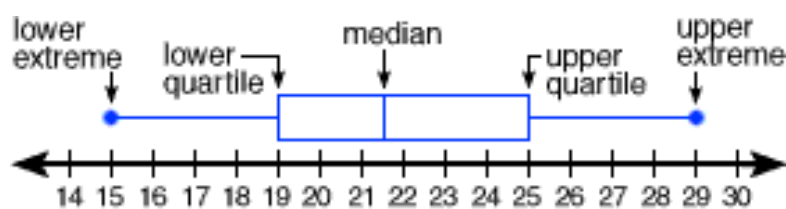
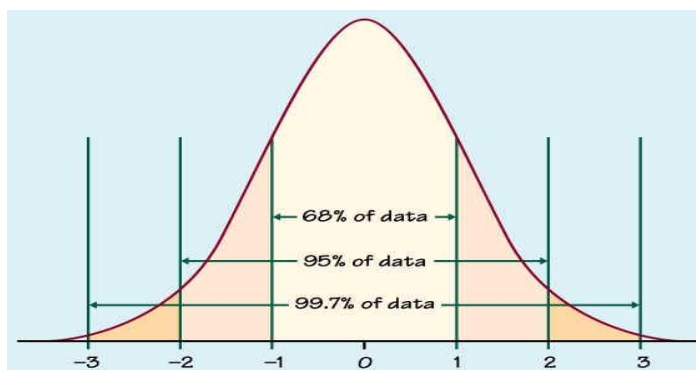


**Box Plot:** A box plot also known as a box-and-whisker plot, is a graphical representation of the Five-Number Summary of a data set. A box is drawn from the lower quartile (Q1) to the upper quartile (Q3); a vertical line across the box indicates the median. Two whiskers are drawn, one from the lower quartile to the minimum and one from the upper quartile to the maximum. Box plots can be used to make graphical representations between data sets and to measure the variation within parts of a data set. (Also see Handout by Jennifer Nord)



**Empirical Guidelines:** Also known as the 68–95–99.7 rule. The Empirical Rule for any normal distribution states that

- 68% of data fall within one standard deviation of the mean.  
 $[\mu - \sigma, \mu + \sigma]$ .
- 95% of data fall within two standard deviations of the mean.  
 $[\mu - 2\sigma, \mu + 2\sigma]$ .
- 99.7% of data fall within three standard deviations of the mean.  
 $[\mu - 3\sigma, \mu + 3\sigma]$ .

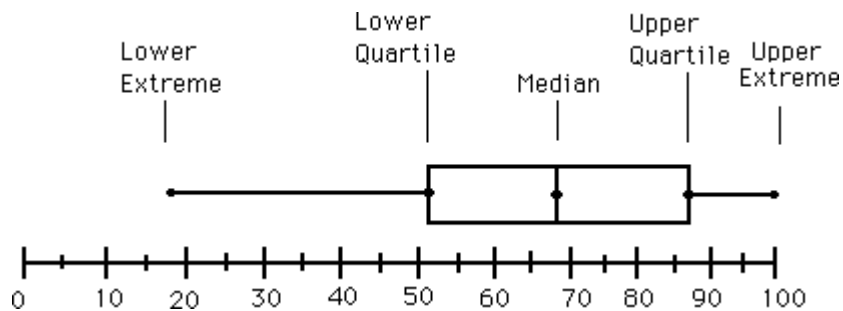


**Five-Number Summary:** The five-number summary of a data set is a five-item list comprising of

- 1.) Minimum value (Lower Extreme)
- 2.) Quartile I (Lower quartile)
- 3.) Median
- 4.) Quartile II (Upper Quartile)
- 5.) Maximum Value (Upper Extreme)

**Note 1:** The Five number summary divides the data set into four sets, each of which contains 25% of the data.

**Note 2:** The Five-Number Summary is used to construct a box-and-whisker plot



**Example:** The body temperatures of 18 adults were measured, resulting in the following values:

98.2 97.2 97.4 97.6 97.8 98.0 98.2 98.2 98.2  
98.2 98.4 98.4 98.5 98.6 98.6 99.0 99.2 99.7

