

## 5. The t-test

In Chapter 3 a sampling distribution, the t-distribution, was introduced. In Chapter 4 you learned how to use the t-distribution to make an important inference, an interval estimate of the population mean. Here you will learn how to use that same t-distribution to make more inferences, this time in the form of hypothesis tests. Before we start to learn about those tests, a quick review of the t-distribution is in order.

### The t-distribution

The t-distribution is a sampling distribution. You could generate your own t-distribution with  $n-1$  degrees of freedom by starting with a normal population, choosing all possible samples of one size,  $n$ , computing a t-score for each sample:

$$t = \frac{\bar{x} - \mu}{\frac{s}{\sqrt{n}}}$$

where:  $\bar{x}$  = the sample mean

$\mu$  = the population mean

$s$  = the sample standard deviation

$n$  = the size of the sample.

When you have all of the samples' t-scores, form a relative frequency distribution and you will have your t-distribution. Luckily, you do not have to generate your own t-distributions because any statistics book has a table that shows the shape of the t-distribution for many different degrees of freedom. Exhibit 1 reproduces a portion of a typical t-table. See below.

## 5. The t-test



df	$p = .10$	$p = .05$	$p = .01$
	upper tail = .05	upper tail = 0.025	upper tail = .005
5	2.02	2.57	4.03
10	1.81	2.23	3.17
15	1.75	2.13	2.95
20	1.72	2.09	2.85
25	1.71	2.06	2.79
30	1.70	2.04	2.75
35	1.69	2.03	2.72
40	1.68	2.02	2.70
45	1.68	2.01	2.69
50	1.68	2.01	2.68
55	1.67	2.00	2.67
60	1.67	2.00	2.66
65	1.67	2.00	2.65
70	1.67	1.99	2.65
75	1.67	1.99	2.64
80	1.66	1.99	2.64
85	1.66	1.99	2.63
90	1.66	1.99	2.63
95	1.66	1.99	2.63
100	1.66	1.98	2.63
Infinite	1.65	1.96	2.58

Exhibit 7: A portion of a typical t-table

When you look at the formula for the t-score, you should be able to see that the mean t-score is zero because the mean of the  $\bar{x}$ 's is equal to  $\mu$ . Because most samples have  $\bar{x}$ 's that are close to  $\mu$ , most will have t-scores that are close to zero. The t-distribution is symmetric, because half of the samples will have  $\bar{x}$ 's greater than  $\mu$ , and half less. As you can see from the table, if there are 10 df, only .005 of the samples taken from a normal population will have a t-score greater than +3.17. Because the distribution is symmetric, .005 also have a t-score less than -3.17. Ninety-nine per cent of samples will have a t-score between  $\pm 3.17$ . Like the example in Exhibit 1, most t-tables have a picture showing what is in the body of the table. In Exhibit 1, the shaded area is in the right tail, the body of the table shows the t-score that leaves the  $\alpha$  in the right tail. This t-table also lists the two-tail  $\alpha$  above the one-tail where it has  $p = .xx$ . For 5 df, there is a .05 probability that a sample will have a t-score greater than 2.02, and a .10 probability that a sample will have a t score either  $> +2.02$  or  $< -2.02$ .

There are other sample statistics which follow this same shape and which can be used as the basis for different hypothesis tests. You will see the t-distribution used to test three different types of hypotheses in this chapter and that the t-distribution can be used to test other hypotheses in later chapters.

Though t-tables show how the sampling distribution of t-scores is shaped if the original population is normal, it turns out that the sampling distribution of t-scores is very close to the one in the table even if the original population is not quite normal, and most researchers do not worry too much about the normality of the original population. An even more important fact is that the sampling distribution of t-scores is very close to the one in the table even if the original population is not very close to being normal as long as the samples are large. This means that you can safely use the t-distribution to make inferences when you are not sure that the population is normal as long as you are sure that it is bell-shaped. You can also make inferences based on samples of about 30 or more using the t-distribution when you are not sure if the population is normal. Not only does the t-distribution describe

the shape of the distributions of a number of sample statistics, it does a good job of describing those shapes when the samples are drawn from a wide range of populations, normal or not.

