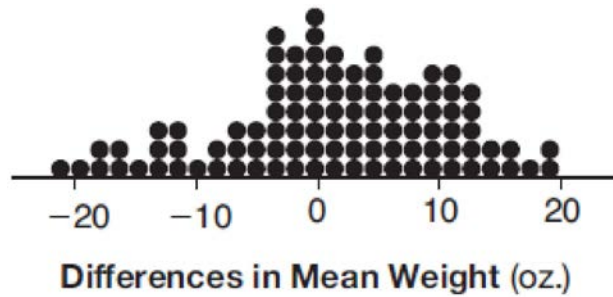


**S.IC.B.5: Analysis of Data**

- 1 Gabriel performed an experiment to see if planting 13 tomato plants in black plastic mulch leads to larger tomatoes than if 13 plants are planted without mulch. He observed that the average weight of the tomatoes from tomato plants grown in black plastic mulch was 5 ounces greater than those from the plants planted without mulch. To determine if the observed difference is statistically significant, he rerandomized the tomato groups 100 times to study these random differences in the mean weights. The output of his simulation is summarized in the dotplot below.



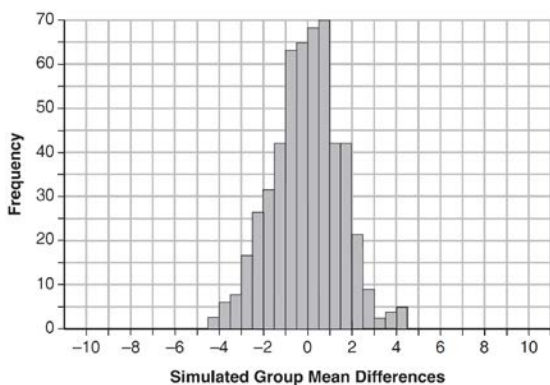
Given these results, what is an appropriate inference that can be drawn?

- |   |   |
|---|---|
| 1) There was no effect observed between the two groups.   | 3) There is strong evidence to support the hypothesis that tomatoes from plants planted in black plastic mulch are larger than those planted without mulch. |
| 2) There was an effect observed that could be due to the random assignment of plants to the groups. | 4) There is strong evidence to support the hypothesis that tomatoes from plants planted without mulch are larger than those planted in black plastic mulch. |

- 2 Seventy-two students are randomly divided into two equally-sized study groups. Each member of the first group (group 1) is to meet with a tutor after school twice each week for one hour. The second group (group 2), is given an online subscription to a tutorial account that they can access for a maximum of two hours each week. Students in both groups are given the same tests during the year. A summary of the two groups' final grades is shown below:

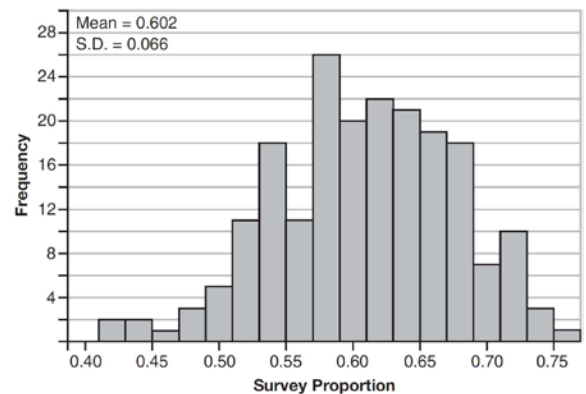
	Group 1	Group 2
$\bar{x}$	80.16	83.8
$S_x$	6.9	5.2

Calculate the mean difference in the final grades (group 1 – group 2) and explain its meaning in the context of the problem. A simulation was conducted in which the students' final grades were rerandomized 500 times. The results are shown below.



Use the simulation to determine if there is a significant difference in the final grades. Explain your answer.

- 3 Fifty-five students attending the prom were randomly selected to participate in a survey about the music choice at the prom. Sixty percent responded that a DJ would be preferred over a band. Members of the prom committee thought that the vote would have 50% for the DJ and 50% for the band. A simulation was run 200 times, each of sample size 55, based on the premise that 60% of the students would prefer a DJ. The approximate normal simulation results are shown below.