Data sorted from smallest to largest:

97.1 97.2 97.4 97.6 97.8 98.0 98.2 98.2 98.2  
98.4 98.4 98.5 98.6 98.6 99.0 99.2 99.7

The Five-Number Summary for this data set is:

Min = 97.1    Q1 = 97.7    Med = 98.2    Q3 = 98.6    Max = 99.7

**Interquartile Range (IQR):** The Interquartile Range (IQR) is the range between the first and third quartiles. It is a measure of statistical dispersion. The interquartile range is a more stable statistic than the (total) range, and is often preferred to the latter statistic.

*Note 1:* The interquartile range is the length of the interval between the lower quartile (Q1) and the upper quartile (Q3). . This interval indicates the central or middle, 50% of a data set.

*Note 2:* The IQR is the length of the box in a box-and-whisker plot.

*Example:* Data Set: {2, 3, 4, 5, 6, 6, 6, 7, 7, 8, 9}  
Third Quartile (Q3) = 7  
First Quartile (Q1) = 4  
IQR (Q3-Q1)  $7 - 4 = 3$

**IQR Test:** The IQR Test is used to determine outliers for median based statistics. An outlier will be classified as any data value that lays more than 1.5 times length of the box from either end of the box.

*Example:* Consider the data set:  
The lower quartile = Q1 = 56.  
The upper quartile = Q3 = 87  
IQR =  $87 - 56 = 31$   
 $Q3 + 1.5IQR = 87 + 1.5(31) = 133.5$   
 $Q1 - 1.5 IQR = 56 - 1.5(31) = 9.5$

Therefore, if any data is more than 133.5 or less than 9.5 it would be considered an outlier for that set of data.

**Lower Quartile (or First Quartile, Q1):** In a distribution, the value that separates the lower quarter or 25% of the data values from the upper three-quarters of the data values. (See Quartiles)

*Note 1:* The lower quartile is the same as the 25<sup>th</sup> percentile

*Note 2:* The lower quartile is the median of the lower half of the data values.

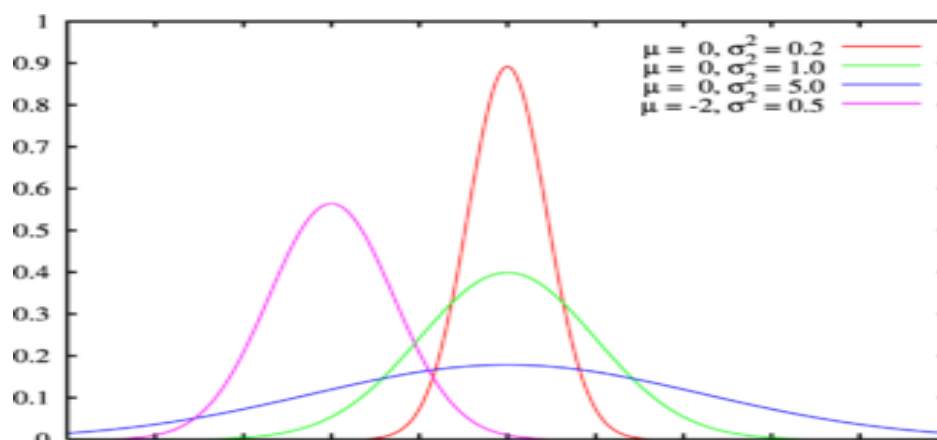
**Measures of Dispersion** are statistics that give an indication of how the data are “spread out” or distributed. Some common measures of dispersion include:

- Range
- Interquartile Range
- Standard Deviation
- Variance

**Normal Distribution:** A useful probability distribution which has a symmetric, bell or mound shape and tails extending infinitely far in both directions.

