Chapter 7: The Normal Probability Distribution
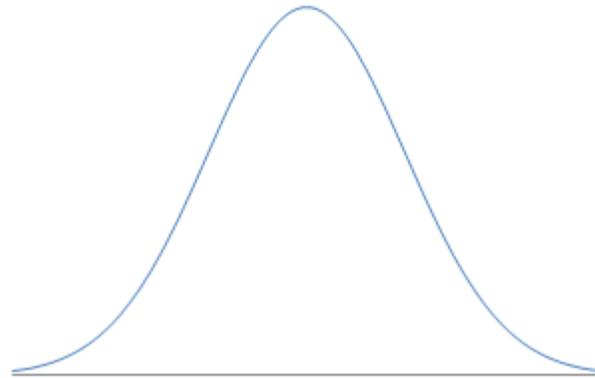
In Chapter 7, we bring together much of the ideas in the previous two on probability. We expand the earlier bell-shaped distribution (we introduced this shape back in Section 2.2) to its more formal name of a **normal curve**.

Many random variables have histograms that follow the normal curve, like an individuals height, the thickness of tree bark, IQs, or the amount of light emitted by a light bulb. This is helpful, because we can use it to answer all kinds of interesting questions.

Source: stock.xchng

- What proportion of individuals are geniuses?
- Is a systolic blood pressure of 110 unusual?
- What percentage of a particular brand of light bulb emits between 300 and 400 lumens?
- What is the 90th percentile for the weights of 1-year-old boys?

If you're ready to begin, just click on the "start" link below, or one of the section links on the left.

# Section 7.1: Properties of the Normal Distribution

7.1 Properties of the Normal Distribution
7.2 Applications of the Normal Distribution
7.3 Assessing Normality

Objectives

By the end of this lesson, you will be able to...

1. use the uniform probability distribution
2. graph a normal curve
3. state the properties of the normal curve
4. explain the role of area in the normal density function

Probability Density Functions

In Chapter 6, we focused on **discrete random variables**, random variables which take on either a finite or countable number of values. **Continuous random variables**, which have infinitely many values, can be a bit more complicated.

Consider the **rand()** function in the computer software Microsoft Excel. It returns a random number between 0 and 1. There are infinitely many possibilities, so each particular value has a probability of 0!

When we consider continuous random variables, we need to instead consider the probability "density", which might not always be the same for each value. Some ranges might be more likely, and hence the probability would be more "dense" near those values. To make this easier to understand, we need a new concept called a **probability density function**.

Let's look at Example 4, from Section 6.1, in which two dice were tossed and X = the sum of the two dice. The histogram below highlights P(X<6).
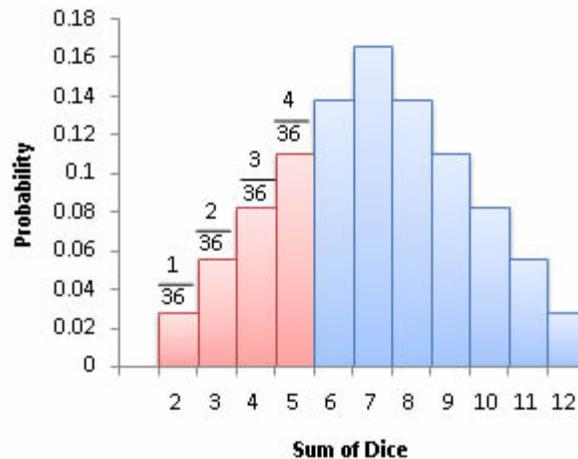

We can see from the histogram that P(X<6) = P(X=2) + P(X=3) + P(X=4) + P(X=5), but let's look at things a little differently. Instead of focusing on the probabilities, let's look at the *area* that's shaded red. The width of each rectangle is 1, so the area of each is its corresponding probability.

This leads us to another interpretation of P(X<6) - we could think of it as the *area* from 2 to 5. Extending that idea, we can now give a definition of a **probability density function**.

A **probability density function** is an equation used to compute probabilities of continuous random variables. The equation must satisfy the following two properties:

1. The total area under the graph of the equation over all possible values of the random variable must equal 1.
2. The height of the graph of the equation must be greater than or equal to 0 for all possible values of the random variable.

If we go back and consider the earlier example of the **rand()** function in Excel. Our probability density function would be fairly simple:
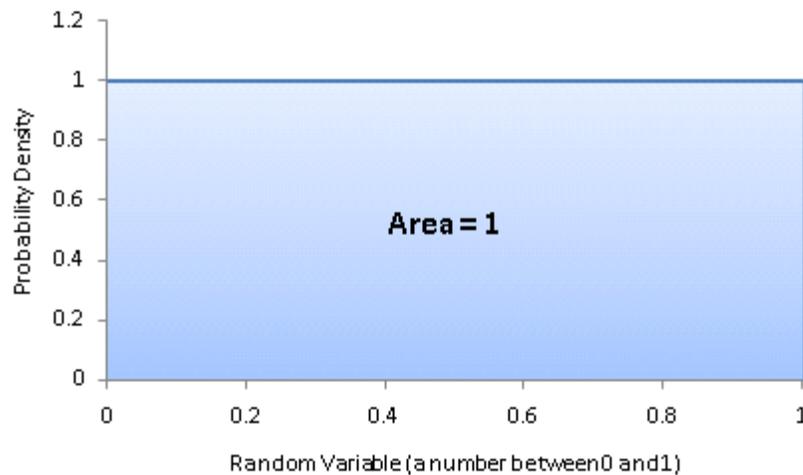

Probabilities as Areas

Now that we have the basic connection between area underneath the probability density function and the probability of that random variable, let's do a little further exploration.

In general, the **area under a probability density function** over a particular interval of values can have two interpretations:

- the proportion of the population with the characteristic
- the probability that a randomly selected individual will be within the interval