### A simple test: does this sample come from a population with that mean?

Imagine that you have taken all of the samples with  $n=10$  from a population that you knew the mean of, found the t-distribution for 9 df by computing a t-score for each sample and generated a relative frequency distribution of the t's. When you were finished, someone brought you another sample ( $n=10$ ) wondering if that new sample came from the original population. You could use your sampling distribution of t's to test if the new sample comes from the original population or not. To conduct the test, first hypothesize that the new sample comes from the original population. With this hypothesis, you have hypothesized a value for  $\mu$ , the mean of the original population, to use to compute a t-score for the new sample. If the t for the new sample is close to zero—if the t-score for the new sample could easily have come from the middle of the t-distribution you generated—your hypothesis that the new sample comes from a population with the hypothesized mean seems reasonable and you can conclude that the data supports the new sample coming from the original population. If the t-score from the new sample was far above or far below zero, your hypothesis that this new sample comes from the original population seems unlikely to be true, for few samples from the original population would have t-scores far from zero. In that case, conclude that the data gives support to the idea that the new sample comes from some other population.

This is the basic method of using this t-test. Hypothesize the mean of the population you think a sample might come from. Using that mean, compute the t-score for the sample. If the t-score is close to zero, conclude that your hypothesis was probably correct and that you know the mean of the population from which the sample came. If the t-score is far from zero, conclude that your hypothesis is incorrect, and the sample comes from a population with a different mean.

Once you understand the basics, the details can be filled in. The details of conducting a "hypothesis test of the population mean", testing to see if a sample comes from a population with a certain mean—are of two types. The first type concerns how to do all of this in the formal language of statisticians. The second type of detail is how to decide what range of t-scores implies that the new sample comes from the original population.

You should remember from the last chapter that the formal language of hypothesis testing always requires two hypotheses. The first hypothesis is called the "null hypothesis", usually denoted  $H_o$  :. It states that there is no difference between the mean of the population from which the sample is drawn and the hypothesized mean. The second is the "alternative hypothesis", denoted  $H_1$  : or  $H_a$  :. It states that the mean of the population from which the sample comes is different from the hypothesized value. If your question is simply "does this sample come from a population with **this** mean?", your  $H_a$  : is simply  $\mu \neq \text{the hypothesized value}$  . If your question is "does this sample come from a population with a mean **greater than** some value", then your  $H_a$  : becomes  $\mu > \text{the hypothesized value}$  .

The other detail is deciding how "close to zero" the sample t-score has to be before you conclude that the null hypothesis is probably correct. How close to zero the sample t-score must be before you conclude that the data supports  $H_o$  : depends on the df and how big a chance you want to take that you will make a mistake. If you decide to conclude that the sample comes from a population with the hypothesized mean only if the sample t is very, very close to zero, there are many samples actually from the population that will have t-scores that would lead

## 5. The t-test

you to believe they come from a population with some other mean—it would be easy to make a mistake and conclude that these samples come from another population. On the other hand, if you decide to accept the null hypothesis even if the sample t-score is quite far from zero, you will seldom make the mistake of concluding that a sample from the original population is from some other population, but you will often make another mistake—concluding that samples from other populations are from the original population. There are no hard rules for deciding how much of which sort of chance to take. Since there is a trade-off between the chance of making the two different mistakes, the proper amount of risk to take will depend on the relative costs of the two mistakes. Though there is no firm basis for doing so, many researchers use a 5 per cent chance of the first sort of mistake as a default. The level of chance of making the first error is usually called "alpha" ( $\alpha$ ) and the value of alpha chosen is usually written as a decimal fraction—taking a 5 per cent chance of making the first mistake would be stated as " $\alpha = .05$ ". When in doubt, use  $\alpha = .05$ .

If your alternative hypothesis is "not equal to", you will conclude that the data supports  $H_a$ : if your sample t-score is either well below or well above zero and you need to divide  $\alpha$  between the two tails of the t-distribution. If you want to use  $\alpha = .05$ , you will support  $H_a$ : if the t is in either the lowest .025 or the highest .025 of the distribution. If your alternative is "greater than", you will conclude that the data supports  $H_a$ : only if the sample t-score is well above zero. So, put all of your  $\alpha$  in the right tail. Similarly, if your alternative is "less than", put the whole  $\alpha$  in the left tail.