*Note 1:* To speak specifically of any normal distribution, two quantities have to be specified: the mean,  $\mu$ , where the peak of the distribution occurs, and the standard deviation,  $\sigma$ , which indicates the spread or girth of the bell curve.

*Note 2:* Because normal distributions are symmetric and unimodal, the mean, the median and the mode are all identical.



**Outlier**: An outlier is an extreme value of the data. It is an observation value that is significantly different from the rest of the data.

*Note 1*: For median based statistics an outlier is defined to be any value of the data that is more than 1.5 IQR units below the lower quartile or 1.5 IQR units above the upper quartile. See IQR test.

*Note 2*: For a normal distribution an outlier would typically be a data value that is more than two standard deviations away from the mean.

**Percentile**: Any of the 100 equal parts into which the range of values of a set of data can be divided in order to show the distribution of those values. The percentile of a given value is determined by the percentage of the values that are smaller than that value. A score in the  $x$ th percentile is a score that is greater than  $x\%$  of the scores attained within a given set.

*Example 1*: If the 10th percentile of a set of 100 numbers has the value 27, then 10 percent of the numbers (10 numbers) are less than 27

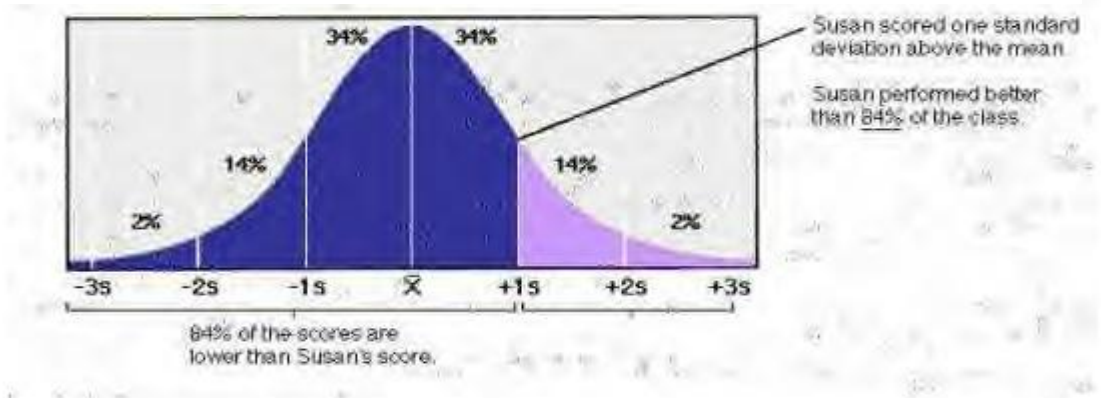
*Example 2*: If a baby is in the 70<sup>th</sup> percentile for height and the 45<sup>th</sup> percentile for weight then the baby is taller than 70% and heavier than 45% of the babies of the same age.

