Solution

Look down the column headed "0.025" to the row corresponding to 12 degrees of freedom. The value for t_{12} is 2.179.

Example
5.18

Find the value for the t_{14} distribution whose lower-tail probability is 0.01.

Solution

Look down the column headed "0.01" to the row corresponding to 14 degrees of freedom. The value for t_{14} is 2.624. This value cuts off an area, or probability, of 1% in the upper tail. The value whose lower-tail probability is 1% is -2.624.

Don't Use the Student's *t* Statistic If the Sample Contains Outliers

For the Student's *t* statistic to be valid, the sample must come from a population that is approximately normal. Such samples rarely contain outliers. Therefore, methods involving the Student's *t* statistic should *not* be used for samples that contain outliers.

Confidence Intervals Using the Student's *t* Distribution

When the sample size is small, and the population is approximately normal, we can use the Student's *t* distribution to compute confidence intervals. We illustrate this with an example.

A metallurgist is studying a new welding process. He manufactures five welded joints and measures the yield strength of each. The five values (in ksi) are 56.3, 65.4, 58.7, 70.1, and 63.9. Assume that these values are a random sample from an approximately normal population. The task is to find a confidence interval for the mean strength of welds made by this process.

When the sample size is large, we don't need to worry much about the nature of the population, because the Central Limit Theorem guarantees that the quantity $\overline{X}$ will be approximately normally distributed. When the sample is small, however, the distribution of the population must be approximately normal.

The confidence interval in this situation is constructed much like the ones in Section 5.1, except that the *z*-score is replaced with a value from the Student's *t* distribution. The quantity

$$\frac{\overline{X} - \mu}{s/\sqrt{n}}$$

has a Student's *t* distribution with $n - 1$ degrees of freedom. Figure 5.12 (page 348) shows the t_4 distribution. From the Student's *t* table, we find that 95% of the area under the curve is contained between the values $t = -2.776$ and $t = 2.776$. It follows that for 95% of all the samples that might have been chosen,

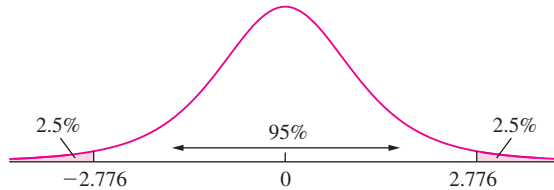$$-2.776 < \frac{\overline{X} - \mu}{s/\sqrt{n}} < 2.776$$

FIGURE 5.12 The Student's t distribution with four degrees of freedom. Ninety-five percent of the area falls between $t = -2.776$ and $t = 2.776$.

Put another way, for 95% of all the samples that might have been chosen, it is the case that

$$-2.776\frac{s}{\sqrt{n}} < \overline{X} - \mu < 2.776\frac{s}{\sqrt{n}}$$

Multiplying by -1 and adding $\overline{X}$ across the inequality, we obtain a 95% confidence interval for μ:

$$\overline{X} - 2.776\frac{s}{\sqrt{n}} < \mu < \overline{X} + 2.776\frac{s}{\sqrt{n}}$$

In this example, the sample mean is $\overline{X} = 62.88$, and the sample standard deviation $s = 5.4838$. The sample size is $n = 5$. Substituting values for $\overline{X}$, s, and n, we find that a 95% confidence interval for μ is $62.88 - 6.81 < \mu < 62.88 + 6.81$, or $(56.07,\ 69.69)$.

In general, to produce a level $100(1 - \alpha)\%$ confidence interval, let $t_{n-1,\alpha/2}$ be the $1 - \alpha/2$ quantile of the Student's t distribution with $n - 1$ degrees of freedom, that is, the value which cuts off an area of $\alpha/2$ in the right-hand tail. For example, earlier we found that $t_{4,.025} = 2.776$. Then a level $100(1 - \alpha)\%$ confidence interval for the population mean μ is $\overline{X} - t_{n-1,\alpha/2}(s/\sqrt{n}) < \mu < \overline{X} + t_{n-1,\alpha/2}(s/\sqrt{n})$, or $\overline{X} \pm t_{n-1,\alpha/2}(s/\sqrt{n})$.

Summary

Let $X_1, \ldots, X_n$ be a *small* random sample from a *normal* population with mean μ. Then a level $100(1 - \alpha)\%$ confidence interval for μ is

$$\overline{X} \pm t_{n-1,\alpha/2}\frac{s}{\sqrt{n}} \tag{5.9}$$

How Do I Determine Whether the Student's t Distribution Is Appropriate?

The Student's t distribution is appropriate whenever the sample comes from a population that is approximately normal. Sometimes one knows from past experience whether a process produces data that are approximately normally distributed. In many cases, however, one must decide whether a population is approximately normal by examining the sample. Unfortunately, when the sample size is small, departures from normality may be hard to detect. A reasonable way to proceed is to construct a boxplot or dotplot of

the sample. If these plots do not reveal a strong asymmetry or any outliers, then in most cases the Student's t distribution will be reliable. In principle, one can also determine whether a population is approximately normal by constructing a probability plot. With small samples, however, boxplots and dotplots are easier to draw, especially by hand.