The table itself can be confusing even after you know how many degrees of freedom you have and if you want to split your  $\alpha$  between the two tails or not. Adding to the confusion, not all t-tables look exactly the same. Look at the typical t-table above and notice that it has three parts: column headings of decimal fractions, row headings of whole numbers, and a body of numbers generally with values between 1 and 3. The column headings are labeled p or "area in the right tail," and sometimes are labeled " $\alpha$ ." The row headings are labeled "df," but are sometimes labeled "v" or "degrees of freedom". The body is usually left unlabeled and it shows the t-score which goes with the " $\alpha$ " and "degrees of freedom" of that column and row. These tables are set up to be used for a number of different statistical tests, so they are presented in a way that is a compromise between ease of use in a particular situation and the ability to use the same table for a wide variety of tests. My favorite t tables are available online at <http://www.itl.nist.gov/div898/handbook/eda/section3/eda3672.htm>

In order to use the table to test to see if "this sample comes from a population with a certain mean" choose  $\alpha$  and find the number of degrees of freedom. The number of degrees of freedom in a test involving one sample mean is simply the size of the sample minus one ( $df = n-1$ ). The  $\alpha$  you choose may not be the  $\alpha$  in the column heading. The column headings show the "right tail areas"—the chance you'll get a t-score *larger* than the one in the body of the table. Assume that you had a sample with ten members and chose  $\alpha = .05$ . There are nine degrees of freedom, so go across the 9 df row to the .025 column since this is a two-tail test, and find the t-score of 2.262. This means that in any sampling distribution of t-scores, with samples of ten drawn from a normal population, only 2.5 per cent (.025) of the samples would have t-scores greater than 2.262—any t-score greater than 2.262 probably occurs because the sample is from some other population with a larger mean. Because the t-distributions are symmetrical, it is also true that only 2.5 per cent of the samples of ten drawn from a normal population will have t-scores less than -2.262. Putting the two together, 5 per cent of the t-scores will have an absolute value greater the 2.262. So if you choose

$\alpha = .05$ , you will probably be using a t-score in the .025 column. The picture that is at the top of most t-tables shows what is going on. Look at it when in doubt.

LaTonya Williams is the plant manager for Eileen's Dental Care Company (EDC) which makes dental floss. EDC has a good, stable work force of semi-skilled workers who work packaging floss, paid by piece-work, and the company wants to make sure that these workers are paid more than the local average wage. A recent report by the local Chamber of Commerce shows an average wage for "machine operators" of USD 8.71. LaTonya needs to decide if a raise is needed to keep her workers above the average. She takes a sample of workers, pulls their work reports, finds what each one earned last week and divides their earnings by the hours they worked to find average hourly earnings.

That data appears below:

Smith 9.01

Wilson 8.67

Peterson 8.90

Jones 8.45

Gordon 8.88

McCoy 9.13

Bland 8.77

LaTonya wants to test to see if the mean of the average hourly earnings of her workers is greater than USD 8.71. She wants to use a one-tail test because her question is "greater than" not "unequal to". Her hypotheses are:

$$H_0: \mu \leq 8.71 \quad \text{and} \quad H_a: \mu > 8.71$$

As is usual in this kind of situation, LaTonya is hoping that the data supports  $H_a$ ; but she wants to be confident that it does before she decides her workers are earning above average wages. Remember that she will compute a t-score for her sample using USD 8.71 for  $\mu$ . If her t-score is negative or close to zero, she will conclude that the data supports  $H_0$ . Only if her t-score is large and positive will she go with  $H_a$ . She decides to use

$\alpha = .025$  because she is unwilling to take much risk of saying the workers earn above average wages when they really do not. Because her sample has  $n=7$ , she has 6 df. Looking at the table, she sees that the data will support  $H_a$ ; the workers earn more than average, only if the sample t-score is greater than 2.447.

Finding the sample mean and standard deviation,  $\bar{x} = \$8.83$  and  $s = .225$ , LaTonya computes her sample t-score:

$$t = \frac{\bar{x} - \mu}{\frac{s}{\sqrt{n}}} = \frac{8.83 - 8.71}{\frac{.225}{\sqrt{7}}}$$

$$t = \frac{.12}{.085} = 1.41$$

Because her sample t is not greater than +2.447, LaTonya concludes that she will have to raise the piece rates EDC pays in order to be really sure that mean hourly earnings are above the local average wage.

## 5. The t-test

If LaTonya had simply wanted to know if EDC's workers earned the same as other workers in the area, she would have used a two-tail test. In that case her hypotheses would have been:

$$H_o: \mu = 8.71 \quad \text{and} \quad H_a: \mu \neq 8.71$$

Using  $\alpha = .10$ , LaTonya would split the .10 between the two tails since the data supports  $H_a$  : if the sample t-score is either large and negative or large and positive. Her arithmetic is the same, her sample t-score is still 1.41, but she now will decide that the data supports  $H_a$  : only if it outside  $\pm 1.943$ .

