S.IC.B.5: Analysis of Data

1 A group of students was trying to determine the proportion of candies in a bag that are blue. The company claims that 24% of candies in bags are blue. A simulation was run 100 times with a sample size of 50, based on the premise that 24% of the candies are blue. The approximately normal results of the simulation are shown in the dot plot below.

The simulation results in a mean of 0.254 and a standard deviation of 0.060. Based on this simulation, what is a plausible interval containing the middle 95% of the data?

1) $(0.194, 0.314)$ 3) $(-0.448, 0.568)$

2) $(0.134, 0.374)$ 4) $(0.254, 0.374)$

2 The J& B candy company claims that 45% of the candies it produces are blue, 30% are brown, and 25% are yellow. Each bag holds 65 candies. A simulation was run 200 times, each of sample size 65, based on the premise that 45% of the candies are blue. The results of the simulation are shown below.

Bonnie purchased a bag of J& B's candy and counted 24 blue candies. What inference can be made regarding a bag of J& B's with only 24 blue candies?

1) The company is not meeting their production standard.

2) Bonnie's bag was a rarity and the company should not be concerned.

3) The company should change their claim to 37% blue candies are produced.

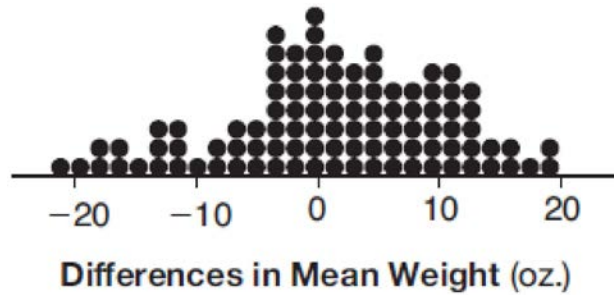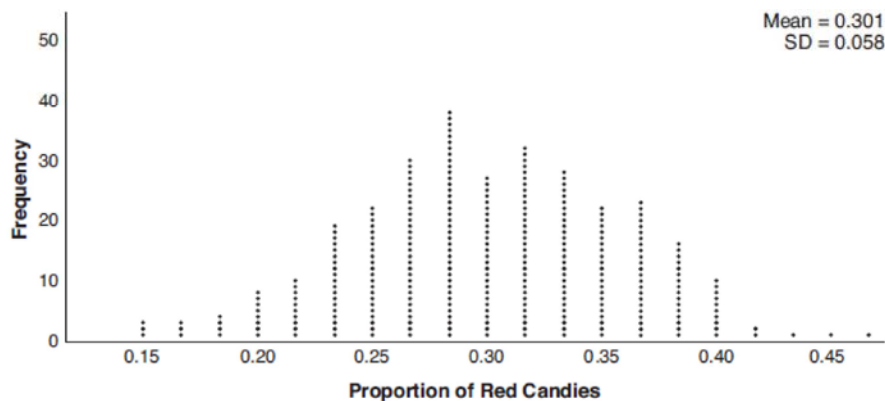4) Bonnie's bag is within the middle 95% of the simulated data supporting the company's claim.

3   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.

Differences in Mean Weight (oz.)

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

1) There was no effect observed between the two groups.

2) There was an effect observed that could be due to the random assignment of plants to the 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.

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.

4   Mary bought a pack of candy. The manufacturer claims that 30% of the candies manufactured are red. In her pack, 14 of the 60 candies are red. She ran a simulation of 300 samples, assuming the manufacturer is correct. The results are shown below.

Mean = 0.301
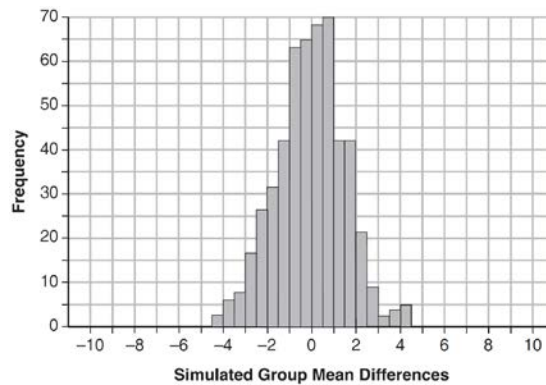SD = 0.058

Proportion of Red Candies

Based on the simulation, determine the middle 95% of plausible values that the proportion of red candies in a pack is within. Based on the simulation, is it unusual that Mary's pack had 14 red candies out of a total of 60? Explain.

