S.ID.A.2 Central Tendency and Dispersion

GRAPHS AND STATISTICS

S.ID.A.2: Central Tendency and Dispersion

A. Summarize, represent, and interpret data on a single count or measurement variable

2. Use statistics appropriate to the shape of the data distribution to compare center (median, mean) and spread (interquartile range, standard deviation) of two or more different data sets.

Overview of Lesson
- activate prior knowledge and review learning objectives (see above)
- explain vocabulary and/or big ideas associated with the lesson
- connect assessment practices with curriculum
- model an assessment problem and solution strategy
- facilitate guided discussion of student activity
- facilitate guided practice of student activity

Selected problem set(s)

- facilitate a summary and share out of student work

Homework – Write the Math Assignment

Measures of Central Tendency

A **measure of central tendency** is a *summary statistic* that indicates the typical value or center of an organized data set. The three most common measures of central tendency are the *mean*, *median*, and *mode*.

Mean A measure of central tendency denoted by $\overline{x}$, read "x bar", that is calculated by adding the data values and then dividing the sum by the number of values. Also known as the arithmetic mean or arithmetic average. The algebraic formula for the mean is:

$$\text{Mean} = \frac{Sum\ of\ \ items}{Count} = \frac{x_1 + x_2 + x_3 + \ldots + x_n}{n}$$

Median A measure of central tendency that is, or indicates, the middle of a data set when the data values are arranged in ascending or descending order. ***If there is no middle number, the median is the average of the two middle numbers*.**

Examples:
The **median** of the set of numbers: {2, 4, 5, 6, 7, 10, 13} is 6
The **median** of the set of numbers: {6, 7, 9, 10, 11, 17} is 9.5

Quartiles:
Q1, the **first quartile**, is the middle of the lower half of the data set.
Q2, the **second quartile**, is also known as the **median**.
Q3, the **third quartile**, is the middle of the upper half of the data set.

NOTE: To computer Q1 and Q2, find the middle numbers in the lower and upper halves of the data set. The median itself is not included in either the upper or the lower halves of the data set. When the data set contains an even number of elements, the median is the average of the two middle numbers and is excluded from the lower and upper halves of the data set.

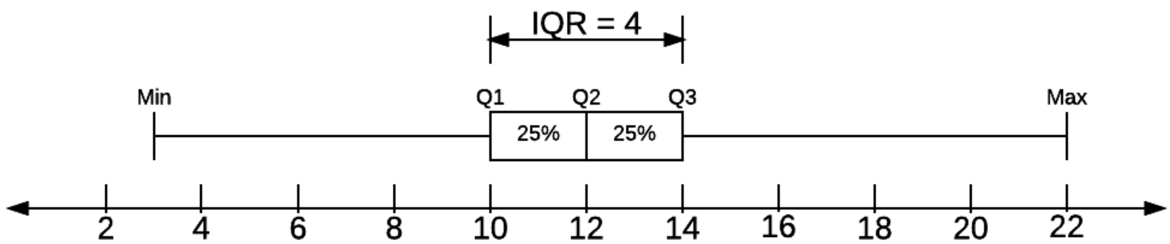
Mode A measure of central tendency that is given by the data value(s) that occur(s) most frequently in the data set.

Examples:
The **mode** of the set of numbers {5, 6, 8, 6, 5, 3, 5, 4} is 5.
The **modes** of the set of numbers {4, 6, 7, 4, 3, 7, 9, 1,10} are 4 and 7.
The **mode** of the set of numbers {0, 5, 7, 12, 15, 3} is none or there is no mode.

Measures of Spread

Interquartile Range: The difference between the first and third quartiles; a measure of variability resistant to outliers.

$$IQR = Q3 - Q1$$



Standard Deviation: A measure of variability. **Standard deviation** measures the average distance of a data element from the mean. There are two types of **standard deviations**: population and sample.