### An alternative to choosing an $\alpha$

Many researchers now report how unusual the sample t-score would be if the null hypothesis was true rather than choosing an  $\alpha$  and stating whether the sample t-score implies the data supports one or the other of the hypotheses based on that  $\alpha$ . When a researcher does this, he is essentially letting the reader of his report decide how much risk to take of making which kind of mistake. There are even two ways to do this. If you look at the portion of the t-table reproduced above, you will see that it is not set up very well for this purpose; if you wanted to be able to find out what part of a t-distribution was above any t-score, you would need a table that listed many more t-scores. Since the t-distribution varies as the df changes, you would really need a whole series of t-tables, one for each df.

The old-fashioned way of making the reader decide how much of which risk to take is to not state an  $\alpha$  in the body of your report, but only give the sample t-score in the main text. To give the reader some guidance, you look at the usual t-table and find the smallest  $\alpha$ , say it is .01, that has a t-value less than the one you computed for the sample. Then write a footnote saying "the data supports the alternative hypothesis for any  $\alpha > .01$ ".

The more modern way uses the capability of a computer to store lots of data. Many statistical software packages store a set of detailed t-tables, and when a t-score is computed, the package has the computer look up exactly what proportion of samples would have t-scores larger than the one for your sample. Exhibit 2 shows the computer output for LaTonya's problem from a typical statistical package. Notice that the program gets the same t-score that LaTonya did, it just goes to more decimal places. Also notice that it shows something called the "P value". The P value is the proportion of t-scores that are larger than the one just computed. Looking at the example, the computed t statistic is 1.41188 and the P value is 0.1038. This means that if there are 6 df, a little over 10 per cent of samples will have a t-score greater than 1.41188. Remember that LaTonya used an  $\alpha = .025$  and decided that the data supported  $H_o$  ; the P value of .1038 means that  $H_o$  : would be supported for any  $\alpha$  less than .1038. Since LaTonya had used  $\alpha = .025$ , this p value means she does not find support for  $H_o$ .

<b>Hypothesis test: Mean</b>
Null Hypothesis: Mean = 8.71
Alternative: greater than
Computed t statistic = 1.41188
P value = 0.1038

Exhibit 8: Output from typical statistical software for LaTonya's problem

The P-value approach is becoming the preferred way to The P-value presents research results to audiences of professional researchers. Most of the statistical research conducted for a business firm will be used directly for decision making or presented to an audience of executives to aid them in making a decision. These audiences will generally not be interested in deciding for themselves which hypothesis the data supports. When you are making a presentation of results to your boss, you will want to simply state which hypothesis the evidence supports. You may decide by using either the traditional  $\alpha$  approach or the more modern P-value approach, but deciding what the evidence says is probably your job.

### Another t-test: do these two (independent) samples come from populations with the same mean?

One of the other statistics that has a sampling distribution that follows the t-distribution is the difference between two sample means. If samples of one size ( $n_1$ ) are taken from one normal population and samples of another size ( $n_2$ ) are taken from another normal population (and the populations have the same standard deviation), then a statistic based on the difference between the sample means and the difference between the population means is distributed like t with  $n_1 + n_2 - 2$  degrees of freedom. These samples are independent because the members in one sample do not affect which members are in the other sample. You can choose the samples independently of each other, and the two samples do not need to be the same size. The t- statistic is:

$$t = \frac{(\bar{x}_1 - \bar{x}_2) - (\mu_1 - \mu_2)}{\sqrt{\frac{s^2}{n_1} + \frac{s^2}{n_2}}}$$

where:  $\bar{x}_i$  = the mean of sample i

$\mu_i$  = the mean of population i

$s^2$  = the pooled variance

$n_i$  = the size of sample i.

The usual case is to test to see if the samples come from populations with the same mean, the case where  $(\mu_1 - \mu_2) = 0$ . The pooled variance is simply a weighted average of the two sample variances, with the weights based on the sample sizes. This means that you will have to calculate the pooled variance before you calculate the t-score. The formula for pooled variance is:

$$s^2 = \frac{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2}{n_1 + n_2 - 2}$$

To use the pooled variance t-score, it is necessary to assume that the two populations have equal variances. If you are wondering about why statisticians make a strong assumption in order to use such a complicated formula, it is because the formula that does not need the assumption of equal variances is even more complicated, and reduces the degrees of freedom in the final statistic. In any case, unless you have small samples, the amount of arithmetic needed means that you will probably want to use a statistical software package for this test. You should also note

## 5. The t-test

that you can test to see if two samples come from populations that are any hypothesized distance apart by setting  $(\mu_1 - \mu_2)$  equal to that distance.

