

Unit 21

Student's t Distribution in Hypotheses Testing

Objectives:

- To understand the difference between the standard normal distribution and the Student's t distributions
- To understand the difference between a hypothesis test about a population proportion λ and a hypothesis test about a population mean μ

Our introduction to hypothesis testing has focused solely on hypothesis tests about a population proportion λ . The z test statistic we use in the hypothesis test $H_0: \lambda = \lambda_0$ vs. $H_1: \lambda \neq \lambda_0$ is

$$z = \frac{\bar{p} - \lambda_0}{\sqrt{\frac{\lambda_0(1 - \lambda_0)}{n}}}$$

Let us now investigate whether we can use the same sort of test statistic in a hypothesis test about a population mean μ .

As an illustration, suppose the manager at a certain manufacturing plant would like to see if there is any evidence that the mean amount of cereal per box from an assembly line is different from 12 oz., which is the advertised amount. He chooses a 0.05 significance level to perform a hypothesis test. The following amounts (in ounces) are recorded for a simple random sample of 16 boxes of cereal selected from the assembly line:

11.660 11.550 12.045 11.616 11.924 12.320 11.616 11.913
12.078 11.836 11.847 11.946 12.232 11.968 11.781 11.748

You can verify that the sample mean is $\bar{x} = 11.880$ oz., and that the sample standard deviation is $s = 0.220$ oz. We shall now attempt to perform the four steps of a hypothesis test.

The first step is to state the null and alternative hypotheses, and to choose a significance level. Remember that the null hypothesis is what we assume to be true unless there is sufficient evidence against it, and the alternative hypothesis is the statement we are looking for evidence to support; also, recall that a null hypothesis is often a statement involving equality, while the alternative hypothesis is a statement involving inequality. In this hypothesis test, we shall assume that the mean amount of cereal per box is 12 oz. unless we find sufficient evidence otherwise; in other words, 12 oz. is our hypothesized value for μ . We complete the first step of the hypothesis test by writing following:

$$H_0: \mu = 12 \quad \text{vs.} \quad H_1: \mu \neq 12 \quad (\alpha = 0.05).$$

The second step is to collect data and calculate the value of the test statistic. Recall that our test statistic in a hypothesis test about λ is the z -score of \bar{p} calculated under the assumption that the hypothesized value of λ is correct. To use the same sort of test statistic in a hypothesis test about a population mean μ , we would calculate the z -score of \bar{x} under the assumption that the hypothesized value of μ was correct. Letting μ_0 represent our hypothesized value for μ , this z -score for \bar{x} is

$$z = \frac{\bar{x} - \mu_x}{\sigma_x} = \frac{\bar{x} - \mu_0}{\frac{\sigma}{\sqrt{n}}}.$$

Calculating this z -score for \bar{x} presents a problem because we do not have the value of σ . We know the value of \bar{x} (since we calculated it from our sample), we know the value of μ_0 (since this represents our hypothesized

mean), and we know the value of n (since this is of course the sample size of our data); however, we do not know the population standard deviation σ , and consequently we are not able to obtain the value of $\sigma_{\bar{x}}$. When calculating the z -score of \bar{p} , we do not have this problem, since the value of $\sigma_{\bar{p}}$ depends only on λ and n .

If it were possible for us to have access to the entire population in order to obtain the population standard deviation σ , then we could also obtain the population mean μ , and there would be no reason for us to be doing a hypothesis test! A very reasonable way to estimate the population standard deviation σ is to use the sample standard deviation s whose value we can easily obtain from our sample. It is then tempting to simply replace σ by s in the z -score which would be our test statistic, and in doing so the denominator of our test statistic would have σ/\sqrt{n} in place of s/\sqrt{n} .

For small sample sizes n , simply using s/\sqrt{n} in place of $\sigma_{\bar{x}} = \sigma/\sqrt{n}$ in the z -score for \bar{x} and still treating this statistic as a z -score may not give accurate results. This was demonstrated in a paper published in 1908 by W. S. Gosset. Gosset showed that instead of associating this test statistic with the standard normal distribution, the test statistic should be associated with the *Student's t distribution with $n - 1$ degrees of freedom*. (In order to prevent the competitors of Gosset's employers from realizing that they could benefit from more accurate statistical methods, Gosset's paper was published anonymously under the name of "Student," because Gosset was studying at the University of London at the time of his discovery.)