Example 5.19

The article "Direct Strut-and-Tie Model for Prestressed Deep Beams" (K. Tan, K. Tong, and C. Tang, *Journal of Structural Engineering*, 2001:1076–1084) presents measurements of the nominal shear strength (in kN) for a sample of 15 prestressed concrete beams. The results are

580	400	428	825	850	875	920	550
575	750	636	360	590	735	950	

Is it appropriate to use the Student's t statistic to construct a 99% confidence interval for the mean shear strength? If so, construct the confidence interval. If not, explain why not.

Solution

To determine whether the Student's t statistic is appropriate, we will make a boxplot and a dotplot of the sample. These are shown in the following figure.
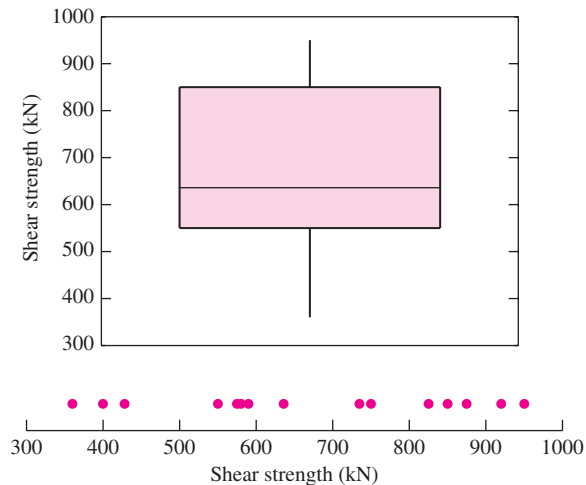
There is no evidence of a major departure from normality; in particular the plots are not strongly asymmetric, and there are no outliers. The Student's t method is appropriate. We therefore compute $\overline{X} = 668.27$ and $s = 192.089$. We use expression (5.9) with $n = 15$ and $\alpha/2 = 0.005$. From the t table with 14 degrees of freedom, we find $t_{14,.005} = 2.977$. The 99% confidence interval is $668.27 \pm (2.977)(192.089)/\sqrt{15}$, or $(520.62, \ 815.92)$.

The following computer output (from MINITAB) presents the confidence interval calculated in Example 5.19.

```
One-Sample T: Strength
Test of mu = 0 vs not = 0
Variable   N      Mean      StDev    SE Mean           99% CI
Strength   15  668.2667  192.0891  49.59718  (520.6159, 815.9175)
```

The output is self-explanatory. The quantity labeled "SE Mean" is the estimated standard deviation of the sample mean, $s/\sqrt{n}$.
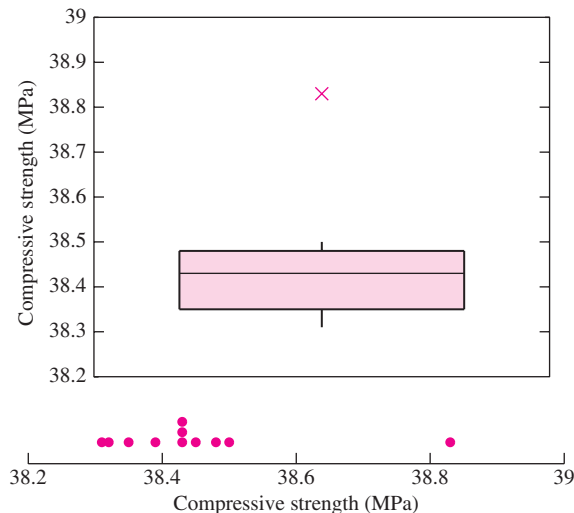
Example
5.20

In the article referred to in Example 5.19, cylindrical compressive strength (in MPa) was measured for 11 beams. The results were

  38.43 38.43 38.39 38.83 38.45 38.35 38.43 38.31 38.32 38.48 38.50

Is it appropriate to use the Student's t statistic to construct a 95% confidence interval for the mean cylindrical compressive strength? If so, construct the confidence interval. If not, explain why not.

Solution

As in Example 5.19, we will make a boxplot and a dotplot of the sample. These are shown in the following figure.

There is an outlier in this sample. The Student's t statistic should not be used.

Example

5.21

An engineer reads a report that states that a sample of 11 concrete beams had an average compressive strength of 38.45 MPa with standard deviation 0.14 MPa. Should the t curve be used to find a confidence interval for the mean compressive strength?

Solution

No. The problem is that there is no way of knowing whether the measurements came from a normal population. For example, if the measurements contained an outlier (as in Example 5.20), the confidence interval would be invalid.

The Student's t distribution can be used to compute one-sided confidence intervals. The formulas are analogous to those used with large samples.

Let $X_1, \ldots, X_n$ be a *small* random sample from a *normal* population with mean μ. Then a level $100(1 - \alpha)\%$ upper confidence bound for μ is

$$\overline{X} + t_{n-1,\alpha} \frac{s}{\sqrt{n}} \qquad (5.10)$$

and a level $100(1 - \alpha)\%$ lower confidence bound for μ is

$$\overline{X} - t_{n-1,\alpha} \frac{s}{\sqrt{n}} \qquad (5.11)$$

Use *z*, Not *t*, If σ Is Known

Occasionally a small sample may be taken from a normal population whose standard deviation σ is known. In these cases, we do not use the Student's t curve, because we are not approximating σ with s. Instead, we use the z table. Example 5.22 illustrates the method.

Example

5.22

Refer to Example 5.19. Assume that on the basis of a very large number of previous measurements of other beams, the population of shear strengths is known to be approximately normal, with standard deviation $\sigma = 180.0$ kN. Find a 99% confidence interval for the mean shear strength.

