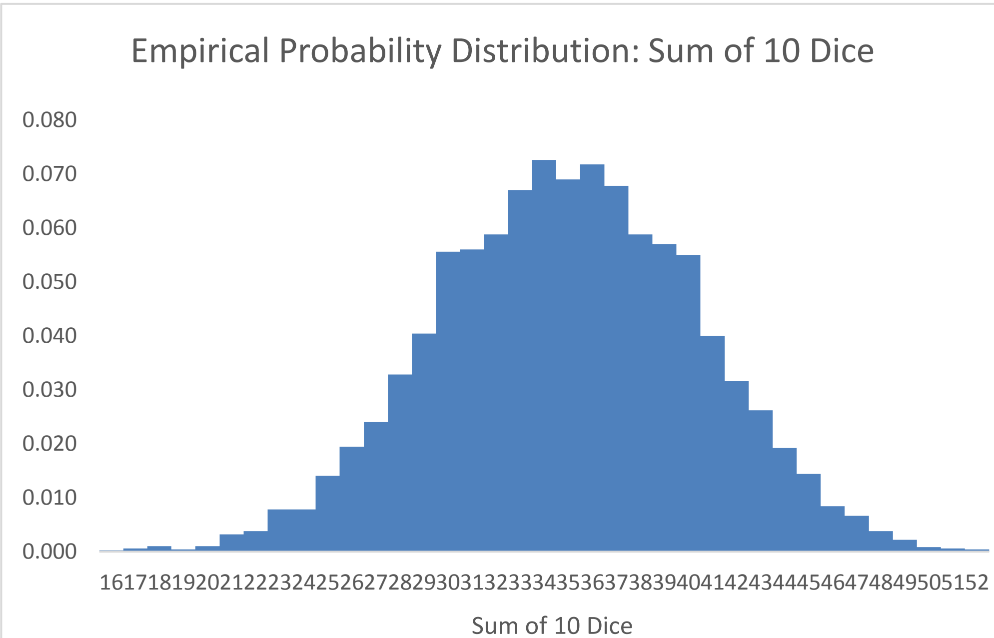
**MTH245 Unit4 Module 4 The Normal Distribution**

If we set up a simulation to roll 10 dice and find the sum of the dice, we will end up with a probability distribution similar to:

