**The normal distribution**

The next probability distribution we'll discuss is the normal distribution. It's one of the most important probability distributions you'll learn about since a countless number of statistical methods rely on it, and it applies to more real-world situations than the distributions we've covered so far.

**What is the normal distribution?**

The normal distribution looks like this. Its shape is commonly referred to as a "bell curve". The normal distribution has a few important properties.

**Symmetrical**

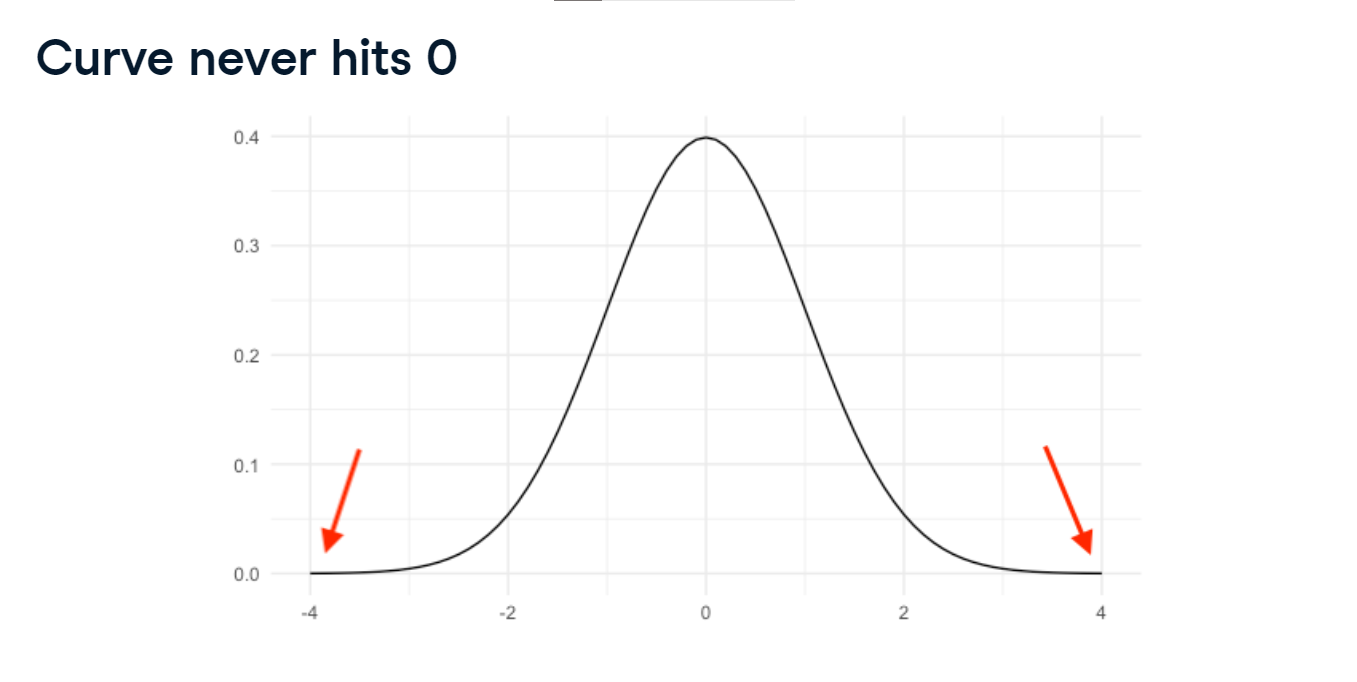
First, it's symmetrical, so the left side is a mirror image of the right.

**Area = 1**

Second, just like any continuous distribution, the area beneath the curve is 1.

**Curve never hits 0**

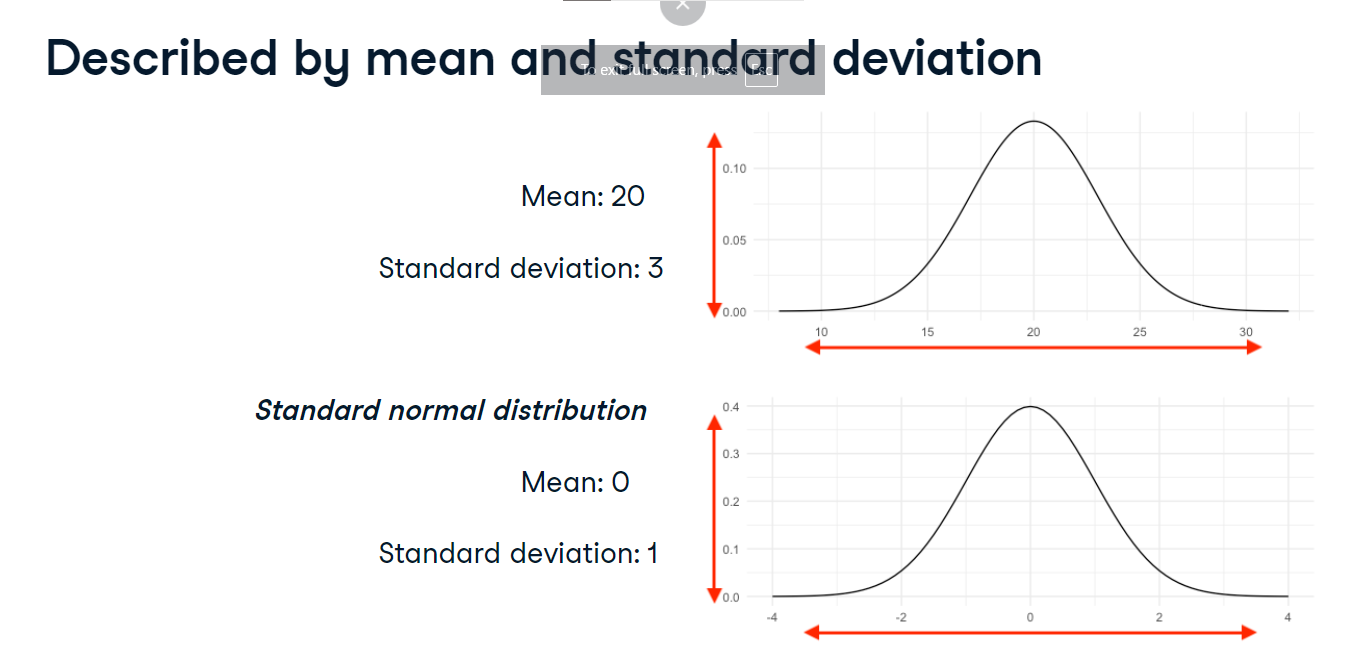
Second, the probability never hits 0, even if it looks like it does at the tail ends. Only 0-point-006% of its area is contained beyond the edges of this graph.



**Described by mean and standard deviation**

The normal distribution is described by its mean and standard deviation. Here is a normal distribution with a mean of 20 and standard deviation of 3, and here is a normal distribution with a mean of 0 and a standard deviation of 1. When a normal distribution has mean 0 and a standard deviation of 1, it's a special distribution called the standard normal distribution Notice how both distributions have the same shape,

but their axes have different scales.

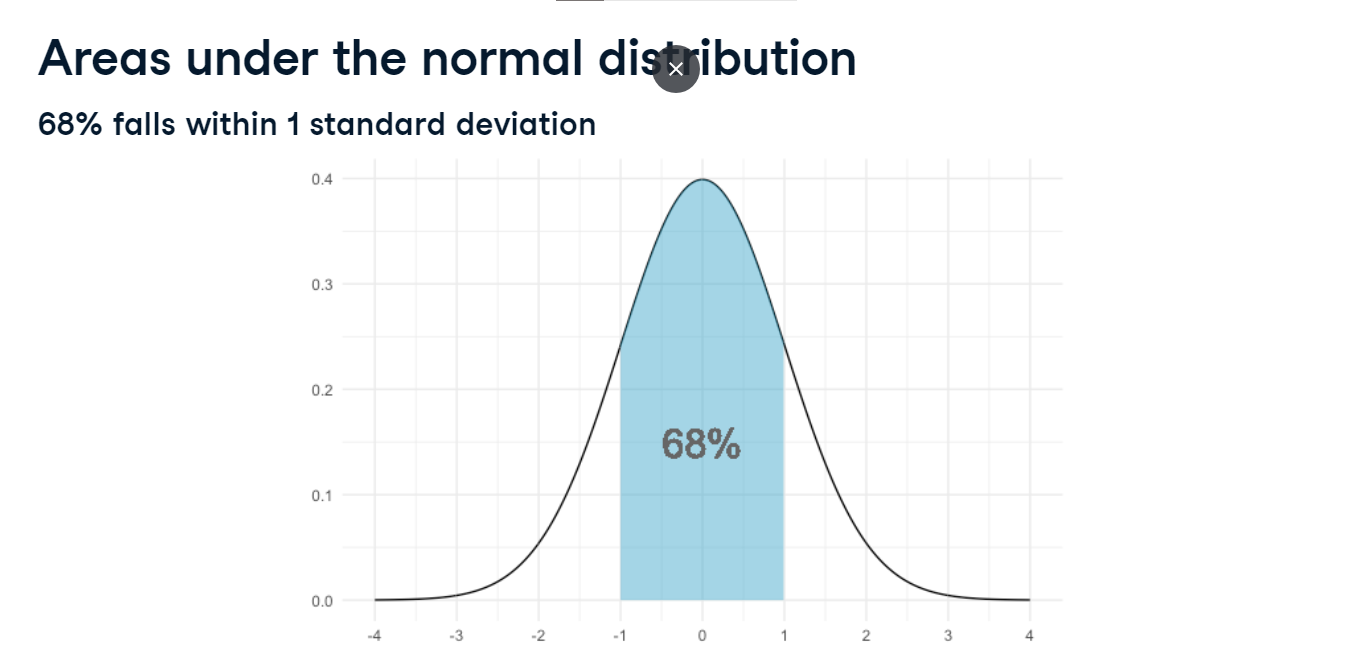


**Areas under the normal distribution**

For the normal distribution, 68% of the area is within 1 standard deviation of the mean. 95% of the area falls within 2 standard deviations of the mean,