Solution

We compute $\overline{X} = 668.27$. We do not need to compute s, because we know the population standard deviation σ. Since we want a 99% confidence interval, $\alpha/2 = 0.005$. Because we know σ, we use $z_{\alpha/2} = z_{.005}$, rather than a Student's t value, to compute the confidence interval. From the z table, we obtain $z_{.005} = 2.58$. The confidence interval is $668.27 \pm (2.58)(180.0)/\sqrt{15}$, or $(548.36, 788.18)$.

It is important to remember that when the sample size is small, the population must be approximately normal, whether or not the standard deviation is known.

Summary

Let $X_1, \ldots, X_n$ be a random sample (of any size) from a *normal* population with mean μ. If the standard deviation σ is known, then a level $100(1 - \alpha)\%$ confidence interval for μ is

$$\overline{X} \pm z_{\alpha/2} \frac{\sigma}{\sqrt{n}} \tag{5.12}$$

Occasionally one has a single value that is sampled from a normal population with known standard deviation. In these cases a confidence interval for μ can be derived as a special case of expression (5.12) by setting $n = 1$.

Summary

Let X be a single value sampled from a *normal* population with mean μ. If the standard deviation σ is known, then a level $100(1 - \alpha)\%$ confidence interval for μ is

$$X \pm z_{\alpha/2}\sigma \tag{5.13}$$

Exercises for Section 5.3

1. Find the value of $t_{n-1,\alpha/2}$ needed to construct a two-sided confidence interval of the given level with the given sample size:
   a. Level 90%, sample size 12.
   b. Level 95%, sample size 7.
   c. Level 99%, sample size 2.
   d. Level 95%, sample size 29.

2. Find the value of $t_{n-1,\alpha}$ needed to construct an upper or lower confidence bound in each of the situations in Exercise 1.

3. Find the level of a two-sided confidence interval that is based on the given value of $t_{n-1,\alpha/2}$ and the given sample size.
   a. $t = 2.776$, sample size 5.
   b. $t = 2.718$, sample size 12.
   c. $t = 5.841$, sample size 4.
   d. $t = 1.325$, sample size 21.
   e. $t = 1.746$, sample size 17.

4. True or false: The Student's t distribution may be used to construct a confidence interval for the mean of any population, so long as the sample size is small.

5. The article "Ozone for Removal of Acute Toxicity from Logyard Run-off" (M. Zenaitis and S. Duff, *Ozone Science and Engineering*, 2002: 83–90) presents chemical analyses of runoff water from sawmills in British Columbia. Included were measurements of pH for six water specimens: 5.9, 5.0, 6.5, 5.6, 5.9, 6.5. Assuming these to be a random sample of water specimens from an approximately normal population, find a 95% confidence interval for the mean pH.

6. The following are summary statistics for a data set. Would it be appropriate to use the Student's t distribution to construct a confidence interval from these data? Explain.

N	Mean	Median	StDev
10	8.905	6.105	9.690

Minimum	Maximum	Q1	Q3
0.512	39.920	1.967	8.103

7. The article "An Automatic Visual System for Marble Tile Classification" (L. Carrino, W. Polini, and S. Turchetta, *Journal of Engineering Manufacture*, 2002:1095–1108) describes a measure for the shade of marble tile in which the amount of light reflected by the tile is measured on a scale of 0–255. A perfectly black tile would reflect no light and measure 0, and a perfectly white tile would measure 255. A sample of nine Mezza Perla tiles were measured, with the following results:

   204.999   206.149   202.102   207.048   203.496
   206.343   203.496   206.676   205.831

   Is it appropriate to use the Student's t statistic to construct a 95% confidence interval for the mean shade of Mezza Perla tile? If so, construct the confidence interval. If not, explain why not.

8. A chemist made eight independent measurements of the melting point of tungsten. She obtained a sample mean of 3410.14 degrees Celsius and a sample standard deviation of 1.018 degrees.

   a. Use the Student's t distribution to find a 95% confidence interval for the melting point of tungsten.

   b. Use the Student's t distribution to find a 98% confidence interval for the melting point of tungsten.

   c. If the eight measurements had been 3409.76, 3409.80, 3412.66, 3409.79, 3409.76, 3409.77, 3409.80, 3409.78, would the confidence intervals above be valid? Explain.

9. Eight independent measurements are taken of the diameter of a piston. The measurements (in inches) are: 3.236, 3.223, 3.242, 3.244, 3.228, 3.253, 3.253, and 3.230.

   a. Make a dotplot of the eight values.

   b. Should the Student's t distribution be used to find a 99% confidence interval for the diameter of this piston? If so, find the confidence interval. If not, explain why not.

   c. Eight independent measurements are taken of the diameter of another piston. The measurements this time are: 3.295, 3.232, 3.261, 3.248, 3.289, 3.245, 3.576, and 3.201. Make a dotplot of these values.

   d. Should the Student's t distribution be used to find a 95% confidence interval for the diameter of this piston? If so, find the confidence interval. If not, explain why not.

10. Fission tracks are trails found in uranium-bearing minerals, left by fragments released during fission events. The article "Yo-yo Tectonics of the Niğde Massif During Wrenching in Central Anatolia" (D. Whitney, P. Umhoefer, et al., *Turkish Journal of Earth Sciences*, 2008:209–217) reports that fifteen tracks on one rock specimen had an average track length of 13 μm with a standard deviation of 2 μm. Assuming this to be a random sample from an approximately normal population, find a 99% confidence interval for the mean track length for this rock specimen.