2

5   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:
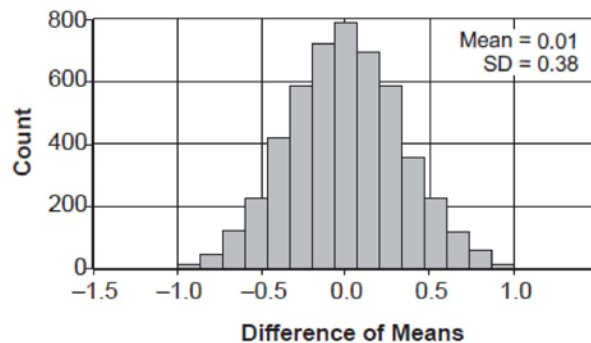
	Group 1	Group 2
$\bar{x}$	80.16	83.8
S_x	6.9	5.2

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.
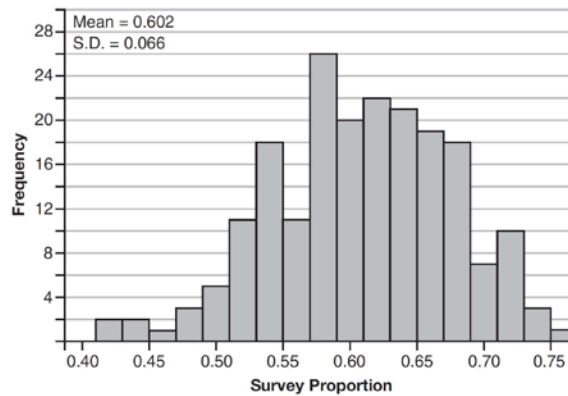

Simulated Group Mean Differences

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

6   Two classes of students were entered into an experiment to see whether using an interactive whiteboard leads to better grades. It was observed that the mean grade of students in the class with the interactive whiteboard was 0.6 points higher than the class without it. To determine if the observed difference is statistically significant, the classes were rerandomized 5000 times to study these random differences in the mean grades. The output of the simulation is summarized in the histogram below.

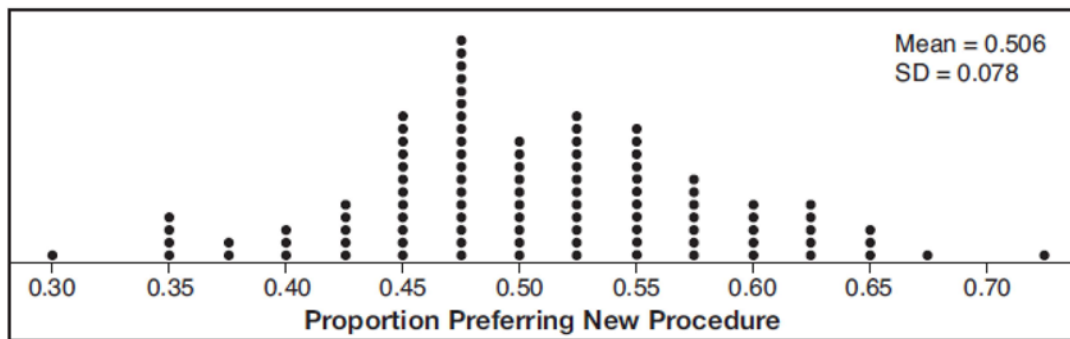Mean = 0.01
SD = 0.38

Difference of Means

Determine an interval containing the middle 95% of the simulation results. Round your answer to the *nearest hundredth*. Does the interval indicate that the difference between the classes' grades is significant? Explain.

7 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.

8 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.
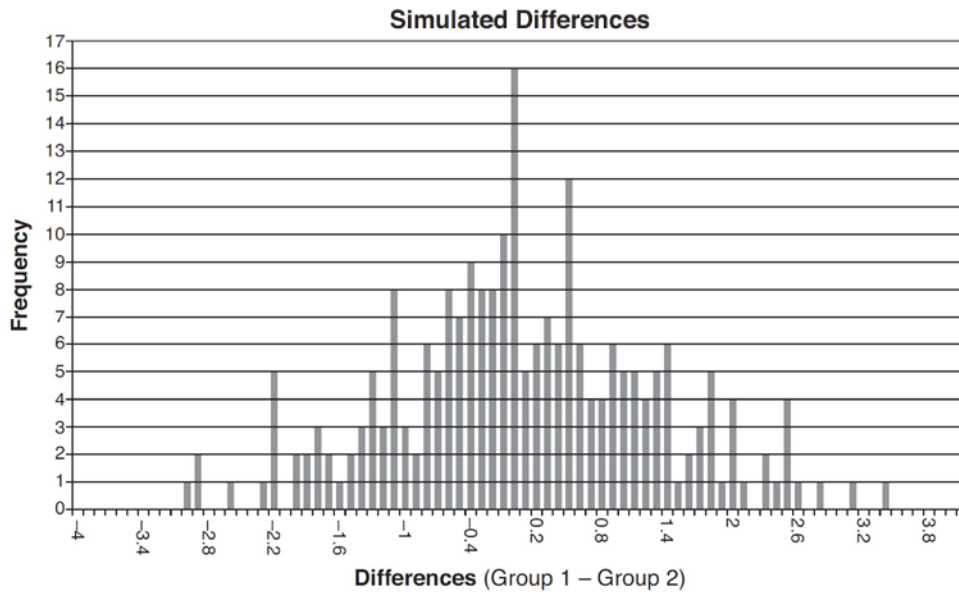

