11.6 Making Inferences from Experiments

Essential Question How can you test a hypothesis about an experiment?

EXPLORATION 1 Resampling Data

Work with a partner. A randomized comparative experiment tests whether water with dissolved calcium affects the yields of yellow squash plants. The table shows the results.

Yield (kilograms)	
Control Group	Treatment Group
1.0	1.1
1.2	1.3
1.5	1.4
0.9	1.2
1.1	1.0
1.4	1.7
0.8	1.8
0.9	1.1
1.3	1.1
1.6	1.8

a. Find the mean yield of the control group and the mean yield of the treatment group. Then find the difference of the two means. Record the results.

b. Write each yield measurement from the table on an equal-sized piece of paper. Place the pieces of paper in a bag, shake, and randomly choose 10 pieces of paper. Call this the "control" group, and call the 10 pieces in the bag the "treatment" group. Then repeat part (a) and return the pieces to the bag. Perform this resampling experiment five times.

c. How does the difference in the means of the control and treatment groups compare with the differences resulting from chance?

EXPLORATION 2 Evaluating Results

Work as a class. To conclude that the treatment is responsible for the difference in yield, you need strong evidence to reject the hypothesis:

Water dissolved in calcium has no effect on the yields of yellow squash plants.

To evaluate this hypothesis, compare the experimental difference of means with the resampling differences.

a. Collect all the resampling differences of means found in Exploration 1(b) for the whole class and display these values in a histogram.

b. Draw a vertical line on your class histogram to represent the experimental difference of means found in Exploration 1(a).

c. Where on the histogram should the experimental difference of means lie to give evidence for rejecting the hypothesis?

d. Is your class able to reject the hypothesis? Explain your reasoning.

MODELING WITH MATHEMATICS

To be proficient in math, you need to identify important quantities in a practical situation, map their relationships using such tools as diagrams and graphs, and analyze those relationships mathematically to draw conclusions.

Communicate Your Answer

3. How can you test a hypothesis about an experiment?

4. The randomized comparative experiment described in Exploration 1 is replicated and the results are shown in the table. Repeat Explorations 1 and 2 using this data set. Explain any differences in your answers.

	Yield (kilograms)									
Control Group	0.9	0.9	1.4	0.6	1.0	1.1	0.7	0.6	1.2	1.3
Treatment Group	1.0	1.2	1.2	1.3	1.0	1.8	1.7	1.2	1.0	1.9

11.6 Lesson

Core Vocabulary

Previous
randomized comparative experiment
control group
treatment group
mean
dot plot
outlier
simulation
hypothesis

What You Will Learn

- Organize data from an experiment with two samples.
- Resample data using a simulation to analyze a hypothesis.
- Make inferences about a treatment.

Experiments with Two Samples

In this lesson, you will compare data from two samples in an experiment to make inferences about a treatment using a method called *resampling*. Before learning about this method, consider the experiment described in Example 1.

EXAMPLE 1 Organizing Data from an Experiment

A randomized comparative experiment tests whether a soil supplement affects the total yield (in kilograms) of cherry tomato plants. The control group has 10 plants and the treatment group, which receives the soil supplement, has 10 plants. The table shows the results.

	Total Yield of Tomato Plants (kilograms)									
Control Group	1.2	1.3	0.9	1.4	2.0	1.2	0.7	1.9	1.4	1.7
Treatment Group	1.4	0.9	1.5	1.8	1.6	1.8	2.4	1.9	1.9	1.7

a. Find the mean yield of the control group, $\bar{x}_{\text{control}}$.

b. Find the mean yield of the treatment group, $\bar{x}_{\text{treatment}}$.

c. Find the experimental difference of the means, $\bar{x}_{\text{treatment}} - \bar{x}_{\text{control}}$.

d. Display the data in a double dot plot.

e. What can you conclude?

SOLUTION

a. $\bar{x}_{\text{control}} = \dfrac{1.2 + 1.3 + 0.9 + 1.4 + 2.0 + 1.2 + 0.7 + 1.9 + 1.4 + 1.7}{10} = \dfrac{13.7}{10} = 1.37$

The mean yield of the control group is 1.37 kilograms.

