Linear regression inherently involves at least two variables. Before investigating several variables at a time though, it is important to understand some things about the theory of a single random variable. These things form the subject of this chapter.

In the first section of the chapter, we examine two of the most commonly encountered distributions of continuous random variables in statistics, namely normal distributions and $t$ distributions. In the following section, we will use these two distributions to introduce statistical hypothesis testing. After that, we will take up the closely related topic of confidence intervals.

### 2.1 Normal distributions

The two distributions of interest to us at present are normal distributions and $t$ distributions. We begin our study of continuous distributions with normal distributions.

Normal distributions are by far the most important continuous distributions in probability and statistics. Because normal distributions are so commonly encountered in statistics, it is helpful to become very familiar
Chapter 2  Univariate Statistics

with their basic properties. This will help build statistical intuition and save large amounts of time that would otherwise have to be spent repeatedly consulting reference material on normal distributions.

Defining normal distributions

We begin our study of normal distributions by defining them. However, note that at this point, the formula for the probability density function is not important to know; rather, the vocabulary being introduced in the definition is what should receive particular attention.

**Definition 2.1.1** A normal distribution is a continuous distribution whose probability density function \( N_{pdf} \) is given by

\[
N_{pdf}(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}
\]

for some real number \( \mu \) and some positive real number \( \sigma \). The real number \( \mu \) is called the mean of the distribution, and the positive real number \( \sigma \) is called the standard deviation. The mean \( \mu \) and standard deviation \( \sigma \) are referred to as the parameters of a normal distribution.

To denote that the distribution of a random variable \( X \) is a normal distribution with mean \( \mu \) and standard deviation \( \sigma \), we write:

\[
X \sim N(\mu, \sigma).
\]

A random variable \( X \) is said to be normally distributed if the distribution of \( X \) is a normal distribution.

The standard normal distribution is defined to be \( N(0, 1) \), the normal distribution with mean 0 and standard deviation 1.

Before we explore normal distributions further, we should note that the parameter names mean and standard deviation are not accidental. Indeed, the random variable mean (which we usually refer to as the expected value) and random variable standard deviation of a normally distributed random variable are equal to the correspondingly named parameters. This is useful to keep in mind in trying to understand what these parameters tell us about the distribution, and it also brings up another term.

**Definition 2.1.2** The variance \( \sigma^2 \) of a normal distribution is defined to be the square of the standard deviation \( \sigma \) of the distribution.
Since the random variable standard deviation of a normally distributed random variable equals its standard deviation parameter, the random variable variance of a normally distributed random variable equals the variance that we have just defined.

We should note that the standard deviation and the variance of a normal distribution carry exactly the same information. As such, we choose one of them to use as the parameter, since using both as parameters would be redundant. In this text, we have chosen the standard deviation as the parameter, in keeping with most of the statistical computer packages. However, in the mathematical and theoretical statistics literature, the variance is usually taken to be the parameter instead. The implication of this discrepancy in terminology and notation is that wherever normal distributions are discussed, it is important to know which alternative is being used. To reiterate, in this text we will consider the standard deviation $\sigma$ to be the parameter, not the variance $\sigma^2$.

Basic properties of normal distributions

Now that we have the requisite notation and terminology, we can proceed in our investigation of normal distributions. The definition of a normal distribution is simple enough, but it begs a number of questions. Perhaps the most immediate of these is: what does the graph a normal distribution look like? In fact, that is a good first question to ask of any distribution. Happily, the technology available today makes it pretty easy to answer the question for most of the commonly encountered distributions. We highly recommend searching the internet for keywords such as "normal distribution" applet (with quotation marks as indicated) to find interactive demonstrations involving normal distributions. Interactive demonstrations are one of the best ways to develop a good intuition for the roles played by parameters in probability distributions. Since a book is necessarily static, however, we will merely give some pictures here to illustrate the basic properties of normal distributions.