Assume the set of data is approximately normal and the dealership wants to be 95% confident of its results. Determine an interval containing the plausible sample values for which the dealership will launch the new procedure. Round your answer to the *nearest hundredth*. Forty customers are selected randomly 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.

Name: _____

9 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.

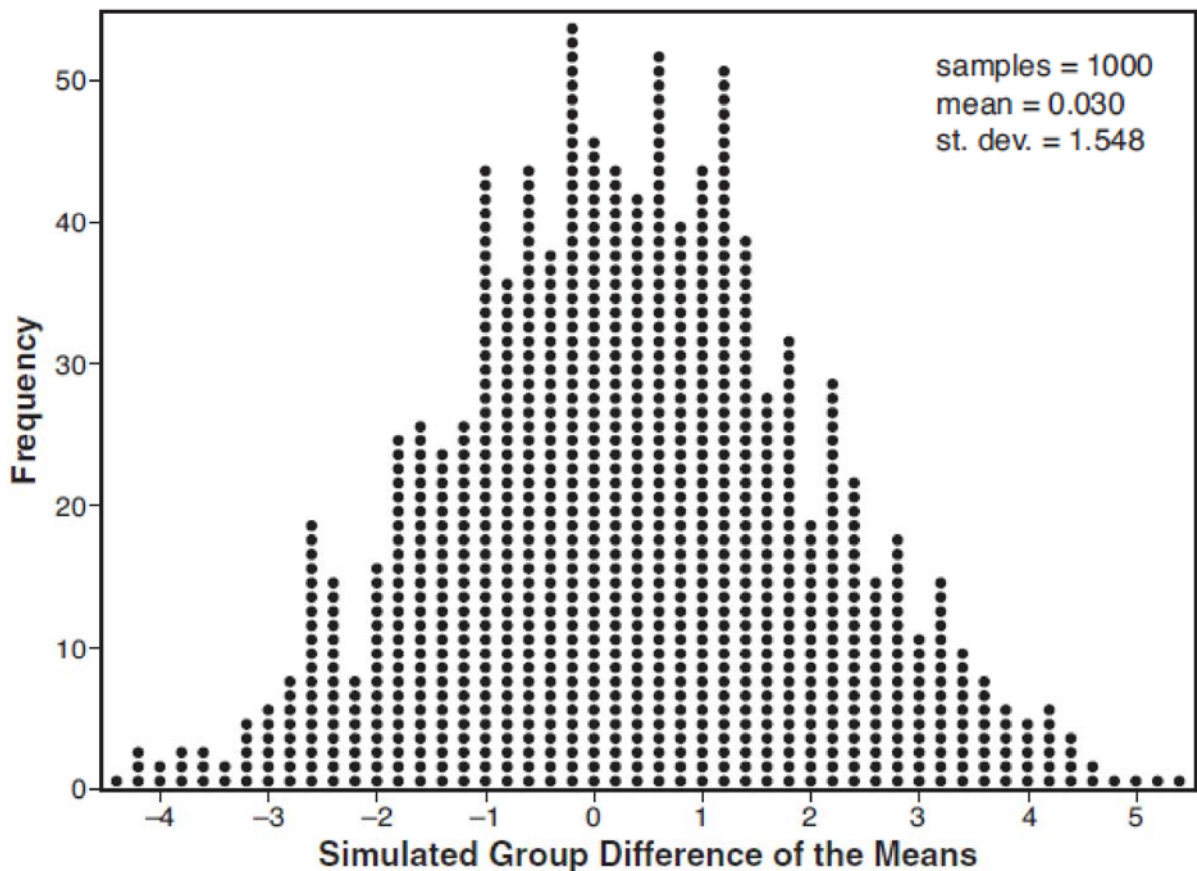**Simulated Differences**

Differences (Group 1 – Group 2)

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.

10 Joseph was curious to determine if scent improves memory. A test was created where better memory is indicated by higher test scores. A controlled experiment was performed where one group was given the test on scented paper and the other group was given the test on unscented paper. The summary statistics from the experiment are given below.