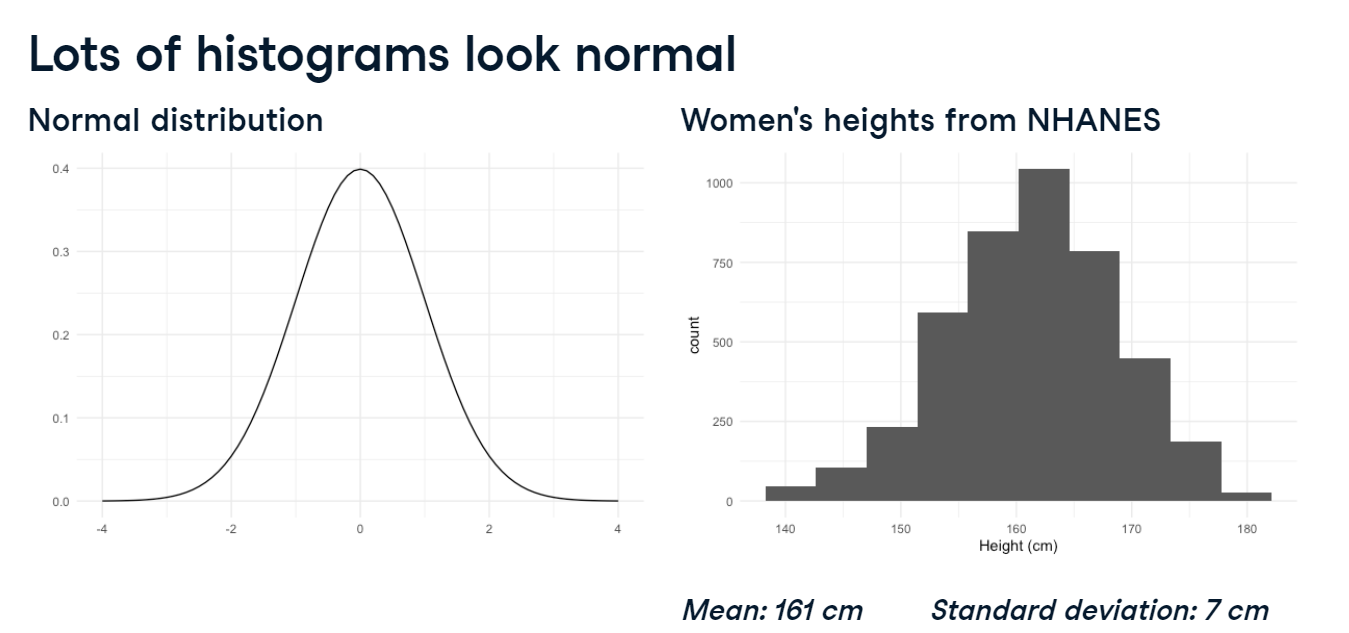
**Areas under the normal distribution**

and 99.7% of the area falls within three standard deviations. This is sometimes called the 68-95-99-point-7 rule.



**Lots of histograms look normal**

There's lots of real-world data shaped like the normal distribution. For example, here is a histogram of the heights of women that participated in the National Health and Nutrition Examination Survey. The mean height is around 161 centimeters and the standard deviation is about 7 centimeters.

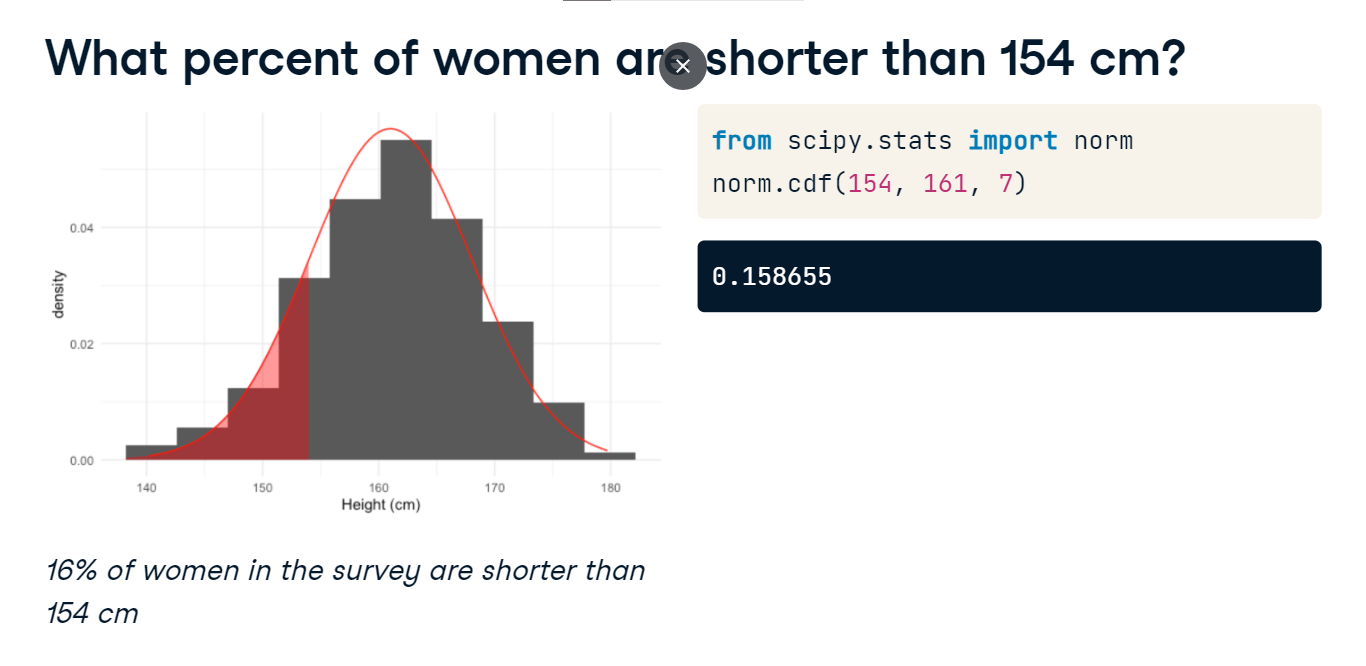


**Approximating data with the normal distribution**

Since this height data closely resembles the normal distribution, we can take the area under a normal distribution with mean 161 and standard deviation 7 to approximate what percent of women fall into different height ranges.

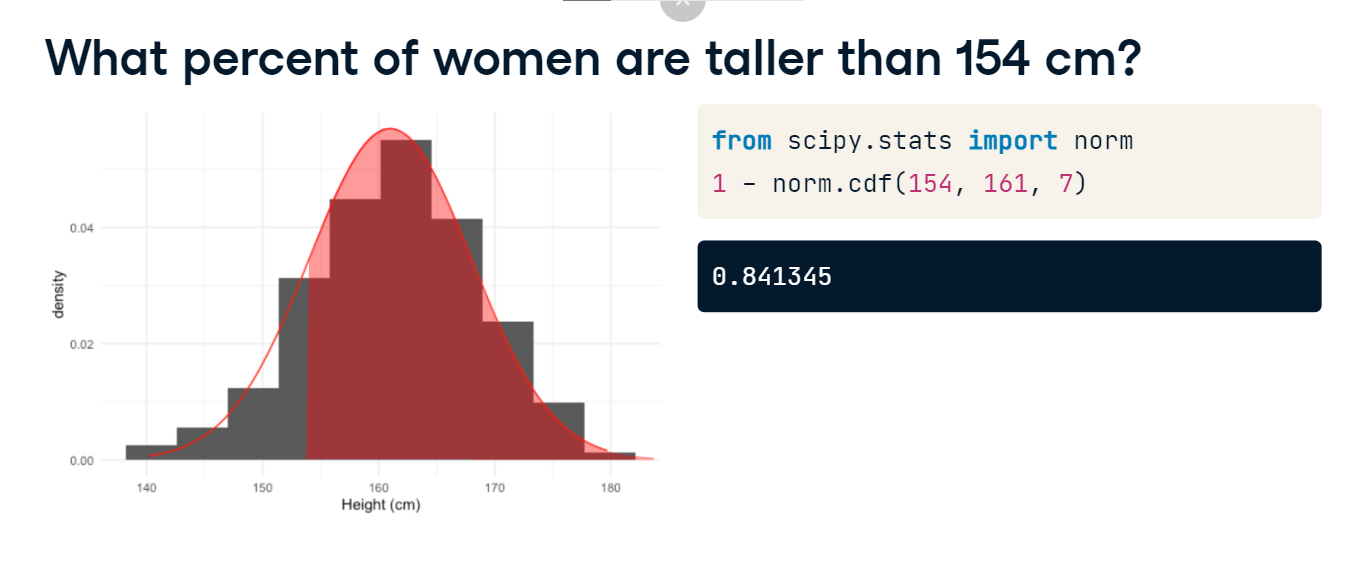
**What percent of women are shorter than 154 cm?**

For example, what percent of women are shorter than 154 centimeters? We can answer this using norm-dot-cdf from scipy-dot-stats, which takes the area of the normal distribution less than some number. We pass in the number of interest, 154, followed by the mean and standard deviation of the normal distribution we're using. This gives us about 16% of women are shorter than 154 centimeters.



