Name _____ Date _____

1. Suppose you wanted to determine whether students who close their eyes are better able to estimate when 30 seconds have passed than students who do not close their eyes. (You ask students to tell you when to stop a stopwatch after they think 30 seconds have passed.) You find the first 50 students arriving at school one day. For those 50, you flip a coin to decide whether or not they will close their eyes during the test. Then, you compare the amounts by which each group overestimated or underestimated.

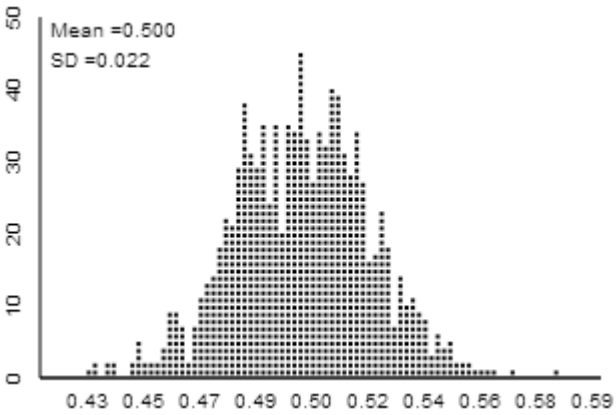
   a. Did this study use *random sampling*? Explain your answer by describing what purpose random sampling serves in such a study.

   b. Did this study make use of *random assignment*? Explain your answer by describing what purpose random assignment serves in such a study.

   c. Would the study described above be an *observational study* or an *experimental study*? Explain how you are deciding.

2. A Gallup poll conducted July 10–14, 2013, asked a random sample of U.S. adults: "How much attention do you pay to the nutritional information that is printed on restaurant menus or posted in restaurants, including calories and sugar and fat content?" The sample results were that 43% of the respondents said they pay a "fair amount" or a "great deal" of attention. Suppose there had been 500 people in the study.

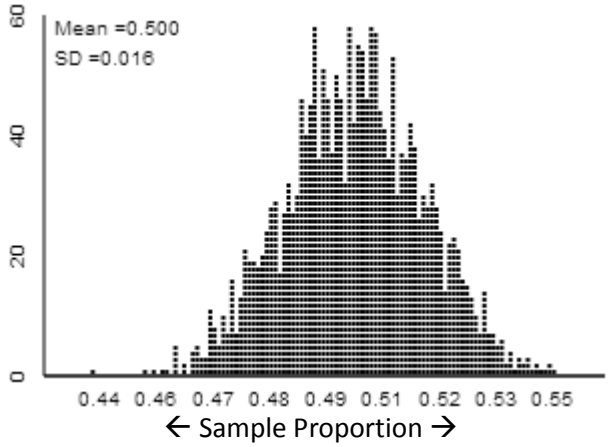
   The following graph displays the results from $1{,}000$ random samples (each with sample size 500) from a very large population where 50% of respondents "pay some attention" and 50% "pay little or no attention."



   a. Based on the simulation results above, are the sample data (43% responding "pay some attention") consistent with the simulation? In other words, do these results cause you to question whether the population is $50/50$ on this issue? Explain.

   Do you believe it is reasonable to generalize the results from this study to all U.S. adults? Explain.

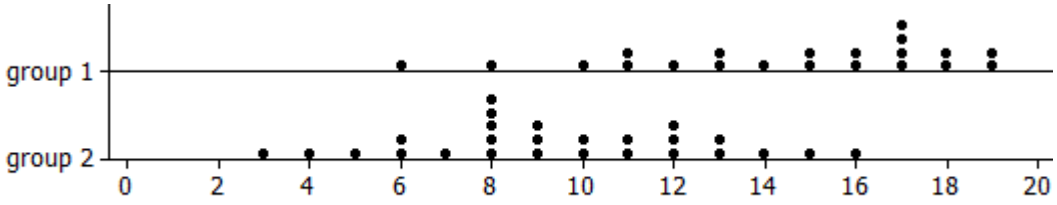
Suppose Gallup plans to conduct a new poll of a random sample of 1,000 U.S. adults on an issue where the population is evenly split between two responses. The following graph displays the results from 2,000 random samples (each with sample size 1,000) from such a population.



b. Based on these simulation results, estimate the expected margin of error for the Gallup poll. Explain how you developed your estimate.

c. Suppose the study used a sample size of 2,000 instead of 1,000. Would you expect the margin of error to be larger or smaller?

