Unit 8, Lesson 13

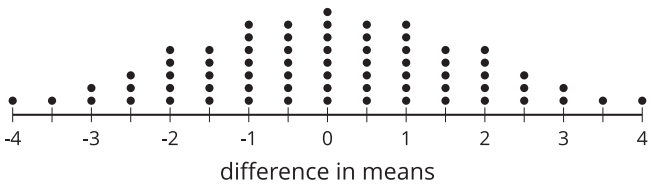
Using Normal Distributions for Experiment Analysis

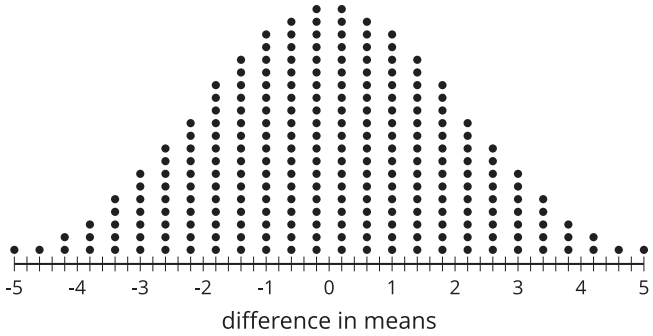
Let's determine when the results from an experiment are significant.

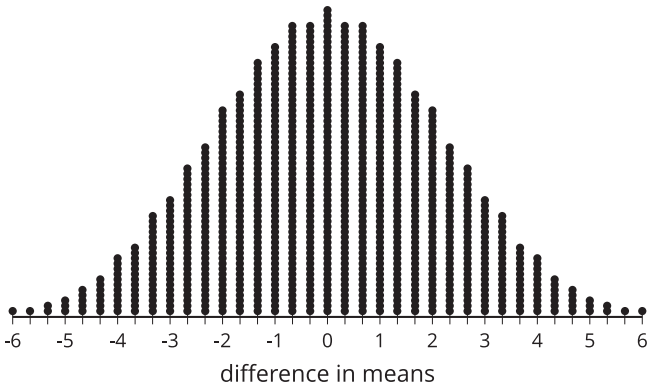
13.1 Notice and Wonder: Some Distributions

What do you notice? What do you wonder?









13.2 A Theoretical Experiment

To see what might be happening when we regroup data, consider an experiment that takes 12 subjects and divides them into 2 groups at random. The control group contains 6 subjects, and the treatment group contains 6 subjects. To explore what's possible, assume the control group results in these data: 1, 3, 4, 6, 8, and 10. The treatment group results in these data: 2, 5, 7, 9, 11, and 12.

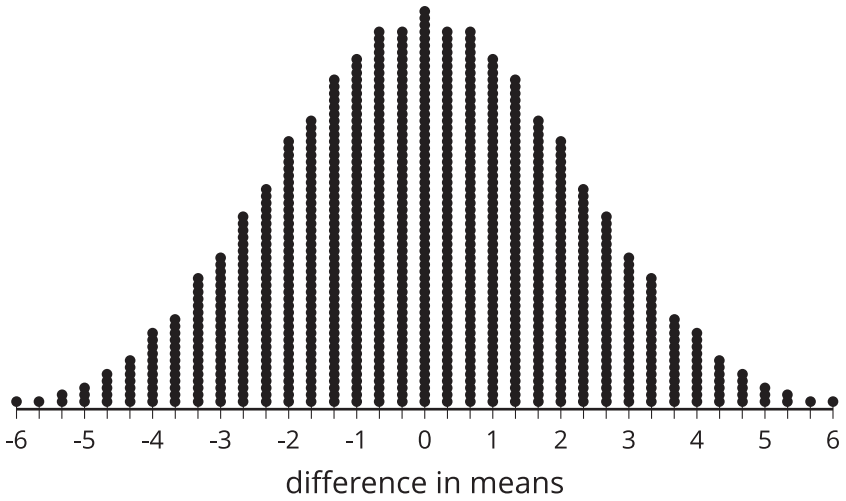
1. Find the difference in means for the original group by subtracting the control group mean from the treatment group's mean.

