

Comparison of two data sets with the same mean

NAME:

We will work with the data sets below to practice finding variance and standard deviation. The first step is to find the means of both sets (add them up and divide by how many numbers you have). Do that now and write your answers in the table below.

<b>Data Set #1</b>		<b>Data Set #2</b>
6		0
8		5
10		10
12		15
14		20
<b>mean =</b>		<b>mean =</b>

The formula for the variance, represented by  $v$ , of a set of numbers is given below. (The variables  $x_1, x_2, \dots, x_n$  represent the actual values in the data set,  $n$  is the number of values in the data set, and  $\bar{x}$  is the mean of the set.)

To find the standard deviation, you take the square root of the variance.

$$v = \frac{(x_1 - \bar{x})^2 + (x_2 - \bar{x})^2 + \dots + (x_n - \bar{x})^2}{n - 1}$$

1a. First, we will find the variance of data set #1. How many values are in the set? This number will be the  $n$  in the formula.

The variance calculation is done on the next page. (This worksheet, by the way, uses the *sample* variance and standard deviation formulas, as though the data was obtained through a sample, as opposed to population data. That formula is slightly different.)

1b. Now, we will follow the formula. It says to take each value, subtract the mean, and square that. We then add those numbers up and divide by  $n - 1$ . The table below will help you do this step by step. The first row is done for you.

<b>Data Set #1</b>		
$x_i$	$(x_i - \bar{x})$	$(x_i - \bar{x})^2$
6	$6 - 10 = -4$	$(-4)^2 = 16$
8		
10		
12		
14		
Remember that $\bar{x} = 10$ .		<b>sum of column =</b>
<b>Here, we find the variance of the data.</b>		<b>divide sum by <math>n - 1 \rightarrow</math></b>

1c. Now take the square root of the variance we just found to find the standard deviation of data set #1.

2. Standard deviation is essentially the average distance each number is from the mean. It tells us how spread out the numbers are. Do you think set #2 will have a bigger or smaller standard deviation than set #1? Why?

3a. Let's find the standard deviation for set #2. Notice it also has 5 values, so  $n$  is 5. We will repeat the table work we did but now use set #2.

<b>Data Set #2</b>		
$x_i$	$(x_i - \bar{x})$	$(x_i - \bar{x})^2$
0		
5		
10		
15		
20		
Remember that $\bar{x} = 10$ .		<b>sum of column =</b>
<b>Here, we find the variance of the data.</b>		<b>divide sum by <math>n - 1 \rightarrow</math></b>

3b. Now take the square root of the variance we just found to find the standard deviation of data set #2. Did you guess correctly in question 2?

4. Draw a number line, with 0 on the far left and 20 on the right. You do *not* need to label every number in between but include enough tick marks so that they are clear. Then plot the numbers for set #1 above this number line and set #2 below the line, labeling the sets. Which set appears more spread out? How do your values for the sets' standard deviations show this?