11. The article "Effect of Granular Subbase Thickness on Airfield Pavement Structural Response" (K. Gopalakrishnan and M. Thompson, *Journal of Materials in Civil Engineering*, 2008:331–342) presents a study of the effect of the subbase thickness on the amount of surface deflection caused by aircraft landing on an airport runway. In six applications of a 160 kN load on a runway with a subbase thickness of 864 mm, the average surface deflection was 2.03 mm with a standard deviation of 0.090 mm. Find a 90% confidence interval for the mean deflection caused by a 160 kN load.

12. Surfactants are chemical agents, such as detergents, that lower the surface tension of a liquid. Surfactants play an important role in the cleaning of contaminated soils. In an experiment to determine the effectiveness of a certain method for removing toluene from sand, the sand was washed with a surfactant, and then rinsed with de-ionized water. Of interest was the amount of toluene that came out in the rinse. In five such experiments, the amounts of toluene removed in the rinse cycle, expressed as a percentage of the total amount originally present, were 5.0, 4.8, 9.0, 10.0, and 7.3. Find a 95% confidence interval for the percentage of toluene removed in the rinse. (This exercise is based on the article "Laboratory Evaluation of the Use of Surfactants for Ground Water Remediation and the Potential for Recycling Them," D. Lee, R. Cody, and B. Hoyle, *Ground Water Monitoring and Remediation*, 2001:49–57.)

13. Twelve samples of water from a particular source have an average lead concentration of 12.5 μg/L with a standard deviation of 2.0 μg/L. Find a 95% confidence interval for the lead concentration in water from this source.

14. The following MINITAB output presents a confidence interval for a population mean.

```
One-Sample T: X
Variable     N       Mean      StDev     SE Mean            95% CI
X           10     6.59635   0.11213    0.03546    (6.51613, 6.67656)
```

    a. How many degrees of freedom does the Student's t distribution have?
    b. Use the information in the output, along with the t table, to compute a 99% confidence interval.

15. The following MINITAB output presents a confidence interval for a population mean, but some of the numbers got smudged and are now illegible. Fill in the missing numbers for (a), (b), and (c).

```
One-Sample T: X
Variable     N       Mean     StDev     SE Mean        99% CI
X           20     2.39374    (a)      0.52640    ( (b), (c) )
```

16. The concentration of carbon monoxide (CO) in a gas sample is measured by a spectrophotometer and found to be 85 ppm. Through long experience with this instrument, it is believed that its measurements are unbiased and normally distributed, with an uncertainty (standard deviation) of 8 ppm. Find a 95% confidence interval for the concentration of CO in this sample.

17. The article "Filtration Rates of the Zebra Mussel (*Dreissena polymorpha*) on Natural Seston from Saginaw Bay, Lake Huron" (D. Fanslow, T. Nalepa, and G. Lang, *Journal of Great Lakes Research* 1995:489:500) reports measurements of the rates (in mL/mg/h) at which mussels filter seston (particulate matter suspended in seawater).

    a. In the year 1992, 5 measurements were made in the Outer Bay; these averaged 21.7 with a standard deviation of 9.4. Find a 95% confidence interval for the mean filtration rate in the Outer Bay.
    b. In the year 1992, 7 measurements were made in the Inner Bay; these averaged 8.6 with a standard deviation of 4.5. Should the Student's t distribution be used to find a 95% confidence interval for the mean filtration rate for the Inner Bay? If so, find the confidence interval. If not, explain why not.

5.4 Confidence Intervals for the Difference Between Two Means

We now investigate examples in which we wish to estimate the difference between the means of two populations. The data will consist of two samples, one from each population. The basic idea is quite simple. We will compute the difference of the sample means and the standard deviation of that difference. Then a simple modification of expression (5.1) (in Section 5.1) will provide the confidence interval. The method we describe is based on the results concerning the sum and difference of two independent normal random variables that were presented in Section 4.5. We review these results here:

Let X and Y be independent, with $X \sim N(\mu_X, \sigma_X^2)$ and $Y \sim N(\mu_Y, \sigma_Y^2)$. Then

$$X + Y \sim N(\mu_X + \mu_Y, \sigma_X^2 + \sigma_Y^2) \qquad (5.14)$$

$$X - Y \sim N(\mu_X - \mu_Y, \sigma_X^2 + \sigma_Y^2) \qquad (5.15)$$

We will now see how to construct a confidence interval for the difference between two population means. As an example, assume that a new design of lightbulb has been developed that is thought to last longer than an old design. A simple random sample of 144 new lightbulbs has an average lifetime of 578 hours and a standard deviation of 22 hours. A simple random sample of 64 old lightbulbs has an average lifetime of 551 hours and a standard deviation of 33 hours. The samples are independent, in that the lifetimes for one sample do not influence the lifetimes for the other. We wish to find a 95% confidence interval for the difference between the mean lifetimes of lightbulbs of the two designs.

We begin by translating the problem into statistical language. We have a simple random sample $X_1, \ldots, X_{144}$ of lifetimes of new lightbulbs. The sample mean is $\overline{X} = 578$ and the sample standard deviation is $s_X = 22$. We have another simple random sample $Y_1, \ldots, Y_{64}$ of lifetimes of old lightbulbs. This sample has mean $\overline{Y} = 551$ and standard deviation $s_Y = 33$. The population means and standard deviations are unknown. Denote the mean of the population of lifetimes of new lightbulbs by μ_X, and the mean of the population of old lightbulbs by μ_Y. Denote the corresponding standard deviations by σ_X and σ_Y. We are interested in the difference $\mu_X - \mu_Y$.

