# Technical Assistance Document

# 2009 Algebra I Standard of Learning A.9

**Acknowledgements**
The Virginia Department of Education wishes to express sincere thanks to J. Patrick Lintner, Donna Meeks, Dr. Marcia Perry, Amy Siepka, Mike Traylor, and Dr. Daniel Yates for their assistance in the development of this document.

Questions about this document may be addressed to Michael Bolling at Michael.Bolling@doe.virginia.gov or (804) 786-6418.

**Virginia Department of Education**
**April 2010**

## Standard of Learning A.9

The student, given a set of data, will interpret variation in real-world contexts and calculate and interpret mean absolute deviation, standard deviation, and z-scores.

## Introduction

The purpose of this document is to assist teachers as they provide instruction and assess students on 2009 Standard of Learning (SOL) A.9. A.9 is intended to extend the study of descriptive statistics beyond the measures of center studied during the middle grades. **Although calculation is included in SOL A.9, the instruction and assessment emphasis should be on understanding and interpreting statistical values associated with a data set including standard deviation, mean absolute deviation, and z-score.** While not explicitly included in this SOL, the arithmetic mean (2009 SOL 5.16), will be integral to the study of descriptive statistics. In Algebra II, students will continue the study of descriptive statistics as they analyze the normal distribution curve and normally distributed data and apply statistical values to determine probabilities associated with areas under the standard normal curve. **The notation and formulas highlighted in shaded boxes will be used on the Algebra I End-of-Course (EOC) SOL assessment and included on the formula sheet for the Algebra I EOC SOL assessment beginning in the spring of 2012.**

## Definitions, formulas, and notation

The study of statistics includes gathering, displaying, analyzing, interpreting, and making predictions about a larger group of data (population) from a sample of those data. Data can be gathered through scientific experimentation, surveys, and/or observation of groups or phenomena. Numerical data gathered can be displayed numerically or graphically (examples would include line plots, histograms, and stem-and-leaf plots).

### Sample vs. Population Data

Sample data can be collected from a defined statistical population. Examples of a statistical population might include *SOL scores of all Algebra I students in Virginia, the heights of every U.S. president,* or *the ages of every mathematics teacher in Virginia*. Sample data can be analyzed to make inferences about the population. A data set, whether a sample or population, is comprised of individual data points referred to as elements of the data set.

> An element of a data set will be represented as $x_i$, where $i$ represents the $i^{\text{th}}$ term of the data set.

When beginning to teach this standard, teachers may want to start with small, defined population data sets of approximately 30 items or less to assist in focusing on development of understanding and interpretation of statistical values and how they are related to and affected by the elements of the data set.

Related to the discussion of samples versus populations of data are discussions about notation and variable use. In formal statistics, the arithmetic mean of a population is represented by the Greek letter $\mu$ (mu), while the calculated arithmetic mean of a sample is represented by $\overline{x}$, read "x bar."

The arithmetic mean of a data set will be represented by $\mu$ .

On both brands of approved graphing calculators in Virginia, the calculated arithmetic mean of a data set is represented by $\overline{x}$ .

**Mean Absolute Deviation vs. Variance and Standard Deviation**

Statisticians like to measure and analyze the dispersion (spread) of the data set about the mean in order to assist in making inferences about the population. One measure of spread would be to find the sum of the deviations between each element and the mean; however, this sum is always zero. There are two methods to overcome this mathematical dilemma: 1) take the absolute value of the deviations before finding the average or 2) square the deviations before finding the average. The mean absolute deviation uses the first method and the variance and standard deviation uses the second. If either of these measures is to be computed by hand, teachers should not require students to use data sets of more than about 10 elements. **Please note that students have not been introduced to summation notation prior to Algebra I.** An introductory lesson on how to interpret the notation will be necessary.

