

## Brief Definition of Variance and Standard Deviation

### An Example

Where the average (or mean) is a measure of the center of a group of numbers, the **variance** is the measure of the spread. The following two sets of numbers have the same mean, 10.

$$S^1 = \{10, 10, 10, 10, 10\}$$

$$S^2 = \{0, 5, 10, 15, 20\}$$

The first set though has a variance of zero; all numbers are the same. The second set has a variance of 50.

### Computing the Variance

$$\text{Var}(S) = \text{Sum}_i (S_i - E(S))^2 / N$$

where  $\text{Sum}_i$  means to sum over all elements of set  $S$ ,  
 $N$  is the number of elements in  $S$ ,  
 $S_i$  is the  $i$ th element of the set  $S$ ,  
and  $E(S)$  is the mean over the values of set  $S$ .

When we are dealing with a *sample* (that is, a subset of the complete population), we cannot of course compute the mean and variance exactly, but rather estimate them. Given a sample  $U$  with  $M$  elements  $U_i, i=1,2,\dots,M$ , we obtain an unbiased estimate of the mean as follows:

$$\mu = \text{Sum}_i U_i / M,$$

while an unbiased estimate  $s^2$  of the variance is obtained from the formula

$$s^2 = \text{Sum}_i (U_i - \mu)^2 / (M-1).$$

The **variance is the Standard deviation squared ( $SD^2$ )**. To calculate with Excel=**VAR(highlight the cells you want)**

### The Standard Deviation

The standard deviation is the square root of the variance.

### In teacher talk:

**In teacher talk: the SD (and the variance) are basically the same-- a measure of how far a set of values (grades) differs or varies from the mean score (average). So a large SD means their (student grades) scores are all over the place—or widely distributed. Some people are getting it, and some people are not. If SD is small (i.e., 5) your students grades are tightly bound (the average is 95 and most of the class got around a 95). If the SD is small (i.e., 5) but your mean is 35—then perhaps it might be best to reconsider teaching this content over—since the average was low and most students got a low grade. So what if the SD is high (a large range of grades) and the average is 63? Well here the activity is distinguishing “who got it and who didn’t”—if the same “who didn’t” students continue-you might need to look into other reasons (reading level, content level, prior knowledge, learning style of student etc...).**