*Note 1*: The median of the set is the 50th percentile

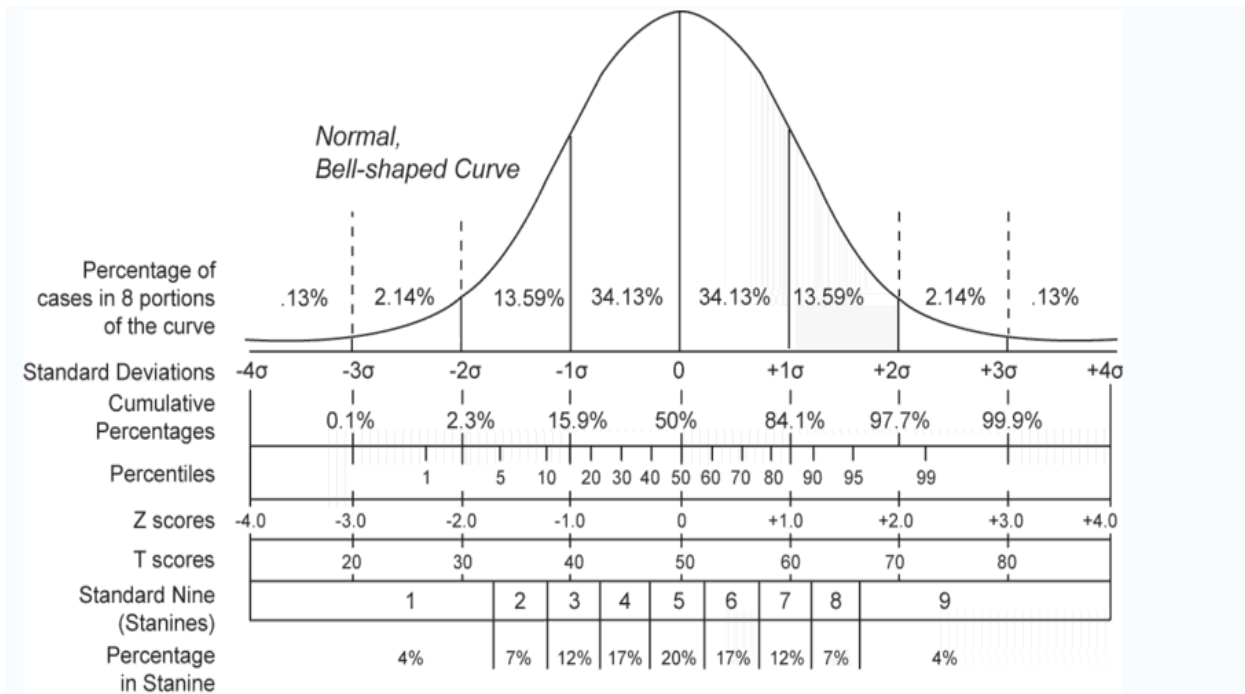
*Note 2*: To find a particular percentile for a normal distribution the **invNorm**( function on the TI83 or TI84 graphing calculator can be used.

**Percentiles on the Normal Curve:** Percentiles on the normal curve are measured from the left all the way to the data that is being measured.

**Example:** Suppose Susan scored one standard deviation above the mean. That is she is 34% above the mean, but to find her percentile also include the 50% below the mean, thus showing that Susan performed better than (34%+50%= 84%) 84% of the class. Susan would be in the 84<sup>th</sup> percentile



**Illustration:** The following diagram shows the relationship between the empirical guidelines and percentiles on the normal distribution:



**Quartile:** Quartiles are numbers that divide an ordered data set into four portions, each containing approximately one-fourth of the data. The total of 100% is broken into four equal parts: 25%, 50%, 75% 100%.

- 25% of the data values come before the first quartile (Q1)
- 50% of the data values come before the median or 2<sup>nd</sup> quartile (Q2).
- 75% of the data values come before the third quartile (Q3)

**Note1:** Quartiles are special types of percentiles.

Lower Quartile or first quartile (Q1) is the 25<sup>th</sup> percentile

Median Quartile or second quartile (Q2) is the 50<sup>th</sup> percentile

Upper Quartile or third quartile (Q3) is the 75<sup>th</sup> percentile.

**Note 2:** The quartiles lower, median and upper are part of the Five Number Summary.

**Note 3:** There are various methods used for determining the upper and lower quartiles of a set of data; refer to your local curriculum for the method that should be used with your students. If there is no recommended method for your district or state, then you can use the following process for identifying the upper and lower quartiles:

- Arrange the data in order from least to greatest, and identify the median.
- Identify the middle term of each half of the data on either side of the median. These values are the upper and lower quartiles.

**Example:**      Data Set                    {6, 47, 49, 15, 43, 41, 7, 39, 43, 41, 36}  
                  Ordered Data Set:    {6, 7, 15, 36, 39, 41, 41, 43, 43, 47, 49}

- The lower half is {6, 7, 15, 36, 39}, and the middle term of that half is 15. Therefore, the lower quartile is 15.
- The median (or 2<sup>nd</sup> quartile) is the middle term, 41.
- The upper half is {41, 43, 43, 47, 49}, and the middle term of that half is 43. Therefore, the upper quartile is 43.