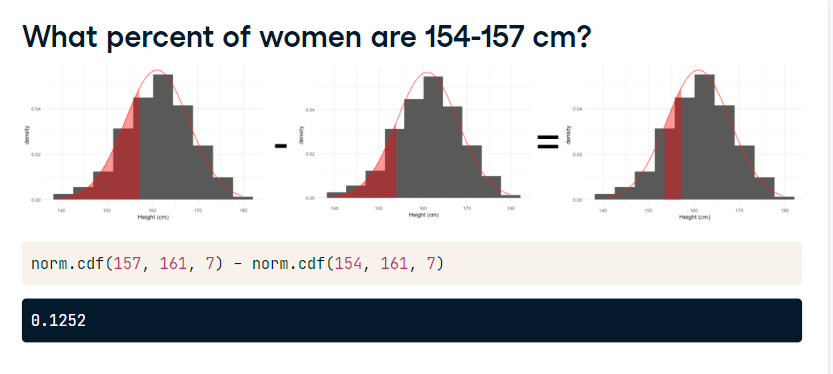
**What percent of women are taller than 154 cm?**

To find the percent of women taller than 154 centimeters, we can take 1 minus the area on the left of 154, which equals the area to the right of 154.



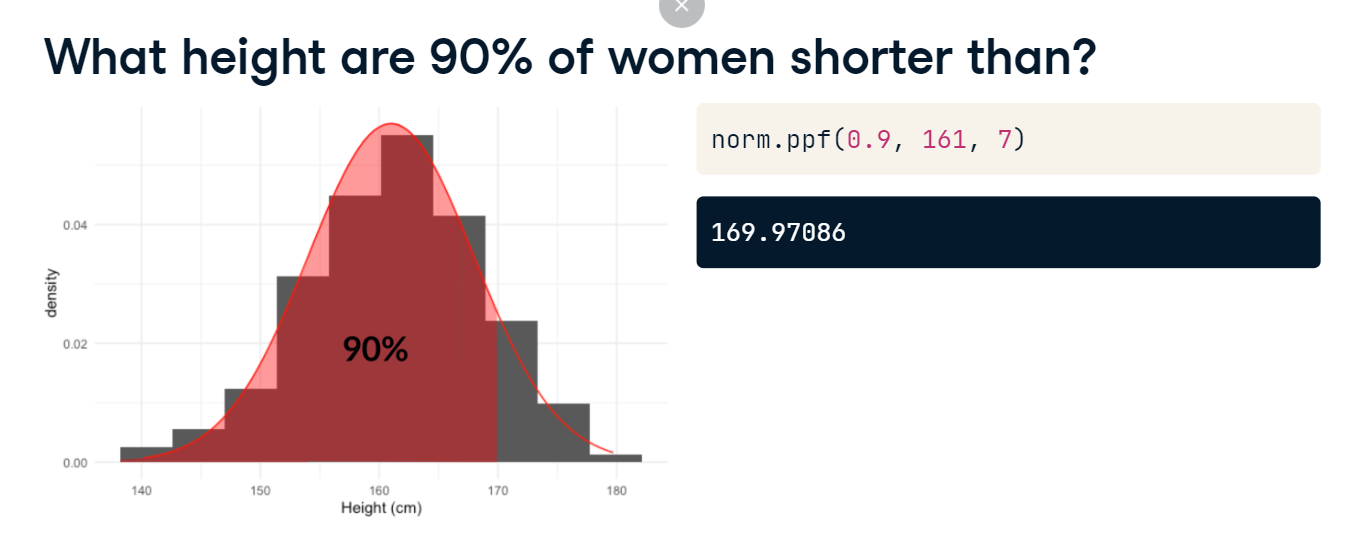
**What percent of women are 154-157 cm?**

To get the percent of women between 154 and 157 centimeters tall we can take the area below 157 and subtract the area below 154, which leaves us the area between 154 and 157.



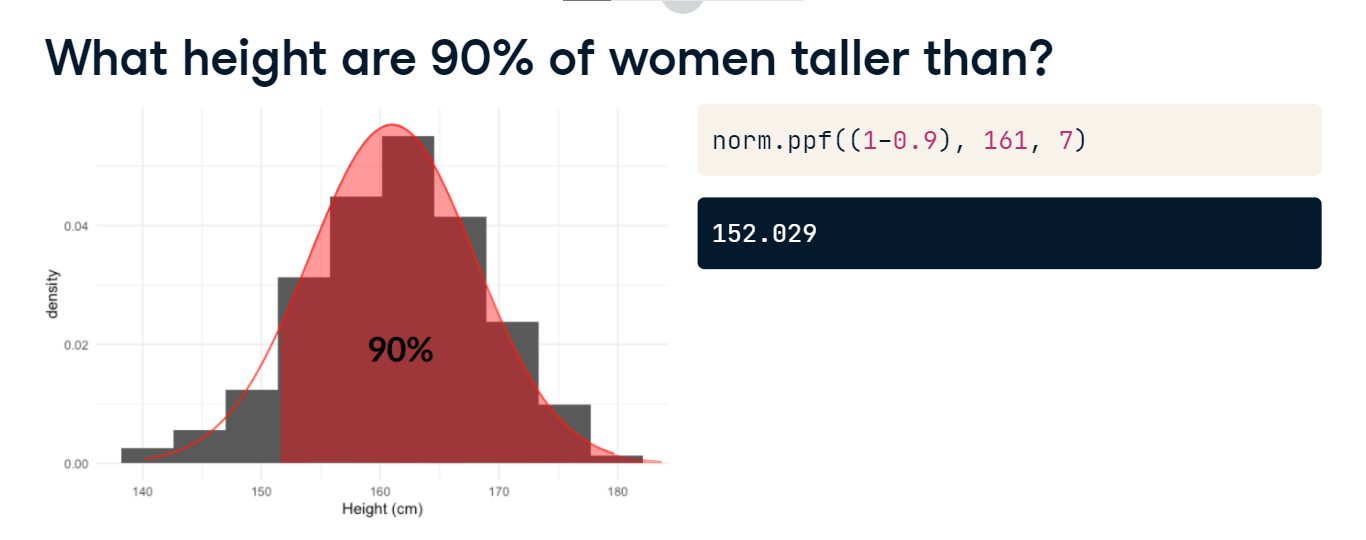
**What height are 90% of women shorter than?**

We can also calculate percentages from heights using norm-dot-ppf. To figure out what height 90% of women are shorter than, we pass 0-point-9 into norm-dot-ppf along with the same mean and standard deviation we've been working with. This tells us that 90% of women are shorter than 170 centimeters tall.



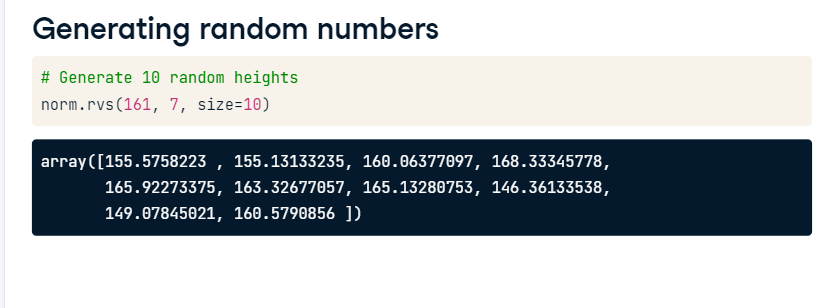
**What height are 90% of women taller than?**

We can figure out the height 90% of women are taller than, since this is also the height that 10% of women are shorter than. We can take 1 minus point-9 to get point-1, which we'll use as the first argument of norm-ppf.



**Generating random numbers**

Just like with other distributions, we can generate random numbers from a normal distribution using norm-dot-rvs, passing in the distribution's mean and standard deviation, as well as the sample size we want. Here, we've generated 10 more random heights.

