

1. To determine if people's preference in dogs had changed in the recent years, organizers of a local dog show asked people who attended the show to indicate which breed was their favorite. This information was compiled by dog breed and gender of the people who responded. The table summarizes the responses.

	Female	Male	Total
Yorkshire Terrier	71	59	132
Dachshund	49	47	96
Golden Retriever	58	33	91
Labrador	37	41	78
Dalmatian	45	28	73
Other	86	67	153
Total	348	275	623

- (a) Identify the variables and tell whether each is categorical or quantitative.
- (b) What percent of the responses were from males who favor Labradors?
- (c) What percent of the male responses favor Labradors?
- (d) What percent of the people who choose Labradors were males?
2. On Monday, a class of students took a big test and the highest score was 86. The next day, a student who had been absent made up the test, scoring 98. Indicate whether adding that student's score to the rest of the data made each of the summary statistics increase, decrease, or stay about the same.
- (a) Mean
- (b) Median
- (c) Range
- (d) IQR
- (e) Standard Deviation

3. The five number summary for the weights (in pounds) of fish caught in a bass tournament is:

Min	Q1	Median	Q3	Max
2.3	2.8	3.0	3.2	4.5

- (a) Would you expect the mean weight of all fish caught to be higher or lower than the median. Explain.
- (b) You caught 3 bass weighing 2.3, 3.9, and 4.2 pounds. Were any of your fish outliers? Explain.
4. Although most of us buy milk by the quart or gallon, farmers measure daily production in pounds. Guernsey cows average 39 pounds of milk a day with a standard deviation of 8 pounds. For Jerseys the mean daily production is 43 pounds with a standard deviation of 5 pounds. When being shown at a state fair a champion Guernsey and a champion Jersey each gave 54 pounds of milk. Which cow's milk production was more remarkable? Explain.
5. The distribution of waiting times at the student health center is bell-shaped with a mean of 11 minutes and a standard deviation of 3.
- (a) Give an interval that is likely to contain about 95% of wait times.
- (b) Find and interpret the z-score of someone who waits 5 minutes. Round your z-score to two decimal places. Be sure to specifically indicate if a wait time of 5 minutes is unusual.

6. A random sample of 500 male athletes from all NCAA universities across the US was selected and each player was tested for use of steroids. Sport was also recorded.
- (a) Suppose researchers find strong evidence that steroid use (recorded as “yes” or “no”) is more prevalent in some sports than others. What is the scope of inference for the study?
 - (b) If we look at just two sports for men – basketball and football – what hypotheses should we test to see if the true proportions of steroid users differ between the two sports?
 - (c) Assume a hypothesis test was conducted in Statkey.
 - In the context of the problem, explain how a dot is created.
 - How is the value computed for the dot (in the context of the problem)?
 - (d) Where would the distribution be centered? Explain fully.
7. 50 randomly selected athletes who had suffered sprained ankles were randomly assigned to one of two exercise programs. The first group returned to sports activities in a mean of 38.7 days. The other group returned to sports in mean of 18.3 days. Why is it important to assign the athletes to the two treatments randomly?

8. Suppose you are hired by a local pizza company “C” and your job is to see if your company can make the advertising claim: “Our large size pizzas are larger in diameter than those of Dominoes”. We’ll use subscripts C and D for the two companies, and the null hypothesis is that the two true mean diameters are equal.