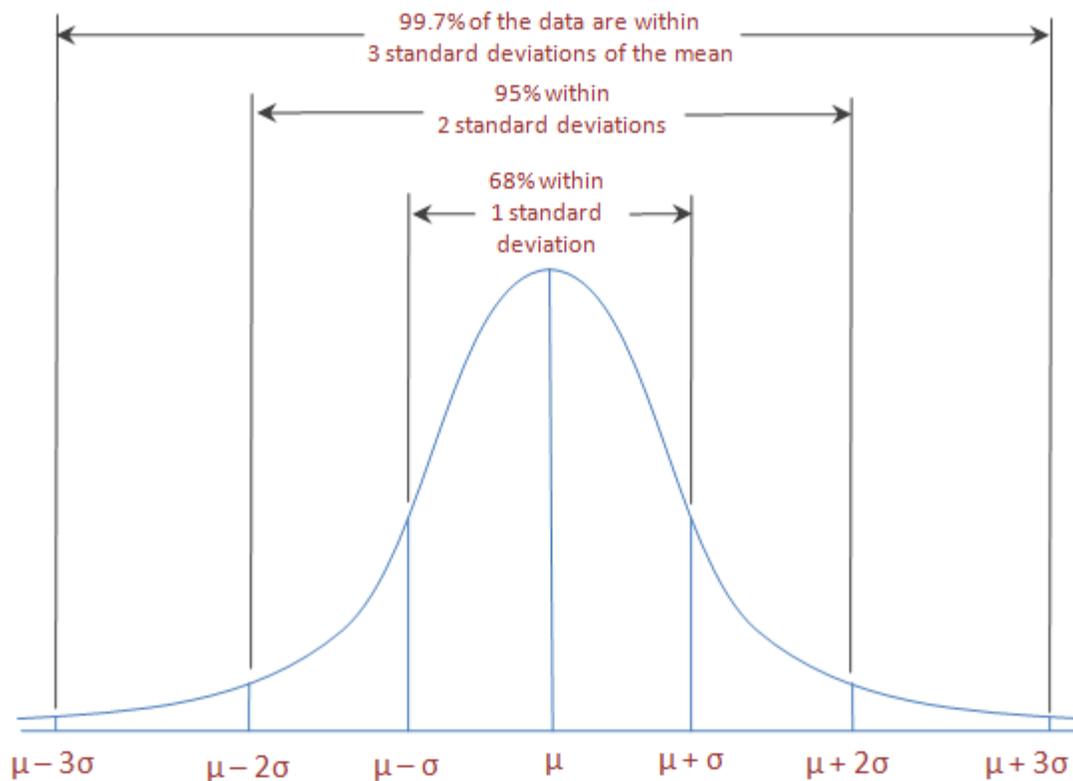## The Normal Curve

Many continuous variables follow a bell-shaped distribution (we introduced this shape back in Section 2.2), like an individuals height, the thickness of tree bark, IQs, or the amount of light emitted by a light bulb. The more formal name of a histogram of this shape is a **normal curve**.
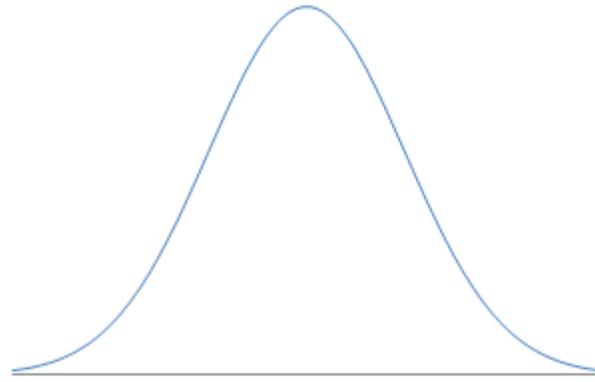
A continuous random variable is **normally distributed** or has a **normal probability distribution** if its relative frequency histogram has the shape of a normal curve.

In Section 3.2, we introduced the Empirical Rule, which said that almost all (99.7%) of the data would be within 3 standard deviations, if the distribution is bell-shaped.

We can extend this idea to the shape of other distributions. If μ = 0 and σ = 1, almost all of the data should be between -3 and 3, with the center at 0. If μ = 0 and σ = 0.5, almost all of the data should be between -1.5 and 1.5.

Exploring the Shape of the Normal Curve

To do some exploring yourself, go to the Demonstrations Project from Wolfram Research, and download the Bell Curves demonstration. If you haven't already, download and install the player by clicking on the image to the right.

Once you have the player installed and the Bell Curves demonstration downloaded,

move the sliders for the mean and standard deviations to get a sense of their effects on the shape.
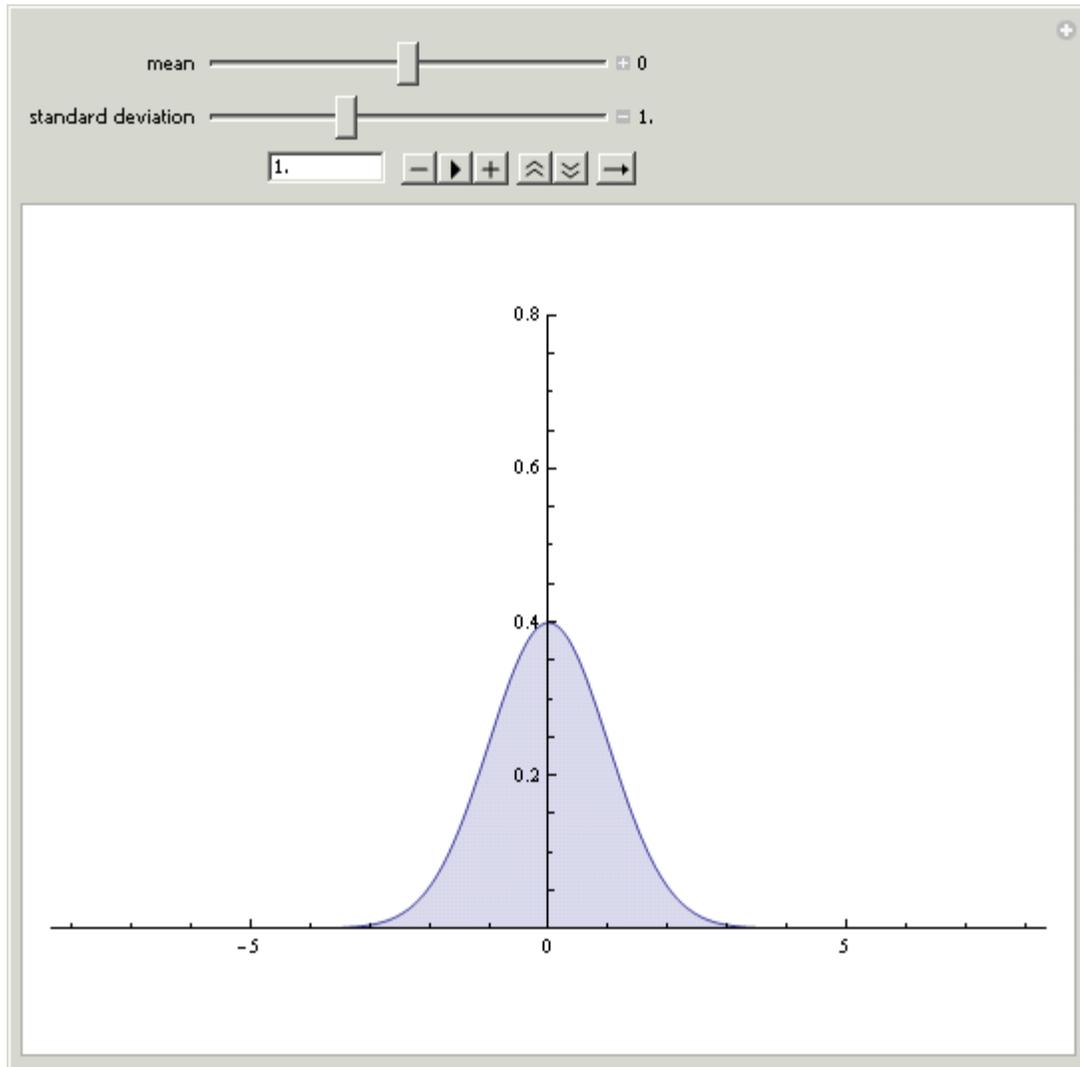
So what effect did you see from moving the mean and standard deviation? You should have seen that moving the mean simply slides the shape left or right - it changes the center, not the spread. The standard deviation, on the other hand, changes the shape.

