Comparing Hand Spans

In this activity, you will learn about the standard deviation, a common measure of variability.

How can you quantify variability and summarize it into a single measure?

1. Measure and record the hand span for each person in your group.

2. Enter the data into a TinkerPlots™ case table. Create a plot of the hand spans for your group. Sketch the plot below. Be sure to appropriately label the x-axis.

3. Compute the mean hand span for your group using TinkerPlots™. Record the mean.

The Standard Deviation

Recall that the mean is a single number that can be used to summarize the data. In this context, it is a description of the typical hand span measurement for your group. Of course, not every student in the sample is at the typical value (in fact all of them might be different from the typical value). Thus, it is also useful to have a single number description of how different the data tends to be from this typical value.

One single number description of the variability in a sample of data is called the **standard deviation** or *SD*. If the word "typical" is substituted for the word "standard" in its name, the name standard deviation (typical deviation) makes more sense. This measure quantifies variability by determining how far data cases typically deviate from the mean value.

- Use TinkerPlots™ to create a new attribute in the table, called *Deviations*, that contains the difference between the observed data (hand spans) and the mean of your group members' hand spans. Use a formula to compute this difference (you can compute these by subtracting the mean from each observation).
- Create a plot of the *Deviations* attribute.

4. How would you interpret the values of the *Deviations* attribute?

5. Sketch the plot below. Make sure to label the x-axis. Circle the mean and record its value.

6. How does the distribution of deviations compare with the distribution of hand span lengths you created in Question 2?

One thing you may not have known about the mean is that it is the value that "balances" the data. In other words, the mean is the value that gets the deviations to sum to zero. This is useful when describing a typical value of the data (it is the "closest" point to all of the cases, on average). If you try to average these deviations, however, you will always get zero. This is not very useful in summarizing variation in a data set, nor in comparing the variation between two data sets. One way to alleviate this problem is to square each of the deviations before you add them together.

- Create another attribute, *SquaredDeviations*, which contains the squared values of the deviations. (Again, use the Formula Editor to create this attribute.)
- Create a plot of the *SquaredDeviations* attribute.

7. Sketch the plot below. Make sure to label the x-axis.

8. How does the distribution of the squared deviations compare to the distributions of the original hand spans? How does it compare to the distribution of *deviations*?

9. Compute the mean of the squared distributions and record its value below. Because the deviations have been squared, this value represents the **typical squared deviation**.

10. Find the square root of the mean value and record its value below.

11. What does this new value represent (i.e., interpret its value)?

Computing the Standard Deviation using TinkerPlots™

Now use TinkerPlots™ to find the standard deviation of the original data directly.

- In the case table where you entered the original hand spans, create a new attribute called standardDeviation.
- Use the Formula Editor to compute the standard deviation of the hand spans by using the stdDev() function.

The value computed using TinkerPlots™ will be similar, albeit higher, than the value you obtained in Question 10. This is because there is a slight adjustment made to the denominator when a standard deviation is computed from a sample of data. From this point forward, you should always use the stdDev() function to compute the standard deviation.

Computing the Standard Deviation of a Plot of Results

Open the TinkerPlots™ file you saved from the *Helper or Hinderer* activity. (If you didn't save your TinkerPlots™ file from this activity, re-run the simulation.)

12. Describe the shape, center, and variation for the distribution of results. This time, rather than giving a more informal description of the variation, compute the standard deviation using the stdDev() function.

13. From statistical theory, we know that most observations in a distribution are within one standard deviation of the mean. Add and subtract one standard deviation from the mean to complete the following sentence:

Most simulated means will be between ______ and ______.

- Highlight the plot of results..
- Click the Divider tool in the upper toolbar. This will add a shaded rectangle to the plot.
- Click the Counts (%) button in the upper toolbar. This will show the percentage of cases that are included in the shaded area and those on both sides of the shaded area.
- Now move the ends of the shaded area to one standard deviation above and below the mean (to the values you computed in Question 13). To do this drag the white rectangles in the upper corners of the shaded rectangle. You can also double-click each of the white rectangles and enter in the value you want to move the line to.

14. What percentage of the results are within one standard deviation of the mean?

15. Now move the ends of the shaded area to two standard deviations above and below the mean. What percentage of the results are within two standard deviation of the mean?

16. How many standard deviations above the mean is the observed result of 14?

Most statisticians define *likely* results as those that are within two standard deviations of the mean. Anything more than two standard deviations from the mean would be called *unlikely*.