Figure 2.1.3 depicts the graph of the normal distribution with mean $\mu$ and standard deviation $\sigma$. This picture shows a few of the fundamental properties of normal distributions:
Figure 2.1.3. The characteristic bell shape of a normal distribution. The mean $\mu$ is the $x$ coordinate of the top of the bell, and the standard deviation $\sigma$ is related to how wide the bell is.
Properties of Normal Distributions

1. The overall shape of a normal distribution is approximately that of a side view of a bell.

2. A normal distribution is symmetric about its mean $\mu$.

3. The high point of the graph of a normal distribution has $x$ coordinate equal to $\mu$.

4. The width of the bell shape of a normal distribution is related to the standard deviation $\sigma$.

To get an even better feeling for how the parameters $\mu$ and $\sigma$ relate to the overall shape of the normal distribution, an interactive demonstration on a computer is highly recommended. However, since this is a book, we illustrate the role the parameters play with some pictures instead (Figure 2.1.4).

The relationship between the standard deviation of a normal distribution and the width of its bell shape is shown in Figure 2.1.4, but it could still use further elaboration. How widely spread out a distribution is can be described in terms of what fraction of its area it lies close to its mean. For normal distributions, this is usually stated in terms of the fraction of the area that lies within 1, 2, and 3 standard deviations of the mean, as described in the following rule.
Figure 2.1.4. Changing the parameters of a normal distribution. Increasing the mean $\mu$ shifts the whole distribution to the right, and increasing the standard deviation $\sigma$ widens the distribution.
Figure 2.1.5. Graphical depiction of the 68-95-99.7 Rule. Since the total area of the distribution equals 1, multiplying each shaded area by 100 gives the percentage of the total area within each shaded region. This gives the rule its name.
Chapter 2  Univariate Statistics

The 68-95-99.7 Rule

As shown in Figure 2.1.5:

1. Approximately 68% of the area of a normal distribution lies within 1 standard deviation of the mean.

2. Approximately 95% of the area of a normal distribution lies within 2 standard deviations of the mean.

3. Approximately 99.7% of the area of a normal distribution lies within 3 standard deviations of the mean.

This rule is key in developing a good intuition for normal distributions and, as such, should be memorized. The rule describes concisely the relationship between the standard deviation $\sigma$ and the width of the bell shape of the normal distribution.

From the definition, it is immediate that there are infinitely many normal distributions, one for each pair of parameter values $\mu$ and $\sigma$. However, normal distributions are all really the same up to a change of units. This can be seen by choosing appropriate units for the random variable in question, a process which corresponds in the graph of the distribution to choosing an appropriate origin and scale for the $x$ axis. Instead of using the original units for the random variable in question (such as centimeters, kilograms, etc.), we use units that correspond to the number of standard deviations a value is from the mean of the distribution.

Definition 2.1.6 For any normally distributed random variable, standardized units are units that give the signed number of standard deviations that a value is from the mean.

By signed, we mean that values smaller than the mean are negative, while those larger than the mean are positive. The relationship between the original units and standardized units is shown in Figure 2.1.7, where the original units are shown on the top $x$ axis and standardized units on the bottom $x$-axis. Notice how much more relevant the standardized units are to the curve being depicted, compared with the original units.

The definition of standardized units implies that any normal distribution $N(\mu, \sigma)$ re-expressed in its standardized units is a standard normal
Figure 2.1.7. A normal distribution with mean 13 and standard deviation 5. The upper $x$ axis uses the original units of the random variable, while the lower $x$ axis is in standardized units.
distribution $N(0, 1)$. This follows since the mean is 0 standardized units from the mean, and the mean plus the standard deviation is 1 standardized unit from the mean.

The following formula shows how to convert a random variable $X$ expressed in arbitrary units to a random variable $Z$ expressed in standardized units.

### CONVERTING TO STANDARDIZED UNITS

For any normally distributed random variable $X$ with mean $\mu$ and standard deviation $\sigma$, the random variable

$$Z = \frac{X - \mu}{\sigma}.$$

is called the **standardized** version of $X$. This random variable $Z$ gives the values of $X$ converted to standardized units. For any value $x$ assumed by $X$, the corresponding value

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

assumed by $Z$ is called the **$z$-score** of $x$.