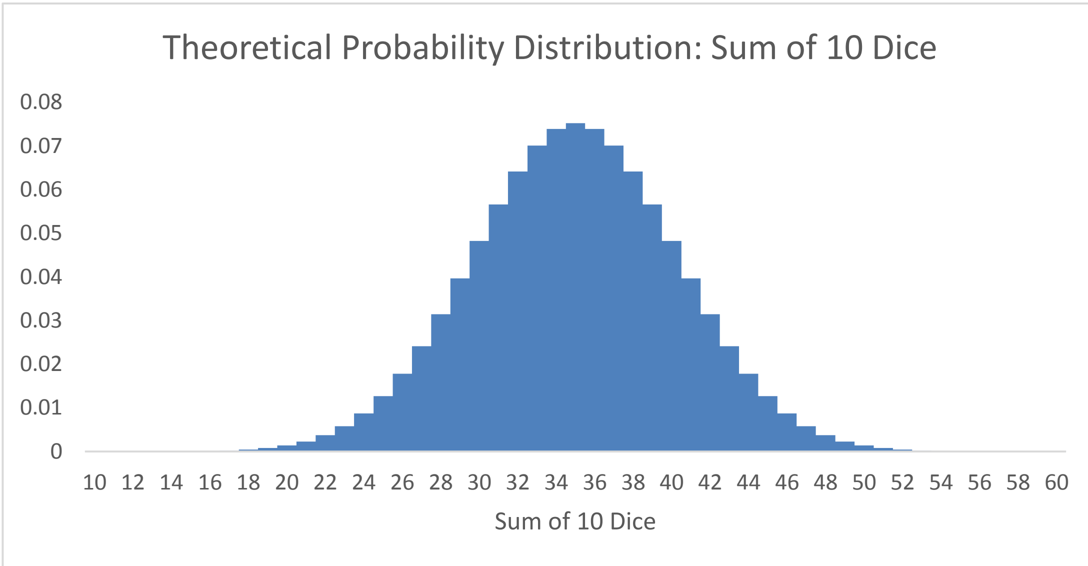



| Row Labels | Count of sum |
|---|---|
| 16 | 0.000 |
| 17 | 0.001 |
| 18 | 0.001 |
| 19 | 0.000 |
| 20 | 0.001 |
| 21 | 0.003 |
| 22 | 0.004 |
| 23 | 0.008 |
| 24 | 0.008 |
| 25 | 0.014 |
| 26 | 0.019 |
| 27 | 0.024 |
| 28 | 0.033 |
| 29 | 0.040 |
| 30 | 0.056 |
| 31 | 0.056 |
| 32 | 0.059 |
| 33 | 0.067 |
| 34 | 0.073 |
| 35 | 0.069 |
| 36 | 0.072 |
| 37 | 0.068 |
| 38 | 0.059 |
| 39 | 0.057 |
| 40 | 0.055 |
| 41 | 0.040 |
| 42 | 0.032 |
| 43 | 0.026 |
| 44 | 0.019 |
| 45 | 0.014 |
| 46 | 0.008 |
| 47 | 0.007 |
| 48 | 0.004 |
| 49 | 0.002 |
| 50 | 0.001 |
| 51 | 0.001 |
| 52 | 0.000 |
| **Grand Total** | **1.000** |

This is an example of the famous bell curve. Many random variables (IQ scores, test scores, height distributions) that occur in the world have a probability distribution that is in this classic shape. It is a symmetric distribution (the area to the left of center mirrors the area right of center), centered at its mean (35 in this example) and the total area under the curve (which is the sum of the probabilities in the table) totals 1 (or 100%).

If we run more simulations we will get slightly different looking versions of this graph, and, the larger our simulation, the “smoother” the graph will look.

Below I have created a theoretical version of the same situation, but it was created using a formula for what is called the **Normal Distribution.**

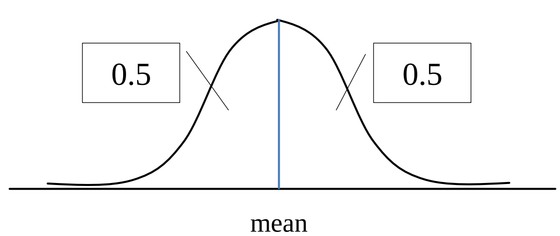



We have been using probability distribution tables in the last few sections, but with the Normal distribution it will be helpful to have a diagram of the distribution, as well as having a table. The way the tables are organized can be a little confusing at first and having the picture to visualize how large or small a probability (represented by the area under the graph) will give us a chance to see whether our results are reasonable.

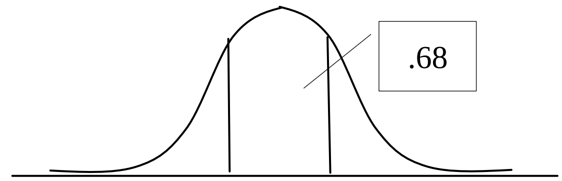
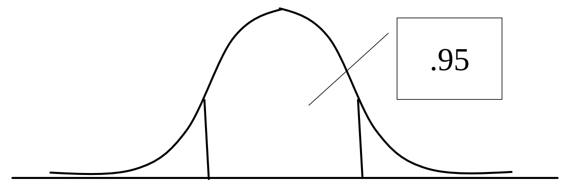
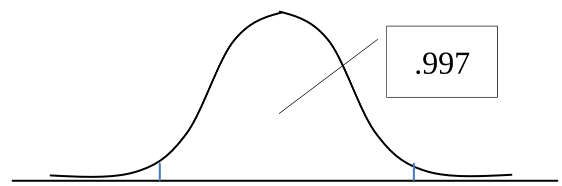
Also, the normal distribution is what is called a continuous distribution rather than a discrete one...all that means is that we will be finding the probability of an interval of values rather than the probability of an individual value. For example, rather than finding the probability that a child is 30 inches tall, we might find the probability that a child is between 29.5 inches and 30.5 inches tall. Or the probability that the child is more than 30 inches tall.

This is a detail to keep in mind: when looking at an illustration of a normal curve the line along the bottom is a number line representing values of x—values of the variable we are interested in. The area under the curve is the probability that the random variable has a value in the interval (*a, b*) is given by the area under the curve between the lines x = a and x = b.