In the hypothesis test $H_0: \mu = \mu_0$ vs. $H_1: \mu \neq \mu_0$, our test statistic shall be

$$t_{n-1} = \frac{\bar{x} - \mu_0}{s/\sqrt{n}},$$

which we refer to as a t statistic with $n - 1$ degrees of freedom. Consequently, in order to perform hypothesis tests about μ , we must first study the t distributions. Recall that Table A.2 provides us with information about the standard normal distribution. We shall now be concerned with Table A.3 which provides us with information about the t distributions.

The figures displaying density curves at the top of Tables A.2 and A.3 indicate that some similarities exist between the standard normal distribution and the t distributions; specifically, these distributions are symmetric and bell-shaped. The difference between the standard normal distribution and the t distributions is that the t distributions have more variation and a flatter bell-shape than the standard normal distribution. The figures at the top of Tables A.2 and A.3 look alike only because they are not drawn to scale. If a standard normal density curve and a Student's t density curve were to be graphed on the same scale, the Student's t density curve would look flatter and more "spread out."

It is important to keep in mind that Table A.2 supplies information about the one and only standard normal density curve, but Table A.3 contains information about several different t distributions. Each different t distribution is identified by something we call *degrees of freedom*, which we shall abbreviate as *df*. We refer to our test statistic in a test about a population mean μ as t_{n-1} where $n - 1$ is the corresponding *df*. It is difficult to explain in detail the concept of *df* in a general context without venturing beyond the mathematical scope of this text.

While we could just think of degrees of freedom (*df*) simply as a way of distinguishing between the different density curves for the various t distributions, there is an intuitive rule of thumb possible to explain degrees of freedom. In any situation where a statistical procedure is applied to data to obtain some information about population parameters, we can identify the number of parameters being estimated and the size of the samples(s) in the data. The rule of thumb that often applies is that the relevant degrees of freedom is equal to the total size of the samples(s) minus the number of parameters being estimated. Suppose a random sample of size n is to be used in a hypothesis test about a population mean μ , then the rule of thumb leads us to believe that the degrees of freedom should be equal to the total sample size n minus the number of parameters being estimated which is one (i.e., the one and only parameter is μ). This justifies the $n - 1$ degrees of freedom associated with the test statistic t_{n-1} defined previously.

Each row of Table A.3 is labeled by a value for *df*. The information contained in each row is exactly the same type of information contained at the bottom of Table A.2. At the bottom of Table A.2, you find values

such as $z_{0.05}$, $z_{0.025}$, and $z_{0.005}$, where the amount of area above the z -score is indicated in the subscript. Each row of Table A.3 contains t -scores, where the column heading indicates the amount of area above each t -score under the corresponding t density curve, etc. The notation we shall use to represent the t -scores in each row of Table A.3 will be similar to the notation we use to represent the z -scores at the bottom of Table A.2. For instance, just as we use $z_{0.01}$ to represent the z -score above which lies 0.01 of the area under the standard normal density curve, we shall use $t_{5;0.01}$ to represent the t -score above which lies 0.01 of the area under the density curve for the Student's t distribution with $df = 5$. When designating t -scores in Table A.3, we include in the subscript the degrees of freedom followed by the area above the t -score, separated by a semicolon.

Pick any column of Table A.3, and look at how the t -scores change from the row for $df = 1$ down through the successive rows. You should notice that as df increase, the corresponding t -scores in two successive rows become closer in value. By the time $df = 30$, the corresponding t -scores in two successive rows are so much alike that Table A.3 jumps from $df = 30$ to $df = 40$ to $df = 60$ to $df = 120$. The t distributions for $df > 120$ are almost identical to the standard normal distribution. Consequently, the t -scores in the last row of Table A.3, labeled $df = \infty$ (where ∞ represents infinity), are exactly equal to the values at the bottom of Table A.2: $z_{0.10}$, $z_{0.05}$, etc.