**Range:** A measure of dispersion found by taking the difference of the minimum and maximum values of a data set.

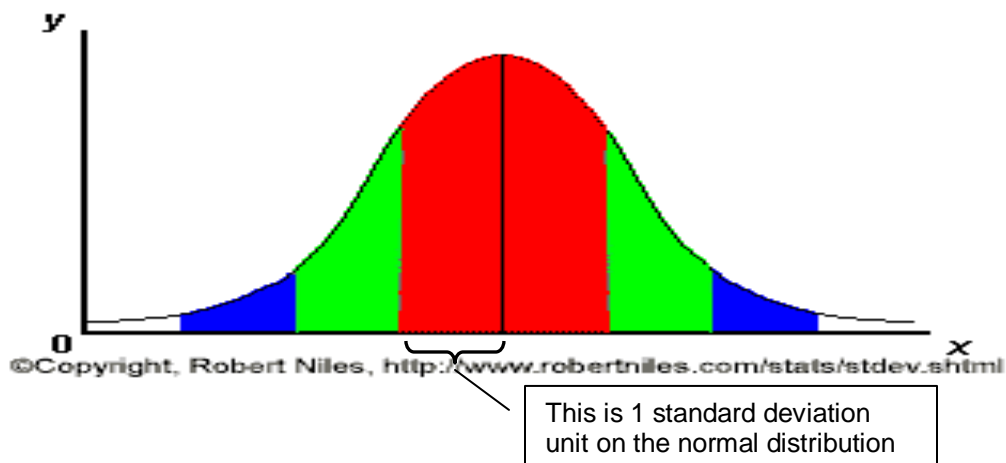
**Standard Deviation**: Standard deviation is a statistical measure of spread or dispersion within a sample or population. It is a way of expressing how different the numbers are from the average (mean). The standard deviation is roughly the amount by which the average of one data point differs from the average of all data points. In general, the greater the standard deviation, the more spread out the data is.

**Note 1:** The standard deviation allows us to compare several data sets to see which is more homogenous.

**Note 2:** The mean and the standard deviation can be used to compare individual student's performances on different tests relative to the class as a whole: (See standard score or z-score)

**Note 3:** The standard deviation is particularly useful in considering a normal distribution.

- If the data points are close to the mean and the bell-shaped curve is steep, the standard deviation is small.
- When the data points are spread far apart from the mean, the bell-shaped curve is relatively flat, that tells you that the standard deviation is large.
- If the data points are equal, then the standard deviation is zero.
- One standard deviation away from the mean in either direction on the horizontal axis accounts for somewhere around 68 percent of the data.
- Two standard deviations away from the mean, account for roughly 95 percent of the data.
- And three standard deviations account for about 99 percent of the people.



**Note 4:** Because the standard deviation is a mean based statistic it is greatly influenced by extreme values in the data and is NOT a resistant statistic.

**Note 5:** To calculate the standard deviation you take the square root of the variance. (See variance)

**Standard Normal Distribution**: The standard normal distribution is a normal distribution with a mean of zero and a standard deviation of one.

*Note*: Any normal distribution can be transformed into a standard normal distribution by using standard scores (z-scores).

**Standard Score**: The standard score or z-score for an observed value tells us how many standard deviations the observed value is from the mean – that is, it tells us how far the observed value is from the mean in standard deviations units. (See z-score)

*Note 1*: When a set of scores are converted to z-scores, the scores are said to be standardized and are referred to as standard scores.

*Note 2*: Standard scores based on the standard normal distribution which has a mean of 0 and a standard deviation of 1.

