

Section 3.4: Measures of Dispersion

The measures of central tendency we studied in the last section (mean, median and mode) of a set of data can be deceiving. Using only mean, median and mode to make a decision may lead to mistakes. Standard deviation is a number that describes how bunched up or spread out a group of data is. Knowing the standard deviation of a group of data may help you make more informed decisions.

For Example:

Introduction to the concept of standard deviation: Consider two fictitious companies that make airplane engines.

Company A makes airplane engines that have an average (mean) life of **1000 hours** flying time before they fail.

Company B makes airplane engines for the **same price** that have an average (mean) life of **950 hours** flying time before they fail.

If all you look at is the average (mean) life span it may appear that Company A's is the better Engine.

Mean life can be a deceiving number. It is important to know more than the mean to make informed decisions.

For example: What if Company A's engine with an average lifespan of 1000 hours and half of the engines they make last 1500 hours and the other half last 500 hours. This would produce a mean life of 1000 hours. Of course if you are lucky and get the 1500 hour engine you would be happy, but if you expected to get 1000 hours and you get the 500 hour engine you would not be happy.

What if Company B's engines with an average lifespan of 950 hours has half of their engines last 975 hours and the other half last 925 hours. This will produce an engine whose mean life is 950 hours.

In the case then Company B's engines are more consistent and consequently more reliable and they may be the better choice if you don't want to gamble with reliability. B's engines don't last as long on average, but you might be willing to accept a shorter average life considering the alternative choice. You more or less can predict what you are buying with B's engines and A's engines are a gamble.

This example illustrates the importance of knowing something about the SPREAD or VARIABILITY of the data.

There are two ways we will measure spread of the data.

Range: the **range** of a set of data is the difference between the highest and lowest values in a set of data; it indicates the total spread of the data.

The range of Company's A engines is $1500 - 500 = 1000$ miles

The range of Company's B engines is $975 - 925 = 50$ miles

Example: What is the range of the following set of data? 23, 18, 30, 17, 24, 21, 5, 20, 23, 19

$$\text{Range} = 30 - 5 = 25$$

The second measure of the spread of the data is called **STANDARD DEVIATION**, it measures how much data differ from the mean. The lower case letter "s" or the Greek (sigma) σ is used to denote standard deviation. The "s" is used when the standard deviation of a sample is calculated. The σ is used when the standard deviation of the entire population is calculated. Just to keep things less confusing than they can get we will just consider the "s" standard deviation. You will learn when to use each standard deviation measure when you take a statistics class.

How to Manually Calculate Standard Deviation:

Here are the formulas for each type of standard deviation, again we will just focus on the s. The formulas are rather impressive to look at, but not that hard to use.

$$s = \sqrt{\frac{\sum(X - M)^2}{n - 1}}$$

$$\sigma = \sqrt{\frac{\sum(X - M)^2}{n}}$$

Basically, the larger the spread of data, the larger the standard deviation. In our example the standard deviation associated with the life of the A Company's engines would be far greater than that of the B companies.

Let me make up some numbers for the engine life of a few of company A and company B's engines and compute some standard deviations.

Example: Consider the following two sets of data of engine life. Find the standard deviation for each engine manufacturer.

A. 500, 500, 500, 1500, 1500, 1500

B. 925, 925, 925, 975, 975, 975

Standard deviation more or less measures how spread out the data is. In this case we would expect the Standard deviation in group A to be larger than the standard deviation in group B. Let's use a formula to compute the standard deviation for company A. This is a tedious computation, that we will ultimately use a calculator for.

Step 1. First, find the mean (average) of the data set.

$$mean = \frac{500+500+500+1500+1500+1500}{6} = 1000$$

Step 2. Find the difference between each data point and the mean

| |
|-------------------|
| Data point - mean |
| 500 - 1000 = -500 |
| 500 - 1000 = -500 |
| 500 - 1000 = -500 |
| 1500 - 1000 = 500 |
| 1500 - 1000 = 500 |
| 1500 - 1000 = 500 |

Step 3. Square the difference between each data point and the mean and sum the result and add up the numbers in the right column.

| Square results in step 2 | Squared difference |
|--------------------------|--------------------|
| $(-500)^2$ | 250000 |
| $(-500)^2$ | 250000 |
| $(-500)^2$ | 250000 |
| $(500)^2$ | 250000 |
| $(500)^2$ | 250000 |
| $(500)^2$ | 250000 |
| Total | 1500000 |

Step 4. Take the square root of the sum of the squares of the differences divided by the number of data points minus 1. (The minus one is part of the formula, and honestly I've never seen a great explanation of why we have to subtract 1, but I just do it anyway.) There are 6 data points and I will divide by $6 - 1 = 5$.

$$\text{standard deviation } s = \sqrt{\frac{1500000}{5}} = 547.72$$

Company A's engines have a standard deviation of 547.72 hours.

Now let's compute company B's engines standard deviation.