The key is area, which we mentioned earlier this section. Since the total area under the curve needs to still be equal to 1, if we make the distribution narrower by decreasing the standard deviation, it needs to get taller to equal the same area.

Fun with Plinko

Have you heard of the game, Plinko, from the game show The Price is Right? In this game, the contestant realeases a small disc on a board covered with pegs, which direct the disc left or right. Here's a video showing a particularly successful contestant.

What's interesting is that the distribution of the Plinko chips at the bottom follows a normal distribution! Here's an example of a Java applet, showing the distribution as it might develop over hundreds of Plinko chips.

Drawing Normal Curves Using StatCrunch

Click on Stat > Calculators > Normal

Enter the mean and standard deviation (and x and the direction of the

inequality, if desired). Then press **Compute**.

To export the image, press **Snapshot** and save the image to your computer. (Don't forget where you saved it to!)

Depending on your word processing software, it's usually fairly straight-forward to insert an image. In Microsoft Word, simply choose **Insert > Picture** and select the file you saved earlier.

You can also go to the video page for links to see videos in either Quicktime or iPod format.

Areas Under a Normal Curve

Let's now connect the concepts of a normal curve and the earlier idea of area under a probability density function.

Example 1

Most tests that gauge one's intelligence quotient (IQ) are designed to have a mean of 100 and a standard deviation of 15. It's also known that IQs are normally distributed. So what would the distribution look like for IQs?
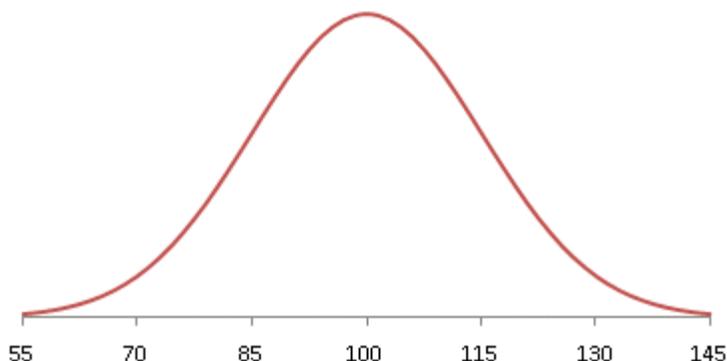
There is no universal agreement on what IQ constitutes a "genius", though in 1916, psychologist Lewis M. Thurman set a guideline of 140 (scaled to 136 in today's tests) for "potential genius".
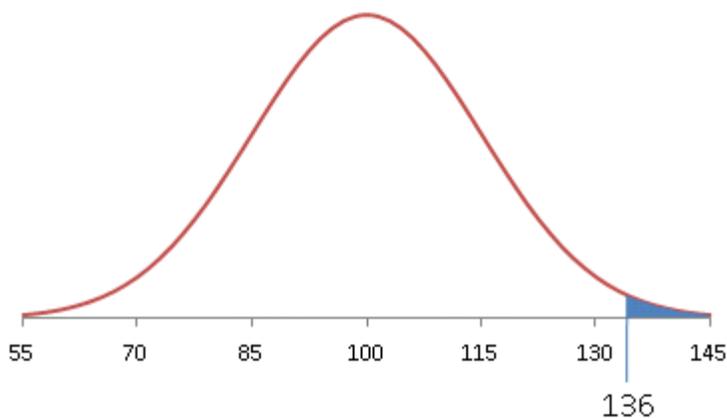
Suppose the area to the right of 136 is about 0.0082. What are two interpretations of that area?

Example 2

Weights of 1-year-old boys are approximately normally distributed, with a mean of 22.8 lbs and a standard deviation of about 2.15. (Source: About.com)

a. Draw a quick sketch of the normal curve for the weights of 1-year-old boys.
b. Shade the area representing the boys who are at least 20 pounds.
c. The area is approximately 0.9036. Give two interpretations of this result.

Source: stock.xchng

The Standard Normal Distribution

Back in Section 3.4, we introduced the idea of a z-score:

The **z-score** represents the number of standard deviations a data value is from the mean.

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

We mentioned then that we'd need to remember the z-score later - this is that moment!

The z-score is important, because if the variable X is normally distributed, Z is as well. This brings us to an important fact:

If X is normally distributed with mean μ and standard deviation σ, then

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

is normally distributed with a mean of 0 and a standard deviation of 1. We say that Z has the **standard normal distribution.**

Exploring the Standard Normal Distribution

To do some exploring yourself, go to the Demonstrations Project from Wolfram Research, and download the Area of a Normal Distribution demonstration. If you haven't already, download and install the player by clicking on the image to the right.

Once you have the player installed and the Area of a Normal Distribution demonstration downloaded, move the sliders for the mean and standard deviation of X and the value of Z to see the relationship between areas under the general normal curve and the areas under the standard normal curve.

Area
0.158655   0.841345

N(0,1)

N(15,2)

