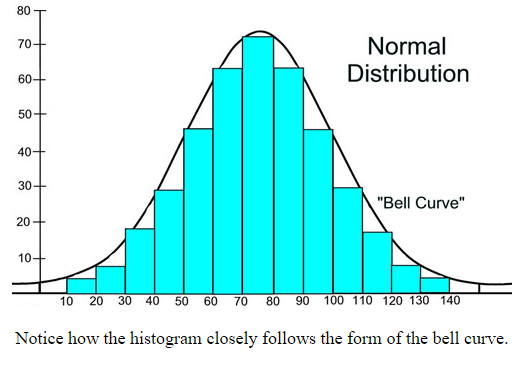
**Statistics Part -2**

**Normal Distribution**

Bell-shaped Curve:

There are certain sets of data where the data, when graphed, are symmetrical with a single central peak at the mean (average) of the data. The shape of the curve is described as bell-shaped with the graph falling off evenly on either side of the mean. Fifty percent of the distribution lies to the left of the mean and fifty percent lies to the right of the mean. Such graphs are called normal curves, and referred to as a normal distribution. The mean, median and mode are all the same in a normal distribution.

****

A normal distribution is the most widely known and used of all distributions. It is an extremely important statistical data distribution pattern occurring in many natural phenomena, such a blood pressure, machined parts, human height, error in measurement, IQ scores, sizes of snowflakes, lifespans of light bulbs, weights of loaves of bread, test scores, milk production in cows, etc. When data pertaining to these phenomena are graphed as histograms with data on the horizontal axis and the amount of data on the vertical axis, a bell-shaped curve (normal curve) may be created.

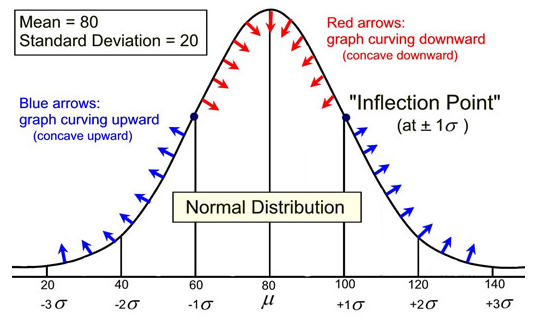
A normal distribution is actually a "*family of distributions*", since the mean and standard deviation, which determine the shape of the distribution, may differ from graph to graph.

|  |  |  |
| --- | --- | --- |
| |  | | --- | | Normal Distribution Basic Properties: 1. symmetric about the mean 2. the mean = the mode = the median 3. the mean divides the data in half 4. defined by mean and standard deviation 5. the curve is unimodal (one peak) 6. the curve approaches, but never touches, the *x*-axis, as it extends farther and farther away from the mean. 7. total area under the curve = 1. (Characteristics of *perfectly* normal distributions.) | | normalbunch2 |
| While the four normal curves shown above (at the right) share all of these basic properties, they are still unique (different from one another) as to mean and standard deviation. | |

Standard Deviation:

|  |  |
| --- | --- |
| A normal distribution can have any mean and any positive standard deviation. The mean determines the line of symmetry of the graph, and the standard deviation determines how much the data are spread out.  The smaller the standard deviation, the more oncentrated the data and narrower the raph. The larger the standard deviation, the more dispersed the data, and wider the graph. |  |
| **Population standard deviation**= *σ* (small case Greek sigma);  **population mean**= *μ* (small case Greek mu). | |

A Closer Look at the Shape of the Normal Curve:

****

The graph shown above states that for this normal distribution the mean is 80 and the standard deviation is 20. The graph also shows that the mean (80) + 1 standard deviation (20) equals 100, the mean (80) + 2 standard deviations (40) = 120, and so on, to both the left and right sides of the mean.

If you were given a normal curve, without being told the mean and the standard deviation, you could approximate this information based upon the shape of the curve.

• The mean in a normal curve divides the curve symmetrically. Therefore, the mean will pass through the highest point on the graph. In this example, it is logical to assume that the mean is 80.

