SAMLab Tip Sheet #6

Calculating the Variance and Standard Deviation

The purpose of this Tip Sheet is two-fold. First, we will use Excel to calculate a variance and standard deviation using the definitional formula for a standard deviation shown below.

$$\sqrt{\frac{\sum (x_i - \bar{x})^2}{N}}$$

Second, we will see how Excel can calculate variances and standard deviations with its functions. For both methods we will use the data below.

	A	B	C	D	E
1	Data				
2	9				
3	7				
4	11				
5	13				
6	13				
7	14				
8	5				
9	9				
10	13				
11	7				

Sheet1 / Sheet2 / Sheet3 /

The Definitional Approach

Technically, the variance is the average squared deviation of a data set, and the standard deviation is the square root of the variance. These are not descriptions most people easily understand. Calculating the variance and standard deviation piece-by-piece gives us a firmer grasp of what exactly the variance and standard deviation are. To begin, we will calculate the average and then subtract it from each of our scores, making a column of **deviation scores**.

Microsoft Excel - Book1

File Edit View Insert Format Tools Data Window
Help Macro Filter

B2   = =A2-A12

	A	B
1	Data	Deviation Scores
2	9	-1.1
3	7	-3.1
4	11	0.9
5	13	2.9
6	13	2.9
7	14	3.9
8	5	-5.1
9	9	-1.1
10	13	2.9
11	7	-3.1
	10.1	

Sheet1 / Sheet2 / SI

This is the average calculated with a paste function. The cell contains the following:
=AVERAGE(A2:A11)
For help with using the function library see Tip Sheet #3.

Notice the static cell reference used in the formula. This is done so that the formula can be pasted into the lower cells without changing the cell reference for the mean. For help on static cell references and pasting formulas see Tip Sheet #2 for an example using deviation scores.

SAMLab Tip Sheet #6

The next step is to square our deviation scores. Also, we will check to make sure we calculated our deviation scores correctly by summing them up in cell B12. The sum of deviation scores always equals 0 when they are calculated properly, which is why the standard deviation is not just the average deviation score. If it were, we would always get 0 for the standard deviation.

To square a number, multiply it by itself or raise it to the 2nd power using the ^ character. See Tip Sheet #1 for help.

Our next step is to calculate the sum of the squared deviation scores (termed **sum of squares**). We then calculate the variance by dividing the sum of squares by the number of scores (N).

Taking the square root of the variance yields the standard deviation shown on the left below. You may have come across a different formula for the variance and standard deviation that divides the sum of squares by "N-1" rather than "N."[1] The formula shown here with "N" is used when treating a set of scores as an entire population, and it yields the variance and standard deviation for these particular data. If we were to treat the scores as just a sample or subset of a population, we would use the formula with "N-1" to estimate the variance and standard deviation of the population from which the sample was taken. The window on the right below shows the results using the "N-1" formula.

Left window — Microsoft Excel - Book1

Formula bar: D10 = =SQRT(D8)

	C	D
1	Squared Deviations	
2	1.21	
3	9.61	
4	0.81	
5	8.41	Sum of Squares
6	8.41	88.9
7	15.21	Variance
8	26.01	8.89
9	1.21	Standard Deviation
10	8.41	2.98
11	9.61	

Sheet1 / Sheet2

Right window — Microsoft Excel - Book1

Formula bar: D8 = =D6/9

	C	D
1	Squared Deviations	
2	1.21	
3	9.61	
4	0.81	
5	8.41	Sum of Squares
6	8.41	88.9
7	15.21	Variance
8	26.01	9.88
9	1.21	Standard Deviation
10	8.41	3.14
11	9.61	