An article in *U. S. News and World Report* (Nov. 1993) lamenting grade inflation in colleges states that economics grades have not been inflated as much as most other grades. Nora Alston chairs the Economics Department at Oaks College, and the dean has sent her a copy of the article with a note attached saying "Is this true here at Oaks? Let me know." Dr Alston is not sure if the Dean would be happier if economics grades were higher or lower than other grades, but the article claims that economics grades are lower. Her first stop is the Registrar's office.

She has the clerk in that office pick a sample of 10 class grade reports from across the college spread over the past three semesters. She also has the clerk pick out a sample of 10 reports for economics classes. She ends up with a total of 38 grades for economics classes and 51 grades for other classes. Her hypotheses are:

$$H_o : \mu_{econ} - \mu_{other} \geq 0$$

$$H_a : \mu_{econ} - \mu_{other} < 0$$

She decides to use  $\alpha = .05$ .

This is a lot of data, and Dr Alston knows she will want to use the computer to help. She initially thought she would use a spreadsheet to find the sample means and variances, but after thinking a minute, she decided to use a statistical software package. The one she is most familiar with is one called SAS. She loads SAS onto her computer, enters the data, and gives the proper SAS commands. The computer gives her the output in Exhibit 3.

The SAS System						
TTFST Procedure						
Variable: GRADE						
Dept	N	Mean	Dev	Std Error	Minimum	Maximum
Econ	38	2.28947	1.01096	0.16400	0	4.00000
Variance	T	DF	Prob>[T]			
Unequal	-2.3858	85.1	0.0193			
Equal	-2.3345	87.0	0.0219			
For HO: Variances are equal, F=1.35 DF=[58.37] Prob>F=0.3485						

Exhibit 9: SAS system software output for Dr Alston's grade study

Dr Alston has 87 df, and has decided to use a one-tailed, left tail test with  $\alpha = .05$ . She goes to her t-table and finds that 87 df does not appear, the table skipping from 60 to 120 df. There are two things she could do. She could try to interpolate the t-score that leaves .05 in the tail with 87 df, or she could choose between the t-value for 60 and 120 in a conservative manner. Using the conservative choice is the best initial approach, and looking at her table she sees that for 60 df .05 of t-scores are less than -1.671, and for 120 df, .05 are less than -1.658. She does not want to conclude that the data supports economics grades being lower unless her sample t-score is far from zero, so she



decides that she will accept  $H_a$  : if her sample  $t$  is to the left of  $-1.671$ . If her sample  $t$  happens to be between  $-1.658$  and  $-1.671$ , she will have to interpolate.

Looking at the SAS output, Dr Alston sees that her  $t$ -score for the equal variances formula is  $-2.3858$ , which is well below  $-1.671$ . She concludes that she will tell the dean that economics grades are lower than grades elsewhere at Oaks College.

Notice that SAS also provides the  $t$ -score and  $df$  for the case where equal variances are not assumed in the "unequal" line. SAS also provides a  $P$  value, but it is for a two-tail test because it gives the probability that a  $t$  with a larger absolute value,  $>|T|$ , occurs. Be careful when using the  $p$  values from software: notice if they are one-tail or two-tail  $p$ -values before you make your report!

### A third t-test: do these (paired) samples come from the sample population?

Managers are often interested in "before and after" questions. As a manager or researcher you will often want to look at "longitudinal" studies, studies that ask about what has happened to an individual as a result of some treatment or across time. Are they different after than they were before? For example, if your firm has conducted a training program you will want to know if the workers who participated became more productive. If the work area has been re-arranged, do workers produce more than before? Though you can use the difference of means test developed earlier, this is a different situation. Earlier, you had two samples that were chosen independently of each other; you might have a sample of workers who received the training and a sample of workers who had not. The situation for this test is different; now you have a sample of workers and for each worker you have measured their productivity before the training or re-arrangement of the work space and you have measured their productivity after. For each worker you have a pair of measures, before and after. Another way to look at this is that for each member of the sample you have a difference between before and after.

You can test to see if these differences equal zero, or any other value, because a statistic based on these differences follows the  $t$ -distribution for  $n-1$   $df$  when you have  $n$  matched pairs. That statistic is:

$$t = \frac{\bar{D} - \delta}{\frac{s_D}{\sqrt{n}}}$$

where:  $\bar{D}$  = the mean of the differences in the pairs in the sample

$\delta$  = the mean of the differences in the pairs in the population

$s_D$  = the standard deviation of the differences in the sample

$n$  = the number of pairs in the sample.

