***Normal Distribution – Determining # of Std. Deviations from Mean***

**The Characteristics of a Normal Distribution**

Think about the popcorn example in the introduction to this Chapter. The amount of popcorn popping starts small, gets larger, and then decreases as time goes by, much like the graph below.

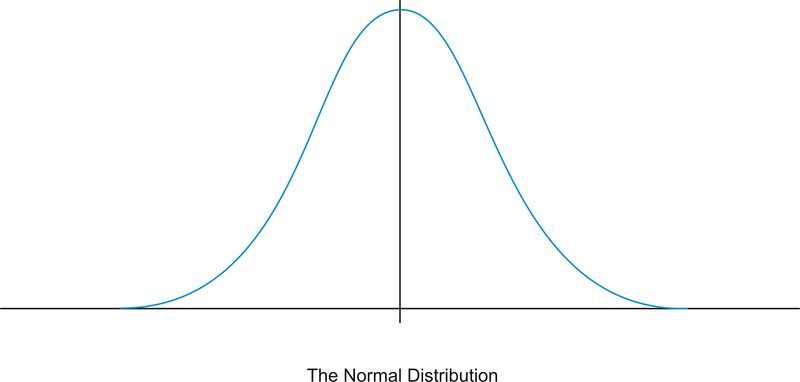


This graph is considered a normal curve, or a "bell curve" because it looks like a bell. Many data sets look similar to this when plotted. But are they all exactly the same? The answer is no; they may be centered at different values, and some may be more spread out than others. While still having this similar bell shape, there may be slight differences in the exact shape.

**Shape**

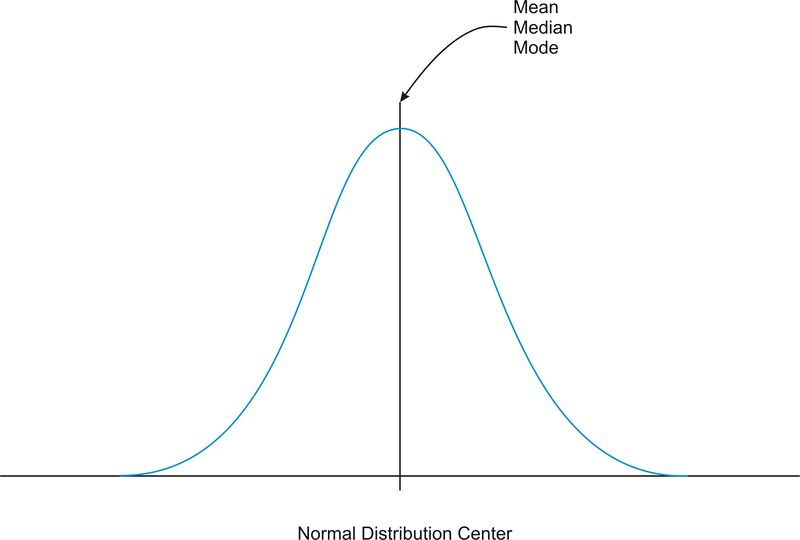
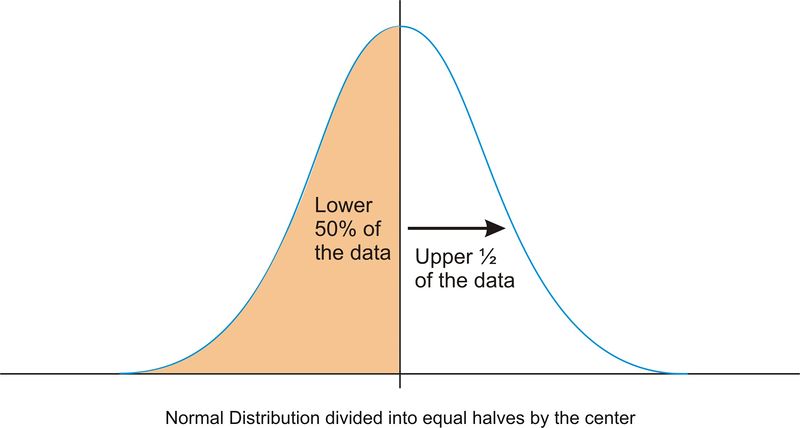
When graphing the data from each of the examples in the introduction, the distributions from each of these situations would be mound-shaped and mostly symmetric. A normal distributionis a perfectly symmetric, mound-shaped distribution.