It should not be surprising that the t distributions become more like a standard normal distribution as df become larger. In the formula for the test statistic t_{n-1} , we are using s to estimate σ , and as the sample size n increases, s tends to become a more accurate estimate of σ , as we should expect to happen. In fact, with sample sizes $n > 120$, s is such a good estimator of σ , that it is reasonable to treat the value of s as if it were the actual value of σ . In other words, for sample sizes $n > 120$, we may treat the test statistic t_{n-1} as if it were the test statistic z . This is why the t -scores in the last row of Table A.3 ($df = \infty$) are exactly equal to the values at the bottom of Table A.2.

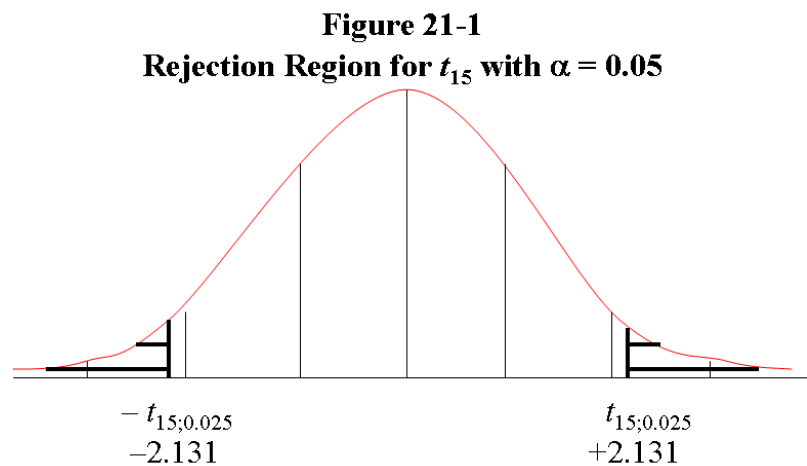
Now let us return to the second step of our hypothesis test concerning the mean amount of cereal per box (i.e., $H_0: \mu = 12$ vs. $H_1: \mu \neq 12$). Since $n = 16$, then $df = n - 1 = 16 - 1 = 15$. Earlier, we found that $\bar{x} = 11.880$ oz. and $s = 0.220$ oz. in the simple random sample of $n = 16$ cereal boxes; since 12 oz. is our hypothesized mean, we can calculate the value of our test statistic t_{15} as follows:

$$t_{15} = \frac{\bar{x} - \mu_0}{\frac{s}{\sqrt{n}}} = \frac{11.880 - 12}{\frac{0.220}{\sqrt{16}}} = -2.182 \quad .$$

The third step of our hypothesis test is to define the rejection region, decide whether or not to reject the null hypothesis, and obtain the p -value of the test. Figure 21-1 displays the rejection region graphically. The shaded area corresponds to values of the test statistic t_{15} which are unlikely to occur if $H_0: \mu = 12$ is true. In order that the total shaded area be equal to $\alpha = 0.05$, the shaded area corresponding to positive values of t_{15} is 0.025, and the shaded area corresponding to negative values of t_{15} is 0.025. We see then that the rejection region is defined by the t -score above which 0.025 of the area lies and the t -score below which 0.025 of the area lies, when $df = 15$. The symmetry of the t distributions of course implies these two t -scores are negatives of each other.

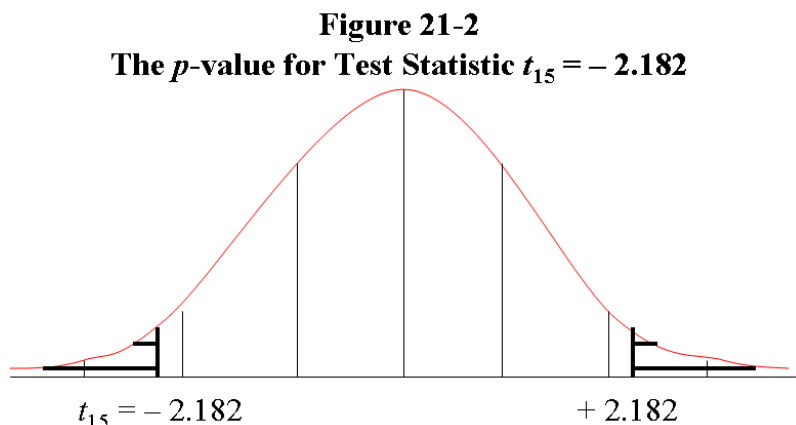
From the row corresponding to $df = 15$ in Table A.3, we find that the t -score above which 0.025 of the area lies is $t_{15;0.025} = 2.131$. We can then define our rejection region algebraically as

$$t_{15} \geq +2.131 \quad \text{or} \quad t_{15} \leq -2.131 \quad (\text{i.e., } |t_{15}| \geq 2.131) .$$



Since our test statistic, calculated in the second step, was found to be $t_{15} = -2.182$ which is in the rejection region, our decision is to reject $H_0: \mu = 12$; in other words, our data provides sufficient evidence against $H_0: \mu = 12$.