The total area under the curve is 1 (or 100%) just as the sum of the probabilities in a probability table is 1 (or 100%). The curve is symmetric about its mean (centered at the mean), the vertical line cuts the distribution in half and
P(x < the mean) = P(x > the mean) = .5

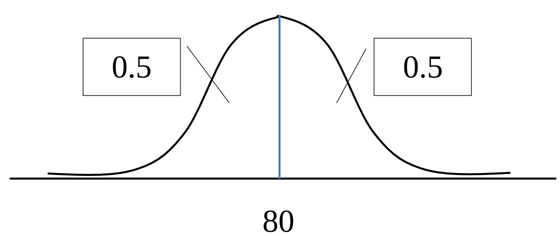


If a probability distribution has a normal distribution we may use the following **Rules of Thumb**:

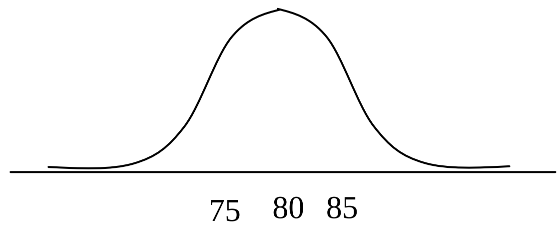
| | |
|---|---|
| 1. About 68% of the values lie within one standard deviation of the mean. |  |
| 2. About 95% of the values lie within two standard deviations of the mean. |  |
| 3. About 99.7% of the values lie within three standard deviations of the mean. |  |

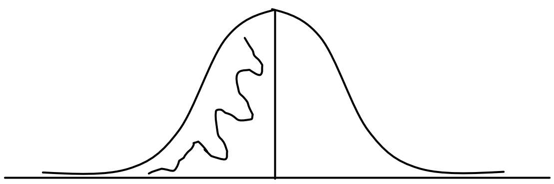
The shape of the curve is determined by its mean and standard deviation. The larger the standard deviation is, the more spread out the bell curve is, but the total area under the curve will be 1 no matter that the standard deviation is.

Feeling a little overwhelmed? Let's look at some specific examples. Say a test score has a normal distribution with a mean of 80 and a standard deviation of 5. With a mean of 80, we know that the normal curve will be centered over the value of 80.

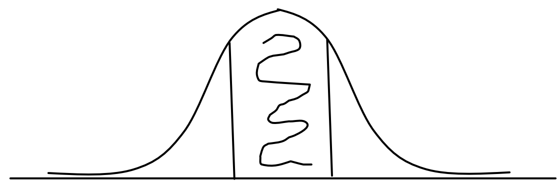


Knowing the standard deviation of 5 gives us an idea of how spread out the test scores will be along the number line.

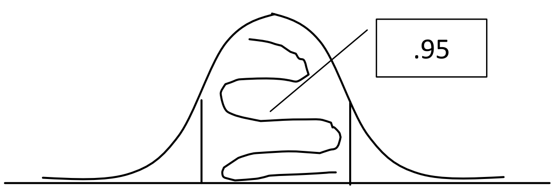


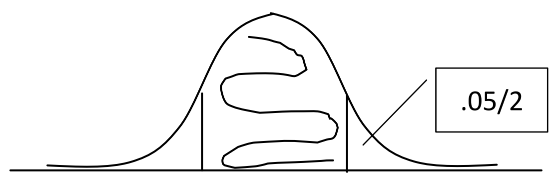
To find the probability that a test score is less than 80 look at the graph: the line above 80 cuts the area in half.
Half of the scores will be less than 80 and half will be more. $P(x < 80) = .5$



To find the probability that a test score is between 75 and 85 we need to note that 75 and 85 are one standard deviation left and right of the mean. Our Rule of Thumb tells us that: $P(75 < x < 85) = .68$




To find the probability that a test score is between 70 and 90 we need to note that 70 and 90 are two standard deviations left and right of the mean. Our Rule of Thumb tells us that $P(70 < x < 90)$=.95




To find the probability that a test score is greater than 90, recall the total area under the curve is 1. So 1-.95 =.05 is the area NOT between 70 and 90. Because the area is symmetric, half of the .05 is left of 70 and half is right of 90: $P(x > 90) = .025$

These general rules of thumb we have been using are great for making mental calculations, and for estimations. To find more specific probabilities, such as how likely is a value of x to be within 1.3 standard deviations of the mean, we need another technique.