2. With a smaller data set like this, we can actually consider all of the different arrangements of the data. There are 924 distinct ways to separate the 12 values into 2 groups of 6. The frequency table shows all non-negative differences in means and how often they occur. Due to symmetry, the negative differences should match. Notice that a difference in means of 4.33 occurs 7 times, so a difference of -4.33 also occurs 7 times. The dot plot shows the same information.

difference in means	6	5.67	5.33	5	4.67	4.33	4
frequency	1	1	2	3	5	7	11

difference in means	3.67	3.33	3	2.67	2.33	2
frequency	13	18	22	28	32	39

difference in means	1.67	1.33	1	0.67	0.33	0
frequency	42	48	51	55	55	58



What proportion of possible groupings have a difference at least as great as the difference in means for the original groups? Explain or show your reasoning.

3. The proportion you calculate represents the probability that, if we assumed the treatment had no impact on the response, we would still see a difference at least as large as what we saw with our original grouping. Based on the proportion you calculated for this situation, which description is most accurate? Explain your reasoning.

    a. Because the proportion is so low, it is unlikely that the difference in means is due to the randomized groupings. This means that the difference in means is most likely caused by the treatment.

    b. Because the proportion is not that low, it is possible that the original difference in means is due to the random groupings. This means that there is not enough evidence to determine that the difference in means is likely caused by the treatment.

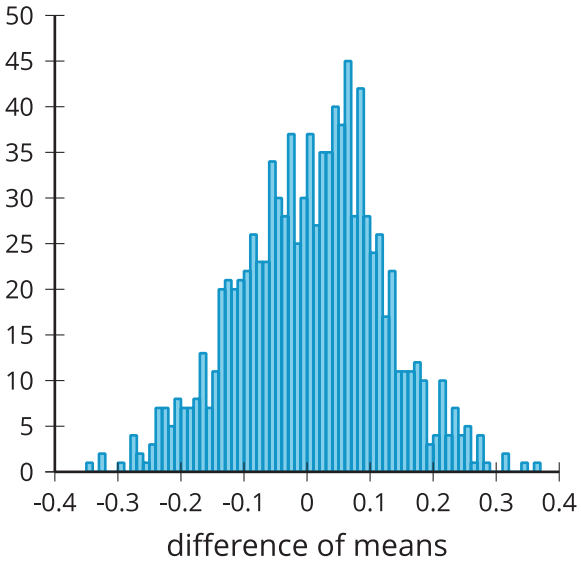
13.3 Simulating to Decide



Researchers want to know the effect of captively raising birds on the weight of the birds. The researchers begin with 100 birds divided into 2 groups of 50 each. One group of 50 is raised in captivity and the other 50 are tagged and released into the wild. After 5 years, all 100 birds are collected and weighed.

There are more than 10^{29} different ways to regroup the 100 birds into groups of 50 again, so looking at all the combinations would be too time consuming to reproduce. In this case, we can run simulations to determine how the original difference in means compares to those from regrouping the data.

The original groups have a difference of means of 0.27 gram. Researchers run 1,000 simulations that regroup the data into 2 groups at random and record the differences in means for the groups in each simulation. The histogram shows the differences in means from the simulations.



The researchers determine that the mean of the differences of means from the simulations is 0.0021 gram and the standard deviation for the differences of means from the simulations is 0.112 gram.

1. What features of the distribution in the histogram let us know that modeling with a normal distribution is reasonable?

2. Model the simulations using a normal distribution with a mean of 0 and a standard deviation of 0.112. What is the area under this normal curve that is greater than 0.27?

3. What does this area mean in this situation?

4. Based on the area under the normal curve, is there evidence that the original difference in means is likely due to where the birds spent the 5 years? Explain your reasoning.

Are you ready for more?