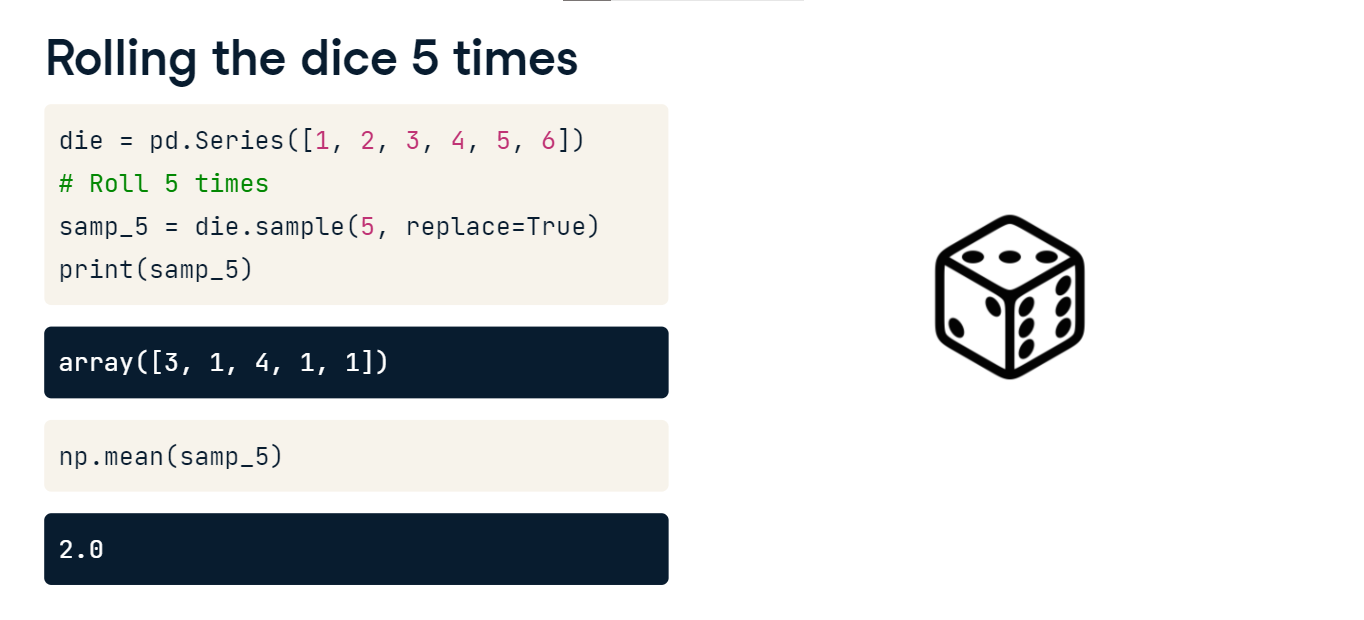


**The central limit theorem**

Now that you're familiar with the normal distribution, it's time to learn about what makes it so important.

**Rolling the dice 5 times**

Let's go back to our dice rolling example. We have a Series of the numbers 1 to 6 called die. To simulate rolling the die 5 times, we'll call die-dot-sample. We pass in the Series we want to sample from, the size of the sample, and set replace to True. This gives us the results of 5 rolls. Now, we'll take the mean of the 5 rolls, which gives us 2.



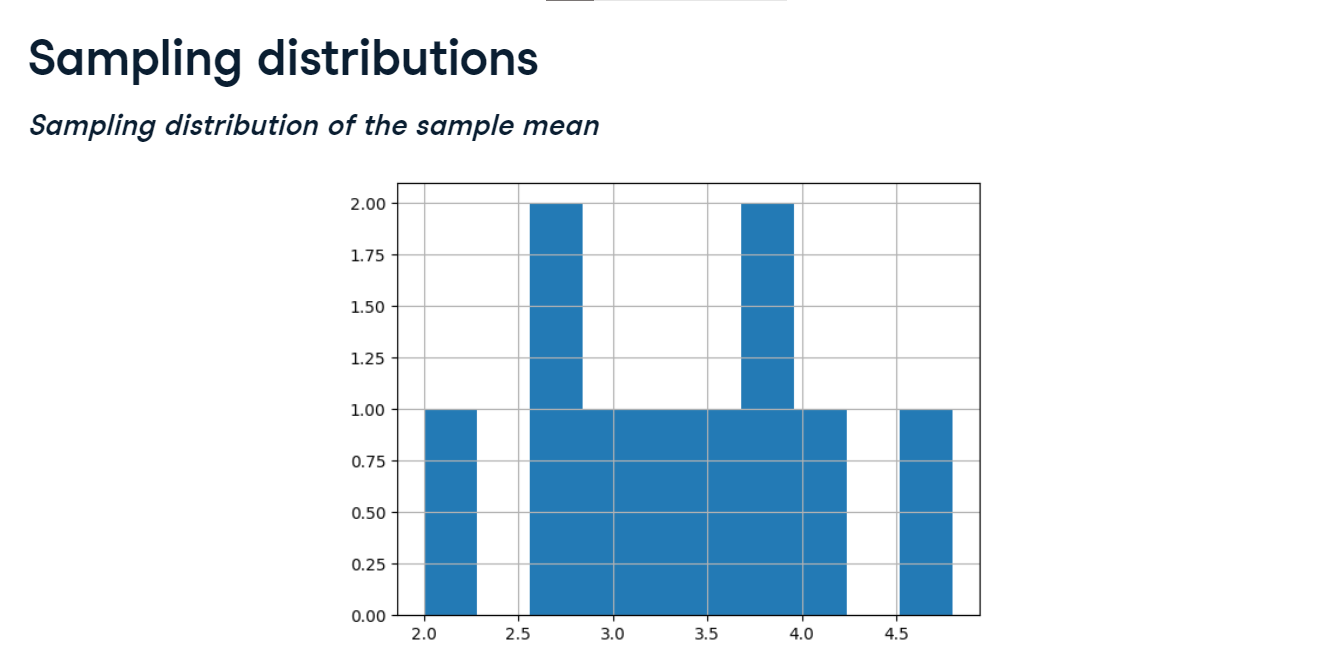
**Rolling the dice 5 times**

If we roll another 5 times and take the mean, we get a different mean. If we do it again, we get another mean.



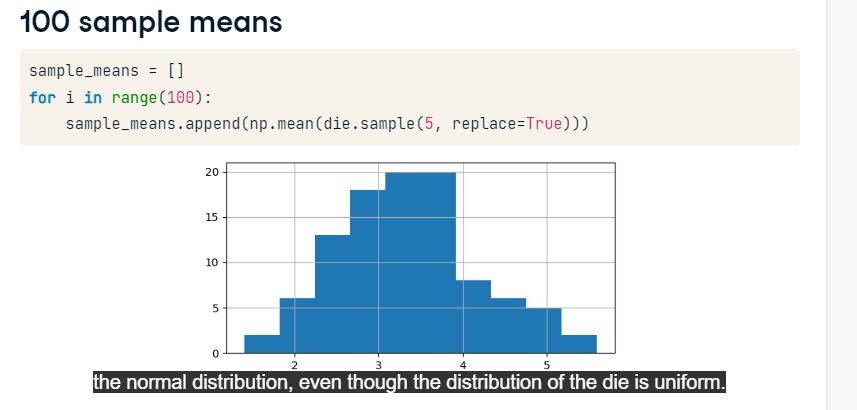
**Rolling the dice 5 times 10 times**

Let's repeat this 10 times: we'll roll 5 times and take the mean. To do this, we'll use a for loop. We start by creating an empty list called sample\_means to hold our means. We loop from 0 to 9 so that the process is repeated 10 times. Inside the loop, we roll 5 times and append the sample's mean to the sample\_means list. This gives us a list of 10 different sample means. Let's plot these sample means.



**Sampling distributions**

A distribution of a summary statistic like this is called a sampling distribution. This distribution, specifically, is a sampling distribution of the sample mean.



**100 sample means**

Now let's do this 100 times. If we look at the new sampling distribution, its shape somewhat resembles the normal distribution, even though the distribution of the die is uniform.

