freshmen who have an unusually low rate of cell phone ownership? We can establish how likely this is to happen by using simulation.

***Exercise* #2:** We would like to determine how likely it is that a sample of 52 out of a population with a proportion of cell phone ownership of 71% or 0.71 would result in a sample proportion of 0.58.

(a) Run the program PSIMUL with a *p* value of 0.71 and a sample size of 52 for 100 simulations. How many of the 100 simulations had a proportion less than or equal to 0.58?

(b) Based on your answer to (a), how likely is it that a sample of 52 from a population with a cell phone ownership of 71% would result in a sample proportion of only 0.58 or less?

(c) Is it possible that a sample of 52 from a population with a cell phone ownership rate of 71% could have a sample proportion of 0.58 or less? Justify your answer.

(d) What conclusion can you make about freshmen cell phone ownership compared to ownership by juniors? Explain.

Inferential statistics is never about proving beyond **any doubt** that a sample either can or cannot come from a certain population. It is about **quantifying how likely it is that it could come from a given population**. Let's continue exploring this question of sample proportions.

***Exercise* #3:** Let's say we have a population with a 0.25 proportion of being 65 years or older. Let's take different sized samples from this population and see how the sample proportions behave. Use the program PSIMUL to simulate a population with a proportion of 0.25 for various sample sizes and 100 simulations.

(a) Fill in the table below.

Sample Size	Low to High $\hat{p}$ values	Range in $\hat{p}$
10		
20		
50		
100		

(b) What was the effect of increasing the sample size on the sample proportions that were simulated? Why does this make sense?

(c) Run PSIMUL one more time with a sample size of 50 but for 200 simulations. Using your results, find the value that represents the 5^{th} percentile of $\hat{p}$ values. Find the result that represents that 95^{th} percentile of the $\hat{p}$ values. Then, write the **90% confidence interval** for this sample size coming from this population.

(d) If researchers surveyed 50 people walking out of a movie and found that 21 of them were 65 years or older, do you believe this sample came from the general population with a $p = 0.25$? Why or why not.

Name: ___________________________________ Date: _________________

SAMPLE PROPORTIONS
COMMON CORE ALGEBRA II HOMEWORK

APPLICATIONS

1. Historically, the proportion of emperor penguins with adult weights above 60 pounds is 0.64. Take this to be the population proportion for this characteristic.

   (a) A sample of 26 emperor penguins in a zoo found that 20 of the penguins had adult weights above 60 pounds. Calculate the sample proportion, $\hat{p}$, for this sample.

   (b) Run PSIMUL with a population proportion of $p = 0.64$ with a sample size of 26. Do 100 simulations. What percent of these simulations resulted in a $\hat{p}$ value at or above what you found in part (a)?

   (c) Do you have enough evidence from (b) to conclude that penguins raised in a zoo have a significantly higher proportion of weights above 60 pounds? Why or why not?

2. Let's stick with our emperor penguins from #1. Out of a sample of 56 penguins from a zoo, it was found that 43 penguins had weights over 60 pounds. Run PSIMUL again, but now with a sample size of 56. Continue to use $p = 0.64$ and 100 simulations. Do you now have stronger evidence that penguins raised in zoos have a higher proportion with weights over 60 pounds? Explain.

3. In general, as sample size increases, the range in the distribution of sample proportions

   (1) increases (3) decreases

   (2) stays the same (4) could increase or decrease _________

4. In a population with a proportion $p = 0.35$, if samples of size 30 were repeatedly taken, then we would expect approximately 90% of those samples proportions to fall within which of the following ranges?

   (1) 0.28 to 0.42 (3) 0.21 to 0.49

   (2) 0.18 to 0.54 (4) 0.31 to 0.39 _________

5. A sample of the graduating high school class was questioned about their plans after college. We worked with this sample of graduating seniors in our unit on probability. The two-way frequency chart below summarizes the results of the questionnaire. The school would like to investigate the effect of gender on the rate that students go to college.

	Gender		
	Male	Female	Total
Going to College	16	13	29
Not Going to College	14	9	23
Total	30	22	52

(a) Calculate the sample proportion of students going to college for the subgroups male and female. Round to the nearest hundredth.

$\hat{p}(\text{males}) = \dfrac{\text{number of males going to college}}{\text{total number of males}}$ $\hat{p}(\text{females}) = \dfrac{\text{number of females going to college}}{\text{total number of females}}$

The proportion for females going to college is higher than for males going to college (a greater percentage of females go to college than males). Is this due to **induced variability** or **sampling variability**? This boils down to asking if the difference is **statistically significant**.

(b) Design a simulation that would test how likely it is for a sample of 30 (the number of men) from a population that has the $\hat{p}(\text{female})$ would result in the $\hat{p}(\text{male})$ **or below**. Explain the simulation and what results you found.

(c) Based on the table above, would you conclude that the overall population proportion of females going to college is greater than the proportion of males? If you believe you have enough evidence from your simulation, explain why. If you do not believe you do, also explain.

(d) Why could this study be an example of both a sample survey and an observational study? Look back at their definitions from the first lesson to fully explain your answer. Also explain how the types of variability introduced.