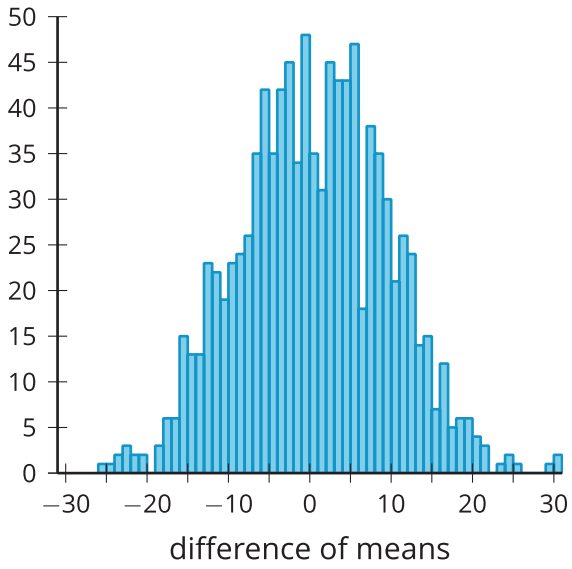
Suppose we decide that, if the probability of observing our difference or a more extreme difference happens less than 5% of the time in simulations, we will conclude that captivity caused the difference, and if the probability is greater than 5%, we will not draw any conclusions. In what ways could the conclusion we make (or decide not to make) be wrong?

Lesson 13 Summary

To analyze the significance of the data collected from an experiment, a randomization distribution can be used. In some cases, in which the number of subjects is small, all of the possible ways to regroup the data can be used to compare the original difference in means. When the difference in means is more extreme than most of the differences seen from the randomized regroupings (usually more than 90%, 95%, or 99%, depending on the situation), we can say that we have evidence that the difference in means is due to the treatment.

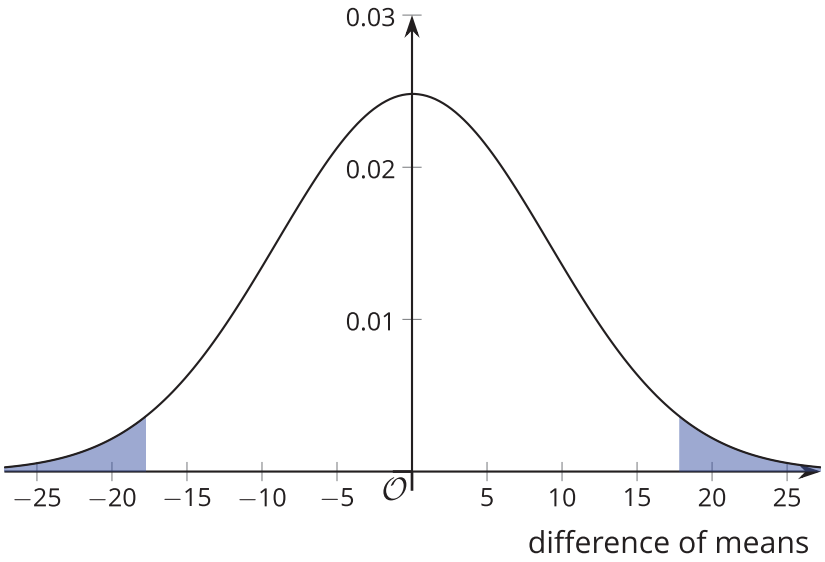
The more subjects included in the experiment, the greater number of possible regroupings. For example, 14 subjects divided into 2 groups of 7 can have their data redistributed into groups 3,432 different ways. When there are 60 subjects divided into 2 groups of 30, there are more than 118 quadrillion (1.18×10^{17}) different ways to redistribute the data into groups of 30. This large number of ways to regroup the data makes looking at the distribution of every possible regrouping difficult.

In these cases, we often do a simulation and redistribute the data many times to get a sense of the true distribution of all possibilities. For example, this histogram shows the difference of means for 1,000 simulations of redistributing 60 data values into 2 groups of 30 each.



The simulations should produce approximately normal distributions with a center near 0. This allows us to use our understanding of normal distributions to estimate the proportion of regroupings that are at least as extreme as the original difference in means from the experiment. When the proportion is small enough, we should conclude that there is enough evidence to say that the difference in means from the original groups is most likely due to the treatment.

For example, using the values from the histogram, the mean is 0.04 and the standard deviation is 9.07. That provides enough information to create a normal distribution that models the data. In the model image, we see the normal distribution and the regions for which the difference of means might be significant since there is only a 5% chance of the original difference in means being in the shaded region (less than -17.78 or greater than 17.78).



If the original difference in means is 20, for example, then we can conclude that there is evidence to show that the difference in means is due to the treatment. On the other hand, if the original difference in means is 10, for example, then we should say that there is not enough evidence to conclude that the difference in means is due to the treatment, because the results are also consistent with a model that assumes the treatment had no impact.