Introduction to Randomization Tests Name: _____ #_____

Yesterday we randomly assigned two groups to take a quiz on prime numbers. One group received a calculator and one group did not receive a calculator. We're going to use that data to determine if using a calculator had a significant impact on quiz scores.

Randomization test by hand:

1. Copy down the data from your class and find the mean quiz score of each group of students. (Do this on Data Sheet)

2. Find the difference of the mean quiz scores. (With Calculator Mean – Without Calculator Mean)

3. If we assume that the calculator has no effect, a student who scored a 5 without a calculator should score what with a calculator?

Because we are assuming that the calculator has no effect we can randomize our data set to get a new set of means.

4. Cut apart the data and shuffle cards all together. Now sort into two piles "with calculator" and "without calculator" and find the means of each. Then find the difference in means.

5. Repeat this process as for 10 minutes.

Randomization using technology:

Now your teacher will use the computer software to demonstrate. We'll repeat this process 1000 times and plot each difference of the means on a histogram. In order for us to see if using the calculator was statistically significant we need to find if the actual difference in the mean falls within the margin of error of our randomized values.

6. Create a histogram to display your differences in means for as many trials as you performed.

7. What is the Margin of Error for a 95% confidence interval in this randomization simulation?

8. Determine if your answer for #2 falls outside of the range of the Margin of Error. What does this tell us about the use of a calculator on this quiz?

Follow Up Questions:

9. Researchers at the Smell and Taste Foundation were interested in the following question.

Can pleasant aromas improve ability to complete a task?

They randomly assigned volunteers to wear an unscented mask or to wear a floral-scented mask. The subjects then completed two pencil-and-paper mazes. The time to complete the two mazes was recorded.

Unscented Mask (in sec)	38.4	72.5	82.8	50.4	32.8	40.9	56.3
Scented Mask (in sec)	38.0	35.0	60.1	44.3	47.9	46.2	

a) Calculate the mean time it takes to complete the maze for each group.

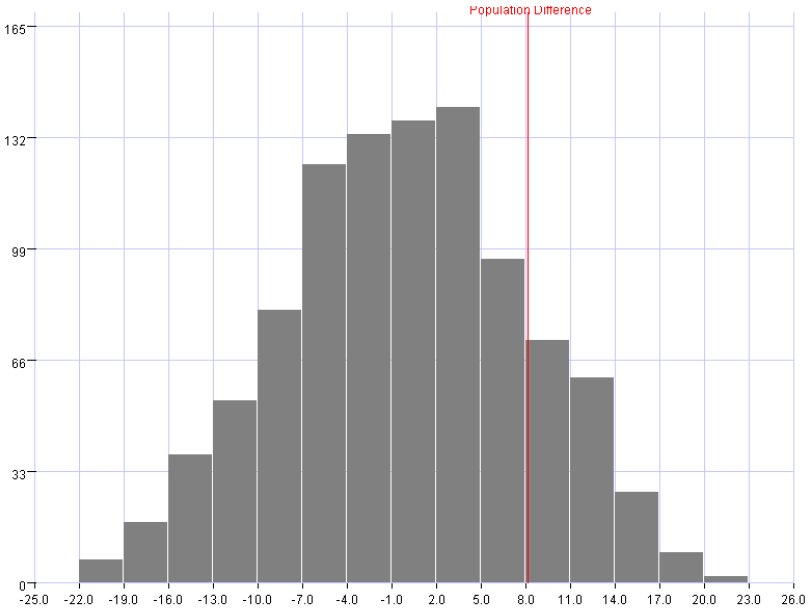
b) Calculate the difference in the means. (unscented – scented)

Assume that the type of mask makes absolutely no difference in how long it takes a person to complete the mazes.

c) Why would you expect there to be a nonzero difference in the mean times for the two treatments?

d) How long should it take the people who wore the unscented masks to complete the mazes if they had worn scented masks instead?

A randomization distribution was created using CPMP-Tools to approximate the distribution of possible differences (unscented mean – scented mean). The display shows 1,000 runs of randomly assigning the 13 times required to complete he mazes to either the unscented mask treatment or the scented mask treatment.



e) The mean of the randomization test is -0.1 with a standard deviation of 7.5. Calculate the Margin of Error for this randomization distribution.

f) Does the population difference in means fall into the interval for the Margin of Error?

g) Which of the following is the best conclusion?