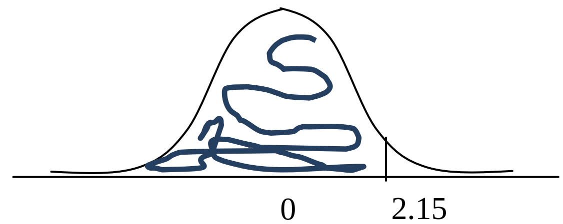
Pretty much every set of data that is normally distributed (has that bell shaped curve for its probability distribution) will have a unique mean and standard deviation.
For now, we will just focus on the standard normal distribution, which has a mean of 0 and a standard deviation of 1.
Obviously, very few applications have a mean of 0 and a standard deviation of 1, but we will learn how to adjust for those later...one step at a time!

**The Standard Normal Distribution (The *Z-curve*: mean = 0 and standard deviation = 1.)**
This table gives values for the probability of variable with a standard normal distribution to be less than some value $z$.

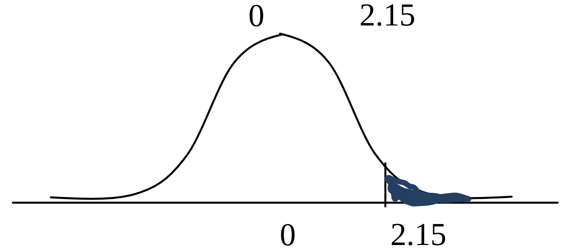
| *z* | *P( < z)* | *z* | *P( < z)* | *z* | *P( < z)* | *z* | *P( < z)* |
|---|---|---|---|---|---|---|---|
| -4.00 | 0.00003 | -2.00 | 0.02275 | 0.00 | 0.50000 | 2.00 | 0.97725 |
| -3.95 | 0.00004 | -1.95 | 0.02559 | 0.05 | 0.51994 | 2.05 | 0.97982 |
| -3.90 | 0.00005 | -1.90 | 0.02872 | 0.10 | 0.53983 | 2.10 | 0.98214 |
| -3.85 | 0.00006 | -1.85 | 0.03216 | 0.15 | 0.55962 | 2.15 | 0.98422 |
| -3.80 | 0.00007 | -1.80 | 0.03593 | 0.20 | 0.57926 | 2.20 | 0.98610 |
| -3.75 | 0.00009 | -1.75 | 0.04006 | 0.25 | 0.59871 | 2.25 | 0.98778 |
| -3.70 | 0.00011 | -1.70 | 0.04457 | 0.30 | 0.61791 | 2.30 | 0.98928 |
| -3.65 | 0.00013 | -1.65 | 0.04947 | 0.35 | 0.63683 | 2.35 | 0.99061 |
| -3.60 | 0.00016 | -1.60 | 0.05480 | 0.40 | 0.65542 | 2.40 | 0.99180 |
| -3.55 | 0.00019 | -1.55 | 0.06057 | 0.45 | 0.67364 | 2.45 | 0.99286 |
| -3.50 | 0.00023 | -1.50 | 0.06681 | 0.50 | 0.69146 | 2.50 | 0.99379 |
| -3.45 | 0.00028 | -1.45 | 0.07353 | 0.55 | 0.70884 | 2.55 | 0.99461 |
| -3.40 | 0.00034 | -1.40 | 0.08076 | 0.60 | 0.72575 | 2.60 | 0.99534 |
| -3.35 | 0.00040 | -1.35 | 0.08851 | 0.65 | 0.74215 | 2.65 | 0.99598 |
| -3.30 | 0.00048 | -1.30 | 0.09680 | 0.70 | 0.75804 | 2.70 | 0.99653 |
| -3.25 | 0.00058 | -1.25 | 0.10565 | 0.75 | 0.77337 | 2.75 | 0.99702 |
| -3.20 | 0.00069 | -1.20 | 0.11507 | 0.80 | 0.78814 | 2.80 | 0.99744 |
| -3.15 | 0.00082 | -1.15 | 0.12507 | 0.85 | 0.80234 | 2.85 | 0.99781 |
| -3.10 | 0.00097 | -1.10 | 0.13567 | 0.90 | 0.81594 | 2.90 | 0.99813 |
| -3.05 | 0.00114 | -1.05 | 0.14686 | 0.95 | 0.82894 | 2.95 | 0.99841 |
| -3.00 | 0.00135 | -1.00 | 0.15866 | 1.00 | 0.84134 | 3.00 | 0.99865 |
| -2.95 | 0.00159 | -0.95 | 0.17106 | 1.05 | 0.85314 | 3.05 | 0.99886 |
| -2.90 | 0.00187 | -0.90 | 0.18406 | 1.10 | 0.86433 | 3.10 | 0.99903 |
| -2.85 | 0.00219 | -0.85 | 0.19766 | 1.15 | 0.87493 | 3.15 | 0.99918 |
| -2.80 | 0.00256 | -0.80 | 0.21186 | 1.20 | 0.88493 | 3.20 | 0.99931 |
| -2.75 | 0.00298 | -0.75 | 0.22663 | 1.25 | 0.89435 | 3.25 | 0.99942 |
| -2.70 | 0.00347 | -0.70 | 0.24196 | 1.30 | 0.90320 | 3.30 | 0.99952 |
| -2.65 | 0.00402 | -0.65 | 0.25785 | 1.35 | 0.91149 | 3.35 | 0.99960 |
| -2.60 | 0.00466 | -0.60 | 0.27425 | 1.40 | 0.91924 | 3.40 | 0.99966 |
| -2.55 | 0.00539 | -0.55 | 0.29116 | 1.45 | 0.92647 | 3.45 | 0.99972 |
| -2.50 | 0.00621 | -0.50 | 0.30854 | 1.50 | 0.93319 | 3.50 | 0.99977 |
| -2.45 | 0.00714 | -0.45 | 0.32636 | 1.55 | 0.93943 | 3.55 | 0.99981 |
| -2.40 | 0.00820 | -0.40 | 0.34458 | 1.60 | 0.94520 | 3.60 | 0.99984 |
| -2.35 | 0.00939 | -0.35 | 0.36317 | 1.65 | 0.95053 | 3.65 | 0.99987 |
| -2.30 | 0.01072 | -0.30 | 0.38209 | 1.70 | 0.95543 | 3.70 | 0.99989 |
| -2.25 | 0.01222 | -0.25 | 0.40129 | 1.75 | 0.95994 | 3.75 | 0.99991 |
| -2.20 | 0.01390 | -0.20 | 0.42074 | 1.80 | 0.96407 | 3.80 | 0.99993 |
| -2.15 | 0.01578 | -0.15 | 0.44038 | 1.85 | 0.96784 | 3.85 | 0.99994 |
| -2.10 | 0.01786 | -0.10 | 0.46017 | 1.90 | 0.97128 | 3.90 | 0.99995 |
| -2.05 | 0.02018 | -0.05 | 0.48006 | 1.95 | 0.97441 | 3.95 | 0.99996 |