Although we do not give a proof of this formula for converting to standardized units here, we do offer some justification of it. First of all, the quantity $x - \mu$ in the numerator tells us how far (and in which direction, positive or negative) $x$ is from the mean $\mu$, still measured in the original units. Dividing by the standard deviation $\sigma$ then tells us how many standard deviations $x$ is from $\mu$. Notice that because the units for $\sigma$ are the original units of the random variable, those original units cancel out of the expression for $z$.

Standardized units really are the “right” units to use when working with normally distributed random variables. The raw distance from a value to the mean tells us almost nothing of interest. However, since we know the 68-95-99.7 Rule, the number of standard deviations a value is from the mean gives us an immediate idea of the probability of observing values closer to (or further from) the mean. This sort of probability estimate turns out to be very handy in statistics.
Computing areas under graphs of normal distributions

We now turn our attention to another aspect of working with normal distributions. Probabilities associated with normally distributed random variables correspond to areas under normal curves. Because of this, such areas frequently need to be computed. Inconveniently enough, this requires the use of a computer (or a calculator or a table) because there is no simple formula giving such areas, meaning that we need to use numerical approximations which are at best impractical to do by hand.

Rather than having a function built in to compute the area under a normal curve between any two values, almost all statistical computer packages (and calculators and tables) instead have only the cumulative distribution function of a normal distributions built in. Since we will usually be using standardized units, only the cumulative distribution function of the standard normal distribution, for which we now introduce some notation, is of interest to us.

**Definition 2.1.8** We denote the cumulative distribution function of a standard normally distributed random variable by $N_{cdf}(z)$.

The standard normal cumulative distribution function can be used to calculate all of the commonly needed areas under normal curves, as we now show. We focus on standard normal distributions now because we will typically use standardized units in this text.

There are four main types of areas that need to be computed when working with standard normally distributed random variables, as pictured in Figure 2.1.9:

1. The area to the left of a given value $z$, known as the **left tail of $z$ (or with endpoint $z$)**.

2. The area to the right of a given value $z$, known as the **right tail of $z$ (or with endpoint $z$)**.

3. The area between two given values $z_{\text{low}}$ and $z_{\text{high}}$ (where $z_{\text{low}} < z_{\text{high}}$), with no particular name.

4. The area to the left of a given value $-z$ together with the area to the right of $z$ (where $z > 0$), known as **symmetric tails of $z$ (or of $-z$)**.

We now show how to compute each of these in terms of the standard normal cumulative distribution function $N_{cdf}(z)$. 
Figure 2.1.9. The four main types of areas under normal curves to be computed. Upper left: A left tail. Upper right: A right tail. Lower left: The area between two values (which need not be negatives of each other, although they are here). Lower right: Symmetric tails.
As the upper left picture in Figure 2.1.9 indicates, the area of the left tail of a value \( z \) is given directly by the cumulative distribution function:

\[
\text{area of the left tail of } z = N_{\text{cdf}} (z).
\]

As the upper right picture in Figure 2.1.9 indicates, the area of a right tail of \( z \) is simply the area under the entire curve (which equals 1) minus the area of the left tail of \( z \):

\[
\text{area of the right tail of } z = 1 - N_{\text{cdf}} (z).
\]

As shown in the lower left picture in Figure 2.1.9, the area under the standard normal curve between \( z_{\text{low}} \) and \( z_{\text{high}} \) can be computed as the area of the left tail of \( z_{\text{high}} \) minus the area of the left tail of \( z_{\text{low}} \):

\[
\text{area between } z_{\text{low}} \text{ and } z_{\text{high}} = N_{\text{cdf}} (z_{\text{high}}) - N_{\text{cdf}} (z_{\text{low}}).
\]

In the lower right picture in Figure 2.1.9, the area of the symmetric tails of a value \( z \) can be expressed either as twice the area of the left tail of \( z \) or as twice the area of the right tail of \( z \):

\[
\text{area of symmetric tails of } z = \text{area to the left of } -z \text{ plus to the right of } z = N_{\text{cdf}} (-z) + (1 - N_{\text{cdf}} (z)) = 2N_{\text{cdf}} (-z) \text{ or } 2(1 - N_{\text{cdf}} (z)).
\]

In the later chapters of this text, we will use the appropriate method from those above without further explanation whenever we need to compute such an area, so it important to become comfortable with computing such areas at this point.

