

MATH 3307

Lesson 3

Another important question we want to answer about data is about its spread or dispersion. Roughly speaking, the **population standard deviation**,  $\sigma$ , tells the average distance that data values fall from the mean. The standard deviation is the square root of the **population variance**,  $\sigma^2$ . So, what is the variance? The variance is the average of the squared differences of the data values from the mean.

If  $N$  is the number of values in a population with mean  $\mu$ , and  $x_i$  represents each individual value in the population, then the variance is found by:

$$\sigma^2 = \frac{\sum_{i=1}^N (x_i - \mu)^2}{N}$$

And the population standard deviation is  $\sigma = \sqrt{\sigma^2}$

Most of the time we are not working with the entire population. Instead, we are working with a sample.

- Sample variance -  $s^2 = \frac{\sum_{i=1}^n (x_i - \bar{x})^2}{n-1}$
- Sample standard deviation -  $s = \sqrt{s^2}$

Example:

1. A statistics teacher wants to decide whether or not to curve an exam. From her class of 300 students, she chose a sample of 10 students and their grades were:

72, 88, 85, 81, 60, 54, 70, 72, 63, 43

Find the mean, variance and standard deviation for this sample.

To copy: 72, 88, 85, 81, 60, 54, 70, 72, 63, 43

In R Studio:

`var(list)` will give the variance

and

`sd(list)` will give the standard deviation.

2. Suppose the statistics teacher decides to curve the grades by adding 10 points to each score. What is the new mean, variance and standard deviation?

We can see from example 2 that adding the same value to all elements does not affect the variance (or standard deviation) of a set of data. What about multiplying?

3. Find the variance and the standard deviation for the following set of data (whose mean is 4.5)

3, 6, 2, 7, 4, 5

Now, multiply each value by 2. What is the new variance and the new standard deviation?