It is a good idea to take a minute and figure out this formula. There are paired samples and the differences in those pairs, the  $D$ 's, are actually a population. The mean of those  $D$ 's is  $\delta$ . Any sample of pairs will also yield a sample of  $D$ 's. If those  $D$ 's are normally distributed, then the  $t$ -statistic in the formula above will follow the  $t$ -distribution. If you think of the  $D$ 's as the same as  $x$ 's in the  $t$ -formula at the beginning of the chapter, and think of  $\delta$  as the population mean, you should realize that this formula is really just that basic  $t$  formula.

Lew Podolsky is division manager for Dairyland Lighting, a manufacturer of outdoor lights for parking lots, barnyards, and playing fields. Dairyland Lighting organizes its production work by teams. The size of the team

## 5. The t-test

varies somewhat with the product being assembled, but there are usually three to six in a team, and a team usually stays together for a few weeks assembling the same product. Dairyland Lighting has a branch plant in the US state of Arizona that serves their west coast customers and Lew has noticed that productivity seems to be lower in Arizona during the summer, a problem that does not occur at the main plant in the US city of Green Bay, Wisconsin. After visiting the Arizona plant in July, August, and November, and talking with the workers during each visit, Lew suspects that the un-air conditioned plant just gets too hot for good productivity. Unfortunately, it is difficult to directly compare plant-wide productivity at different times of the year because there is quite a bit of variation in the number of employees and product mix across the year. Lew decides to see if the same workers working on the same products are more productive on cool days than hot days by asking the local manager, Dave Mueller, to find a cool day and a hot day from last fall and choose ten work teams who were assembling the same products on the two days. Dave sends Lew the following data:

Team leader	Output—cool day	Output—hot day	Difference (cool-hot)
	October 14	October 20	
Martinez	153	149	4
McAlan	167	170	-3
Wilson	164	155	9
Burningtree	183	179	4
Sanchez	177	167	10
Lilly	162	150	12
Cantu	165	158	7

Exhibit 10: Lew Podolsky's data for the air-conditioning decision

Lew decides that if the data support productivity being higher of cool days, he will call in a heating/air-conditioning contractor to get some cost estimates so that he can decide if installing air conditioning in the Arizona plant is cost-effective. Notice that he has matched pairs data--for each team he has production on October 14, a cool day, and on October 20, a hot day. His hypotheses are:

$$H_0: \delta \leq 0 \text{ and } H_a: \delta > 0$$

Using  $\alpha = .05$  in this one-tail test, Lew will decide to call the engineer if his sample t-score is greater than 1.943, since there are 6 df. This sample is small, so it is just as easy to do the computations on a calculator. Lew finds:

$$\bar{D} = 6.1428$$

$$s_D = 5.0142$$

and his sample t-score is:

$$t = \frac{\bar{D} - \delta}{\frac{s_D}{\sqrt{n}}} = \frac{6.14 - 0}{\frac{5.01}{\sqrt{7}}}$$

$$t = \frac{6.14}{1.89} = 3.24$$

Because his sample t-score is greater than 1.943, Lew gets out the telephone book and looks under air conditioning contractors to call for some estimates.

### Summary

The t-tests are commonly used hypothesis tests. Researchers often find themselves in situations where they need to test to see if a sample comes from a certain population, and therefore test to see if the sample probably came from a population with that certain mean. Even more often, researchers will find themselves with two samples and want to know if the samples come from the same population, and will test to see if the samples probably come from populations with the same mean. Researchers also frequently find themselves asking if two sets of paired samples have equal means. In any case, the basic strategy is the same as for any hypothesis test. First, translate the question into null and alternative hypotheses, making sure that the null hypothesis includes an equal sign. Second, choose  $\alpha$ . Third, compute the relevant statistics, here the t-score, from the sample or samples. Fourth, using the tables, decide if the sample statistic leads you to conclude that the sample came from a population where the null hypothesis is true or a population where the alternative is true.

The t-distribution is also used in testing hypotheses in other situations since there are other sampling distributions with the same t-distribution shape. So, remember how to use the t-tables for later chapters.

Statisticians have also found how to test to see if three or more samples come from populations with the same mean. That technique is known as "one-way analysis of variance". The approach used in analysis of variance is quite different from that used in the t-test. It will be covered in chapter, "The F-test and One-Way ANOVA".