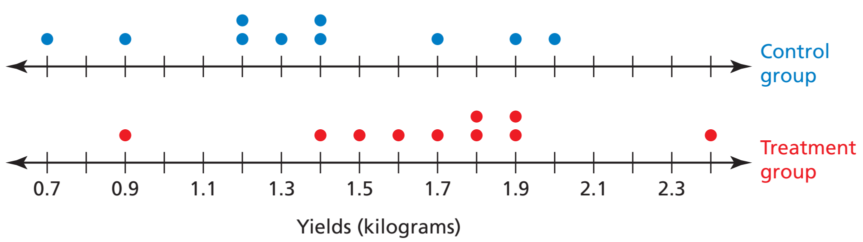
b. $\bar{x}_{\text{treatment}} = \dfrac{1.4 + 0.9 + 1.5 + 1.8 + 1.6 + 1.8 + 2.4 + 1.9 + 1.9 + 1.7}{10} = \dfrac{16.9}{10} = 1.69$

The mean yield of the treatment group is 1.69 kilograms.

c. $\bar{x}_{\text{treatment}} - \bar{x}_{\text{control}} = 1.69 - 1.37 = 0.32$

The experimental difference of the means is 0.32 kilogram.

d.



e. The plot of the data shows that the two data sets tend to be fairly symmetric and have no extreme values (outliers). So, the mean is a suitable measure of center. The mean yield of the treatment group is 0.32 kilogram more than the control group. It appears that the soil supplement might be slightly effective, but the sample size is small and the difference could be due to chance.

Monitoring Progress

Help in English and Spanish at *BigIdeasMath.com*

1. In Example 1, interpret the meaning of $\overline{x}_{\text{treatment}} - \overline{x}_{\text{control}}$ when the difference is (a) negative, (b) zero, and (c) positive.

Resampling Data Using a Simulation

The samples in Example 1 are too small to make inferences about the treatment. Statisticians have developed a method called resampling to overcome this problem. Here is one way to resample: combine the measurements from both groups, and repeatedly create new "control" and "treatment" groups at random from the measurements without repeats. Example 2 shows one resampling of the data in Example 1.

EXAMPLE 2 Resampling Data Using a Simulation

Resample the data in Example 1 using a simulation. Use the mean yields of the new control and treatment groups to calculate the difference of the means.

SOLUTION

Step 1 Combine the measurements from both groups and assign a number to each value. Let the numbers 1 through 10 represent the data in the original control group, and let the numbers 11 through 20 represent the data in the original treatment group, as shown.

original control group →	1.2	1.3	0.9	1.4	2.0	1.2	0.7	1.9	1.4	1.7
assigned number →	1	2	3	4	5	6	7	8	9	10

original treatment group →	1.4	0.9	1.5	1.8	1.6	1.8	2.4	1.9	1.9	1.7
assigned number →	11	12	13	14	15	16	17	18	19	20

Step 2 Use a random number generator. Randomly generate 20 numbers from 1 through 20 *without repeating a number*. The table shows the results.

```
randIntNoRep(1,20)
{14 19 4 3 18 9...
```

14	19	4	3	18	9	5	15	2	7
1	17	20	16	6	8	13	12	11	10

Use the first 10 numbers to make the new control group, and the next 10 to make the new treatment group. The results are shown in the next table.

	Resample of Tomato Plant Yields (kilograms)									
New Control Group	1.8	1.9	1.4	0.9	1.9	1.4	2.0	1.6	1.3	0.7
New Treatment Group	1.2	2.4	1.7	1.8	1.2	1.9	1.5	0.9	1.4	1.7

Step 3 Find the mean yields of the new control and treatment groups.

$$\overline{x}_{\text{new control}} = \frac{1.8 + 1.9 + 1.4 + 0.9 + 1.9 + 1.4 + 2.0 + 1.6 + 1.3 + 0.7}{10} = \frac{14.9}{10} = 1.49$$

$$\overline{x}_{\text{new treatment}} = \frac{1.2 + 2.4 + 1.7 + 1.8 + 1.2 + 1.9 + 1.5 + 0.9 + 1.4 + 1.7}{10} = \frac{15.7}{10} = 1.57$$

▶ So, $\overline{x}_{\text{new treatment}} - \overline{x}_{\text{new control}} = 1.57 - 1.49 = 0.08$. This is less than the experimental difference found in Example 1.

Making Inferences About a Treatment

To perform an analysis of the data in Example 1, you will need to resample the data more than once. After resampling many times, you can see how often you get differences between the new groups that are at least as large as the one you measured.

