

Name:

APPLICATION: Analyzing College Statistics

Level 1

ACTIVITY: [Sampling Methods Card Sort](#)

Follow your teacher's instructions to complete this activity.

1. Match each word description, equation, graph and table of values that represent the same function. Write your answers or paste the cards below.

Match 1	
Match 2	
Match 3	
Match 4	

College Student Aid

You perform a hypothesis test with a null hypothesis that the average percentage of undergraduate students who receive federal student financial aid is 44.5%. Your alternative hypothesis is that more students receive financial aid than that, and you are testing with a significance level of 0.05. You perform your test and come up with a p-value of 0.0224.

- 2. What does this p-value signify in real terms for your test?**

- 3. Give TWO different ways to word your statement to accurately describe the results of your test**

- 4. When you collected data for your sample, you referenced the list of the largest public universities in the nation and used simple random sampling from that list to get 400 data points.**
 - a. What is one possible error in your sample due to selection bias?

 - b. Explain how you might select a better sample to reduce selection bias sampling error.

Level 2

Ethics in Testing

You decide to perform a statistical analysis to test the claim that 62.5% of undergraduate students receive some form of federal, state, or local financial aid for college. First you need to decide how you are getting your sample data.

1. **You have a list of all undergraduate universities in the country available and you decide to use a stratified sampling method to select your sample data. List at least one way you could divide the universities into groups and explain why you chose those groups.**

You collect your data and use a sample of 500 universities out of over 5,000 and perform your tests. You find that the p-value is 0.055 and you were using a significance level of 0.05.

2. **What result would you report for your statistical test?**
3. **List TWO ways you could word the interpretation of the result of your test.**
4. **You decide instead to change how you report your results. They claim that “Our data supports the rejection of the null hypothesis at the 94.5% confidence level.” This is called ‘cherry-picking’ when you change your chosen significance level after performing your test, and it is scientifically unethical. Another group gets a p-value of 0.10 and reports that they reject the null hypothesis with 90% confidence. A third group with p-value of 0.22 publishes their data saying the data shows a statistically significant difference from the null hypothesis at the 78% confidence level. What do you think the problem can be with changing your confidence level after you perform your test analysis?**
5. **You feel you are really close and are still confident that the null hypothesis is incorrect. What changes could you make in a follow up experiment given the results of your original experiment with p-value of 0.055?**

Level 3

Bias in College Costs Reporting

Let's take a look now at the effect of sample size on the accuracy of our estimates. Smaller samples may be easier to obtain but there are trade-offs in how significant your results are. The average tuition for all 53 states and territories in this data is \$19,029. We will see the effect on sampling error with different size samples.

For each of the sample sizes, three different samples of that size were selected. All samples were taken using a simple random sampling method.

1. Find the averages of all values to complete the chart for the three samples with size 3.

SAMPLE SIZE 3	Sample 1	Sample 2	Sample 3	Average of Samples
Mean	\$16,847	\$18,309	\$18,201	
Absolute Value of Difference from Population Mean	\$1,751	\$1,523	\$2,419	

Complete the following chart with your values from question 1. Then, use it to answer the questions.

Sample Size	Average Absolute Value of Difference from the Population Mean	Average of the Square of the Difference from Population Mean
3		\$3,352
10	\$1,243	\$2,443
30	\$605	\$1,771
50	\$80	\$192

2. What do you notice, generally, about the differences as the sample size gets closer to the population size?
3. Why would it be important to measure the squares of the difference as well as the Difference itself?

4. What further advantage does measuring the squares of the values have over just the absolute value?
5. Explain in your own words why sample size is an important part of any study to know and understand.

Below is a chart of data with the average amount of federal and state aid awarded to students for both a population and different samples. Each sample size was 50 to eliminate any sampling error due to size, and different sampling techniques were used for each sample as noted below. Compare how each sampling technique measured up to the true population average of \$8,123.

- Simple Random Sampling - assigned random number and selected first 50 values
- Systematic Sampling - assigned different random numbers and selected every third value
- Stratified Sampling - Divided into 5 groups based on the percentage of undergraduate students awarded federal student loans. The categories were: 0-20%, 21-40%, 41-60%, 61-80%, and 81-100% of students receiving federal loans. Ten values were then selected from each group using simple random sampling on the group.

Simple Random	\$6,549
Systematic	\$8,924
Stratified	\$7900

6. Order the sampling techniques from the easiest/quickest to the most difficult/slow to use.
7. Which technique seems to be the most accurate? Give a reasonable explanation for why that technique might have some more accuracy than the others.
8. Your friend decides that because this technique was the most accurate, they are going to use only this one from now on. "If it is the most accurate this time, it will always be the most accurate," they argue. What counterargument could you offer your friend against only relying on this one sampling technique?