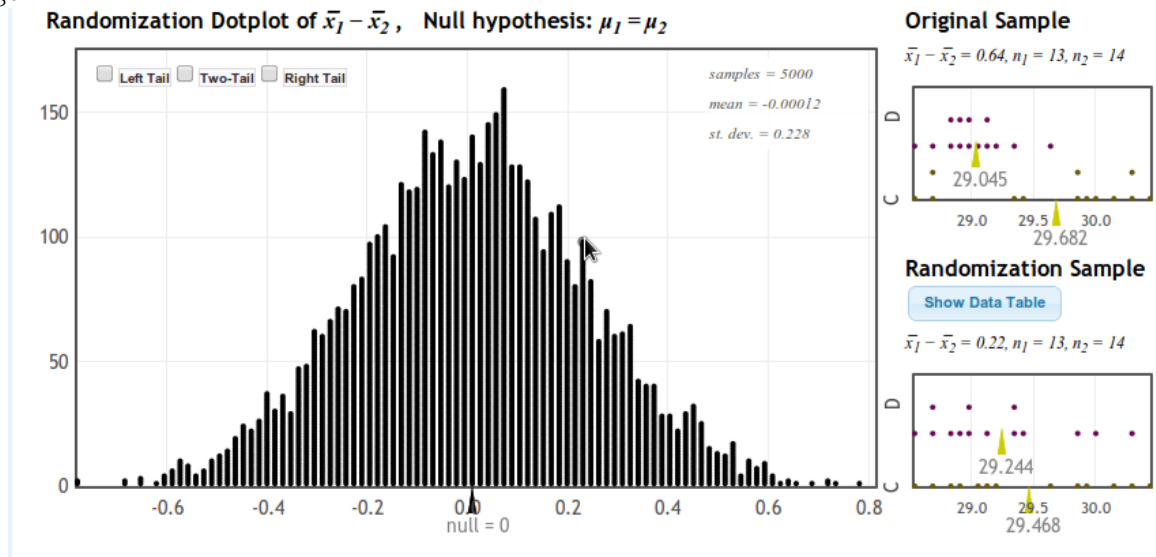
(a) The alternative hypothesis should be: (4 pts)

- A. $\mu_C > \mu_D$ B. $\mu_C < \mu_D$ C. $\mu_C \neq \mu_D$
 D. $\bar{x}_C > \bar{x}_D$ E. $\bar{x}_C < \bar{x}_D$ F. $\bar{x}_C \neq \bar{x}_D$

(b) Explain what a Type I error would mean in the context of the problem.

(c) Explain what a Type II error would mean in the context of the problem.

A random sample of 13 company C large Hawaiian pizzas is selected and measured and the average diameter is 29.68 cm. A random sample of 14 large Hawaiian pizzas from the local Dominoes is measured and the average diameter is 29.045 cm. We load the data into StatKey and create 5000 randomization samples to get the plot on the next page.



(d) Explain why the distribution is centered at zero.

(e) Notice that the pointer arrow is over one of the dots in the plot.

- In the context of the problem, explain how a dot is created.

- How is the value computed for the dot (in the context of the problem)?

(f) What is the value of our observed result (statistic)?

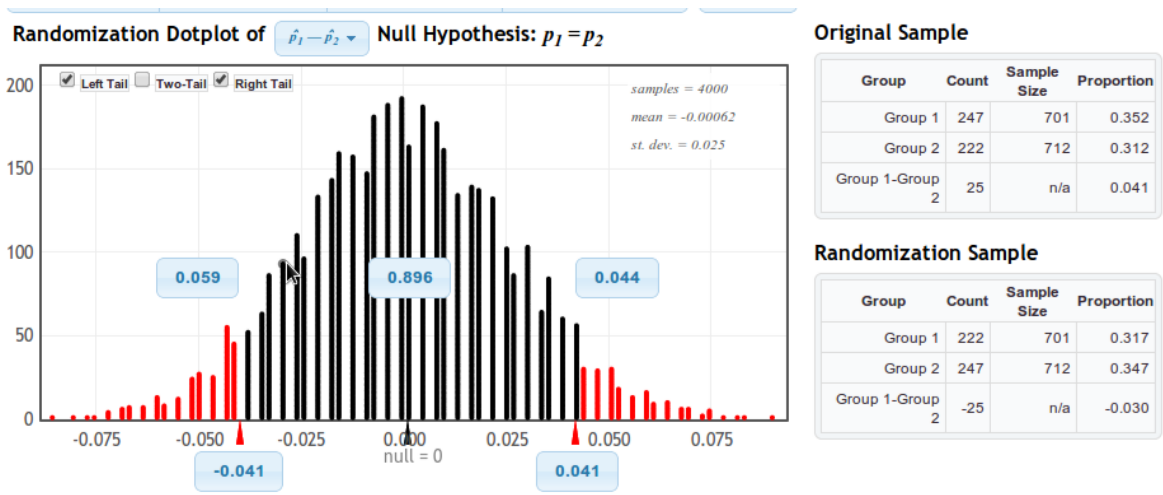
(g) There are 6 dots as or more extreme than the observed result in the right tail and 5 points this extreme or more so in the left tail. Using your alternative hypothesis from part 8a, find the p-value.