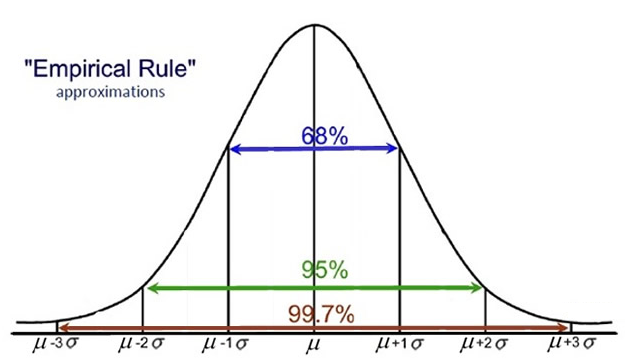
• In a normal curve, the point at which the graph changes from curving downward to curving upward is called an inflection point and occurs at plus (or minus) one standard deviation from the mean. Examining a normal curve for this location will yield an approximation of the value of the standard deviation. In this example, an inflection point can be seen to be occurring around 100, or approximately 20 points above the mean. The standard deviation could be approximated to be 20

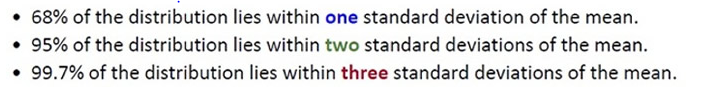
Percentages Under a Normal Curve:

As seen in the previous section, the standard deviation can be used to sub-divide the space (the *area*) under a normal curve, starting from the mean. Each of these sub-divided sections can be used to represent a portion (a *percentage*) of the data falling into these sections of the graph. The normal curve actually shows how likely it is to find a value within a specific distance from the mean.

|  |
| --- |
| Using 1 standard deviation to create the subdivisions: |

The most popular subdivision utilizes distances from the mean in increments of one standard deviation of that specific normal curve. When dealing with a normal curve:  
• *approximately*68% of the data will fall within one standard deviation of the mean  (between the mean minus one standard deviation and the mean plus one standard deviation),  
• *approximately* 95% of the data will fall within two standard deviations of the mean (between the mean minus two standard deviations and the mean plus two standard deviations),and  
• *approximately* 99.7% of the data will fall within three standard deviations of the mean (between the mean minus three standard deviations and the mean plus three standard deviations).  
These three facts make up what is referred to as the Empirical Rule (or the *68-95-99.7 Rule)*

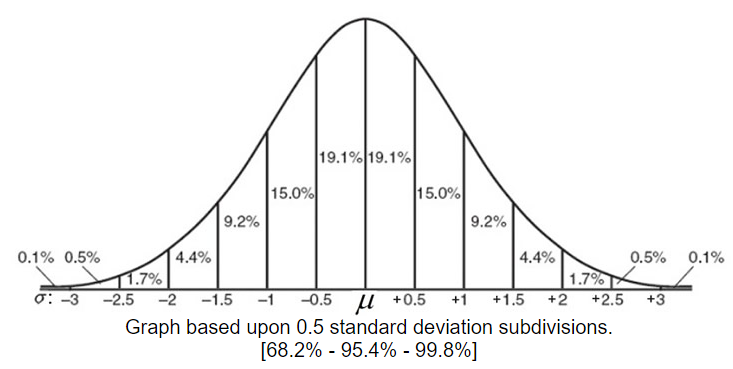
****



*NOTE:*Normal distributions may also be referred to as normal probability distributions.

|  |
| --- |
| Using ½ of 1 standard deviation to create the subdivisions: |

It is possible to subdivide the area under a normal curve into smaller intervals, such as widths of 0.5 standard deviations, as shown in the graph below. The addition of the percentages in this graph will be slightly different from the Empirical Rule values which are rounded approximations. These smaller subdivisions would be used when information presented in a question falls on the increments of one-half of one standard deviation from the mean.



beware   Not all data is normally distributed.