Examples of summation notation

$$\sum_{i=1}^{5} i = 1+2+3+4+5 \qquad \sum_{i=1}^{4} x_i = x_1 + x_2 + x_3 + x_4$$

**Mean Absolute Deviation**

Mean absolute deviation is one measure of spread about the mean of a data set, as it is a way to address the dilemma of the sum of the deviations of elements from the mean being equal to zero. The mean absolute deviation is the arithmetic mean of the absolute values of the deviations of elements from the mean of a data set.

Mean absolute deviation $= \dfrac{\sum_{i=1}^{n} |x_i - \mu|}{n}$, where $\mu$ represents the mean of the data set, $n$ represents the number of elements in the data set, and $x_i$ represents the $i^{\text{th}}$ element of the data set.

The mean absolute deviation is less affected by outlier data than the variance and standard deviation. Outliers are elements that fall at least 1.5 times the interquartile range ($IQR$) below the first quartile ($Q_1$) or above the third quartile ($Q_3$). Graphing calculators identify $Q_1$ and $Q_3$ in the list of computed 1-varible statistics. Mean absolute deviation cannot be directly computed on the graphing calculator as can the standard deviation. The mean absolute deviation must be computed by hand or by a series of keystrokes using computation with lists of data. More information (keystrokes and screenshots) on using graphing calculators to compute this can be found later in this document.

**Variance**

The second way to address the dilemma of the sum of the deviations of elements from the mean being equal to zero is to square the deviations prior to finding the arithmetic mean. The average of the squared deviations from the mean is known as the variance, and is another measure of the spread of the elements in a data set.

> Variance $\left(\sigma^2\right)=\frac{\sum_{i=1}^{n}\left(x_i-\mu\right)^2}{n}$, where $\mu$ represents the mean of the data set, *n* represents the number of elements in the data set, and $x_i$ represents the $i^{\text{th}}$ element of the data set.

The differences between the elements and the arithmetic mean are squared so that the differences do not cancel each other out when finding the sum. When squaring the differences, the units of measure are squared and larger differences are "weighted" more heavily than smaller differences. In order to provide a measure of variation in terms of the original units of the data, the square root of the variance is taken, yielding the standard deviation.

**Standard Deviation**

The standard deviation is the positive square root of the variance of the data set. The greater the value of the standard deviation, the more spread out the data are about the mean. The lesser (closer to 0) the value of the standard deviation, the closer the data are clustered about the mean.

> Standard deviation $\left(\sigma\right)=\sqrt{\frac{\sum_{i=1}^{n}\left(x_i-\mu\right)^2}{n}}$, where $\mu$ represents the mean of the data set, *n* represents the number of elements in the data set, and $x_i$ represents the $i^{\text{th}}$ element of the data set.

Often, textbooks will use two distinct formulas for standard deviation. In these formulas, the Greek letter "$\sigma$", written and read "sigma", represents the standard deviation of a population, and "*s*" represents the sample standard deviation. The population standard deviation can be estimated by calculating the sample standard deviation. The formulas for sample and population standard deviation look very similar except that in the sample standard deviation formula, *n*-1 is used instead of *n* in the denominator. The reason for this is to account for the possibility of greater variability of data in the population than what is seen in the sample. When *n*-1 is used in the denominator, the result is a larger number. So, the calculated value of the sample standard deviation will be larger than the population standard deviation. As sample sizes get larger (*n* gets larger), the difference between the sample standard deviation and the population standard deviation gets smaller. The use of *n*-1 to calculate the sample standard deviation is known as Bessel's correction. As a reminder, in A.9 students and teachers should use the formula for standard deviation with *n* in the denominator as noted in the shaded box above.

When using Casio or Texas Instruments (TI) graphing calculators to compute the standard deviation for a data set, two computations for the standard deviation are given, one for a population (using *n* in the denominator) and one for a sample (using $n-1$ in the denominator). Students should be asked to use the computation of standard deviation for population data in instruction and assessments. On a Casio calculator, it is indicated with "$x\sigma$n" and on a TI graphing calculator as "$\sigma$ x". More information (keystrokes and screenshots) on using graphing calculators as an instructional tool for 2009 SOL A.9 can be found later in this document.