3. A randomized experiment compared the reaction time (in milliseconds) for subjects who had been sleep deprived (group 1) and subjects who had not (group 2).



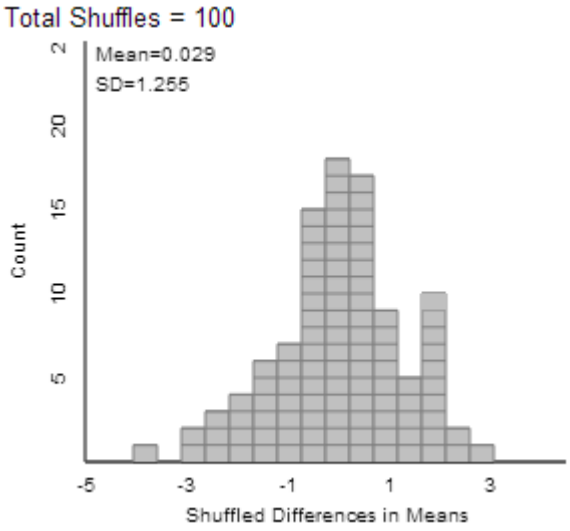
Variable	N	Mean	StDev
group 1	21	14.38	3.64
group 2	26	9.50	3.33

a. Based on the above output, for which group would it be more reasonable to use a normal curve to model the reaction time distribution? Justify your choice.

b. The difference in means is $14.38 - 9.50 = 4.88$. One of the researchers claims that the reaction time if you are sleep deprived is 5 ms greater than the reaction time if you are not sleep deprived. Explain one reason why this claim is potentially misleading.

c. Describe how to carry out a simulation analysis to determine whether the mean reaction time for group 1 is significantly larger than the mean reaction time for group 2.

d. The graph below displays the results of 100 repetitions of a simulation to investigate the difference in sample means when there is no real difference in the treatment means. Use this graph to determine whether the observed mean reaction time for group 1 is significantly larger than the observed mean reaction time for group 2. Explain your reasoning.



A Progression Toward Mastery					
Assessment Task Item		**STEP 1** Missing or incorrect answer and little evidence of reasoning or application of mathematics to solve the problem.	**STEP 2** Missing or incorrect answer but evidence of some reasoning or application of mathematics to solve the problem.	**STEP 3** A correct answer with some evidence of reasoning or application of mathematics to solve the problem. OR An incorrect answer with substantial evidence of solid reasoning or application of mathematics to solve the problem.	**STEP 4** A correct answer supported by substantial evidence of solid reasoning or application of mathematics to solve the problem.
1	a S-IC.B.3	Student provides an answer that is not germane to the design of the study.	Student confuses random sampling with other random forms such as random assignment.	Student indicates that the study did not use random sampling, but the description of the purpose of random sampling is inadequate (e.g., to draw cause-and-effect conclusions).	Student indicates that the study did not use random sampling and describes the purpose as obtaining a representative sample.
	b S-IC.B.3	Student provides an answer that is not germane to the design of the study.	Student confuses random assignment with other random methods such as random sampling.	Student indicates that the study did use random assignment, but the description of the purpose is inadequate (e.g., only to draw cause-and-effect conclusions but with no explanation why that works).	Student indicates that the study did use random assignment and describes the purpose in terms of balancing out confounding variables.
	c S-IC.B.3	Student is not consistent with a choice between the two types of studies.	Student identifies the study as observational, and/or the justification contradicts student's choice.	Student identifies the study as experimental, but the justification is inadequate (e.g., "They controlled variables").	Student identifies the study as experimental and justifies based on the imposition of the eyes being open or closed and/or use in random assignment.