Because so many real data sets closely approximate a normal distribution, we can use the idealized normal curve to learn a great deal about such data. With a practical data collection, the distribution will never be exactly symmetric, so just like situations involving probability, a true normal distribution only results from an infinite collection of data. Also, it is important to note that the normal distribution describes a continuous random variable.



**Center**

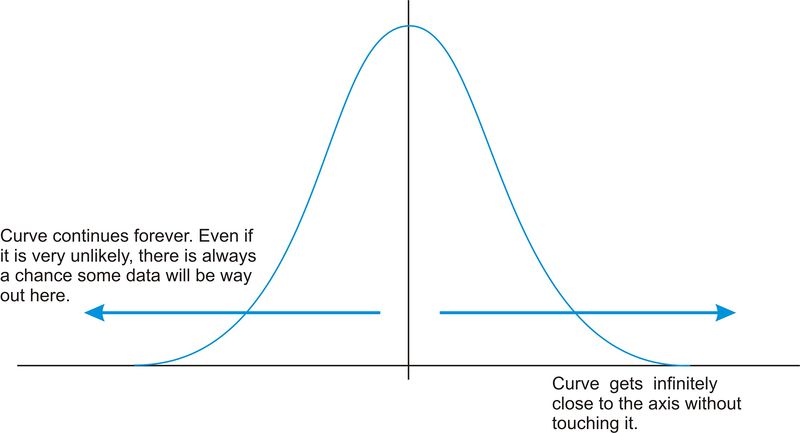
Due to the exact [symmetry](http://www.ck12.org/analysis/Symmetry) of a normal curve, the center of a normal distribution, or a data set that approximates a normal distribution, is located at the highest point of the distribution, and all the statistical measures of center we have already studied (the mean, [median](http://www.ck12.org/statistics/Median), and mode) are equal.



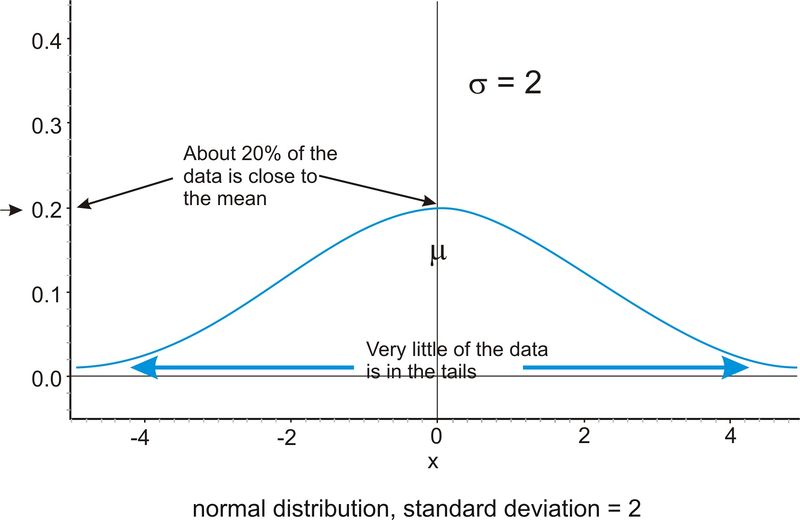
It is also important to realize that this center peak divides the data into two equal parts.