Figure 21-2 graphically illustrates the p -value. The p -value of this hypothesis test is the probability of obtaining a test statistic value t_{15} which represents a larger distance away from the hypothesized mean than the observed $t_{15} = -2.182$. In Figure 21-2, the observed test statistic value $t_{15} = -2.182$ has been located on the horizontal axis along with the value $+2.182$. The two shaded areas together correspond to test statistic values which represent a larger distance away from the hypothesized mean than the observed $t_{15} = -2.182$; in other words, the total shaded area in Figure 21-2 is the p -value of the hypothesis test.



To obtain the shaded area exactly from a table, we would need a table of areas under a Student's t density curve as detailed as Table A.2 for the standard normal distribution. However, we can use Table A.3 to obtain upper and lower bounds for the p -value, and such bounds are satisfactory in practice. In the row of Table A.3 corresponding to $df = 15$, we find that the observed test statistic $t_{15} = -2.182$ is between $t_{15;0.025} = 2.131$ and $t_{15;0.01} = 2.602$. This tells us that the area above the t -score $+2.182$ is between 0.01 and 0.025, and the area below the t -score -2.182 is between 0.01 and 0.025. Therefore, the total shaded area in Figure 21-2, which is the p -value, must be between $0.02 = 0.01 + 0.01$ and $0.05 = 0.025 + 0.025$. We indicate this by writing $0.02 < p\text{-value} < 0.05$.

Finding that $0.02 < p\text{-value} < 0.05$ confirms to us that H_0 is rejected with $\alpha = 0.05$. However, it also provides the reader of these results with some additional information. Remember that the choice of $\alpha = 0.05$ is made by the person(s) performing the hypothesis test. Those who read the results of a hypothesis test may wish to know whether or not the conclusion would have been the same with a different significance level. The p -value provides this information. For instance, suppose someone looking at the results of this present hypothesis test wants to know whether or not the H_0 would have been rejected with $\alpha = 0.10$ or with $\alpha = 0.01$. Knowing that $0.02 < p\text{-value} < 0.05$ tells us that H_0 would be rejected with $\alpha = 0.10$ but not with $\alpha = 0.01$. In other words, the evidence against H_0 is sufficiently strong with $\alpha = 0.10$ or $\alpha = 0.05$ but not with $\alpha = 0.01$.

To complete the fourth step of the hypothesis test, we can summarize the results of the hypothesis test as follows:

Since $t_{15} = -2.182$ and $t_{15;0.025} = 2.131$, we have sufficient evidence to reject H_0 . We conclude that the mean amount of cereal per box is different from 12 oz. ($0.02 < p\text{-value} < 0.05$). The data suggest that the mean amount of cereal per box is less than 12 oz.

Since we rejected the H_0 , thus concluding that the population mean was not equal to the hypothesized 12 oz., we are left with the question of whether the mean is less than or greater than the hypothesized 12 oz. We can answer this question by noting that the observed test statistic $t_{15} = -2.182$ is in the half of the rejection region corresponding to negative values of t_{15} , because the sample mean $\bar{x} = 11.880$ oz. was less than the hypothesized 12 oz. It is for this reason that we included as part of our statement of results a sentence indicating that the mean amount of cereal per box appears to be less than 12 oz. Consequently, an adjustment appears to be necessary in the assembly line from which the boxes of cereal come.

It is often useful to include a graphical display with the results of a hypothesis test. Figure 21-3 is a stem-and-leaf display for the amounts of cereal recorded for the sample of 16 boxes used in the hypothesis test just completed. From this stem-and-leaf display, we can easily find the five-number summary: 11.550, 11.704, 11.880, 12.0065, 12.320. This five-number summary was used to construct the box plot displayed as Figure 21-4.