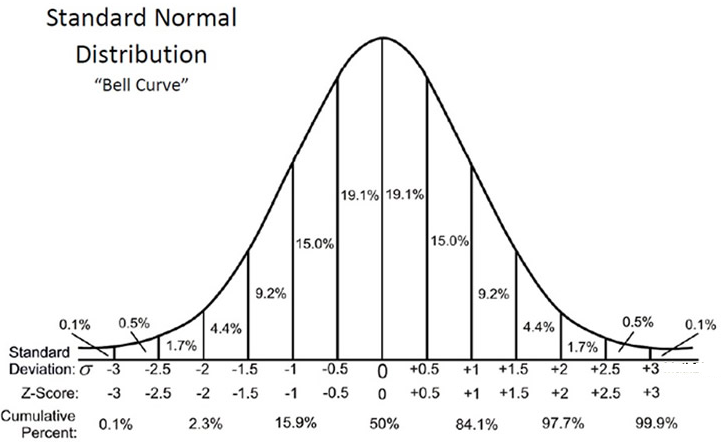
**Standard Normal Distribution**

|  |  |
| --- | --- |
| There are many different normal distributions, with each one depending on two parameters: the population mean, *μ*, and the population standard deviation, *σ*. Rather than performing computations on each new set of parameters for a variety of normal curves, it is easier to work in reference to the "simplest case" of the normal curves, called the *standard normal distribution*. A parameter is a numerical measurement describing some characteristic of a population. | normalbunch1 |

|  |  |  |
| --- | --- | --- |
| |  |  | | --- | --- | | definition | A standard normal distribution is a special case of the normal distribution. It is a normal distribution with a mean of zero and a standard deviation equal to one. The total area under its density curve is equal to 1. | |

In a standard normal distribution, the random variable, x, is called a standard score, or a z-score. The horizontal scale of the graph of the standard normal distribution corresponds to z-scores.

All normal distributions are equivalent to the standard normal distribution when the unit of measure is changed to measure standard deviations from the mean. The standard normal distribution can be used to deal with any problem involving any normal distribution. The process of converting a normal distribution into an equivalent standard normal distribution is covered under "[Understanding z-scores](http://mathbitsnotebook.com/Algebra2/Statistics/STzScores.html)".



Reading from the chart, it can be seen that approximately 19.1% of normally distributed data is located between the mean (the peak) and 0.5 standard deviations to the right (or left) of the mean.

Empirical Rule Still True!

As seen in the normal curve, the Empirical Rule *(68-95-99.7 Rule),*statesthat *approximately:*

|  |
| --- |
| *•*68% of the data will fall within one standard deviation of the mean. • 95% of the data will fall within two standard deviations of the mean. • 99.7% will fall within three standard deviations of the mean |

50% Data in the Middle

50% of the distribution lies within 0.67448 standard deviations of the mean (that is, "centered about the mean", or "in the middle") .

If you are asked for the interval about the mean containing 50% of the data, you are actually being asked for the interquartile range, IQR. When working with box plots, the IQR is computed by subtracting the first quartile from the third quartile.

In a standard normal distribution (with mean 0 and standard deviation 1), the first and third quartiles are located at -0.67448 and +0.67448 respectively. Thus the interquartile range (IQR) is 1.34896.

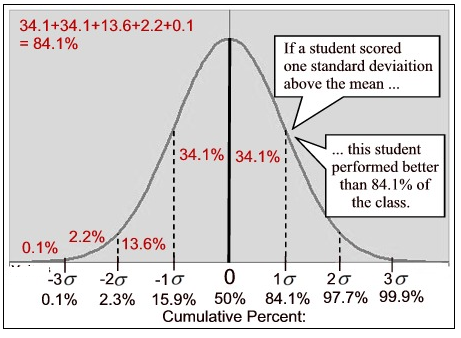
|  |
| --- |
| In a standard normal distribution: IQR = Q3 - Q1 = 0.67448- (-0.67448)= 1.34896 In **ANY** normal distribution: IQR = Q3 - Q1 = 0.67448*σ* - (-0.67448*σ*)= 1.34896*σ*  (Interquartile range = 1.34896 x standard deviation)  This will be the *population* IQR |

Percentiles and the Normal Curve

The mean (at the center peak of the curve) is the 50th percentile (50%).

The term "percentile rank" refers to the area (probability) to the left of the value.

Adding the given percentages from the chart will let you find certain percentiles along the curve.