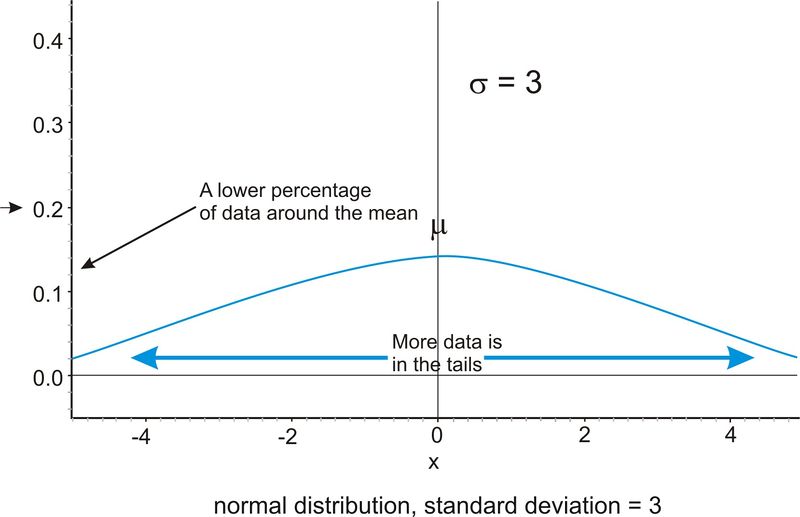
**Spread**

Let’s go back to our popcorn example. The bag advertises a certain time, beyond which you risk burning the popcorn. From experience, the manufacturers know when most of the popcorn will stop popping, but there is still a chance that there are those rare kernels that will require more (or less) time to pop than the time advertised by the manufacturer. The directions usually tell you to stop when the time between popping is a few seconds, but aren’t you tempted to keep going so you don’t end up with a bag full of un-popped kernels? Because this is a real, and not theoretical, situation, there will be a time when the popcorn will stop popping and start burning, but there is always a chance, no matter how small, that one more kernel will pop if you keep the microwave going. In an idealized normal distribution of a continuous random variable, the distribution continues infinitely in both directions.



Because of this infinite spread, the range would not be a useful statistical measure of spread. The most common way to measure the spread of a normal distribution is with the standard deviation, or the typical distance away from the mean. Because of the [symmetry](http://www.ck12.org/analysis/Symmetry) of a normal distribution, the standard deviation indicates how far away from the maximum peak the data will be. Here are two normal distributions with the same center (mean):

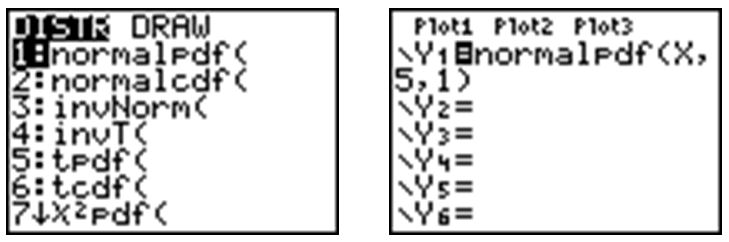




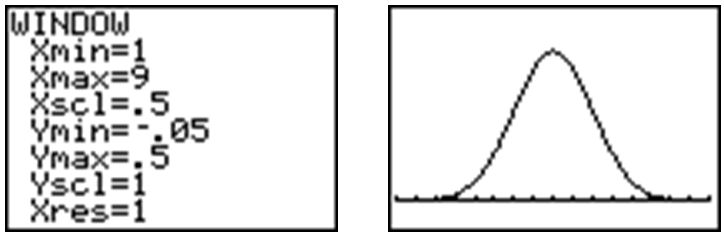
The first distribution pictured above has a smaller standard deviation, and so more of the data are heavily concentrated around the mean than in the second distribution. Also, in the first distribution, there are fewer data values at the extremes than in the second distribution. Because the second distribution has a larger standard deviation, the data are spread farther from the mean value, with more of the data appearing in the tails.