While Tables A.2 and A.3 provide us with detailed information about the standard normal distribution and the t distributions, this information is also often available from computer software such as spreadsheets and statistical packages, and also from some calculators such as programmable calculators. However, when it is necessary to consult Table A.3, the desired df is not always available. For instance, $t_{45;0.05}$ and $t_{50;0.05}$ are not directly available from Table A.3, but this really presents no problem. The df in the table closest to $df = 45$ is $df = 40$; since $t_{40;0.05} = 1.684$ and $t_{60;0.05} = 1.671$ are close in value, we may simply choose to say $t_{45;0.05} \approx t_{40;0.05} = 1.684$. Since $df = 50$ is exactly in between $df = 40$ and $df = 60$, we can approximate $t_{50;0.05}$ as the average of $t_{40;0.05}$ and $t_{60;0.05}$; that is, we can say $t_{50;0.05} \approx (t_{40;0.05} + t_{60;0.05}) / 2 = (1.684 + 1.671) / 2 = 1.6775$.

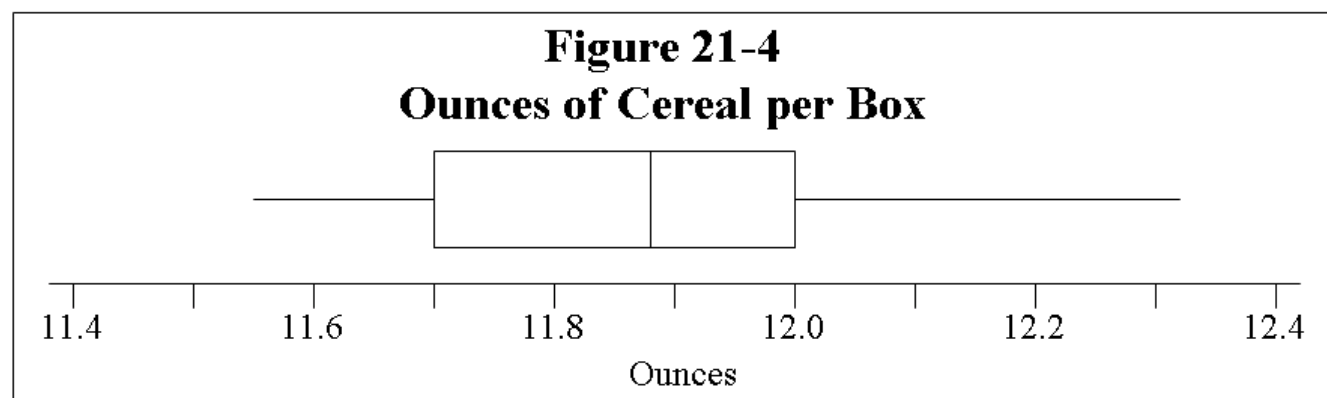
The use of the test statistic t_{n-1} in a hypothesis test about μ is based on the assumption that the data is a simple random sample selected from a population having a normal distribution. When the population from which the simple random sample is selected does not have a normal distribution, the test statistic t_{n-1} will still be appropriate if the sample size n is sufficiently large. We have previously seen how we can easily verify that the use of the test statistic z is appropriate in a test about λ , but it is not as easy to verify that the use of the test statistic t_{n-1} is appropriate in a test about μ .

One of the best ways to decide whether or not the population from which a simple random sample is selected has a normal distribution is to use the *Shapiro-Wilk test*. We will not discuss the Shapiro-Wilk test in detail here, since it is somewhat complex and generally requires the use of a statistical software

package on a computer. We shall instead just say that the Shapiro-Wilk test is based on what is called a *normal probability plot*. A normal probability plot is a scatter plot constructed by labeling one axis with the observations from a simple random sample and labeling the other axis with the z -scores corresponding to the percentile ranks of the observations in the sample. If the points lie sufficiently close to a straight line, then we conclude that the simple random sample came from a normal distribution, in which case we can feel confident that the use of the test statistic t_{n-1} is appropriate; otherwise, we conclude that the sample was selected from a non-normal distribution. Generally, it is not easy to construct a normal probability plot without access to computer software or a graphing calculator.