**z-Scores**

A z-score, also called a standard score, is a measure of position derived from the mean and standard deviation of the data set. In Algebra I, the z-score will be used to determine how many standard deviations an element is above or below the mean of the data set. It can also be used to determine the value of the element, given the z-score of an unknown element and the mean and standard deviation of a data set. The z-score has a positive value if the element lies above the mean and a negative value if the element lies below the mean. A z-score associated with an element of a data set is calculated by subtracting the mean of the data set from the element and dividing the result by the standard deviation of the data set.

> z-score$\left(z\right)=\dfrac{x-\mu}{\sigma}$, where $x$ represents an element of the data set, $\mu$ represents the mean of the data set, and $\sigma$ represents the standard deviation of the data set.

A z-score can be computed for any element of a data set; however, they are most useful in the analysis of data sets that are normally distributed. In Algebra II, z-scores will be used to determine the relative position of elements within a normally distributed data set, to compare two or more distinct data sets that are distributed normally, and to determine percentiles and probabilities associated with occurrence of data values within a normally distributed data set.

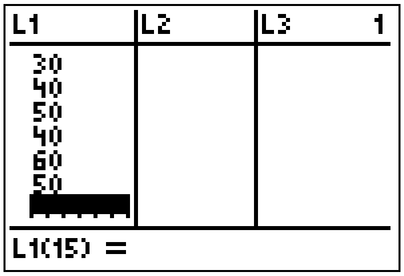
## Computation of descriptive statistics using graphing calculators – Example 1

> **Maya's company produces a special product on 14 days only each year. Her job requires that she report on production at the end of the 14 days. She recorded the number of products produced each day (below) and decided to use descriptive statistics to report on product production.**
>
> **{40, 30, 50, 30, 50, 60, 50, 50, 30, 40, 50, 40, 60, 50}**

### Computation of Variance and Standard Deviation

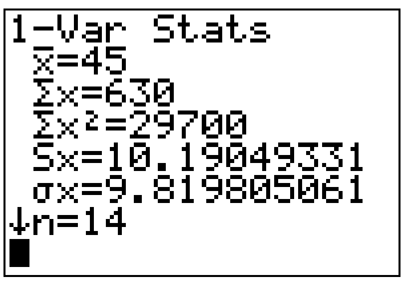
| **Texas Instruments (TI-83/84)** | **Casio (9750/9850/9860)** |
|---|---|
| <u>**Enter the data into L1**</u><br><br>• From the home screen click on [STAT].<br>• To enter data into lists choose option "1:Edit" or [ENTER].<br>• In L1, enter each element of the data set. Enter the first element and press [ENTER] to move to the next line and continue | <u>**Enter the data into List 1**</u><br><br>• From the menu screen select [STAT] and press [EXE].<br>• In List 1, enter the data set. Enter the first element and press [EXE] to move to the next line and continue until all the elements have been entered into List 1. |

until all the elements have been entered into L1.



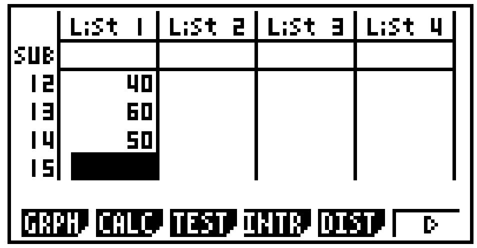

**Calculate variance and standard deviation by computing 1-Variable Statistics for L1**

- Press STAT ▸
- Choose option "1: 1-Var Stats" or ENTER
- Press ENTER to compute the 1-variable statistics (defaults to L1, enter list name after "1-Var Stats" if data are in another list).