We will use the table to calculate the probabilities, but I strongly encourage you to also sketch the curve and shade the area that represents the probability that you are asked to find. Why? There are some mental gymnastics involved if it isn't $P(< z)$ that we actually want… if it is $P(> z)$, $P(b < z < c)$, we will need to subtract values from 1, or subtract two values from each other. This is not difficult to do, but it is easier to keep it all straight with a sketch.

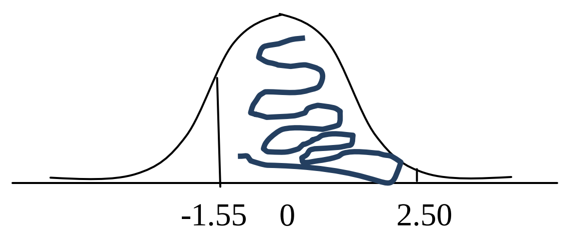
Examples:
Let $x$ be a continuous random variable with standard normal distribution.
Use the table to find:
a) $P( x \leq 2.15)$
This is the type of probability our table is exactly set up for. All we need to do is look in the "z" column for the value 2.15. In the column to its right, the column labeled $P(< z)$ is the value 0.98422. That tells us that the $P( x \leq 2.15)$ = 0.98422. And this probability corresponds to the percent of the area that is shaded.