**Upper Quartile(or Third Quartile, Q3)**: In a distribution, the value that separates the lower three-quarters or 75% of the data values from the upper quarter of the data values. (See Quartiles)

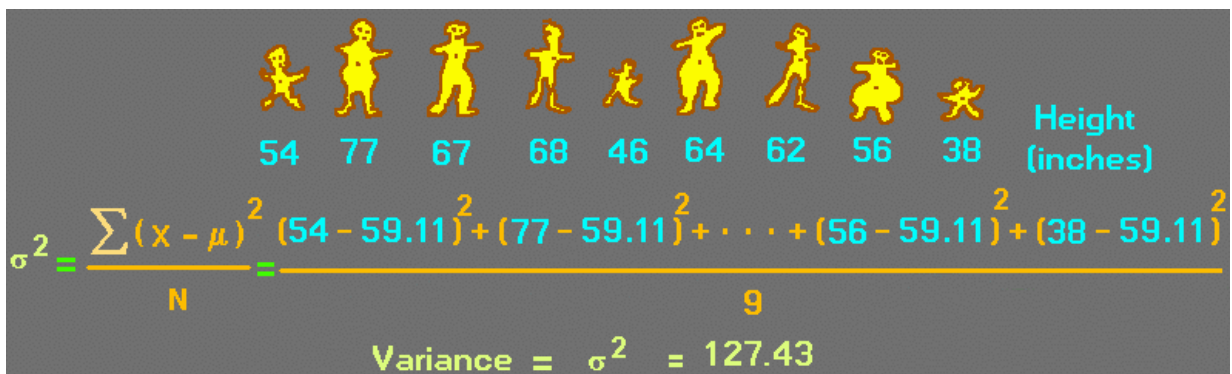
*Note 1*: The upper quartile is the same as the 75<sup>th</sup> percentile

*Note 2*: The upper quartile is the median of the upper half of the data values.

**Variance:** The variance is one of several indices of variability that statisticians use to characterize the dispersion among the measures in a given population.

To calculate the variance of a given population,

- Calculate the mean of the scores ( $\mu$ )
- Measure the amount that each score deviates from the mean ( $x - \mu$ )
- Square that deviation (by multiplying it by itself).  $(x - \mu)^2$
- Find average of the squared deviations from the mean.



**Note:** Variance is the square of standard deviation or standard deviation is the square root of variance.

**Z-score**: The  $z$ -score for an item, indicates how far and in what direction, that item deviates from its distribution's mean, expressed in units of its distribution's standard deviation.

**Note 1**: The  $z$ -score or **standard score** for an observed value tells us how many standard deviations the observed value is from the mean – that is, it tells us how far above or below the observed value is from the mean in standard deviation units.

**Note 2**: Numbers above the mean have positive  $z$ -scores,  
Numbers below the mean have negative  $z$ -score  
The mean has a  $z$ -score of zero.

**Note 3**: When a set of scores are converted to  $z$ -scores, the scores are said to be standardized and are referred to as **standard scores**.

**Note 4**: If  $x$  is the raw score to be standardized,  $\sigma$  the standard deviation of the population, and  $\mu$  the mean of the population, then the  $z$ -score can be found by:

$$z = \frac{x - \mu}{\sigma}$$

**Note 5**: The mathematics of the  $z$ -score transformation are such that if every item in a distribution is converted to its  $z$ -score, the transformed scores will necessarily have a mean of zero and a standard deviation of one.

**Note 6**:  $Z$ -scores allow us to compare the results of different normal distributions.

**Example**: Among college applicants who take the SAT, scores have a mean of 896 and a standard deviation of 174. Among college applicants who take the ACT, scores have a mean of 20.6 and a standard deviation of 5.2. Bobby scored 1080 on the SAT while Kathy score 28 on the ACT.

Bobby's  $z$ -score is  $(1080 - 896) / 174$  or approximately 1.05

Kathy's  $z$ -score is  $(28 - 20.6) / 5.2$  or approximately 1.42

Thus Kathy performed higher on a standardized scale.