EXAMPLE 3 Making Inferences About a Treatment

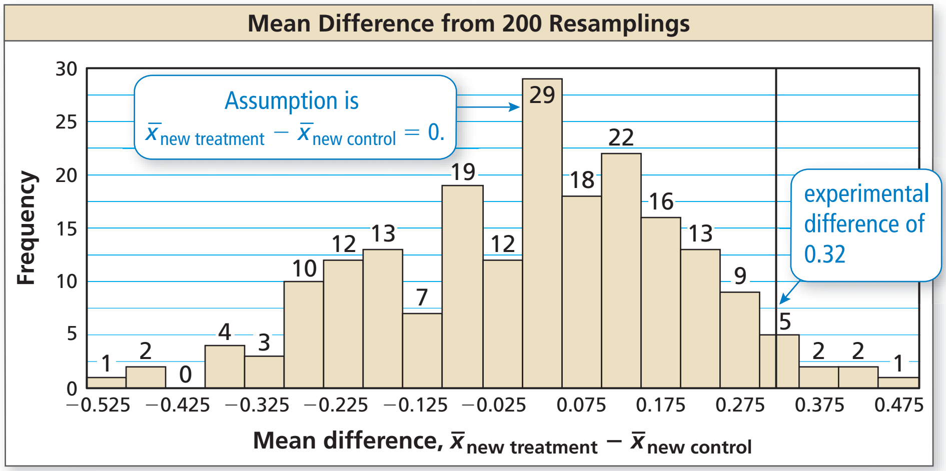
To conclude that the treatment in Example 1 is responsible for the difference in yield, you need to analyze this hypothesis:

The soil nutrient has no effect on the yield of the cherry tomato plants.

Simulate 200 resamplings of the data in Example 1. Compare the experimental difference of 0.32 from Example 1 with the resampling differences. What can you conclude about the hypothesis? Does the soil nutrient have an effect on the yield?

SOLUTION

The histogram shows the results of the simulation. The histogram is approximately bell-shaped and fairly symmetric, so the differences have an approximately normal distribution.



Note that the hypothesis assumes that the difference of the mean yields is 0. The experimental difference of 0.32, however, lies close to the right tail. From the graph, there are about 5 to 10 values out of 200 that are greater than 0.32, which is at most 5% of the values. Also, the experimental difference falls outside the middle 90% of the resampling differences. (The middle 90% is the area of the bars from −0.275 to 0.275, which contains 180 of the 200 values, or 90%.) This means it is unlikely to get a difference this large when you assume that the difference is 0, suggesting the control group and the treatment group differ.

INTERPRETING MATHEMATICAL RESULTS

With this conclusion, you can be 90% confident that the soil supplement does have an effect.

▶ You can conclude that the hypothesis is most likely false. So, the soil nutrient *does* have an effect on the yield of cherry tomato plants. Because the mean difference is positive, the treatment *increases* the yield.

Monitoring Progress

Help in English and Spanish at *BigIdeasMath.com*

2. In Example 3, what are the consequences of concluding that the hypothesis is false when it is actually true?

11.6 Exercises

Dynamic Solutions available at *BigIdeasMath.com*

Vocabulary and Core Concept Check

1. **COMPLETE THE SENTENCE** A method in which new samples are repeatedly drawn from the data set is called ____________.

2. **DIFFERENT WORDS, SAME QUESTION** Which is different? Find "both" answers.

What is the experimental difference of the means?

What is $\bar{x}_{\text{treatment}} - \bar{x}_{\text{control}}$?

	Weight of Tumor (grams)					
Control Group	3.3	3.2	3.7	3.5	3.3	3.4
Treatment Group	0.4	0.6	0.5	0.6	0.7	0.5

What is the square root of the average of the squared differences from −2.85?

What is the difference between the mean of the treatment group and the mean of the control group?

Monitoring Progress and Modeling with Mathematics

3. **PROBLEM SOLVING** A randomized comparative experiment tests whether music therapy affects the depression scores of college students. The depression scores range from 20 to 80, with scores greater than 50 being associated with depression. The control group has eight students and the treatment group, which receives the music therapy, has eight students. The table shows the results. *(See Example 1.)*

	Depression Score			
Control Group	49	45	43	47
Treatment Group	39	40	39	37

Control Group	46	45	47	46
Treatment Group	41	40	42	43

a. Find the mean score of the control group.

b. Find the mean score of the treatment group.

c. Find the experimental difference of the means.

d. Display the data in a double dot plot.

e. What can you conclude?

4. **PROBLEM SOLVING** A randomized comparative experiment tests whether low-level laser therapy affects the waist circumference of adults. The control group has eight adults and the treatment group, which receives the low-level laser therapy, has eight adults. The table shows the results.

	Circumference (inches)			
Control Group	34.6	35.4	33	34.6
Treatment Group	31.4	33	32.4	32.6

Control Group	35.2	35.2	36.2	35
Treatment Group	33.4	33.4	34.8	33

a. Find the mean circumference of the control group.

b. Find the mean circumference of the treatment group.

c. Find the experimental difference of the means.

d. Display the data in a double dot plot.

e. What can you conclude?

5. **ERROR ANALYSIS** In a randomized comparative experiment, the mean score of the treatment group is 11 and the mean score of the control group is 16. Describe and correct the error in interpreting the experimental difference of the means.