−5   0   5   10   15   20

z = −1   x = 13
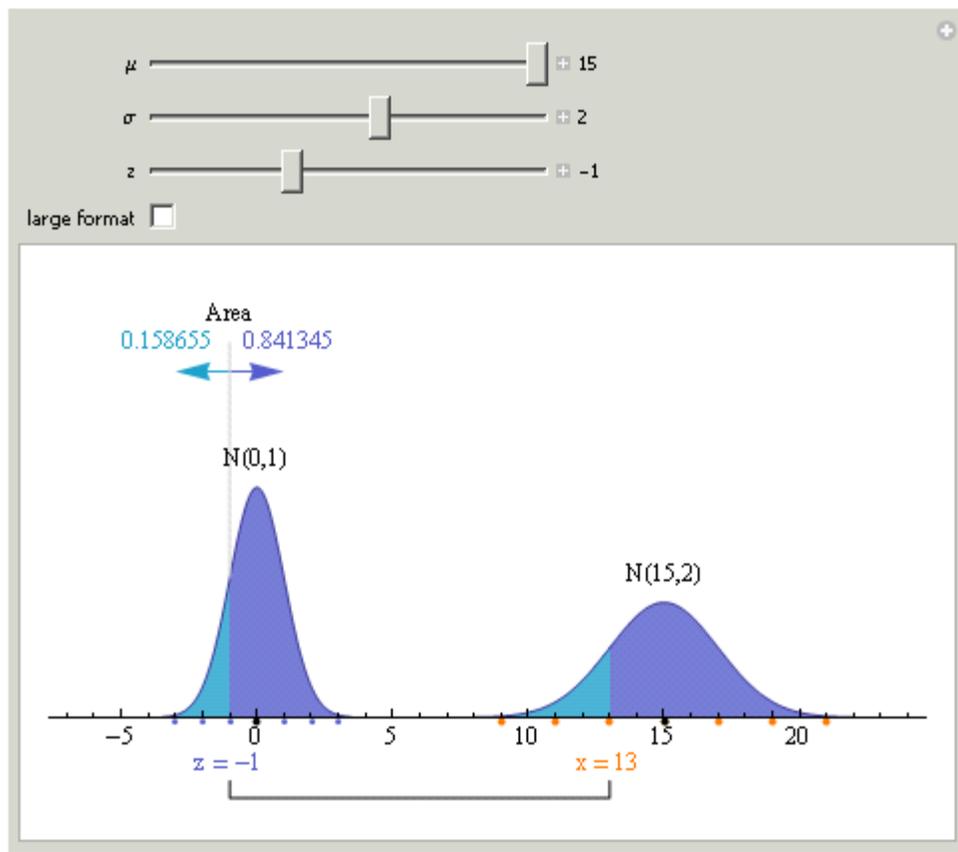
The idea here is that the area under the normal curve on the right is equal to the area under the standard normal curve on the left.

Section 7.2: Applications of the Normal Distribution

7.1 Properties of the Normal Distribution
7.2 Applications of the Normal Distribution
7.3 Assessing Normality

Objectives

By the end of this lesson, you will be able to...

1. find and interpret the area under a normal curve
2. find the value of a normal random variable

Finding Areas Using a Table

Once we have the general idea of the Normal Distribution, the next step is to learn how to find areas under the curve. We'll learn two different ways - using a table and using technology.

Since every normally distributed random variable has a slightly different distribution shape, the only way to find areas using a table is to *standardize* the variable - transform our variable so it has a mean of 0 and a standard deviation of 1. How do we do that? **Use the z-score!**

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

As we noted in Section 7.1, if the random variable X has a mean μ and standard deviation σ, then transforming X using the z-score creates a random variable with mean 0 and standard deviation 1! With that in mind, we just need to learn how to find areas under the **standard normal curve**, which can then be applied to any normally distributed random variable.

Finding Area under the Standard Normal Curve to the Left

Before we look a few examples, we need to first see how the table works. Before we start the section, you need a copy of the table. You can download a printable copy of this table, or use the table in the back of your textbook. It should look something like this:
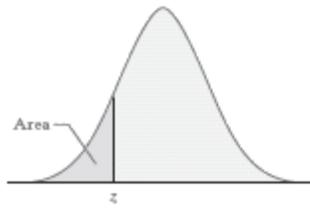
Area ↘
z

Table IV

Standard Normal Distribution

z	.00	.01	.02	.03	.04	.05	.06	.07	.08	.09
−3.4	0.0003	0.0003	0.0003	0.0003	0.0003	0.0003	0.0003	0.0003	0.0003	0.0002
−3.3	0.0005	0.0005	0.0005	0.0004	0.0004	0.0004	0.0004	0.0004	0.0004	0.0003
−3.2	0.0007	0.0007	0.0006	0.0006	0.0006	0.0006	0.0006	0.0005	0.0005	0.0005
−3.1	0.0010	0.0009	0.0009	0.0009	0.0008	0.0008	0.0008	0.0008	0.0007	0.0007
−3.0	0.0013	0.0013	0.0013	0.0012	0.0012	0.0011	0.0011	0.0011	0.0010	0.0010
−2.9	0.0019	0.0018	0.0018	0.0017	0.0016	0.0016	0.0015	0.0015	0.0014	0.0014
−2.8	0.0026	0.0025	0.0024	0.0023	0.0023	0.0022	0.0021	0.0021	0.0020	0.0019
−2.7	0.0035	0.0034	0.0033	0.0032	0.0031	0.0030	0.0029	0.0028	0.0027	0.0026
−2.6	0.0047	0.0045	0.0044	0.0043	0.0041	0.0040	0.0039	0.0038	0.0037	0.0036
−2.5	0.0062	0.0060	0.0059	0.0057	0.0055	0.0054	0.0052	0.0051	0.0049	0.0048
−2.4	0.0082	0.0080	0.0078	0.0075	0.0073	0.0071	0.0069	0.0068	0.0066	0.0064
−2.3	0.0107	0.0104	0.0102	0.0099	0.0096	0.0094	0.0091	0.0089	0.0087	0.0084
−2.2	0.0139	0.0136	0.0132	0.0129	0.0125	0.0122	0.0119	0.0116	0.0113	0.0110
−2.1	0.0179	0.0174	0.0170	0.0166	0.0162	0.0158	0.0154	0.0150	0.0146	0.0143
−2.0	0.0228	0.0222	0.0217	0.0212	0.0207	0.0202	0.0197	0.0192	0.0188	0.0183
−1.9	0.0287	0.0281	0.0274	0.0268	0.0262	0.0256	0.0250	0.0244	0.0239	0.0233
−1.8	0.0359	0.0351	0.0344	0.0336	0.0329	0.0322	0.0314	0.0307	0.0301	0.0294
−1.7	0.0446	0.0436	0.0427	0.0418	0.0409	0.0401	0.0392	0.0384	0.0375	0.0367
−1.6	0.0548	0.0537	0.0526	0.0516	0.0505	0.0495	0.0485	0.0475	0.0465	0.0455
−1.5	0.0668	0.0655	0.0643	0.0630	0.0618	0.0606	0.0594	0.0582	0.0571	0.0559
−1.4	0.0808	0.0793	0.0778	0.0764	0.0749	0.0735	0.0721	0.0708	0.0694	0.0681
−1.3	0.0968	0.0951	0.0934	0.0918	0.0901	0.0885	0.0869	0.0853	0.0838	0.0823
−1.2	0.1151	0.1131	0.1112	0.1093	0.1075	0.1056	0.1038	0.1020	0.1003	0.0985
−1.1	0.1357	0.1335	0.1314	0.1292	0.1271	0.1251	0.1230	0.1210	0.1190	0.1170
−1.0	0.1587	0.1562	0.1539	0.1515	0.1492	0.1469	0.1446	0.1423	0.1401	0.1379
−0.9	0.1841	0.1814	0.1788	0.1762	0.1736	0.1711	0.1685	0.1660	0.1635	0.1611
−0.8	0.2119	0.2090	0.2061	0.2033	0.2005	0.1977	0.1949	0.1922	0.1894	0.1867
−0.7	0.2420	0.2389	0.2358	0.2327	0.2296	0.2266	0.2236	0.2206	0.2177	0.2148
−0.6	0.2743	0.2709	0.2676	0.2643	0.2611	0.2578	0.2546	0.2514	0.2483	0.2451
−0.5	0.3085	0.3050	0.3015	0.2981	0.2946	0.2912	0.2877	0.2843	0.2810	0.2776
−0.4	0.3446	0.3409	0.3372	0.3336	0.3300	0.3264	0.3228	0.3192	0.3156	0.3121
−0.3	0.3821	0.3783	0.3745	0.3707	0.3669	0.3632	0.3594	0.3557	0.3520	0.3483
−0.2	0.4207	0.4168	0.4129	0.4090	0.4052	0.4013	0.3974	0.3936	0.3897	0.3859
−0.1	0.4602	0.4562	0.4522	0.4483	0.4443	0.4404	0.4364	0.4325	0.4286	0.4247
−0.0	0.5000	0.4960	0.4920	0.4880	0.4840	0.4801	0.4761	0.4721	0.4681	0.4641