Figure 21-3
Ounces of Cereal per Box

11.5	50
11.6	60, 16, 16
11.7	81, 48
11.8	36, 47
11.9	24, 13, 46, 68
12.0	45, 78
12.1	
12.2	32
12.3	20



If one cannot easily perform the Shapiro-Wilk test or does not have easy access to a normal probability plot, a box plot or histogram can perhaps give one some information about whether or not there appears to be any skewness or outliers. Of course, examination of a normal probability plot or the results of the Shapiro-Wilk test are always preferable. If we should find that the population from which our sample comes does not have a normal distribution, the t_{n-1} statistic may still be appropriate with a sufficiently large sample size n . In cases where we do not think the sample size is sufficiently large, or where we think the non-normality of the

distribution is too great, there are alternative test statistics available in place of the t_{n-1} statistic. These alternative test statistics are often available with statistical software and on certain calculators, and we will not take the time to discuss these in detail here.

The points in a normal probability plot of the data displayed in Figure 21-3 will all lie relatively close to a straight line, which you can verify if you were to construct this normal probability plot. This suggests that the use of the test statistic t_{n-1} seems appropriate. Without access to this normal probability plot, one could look at the box plot displayed as Figure 21-4. Since there are no outliers and there appears to be only a slight skewness in the data, one might think that it is not unreasonable to assume that the sample size $n = 16$ is sufficiently large for the test statistic t_{15} to be appropriate.

We have now discussed the z test about a population proportion λ and the t_{n-1} test about a population mean μ . One assumption on which each of these two hypothesis tests is based is that a simple random sample has been selected. Back when we discussed the topic of sampling, we observed how a true simple random sample may often not be available in practice. However, there are situations in which we may reasonably treat a sample which is not truly a simple random sample as if it were a simple random sample. When conducting any hypothesis test, it is important to assess how the sampling procedure used may affect the results of the test.

As we study many different types of hypothesis tests, we shall find that even though the purpose behind various hypothesis tests may differ, the general procedure can always be summarized in terms of the four steps we have been using. The primary difference is the test statistic which is used and the distribution on which the rejection region is based.

Self-Test Problem 21-1. For many years, the mean gas mileage on a long trip was known to be 26.5 miles per gallon for a certain type of automobile. When a newly designed engine is incorporated into the automobile, a 0.05 significance level is chosen for a hypothesis test to see if there is any evidence that the mean gas mileage with the new design is different from 26.5 miles per gallon. Each automobile in a simple random sample of those with the new design is driven 100 miles and the following gas mileages are recorded:

26.2 26.5 27.2 27.3 27.1 27.6 25.9 27.5 26.8.

- Complete the four steps of the hypothesis test by completing the table titled *Hypothesis Test for Self-Test Problem 21-1*.
- Why would a box plot be an appropriate graphical display for the data used in this hypothesis test?
- Construct the box plot, and comment on whether there is any reason to think the t statistic is not appropriate.
- Considering the results of the hypothesis test, decide which of the Type I or Type II errors is possible, and describe this error.
- Decide whether H_0 would have been rejected or would not have been rejected with each of the following significance levels: (i) $\alpha = 0.01$, (ii) $\alpha = 0.10$.

Self-Test Problem 21-2. A manufacturer of light bulbs claims that the mean lifetime of the bulbs is 500 hours. A consumer group chooses a 0.10 significance level for a hypothesis test to see if there is any evidence that the mean lifetime is different from 500 hours. In a simple random sample of 40 light bulbs, the mean lifetime is 483.75 hours and the standard deviation is 26.35 hours.

- Complete the four steps of the hypothesis test by completing the table titled *Hypothesis Test for Self-Test Problem 21-2*.
- If the data were available, decide which of the following would be best as a graphical display for the data and say why: (i) bar or pie chart, (ii) scatter plot, (iii) box plot.
- Considering the results of the hypothesis test, decide which of the Type I or Type II errors is possible, and describe this error.
- Decide whether H_0 would have been rejected or would not have been rejected with each of the following significance levels: (i) $\alpha = 0.01$, (ii) $\alpha = 0.05$.
- What would the presence of one or more outliers in the data suggest about using the t statistic?

Self-Test Problem 21-3. Figure 21-5 is the same stem-and-leaf display as Figure 5-1. The data are from Self-Test Problem 5-2 where the breaking strength in pounds is recorded for several pieces of a type of rope. Find the five-number summary for the data, determine whether or not there are any potential outliers, and give a reason why the test statistic t_{n-1} might not be considered appropriate for a hypothesis test. (Optional: Construct a normal probability plot.)

