Name: _____ Date: _____

VARIATION WITHIN A DATA SET
COMMON CORE ALGEBRA I

Measures of central tendency give us numbers that describe the typical data value in a given data set. But, they do not let us know how much **variation** there is in the data set. Two data sets can have the same mean but look radically different depending on how varied the numbers are in the set.

Exercise #1: The two data sets below each have equal means but differ in the variation within the data set. Use your calculator to determine the **Interquartile Range (IQR)** of each data set. The IQR is defined as the difference between the third quartile value and the first quartile value.

Data Set #1: 3, 3, 4, 4, 5, 5, 6, 6, 7, 8, 8, 9, 9, 10, 10, 11, 11

Data Set #2: 5, 5, 6, 6, 7, 7, 8, 8, 9, 9

The interquartile range gives a good measure of how spread out the data set is. But, the best measure of variation within a data set is the **standard deviation**. The actual calculation of standard deviation is complex and we will not go into it here. We will rely on our calculators for its calculation.

Exercise #2: Using the same data sets above, use your calculator to produce the standard deviation (shown as σ_x on the calculator) of the two data sets. Round your answers to the nearest tenth.

Data Set #1:   Data Set #2:

Standard Deviation

The **standard deviation** of a data set tells us, on **average**, how **far** a **data point** is **away** from the **mean** of the data set. The **larger** the **standard deviation**, the **greater** the **variation** within the data set.

Exercise #3: A farm is studying the weight of baby chickens (chicks) after 1 week of growth. They find the weight, in ounces, of 20 chicks. The weights are shown below. Find the mean, the interquartile range and the standard deviation for this data set. Round any non-integer values to the nearest tenth. Include appropriate units in your answers. Give an interpretation of the standard deviation.

2, 1, 3, 4, 2, 2, 3, 1, 5, 3, 4, 4, 5, 6, 3, 8, 5, 4, 6, 3

mean   interquartile range (IQR)   standard deviation

Exercise #4: A marketing company is trying to determine how much diversity there is in the age of people who drink different soft drinks. They take a **sample** of people and ask them which soda they prefer. For the two sodas, the age of those people who preferred them is given below.

Soda A: 18, 16, 22, 16, 28, 18, 21, 38, 22, 29, 25, 44, 36, 27, 40

Soda B: 25, 22, 18, 30, 27, 19, 22, 28, 25, 19, 23, 29, 26, 18, 20

(a) Explain why standard deviation is a better measure of the diversity in age than the mean.

(b) Which soda appears to have a greater diversity in the age of people who prefer it? How did you decide on this?

(c) Use your calculator to determine the **sample standard deviation**, normally given as s_x, for both data sets. Round your answers to the nearest tenth. Did this answer reinforce your pick from (b)? How?

### Population Versus Sample Standard Deviation

When we are working with **every possible data point** of interest, we call this a **population** and use the population standard deviation, σ. When we have only a **sample** of all possible values we use the sample standard deviation, s. The formulas for these two differ very slightly, so their values tend to be slightly different.

Exercise #5: Which of the following data sets would have a standard deviation (population) closest to zero? Do this without your calculator. Explain how you arrived at your answer.

(1) $\{-5, -2, -1, 0, 1, 2, 5\}$ (3) $\{11, 11, 12, 13, 13\}$

(2) $\{5, 8, 10, 16, 20\}$ (4) $\{3, 7, 11, 11, 11, 18\}$

Name: _____   Date: _____

VARIATION WITHIN A DATA SET
COMMON CORE ALGEBRA I HOMEWORK

1. For each of the following data sets, use your calculator to help find the interquartile range and the population standard deviation. Show your calculation for the IQR. Round all non-integer values to the nearest *tenth*.

   (a) 4, 6, 8, 10, 15, 19, 22, 25   (b) 3, 3, 4, 5, 5, 6, 6, 7, 7, 8

2. For the data set shown in the dot plot below, which of the following is closest to its population standard deviation?

   (1) 2.7   (3) 3.3

   (2) 4.2   (4) 5.8

3. What is the interquartile range of the data set represented in the box plot shown below?

   (1) 24   (3) 8

   (2) 14   (4) 12

4. Which of the following best measures the average distance that a data value lies away from the mean?

   (1) mean   (3) median

   (2) standard deviation   (4) range

5. Which of the following data sets would have the largest standard deviation?

   (1) $\{3, 3, 4, 5, 5\}$ (3) $\{2, 8, 18, 26, 35\}$

   (2) $\{72, 73, 74, 75, 76\}$ (4) $\{8, 10, 12, 14, 16\}$

6. We are going to revisit our survey of households that have video enabled devices (televisions, smart phones, tablets, etcetera). Recall that two surveys were done, each with 30 participants. In the first case (Survey A), the survey was random, in the second case (Survey B), the survey only included families with at least one teenager. The dot plots of the results are shown below.
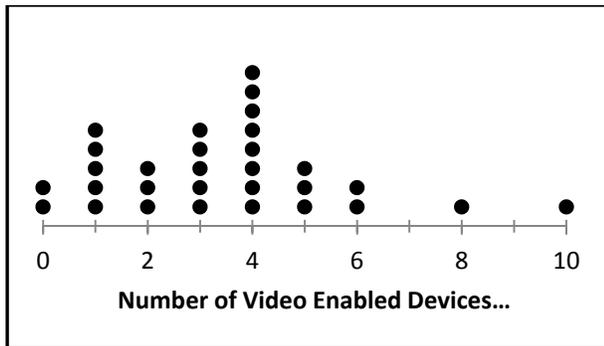
Survey A
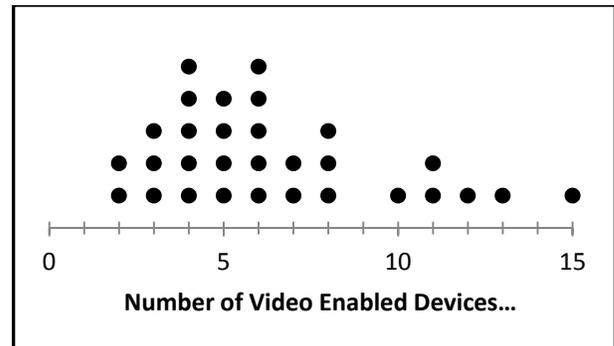
Survey B

(a) Enter the data into your calculator and use it to calculate the mean number of devices, the interquartile range, and the standard deviation of both data sets. Round all non-integers to the nearest tenth. Remember, you will have to enter a given data point more than once. For example, in Survey A, you will need to enter 2-0's, 5-1's, 3-2's, etcetera. Use the **sample standard deviation**.

Survey A Statistics: **Survey B Statistics:**

(b) Which of these two survey data sets had the greatest variation in the data? Explain based on the statistics you found in part (a).

(c) How many of the 30 values in Survey B fall within one standard deviation of the mean? To do this calculation, add the standard deviation and subtract the standard deviation from the mean and then count the number of values between the results of this addition and subtraction.