(h) Can we conclude the company is the cause of the difference in pizza size? Explain.

9. A study was done at a major medical center in Baltimore on completion rates for a series of three vaccinations needed to protect young women from HPV, human papilloma virus. Of 1413 women who received the first shot, only 469, or 33% completed the sequence within one year. The data were collected by sifting through a convenience sample of records and women were categorized as younger (ages 11 to 17. We use “Y” to label the 701 women in this group) or older (18 to 26. We use “O” to label these 712 women). Researchers want to know if there is any difference in completion rates for the two age groups.

(a) What is the alternative hypothesis?

- | | | |
|----------------------------|----------------------------|-------------------------------|
| A. $p_y > p_o$ | B. $p_y < p_o$ | C. $p_y \neq p_o$ |
| D. $\hat{p}_y > \hat{p}_o$ | E. $\hat{p}_y < \hat{p}_o$ | F. $\hat{p}_y \neq \hat{p}_o$ |



- (b) What is the observed result?
- (c) How unusual is the observed result under the null hypothesis if the alternative does not specify which proportion is expected to be larger?
- (d) Give your decision and conclusion in the context of the problem using appropriate scope of inference at the $\alpha = 0.05$ level.

10. Fill in the table below.

Variable(s)	Estimating	Parameter	Statistic	H_0	Center of ?
1 Quant. 1 Quant. 1 Quant.					Sampling Distribution Bootstrap Resampling Distribution Randomization Distribution
1 Cat. 1 Cat. 1 Cat.					Sampling Distribution Bootstrap Resampling Distribution Randomization Distribution
2 Cat. 2 Cat. 2 Cat.					Sampling Distribution Bootstrap Resampling Distribution Randomization Distribution
1 of Each 1 of Each 1 of Each					Sampling Distribution Bootstrap Resampling Distribution Randomization Distribution

11. How much time do the Ottawa Senators spend in the penalty box? Researchers took a random sample of 24 Senators' games and recorded the number of penalty minutes for each to answer this question.
- (a) Should the researchers use a hypothesis test, confidence interval, or neither to answer their question?
- (b) The sample had a mean = 49.583 and a standard deviation of 49.138. If a bootstrap distribution were created in StatKey,
- Where would the distribution be centered? Explain.
 - In the context of the problem, explain how one dot would be created.
 - How is the value computed for one dot (in the context of the problem)?
- (c) A 95% confidence interval was found to be (32.2, 70.1). Interpret the interval in the context of the problem.
- (d) What does it mean to be "95% confident" in the interval?

12. In a 2010 study, 23 subjects ages 18 to 31 were photographed twice, once after a good night's sleep and once after being kept awake for 31 hours, the order of which was randomized. Hair, make-up, clothing, and lighting were the same for both photographs. Observers then rated the photographs for attractiveness. The researcher's report in the *British Medical Journal* that "Our findings show that sleep-deprived people appear significantly less attractive compared with when they are well rested."
- (a) What is the explanatory variable? What is the response variable?
 - (b) What type of study is this? Observational study, randomized comparative experiment, or matched pairs experiment?
 - (c) Define the parameter of interest in words and notation.
 - (d) Can we conclude that sleep deprivation **causes** people to look less attractive? Why or why not?
 - (e) If a confidence interval were created for the parameter of interest (at the appropriate level of the significance level), would 0 be included in the interval? Why or why not?
 - (f) To create that confidence interval, a bootstrap distribution would be needed. What method would you use to create one bootstrap resample?
 - A. Combine the groups, draw with replacement from the pile, and deal into two new piles.
 - B. Combine the groups, the shuffle and deal into two new piles.
 - C. Draw with replacement from the list of differences.
 - D. Draw with replacement within each group.