**Technology Note: Investigating the Normal Distribution on a TI-83/84 Graphing Calculator**

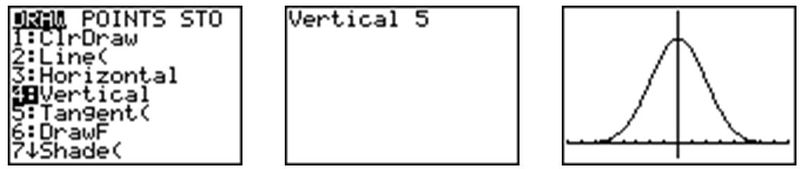
We can graph a normal curve for a probability distribution on the TI-83/84 calculator. To do so, first press **[Y=]**. To create a normal distribution, we will draw an idealized curve using something called a density function. The command is called 'normalpdf(', and it is found by pressing **[2nd][DISTR][1]**. Enter an X to represent the random variable, followed by the mean and the standard deviation, all separated by commas. For this example, choose a mean of 5 and a standard deviation of 1.



Adjust your window to match the following settings and press **[GRAPH]**.



Press **[2ND][QUIT]**to go to the home screen. We can draw a vertical line at the mean to show it is in the center of the distribution by pressing **[2ND][DRAW]**and choosing 'Vertical'. Enter the mean, which is 5, and press **[ENTER]**.

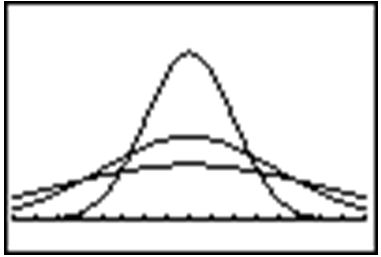


Remember that even though the graph appears to touch the x -axis, it is actually just very close to it.

In your **Y=**Menu, enter the following to graph 3 different normal distributions, each with a different standard deviation:

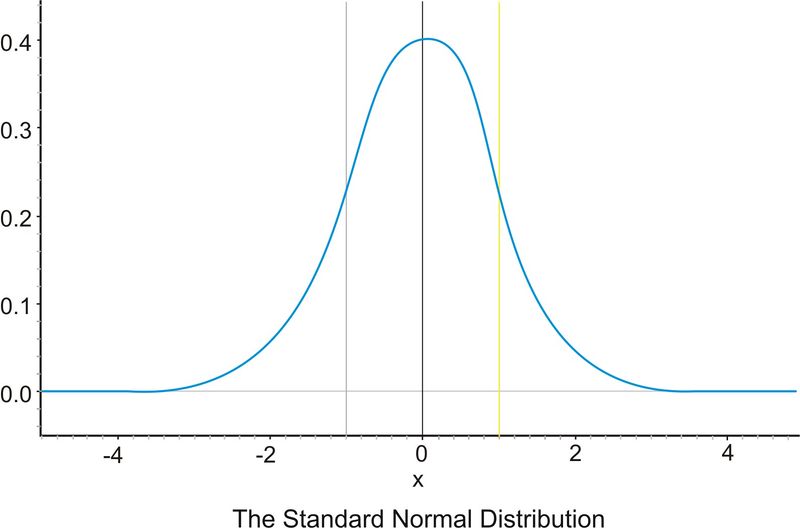


This makes it easy to see the change in spread when the standard deviation changes.