**1000 sample means**

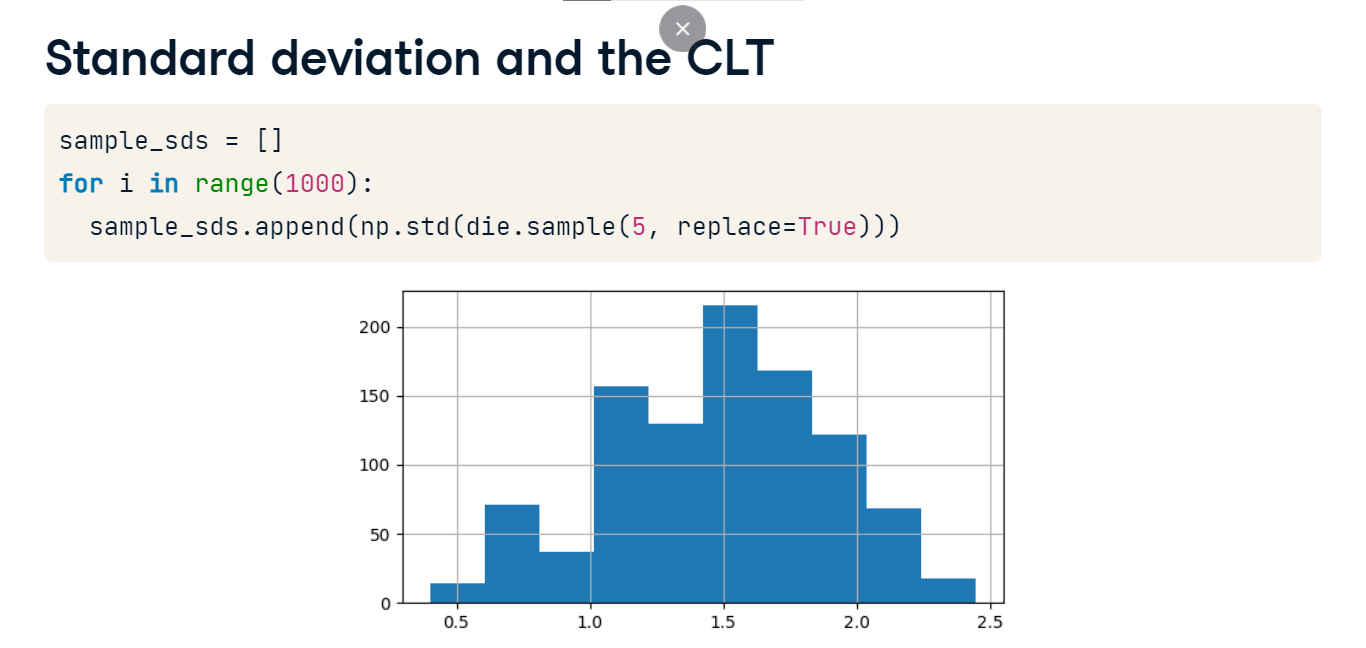
Let's take 1000 means. This sampling distribution more closely resembles the normal distribution.

**Central limit theorem**

This phenomenon is known as the central limit theorem, which states that a sampling distribution will approach a normal distribution as the number of trials increases. In our example, the sampling distribution became closer to the normal distribution as we took more and more sample means. It's important to note that the central limit theorem only applies when samples are taken randomly and are independent, for example, randomly picking sales deals with replacement.

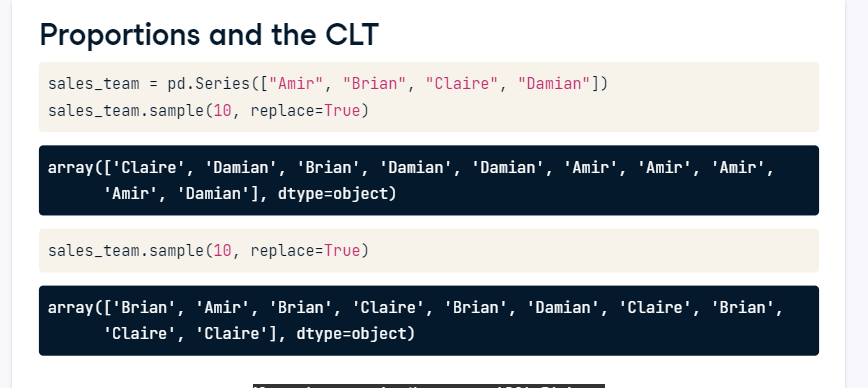
**Standard deviation and the CLT**

The central limit theorem, or CLT, applies to other summary statistics as well. If we take the standard deviation of 5 rolls 1000 times, the sample standard deviations are distributed normally, centered around 1-point-9, which is the distribution's standard deviation.



**Proportions and the CLT**

Another statistic that the CLT applies to is proportion. Let's sample from the sales team 10 times with replacement and see how many draws have Claire as the outcome. In this case, 10% of draws were Claire. If we draw again, there are 40% Claires.

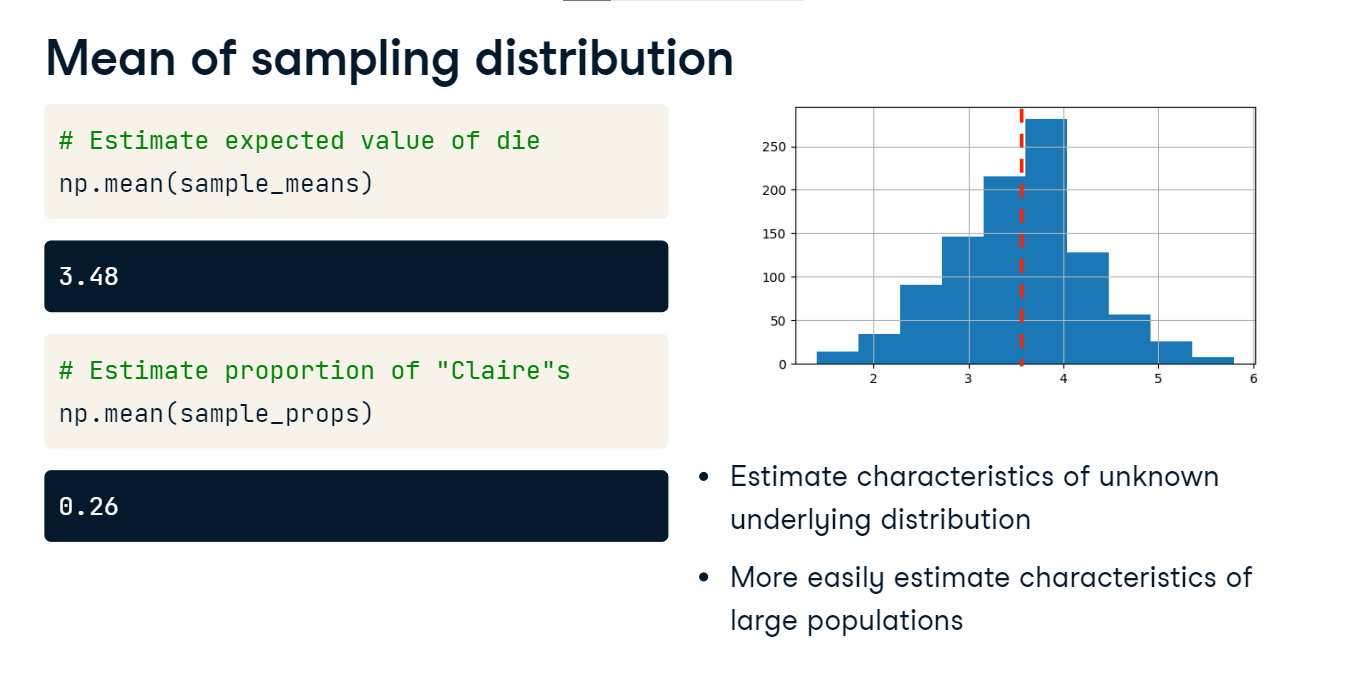


**Sampling distribution of proportion**

If we repeat this 1000 times and plot the distribution of the sample proportions, it resembles a normal distribution centered around 0-point-25, since Claire's name was on 25% of the tickets.

**Mean of sampling distribution**

Since these sampling distributions are normal, we can take their mean to get an estimate of a distribution's mean, standard deviation, or proportion. If we take the mean of our sample means from earlier, we get 3-point-48. That's pretty close to the expected value, which is 3-point-5! Similarly, the mean of the sample proportions of Claires isn't far off from 0-point-25. In these examples, we know what the underlying distributions look like, but if we don't, this can be a useful method for estimating characteristics of an underlying distribution. The central limit theorem also comes in handy when you have a huge population and don't have the time or resources to collect data on everyone. Instead, you can collect several smaller samples and create a sampling distribution to estimate what the mean or standard deviation is.

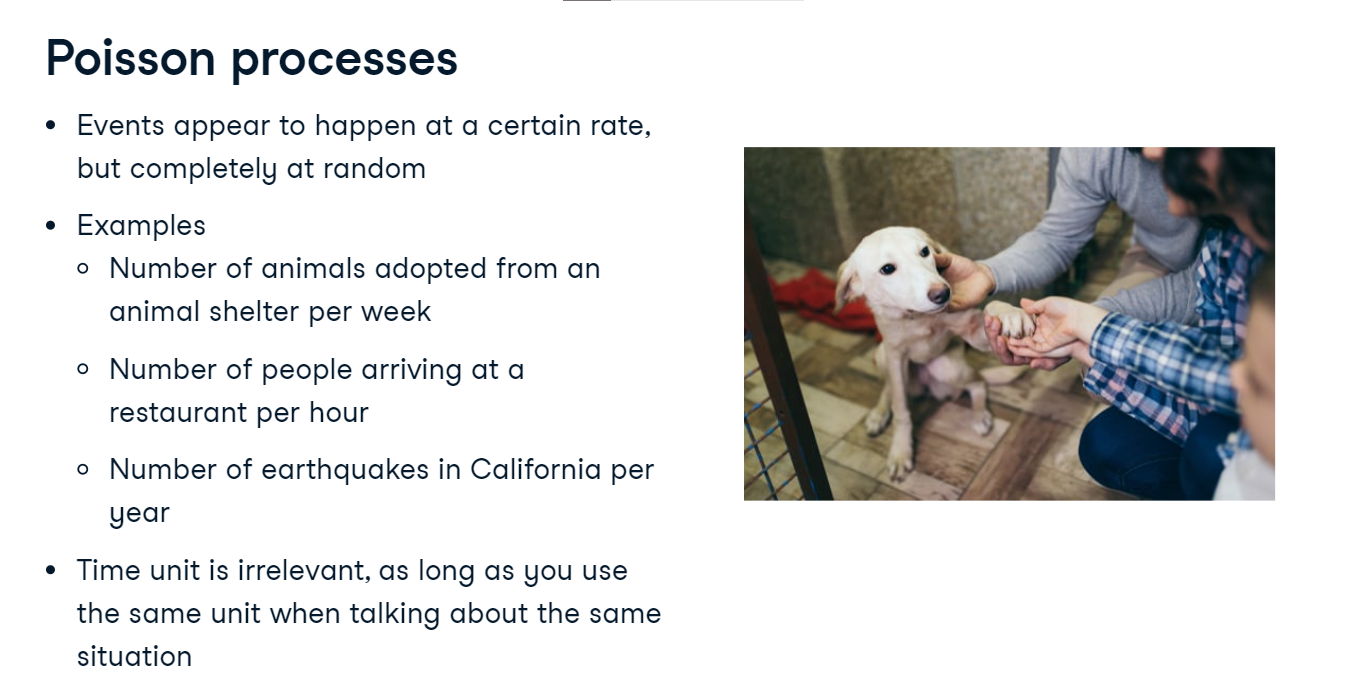


**The Poisson distribution**

In this video, we'll talk about another probability distribution called the Poisson distribution.