$\bar{x}$ = arithmetic mean of the data set
$\Sigma x$ = sum of the $x$ values
$\Sigma x^2$ = sum of the $x^2$ values
$Sx$ = sample standard deviation
$\sigma x$ = population standard deviation
$n$ = number of data points (elements)

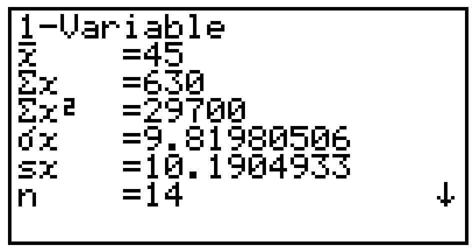
**NOTE: "$\sigma x$" will represent the standard deviation ($\sigma$). Squaring $\sigma$ will yield the variance ($\sigma^2$).**




Note: Screenshot from 9860/9750. The 9850 would not show the "SUB" row.

**Calculate variance and standard deviation by computing 1-Variable Statistics for List 1**

- Press F2 (CALC)
- Press F1 (1VAR)




$\bar{x}$ = arithmetic mean of the data set
$\Sigma x$ = sum of the $x$ values
$\Sigma x^2$ = sum of the $x^2$ values
$\sigma x$ = population standard deviation
$sx$ = sample standard deviation
$n$ = number of data points (elements)

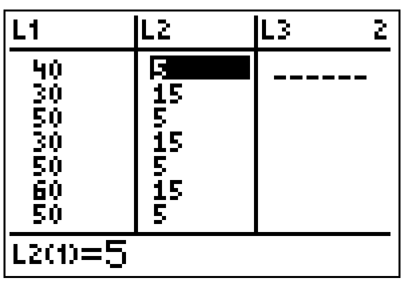
**NOTE: "$\sigma x$" will represent the standard deviation ($\sigma$). Squaring $\sigma$ will yield the variance ($\sigma^2$).**

## Computation of Mean Absolute Deviation

### Texas Instruments (TI-83/84)

**<u>Compute the mean absolute deviation using the data in L1</u>**

- From the home screen click on [STAT].
- Choose option "1:Edit" or [ENTER].
- Press [▸] to move to L2.
- Press [▴] to highlight L2.
- Press [ENTER] (to get "L2 =" at the bottom of the screen) [MATH] [▸]
- Choose option "1: abs (" or [ENTER]
- Press [2nd][1] (to get "L1")[−][VARS] choose option "5: Statistics..." and then choose option "2: $\bar{x}$"(to get "$\bar{x}$" or type in value)
- Press [)][ENTER] (L2 will automatically fill with data)




L2 now contains the $|x_i - \mu|$ part of the mean absolute deviation formula. In order to complete the calculation of the mean absolute deviation, find the arithmetic mean of L2 by calculating the 1-variable statistics of L2.

- Press [STAT][▸]
- Choose option "1: 1-Var Stats" or [ENTER]
- Press [2nd][2](to get "L2") [ENTER] to compute the 1-variable statistics for L2.

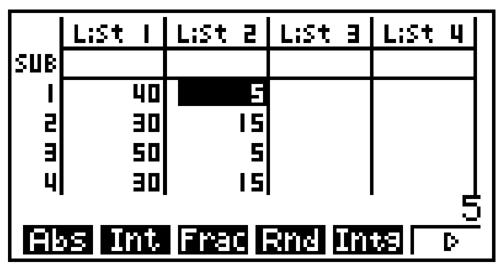
### Casio (9750/9850/9860)

**<u>Compute the mean absolute deviation using the data in List 1</u>**

- Continuing from the calculation of standard deviation, press [EXIT] until the function button menus read "GRAPH CALC TEST INTR DIST".
- Press [▶] to move the cursor to List 2.
- Press [▲] until "List 2" is highlighted.
- Press [OPTN]

  **[9750/9860]** [F6] (more-arrow right) [F1] (NUM) [F1] (Abs)
  **[9850]** [F4] (NUM) [F1] (Abs)

- Press [(] [OPTN] [F1] [F1] (to get "List") [1] [−] [VARS] [F3] [F5] [F2] (to get "$\bar{x}$" or type in value) [)] [EXE] (List 2 will automatically fill with data)