B. 925, 925, 925, 975, 975, 975

Step 1. First, find the mean (average) of the data set.

$$\text{mean} = \frac{925+925+925+975+975+975}{6} = 950$$

Step 2. Find the difference between each data point and the mean

| |
|-------------------|
| Data point - mean |
| $925 - 950 = -25$ |
| $925 - 950 = -25$ |
| $925 - 950 = -25$ |
| $975 - 950 = 25$ |
| $975 - 950 = 25$ |
| $975 - 950 = 25$ |

Step 3. Square the difference between each data point and the mean and sum the numbers in the right column.

| Square results in step 2 | Squared difference |
|--------------------------|--------------------|
| $(-25)^2$ | 625 |
| $(-25)^2$ | 625 |
| $(-25)^2$ | 625 |
| $(25)^2$ | 625 |
| $(25)^2$ | 625 |
| $(25)^2$ | 625 |
| Total | 3750 |

Step 4. Take the square root of the sum of the squares of the differences divided by the number of data points minus 1. Again I will divide by $6 - 1 = 5$.

$$\text{standard deviation } s = \sqrt{\frac{3750}{5}} = 27.39$$

Company B's engines have a standard deviation of 27.39 hours.

Even though company A's engines last 50 hours longer on average the standard deviation in the lives of the A engines of (547.72) hours far exceeds that of the B engines (27.39) hours. Company B's engines may be the better choice. Of course there are many other factors, such as cost that need to go into the decision.

The standard deviation show a difference that the mean just can't show.

We really won't need to do this computation by hand, as the TI-83 and TI-84 calculators can do it. I will show you how to use your calculator to compute standard deviation on my video. Here are the steps I will follow. I'm not sure they will make much sense without watching me.

Example: Use your calculator to find the standard deviation of A engine data. We should get 547.72 for an answer. We did this computation by hand and our calculator will give exactly the same answer.

500, 500, 500, 1500, 1500, 1500

Using a TI 83+ to Calculate Standard Deviation.

Step 1. Press **STAT** **EDIT** **1:Edit**. (Press STAT and hit enter on Edit)

Step 2. Move your cursor to the top of the "L1" column and hit

CLEAR

Step 3. Enter your data in the "L1" column, pressing **ENTER** after every data entry.
I will enter 500, 500, 500, 1500, 1500, 1500

Step 4. Press **STAT** **CALC** **1:1-Var Stats** **ENTER** **ENTER**

Step 5. Scroll down to the lower case symbol for the s.

You will see we get $s = 547.72$ just like we did when we did this by hand.

Answer: Standard deviation = 547.72 hours

Example: Compute the standard deviation of the B data. The answer should be the same answer that we got when we did this by hand. The answer should be 27.39.

Remember B. 925, 925, 925, 950, 950, 950

Follow the steps above.

Answer: Standard deviation = 27.39 hours

Sometimes only a very small standard deviation is desirable or acceptable.

Another short discussion about the importance of standard deviation.

If a cereal box is supposed to contain 20 oz. of cereal then consumers expect 20 oz.

If the amount varies too much from box to box- sometimes under filling, sometimes overfilling- the manufacturer will soon be in trouble with consumer groups (if under filling) or the investors in the company (if overfilling...and losing potential profit).

Other times a larger spread of data might be expected...for example IQ scores are expected to exhibit a considerable spread about the mean because everyone is different.

Homework #1-6: Determine the range and standard deviation of the set of data, round to the nearest hundredth when applicable. You may use your calculator to calculate standard deviation. Remember to select the "s".

- 1) 10, 10, 14, 16, 8, 8
- 2) 3, 5, 7, 9, 11, 13, 15
- 3) 50, 75, 100, 125, 150, 175, 200
- 4) 1, 3, 71, 8, 92, 63
- 5) 5, 10, 20, 200, 300, 400
- 6) 0, 3, 5, 9, 9, 10, 11, 13

7) At the Steakhouse Grill, steaks are cut into 12-ounce portions when they are ordered by a customer. For the past month, the manager of Steakhouse Grill has been trying to determine which cook can mostly accurately portion out the meat. The manager collected the mean of sample meat portions for a few weeks for both of his cooks. The results are summarized in the table below.

| | Week 1 | Week 2 | Week 3 | Week 4 |
|---------|--------|--------|--------|--------|
| Opie | 12.5 | 12.3 | 11.8 | 11.2 |
| Maynard | 12.0 | 12.2 | 11.8 | 12.4 |

- a) Compute the mean portion size for each cook.
- b) Compute the standard deviations for each cook.
- c) The manager needs to choose a cook to cut portions of meat throughout the week. Who should he choose? Why?

8) For a summer job, you were working in the quality control department for a computer company that manufactures computer parts. The specific part that you are to evaluate the quality of is supposed to be 8 micrometers in thickness. You obtained samples of four of these parts manufactured by the day shift and four parts manufactured by the night shift workers.

Here are the findings:

Day Shift 7.9, 8.0, 8.2, 8.3

Night Shift 2, 4, 12, 14

- a) Compute the mean thickness for each shift.
- b) Compute the standard deviations for each shift.
- c) Which crew is doing a better job? Why?

Answers: 1) Range 8, standard deviation 3.29 3) Range 150 standard deviation 54.01
5) Range 395 standard deviation 170.19 7a) Opie mean 11.95 oz. Maynard mean 12.1 oz.
7b) Opie standard deviation .58 ounces Maynard standard deviation .26 oz
7c) Maynard's cuts are more consistent and closer to the desired average. I would pick Maynard. I could see an argument for Opie. If you selected Opie and gave a good reason, I would give you credit.