It's pretty overwhelming at first, but if you look at the picture at the top (take a minute and check it out), you can see that it is indicating the area to the left. That's the key - the values in the middle represent areas to the left of the corresponding z-value. To determine which z-value it's referring to, we look to the left to get the first two digits and above to the columns to get the hundredths value. (Z-values with more accuracy need to be rounded to the hundredths in order to use this table.)

Say we're looking for the area left of -2.84. To do that, we'd start on the -2.8 row and go across until we get to the 0.04 column. (See picture.)

z	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
-3.4	0.0003	0.0003	0.0003	0.0003	0.0003	0.0003	0.0003	0.0003	0.0003	0.0002
-3.3	0.0005	0.0005	0.0005	0.0004	0.0004	0.0004	0.0004	0.0004	0.0004	0.0003
-3.2	0.0007	0.0007	0.0006	0.0006	0.0006	0.0006	0.0006	0.0005	0.0005	0.0005
-3.1	0.0010	0.0009	0.0009	0.0009	0.0008	0.0008	0.0008	0.0008	0.0007	0.0007
-3.0	0.0013	0.0013	0.0013	0.0012	0.0012	0.0011	0.0011	0.0011	0.0010	0.0010
-2.9	0.0019	0.0018	0.0018	0.0017	0.0016	0.0016	0.0015	0.0015	0.0014	0.0014
-2.8	0.0026	0.0025	0.0024	0.0023	0.0023	0.0022	0.0021	0.0021	0.0020	0.0019
-2.7	0.0035	0.0034	0.0033	0.0032	0.0031	0.0030	0.0029	0.0028	0.0027	0.0026
-2.6	0.0047	0.0045	0.0044	0.0043	0.0041	0.0040	0.0039	0.0038	0.0037	0.0036
-2.5	0.0062	0.0060	0.0059	0.0057	0.0055	0.0054	0.0052	0.0051	0.0049	0.0048

From the picture, we can see that the area left of -2.84 is 0.0023.

Finding Areas Using StatCrunch

Click on **Stat** > **Calculators** > **Normal**

Enter the mean, standard deviation, x, and the direction of the inequality. Then press Compute. The image below shows P(Z < 1.23).

Mean: 0	Std. Dev.: 1
Prob(X <= ▼ 1.23) =	0.89065146
Close Compute	

Let's try some examples.

Example 1

a. Find the area left of Z = -0.72

[reveal answer]

b. Find the area left of Z = 1.90

[reveal answer]

Finding Area under the Standard Normal Curve to the Right
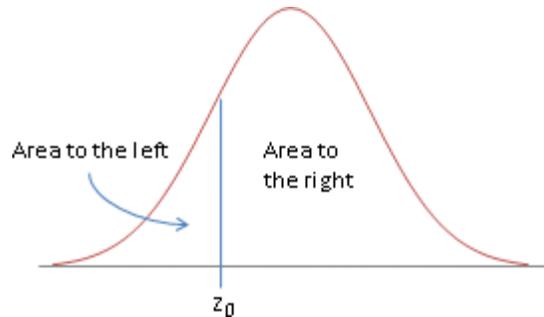
To find areas to the right, we need to remember the complement rule. Take a minute and look back at the rule from .

Since we know the entire area is 1,

> (Area to the right of z_0) = 1 - (Area to the left of z_0)

Example 2

a. Find the area to the right of Z = -0.72

[reveal answer]

b. Find the area to the right of Z = 2.68

[reveal answer]

An alternative idea is to use the symmetric property of the normal curve. Instead of looking to the right of Z=2.68 in Example 2 above, we could have looked at the area *left of -2.68*. Because the curve is symmetric, those areas are the same.

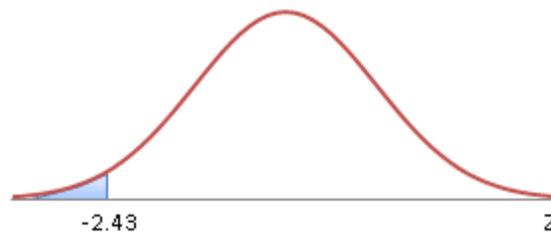## Finding Area under the Standard Normal Curve Between Two Values

To find the area between two values, we think of it in two pieces. Suppose we want to find the area between $Z = -2.43$ and $Z = 1.81$.
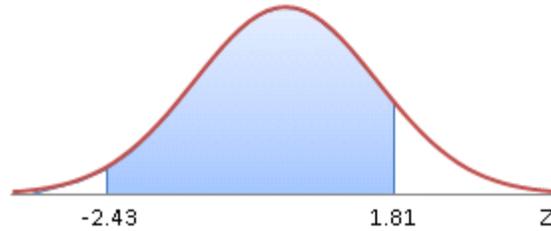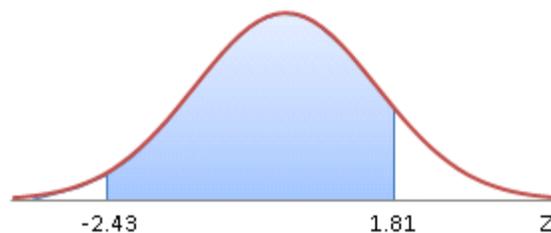

What we do instead, is find the area left of 1.81, and then subtract the area left of -2.43. Like this:
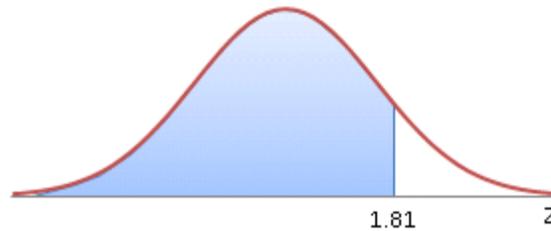
So the area between -2.43 and 1.81 = 0.9649 - 0.0075 = 0.9574

> **Note:** StatCrunch is unable to draw or calculate the "between" calculations. You'll need to perform them similarly to the one above.