2	**a** **S-IC.A.2**	Student makes no use of the simulated distribution.	Student answers no because the center of the simulated distribution is close to 0.50.	Student attempts to integrate information from the graph and the observed statistic, but the response is inconsistent (e.g., "0.43 is on the graph, so I believe the population is $50/50$"), or the response is too definitive (e.g., "We have proven that the population cannot be $50/50$").	Student uses the observed statistic (0.43) and the simulated distribution to address the question of whether the population is $50/50$ (e.g., "Yes, because 0.43 is far in the tail of this distribution").
	b **S-IC.A.1**	Student does not include statistical justification (e.g., "I was not surveyed," or anecdotal evidence about who pays attention).	Student says no based on the sample size being too small or that only adults were surveyed or says the task of representing the entire population is too difficult.	Student attempts to use the random sampling as justification but is concerned by the 43% being too different from 50%.	Student appeals to the use of a random sample of U.S. adults to argue that it would be appropriate to generalize the results to the larger population of U.S. adults.
	c **S-IC.B.4**	Student does not calculate a margin of error but discusses possible sampling or nonsampling errors.	Student reports a margin of error, but it is not clear how that value is obtained, or uses an incorrect method (e.g., the number 2) or claims not enough information is provided.	Student eyeballs that most of the sample proportions are within 0.3 of the mean. OR Student uses $1SD$ as the margin of error or uses $SD = 0.022$ from the first graph.	Student calculates $2 \times SD = 0.032$ as the margin of error.
	d **S-IC.B.4**	Student does not make a claim about the change in the margin of error.	Student expects the margin of error to not change or to increase with the larger sample size.	Student expects the margin of error to be smaller with the larger sample size, but student's explanation reveals an inconsistency in logic.	Student expects the margin of error to be smaller with the larger sample size.
3	**a** **S-ID.A.4**	Student does not differentiate between the two distributions.	Student picks group 2 based on the mean being closer to the middle of the graph.	Student picks group 1 and gives a statistical argument such as $\text{mean} - 3SD$ is too close to zero for group 2.	Student picks group 2 based on the symmetric, bell-shaped nature of the distribution.

	b **S-IC.B.6**	Student does not describe a misleading aspect to the claim.	Student focuses on potential sampling or nonsampling errors in the data collection.	Student references the skewed nature of group 1 and indicates that means would not be an appropriate comparison. OR Student focuses on the discrepancy between 4.88 and 5.	Student recognizes that the statement does not reflect that these means are estimates. The statement also mixes *mean* and *individual* reaction times.
	c **S-IC.B.5**	Student does not describe a simulation approach.	Student focuses on random sampling rather than random assignment, and it is not clear on how a decision of significance will be made.	Student shows an understanding of how to use the simulation to evaluate 4.88, but details of the simulation are not clear.	Student includes random reassignment of the reaction times to two groups, recomputing of the difference in sample means, and use of the simulated distribution to judge how unlikely a result of 4.88 or more extreme is to occur. (How to use the simulation results could be described in part (d).)
	d **S-IC.B.5**	Student makes no use of the simulated distribution.	Student answers no because the center of the simulated distribution is close to 0.	Student attempts to integrate information from the graph and the observed statistic, but the response is inconsistent (e.g., "4.88 is not on the graph, so there is not a significant difference").	Student uses the observed statistic (4.88) and the simulated distribution to address the question of significance (e.g., "Yes, because 4.88 is not listed on the graph").

Name _____ Date _____

1. Suppose you wanted to determine whether students who close their eyes are better able to estimate when 30 seconds have passed than students who do not close their eyes. (You ask students to tell you when to stop a stopwatch after they think 30 seconds have passed.) You find the first 50 students arriving at school one day. For those 50, you flip a coin to decide whether or not they will close their eyes during the test. Then, you compare the amounts by which each group overestimated or underestimated.

   a. Did this study make use of *random sampling*? Explain your answer by describing what purpose random sampling serves in such a study.

