TutorTube: Sum or Squares, Standard Deviation, and Variance

Introduction

Hello, welcome to another edition of TutorTube, where the Learning Center's Lead Tutors help you understand challenging course concepts with easy to understand videos. My name is Kelly Schmidt, Lead Tutor for statistics at the Learning Center. In today's video, we will go through the process of finding the sum of squares, standard deviation, and variance of a dataset using hand calculations (with some assistance from Excel). If you have never used Excel and are brand new to statistics, this video is for you. Let's get started!

Sum of Squares, Standard Deviation, Variance

Before we jump into calculating these values, let me give you some idea of the general idea behind what we are trying to do when we find the sum or squares, variance, or standard deviation of a dataset. Let's say we have a set of numbers which represent test scores on an English test for a certain class of 8 high school students.

Table 1. Class A

Student	Score
1	45
2	65
3	70
4	75
5	75
6	80
7	92
8	98

We can find the average score by adding up all the numbers and dividing by the number of students. This gives us:

$$\frac{70 + 80 + 75 + 98 + 92 + 45 + 75 + 65}{8} = \frac{600}{8} = \mathbf{75}$$

So, we can say that the average score on this test was a 75%. However, the average value only tells us about one aspect of this class of students. We know that overall, the class had a C average on the test, but that number alone doesn't give us any information about how spread out the scores were. For example, let's say we had another class of high schoolers taking the same test and got the scores below:

Table 2. Class B

Student	Score
1	100
2	100
3	100
4	98
5	92
6	0
7	75
8	35

If we take the average here, we get:

$$\frac{100 + 100 + 100 + 98 + 92 + 0 + 75 + 35}{8} = \frac{600}{8} = \mathbf{75}$$

This class has the same average as the first class, but if we look at the scores themselves, they look a lot different. The first class had scores that ranged between 45 to 98, whereas this class had scores ranging from 0 to 100. In order to get an idea of the spread of these two groups, we need another type of statistic; this is where the idea of sum of squares, variance, and standard deviation come into play. For a **sample**, the formulas for each are shown here:

$$Sum\ of\ Squares\ (SS) = \sum (x_i - \bar{x})^2$$

$$Sample\ Variance\ (s^2) = \frac{1}{n-1} \sum (x_i - \bar{x})^2$$

$$Sample\ Standard\ Deviation\ (s) = \sqrt{\frac{1}{n-1} \sum (x_i - \bar{x})^2}$$

Notice how similar each of the formulas are. In fact, we will use the sum of squares to calculate variance, and then use the variance to calculate the standard deviation. Also, remember that these are formulas we use when dealing with SAMPLES. If we were working with a POPULATION, the formulas

would change just a bit. If you've never worked with this kind of mathematical notation before, don't worry about it. I will walk through the process of finding them step by step.

Sum of Squares

Start by opening a blank sheet and then entering in our data for the first class of students.

Figure 1. Initial Data in Excel

Step 1: Find the average, $\bar{x}$

First, we need to find the mean of the scores. Click into a blank cell anywhere in the sheet and type =AVERAGE(. Next, click and drag to select all the scores in our class. Then type a closed parenthesis ")" and hit Enter. This will spit out the mean or average value of our data set, which we know should be 75.
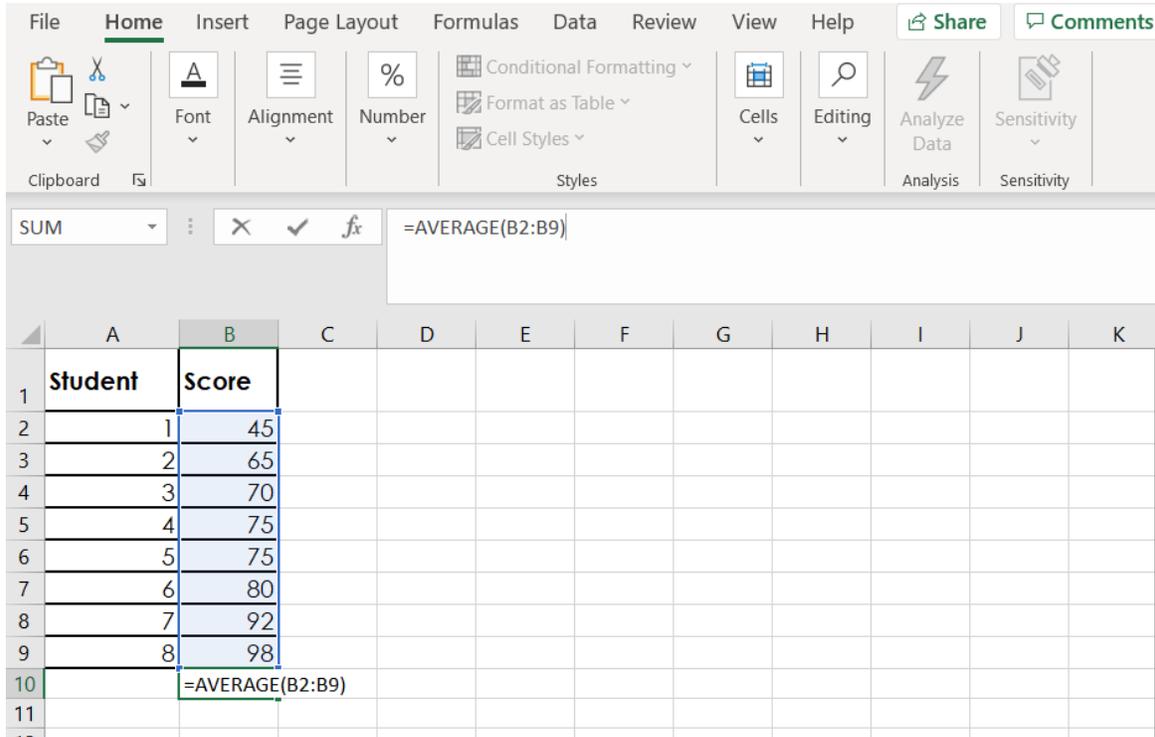
Figure 2. Average of Score

