We will study a specific bell-shaped distribution, which we first saw in chapter 3.

General Education Statistics
Class Notes
The Normal Probability Distribution (Section 7.1)

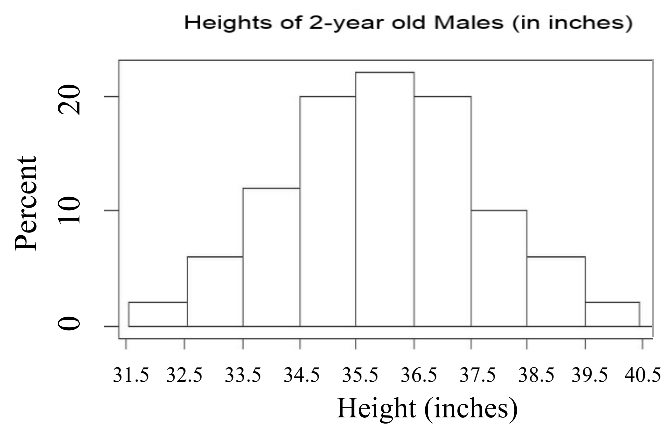
In this chapter, we will be discussing *continuous* variables. There is a big difference between discrete and continuous variables that will change up how we do things.

Many variables are considered to be normally distributed, which we will define later in this section. Heights of people, birth weights, and monthly charges for cell phone plans are all examples of continuous variables that are said to be normally distributed (or that they take on a normal distribution).

What does the Normal Curve do for us?

expl 1a: To the right is a histogram that shows sample data for the heights of 50 two-year-old boys. The mean is 36 inches as you will see in the middle of the histogram.

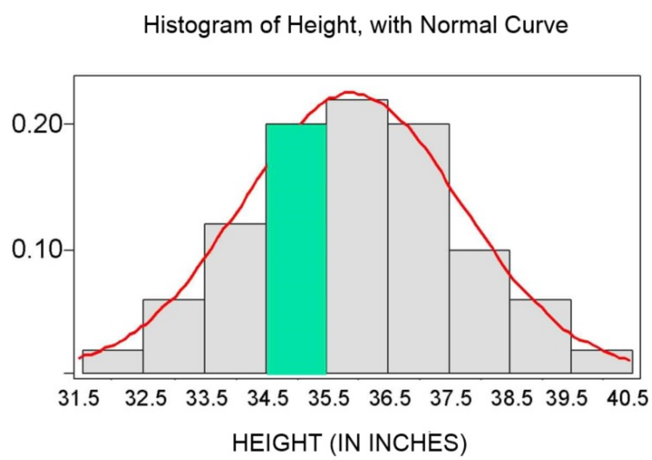
Shade the rectangle that gives us the probability that a randomly selected boy would be between 34.5 and 35.5 inches tall. Estimate this probability. (Recall this is the same as asking what percent of these boys are between 34.5 and 35.5 inches tall.)



expl 1b: Consider this same histogram with the normal curve superimposed on top.

Notice how the area under the curve from the *x*-value of 34.5 to the *x*-value of 35.5 would *approximate* the area of this single rectangle. Shade this area *under the curve*.

The fact that the area under the curve *approximates* the area of the histogram is why we can use the normal curve, and *not* bother with the underlying histograms, to find these probabilities.



We will formally define this normal curve in a bit.

The Probability of a Particular Value is Zero:

One big difference between discrete and continuous variables is that we say that a discrete variable *can* take on a single value. For instance, a person can stream four movies in a month from Netflix. We can find the probability that a randomly selected Netflix customer has streamed four movies that month.

But can we find the probability that a two-year-old boy is exactly 35.5 inches tall?

Again, our variable is continuous. Because an infinite number of possibilities exist, the probability of *one particular* value is said to be 0. This comes from our classical understanding of probability as the number of successes divided by the number of possibilities. But, again, we have an infinite number of possibilities. If you divide by infinity, you get 0.

