



Essential Math for Data Analysis Using Excel Online

Module 1, Lab 4: Standard Deviation

Learning Objectives

- Practice finding the SD in Excel using “fill series” (manually building a formula) and using the canned Excel formula.

Description

Learners will work with a set of numbers and practice finding the standard deviation. They will create columns in Excel to execute the steps of the SD formula (across a series of columns, to illustrate the steps, and dragging down using “fill series”). They will then use the canned formula in Excel.

Data set

Mod1Lab4.csv

Overview

The standard deviation of a set of values is a measure of how spread out those values are. The formula for finding the standard deviation is pretty gnarly, so in this lab, we’ll go through it step by step. Once we’ve seen each step in action, we’ll speed things up with a canned formula in Excel. Our data set will involve everyone’s favorite morning stimulant: coffee.

What You’ll Need

To complete the lab, you will need the online version of Microsoft Excel.

Exercise 1: Find the Standard Deviation Step-by-Step

1. Open the data set in Excel. Here’s what the data look like:

	A	B
1	day	coffees
2	1	2
3	2	4
4	3	4
5	4	4
6	5	5
7	6	6
8	7	6
9	8	5
10	9	5
11	10	3

2. Jot down the standard deviation formula to keep track of all the operations you're doing:





$$\hat{\sigma} = \sqrt{\frac{\sum (x - \bar{x})^2}{n - 1}}$$

3. First, you'll need the mean number of coffees per day (that's \bar{x} in the formula). Create a new column and use the AVERAGE function to find the mean, or average, of all the values in column B.

fx =AVERAGE(B2:B11)			
	A	B	C
1	day	coffees	mean
2	1	2	4.4
3	2	4	
4	3	4	
5	4	4	
6	5	5	
7	6	6	
8	7	6	
9	8	5	
10	9	5	
11	10	3	

This person drank an average of 4.4 cups of coffee per day.

4. Now you want to find the difference between each value and the mean (x minus \bar{x}). In other words, take the actual number of coffees from Day 1 and subtract the mean value from that. To do this, set up a new column for the difference. In the first cell of that column, enter **=B2-4.4**, which is the difference between the actual coffees and the mean coffees.

D2	 	 	f_x	=B2-4.4
	A	B	C	D
1	day	coffees	mean	difference
2	1	2	4.4	-2.4
3	2	4		
4	3	4		
5	4	4		
6	5	5		
7	6	6		
8	7	6		
9	8	5		
10	9	5		
11	10	3		
12				

5. You want that same difference for each day, so use the “fill series” trick to populate the other cells in the difference column. Click and drag the green square in the lower-right corner of cell D2 down all the way to D11. That will automatically calculate the difference between each entry in column B and the mean of 4.4.

D2				\times	\checkmark	fx	=B2-4.4
	A	B	C	D			
1	day	coffees	mean	difference			
2	1	2	4.4	-2.4			
3	2	4		-0.4			
4	3	4		-0.4			
5	4	4		-0.4			
6	5	5		0.6			
7	6	6		1.6			
8	7	6		1.6			
9	8	5		0.6			
10	9	5		0.6			
11	10	3		-1.4			

6. Now you want to square each of those new values in the difference column. Set up a new column for the squared differences and enter the formula **=D2^2**. That'll give the value of $(-2.4)^2$.

E2					$=D2^2$
	A	B	C	D	E
1	day	coffees	mean	difference	square of diff
2	1	2	4.4	-2.4	5.76
3	2	4		-0.4	
4	3	4		-0.4	
5	4	4		-0.4	
6	5	5		0.6	
7	6	6		1.6	
8	7	6		1.6	
9	8	5		0.6	
10	9	5		0.6	
11	10	3		-1.4	

7. Use “fill series” again to drag this formula into every cell of column E, which will automatically square each value from column D.

E2					$=D2^2$
	A	B	C	D	E
1	day	coffees	mean	difference	square of diff
2	1	2	4.4	-2.4	5.76
3	2	4		-0.4	0.16
4	3	4		-0.4	0.16
5	4	4		-0.4	0.16
6	5	5		0.6	0.36
7	6	6		1.6	2.56
8	7	6		1.6	2.56
9	8	5		0.6	0.36
10	9	5		0.6	0.36
11	10	3		-1.4	1.96

8. Next up, find the *variance*, which is the sum of those squared values divided by $n - 1$, where n is the number of values ($n = 10$ in this case because there are 10 observations/days). You guessed it: You’ll use a new column for this. Use the SUM formula to add up all 10 values in column E, then divide that sum by $10 - 1 = 9$. Translation: Type **`=SUM(E2:E11)/9`** into cell F2.

	A	B	C
1	day	coffees	std dev
2	1	2	
3	2	4	
4	3	4	
5	4	4	
6	5	5	
7	6	6	
8	7	6	
9	8	5	
10	9	5	
11	10	3	

2. In cell C2, type in the STDEV formula. The syntax is **=STDEV(first value:last value)**. You want the standard deviation of all the values in column B, so that's **=STDEV(B2:B11)**.

fx =STDEV(B2:B11)			
	A	B	C
1	day	coffees	std dev
2	1	2	1.264911
3	2	4	
4	3	4	
5	4	4	
6	5	5	
7	6	6	
8	7	6	
9	8	5	
10	9	5	
11	10	3	

Awesome. That's the same standard deviation we got using the longer methods. (Note: The last couple digits of your final answer might be slightly different due to rounding, depending on how wide the cells in column C are.)