****

**Understanding Z-Scores**

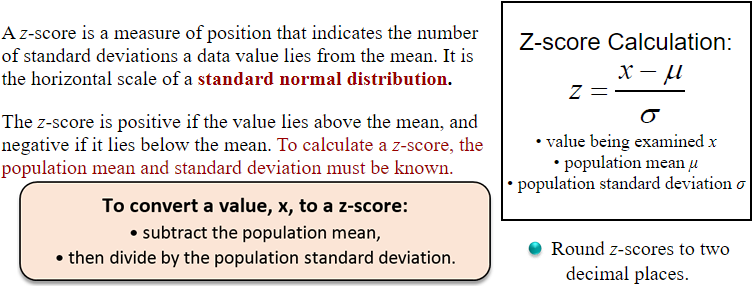
There are many different normal distributions, with each one depending on two parameters: the population mean, μ, and the population standard deviation, σ. Rather than performing computations on each new set of parameters for a variety of normal curves, it is easier to work in reference to the "simplest case" of the normal curves, called the standard normal distribution.

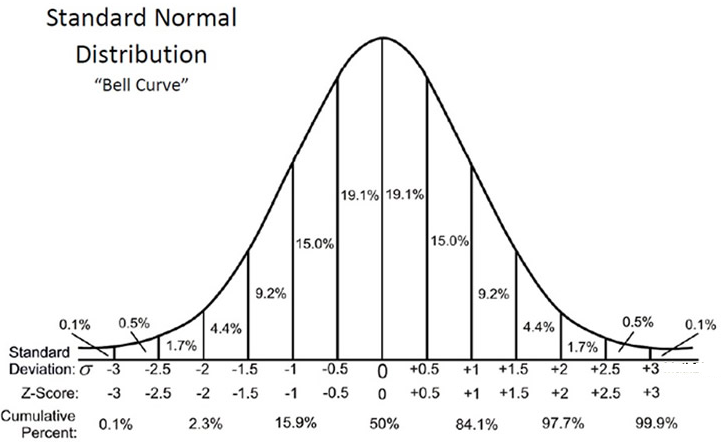
The process of *converting a value from a normal distribution to a value for the standard normal distribution* is called "standardizing" and requires the use of *z*-scores.

|  |  |
| --- | --- |
| definition | A *z*-score (or *standard score*) represents the number of standard deviations a given value *x* falls from the mean, *μ* |

A *z*-score is a measure of position that indicates the number of standard deviations a data value lies from the mean. It is the horizontal scale of a **standard normal distribution.**

The *z*-score is positive if the value lies above the mean, and negative if it lies below the mean. To calculate a *z*-score, the population mean and standard deviation must be known





Remember that z-scores tell us how far a value is from the mean. When you "standardize" a variable, its mean becomes zero and its standard deviation becomes one.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| |  | | --- | | AREA under a Normal Curve: |  |  |  |  | | --- | --- | --- | | |  |  | | --- | --- | | note | Areas under all normal curves are related. For example, the area percentage to the right of 1.5 standard deviations above the mean is identical for **all** normal curves. (The term "area" will refer to "area percentage".) | |   The fact stated above is the reason we can find an area over an interval for any normal curve by finding the **corresponding area**under a standard normal curve (with a mean of 0 and a standard deviation of 1).  We have seen that the [Empirical Rule](http://mathbitsnotebook.com/Algebra2/Statistics/STnormalDistribution.html) (68% - 95% - 99.7%) subdivides the area under a normal distribution into sections with widths of one standard deviation. These subdivisions are fine for determining percentages as long as we are dealing with values that fall at these exact subdivision locations.  What do we do when the value does not fall at an Empirical Rule subdivision? By using *z*-scores, we have the ability to locate a percentage (or area) under a standard normal distribution at **any** location. Z-scores allow for the calculation of area percentages (also called *proportions* or *probabilities*) anywhere along a standard normal distribution curve (and, consequently along the corresponding normal distribution).  The area percentage (proportion, probability) calculated using a z-score will be a decimal value between 0 and 1, and will appear in a Z-Score Table. The total area under any normal curve is 1 (or 100%). Since the normal curve is symmetric about the mean, the area on either sides of the mean is 0.5 (or 50%).  To find a specific area under a normal curve, find the z-score of the data value and use a Z-Score Table to find the area. A Z-Score Table, is a table that shows the percentage of values (or area percentage) to the left of a given *z*-score on a standard normal distribution. |

|  |  |  |  |
| --- | --- | --- | --- |
| |  | | --- | | [Positive Z-Score Table](http://mathbitsnotebook.com/Algebra2/Statistics/PositiveZScores.pdf) | | |  | | --- | | [Negative Z-Score Table](http://mathbitsnotebook.com/Algebra2/Statistics/NegativeZScores.pdf) | |