Example 3

a. Find the area between Z = 0.23 and Z = 1.64.

[reveal answer]

b. Find the area between Z = -3.5 and Z = -3.0.

[reveal answer]

Finding Areas Under a Normal Curve Using the Table

1. Draw a sketch of the normal curve and shade the desired area.
2. Calculate the corresponding Z-scores.
3. Find the corresponding area under the standard normal curve.

If you remember, this is exactly what we saw happening in the Area of a Normal Distribution demonstration. Follow the link and explore again the relationship between the area under the standard normal curve and a non-standard normal curve.
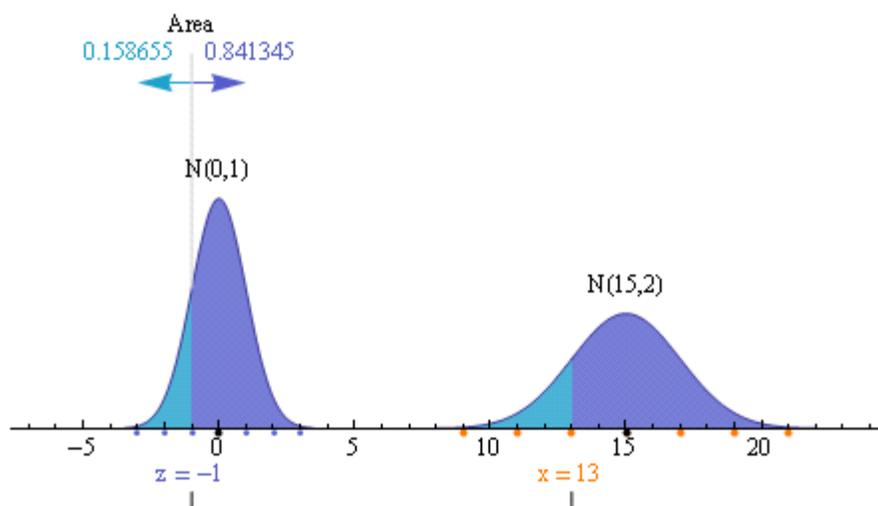
Finding Areas Under a Normal Curve Using StatCrunch

Even though there's no "standard" in the title here, the directions are actually exactly the same as those from above!

Click on **Stat** > **Calculators** > **Normal**

Enter the mean, standard deviation, x, and the direction of the inequality. Then press Compute. The image below shows P(Z < 1.23).

Now we finally get to the real reason we study the normal distribution. We want to be able to answer questions about variables that are normally distributed. Questions like..

*   What proportion of individuals are geniuses?
*   Is a systolic blood pressure of 110 unusual?

What percentage of a particular brand of light bulb emits between 300 and 400 lumens?

- What is the 90th percentile for the weights of 1-year-old boys?

All of these questions can be answered using the normal distribution!

Example 4

Let's consider again the distribution of IQs that we looked at in Example 1 in Section 7.1.

We saw in that example that tests for an individual's intelligence quotient (IQ) are designed to be normally distributed, with a mean of 100 and a standard deviation of 15.

We also saw that in 1916, psychologist Lewis M. Thurman set a guideline of 140 (scaled to 136 in today's tests) for "potential genius".

Using this information, what percentage of individuals are "potential geniuses"?

Solution:

1. Draw a sketch of the normal curve and shade the desired area.
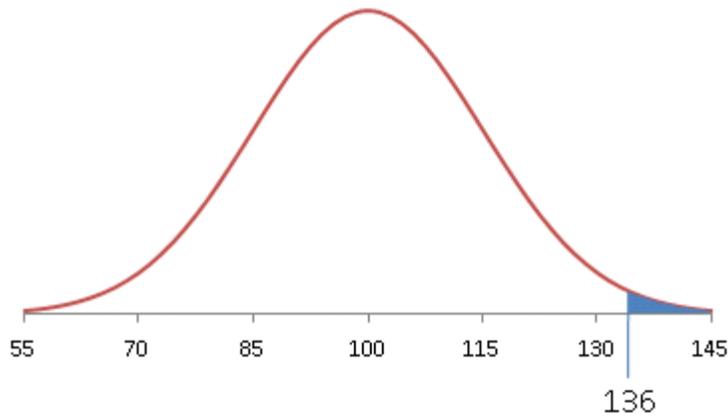
2. Calculate the corresponding Z-scores.

$$Z = \frac{X - \mu}{\sigma} = \frac{136 - 100}{15} = 2.4$$

3. Find the corresponding area under the standard normal curve.
   $P(Z>2.4) = P(Z<-2.4) = 0.0082$.

Based on this, it looks like about 0.82% of individuals can be characterized as "potential geniuses" according to Dr. Thurman's criteria.

Example 5

In Example 2 in Section 7.1, we were told that weights of 1-year-old boys are approximately normally distributed, with a mean of 22.8 lbs and a standard deviation of about 2.15. (Source: About.com)
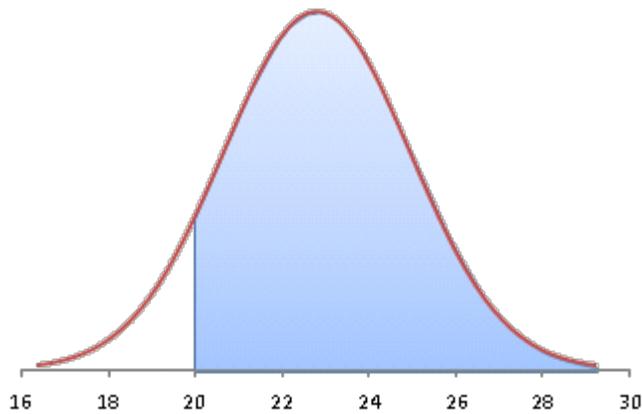
If we randomly select a 1-year-old boy, what is the probability that he'll weigh at least 20 pounds?

Solution:

Let's do this one using technology. We should still start with a sketch:

Using StatCrunch, we get the following result:

Mean: 22.8	Std. Dev.: 2.15
Prob(X => ▼ 20) =	0.9035975
Close Compute	

According to these results, it looks like there's a probability of about 0.9036 that a randomly selected 1-year-old boy will weigh more than 20 lbs.

Why don't you try a couple?

Example 6

Suppose that the volume of paint in the 1-gallon paint cans produced by Acme Paint Company is approximately normally distributed with a mean of 1.04 gallons and a standard deviation of 0.023 gallons.

Photo: A Syed

What is the probability that a randomly selected 1-gallon can will actually contain at least 1 gallon of paint?

