Standard Deviation and Variance

- Another important question we want to answer about data is about its spread or dispersion.
- Roughly speaking, the **population standard deviation**, σ, tells the average distance that data values fall from the mean.
- The **population variance**, σ^2, is the square of the population standard deviation.

- So, how do we find 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 μ, and x_i represents each individual value in the population, then the variance is found by:

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

- The population standard deviation is found by square-rooting the variance so,

$$\sigma = \sqrt{\sigma^2}$$

- 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 -

$$s = \sqrt{s^2}$$

Examples:

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.

$$\bar{x} = \frac{72+88+85+81+60+54+70+72+63+43}{10} = 68.8$$

$$s^2 = \frac{(72-68.8)^2+(88-68.8)^2+(85-68.8)^2+\cdots+(43-68.8)^2}{(10-1)} \approx 199.7$$

$$s = \sqrt{199.7} \approx 14.13$$

Examples:

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?

new $\bar{x} = 78.8$ $(68.8 + 10)$

new $s \approx 14.13$ } no change

new $s^2 \approx 199.7$

- 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?

Examples:

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

$\bar{x} = 4.5$ $s \approx 1.87$ $s^2 = 3.5$

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

new $\bar{x} = 9$ new $s \approx 3.74$ new $s^2 = 14$

$2(4.5)$ $2(1.87)$ $2^2(3.5) = 4(3.5)$

- Sometimes we want to compare the variation between two groups. The **coefficient of variation** can be used for this.
- The coefficient of variation is the **ratio** of the standard deviation to the mean. A smaller ratio will indicate less variation in the data.

Examples:

4. The following statistics were collected on two different groups of stock prices:

	Portfolio A	Portfolio B
Sample size	10	15
Sample mean	$52.65	$49.80
Sample standard deviation	$6.50	$2.95

What can be said about the variability of each portfolio?

A : $\frac{6.50}{52.65} = .123$

B : $\frac{2.95}{49.80} = .059$ → less variation