Name Key

HW 14-10

1. In a random sample of 250 men in the United States, age 21 or older, 139 are married. The graph below simulated samples of 250 men, 200 times, assuming that 139 of the men are married.

Jan. 2018 Regents

a) Based on the simulation, create an interval in which the middle 95% of the number of married men may fall. Round your answer to the nearest integer.

$$CI = 138.905 \pm 2(7.95)$$
$$= (123, 155)$$

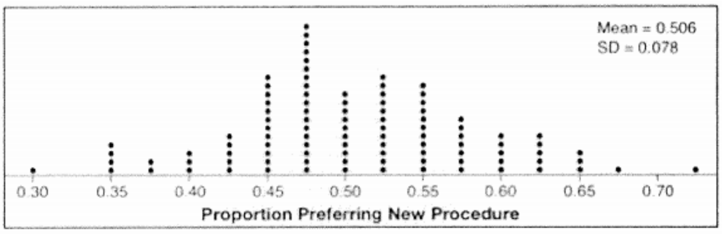
b) A study claims "50 percent of men 21 and older in the United States are married." Do your results from part a contradict this claim? Explain.

50% men 21+ = Half of 250 = 125

125 is within the CI, so my results verify part a), not contradict

June 2017 Regents

2. 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.

$$CI = .506 \pm 2(.078)$$
$$= (.35, .66)$$

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.

Since there are 32.5% of customers that prefer the new procedure, this is below the 95% confidence interval of the population proportion.

c. Compute the probability of getting a "Diff" value as extreme or more extreme than the "Diff" value you obtained in part a.

$$\frac{1}{100} = .01$$

d. Make a conclusion in context based on your probability you found. $\rightarrow \therefore$ stat. sign.

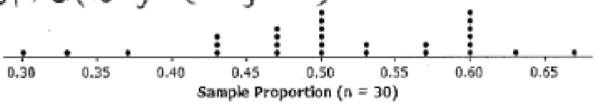
Since there's only a 1% chance of getting a "Diff" value of -1.4 or lower if the two changes in pain scores were the same, we have evidence that group ~~A~~ has a smaller change in pain score.

Name: Key

Unit 14 Review
Algebra II CC

1. A group of eleventh graders wanted to estimate the population proportion of students in their high school who have their driver's licenses. Each student selected a different random sample of 30 students from the high school and calculated the proportion that have their driver's licenses. The dot plot below shows the sampling distribution. This distribution has a mean of 0.51 and a standard deviation of 0.09.

$CI = .51 + 2(.09) = (.33, .69)$



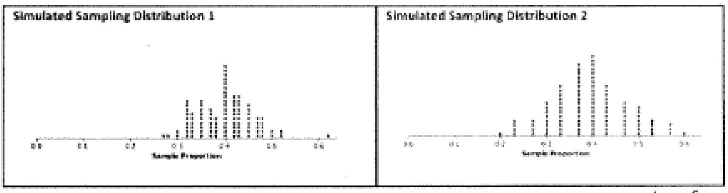
a. Does a sample proportion of 0.30 have a high or low probability of coming from a population proportion of 0.51? Explain your reasoning.

Low probability b/c it's far away from .51 ($\frac{1}{28} \approx .04$ chance of happening). Most likely .30 came from a diff. pop. prop. than 0.51. Also, .3 is outside of CI ∴ low prob.

b. Does a sample proportion of 0.47 have a high or low probability of coming from a population proportion of 0.51? Explain your reasoning.

High probability b/c it's close to 0.51 ($\frac{10}{28} \approx .36$ chance of happening). Most likely 0.47 came from a pop. prop. of 0.51. Also, .47 is inside of CI ∴ high prob.

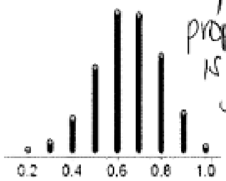
2. Below are two sampling distributions for the sample proportion of smokers in random samples from across the nation. One sample came from a random sample of size 55 while another one came from a random sample of size 102. Which sampling distribution corresponds to the size 102? Explain your choice.