✗ $\bar{x}_{\text{control}} - \bar{x}_{\text{treatment}} = 16 - 11 = 5$

So, you can conclude the treatment increases the score.

6. **REASONING** In Exercise 4, interpret the meaning of $\overline{x}_{\text{treatment}} - \overline{x}_{\text{control}}$ when the difference is positive, negative, and zero.

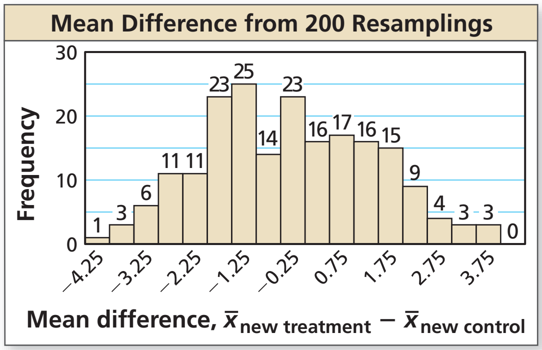
7. **MODELING WITH MATHEMATICS** Resample the data in Exercise 3 using a simulation. Use the means of the new control and treatment groups to calculate the difference of the means. *(See Example 2.)*

8. **MODELING WITH MATHEMATICS** Resample the data in Exercise 4 using a simulation. Use the means of the new control and treatment groups to calculate the difference of the means.

9. **DRAWING CONCLUSIONS** To analyze the hypothesis below, use the histogram which shows the results from 200 resamplings of the data in Exercise 3.

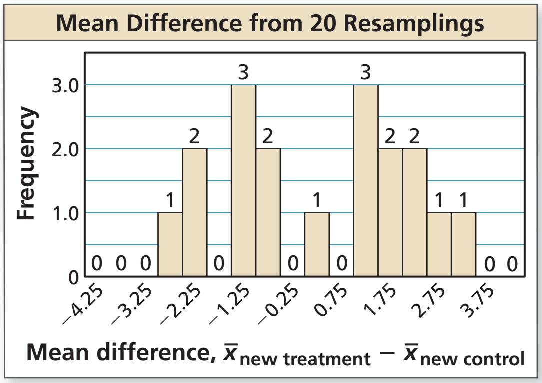
Music therapy has no effect on the depression score.

Compare the experimental difference in Exercise 3 with the resampling differences. What can you conclude about the hypothesis? Does music therapy have an effect on the depression score? *(See Example 3.)*

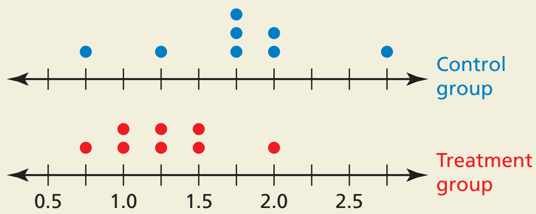


10. **DRAWING CONCLUSIONS** Suppose the experimental difference of the means in Exercise 3 had been −0.75. Compare this experimental difference of means with the resampling differences in the histogram in Exercise 9. What can you conclude about the hypothesis? Does music therapy have an effect on the depression score?

11. **WRITING** Compare the histogram in Exercise 9 to the histogram below. Determine which one provides stronger evidence against the hypothesis, *Music therapy has no effect on the depression score*. Explain.



12. **HOW DO YOU SEE IT?** Without calculating, determine whether the experimental difference, $\overline{x}_{\text{treatment}} - \overline{x}_{\text{control}}$, is positive, negative, or zero. What can you conclude about the effect of the treatment? Explain.



13. **MAKING AN ARGUMENT** Your friend states that the mean of the resampling differences of the means should be close to 0 as the number of resamplings increase. Is your friend correct? Explain your reasoning.

14. **THOUGHT PROVOKING** Describe an example of an observation that can be made from an experiment. Then give four possible inferences that could be made from the observation.

15. **CRITICAL THINKING** In Exercise 4, how many resamplings of the treatment and control groups are theoretically possible? Explain.

Maintaining Mathematical Proficiency

Reviewing what you learned in previous grades and lessons

Factor the polynomial completely. *(Section 4.4)*

16. $5x^3 - 15x^2$
17. $y^3 - 8$
18. $z^3 + 5z^2 - 9z - 45$
19. $81w^4 - 16$

Determine whether the inverse of *f* is a function. Then find the inverse. *(Section 7.5)*

20. $f(x) = \dfrac{3}{x + 5}$
21. $f(x) = \dfrac{1}{2x - 1}$
22. $f(x) = \dfrac{2}{x} - 4$
23. $f(x) = \dfrac{3}{x^2} + 1$