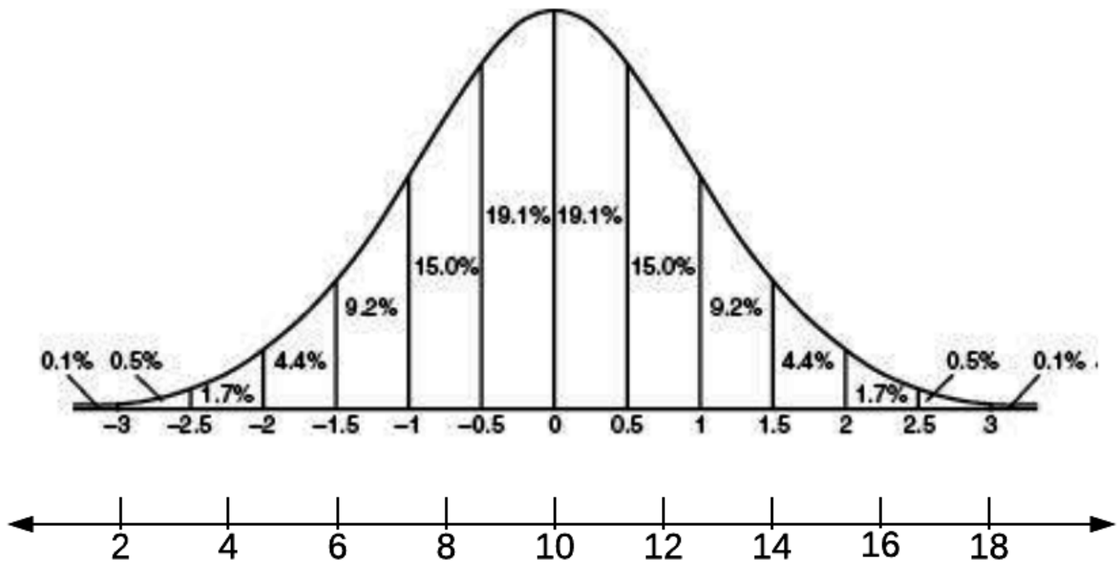
Population Standard Deviation: If data is taken from the *entire population*, divide by n when averaging the squared deviations. The following is the formula for **population standard deviation**:

$$\sigma = \sqrt{\frac{\sum (x_i - \overline{x})^2}{n}}$$

Sample Standard Deviation: If data is taken from a *sample* instead of the *entire population*, divide by $n-1$ when averaging the squared deviations. This results in a larger standard deviation. The following is the formula for **sample standard deviation**:

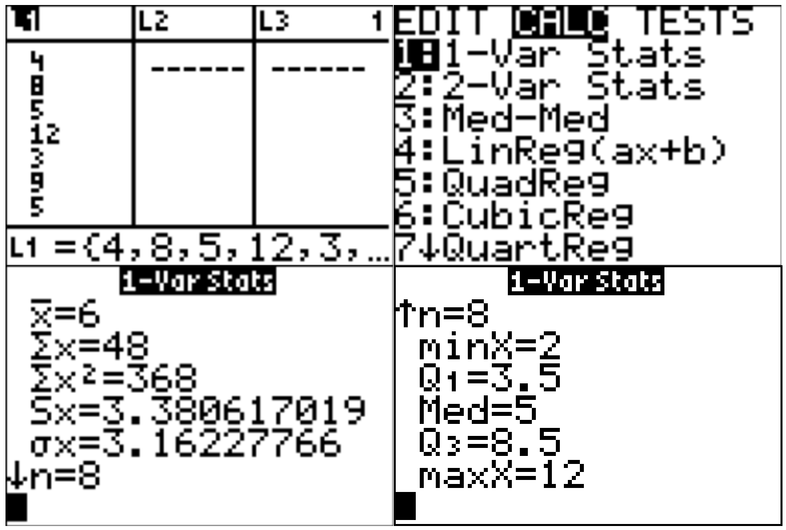
$$s = \sqrt{\frac{\sum (x_i - \overline{x})^2}{n-1}}$$

Standard Deviations and the Normal Curve



- When a data set is normally distributed, there are more elements closer to the mean and fewer elements further away from the mean.
- The normal curve shows the distribution of elements based on their distance from the mean.
- Three standard deviation units above the mean and three standard deviation units below the mean will include approximately 98.8% of all elements in a normally distributed data set.
- Each standard deviation above or below the mean corresponds to a specific value in the data set.
    - In the above example, the distance associated with each standard deviation unit corresponds to a distance of approximately $2\frac{2}{3}$ units on the scale below the curve.
- Many things in nature, such as height, weight, and intelligence, are normally distributed.

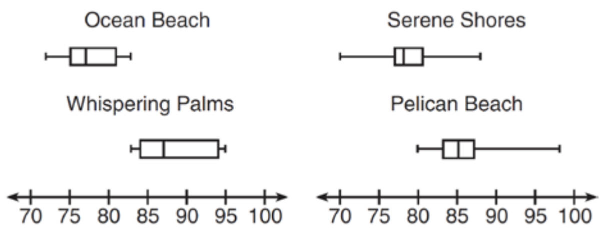
<u>Tips for Computing Standard Deviations</u>:
Use the STATS funtion of a graphing calculator to calculate standard deviation. Remember that the sample standard deviation (s) will be larger than the population standard deviation (σ).
INPUT VALUES: 4, 8, 5, 12, 3, 9, 5, 2
1. Use STATS EDIT to input the data set.
2. Use STATS CALC 1-Var Stats to calculate standard deviations.