- The difference of 8.19 from the actual experiment is statistically significant, so you should abandon your supposition that the scent made no difference.
- It is quite plausible that the scent does not affect the time to complete the maze. In other words, a difference of 8.19 seconds would not be unusual if the scent made no difference and you randomly divide the subjects into two groups.

10. Chrysanthemums with long stems are likely to have smaller flowers than chrysanthemums with shorter stems. An experiment was conducted at the University of Florida to compare growth inhibitors designed to reduce the length of the stems, and so, increase the size of the flowers. Growth inhibitor A was given to 10 randomly selected plants. Growth inhibitor B was given to the remaining 10 plants. The plants were grown under nearly identical conditions except for the growth inhibitor used. The table below gives the amount of growth during the subsequent 10 weeks.

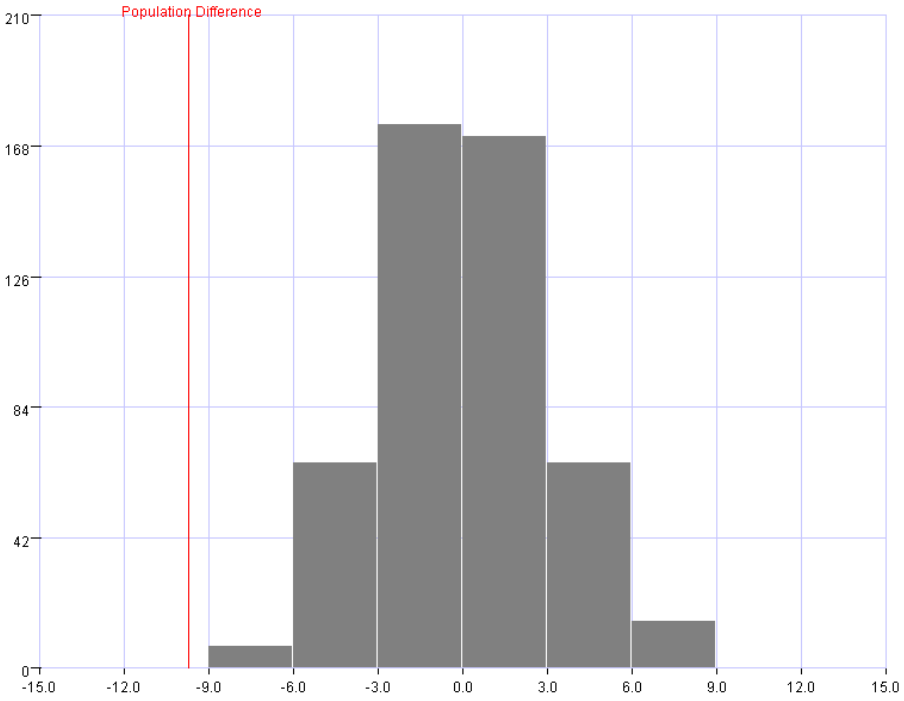
Growth by Plants Given A (in cm)	46	41.5	45	44	41.5	50	45	43	44	30.5
Growth by Plants Given B (in cm)	51	55	57	57.5	53	45.5	53	54.5	55.5	45.5

a) Examine the following summary statists. Which growth inhibitor treatment appears to be better?

Treatment	Mean	Standard Deviation
A	43.05	5.04
B	52.75	4.28

b) Describe as specifically as possible how to use a randomization test to decide whether, on average, one growth inhibitor works better than the other.

A randomization distribution was created using CPMP-Tools showing the approximate distribution of possible differences (growth inhibitor A mean - growth inhibitor B mean). 500 random assignments were run. The results are shown in the histogram to the right.



c) The mean of the randomization test is 0.06 with a standard deviation of 3. Calculate the margin of error for this randomization distribution.

d) Is the population difference in the growth inhibitors statistically significant? Explain how you know.

e) What can you conclude from this randomization test?

11. Explain why this statement is true: Even if the response for each subject would be the same no matter which treatment he or she receives, there is almost always a nonzero difference in the means of the actual responses from the treatments.