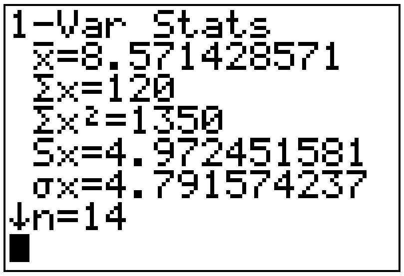



List 2 now contains the $|x_i - \mu|$ part of the mean absolute deviation formula. In order to complete the calculation of the mean absolute deviation, find the arithmetic mean of List 2 by calculating the 1-variable statistics of List 2.

- Continue to press [EXIT] until the function button menus read "GRAPH CALC TEST INTR DIST"
- Press [F2] (CALC) [F6] (SET)

  **[9750/9860]** [F1] [2] (to set it to List 2)
  **[9850]** [F2] (List 2)

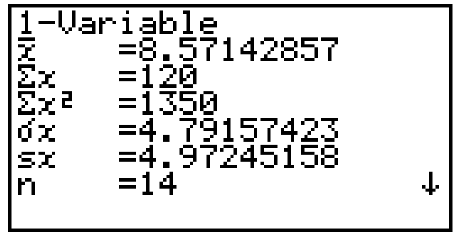
- Press [EXE] [EXIT].




$\bar{x}$ = 8.571428571 from the 1-variable statistics of L2 represents the mean absolute deviation of the original data set recorded in L1.

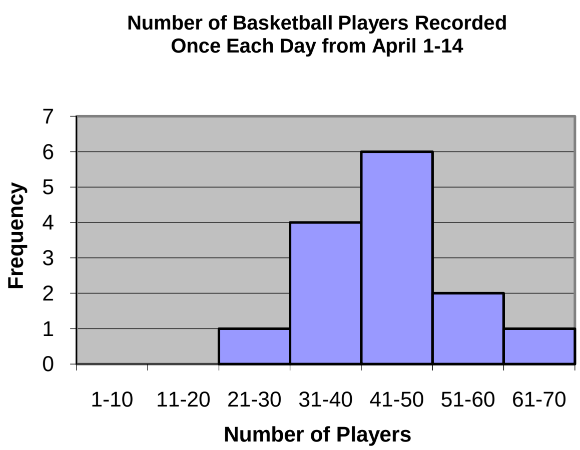
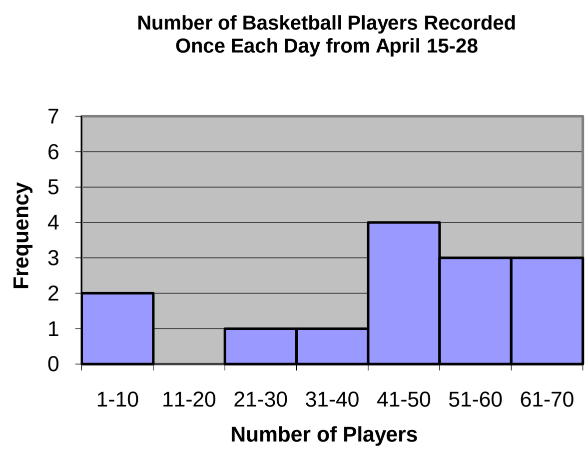
- Press

  **[9750/9860]** F1 (1VAR)
  **[9850]** F2 (CALC) F1 (1VAR)




$\bar{x}$ = 8.57142857 from the 1-variable statistics of List 2 represents the mean absolute deviation of the original data set recorded in List 1.