We can construct the confidence interval for $\mu_X - \mu_Y$ by determining the distribution of $\overline{X} - \overline{Y}$. By the Central Limit Theorem, $\overline{X}$ comes from a normal distribution with mean μ_X and standard deviation $\sigma_X/\sqrt{144}$, and $\overline{Y}$ comes from a normal distribution with mean μ_Y and standard deviation $\sigma_Y/\sqrt{64}$. Since the samples are independent, it follows from expression (5.15) that the difference $\overline{X} - \overline{Y}$ comes from a normal distribution with mean $\mu_X - \mu_Y$ and variance $\sigma^2_{\overline{X}-\overline{Y}} = \sigma^2_X/144 + \sigma^2_Y/64$. Figure 5.13 illustrates the distribution of $\overline{X} - \overline{Y}$ and indicates that the middle 95% of the curve has width $\pm 1.96\sigma_{\overline{X}-\overline{Y}}$.
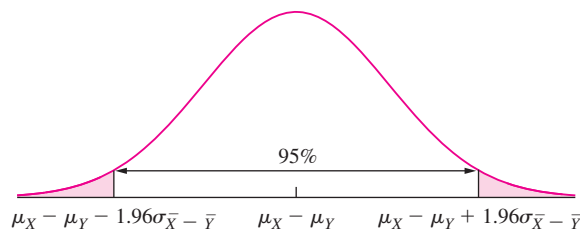

FIGURE 5.13 The observed difference $\overline{X} - \overline{Y} = 27$ is drawn from a normal distribution with mean $\mu_X - \mu_Y$ and standard deviation $\sigma_{\overline{X}-\overline{Y}} = \sqrt{\sigma^2_X/144 + \sigma^2_Y/64}$.

Estimating the population standard deviations σ_X and σ_Y with the sample standard deviations $s_X = 22$ and $s_Y = 33$, respectively, we estimate $\sigma_{\overline{X}-\overline{Y}} \approx \sqrt{22^2/144 + 33^2/64} = 4.514$. The 95% confidence interval for $\mu_X - \mu_Y$ is therefore $578 - 551 \pm 1.96(4.514)$, or 27 ± 8.85.

6.4 Small-Sample Tests for a Population Mean

In Section 6.1, we described a method for testing a hypothesis about a population mean, based on a large sample. A key step in the method is to approximate the population standard deviation σ with the sample standard deviation s. The normal curve is then used to find the P-value. When the sample size is small, s may not be close to σ, which invalidates this large-sample method. However, when the population is approximately normal, the Student's t distribution can be used. We illustrate with an example.

Spacer collars for a transmission countershaft have a thickness specification of 38.98–39.02 mm. The process that manufactures the collars is supposed to be calibrated so that the mean thickness is 39.00 mm, which is in the center of the specification window. A sample of six collars is drawn and measured for thickness. The six thicknesses are 39.030, 38.997, 39.012, 39.008, 39.019, and 39.002. Assume that the population of thicknesses of the collars is approximately normal. Can we conclude that the process needs recalibration?

Denoting the population mean by μ, the null and alternate hypotheses are

$$H_0 : \mu = 39.00 \quad \text{versus} \quad H_1 : \mu \neq 39.00$$

Note that H_0 specifies a single value for μ, since calibration requires that the mean be equal to the correct value. To construct the test statistic, note that since the population is assumed to follow a normal distribution, the quantity

$$t = \frac{\overline{X} - \mu}{s/\sqrt{n}}$$

has a Student's t distribution with $n - 1 = 5$ degrees of freedom. This is the test statistic.

In this example the observed values of the sample mean and standard deviation are $\overline{X} = 39.01133$ and $s = 0.011928$. The sample size is $n = 6$. The null hypothesis specifies that $\mu = 39$. The value of the test statistic is therefore

$$t = \frac{39.01133 - 39.00}{0.011928/\sqrt{6}} = 2.327$$

The P-value is the probability of observing a value of the test statistic whose disagreement with H_0 is as great as or greater than that actually observed. Since H_0 specifies that $\mu = 39.00$, this is a two-tailed test, so values both above and below 39.00 disagree with H_0. Therefore the P-value is the sum of the areas under the curve corresponding to $t > 2.327$ and $t < -2.327$.

Figure 6.7 illustrates the null distribution and indicates the location of the test statistic. From the t table (Table A.3 in Appendix A) the row corresponding to 5 degrees of freedom indicates that the value $t = \pm 2.015$ cuts off an area of 0.05 in each tail, for a total of 0.10, and that the value $t = \pm 2.571$ cuts off an area of 0.025 in each tail, for a total of 0.05. Thus the P-value is between 0.05 and 0.10. While we cannot conclusively state that the process is out of calibration, it doesn't look too good. It would be prudent to recalibrate.
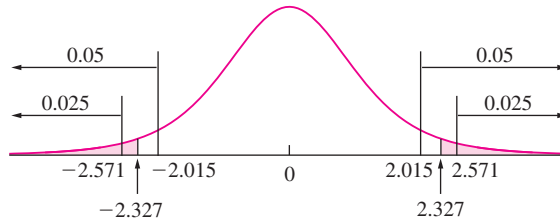
FIGURE 6.7 The null distribution of $t = (\overline{X} - 39.00)/(s/\sqrt{6})$ is Student's t with five degrees of freedom. The observed value of t, corresponding to the observed values $\overline{X} = 39.01133$ and $s = 0.011928$, is 2.327. If H_0 is true, the probability that t takes on a value as extreme as or more extreme than that observed is between 0.05 and 0.10. Because H_0 specified that μ was *equal* to a specific value, both tails of the curve contribute to the P-value.