The outputs include:
$\overline{X}$, which is the mean (average),
$\sum x$, which is the sum of the data set.
$\sum x^2$, which is the sum of the squares of the data set.
Sx, which is the sample standard deviation.
σx, which is the population standard deviation.
n, which is the number of elements in the data set
minX, which is the minimum value
Q2, which is the first quartile
Med, which is the median (second quartile)
Q3, which is the third quartile
maxX, which is the maximum value

REGENTS PROBLEMS TYPICAL OF THIS STANDARD

1. Corinne is planning a beach vacation in July and is analyzing the daily high temperatures for her potential destination. She would like to choose a destination with a high median temperature and a small interquartile range. She constructed box plots shown in the diagram below.



Which destination has a median temperature above 80 degrees and the smallest interquartile range?

a. Ocean Beach
b. Whispering Palms
c. Serene Shores
d. Pelican Beach

2. Christopher looked at his quiz scores shown below for the first and second semester of his Algebra class.
Semester 1: 78, 91, 88, 83, 94
Semester 2: 91, 96, 80, 77, 88, 85, 92
Which statement about Christopher's performance is correct?

a. The interquartile range for semester 1 is greater than the interquartile range for semester 2.
b. The median score for semester 1 is greater than the median score for semester 2.
c. The mean score for semester 2 is greater than the mean score for semester 1.
d. The third quartile for semester 2 is greater than the third quartile for semester 1.

3. The two sets of data below represent the number of runs scored by two different youth baseball teams over the course of a season.

Team *A*: 4, 8, 5, 12, 3, 9, 5, 2
Team *B*: 5, 9, 11, 4, 6, 11, 2, 7

Which set of statements about the mean and standard deviation is true?

a. mean A < mean B; standard deviation A > standard deviation B
b. mean A > mean B; standard deviation A < standard deviation B
c. mean A < mean B; standard deviation A < standard deviation B
d. mean A > mean B; standard deviation A > standard deviation B

S.ID.A.2 Central Tendency and Dispersion
Answer Section

1. ANS: D
Strategy: Eliminate wrong answers based on daily high temperatures, then eliminate wrong answers based on size of interquartile ranges.

Ocean Breeze and Serene Shores can be eliminated because they *do not* have median high temperatures above 80 degrees. Whispering Palms and Pelican Beach *do* have median high temperatures above 80 degrees, so the correct answer must be either Whispering Palms or Pelican Beach.

The interquartile range is defined as the differfence between the first and third quartiles. Pelican Beach has a much smaller interquartile range than Whispering Palms, so Pelican Beach is the correct choice.

PTS: 2 REF: 011514ai NAT: S.ID.A.2 TOP: Central Tendency and Dispersion

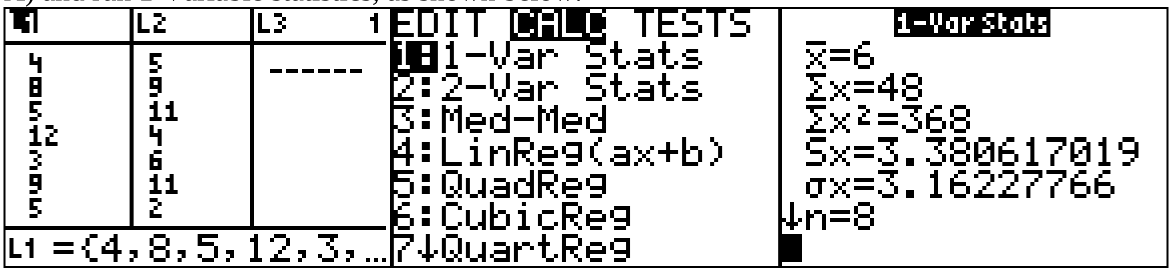
2. ANS: C
Strategy: Compute the mean, Q1, Q2, Q3, and interquartile range for each semester, then choose the correct answer based on the data.

	Mean	Q1	Median (Q2)	Q3	IQR
Semester 1	86.8	80.5	88	92.5	12
Semester 2	87	80	88	92	12

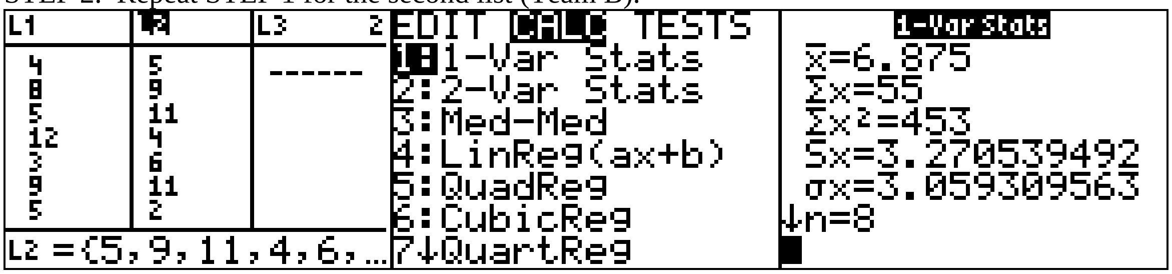
PTS: 2 REF: 061419ai NAT: S.ID.A.2 TOP: Central Tendency and Dispersion

3. ANS: A
Strategy: Compute the mean and standard deviations for both teams, then select the correct answer.

STEP 1. Enter the two sets of data into the STAT function of a graphing calculator, then select the first list (Team A) and run 1-Variable statistics, as shown below:



STEP 2. Repeat STEP 1 for the second list (Team B).