**Poisson processes**

Before we talk about probability, let's define a Poisson process. A Poisson process is a process where events appear to happen at a certain rate, but completely at random. For example, the number of animals adopted from an animal shelter each week is a Poisson process - we may know that on average there are 8 adoptions per week, but this number can differ randomly. Other examples would be the number of people arriving at a restaurant each hour, or the number of earthquakes per year in California. The time unit like, hours, weeks, or years, is irrelevant as long as it's consistent.

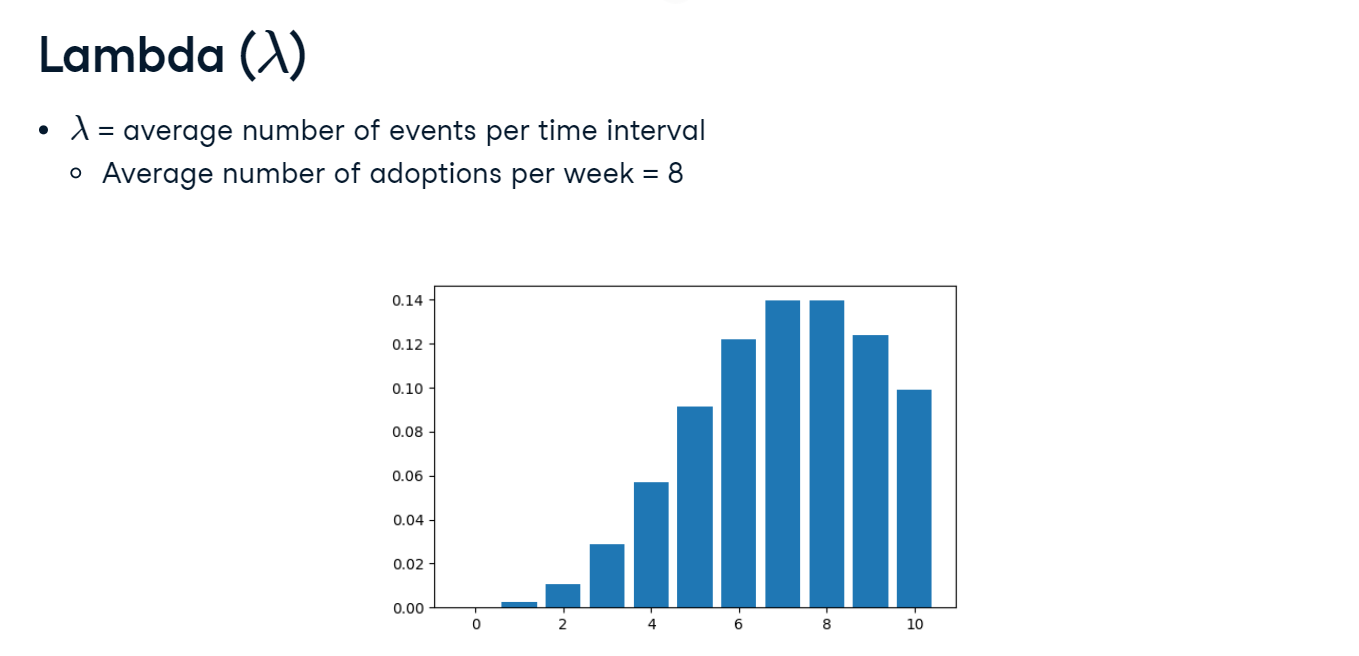


**Poisson distribution**

The Poisson distribution describes the probability of some number of events happening over a fixed period of time. We can use the Poisson distribution to calculate the probability of at least 5 animals getting adopted in a week, the probability of 12 people arriving in a restaurant in an hour, or the probability of fewer than 20 earthquakes in California in a year.

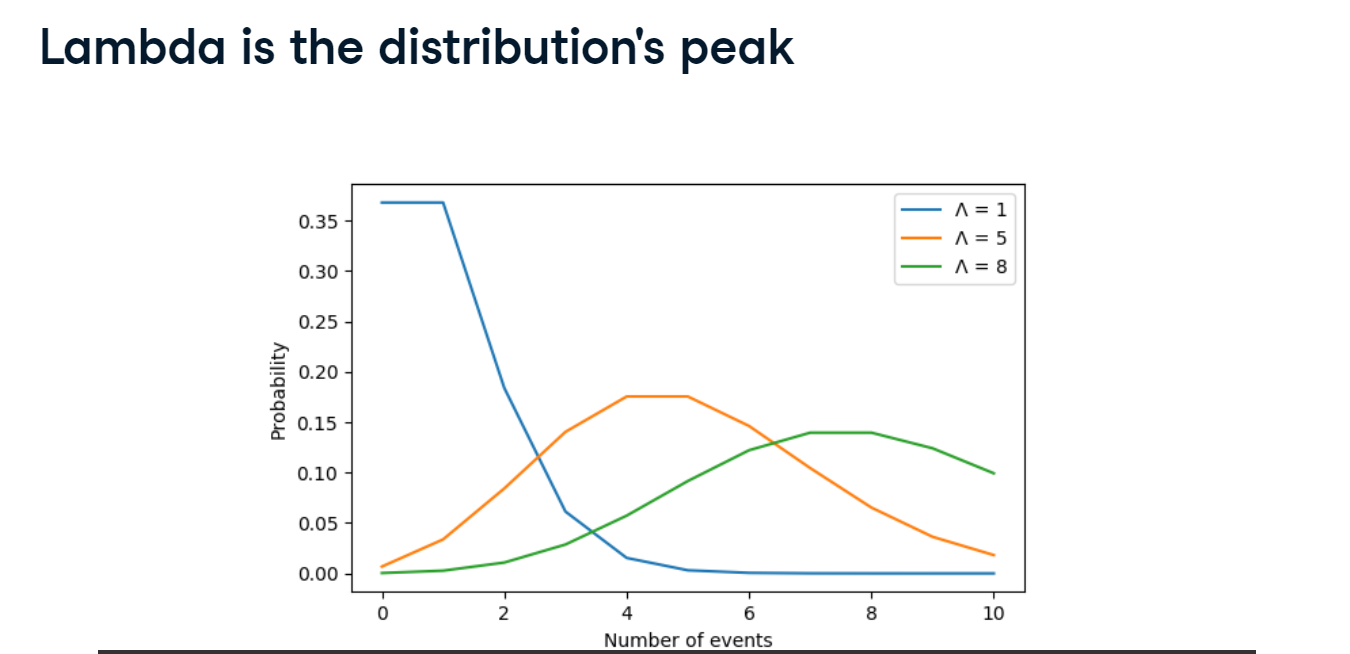
**Lambda ($\lambda$)**

The Poisson distribution is described by a value called lambda, which represents the average number of events per time period. In the animal shelter example, this would be the average number of adoptions per week, which is 8. This value is also the expected value of the distribution! The Poisson distribution with lambda equals 8 looks like this. Notice that it's a discrete distribution since we're counting events, and 7 and 8 are the most likely number of adoptions to happen in a week.



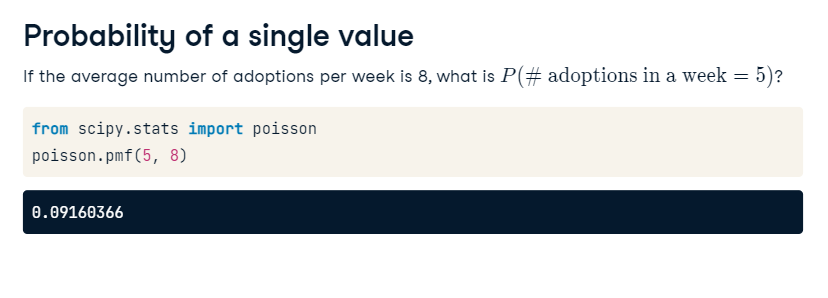
**Lambda is the distribution's peak**

Lambda changes the shape of the distribution, so a Poisson distribution with lambda equals 1, in blue, looks quite different than a Poisson distribution with lambda equals 8, in green, but no matter what, the distribution's peak is always at its lambda value.