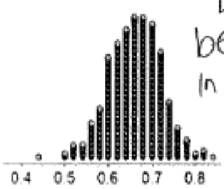
Sampling distribution 1 because it's less spread out/variable. Larger sample size = smaller spread/variability

3. Anyone who plays or watches sports has heard of the "home field advantage." Teams tend to win more often when they play at home. Or do they? If there were no home field advantage, the home teams would win about half of all games played. In the 2007 Major League Baseball season, there were 2431 regular-season games. (Tied at the end of the regular season, the Colorado Rockies and San Diego Padres played an extra game to determine who won the Wild Card playoff spot.) It turns out that the home team won 1319 of the 2431 games, or 54.26% of the time.

   a. Based on the simulation results below, is the sample data (54.26% or .5426) consistent with the simulation? Explain. Yes, much of the simulated sampling distr proportions are around 0.6 and 0.5426 is close to 0.6, it wouldn't be surprising to get a sample prop. as extreme as 0.5426 if pop. prop. is 0.60.



   b. Do you think the below simulation is from a larger or smaller number of games than the 2431 given above? Explain. Larger # of games than 2431 because there's less variability (spread) in this simulation than in a.) Implying there was a larger sample size.



   c. Given that the mean of the above distribution is 0.65 and the standard deviation is 0.075, find:

      i. the margin of error for home team wins. $ME = 2(.075) = .15$

      ii. the confidence interval for home team wins. $CI = .65 \pm .15 = (.5, .8)$

      iii. Would a winning home team win prop. of 0.9 be stat. sign? Explain. Yes, since 0.9 is not w/in the CI.

4. I randomly selected 10 dogs in shelters across Syracuse. Their weights are

12.5, 20.1, 30.4, 7.2, 25.6, 67, 50.3, 42.7, 11.7, 40.5

I want to use this sample to estimate the average weight of all dogs in shelters in Syracuse. The mean of my sample is 30.8 pounds with a standard deviation of 6.1 pounds. Create a 95% confidence interval for the average weight of all dogs in shelters in Syracuse.

$CI = 30.8 \pm 2(6.1) = (18.6, 43)$

5. The temperatures in Baldwinsville, NY were recorded for 200 days out of a year and are normally distributed, with a mean of 52°F and a standard deviation of 2.4°F. Round all answers to the nearest tenth.

   a. What's the margin of error? $ME = 2(2.4) = 4.8$

   b. Create a 95% confidence interval for the average temperature in Baldwinsville, NY.

   $CI = 52 \pm 4.8 = (47.2, 56.8)$

6. 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 $\bar{x} = 82$

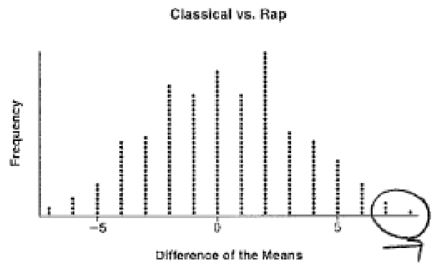
   Rap: 77, 80, 78, 74, 69, 72, 78, 69 $\bar{x} = 74.625$

   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.

   $Diff = 82 - 74.625 = 7.375 \approx 7$

   John found the mean for each group and then subtracted them. 7 represents the classical group mean was 7 higher than the rap group.

   To determine if there is any significance in this value, John rerandomized the 16 scores into two group of 8, calculated the difference of the means, and simulated this process 250 times as shown below.



   $\frac{4}{250} = .016$

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

   Yes, because there's less than a 5% chance of getting a diff. of 7 or more (1.6%) there may be a sign. diff. in quiz scores.

7. The mean length of a sample of 120 lizards is 88 cm with a standard deviation of 2.19 cm. Round your final interval to the nearest tenth of a cm, create a 95% confidence interval for the population mean.

$$CI = 88 \pm 2(2.19) = (83.6, 92.4)$$

8. A pumpkin farmer wants to determine whether the presence of Nutrient Z in the soil promotes the growth of larger pumpkins. The farmer grows half of the pumpkins in typical soil and uses Nutrient Z to treat the soil used to grow the other half of the pumpkin crop. The farmer weighs the pumpkins at harvest time and records the data. Based on the tables below, which of the farmer's experiments has data with the greatest Diff value?