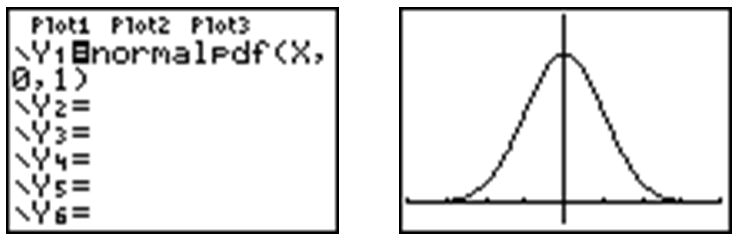


**The Empirical Rule**

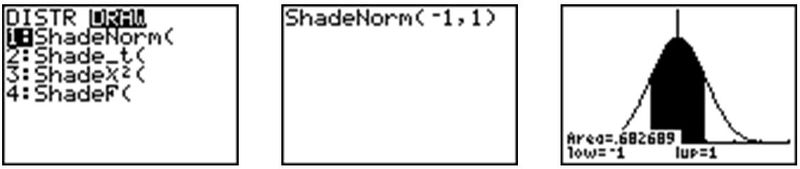
Because of the similar shape of all normal distributions, we can measure the percentage of data that is a certain distance from the mean no matter what the standard deviation of the data set is. The following graph shows a normal distribution with \mu=0 and \sigma=1 . This curve is called a standard normal curve. In this case, the values of x represent the number of standard deviations away from the mean.



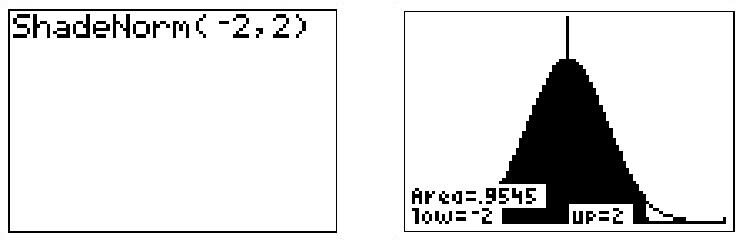
Notice that vertical lines are drawn at points that are exactly one standard deviation to the left and right of the mean. We have consistently described standard deviation as a measure of the typical distance away from the mean. How much of the data is actually within one standard deviation of the mean? To answer this question, think about the space, or area, under the curve. The entire data set, or 100% of it, is contained under the whole curve. What percentage would you estimate is between the two lines? To help estimate the answer, we can use a graphing calculator. Graph a standard normal distribution over an appropriate window.



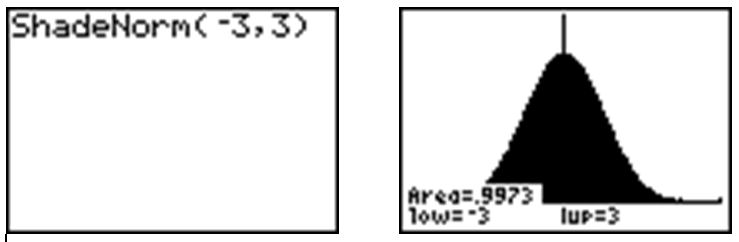
Now press **[2ND][DISTR]**, go to the **DRAW**menu, and choose 'ShadeNorm('. Insert ' -1 , 1' after the 'ShadeNorm(' command and press **[ENTER]**. It will shade the area within one standard deviation of the mean.



The calculator also gives a very accurate estimate of the area. We can see from the rightmost screenshot above that approximately 68% of the area is within one standard deviation of the mean. If we venture to 2 standard deviations away from the mean, how much of the data should we expect to capture? Make the following changes to the 'ShadeNorm(' command to find out:



Notice from the shading that almost all of the distribution is shaded, and the percentage of data is close to 95%. If you were to venture to 3 standard deviations from the mean, 99.7%, or virtually all of the data, is captured, which tells us that very little of the data in a normal distribution is more than 3 standard deviations from the mean.