b) $P(x > 2.15 )$
Here we want exactly the opposite information as what our table is set up to provide. But we know that the total area under the curve is 1, so the area to the right of 2.15 will be 1–.98422 and the $P(x > 2.15 ) = 0.1578$

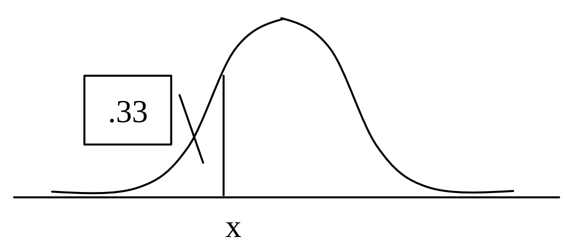



c) $P(-1.55 \leq x \leq 2.50)$




We want the likelihood of being between two values. So we can take the $P(x \leq 2.5)$ which is 0.99379, and then subtract the area left of -1.55-- subtract $P(x \leq -1.55)$, which is 0.06057, and we are left with $P(-1.55 \leq x \leq 2.50) = .93322$

d) What value of $x$ should be used if we want $P(x < z)$ to be about 1/3?
To do this, we will start off looking at the columns in our table labeled $P(< z)$ for a value that is about .33. When z is -0.45 $P(< z) = .32636$ and when z is -0.40 $P(< z) = .34458$.
So $P(x < -.425)$would be approximately 1/3.



**Non-Standard Normal Distribution**

What do we do if we have a population that is normally distributed but does not have a mean of 0 and a standard deviation of 1?

Pretty much every set of data that is normally distributed (has that bell shaped curve for its probability distribution) will have a unique mean and standard deviation.
To make calculations less complicated we have a way of "standardizing" values with z-scores. Whatever the mean of a particular distribution is, we are going to scale the data set so its mean is zero. That way, any value greater than the mean has a positive z-score and any value less than the mean has a negative z-score.
The z-score itself is how many standard deviations above or below the mean a particular value is.

Say we have a set of test scores with a mean of 80 and a standard deviation of 5.
A test score of 85 is one standard deviation greater than the mean so it would have a z-score of 1.
A test score of 70 is two standard deviations less than the mean so it will have a z-score of -2.
This can all be easily calculated with a formula.

If we have a normally distributed random variable $x$ with mean $\mu$ and standard deviation $\sigma$, we can transform it to *standard* normal using $z = \frac{x-\mu}{\sigma}$,
in words: $z = \frac{actual\ value - mean}{st.deviation}$

We use $z$-scores to "standardize" the values we are interested in, into values that we can use with the standard normal table and then find the probability of getting a value in a particular range, like this:

$$P(a < x < b) = P\left(\frac{a-\mu}{\sigma} < z < \frac{b-\mu}{\sigma}\right)$$

.

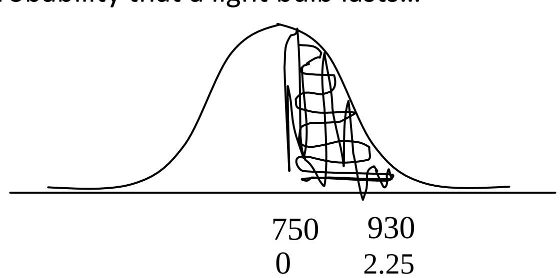
Example:
A large manufacturer of light bulbs finds that the life expectancy of its 100 watt light bulbs has a distribution that is approximately normal with a mean value of 750 hours and a standard deviation of 80 hours.
Use the standard normal distribution to find the probability that a light bulb lasts…

1. between 750 and 930 hours

We need to scale the two values using $z = \frac{x-\mu}{\sigma}$
and we know that $\mu = 750$ and $\sigma = 80$

$z_1 = \frac{750-750}{80} = \frac{0}{80} = 0$ and

$$z_2 = \frac{930-750}{80} = \frac{180}{80} = 2.25$$



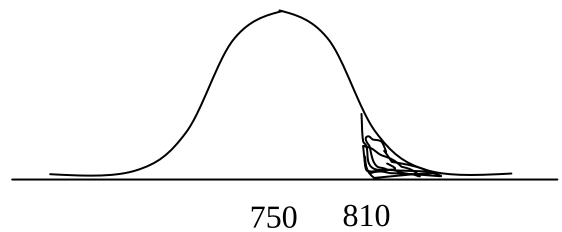
The calculated values of 0 and 2.25 are z-scores that are scaled to have a mean of 0 and standard deviation of 1, so we can now turn to our table to look up the $P(< z)$ for 0 and 2.25, which are 0.5 and 0.98778. We then subtract the two values to get the area in between, and we find that $P(750 < x < 930) = .48778$. The probability of a light bulb lasting between 750 and 930 hours is 48.778%.