Sheet1 / Sheet2

Using the Paste Functions

A quicker way to calculate the variance and standard deviation is to use the paste function library. There are functions for both the "N" and "N-1" formulas. We'll put each along side the values we calculated using the definitional formula to make sure we did them right. Below on the left is the paste function output for the "N" variance and on the right is the "N" standard deviation. Check the formula bars for the names of the functions.
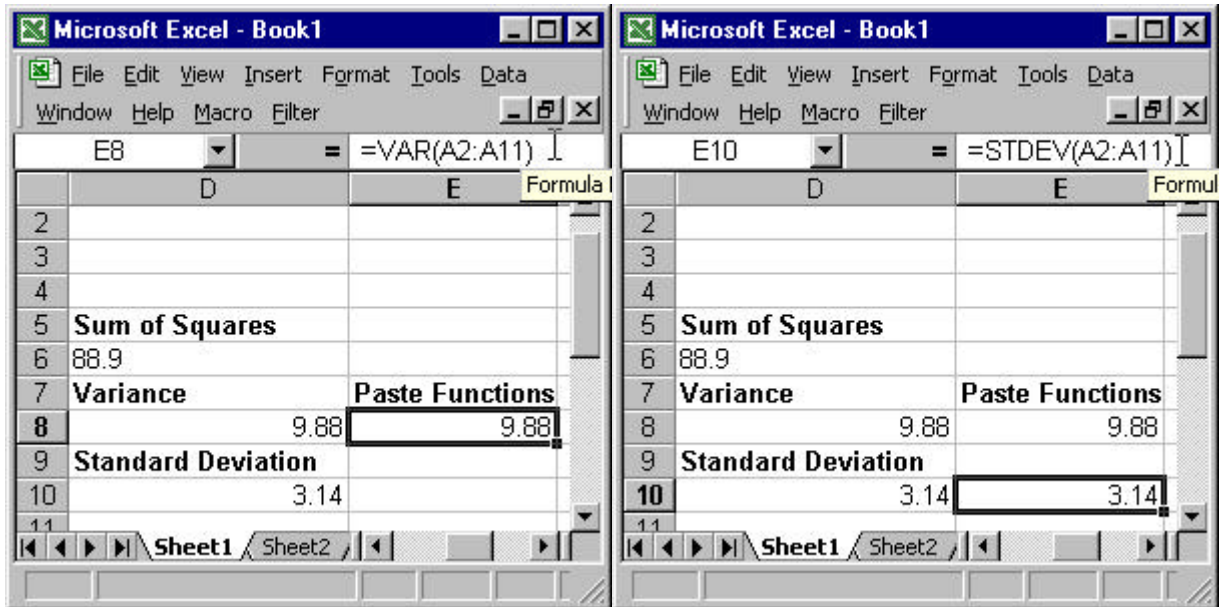
Left window — Microsoft Excel - Book1

Formula bar: E8 = =VARP(A2:A11)

	D	E
1		
2		
3		
4		
5	Sum of Squares	
6	88.9	
7	Variance	Paste Functions
8	8.89	8.89
9	Standard Deviation	
10	2.98	

Sheet1 / Sheet2

Right window — Microsoft Excel - Book1

Formula bar: E10 = =STDEVP(A2:A11)

	D	E
2		
3		
4		
5	Sum of Squares	
6	88.9	
7	Variance	Paste Functions
8	8.89	8.89
9	Standard Deviation	
10	2.98	2.98

Sheet1 / Sheet2

SAMLab Tip Sheet #6

Notice that in order to calculate the variance and standard deviation we only have to insert the function and select the original data. We do not need to calculate the deviation scores or the sums of squares. Below on the left you'll find the paste function output for the estimated variance and on the right you'll find the formula for the estimated standard deviation. Note that the only difference in the function names displayed in the formula bar is that these, the "N-1" formulas, do not have the "P" at the end of their names, meaning that the data are not treated as constituting a population.

	D	E	
Microsoft Excel - Book1			_ □ X
File Edit View Insert Format Tools Data Window Help Macro Filter			_ ⊟ X
E8	▼	= =VAR(A2:A11)	
	D	E	Formula
2			
3			
4			
5	Sum of Squares		
6	88.9		
7	Variance	**Paste Functions**	
8	9.88	9.88	
9	Standard Deviation		
10	3.14		
11			
Sheet1 Sheet2			

	D	E	
Microsoft Excel - Book1			_ □ X
File Edit View Insert Format Tools Data Window Help Macro Filter			_ ⊟ X
E10	▼	= =STDEV(A2:A11)	
	D	E	Formul
2			
3			
4			
5	Sum of Squares		
6	88.9		
7	Variance	**Paste Functions**	
8	9.88	9.88	
9	Standard Deviation		
10	3.14	3.14	
11			
Sheet1 Sheet2			

Notes

[1]Here is the "N-1" formula for the standard deviation. $\sqrt{\dfrac{\sum (x_i - \bar{x})^2}{N-1}}$ The formula that uses "N – 1" is often called the inferential standard deviation (because it's used to estimate a population standard deviation in statistical inference), and the "N" formula is often called the descriptive standard deviation (because is describes a particular sample).