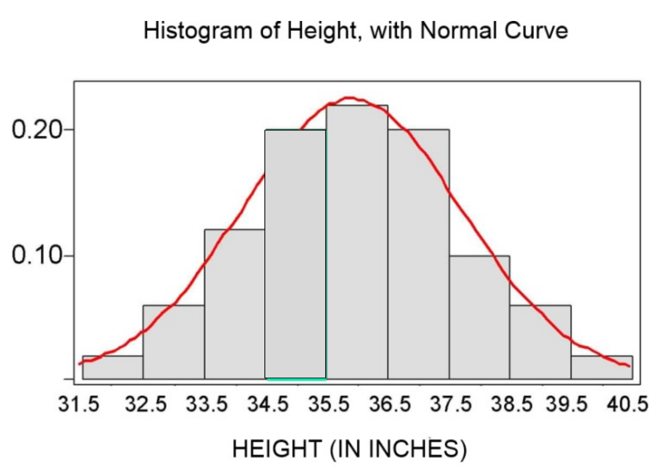
To resolve this problem, we will always calculate these probabilities for intervals of values.

The area under the graph represents the probability of observing a value in that interval. There is *no* area for a single value.

Definition: A continuous random variable is **normally distributed**, or has a **normal probability distribution**, if its relative frequency histogram of the random variable has the shape of a normal curve (bell-shaped and symmetric).

Consider, again, the normal curve to the right.

expl 1c: Shade the area *under the curve* that represents the percent of these two-year-old boys that are between 33.5 and 36.5 inches tall. Estimate this area.

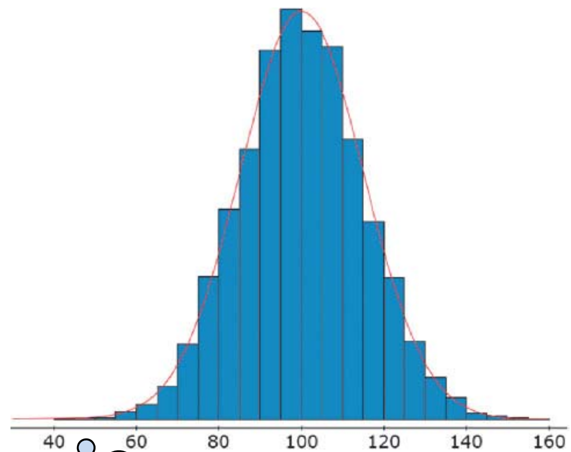


expl 1d: Can we shade the area that would represent the percent of these boys that are *exactly* 38.5 inches tall? Explain.

The Normal Distribution:

We have seen bell-shaped, symmetric distributions before in chapter 3. Here, we look at a special one called the normal distribution. Some examples of variables that take on this shape are IQ, distances a golf ball travels, incubation times for eggs, and many more.

Consider this histogram. It represents the IQ's of people taken in a sample of 10,000.



Notice the greatest number of people in a single class is near the middle. In fact, most people are near to the middle. As you get further from this middle, on either side, you have fewer and fewer people.

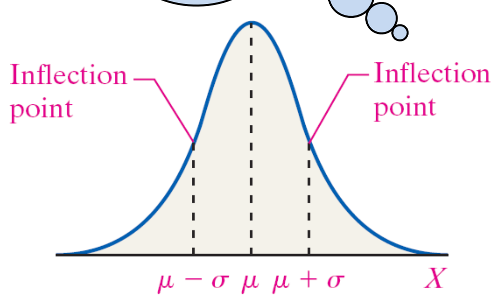
The red (skinny) curve follows the shape of the histogram. In fact, as the number of classes increases, the curve fits the histogram better.