In this example, the test statistic was a t statistic rather than a z-score. For this reason, this test is referred to as a t test.

Example
6.7

Before a substance can be deemed safe for landfilling, its chemical properties must be characterized. The article "Landfilling Ash/Sludge Mixtures" (J. Benoit, T. Eighmy, and B. Crannell, *Journal of Geotechnical and Geoenvironmental Engineering*, 1999: 877–888) reports that in a sample of six replicates of sludge from a New Hampshire wastewater treatment plant, the mean pH was 6.68 with a standard deviation of 0.20. Can we conclude that the mean pH is less than 7.0?

Solution
Let μ denote the mean pH for this type of sludge. The null and alternate hypotheses are

$$H_0 : \mu \geq 7.0 \quad \text{versus} \quad H_1 : \mu < 7.0$$

Under H_0, the test statistic

$$t = \frac{\overline{X} - 7.0}{s/\sqrt{n}}$$

has a Student's t distribution with five degrees of freedom. Substituting $\overline{X} = 6.68$, $s = 0.20$, and $n = 6$, the value of the test statistic is

$$t = \frac{6.68 - 7.00}{0.20/\sqrt{6}} = -3.919$$

Consulting the t table, we find that the value $t = -3.365$ cuts off an area of 0.01 in the left-hand tail, and the value $t = -4.033$ cuts off an area of 0.005 (see Figure 6.8 on page 420). We conclude that the P-value is between 0.005 and 0.01. There is strong evidence that the mean pH is less than 7.0.
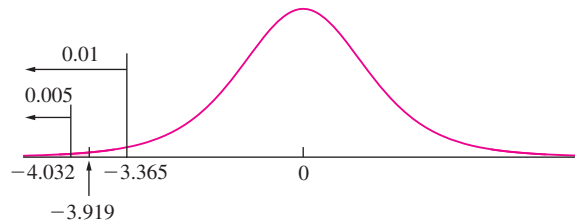
FIGURE 6.8 Solution to Example 6.7. The null distribution is Student's t with five degrees of freedom. The observed value of t is -3.919. If H_0 is true, the probability that t takes on a value as extreme as or more extreme than that observed is between 0.005 and 0.01.

The following computer output (from MINITAB) presents the results from Example 6.7.

```
One-Sample T: pH
Test of mu = 7 vs < 7
                                                  95%
                                                Upper
Variable    N     Mean    StDev    SE Mean     Bound        T        P
pH          6    6.680    0.200   0.081665   6.84453    -3.92    0.006
```

Note that the upper 95% confidence bound provided in the output is consistent with the alternate hypothesis. This indicates that the P-value is less than 5%.

Use *z*, Not *t*, If σ Is Known

Occasionally a small sample may be taken from a normal population whose standard deviation σ is known. In these cases, we do not use the Student's t curve, because we are not approximating σ with s. Instead, we use the z table and perform a z test. Example 6.8 illustrates the method.

Example **6.8**

At the beginning of this section, we described a sample of six spacer collars, whose thicknesses (in mm) were 39.030, 38.997, 39.012, 39.008, 39.019, and 39.002. We denoted the population mean thickness by μ and tested the hypotheses

$$H_0 : \mu = 39.00 \quad \text{versus} \quad H_1 : \mu \neq 39.00$$

Now assume that these six spacer collars were manufactured just after the machine that produces them had been moved to a new location. Assume that on the basis of a very large number of collars manufactured before the move, the population of collar thicknesses is known to be very close to normal, with standard deviation $\sigma = 0.010$ mm, and that it is reasonable to assume that the move has not changed this. On the basis of the given data, can we reject H_0?

We compute $\overline{X} = 39.01133$. We do not need the value of s, because we know that $\sigma = 0.010$. Since the population is normal, $\overline{X}$ is normal even though the sample size is small. The null distribution is therefore

$$\overline{X} \sim N(39.00, \ 0.010^2)$$

The z-score is

$$z = \frac{39.01133 - 39.000}{0.010/\sqrt{6}} = 2.78$$

The P-value is 0.0054, so H_0 can be rejected.

Summary

Let $X_1, \ldots, X_n$ be a sample from a *normal* population with mean μ and standard deviation σ, where σ is unknown.

To test a null hypothesis of the form $H_0 : \mu \leq \mu_0$, $H_0 : \mu \geq \mu_0$, or $H_0 : \mu = \mu_0$:

■ Compute the test statistic $t = \dfrac{\overline{X} - \mu_0}{s/\sqrt{n}}$.

■ Compute the P-value. The P-value is an area under the Student's t curve with $n - 1$ degrees of freedom, which depends on the alternate hypothesis as follows:

Alternate Hypothesis	**P-value**
$H_1 : \mu > \mu_0$	Area to the right of t
$H_1 : \mu < \mu_0$	Area to the left of t
$H_1 : \mu \neq \mu_0$	Sum of the areas in the tails cut off by t and $-t$

■ If σ is known, the test statistic is $z = \dfrac{\overline{X} - \mu_0}{\sigma/\sqrt{n}}$, and a z test should be performed.