	Scented Paper	Unscented Paper
$\bar{x}$	23	18
s_x	2.898	2.408

Calculate the difference in means in the experimental test grades (scented -unscented). A simulation was conducted in which the subjects' scores were rerandomized into two groups 1000 times. The differences of the group means were calculated each time. The results are shown below.
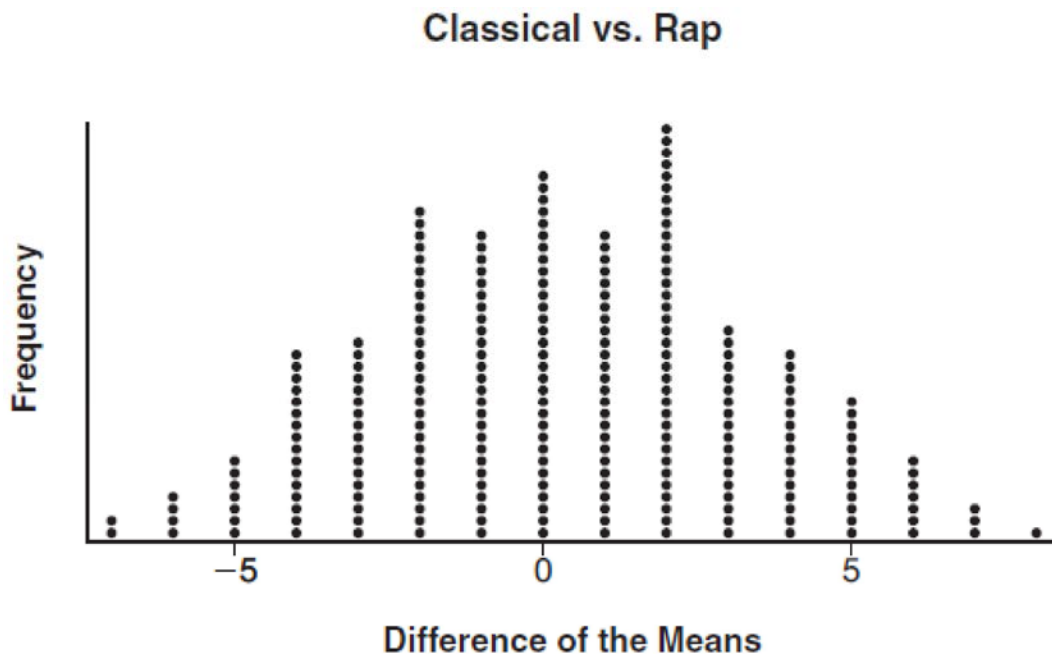

Use the simulation results to determine the interval representing the middle 95% of the difference in means, to the *nearest hundredth*. Is the difference in means in Joseph's experiment statistically significant based on the simulation? Explain.

11 To determine if the type of music played while taking a quiz has a relationship to results, 16 students were randomly assigned to either a room softly playing classical music or a room softly playing rap music. The results on the quiz were as follows:

Classical: 74, 83, 77, 77, 84, 82, 90, 89
Rap: 77, 80, 78, 74, 69, 72, 78, 69

John correctly rounded the difference of the means of his experimental groups as 7. How did John obtain this value and what does it represent in the given context? Justify your answer. To determine if there is any significance in this value, John rerandomized the 16 scores into two groups of 8, calculated the difference of the means, and simulated this process 250 times as shown below.

Classical vs. Rap

Difference of the Means

Does the simulation support the theory that there may be a significant difference in quiz scores? Explain.

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

1 ANS: 2
$0.254 \pm 2(0.060) \rightarrow (0.134, 0.374)$

REF: 061913aii

2 ANS: 4   REF: 012014aii
3 ANS: 2   REF: 011709aii
4 ANS:

$0.301 \pm 2(0.058) \rightarrow 0.185 - 0.417 \quad \frac{14}{60} \approx 0.23.$ It is not unusual because 0.23 falls within this interval.

REF: 081935aii

5 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: fall1514aii

6 ANS:
$0.01 \pm 2 \cdot 0.38 = -0.75 - 0.77.$ No, since 0.6 falls within the 95% interval.

REF: 082336aii

7 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: 061635aii

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

REF: 061736aii

9 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: 081636aii

10 ANS:

$23 - 18 = 5$, $\bar{x} \pm 2\sigma = -3.07 - 3.13$, Yes, a difference of 5 or more occurred three times out of a thousand, which is statistically significant.

REF: 061834aii

11 ANS:

John found the means of the scores of the two rooms and subtracted the means. The mean score for the classical room was 7 higher than the rap room (82-75). Yes, there is less than a 5% chance this difference occurring due to random chance. It is likely the difference was due to the music.

REF: 081836aii