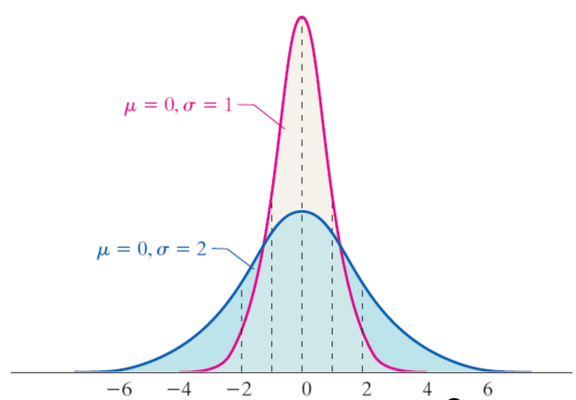
Again, we will find probabilities by approximating the area of the histogram by finding the area under this curve.

The red (skinny) curve is the normal density function. It has an actual formula but we do *not* bother with it here.

A normal curve looks like this. The exact shape is determined by its mean and standard deviation.

The curve flares out a bit, *unlike* parabolas you may have studied.





The general shape is the same. A smaller standard deviation makes it skinnier. Do you know why?

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

1. The total area under its graph must equal 1.
2. The graph will *not* dip below the x-axis.

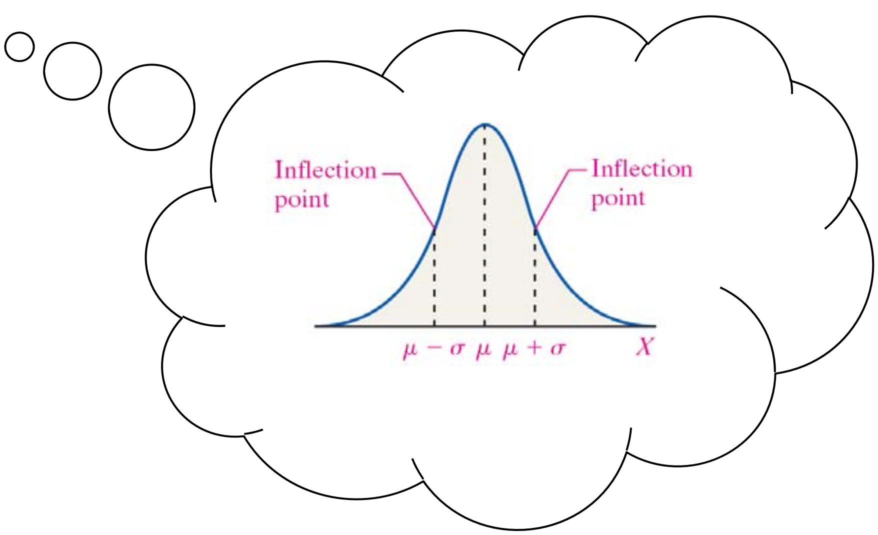
We have seen these ideas before in probability.

To find probabilities, we graph this equation and look for the area under it. In practice, we draw a representative curve, shade the area of interest, and find this area using technology or a table.

The *Normal* Density Function is one of many pdf's. The graph of this function is what we call the normal density curve.

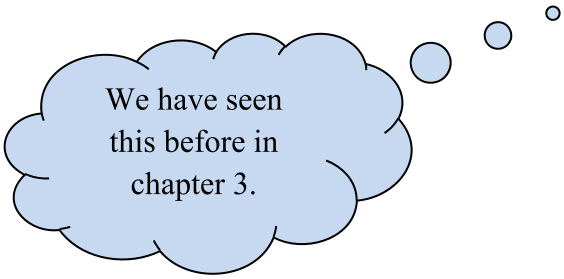
Properties of the Normal Density Curve:

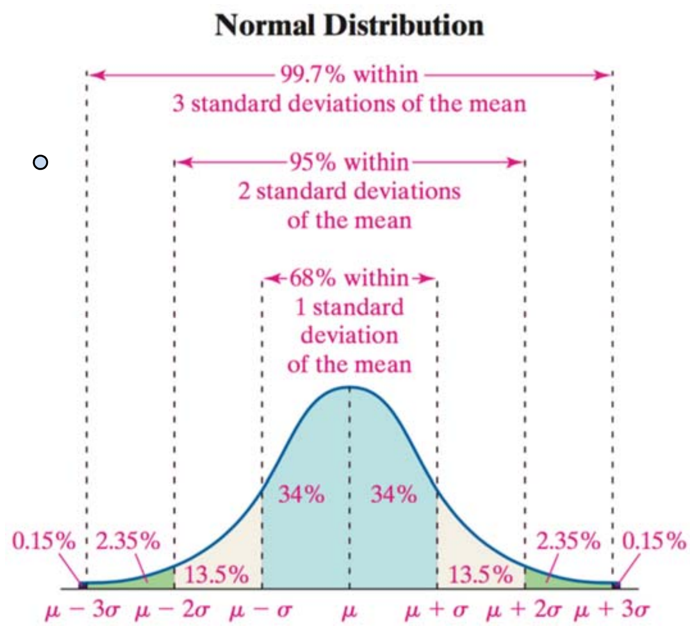
1. It is symmetric about its mean, μ.
2. Because mean = median = mode, the curve has a single peak and the highest point occurs at $x = \mu$.
3. It has inflection points at $\mu + \sigma$ and $\mu - \sigma$. This gives the normal curve a little bit of flare in the middle.
4. The total area under the curve is 1. (This is seen above as pdf property 1.)
5. The area under the curve to the right of μ is equal to the area under the curve to the left of μ. This area is ½.
6. As x increases without bound (gets larger and larger), the graph approaches, but never reaches, the horizontal axis. As x decreases without bound (gets more and more negative), the graph approaches, but never reaches, the horizontal axis. (This is seen above as pdf property 2.)



The Empirical Rule, Again:

Approximately 68% of the area under the normal curve is between $x = \mu - \sigma$ and $x = \mu + \sigma$.
Approximately 95% of the area is between $x = \mu - 2\sigma$ and $x = \mu + 2\sigma$.
Approximately 99.7% of the area is between $x = \mu - 3\sigma$ and $x = \mu + 3\sigma$.





expl 2: The birth weights of full-term babies are normally distributed with mean μ = 3400 grams (7.5 pounds) and σ = 505 grams. Answer the following questions.

a.) Draw a normal curve with the mean and plus or minus two standard deviations marked.

b.) Use the Empirical Rule to *estimate* the probability that a random full-term baby would weigh more than 4410 grams. On your curve, shade the area we are interested in.

c.) Suppose the actual area under the normal curve is found to be .0228 (using methods discussed in the next section and on this section's worksheet). Write a sentence to interpret this result.

Interpreting the Area under a Normal Curve:
Suppose that a random variable X is normally distributed with mean μ and standard deviation σ. The area under the normal curve for any interval of values of the random variable X represents either

1. the percentage (proportion) of the population with the characteristic described by the interval of values, or
2. the probability that a randomly selected individual from the population will have the characteristic described by the interval of values.

expl 3: The lives of refrigerators are normally distributed with mean μ = 14 years and standard deviation σ = 2.5 years.
a.) Draw a normal curve with the parameters labeled.

b.) Shade the region under the curve that represents the proportion of refrigerators that have a life span between 15 and 18 years. (Notice these values do *not* coincide with the values mentioned in the Empirical Rule.)

c.) Suppose this area is found to be 0.2898. Interpret this result.

We will learn how to find these areas on the following worksheet and in the next section.

Optional Worksheet: Normal distribution:
This worksheet works on the idea of using histograms to find probabilities and how the normal curve approximates those probabilities. To start us off, we first study a symmetric histogram and notice how the bell-shaped curve fits it nicely. The areas of the bars of a histogram can give you probabilities. This will get us used to the idea of equating probabilities to the area under a probability curve. We will find probabilities for what we call normal probability distributions. This worksheet covers material covered in sections 7.1 and 7.2. Numbers one through five should be done for this section.