2. More than 810 hours

$$z = \frac{810 - 750}{80} = \frac{60}{80} = 0.75$$

When we look up 0.75 in the table, we find $P(z < .75)$ is .77337. But we need the $P(z > .75) = 1 - .77337 = .22663$

The probability of a bulb lasting more than 810 hours will be 22.663%




3. between 730 and 850 hours

$$z_1 = \frac{730 - 750}{80} = \frac{-20}{80} = -0.25$$

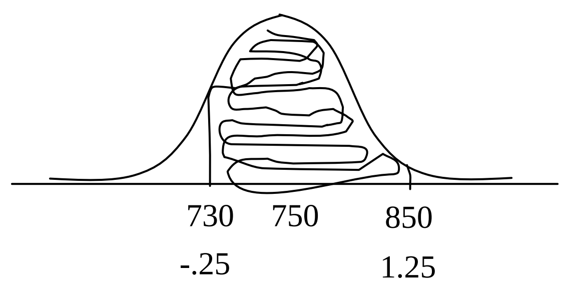
$$z_2 = \frac{850 - 750}{80} = \frac{100}{80} = 1.25$$

From our table:

$$P(z < -.25) = .40129$$

and $P(z < 1.25) = .89435$. Taking the difference: .89435-.40129=.49306.

The probability of one of these light bulbs lasting between 730 and 850 hours is about 49.31%

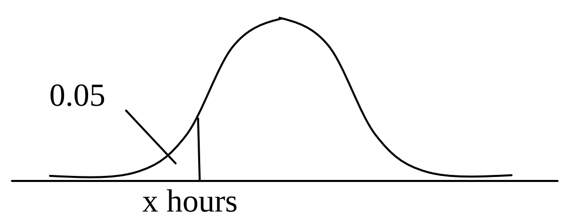



4. If there are 800 light bulbs in a shipment, how many of those would you expect to last between 730 and 850 hours?

We would expect about 49.31% to last that long, and 49.31% of 800 would be about 395 bulbs.

5. The manufacturer wants to offer a guarantee that the bulbs will last at least ___ hours. They do not want to have to replace more than 5% of the bulbs. How many hours do you expect 95% of the light bulbs to last?

Let's think about this one a minute. If the bulbs were guaranteed to last 750 hours, we would expect about half of them to burn about by 750 hours and fall under the warranty. So our answer should be less than 750 hours!



We need to turn to our table and look for a z value where $P(< z) \ = \ 0.05$, Be sure you look in the columns labeled $P(< z)$, not in the $z$ columns! There is not a value that is exactly 0.05, but we can get pretty close with 0.04947. This probability corresponds to a z value of -1.65. Now we have to solve for x hours:

$$
\begin{aligned}
-1.65 &= \frac{x-750}{80} \\
-1.65 \cdot 80 &= \frac{x-750}{80} \cdot 80 \\
-132 \ &= \ x - 750 \\
618 &= x
\end{aligned}
$$

If the bulbs are guaranteed to last at least 618 hours, we would expect only about 5% of them to burn out that early.

Excel can be used rather than the table. The equation =NORMSDIST(z) will return the P(<z) for a value of z. And the equation =NORMSINV(probability) will return a value of z where P(<z) is the probability. For example =NORMSDIST(2) will return .97725 and =NORMSINV(.05) will return -1.64485. When using Excel you are not limited to the values given in the table and will be able to calculate more accurate values if you do not have a “nice” value.

Also in Excel are the =NORMDIST(x, mean, standard_dev, cumulative) and =NORMINV(probability, mean, standard_dev) functions. With these two functions, Excel will do the intermediate calculations of z.