[reveal answer]

Example 7

Suppose the amount of light (in lumens) emitted by a particular brand of 40W light bulbs is normally distributed with a mean of 450 lumens and a standard deviation of 20 lumens.

What percentage of bulbs emit between 425 and 475 lumens?

[reveal answer]

Finding Values

The next type of question comes from the other direction. Instead of giving values and asking for the probability, we'll now be looking at problems where the probability is known, but the values are not. Questions like:

- What is the 90th percentile for the weights of 1-year-old boys?
- What IQ score is below 80% of all IQ scores?
- What weight does a 1-year-old boy need to be so all but 5% of 1-year-old boys weight less than he does?

As with the previous types of problems, we'll learn how to do this using both the table and technology. **Make sure you know both methods - they're both used in many fields of study!**

Finding Z-Scores Using the Table

The idea here is that the values in the table represent area to the left, so if we're asked to find the value with an area of 0.02 to the left, we look for 0.02 on the *inside* of the table and find the corresponding Z-score.

z	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
-3.4	0.0003	0.0003	0.0003	0.0003	0.0003	0.0003	0.0003	0.0003	0.0003	0.0002
-3.3	0.0005	0.0005	0.0005	0.0004	0.0004	0.0004	0.0004	0.0004	0.0004	0.0003
-3.2	0.0007	0.0007	0.0006	0.0006	0.0006	0.0006	0.0006	0.0005	0.0005	0.0005
-2.2	0.0139	0.0136	0.0132	0.0129	0.0125	0.0122	0.0119	0.0116	0.0113	0.0110
-2.1	0.0179	0.0174	0.0170	0.0166	0.0162	0.0158	0.0154	0.0150	0.0146	0.0143
-2.0	0.0228	0.0222	0.0217	0.0212	0.0207	0.0202	0.0197	0.0192	0.0188	0.0183
-1.9	0.0287	0.0281	0.0274	0.0268	0.0262	0.0256	0.0250	0.0244	0.0239	0.0233

Since we don't have an area of *exactly* 0.02, we have to think a bit. We have two choices: (1) take the closest area, or (2) average the two values if it's equidistant from the two areas.

In this case, it's almost equidistant, so we'll take the average and say that the Z-score corresponding to this area is the average of -2.05 and -2.06, so **-2.055**.

Finding Z-Scores Using StatCrunch

Click on **Stat** > **Calculators** > **Normal**

Enter the mean, standard deviation, the direction of the inequality, and the probability (leave X blank). Then press Compute. The image below shows the Z-score with an area of 0.05 to the right.

Mean: 0   Std. Dev.: 1

Prob(X => ▼ 1.6448536) = 0.05

Close   Compute

Let's try a few!

Example 8

a. Find the Z-score with an area of 0.90 to the left.

[reveal answer]

b. Find the Z-score with an area of 0.10 to the right.

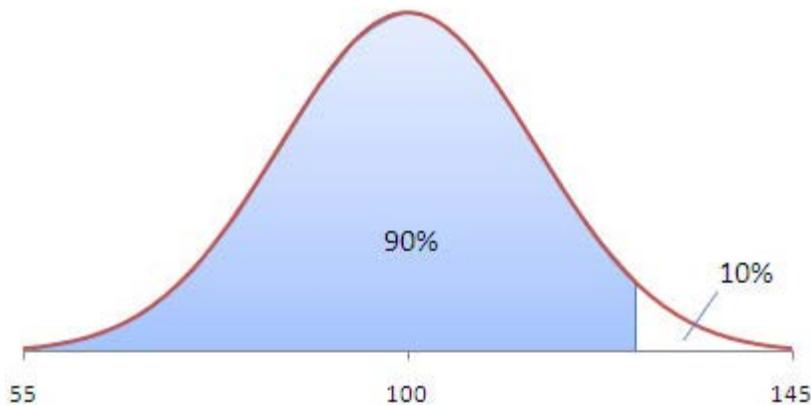c. Find the Z-score such that P(Z < z_0) = 0.025.

So we've talked about how to find a z-score given an area. If you remember, the technology instructions didn't specify that the distribution needed to be the *standard* normal - we actually find values in *any* normal distribution that correspond to a given area/probability using those same techniques.

Example 9

Referring to IQ scores again, with a mean of 100 and a standard deviation of 15. Find the 90th percentile for IQ scores.

Solution:

First, we need to translate the problem into an area or probability. In Section 3.4, we said the **kth percentile** of a set of data divides the lower *k*% of a data set from the upper (100-*k*)%. So the 90th percentile divides the lower 90% from the upper 10% - meaning it has about 90% below and about 10% above.

Using StatCrunch, we get the following result:

Mean: 100   Std. Dev.: 15

Prob(X <= ▼ 119.223274) = 0.9

Therefore, the 90th percentile for IQ scores is about 119.

Example 10

Suppose that the volume of paint in the 1-gallon paint cans produced by Acme Paint Company is approximately normally distributed with a mean of 1.04 gallons and a standard deviation of 0.023 gallons.

Photo: A Syed

What volume can the Acme Paint Company say that 95% of their cans exceed?

Example 11

Referring to the weights of 1-year-old boys again. (The weights of 1-year-old boys are approximately normally distributed, with a mean of 22.8 lbs and a standard deviation of about 2.15.)

What weight does a 1-year-old boy need to be so all but 5% of 1-year-old boys weight less than he does?

[reveal answer]

Finding z_α

The notation z_α ("z-alpha") is the Z-score with an area of α to the right.

α area to the right

Z_α
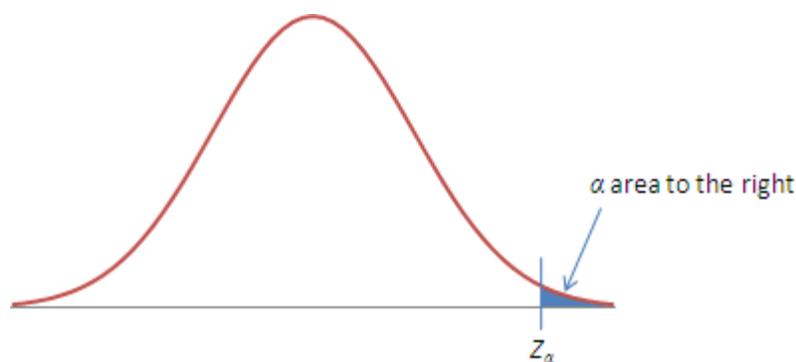
The concept of z_α is used extensively throughout the remainder of the course, so it's an important one to be comfortable with. The applications won't be immediately obvious, but the essence is that we'll be looking for events that are unlikely - and so have a very small probability in the "tail".

Let's try some examples.