## Interpretation of descriptive statistics – Example 2

| Data set 1 | Data set 2 |
|---|---|
|   |   |
| **Mean = 45.0**<br>**Variance = 106.1**<br>**Standard Deviation = 10.3**<br>**Mean Absolute Deviation = 9.1** | **Mean = 45.0**<br>**Variance = 420.3**<br>**Standard Deviation = 20.5**<br>**Mean Absolute Deviation = 16** |

**What types of inferences can be made about the data set with the given information?**

- The standard deviation of data set 1 is less than the standard deviation of data set 2. That tells us that there was less variation (more consistency) in the number of people playing basketball during April 1-14 (data set 1).
- When comparing the variance of the two data sets, the difference between the two indicates that data set 2 has much more dispersion of data than data set 1.
- The standard deviation of data set 2 is almost twice the standard deviation of data set 1, indicating that the elements of data set 2 are more spread out with respect to the mean.

- Given a standard deviation and graphical representations of different data sets, the standard deviation could be matched to the appropriate graph by comparing the spread of data in each graph.

- When a data set contains clear outliers (the elements with values of 1 and 2 in data set 2), the outlying elements have a lesser affect on the calculation of the mean absolute deviation than on the standard deviation.

**How can z-scores be used to make inferences about data sets?**

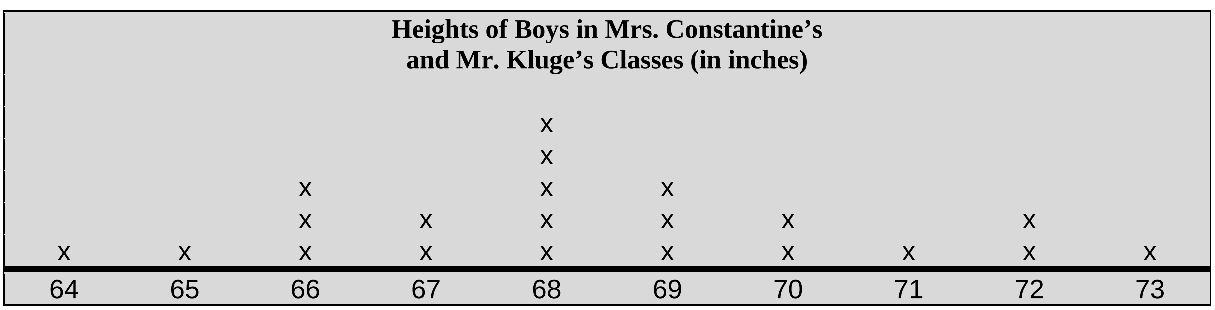
A z-score can be calculated for a specific element's value within the set of data. The z-score for an element with value of 30 can be computed for data set 2.

$$z = \frac{30 - 45}{20.5} = -0.73$$

The value of –0.73 indicates that the element falls just under one standard deviation below (negative) the mean of the data set. If the mean, standard deviation, and z-score are known, the value of the element associated with the z-score can be determined. For instance, given a standard deviation of 2.0 and a mean of 8.0, what would be the value of the element associated with a z-score of 1.5? Since the z-score is positive, the associated element lies above the mean. A z-score of 1.5 means that the element falls 1.5 standard deviations above the mean. So, the element falls 1.5(2.0) = 3.0 points above the mean of 8. Therefore, the z-score of 1.5 is associated with the element with a value of 11.0.

**Interpretation of descriptive statistics – Example 3**

**Maya represented the heights of boys in Mrs. Constantine's and Mr. Kluge's classes on a line plot and calculated the mean and standard deviation.**




**Mean = 68.4** **Standard Deviation = 2.3**

Note: In this problem, a small, defined population of the boys in Mrs. Constantine's and Mr. Kluge's classes is assumed.

**How many elements are above the mean?**
There are 9 elements above the mean value of 68.4.

**How many elements are below the mean?**
There are 12 elements below the mean value of 68.4.

**How many elements fall within one standard deviation of the mean?**
There are 12 elements that fall within one standard deviation of the mean. The values of the mean plus one standard deviation and the mean minus one standard deviation ( $\mu - \sigma = 66.1$ and $\mu + \sigma = 70.7$) determine how many elements fall within one standard deviation of the mean. In other words, all 12 elements between $\mu - \sigma$ and $\mu + \sigma$ (boys that measure 67", 68", 69", or 70") are within one standard deviation of the mean.