Step 2: Find $x_i - \bar{x}$ for each score

Ok, now we make a new column which I'll call SCORE – MEAN. In this step, we build out the middle part of the complicated SS formula by taking each person's score and subtracting the overall mean. The idea is that we want to find the distance of each score from the average score of the group, which was 75. In the second box, type an equal sign then click on the box holding your first score, 45. Next, type a minus sign. Then type in your mean, 75, and hit Enter.
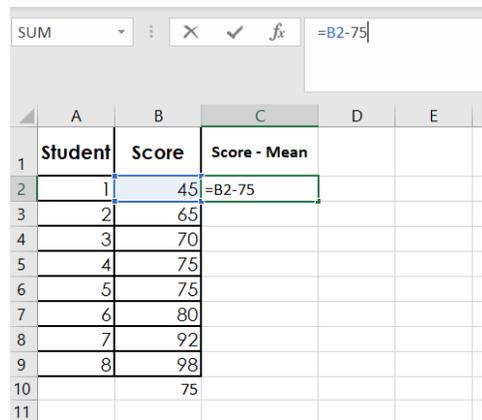

Figure 3. Score - Mean

Excel took the score of 45 and subtracted the mean of 75 to give us a value of -30. We could repeat this for each of our observations all down the column, but Excel has a shortcut. Click on the cell with the -30 inside and then put your cursor over the little black box at the bottom right. When you see it turn into a black plus sign, click and drag down the column.

C2			×	✓	f_x	=B2-75

	A	B	C	D
1	Student	Score	Score - Mean	
2	1	45	-30	
3	2	65	-10	
4	3	70	-5	
5	4	75	0	
6	5	75	0	
7	6	80	5	
8	7	92	17	
9	8	98	23	
10		75		
11				

Figure 4. Score - Mean Column Entries

This copies the formula into each of the cells.

Step 3: Find $(x_i - \bar{x})^2$ for each score

Now that we have our full column of scores, we need to square each of them. I'll make a new column called (SCORE – MEAN)^2. Just like we did before, we click into the first cell and type an equal sign and click on the cell to the left, which now contains our -30. This time instead of subtracting 75, we want to square the value, so we type a ^2 and hit enter.

Figure 5. (Score – Mean)^2

Negative 30 squared gives us a value of 900. Now we drag the cell down to apply the formula to the full column.

Figure 6. (Score - Mean)^2 Column Entries

Step 4: Find $\sum(x_i - \bar{x})^2$, the sum of squares (SS)

In the next step, we need to sum up all of these values. Again, we could do this by hand by just adding them up, but Excel has a shortcut function. Click on any empty cell and type =SUM(. Then click and drag along the column of values we just calculated. Type a closed parenthesis and hit Enter.

Figure 7. Sum of (Score - Mean)^2 Column

This tells us that the sum of all the numbers in the column is 1868. This value is our sum of squares (SS).

Step 5: Find $\frac{1}{n-1}\sum(x_i - \bar{x})^2$, the sample variance ($s^2$)

If we want to find our sample variance, s^2 we divide our sum of squares value by n-1, which is sample size minus one. Our sample size here is 8 because we had eight students in this class. So, we need to divide 1868 by 8-1, which is 7. To do this in Excel, choose any empty cell and type =1868/7.

Figure 8. Sample Variance is $\frac{SS}{n-1}$

Hit enter, and we have our sample variance of roughly 266.86.

Step 6: Find $\sqrt{\frac{1}{n-1}\sum(x_i - \bar{x})^2}$**, the sample standard deviation** (s)

Finally, we can calculate the standard deviation for this sample. To do this, we just need to take the square root of the variance. Choose any empty cell in Excel and type =SQRT(. Then click on the cell containing the variance value of 266.86 that we just calculated and type a closed parenthesis. Hit enter, and we have our value of 16.34 for the standard deviation.

SUM	⋮	✕ ✓ fx	=SQRT(D12)

	A	B	C	D
1	Student	Score	Score - Mean	(Score - Mean)^2
2	1	45	-30	900
3	2	65	-10	100
4	3	70	-5	25
5	4	75	0	0
6	5	75	0	0
7	6	80	5	25
8	7	92	17	289
9	8	98	23	529
10		75		
11			SS =	1868
12			s^2 =	266.8571429
13				=SQRT(D12)
14				

Figure 9. Standard Deviation is $\sqrt{s^2}$

And that is all there is to it. We found our value for sum of squares, variance, and standard deviation. If you want to check your work, you can also use Excel's shortcut formula to get the standard deviation in one step. We can type =STDEV.S(in an empty cell and highlight our original list of values. Close the parentheses and hit Enter.

Figure 10. Shortcut Formula for Sample Standard Deviation

This should be the exact same value that we got by building out step in the formula by hand. Since it matches, we know that we did all of our steps correctly!

Outro

Thank you for watching this TutorTube presentation! I hope you enjoyed this video. Please subscribe to our channel for more exciting videos. Check out the links in the description below for more information about The Learning Center and follow us on social media. See you next time!