Diff = .5

a.

Treatment Group	8.4	9.2	9.7	10.1	10.5	10.9	12.5	13.0
Control Group	8.5	8.9	9.6	9.9	10.3	10.5	10.9	11.7

$\bar{x} = 10.5375$
$\bar{x} = 10.0375$

Diff = .75

b.

Treatment Group	8.9	9.4	9.9	10.1	10.6	11.2	12.4	13.0
Control Group	8.2	8.9	9.6	9.7	10.2	10.8	11.0	11.1

$\bar{x} = 10.6875$
$\bar{x} = 9.9375$

Diff = .25

c.

Treatment Group	8.4	8.9	9.6	10.1	10.3	10.8	12.6	12.9
Control Group	8.5	9.2	9.7	9.9	10.5	10.6	11.5	11.7

$\bar{x} = 10.45$
$\bar{x} = 10.2$

Diff = 1

d.

Treatment Group	8.8	9.2	9.8	10.3	10.7	11.0	12.7	13.6
Control Group	8.0	8.2	9.1	9.5	9.8	10.9	11.2	11.4

$\bar{x} = 10.7625$
$\bar{x} = 9.7625$

9. Researchers wanted to know if a new workout routine would help athletes run faster in their 400 meter run times. 100 people are randomly divided into two equally-sized groups. *Group A* trains with the new workout routine. *Group B* trains without the new workout routine.
A summary of the two groups' 400 meter run times is shown below:

	Group A	Group B
$\bar{x}$	51.5 sec	60.75 sec
S_x	4.07 sec	2.92 sec

Calculate the mean difference in the final grades (Group A - Group B) and explain its meaning in the context of the problem. Diff = 51.5 - 60.75 = -9.25

The athletes that trained with the new workout routine were 9.25 sec. faster than those who did not train with the new workout routine.

10. Lincoln conducts an experiment to determine if studying for an hour impacts the number of questions correct on the learner's permit (driving) test. He randomly assigns the students in his homeroom who have not yet taken the test to two groups. Group A agrees to study for an hour before taking the permit test. Group B agrees to take the test without studying. The number of questions each student answered correctly is recorded in the table below. Note: There are 20 total questions on the test.

Group A	
Student	Number Correct
1	16
2	19
3	20
4	17
5	15
6	19
7	18
8	14
9	14
10	20

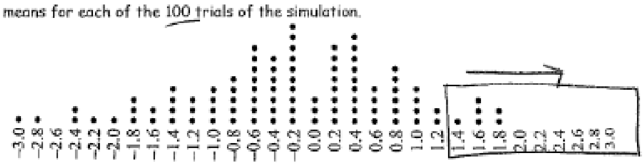
Group B	
Student	Number Correct
11	14
12	17
13	18
14	15
15	19
16	14
17	12
18	16
19	18
20	15

a. Find the sample mean for each group and the difference between the sample means.

$\bar{X}_A = 17.2$ Diff $= 17.2 - 15.8$

$\bar{X}_B = 15.8$ $= 1.4$

To decide if the difference between the sample means is significant, Lincoln takes the 20 data values for the number of questions correct on the test, and randomly assigns them to two groups. He finds the mean of the two groups and their difference. Lincoln repeats this process 100 times. The dot plot below displays the difference between the means for each of the 100 trials of the simulation.



$\frac{6}{100} = .06$

b. Based on the dot plot, is the difference in sample means between Group A and Group B significant? Why or why not? Since there's a 6% chance of getting a diff of 1.4 or higher, the difference between Group A + B are not significant (not < 5% chance).

Note: Since 6% is close to 5%, some may say results ARE stat. sign.

c. Lincoln still believes that there is a significant difference between the group that studied and the one that did not. If he wants to redo the experiment, what could he do to increase his chances of finding a significant difference if there is one?

Increase the # of students in each group (Increase sample size)