**Quantiles of normal distributions**

Now that we know how to compute the area of a left tail under a standard normal curve given the endpoint defining the region, we consider what is in some sense the reverse problem: given an area (between 0 and 1), find the endpoint of the left tail of a standard normal curve with that area. As we saw in the previous chapter, the solution to this problem is given by the quantile function.
Definition 2.1.10 The quantile function of a standard normally distributed random variable is denoted by \( N^{qf}(p) \).

As we know from the previous chapter, the standard normal cumulative distribution function \( N^{cdf}(z) \) and the standard normal quantile function \( N^{qf}(p) \) are inverses of each other in the sense that

\[
N^{cdf}(z) = p \quad \text{is the same as} \quad N^{qf}(p) = z.
\]

As you might imagine, computing values of the standard normal quantile function by hand is not feasible. However, almost all statistical computer packages have the standard normal quantile function built in.

Our main use for the standard normal quantile function in this text is not going to be to find endpoints of left tails, but rather to find endpoints of a slightly different kind of region which we call a central quantile interval.\(^1\)

Definition 2.1.11 For any real number \( c \) between 0 and 1, the central quantile interval of \( c \) is the interval symmetric about 0 with the property that the area under a standard normal curve above this interval equals \( c \). The endpoints of the level \( c \) central quantile interval are denoted by \(-z_c^*\) and \( z_c^*\), and the nonnegative endpoint \( z_c^*\) is called the central quantile value of \( c \).

Figure 2.1.12 illustrates the central quantile interval of \( c = 0.75 \), including its right endpoint \( z_{0.75}^* \), the central quantile value of 0.75.

Central quantile values are used commonly in statistical inference, so we should know how to compute them. Most statistical computer packages do not have built-in functions to compute central quantile values directly, since central quantile values can be computed readily in terms of the standard normal quantile function (which generally is built in). We now explain two ways to accomplish this.

First, the lower endpoint of the central quantile interval of \( c \) has the property that its left tail has area \((1 - c)/2\), as indicated in Figure 2.1.13. This means that the lower endpoint \(-z_c^*\) of the central quantile interval of \( c \) is given by:

\[
-z_c^* = N^{qf}((1 - c)/2).
\]

Taking the negative of both sides gives us that

\[
z_c^* = -N^{qf}((1 - c)/2),
\]

\(^1\)This is not a standard term, but we will find it convenient to use in this text.
Section 2.1 Normal distributions

Figure 2.1.12. How the central quantile value $z_c^*$ (for $c = 0.75$ in this case) is determined.

Figure 2.1.13. Computing a central quantile value using the quantile function.
which is the first way to compute $z^*_c$ in terms of the standard normal quantile function.

The second way to compute $z^*_c$ is to note that the left tail of the upper endpoint $z^*_c$ has area

$$\frac{1 - c}{2} + c = \frac{1 + c}{2}.$$  

This means that the upper endpoint $z^*_c$ (also known as the central quantile value of $c$) is given by:

$$z^*_c = N_{\text{qf}}((1 + c)/2),$$

which gives us the second way to compute $z^*_c$ in terms of the standard normal quantile function. To speed mental calculation, note that $(1 + c)/2$ is simply halfway between $c$ and 1.

In summary then, to compute a central quantile value using the quantile function, we use either of the following two formulas.

### Formulas for Computing Central Quantile Values

If $N_{\text{qf}}(p)$ is the quantile function for the standard normal distribution and $0 < c < 1$, then the central quantile value $z^*_c$ can be computed by either of the following:

$$z^*_c = -N_{\text{qf}}((1 - c)/2) = N_{\text{qf}}((1 + c)/2).$$

We now have sufficient familiarity with normal distributions for the present, so we proceed next to explore $t$ distributions.