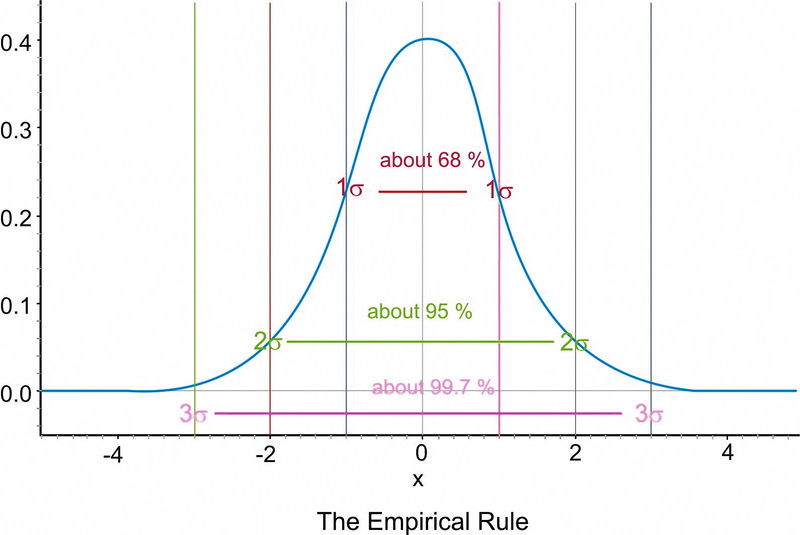


Notice that the calculator actually makes it look like the entire distribution is shaded because of the limitations of the screen resolution, but as we have already discovered, there is still some [area under the curve](http://www.ck12.org/analysis/Area-Under-the-Curve) further out than that. These three approximate percentages, 68%, 95%, and 99.7%, are extremely important and are part of what is called the Empirical Rule.

The [Empirical Rule](http://www.ck12.org/probability/Empirical-Rule) states that the percentages of data in a normal distribution within 1, 2, and 3 standard deviations of the mean are approximately 68%, 95%, and 99.7%, respectively.

**On the Web**

<http://tinyurl.com/2ue78u>Explore the [Empirical Rule](http://www.ck12.org/probability/Empirical-Rule).



* **-Scores**

A *z*-scoreis a measure of the number of standard deviations a particular data point is away from the mean. For example, let’s say the mean score on a test for your statistics class was an 82, with a standard deviation of 7 points. If your score was an 89, it is exactly one standard deviation to the right of the mean; therefore, your  z -score would be 1. If, on the other hand, you scored a 75, your score would be exactly one standard deviation below the mean, and your z -score would be  -1 . All values that are below the mean have negative z -scores, while all values that are above the mean have positive z -scores. A z -score of -2 would represent a value that is exactly 2 standard deviations below the mean, so in this case, the value would be  82 - 14 = 68 .

To calculate a z -score for which the numbers are not so obvious, you take the deviation and divide it by the standard deviation.

z = \frac{\text{Deviation}}{\text{Standard Deviation}}

You may recall that deviation is the mean value of the variable subtracted from the observed value, so in symbolic terms, the z -score would be:

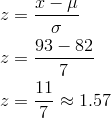
z=\frac{x-\mu}{\sigma}

As previously stated, since \sigma is always positive, z will be positive when x is greater than \mu and negative when x is less than  \mu . A z -score of zero means that the term has the same value as the mean. The value of z represents the number of standard deviations the given value of x is above or below the mean.

#### Example A

What is the z -score for an A on the test described above, which has a mean score of 82? (Assume that an  A is a 93.)

The z -score can be calculated as follows:

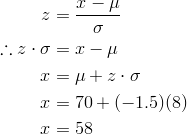


If we know that the test scores from the last example are distributed normally, then a z -score can tell us something about how our test score relates to the rest of the class. From the Empirical Rule, we know that about 68% of the students would have scored between a z -score of -1 and 1, or between a 75 and an 89, on the test. If 68% of the data is between these two values, then that leaves the remaining 32% in the tail areas. Because of [symmetry](http://www.ck12.org/analysis/Symmetry), half of this, or 16%, would be in each individual tail.

#### Example B

On a college entrance exam, the mean was 70, and the standard deviation was 8. If Helen’s*z*-score was *-1.5*, what was her exam score?

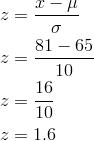
**Solution:**



### Guided Practice

On a nationwide math test, the mean was 65 and the standard deviation was 10. If Robert scored 81, what was his*z*-score?

**Solution:**



Robert's z -score is 1.6, which means that he scored 1.6 standard deviations above the mean.