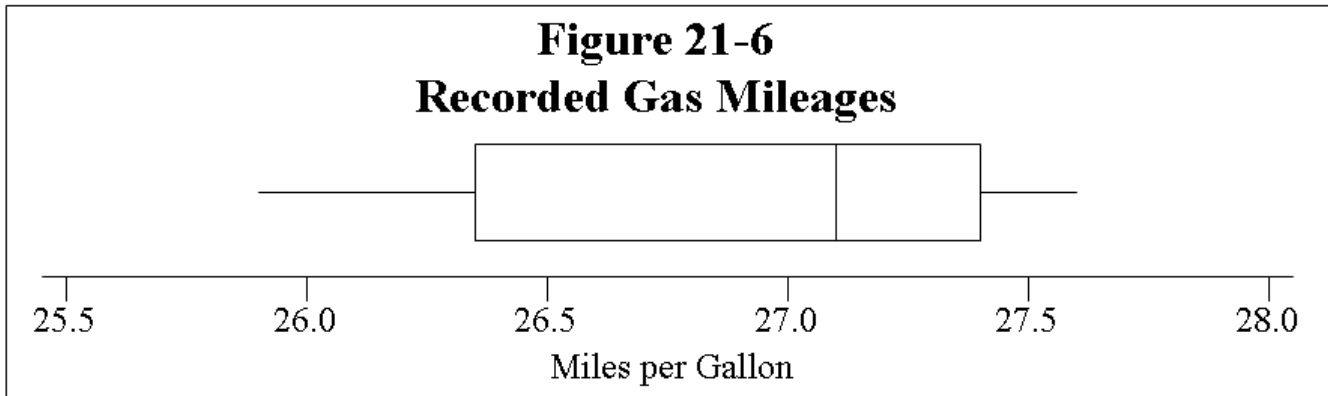
Figure 21-5
Stem-and-Leaf Display for
Rope Breaking Strengths

6		0 2
7		5
8		4
9		3 6 8
10		2 2 5 6 8
11		0 3 4 7
12		4 8

Answers to Self-Test Problems

- 21-1** (a) Step 1: $H_0: \mu = 26.5$, $H_1: \mu \neq 26.5$, $\alpha = 0.05$
Step 2: $n = 9$, $\bar{x} = 26.9$ mpg, $s = 0.5916$ mpg, and $t_8 = + 2.028$
Step 3: The rejection is $t_8 \geq + 2.306$ or $t_8 \leq - 2.306$ (which can be written as $|t_8| \geq 2.306$).
 H_0 is not rejected; $0.05 < p\text{-value} < 0.10$.
Step 4: Since $t_8 = + 2.028$ and $t_{8;0.025} = 2.306$, we do not have sufficient evidence to reject H_0 .
We conclude that the mean gas mileage with the new design is not different from 26.5 miles per gallon ($0.05 < p\text{-value} < 0.10$).
- (b) Since gas mileage is a quantitative variable, a box plot is an appropriate graphical display.
(c) The five-number summary is 25.9, 26.35, 27.1, 27.4, 27.6, and Figure 21-6 is the box plot. Since there are no potential outliers, and the box plot does not look extremely skewed, there is no reason to think that the t statistic is not appropriate.
(d) Since H_0 is not rejected, a Type II error is possible, which is concluding that $\mu = 26.5$ when really $\mu \neq 26.5$.
(e) H_0 would not have been rejected with $\alpha = 0.01$ but would have been rejected with $\alpha = 0.10$.
- 21-2** (a) Step 1: $H_0: \mu = 500$, $H_1: \mu \neq 500$, $\alpha = 0.10$
Step 2: $n = 40$, $\bar{x} = 483.75$ hours, $s = 26.35$ hours, and $t_{39} = - 3.900$
Step 3: The rejection is $t_{39} \geq + 1.684$ or $t_{39} \leq - 1.684$ (which can be written as $|t_{39}| \geq 1.684$).
 H_0 is rejected; $p\text{-value} < 0.001$.
Step 4: Since $t_{39} = - 3.900$ and $t_{39;0.05} = 1.684$, we have sufficient evidence to reject H_0 . We conclude that the mean lifetime is different from 500 hours ($p\text{-value} < 0.001$). The data suggest that the mean lifetime is less than 500 hours.
- (b) Since light bulb lifetime is a quantitative variable, a box plot is an appropriate graphical display.
(c) Since H_0 is rejected, a Type I error is possible, which is concluding that $\mu \neq 26.5$ when really $\mu = 26.5$.
(d) H_0 would have been rejected with both $\alpha = 0.01$ and $\alpha = 0.05$.
(e) The presence of one or more outliers would suggest that the t statistic may not be appropriate.
- 21-3** The five-number summary is 60, 93, 103.5, 113, 128; $IQR = 113 - 93 = 20$ and each of 60 and 62 is a potential outlier. Since there are potential outliers (and the points on a normal probability plot do not appear to lie close to a straight line), the test statistic t_{n-1} may not be appropriate.