Exercises for Section 6.4

1. Each of the following hypothetical data sets represents some repeated weighings of a standard weight that is known to have a mass of 100 g. Assume that the readings are a random sample from a population that follows the normal curve. Perform a t test to see whether the scale is properly calibrated, if possible. If impossible, explain why.

   a. 100.02, 99.98, 100.03

   b. 100.01

2. A geologist is making repeated measurements (in grams) on the mass of a rock. It is not known whether the measurements are a random sample from an approximately normal population. Below are three sets of replicate measurements, listed in the order they were made. For each set of readings, state whether the assumptions necessary for the validity of the t test appear to be met. If the assumptions are not met, explain why.

   a. 213.03 212.95 213.04 213.00 212.99
      213.01 221.03 213.05

   b. 213.05 213.00 212.94 213.09 212.98
      213.02 213.06 212.99

that is supposed to be easier to find when needed. Of the 500 husbands, 62% said that the new remote was easier to find than their old one. Of the 500 wives, only 54% said the new remote was easier to find. Let p_1 be the population proportion of married men who think

that the new remote is easier to find, and let p_2 be the corresponding proportion of married women. Can the statistic $\widehat{p}_1 - \widehat{p}_2 = 0.62 - 0.54$ be used to test $H_0 : p_1 - p_2 = 0$ versus $H_1 : p_1 - p_2 \neq 0$? If so, perform the test and compute the P-value. If not, explain why not.

14. The following MINITAB output presents the results of a hypothesis test for the difference $p_1 - p_2$ between two population proportions.

```
Test and CI for Two Proportions

Sample      X       N      Sample p
1          41       97      0.422680
2          37       61      0.606557

Difference = p (1) − p (2)
Estimate for difference: −0.183877
95% CI for difference: (−0.341016, −0.026738)
Test for difference = 0 (vs not = 0): Z = −2.25   P-Value = 0.024
```

a. Is this a one-tailed or two-tailed test?
b. What is the null hypothesis?
c. Can H_0 be rejected at the 5% level? How can you tell?

15. The following MINITAB output presents the results of a hypothesis test for the difference $p_1 - p_2$ between two population proportions. Some of the numbers are missing. Fill in the numbers for (a) through (d).

```
Test and CI for Two Proportions

Sample       X        N      Sample p
1           101      153       (a)
2           (b)       90      0.544444

Difference = p (1) − p (2)
Estimate for difference: 0.115686
95% CI for difference: (−0.0116695, 0.243042)
Test for difference = 0 (vs not = 0): Z = (c) P-Value = (d)
```

6.7 Small-Sample Tests for the Difference Between Two Means

The *t* test can be used in some cases where samples are small, and thus where the Central Limit Theorem does not apply. We present an example.

The article "The Achondroplasia Paternal Age Effect Is Not Explained By an Increase in Mutant Frequency" (I. Tiemann-Boege, W. Navidi, et al., *Proceedings of the National Academy of Sciences*, 2002:14952–14957) describes an experiment in which

a number of DNA molecules is counted, and it needs to be determined whether these molecules contain a certain sequence of nucleotides. This is done by repeating the experiment with an added enzyme that digests the sequence of interest. If the mean count is lower with the enzyme present, then it can be concluded that the molecules being counted do indeed contain the sequence.

Assume that in six identically prepared specimens without the enzyme present, the numbers of molecules counted are 33, 30, 26, 22, 37, and 34. Assume that in four identically prepared specimens with the enzyme present, the counts were 22, 29, 25, and 23. Can we conclude that the counts are lower when the enzyme is present?

We have only a few observations for each process, so the Central Limit Theorem does not apply. If both populations are approximately normal, the Student's t distribution can be used to construct a hypothesis test.

Let $X_1, \ldots, X_6$ represent the counts obtained without the enzyme, and let $Y_1, \ldots, Y_4$ represent the counts obtained with the enzyme. Let μ_X and μ_Y be the means of the populations from which these samples are drawn; let n_X and n_Y denote the sample sizes. The null and alternate hypotheses are

$$H_0 : \mu_X - \mu_Y \leq 0 \quad \text{versus} \quad H_1 : \mu_X - \mu_Y > 0$$

By assumption, both populations follow normal distributions. Therefore (as discussed in Section 5.6) the quantity

$$\frac{(\overline{X} - \overline{Y}) - (\mu_X - \mu_Y)}{\sqrt{s_X^2/n_X + s_Y^2/n_Y}} \tag{6.5}$$

has an approximate Student's t distribution with ν degrees of freedom, where

$$\nu = \frac{\left(\dfrac{s_X^2}{n_X} + \dfrac{s_Y^2}{n_Y}\right)^2}{\dfrac{(s_X^2/n_X)^2}{n_X - 1} + \dfrac{(s_Y^2/n_Y)^2}{n_Y - 1}} \qquad \text{rounded down to the nearest integer.}$$