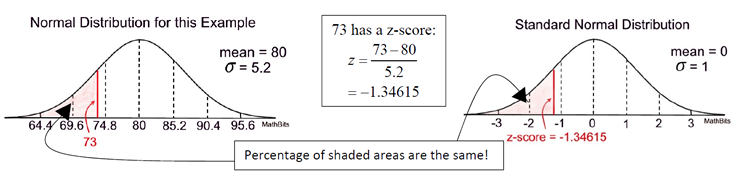
*You need both tables!*

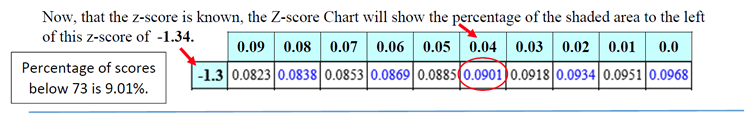
|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| • These tables are designed only for the standard normal distribution, which has a mean of 0 and a standard deviation of 1.  • The left most column is how many standard deviations above (or below) the mean to one decimal place. (The label in the row contains the integer part and the first decimal of the z-score.)  • The part of the *z*-score denoting hundredths is found across the top row of the table. (The label for columns contains the second decimal of the z-score.)  • The intersection of the rows and columns gives the probability or area under the normal curve. Each value in the body of the table is a cumulative area.  beware Z-Score Tables come in different formats, determined by where the computations were started. Consider these two most popular formats: **1.** One form of the table yields probability or area **starting at the mean** and going to the right of the mean up to the needed *z*-score. These tables are usually labeled "cumulative from mean". This table basically works with half of the area under the normal curve, and the user must take this into consideration and make adjustments when using this table. This type of table lists positive z-scores only. **2.**Another form of the table yields probability or area **starting from negative infinity**(the farthest left) and going to the right up to the needed *z*-score. These tables are usually labeled "cumulative from the left". This table works with the entire area under the normal curve, and requires less adjustments than the first option. This table lists both positive and negative z-scores. Most beginning statistical textbooks include this Z-Score Table, and this site will be using this format.   |  |  |  | | --- | --- | --- | | |  | | --- | | **Example 1:**Find the probability that a variable has a *z*-score of less than 0.36. |   **Solution:** Find the *z*-score in the table below. The intersection shows 0.6406. The probability is 64.06% (or the area percentage of the yellow region is 0.6406). | zscore36 | |  |  |   ztableex |

divider dash

|  |  |
| --- | --- |
| |  | | --- | | **Example 2:** A normally distributed population of test scores has a mean of 80 and a standard deviation of 5.2. **a)** Find the percentage of scores that lies below 73. | |

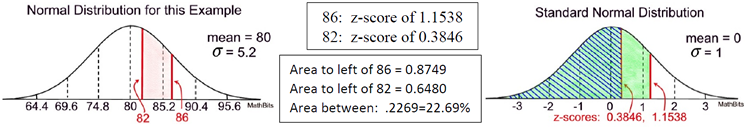
Start by setting up a sketch of the situation. This will give you a better understanding of where your values lie and whether your final answer will be reasonable. Remember that the area under a normal curve is read as an accumulation starting from the left side of the graph.





|  |
| --- |
| **b)** Find the percentage of scores that lies between 82 and 86. |

To find the percentage of the area between two values, such as 82 and 86, find the z-scores for each value. From the Z-Score Table, find the area to the left of 86 and subtract the area to the left of 82.



|  |
| --- |
| **c)** Find the percentage of scores that lies above 73. |

To find the percentage of the area that lies "above" the z-score, take thetotal area under a normal curve (which is 1) and subtract the cumulativearea to the left of the z-score.

In part a, 73 had a z-score of -1.34615 with a cumulative area to the left of 0.0901 or 9.01%.The area to the right of this z-score will be 1 - 0.0901 = 0.9099 or 90.99%.