   No. This study used a convenience sample of the first 50 students arriving at school. A random sample would give us a better representative sample by giving all students an equal chance to be selected for the sample rather than only the early risers.

   b. Did this study make use of *random assignment*? Explain your answer by describing what purpose random assignment serves in such a study.

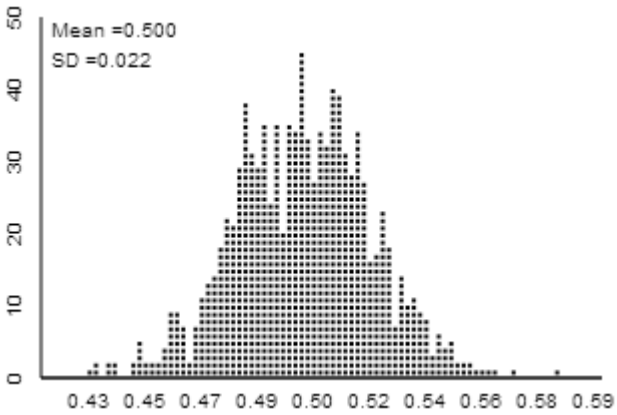
   Yes. A coin was flipped to decide in which treatment condition each student would be placed. The goal is to create two groups that are "equal" with respect to all potential confounding variables. This will allow us to potentially draw cause-and-effect conclusions between the two treatments (i.e., eyes open or closed).

   c. Would the study described above be an *observational study* or an *experimental study*? Explain how you are deciding.

   This is an experiment because the explanatory variable (i.e., eyes open or closed) was randomly imposed by the researcher. An observational study would not control which method students use but would rather let the students choose for themselves (and would not even be told of different possible strategies).

2. A Gallup poll conducted July 10–14, 2013, asked a random sample of U.S. adults: "How much attention do you pay to the nutritional information that is printed on restaurant menus or posted in restaurants, including calories and sugar and fat content?" The sample results were that 43% of the respondents said they pay a "fair amount" or a "great deal" of attention. Suppose there had been 500 people in the study.

   The following graph displays the results from $1{,}000$ random samples (each with sample size 500) from a very large population where 50% of the respondents "pay some attention" and 50% "pay little or no attention."



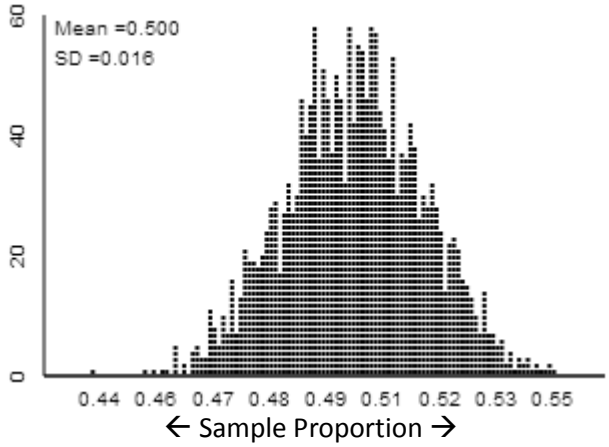
   a. Based on the simulation results above, are the sample data (43% responding "pay some attention") consistent with the simulation? In other words, do these results cause you to question whether the population is $50/50$ on this issue? Explain.

      We see very few of the simulated sample proportions around 0.43. It falls in the tail of the distribution. So, this unusual outcome would lead us to question whether the actual sample proportion came from a population that was 50/50 on the issue. It would be surprising to get a sample proportion as extreme as 0.43 if the population proportion was indeed 0.50.

   b. Do you believe it is reasonable to generalize the results from this study to all U.S. adults? Explain.

      The Gallup poll claims to have collected a random sample, so I would feel comfortable generalizing these results to U.S. adults.

Suppose Gallup plans to conduct a new poll of a random sample of 1,000 U.S. adults on an issue where the population is evenly split between two responses. The following graph displays the results from 2,000 random samples (each with sample size of 1,000) from such a population.