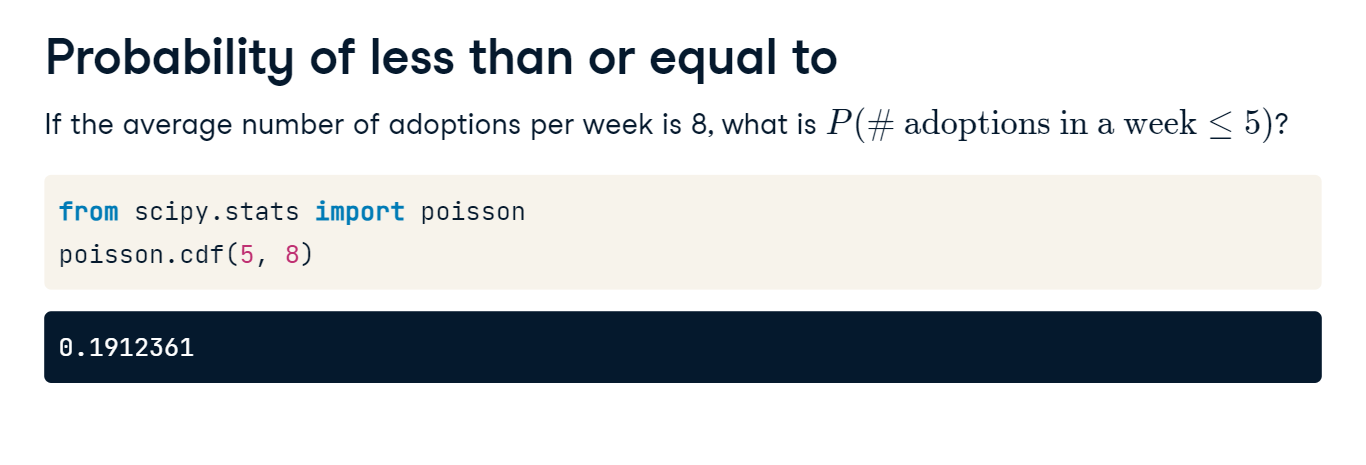
**Probability of a single value**

Given that the average number of adoptions per week is 8, what's the probability of 5 adoptions in a week? Just like the other probability distributions, we can import poisson from scipy-dot-stats. We'll use the poisson-dot-pmf function, passing 5 as the first argument and 8 as the second argument to indicate the distribution's mean. This gives us about 9%.



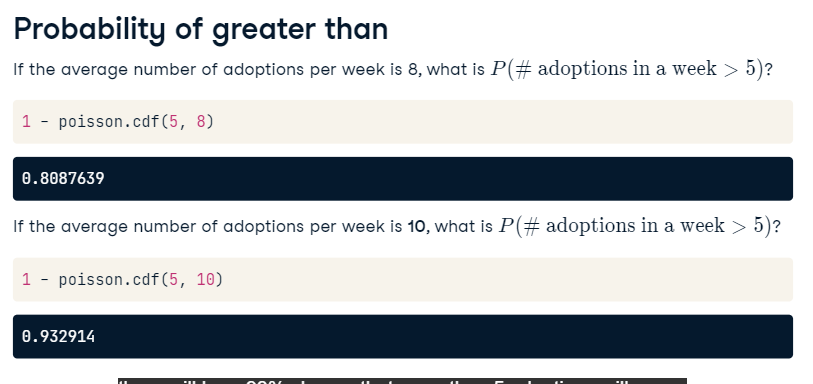
**Probability of less than or equal to**

To get the probability that 5 or fewer adoptions will happen in a week, use the poisson-dot-cdf function, passing in the same numbers. This gives us about 20%.



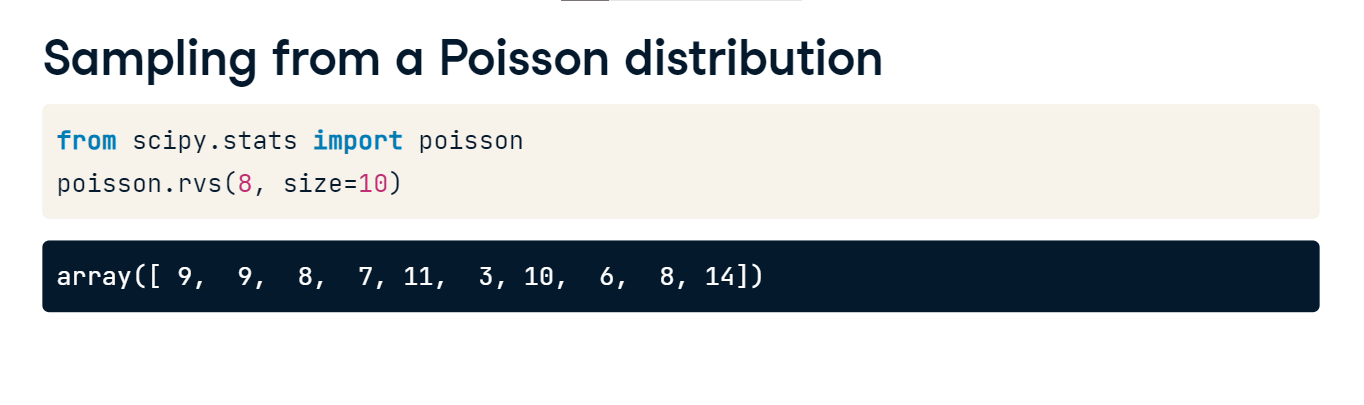
**Probability of greater than**

Just like other probability functions you've learned about so far, take 1 minus the "less than or equal to 5" probability to get the probability of more than 5 adoptions. There's an 81% chance that more than 5 adoptions will occur. If the average number of adoptions rises to 10 per week, there will be a 93% chance that more than 5 adoptions will occur.



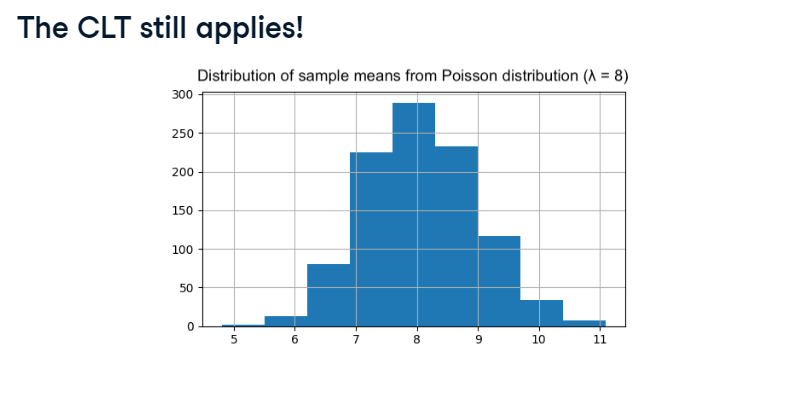
**Sampling from a Poisson distribution**

Just like other distributions, we can take samples from Poisson distributions using poisson-dot-rvs. Here, we'll simulate 10 different weeks at the animal shelter. In one week, there are 14 adoptions, but only 6 in another.



**The CLT still applies!**

Just like other distributions, the sampling distribution of sample means of a Poisson distribution looks normal with a large number of samples.



**More probability distributions**

In this lesson, we'll discuss a few other probability distributions.

**Exponential distribution**

The first distribution is the exponential distribution, which represents the probability of a certain time passing between Poisson events. We can use the exponential distribution to predict, for example, the probability of more than 1 day between adoptions, the probability of fewer than 10 minutes between restaurant arrivals, and the probability of 6-8 months passing between earthquakes. Just like the Poisson distribution, the time unit doesn't matter as long as it's consistent. The exponential distribution uses the same lambda value, which represents the rate, that the Poisson distribution does. Note that lambda and rate mean the same value in this context. It's also continuous, unlike the Poisson distribution, since it represents time.