STEP 3. Use the data from the graphing calculator to choose the correct answer.

Choice a: mean A < mean B

$6 < 6.875$

standard deviation A > standard deviation B

$3.16227766 > 3.059309563$

Both statements in choice A are true.

A: $\bar{x} = 6$; $\sigma_x = 3.16$ B: $\bar{x} = 6.875$; $\sigma_x = 3.06$

PTS: 2 REF: 081519ai NAT: S.ID.A.2 TOP: Central Tendency and Dispersion

Homework - Write the Math Assignment

START	Write your name, date, topic of lesson, and class on your paper.	
	NAME:	Mohammed Chen
	DATE:	December 18, 2015
	LESSON:	Missing Number in the Average
	CLASS:	Z

PART 1a. Copy **the problem** from the lesson and underline/highlight key words.
PART 1b. State your understanding of **what the problem is asking**.
PART 1c. **Answer** the problem.
PART 1d. Explanation of **strategy** with all work shown.

PART 2a. Create **a new problem** that addresses the same math idea.
PART 2b. State your understanding of **what the new problem is asking**.
PART 2c. **Answer** the new problem.
PART 2d. Explanation of **strategy** used in solving the new problem with all work shown.

Clearly label each of the eight parts.

Grading Rubric

Each homework writing assignment is graded using a four point rubric, as follows:

Part 1. The Original Problem	**Up to 2** points will be awarded for: a) correctly restating the original problem; b) explicitly stating what the original problem is asking; c) answering the original problem correctly; and d) explaining the math.
Part 2. My New Problem	**Up to 2** points will be awarded for: a) creating a new problem similar to the original problem; b) explicitly stating what the new problem is asking; c) answering the new problem correctly; and d) explaining the math.

This assignment/activity is designed to incorporate elements of Polya's four step universal algorithm for problem solving with the idea that writing is thinking. Polya's four steps for solving any problem are:

1. Read and understand the problem.
2. Develop a strategy for solving the problem.
3. Execute the strategy.
4. Check the answer for reasonableness.

EXEMPLAR OF A WRITING THE MATH ASSIGNMENT

Part 1a. **The Problem**

TOP Electronics is a small business with five employees. The mean (average) weekly salary for the five employees is $360. If the weekly salaries of four of the employees are $340, $340, $345, and $425, what is the salary of the fifth employee?

Part 1b. **What is the problem asking?**

Find the salary of the fifth employee.

Part 1c. **Answer**

The salary of the fifth employee is $350 per week.

Part 1d. **Explanation of Strategy**

The arithmetic mean or average can be represented algebraically as:

$$\overline{X} = \frac{x_1 + x_2 + ... + x_n}{n}$$

I put information from the problem into the formula. The problem says there are 5 employees, so $n = 5$. The problem also gives the mean (average) salary and the salaries of 4 of the employees. These numbers can be substituted into the formula as follows:

$$360 = \frac{340 + 340 + 345 + 425 + x_5}{5}$$

$$1800 = 340 + 340 + 345 + 425 + x_5$$

$$1800 = 1450 + x_5$$

$$1800 - 1450 = x_5$$

$$350 = x_5$$

Check: $360 = \frac{340 + 340 + 345 + 425 + 350}{5} = \frac{1800}{5} = 360$

Part 2a. **A New Problem**

Joseph took five math exams this grading period and his average score on all of the exams is 88. He remembers that he received test scores of 78, 87, 94, and 96 on four of the examinations, but he has lost one examination and cannot remember what he scored on it. What was Joseph's score on the missing exam?

Part 2b **What is the new problem asking?**

Find Joseph's score on the missing exam.

Part 2c. **Answer to New Problem**

Joseph received a score of 85 on the missing examination.

Part 2d. **Explanation of Strategy**

I substitute information from the problem into the formula for the arithmetic mean, as follows:

$$88 = \frac{78 + 87 + 94 + 96 + x_5}{5}$$

$$440 = 355 + x_5$$

$$85 = x_5$$

$$88 = \frac{78 + 87 + 94 + 96 + 85}{5} = \frac{440}{5} = 88$$

The answer makes sense.