Example 12

a. Find $z_{0.01}$

[reveal answer]

b. Find $z_{0.05}$

[reveal answer]

c. Find $z_{0.025}$

[reveal answer]

# Section 7.3: Assessing Normality

7.1 Properties of the Normal Distribution
7.2 Applications of the Normal Distribution
**7.3 Assessing Normality**

Objectives

By the end of this lesson, you will be able to...

1. find and interpret the area under a normal curve
2. find the value of a normal random variable

---

Earlier in the course, in Section 2.2, we learned that we can characterize the distribution shape of a random variable using a histogram. One of those distribution shapes was **bell-shaped (symmetric)**.
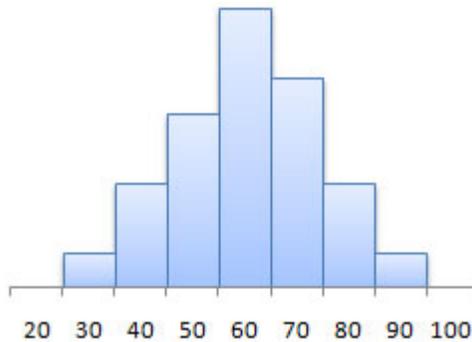
Later, in Section 7.1, we defined a **normally distributed random variable** to be one whose histogram follows the normal (bell-shaped) curve:
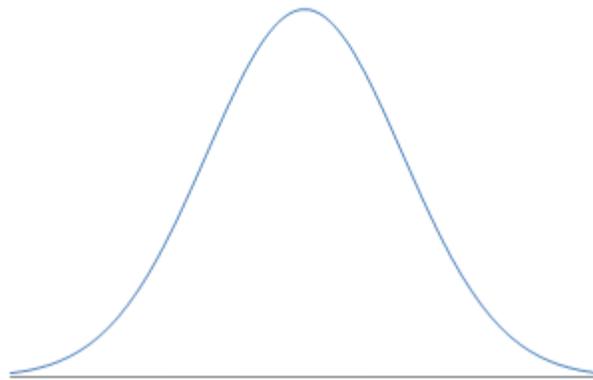
So if we have the histogram, we can determine whether or not the random variable follows the normal distribution.
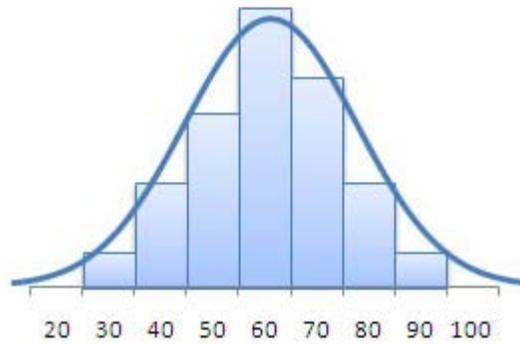
What happens, though, when the sample size is so small that we can't really see the distribution shape in the histogram? We need another method, which brings us to the topic for this section.

The Normal Probability Plot

A normal probability plot is a graph that plots the observed data versus the *normal score*, which is what we would expect if the data actually followed the standard normal distribution.

In other words, if we have 15 observations, the 10th normal score would be the *expected 10th value if the data followed the standard normal distribution*.

We know from earlier this section that

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

If we solve this equation for X, we get $X = \mu + \sigma Z$, which is the equation for a line. This gets us to the key result:

> If sample data are taken from a population that is normally distributed, a normal probability plot **should be approximately linear**.

Constructing a Normal Probability Plot Using Technology

> Unfortunately, StatCrunch doesn't have a method of producing this plot, so we'll instead be doing a Q-Q plot, which is different but offers similar results.

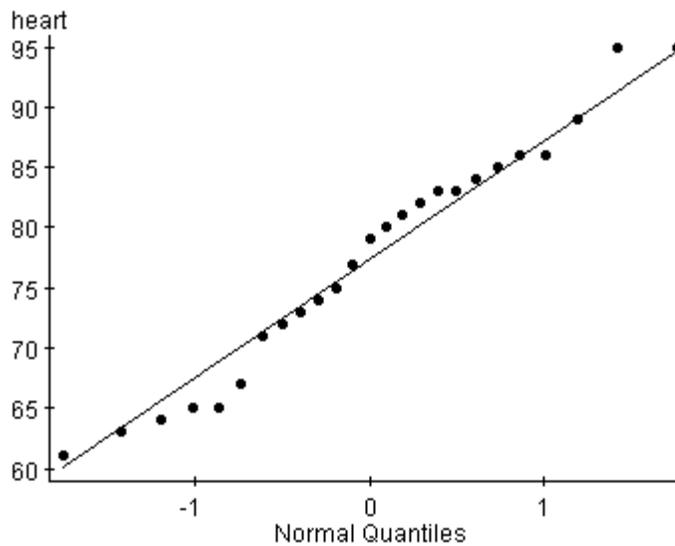You can also go to the video page for links to see videos in either Quicktime or iPod format.

Example 1

Suppose we wish to know whether the resting heart rates of a sample of Mth120 students are normally distributed.

heart rate

| 61	63	64	65	65
67	71	72	73	74
75	77	79	80	81
82	83	83	84	85
86	86	89	95	95

Based on this plot, it does appear as though the resting heart rates are approximately normally distributed. The plot is fairly linear, with just a couple points straying from the line.

Example 2

Suppose we wish to know whether the number of children that students in a particular Mth120 class have in their family is normally distributed.

number of children

3	4	3	1	5
3	2	4	2	5
9	2	3	2	7
3	1	2	6	2
4	3	1	2	2

Number
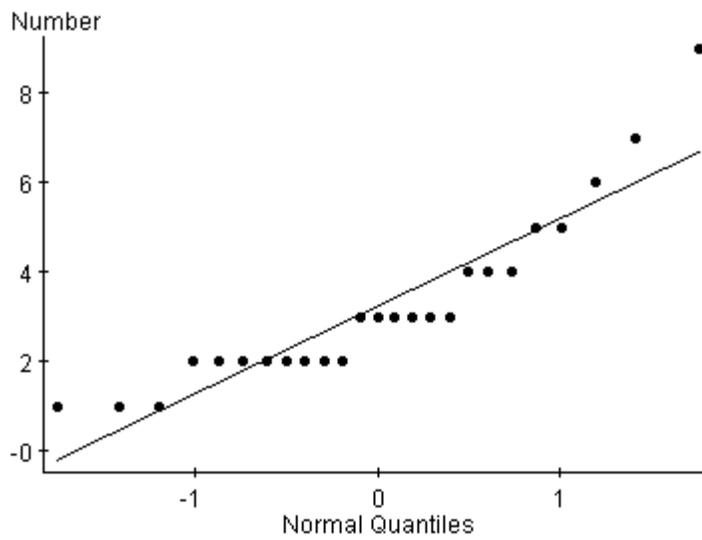
This plot is clearly *not* linear, so the data do not come from a normally distributed population.