## 2.2 $t$ distributions

Although they pale in comparison to normal distributions, $t$ distributions comprise the second most commonly encountered type of continuous distribution in statistics. Since $t$ distributions are similar in many ways to normal distributions, our introduction to $t$ distributions here will roughly parallel our introduction to normal distributions.
Defining \( t \) distributions

We begin with the definition of a \( t \) distribution. As with normal distributions, at this point, the formula for the probability density function is not important to know. Rather, some familiarity with the basic terminology and properties of \( t \) distributions are more important at present.

**Definition 2.2.1** A \( t \) distribution is a continuous distribution whose probability density function \( f(x) \) is given by

\[
f(x) = C \left(1 + \frac{x^2}{d}\right)^{-(d+1)/2}.
\]

for some positive real number \( d \), where the constant \( C \) depends only on \( d \) and is included only to ensure that the total area of the distribution equals 1.\(^2\) The positive real number \( d \) is called the (number of) degrees of freedom of the distribution. The number of degrees of freedom \( d \) is also referred to as the only parameter of a \( t \) distribution.

To denote that the distribution of a random variable \( X \) is a \( t \) distribution with \( d \) degrees of freedom, we write:

\[ X \sim t_d.\]

A random variable \( X \) is said to be \( t \) distributed if the distribution of \( X \) is a \( t \) distribution.

Note that \( t \) distributions have only one parameter, while normal distributions have two. There is no parameter in a \( t \) distribution to shift the center of the distribution, as there is with normal distributions (the mean). There is actually a generalization of a \( t \) distribution called a non-central \( t \) distribution that has a second parameter to shift the center, but it is not of much use to us in this text, so we will merely refer the interested reader to a more advanced text on mathematical statistics for a discussion of it.

Although the random variable mean and random variable standard deviation were natural measures of center and spread for normal distributions (and even coincided with the parameters of the distributions), they are not natural to use for \( t \) distributions and are, in fact, rarely discussed.\(^3\)

\(^2\)We omit the complicated formula for \( C \) here; suffice it to say that \( C \) is not a parameter but rather a positive real number depending only on \( d \).

\(^3\)Except perhaps to note that a \( t \) distributed random variable with 1 degree of freedom has no expected value, since the defining integral diverges.
Suffice it to say, the random variable median of any $t$ distribution is 0, as is the random variable mean when its defining integral converges (namely when $d > 1$). The random variable standard deviation can be computed but is not a particularly informative quantity for $t$ distributions.

**Basic properties of $t$ distributions**

As with normal distributions, the first question to address regarding $t$ distributions is how their graphs look. Again we would highly recommend searching the internet for keywords such as "$t$ distribution" applet (with quotation marks as indicated) to find interactive demonstrations of $t$ distributions. Since a book cannot be interactive the way that a website can, however, we simply give some pictures here to illustrate the basic properties of $t$ distributions.

Figure 2.2.2 depicts the graph of several $t$ distributions with different numbers of degrees of freedom. This picture shows some of the basic properties of $t$ distributions:
Properties of $t$ distributions

1. The overall shape of a $t$ distribution is very similar to that of a standard normal distribution. However, the region near 0 is shorter and the tails are thicker in $t$ distributions.

2. A $t$ distribution is symmetric about 0, which is its highest point.

3. In a $t$ distribution, the region near 0 is shorter (less probable) than in a standard normal distribution.

4. In a $t$ distribution, the tails are thicker (more probable) than in a standard normal distribution.

5. The higher the number of degrees of freedom, the closer a $t$ distribution is to a standard normal distribution.

It is important to note that the 68-95-99.7 rule applied only to normal distributions, and not to $t$ distributions (or any other distributions). However, because $t$ distributions have a shape similar to the standard normal distribution, the 68-95-99.7 rule can be used to give a ballpark intuition for probabilities related to $t$ distributions. For example, while it is not true that 95% of the area under a $t$ distribution lies between $-2$ and $2$ (the region within 2 standard deviations of the mean in a standard normal distribution), we should expect somewhere around that percentage, with the approximation being better for higher numbers of degrees of freedom. Checking this on the computer for a couple of values, we find that with $d = 1$, the approximation isn’t very good: the actual percentage of the total area between $-2$ and $2$ is about 70%. However, with $d = 20$, the actual percentage is about 94%, so a ballpark estimate based on the standard normal distribution would be pretty good.