Figure 21-6
Recorded Gas Mileages



Summary

The test statistic in a hypothesis test about λ is the z -score of \bar{p} calculated under the assumption that the hypothesized value of λ is correct, that is, the test statistic is

$$z = \frac{\bar{p} - \lambda_0}{\sigma_{\bar{p}}} = \frac{\bar{p} - \lambda_0}{\sqrt{\frac{\lambda_0(1 - \lambda_0)}{n}}},$$

where λ_0 represents the hypothesized value for λ . In order to use the same sort of test statistic in a hypothesis test about a population mean μ , we must know the population standard deviation σ ; when σ is not known, which is generally the case, the test statistic in a hypothesis test a population mean μ is

$$t_{n-1} = \frac{\bar{x} - \mu_0}{s / \sqrt{n}},$$

where μ_0 represents the hypothesized value for μ . We refer to this test statistic as a t statistic with $df = n - 1$, where df is *degrees of freedom*. The rule of thumb that often applies is that the relevant degrees of freedom is equal to the total size of the samples(s) minus the number of parameters being estimated.

In hypothesis test about the mean, the test statistic t_{n-1} must be compared to the *Student's t distribution with $n - 1$ degrees of freedom*. Some similarities exist between the standard normal distribution and the t distributions in that these distributions are all symmetric and bell-shaped; the differences are that the t distributions have more variation and a flatter bell-shape than the standard normal distribution. If a standard normal density curve and a Student's t density curve were to be graphed on the same scale, the Student's t density curve would look flatter and more "spread out."

In the formula for the test statistic t_{n-1} , σ is estimated by s . As the sample size n increases, s tends to become a more accurate estimate of σ . With sample sizes $n > 120$, s is such a good estimator of σ , that it is reasonable to treat the value of s as if it were the actual value of σ . Consequently, we can treat a Student's t distribution with $df > 120$ as if it were a standard normal distribution.

The steps in a hypothesis test about a population proportion λ and a hypothesis test about a population mean μ are the same except for the different test statistics that are used. In virtually all hypothesis tests, the same basic steps are performed with differing test statistics and differing purposes behind various hypothesis tests.

One of the best ways to decide whether or not the population from which a simple random sample is selected has a normal distribution is to use the *Shapiro-Wilk test* which is based on what is called a *normal probability plot*. A normal probability plot is a scatter plot constructed by labeling one axis with the observations from a simple random sample and labeling the other axis with the z -scores corresponding to the percentile ranks of the observations from the sample. If the points do not lie sufficiently close to a straight line,

then Shapiro-Wilk test will conclude that the sample was selected from a non-normal distribution. One might look at a box plot or histogram of the sample to see if there is any indication of skewness or if there are any outliers, but the Shapiro-Wilk test and/or examination of a normal probability plot is preferable.

If we should find that the population from which our sample comes does not have a normal distribution, the t_{n-1} statistic may still be appropriate with a sufficiently large sample size n . In cases where we do not think the sample size is sufficiently large, or where we think the non-normality of the distribution is too great, there are alternative test statistics available in place of the t_{n-1} statistic.

The z test about a population proportion λ and the t test are both based on the assumption that a simple random sample has been selected. While a true simple random sample may often not be available in practice, there are situations in which we may reasonably treat a sample which is not a simple random sample as if it were a simple random sample. When conducting any hypothesis test, it is important to assess how the sampling procedure used may affect the results of the test.