12. What does it mean if the results of an experiment are called "statistically significant"?

13. Forty-nine volunteer college students were randomly assigned to two treatments. Twenty-five students were told that they would view a video of a teacher who other students thought was "charismatic": lively, stimulating, and encouraging. The remaining twenty-four students were told that the instructor they would view was thought to be "punitive": not helpful, not interested in students, and a hard grader. Then all students watched the same twenty-minute lecture given by the same instructor. Following the lecture, subjects rated the lecturer. The summary statistcs from the students' ratings are given below. Higher ratings are better.

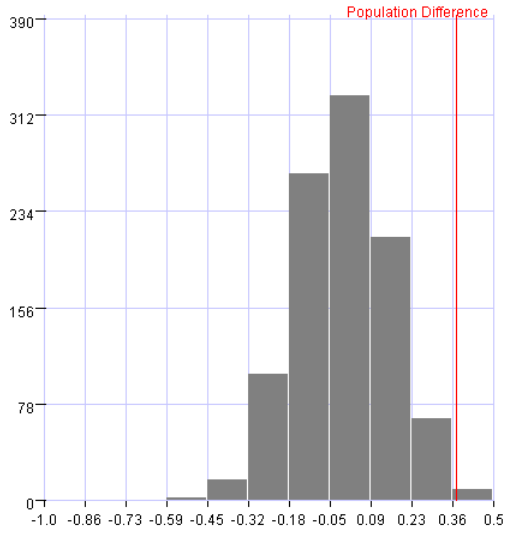
Treatment	Mean	Standard Deviation
Charismatic	2.61	0.53
Punitive	2.24	0.54

a) What are the two treatments?

b) From the summary statistics, does it look like the two treatments cause different responses? Explain?

c) Describe how to perform one run for a randomization test to decide whether the two different treatments result in different mean ratings.

The randomization distribution below shows mean charismatic- mean punitive for 100 runs.



d) The mean of the randomization distribution is 0. The standard deviation is .18. Calculate the margin of error

e) Was the difference in the two treatments statistically significant? Explain how you know.

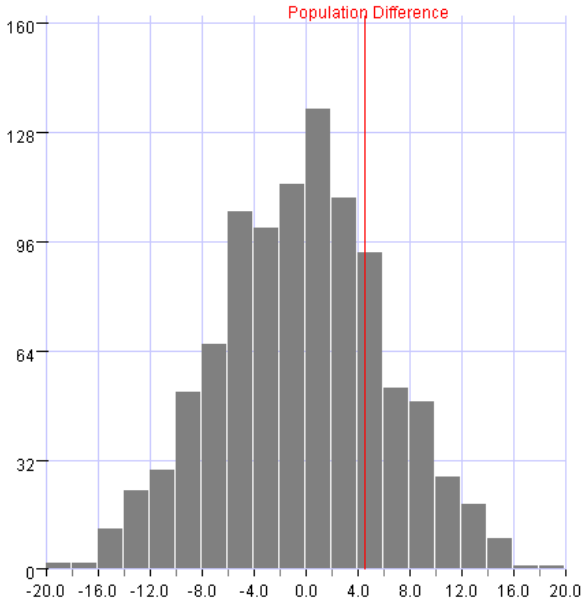
f) What can you conclude from this experiment?

14. For a science project, Brian wanted to determine whether eleventh-graders did better when they took a math test in silence or when Mozart was being played. Twenty-six students were randomly divided into the two treatment groups. Parts of Brian's results are in the table below.

Mozart (% correct)	65	80	72	68	38	58	45	42	58	81	40	41	27
Silence (% correct)	44	70	68	58	58	47	54	44	61	61	9	52	30

a) What is the difference mean response for Mozart group – mean response for silence group?

The randomization distribution below shows mean response for Mozart group – mean response for silence group for 500 runs



b) The mean of the randomization distribution is -0.24 and the standard deviation is 6.7. Calculate the margin of error.

c) Was the difference in the two treatments statistically significant? Explain how you know.

d) What can you conclude from this experiment?

e) Brian concluded that students who listen to Mozart during a test tend to do better. Do you agree with this conclusion or do you think the difference can be reasonably be attributed to random assignment alone?