Note that because $t$ distributions are all centered at 0 and because their random variable standard deviations are not particularly informative, there are no commonly used standardized units for $t$ distributions as there are for normal distributions.
Computing with $t$ distributions

We now turn our attention to another aspect of working with normal distributions. Probabilities associated with normally distributed random variables correspond to areas under normal curves. Because of this, such areas frequently need to be computed. Inconveniently enough, this requires the use of a computer (or a calculator or a table) because there is no simple formula giving such areas, meaning that we need to use numerical approximations which are at best impractical to do by hand.

Rather than having a function built in to compute the area under a normal curve between any two values, almost all statistical computer packages (and calculators and tables) instead have only the cumulative distribution function of a normal distributions built in. Since we will usually be using standardized units, only the cumulative distribution function of the standard normal distribution, for which we now introduce some notation, is of interest to us.

**Definition 2.2.3** We denote the cumulative distribution function of a standard normally distributed random variable by $t_{cdf}$ ($x$).

The $t$ cumulative distribution functions can be used to calculate all of the commonly needed areas under $t$ distributions, as we now show.

There are four main types of areas that need to be computed when working with $t$ distributed random variables, as pictured in Figure 2.2.4:

1. The area to the left of a given value $x$, known as the left tail of $x$ (or with endpoint $x$).

2. The area to the right of a given value $x$, known as the right tail of $x$ (or with endpoint $x$).

3. The area between two given values $x_{low}$ and $x_{high}$ (where $x_{low} < x_{high}$), with no particular name.

4. The area to the left of a given value $-x$ together with the area to the right of $x$ (where $x > 0$), known as symmetric tails of $x$ (or of $-x$).

We now show how to compute each of these in terms of the $t$ cumulative distribution functions $t_{cdf}$ ($x$).

As the upper left picture in Figure 2.1.9 indicates, the area of the left tail of a value $x$ is given directly by the cumulative distribution function:

$$\text{area of the left tail of } x = t_{cdf} (x).$$
Figure 2.2.4. The four main types of areas under $t$ curves to be computed. Upper left: A left tail. Upper right: A right tail. Lower left: The area between two values (which need not be negatives of each other, although they are here). Lower right: Symmetric tails.
As the upper right picture in Figure 2.1.9 indicates, the area of a right tail of \( x \) is simply the area under the entire curve (which equals 1) minus the area of the left tail of \( x \):

\[
\text{area of the right tail of } x = 1 - t_{d}^{\text{cdf}} (x).
\]

As shown in the lower left picture in Figure 2.1.9, the area under the \( t \) distribution with \( d \) degrees of freedom between \( x_{\text{low}} \) and \( x_{\text{high}} \) can be computed as the area of the left tail of \( x_{\text{high}} \) minus the area of the left tail of \( x_{\text{low}} \).

\[
\text{area between } x_{\text{low}} \text{ and } x_{\text{high}} = t_{d}^{\text{cdf}} (x_{\text{high}}) - t_{d}^{\text{cdf}} (x_{\text{low}}).
\]

In the lower right picture in Figure 2.1.9, the area of the symmetric tails of a value \( x \) can be expressed either as twice the area of the left tail of \( x \) or as twice the area of the right tail of \( x \):

\[
\text{area of symmetric tails of } x = \text{area to the left of } -x \text{ plus to the right of } x \\
= t_{d}^{\text{cdf}} (-x) + (1 - t_{d}^{\text{cdf}} (x)) \\
= 2t_{d}^{\text{cdf}} (-x) \text{ or } 2(1 - t_{d}^{\text{cdf}} (z)).
\]

In the later chapters of this text, we will use the appropriate method from those above without further explanation whenever we need to compute such an area, so it important to become comfortable with computing such areas at this point.

Quantiles of \( t \) distributions

Now that we know how to compute the area of a left tail under a \( t \) distribution given the endpoint defining the region, we consider what is in some sense the reverse problem: given an area (between 0 and 1), find the endpoint of the left tail of a particular \( t \) distribution with that area. As we saw with normal distributions, the solution to this problem is given by the quantile function.