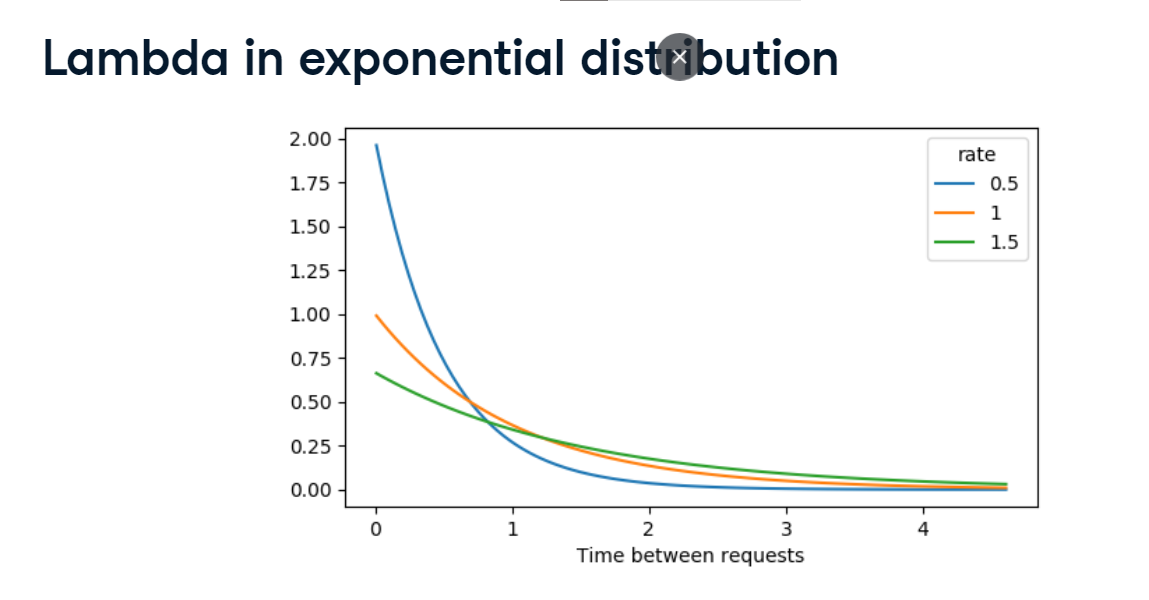
**Customer service requests**

For example, let's say that one customer service ticket is created every 2 minutes. We can rephrase this so it's in terms of a time interval of one minute, so half of a ticket is created each minute. We'll use 0-point-5 as the lambda value. The exponential distribution with a rate of one half looks like this.



**Lambda in exponential distribution**

The rate affects the shape of the distribution and how steeply it declines.



**Expected value of exponential distribution**

Recall that lambda is the expected value of the Poisson distribution, which measures frequency in terms of rate or number of events. In our customer service ticket example, this means that the expected number of requests per minute is point-5. The exponential distribution measures frequency in terms of time between events. The expected value of the exponential distribution can be calculated by taking 1 divided by lambda. In our example, the expected time between requests is 1 over one half, which is 2, so there is an average of 2 minutes between requests.

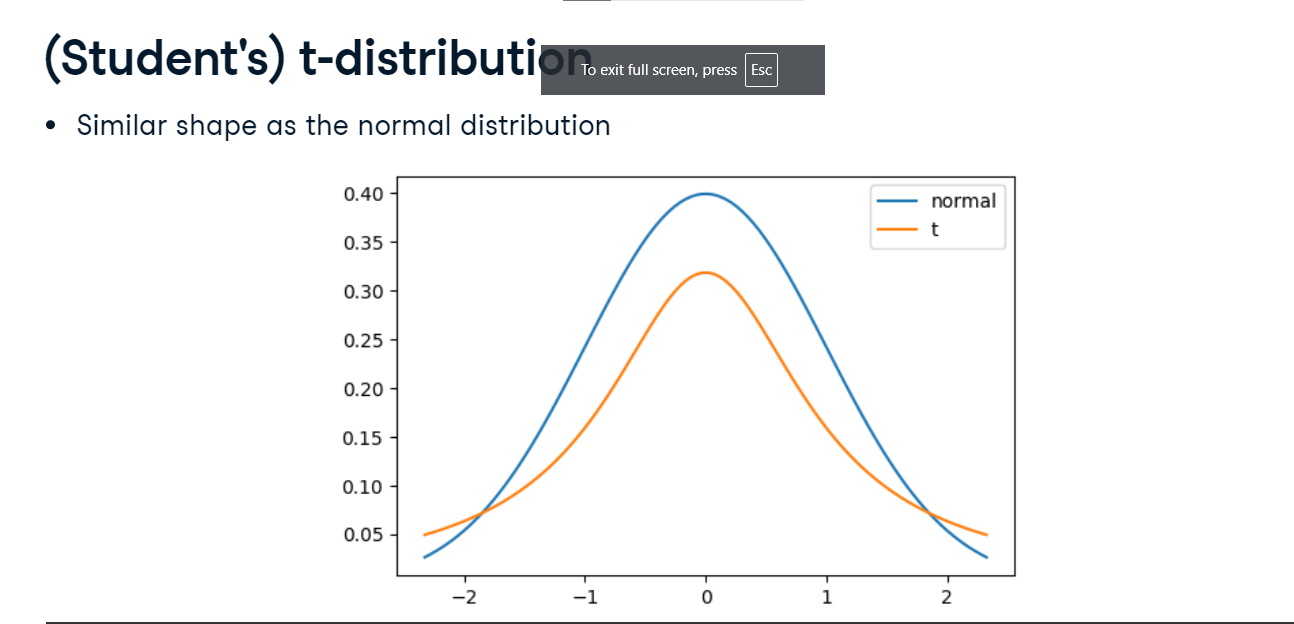
**How long until a new request is created?**

Similar to other continuous distributions, we can use expon-dot-cdf to calculate probabilities. The probability of waiting less than 1 minute for a new request is calculated using expon-cdf, passing in 1 followed by a 2, which gives us about an 40% chance. Note that we're passing in 2, not the lambda value which is point-5. The probability of waiting more than 4 minutes can be found using 1 minus expon-cdf of 4, 2, giving a 13% chance. Finally, the probability of waiting between 1 and 4 minutes can be found by taking expon-cdf of 4 and subtracting expon-cdf of 1. There's a 50% chance you'll wait between 1 and 4 minutes.



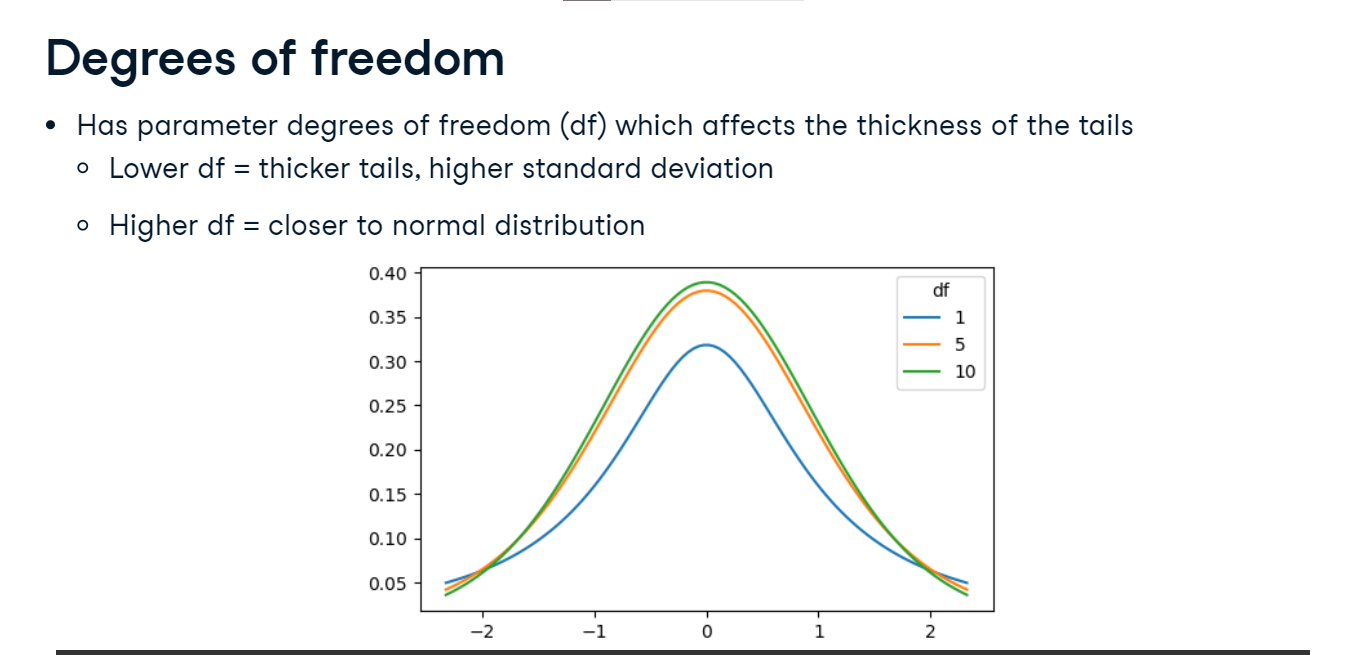
**(Student's) t-distribution**

The next distribution is the t-distribution, which is also sometimes called Student's t-distribution. Its shape is similar to the normal distribution, but not quite the same. If we compare the normal distribution, in blue, with the t-distribution with one degree of freedom, in orange, the t-distribution's tails are thicker. This means that in a t-distribution, observations are more likely to fall further from the mean.



**Degrees of freedom**

The t-distribution has a parameter called degrees of freedom, which affects the thickness of the distribution's tails. Lower degrees of freedom results in thicker tails and a higher standard deviation. As the number of degrees of freedom increases, the distribution looks more and more like the normal distribution.



**Log-normal distribution**

The last distribution we'll discuss is the log-normal distribution. Variables that follow a log-normal distribution have a logarithm that is normally distributed. This results in distributions that are skewed, unlike the normal distribution. There are lots of real-world examples that follow this distribution, such as the length of chess games, blood pressure in adults, and the number of hospitalizations in the 2003 SARS outbreak.