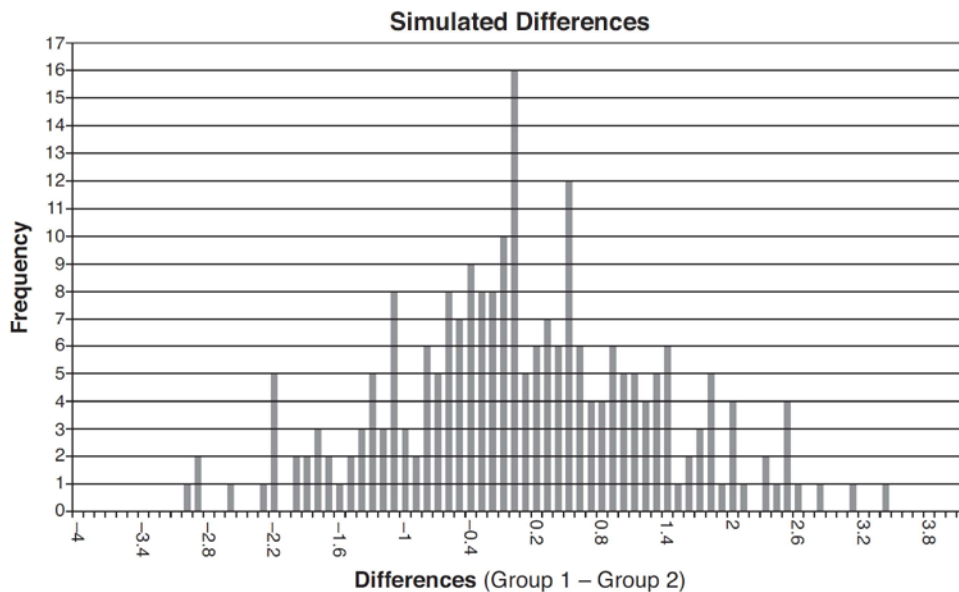


Using the results of the simulation, determine a plausible interval containing the middle 95% of the data. Round all values to the *nearest hundredth*. Members of the prom committee are concerned that a vote of all students attending the prom may produce a 50% – 50% split. Explain what statistical evidence supports this concern.

- 4 Ayva designed an experiment to determine the effect of a new energy drink on a group of 20 volunteer students. Ten students were randomly selected to form group 1 while the remaining 10 made up group 2. Each student in group 1 drank one energy drink, and each student in group 2 drank one cola drink. Ten minutes later, their times were recorded for reading the same paragraph of a novel. The results of the experiment are shown below.

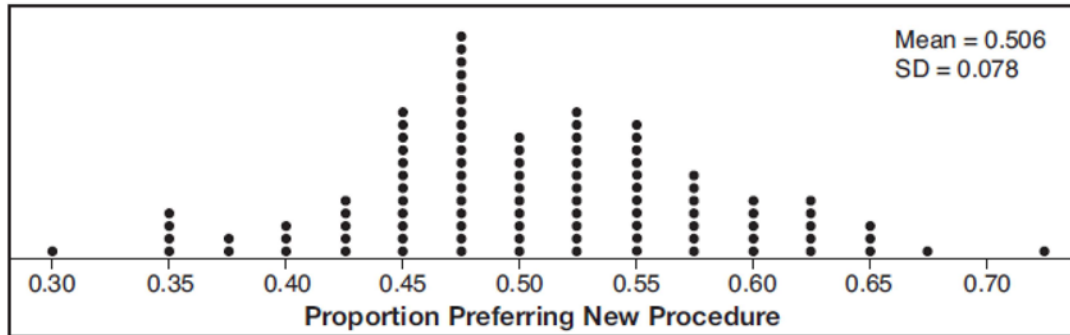
Group 1 (seconds)	Group 2 (seconds)
17.4	23.3
18.1	18.8
18.2	22.1
19.6	12.7
18.6	16.9
16.2	24.4
16.1	21.2
15.3	21.2
17.8	16.3
19.7	14.5
Mean = 17.7	Mean = 19.1

Ayva thinks drinking energy drinks makes students read faster. Using information from the experimental design or the results, explain why Ayva’s hypothesis may be *incorrect*. Using the given results, Ayva randomly mixes the 20 reading times, splits them into two groups of 10, and simulates the difference of the means 232 times.



Ayva has decided that the difference in mean reading times is not an unusual occurrence. Support her decision using the results of the simulation. Explain your reasoning.

- 5 Charlie's Automotive Dealership is considering implementing a new check-in procedure for customers who are bringing their vehicles for routine maintenance. The dealership will launch the procedure if 50% or more of the customers give the new procedure a favorable rating when compared to the current procedure. The dealership devises a simulation based on the minimal requirement that 50% of the customers prefer the new procedure. Each dot on the graph below represents the proportion of the customers who preferred the new check-in procedure, each of sample size 40, simulated 100 times.



Assume the set of data is approximately normal and the dealership wants to be 95% confident of its results. Determine an interval containing the plausible sample values for which the dealership will launch the new procedure. Round your answer to the *nearest hundredth*. Forty customers are selected randomly to undergo the new check-in procedure and the proportion of customers who prefer the new procedure is 32.5%. The dealership decides *not* to implement the new check-in procedure based on the results of the study. Use statistical evidence to explain this decision.

**S.IC.B.5: Analysis of Data****Answer Section**

1 ANS: 2 REF: 011709aai

2 ANS:

The mean difference between the students' final grades in group 1 and group 2 is  $-3.64$ . This value indicates that students who met with a tutor had a mean final grade of 3.64 points less than students who used an on-line subscription. One can infer whether this difference is due to the differences in intervention or due to which students were assigned to each group by using a simulation to rerandomize the students' final grades many (500) times. If the observed difference  $-3.64$  is the result of the assignment of students to groups alone, then a difference of  $-3.64$  or less should be observed fairly regularly in the simulation output. However, a difference of  $-3$  or less occurs in only about 2% of the rerandomizations. Therefore, it is quite unlikely that the assignment to groups alone accounts for the difference; rather, it is likely that the difference between the interventions themselves accounts for the difference between the two groups' mean final grades.

REF: fall1514aai

3 ANS:

$0.602 \pm 2 \cdot 0.066 = 0.47 - 0.73$ . Since 0.50 falls within the 95% interval, this supports the concern there may be an even split.

REF: 061635aai

4 ANS:

Some of the students who did not drink energy drinks read faster than those who did drink energy drinks.

$17.7 - 19.1 = -1.4$  Differences of -1.4 and less occur  $\frac{25}{232}$  or about 10% of the time, so the difference is not unusual.

REF: 081636aai

5 ANS:

$0.506 \pm 2 \cdot 0.078 = 0.35 - 0.66$ . The 32.5% value falls below the 95% confidence level.

REF: 061736aai