Definition 2.2.5 The quantile function of a \( t \) distributed random variable with \( d \) degrees of freedom is denoted by \( t_{d}^{qf} (p) \).
The $t$ cumulative distribution function $t_{d}^{\text{cdf}}(x)$ and the $t$ quantile function $t_{d}^{\text{qf}}(p)$ are inverses of each other in the sense that

$$t_{d}^{\text{cdf}}(x) = p \quad \text{is the same as} \quad t_{d}^{\text{qf}}(p) = x.$$ 

As you might imagine, computing values of $t$ quantile functions by hand is not feasible. However, almost all statistical computer packages have the $t$ quantile function built in.

Our main use for the $t$ quantile function in this text is not going to be to find endpoints of left tails, but rather to find endpoints of a slightly different kind of region which we call a central quantile interval, as with standard normal distributions earlier.

**Definition 2.2.6** For any real number $c$ between 0 and 1, the $t$ central quantile interval of $c$ with $d$ degrees of freedom is the interval symmetric about 0 with the property that the area under a $t$ distribution with $d$ degrees of freedom above this interval equals $c$. The endpoints of the level $c$ central quantile interval are denoted by $-t_{d}^{*c}$ and $t_{d}^{*c}$, and the nonnegative endpoint $t_{d}^{*c}$ is called the central quantile value of $c$.

Exactly as with standard normal distributions, to compute a $t$ central quantile value using $t$ quantile functions, we use either of the following two formulas.

**Formulas for computing $t$ central quantile values**

If $t_{d}^{\text{qf}}(p)$ is the quantile function for the standard normal distribution and $0 < c < 1$, then the central quantile value $t_{d}^{*c}$ can be computed by either of the following:

$$t_{d}^{*c} = -t_{d}^{\text{qf}}((1 - c)/2) = t_{d}^{\text{qf}}((1 + c)/2).$$

**2.3 Hypothesis tests and confidence intervals**

We will cover this in class.
2.4 Problems

1. Suppose you have a test statistic that is standard normally distributed under the null hypothesis $H_0$, and that you compute the test statistic’s value to be $-2.43$ for your data.

   (a) Compute the 2-sided $P$-value of the data.
   (b) Compute the 1-sided to the right $P$-value of the data.
   (c) Compute the 1-sided to the left $P$-value of the data.
   (d) Interpret your result in terms of evidence about the null hypothesis, using the traditional significance level $\alpha = 0.05$.

2. Repeat Problem 1, only for a test statistic whose distribution under the null hypothesis $H_0$ is a $t$ distribution with 4 degrees of freedom. Continue to assume that the test statistic’s value is $-2.43$.

3. Suppose you have a test statistic that is standard normally distributed under the null hypothesis $H_0$, and that you compute the test statistic’s value to be $1.98$ for your data.

   (a) Compute the 2-sided $P$-value of the data.
   (b) Compute the 1-sided to the right $P$-value of the data.
   (c) Compute the 1-sided to the left $P$-value of the data.
   (d) Interpret your result in terms of evidence about the null hypothesis, using the traditional significance level $\alpha = 0.05$.

4. Repeat Problem 3, only for a test statistic whose distribution under the null hypothesis $H_0$ is a $t$ distribution with 11 degrees of freedom. Continue to assume that the test statistic’s value is $1.98$.

5. Suppose you have a test statistic $X$ that is normally distributed with unknown mean $\mu$ and standard deviation $\sigma = 0.74$, and that you have computed the value of $X$ to be $-2.43$ for your data.

   (a) Find a level 0.95 confidence interval for $\mu$.
   (b) Find a level 0.90 confidence interval for $\mu$. 
6. Repeat both parts of Problem 5, only now assume that \( \sigma \) is not known, but that the sample standard deviation is \( s = 0.74 \). The number of observations in the sample is 12 and the number of parameters used in describing \( \mu \) is 1, so the number of degrees of freedom for the problem equals \( 12 - 1 = 11 \). Continue to assume that the test statistic’s value is \(-2.43\).