## Application scenarios

1. Dianne oversees production of ball bearings with a diameter of 0.5 inches at three locations in the United States. She collects the standard deviation of a sample of ball bearings each month from each location to compare and monitor production.

**Standard deviation of 0.5 inch diameter ball bearing production (in inches)**

| | July | August | September |
|---|---|---|---|
| Plant location #1 | 0.01 | 0.01 | 0.02 |
| Plant location #2 | 0.02 | 0.04 | 0.05 |
| Plant location #3 | 0.02 | 0.01 | 0.01 |

Compare and contrast the standard deviations from each plant location, looking for trends or potential issues with production. What conclusions or questions might be raised from the statistical data provided? What other statistical information and/or other data might need to be gathered in order for Dianne to determine next steps?

*Sample response: Plant #1 and Plant #3 had standard deviations that seemed steady, but one would be wise to keep an eye on Plant #1 in the coming months, because it had increases in August and September. A potential concern with the standard deviation of Plant #2 exists. The growing standard deviation indicates that there might be an issue with growing variability in the size of the ball bearings. Plant #2 should be asked to take more frequent samples and continue to monitor, to check the calibration of the equipment, and/or to check for an equipment problem.*

2. Jim needs to purchase a large number of 20-watt florescent light bulbs for his company. He has narrowed his search to two companies offering the 20-watt bulbs for the same price. The Bulb Emporium and Lights-R-Us claim that their 20-watt bulbs last for 10,000 hours. Which descriptive statistic might assist Jim in making the best purchase? Explain why it would assist him.

*Sample response: The standard deviation of the lifespan of each company's 20-watt bulbs should be compared. The bulbs with the lowest lifespan standard deviation will have the slightest variation in number of hours that the bulbs last. The bulbs with the slightest variation in the number of hours that they last means that they are more likely to last close to 10,000 hours.*

3. In a school district, Mr. Mills is in charge of SAT testing. In a meeting, the superintendent asks him how many students scored less than one standard deviation below the mean on the mathematics portion of the SAT in 2009. He looks through his papers and finds that the mean of the scores is 525 and 1653 students took the SAT in 2009. He also found a chart with percentages of z-scores on the SAT in 2009 as follows:

| **z-score (mathematics)** | **Percent of students** |
|---|---|
| z < - 3 | 0.1 |
| -3 ≤ z < -2 | 2.1 |
| -2 ≤ z < -1 | 13.6 |
| -1 ≤ z < 0 | 34.0 |
| 0 ≤ z < 1 | 34.0 |
| 1 ≤ z < 2 | 13.6 |
| 2 ≤ z < 3 | 2.1 |
| z > 3 | 0.1 |

How can Mr. Mills determine the number of students that scored less than one standard deviation below the mean on the mathematics portion of the SAT?

*Sample response: The z-score tells you how many standard deviations an element (in this case a score) is from the mean. If the z-score of a score is -1, then that score is 1 standard deviation below the mean. There are 15.8% (13.6% + 2.1% + 0.1%) of the scores that have a z-score < -1, so there are about 261 (0.158* x *1653) students that scored less than one standard deviation below the mean.*

## Questions to explore with students

1. Given a frequency graph, a standard deviation of _______, and a mean of _______, how many elements fall within _______ standard deviation(s) from the mean? Why?

2. Given the standard deviation and mean or mean absolute deviation and mean, which frequency graph would most likely represent the situation and why?

3. Given two data sets with the same mean and different spreads, which one would best match a data set with a standard deviation or mean absolute deviation of _______? How do you know?

4. Given two frequency graphs, explain why one might have a larger standard deviation.

5. Given a data set with a mean of _______, a standard deviation of _______, and a z-score of _______, what is the value of the element associated with the z-score?

6. What do z-scores tell you about position of elements with respect to the mean? How do z-scores relate to their associated element's value?

7. Given the standard deviation, the mean, and the value of an element of the data set, explain how you would find the associated z-score.