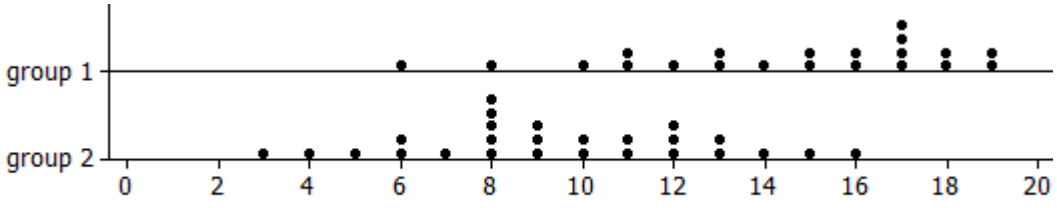
c. Based on these simulation results, estimate the expected margin of error for the Gallup poll. Explain how you developed your estimate.

The standard deviation of the simulated sample proportions is approximately 0.016. So, the approximate margin of error is two standard deviations or about 0.032.

d. Suppose the study used a sample size of 2,000 instead of 1,000. Would you expect the margin of error to be larger or smaller?

A larger sample size should lead to a smaller margin of error.

3. A randomized experiment compared the reaction time (in milliseconds) for subjects who had been sleep deprived (group 1) and subjects who had not (group 2).



Variable	N	Mean	StDev
group 1	21	14.38	3.64
group 2	26	9.50	3.33

a. Based on the above output, for which group would it be more reasonable to use a normal curve to model the reaction time distribution? Justify your choice.

Group 2. The observations for group 2 appear to be more symmetric and mound shaped. Group 1 does not appear to be symmetric.

b. The difference in means is $14.38 - 9.50 = 4.88$. One of the researchers claims that the reaction time if you are sleep deprived is 5 ms greater than the reaction time if you are not sleep deprived. Explain one reason why this claim is potentially misleading.

The researcher is generalizing from the observed difference in means. It's possible that the actual difference in means is a bit smaller than 5. The researcher is also comparing reaction times rather than reflecting the conclusion as a comparison of means.

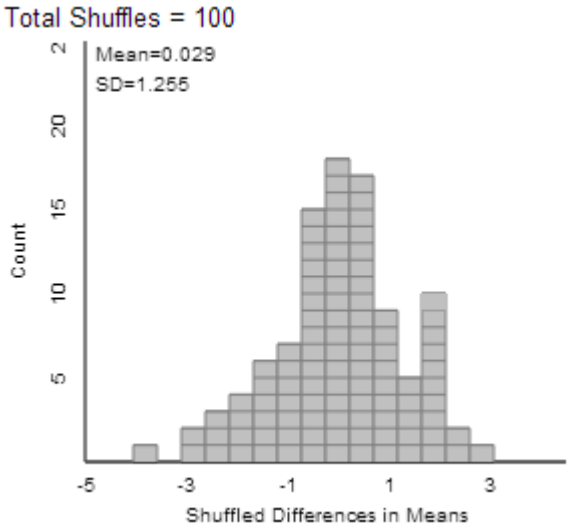
c. Describe how to carry out a simulation analysis to determine whether the mean reaction time for group 1 is significantly larger than the mean reaction time for group 2.

You could put all of the reaction times on index cards and then shuffle the cards and randomly deal them to two groups, one with 21 cards and one with 26 cards. You could also put the response values in a list and then flip a coin to put "group 1" or "group 2" by each value until you fill one of the groups.

Then, calculate the difference in the mean reaction time of the two groups.

Repeat this many times, and see how often the difference is as large as or larger than 4.88.

d. The graph below displays the results of 100 repetitions of a simulation to investigate the difference in sample means when there is no real difference in the treatment means. Use this graph to determine whether the observed mean reaction time for group 1 is significantly larger than the observed mean reaction time for group 2. Explain your reasoning.



We see that 4.88 is in the tail of the distribution. We had no simulated differences in means that were as large as 4.88. This gives us reason to believe that our observed difference came from a process where there was a genuine difference between the treatments. Therefore, we have strong evidence that the average reaction time for those who are sleep deprived is larger than for those who are not sleep deprived (for subjects similar to those in this study).