The observed values for the sample means and standard deviations are $\overline{X} = 30.333$, $\overline{Y} = 24.750$, $s_X = 5.538$, and $s_Y = 3.096$. The sample sizes are $n_X = 6$ and $n_Y = 4$. Substituting the values for the sample standard deviations and sample sizes, we compute $\nu = 7.89$, which we round down to 7. Under H_0, $\mu_X - \mu_Y = 0$. The test statistic is therefore

$$t = \frac{(\overline{X} - \overline{Y}) - 0}{\sqrt{s_X^2/n_X + s_Y^2/n_Y}}$$

Under H_0, the test statistic has a Student's t distribution with seven degrees of freedom. Substituting values for $\overline{X}, \overline{Y}, s_X, s_Y, n_X,$ and n_Y, we compute the value of the test statistic to be

$$t = \frac{5.583 - 0}{2.740} = 2.038$$

Consulting the t table with seven degrees of freedom, we find that the value cutting off 5% in the right-hand tail is 1.895, and the value cutting off 2.5% is 2.365. The P-value

is therefore between 0.025 and 0.05 (see Figure 6.13). We conclude that the mean count is lower when the enzyme is present.
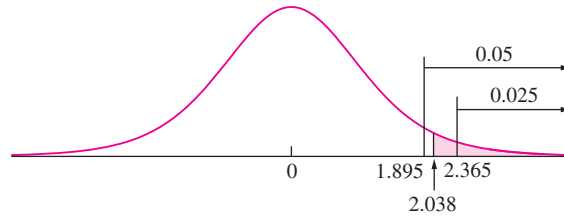

FIGURE 6.13 The null distribution is Student's t with seven degrees of freedom. The observed value of the test statistic is 2.038. If H_0 is true, the probability that t takes on a value as extreme as or more extreme than that observed is between 2.5% and 5%.

Example
6.12

Good website design can make Web navigation easier. The article "The Implications of Visualization Ability and Structure Preview Design for Web Information Search Tasks" (H. Zhang and G. Salvendy, *International Journal of Human-Computer Interaction*, 2001:75–95) presents a comparison of item recognition between two designs. A sample of 10 users using a conventional Web design averaged 32.3 items identified, with a standard deviation of 8.56. A sample of 10 users using a new structured Web design averaged 44.1 items identified, with a standard deviation of 10.09. Can we conclude that the mean number of items identified is greater with the new structured design?

Solution

Let $\overline{X} = 44.1$ be the sample mean for the structured Web design. Then $s_X = 10.09$ and $n_X = 10$. Let $\overline{Y} = 32.3$ be the sample mean for the conventional Web design. Then $s_Y = 8.56$ and $n_Y = 10$. Let μ_X and μ_Y denote the population mean measurements made by the structured and conventional methods, respectively. The null and alternate hypotheses are

$$H_0 : \mu_X - \mu_Y \le 0 \quad \text{versus} \quad H_1 : \mu_X - \mu_Y > 0$$

The test statistic is

$$t = \frac{(\overline{X} - \overline{Y}) - 0}{\sqrt{s_X^2/n_X + s_Y^2/n_Y}}$$

Substituting values for $\overline{X}$, $\overline{Y}$, s_X, s_Y, n_X, and n_Y, we compute the value of the test statistic to be $t = 2.820$. Under H_0, this statistic has an approximate Student's t distribution, with the number of degrees of freedom given by

$$\nu = \frac{\left(\dfrac{10.09^2}{10} + \dfrac{8.56^2}{10}\right)^2}{\dfrac{(10.09^2/10)^2}{9} + \dfrac{(8.56^2/10)^2}{9}} = 17.53 \approx 17$$

Consulting the t table with 17 degrees of freedom, we find that the value cutting off 1% in the right-hand tail is 2.567, and the value cutting off 0.5% in the right-hand tail is 2.898. Therefore the area in the right-hand tail corresponding to values as extreme as or more extreme than the observed value of 2.820 is between 0.005 and 0.010. Therefore $0.005 < P < 0.01$ (see Figure 6.14). There is strong evidence that the mean number of items identified is greater for the new design.
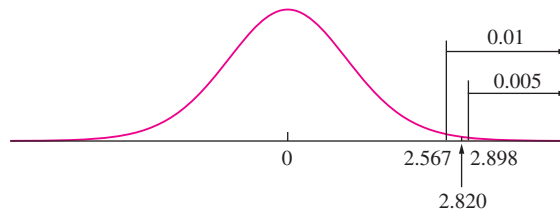

FIGURE 6.14 Solution to Example 6.12. The P-value is the area in the right-hand tail, which is between 0.005 and 0.01.

The following computer output (from MINITAB) presents the results from Example 6.12.

```
Two-Sample T-Test and CI: Struct, Conven

Two-sample T for C1 vs C2

              N        Mean       StDev      SE Mean
Struct       10       44.10       10.09      3.19074
Conven       10       32.30        8.56      2.70691

Difference = mu (Struct) — mu (Conven)
Estimate for difference: 11.8000
95% lower bound for difference: 4.52100
T-Test of difference = 0 (vs >):
T-Value = 2.82   P-Value = 0.006   DF = 17
```

Note that the 95% lower confidence bound is consistent with the alternate hypothesis. This indicates that the P-value is less than 5%.

The methods described in this section can be used to test a hypothesis that two population means differ by a specified constant. Example 6.13 shows how.

Example **6.13**

Refer to Example 6.12. Can you conclude that the mean number of items identified with the new structured design exceeds that of the conventional design by more than 2?

Solution

The null and alternate hypotheses are

$$H_0: \mu_X - \mu_Y \leq 2 \quad \text{versus} \quad H_1: \mu_X - \mu_Y > 2$$

We observe $\overline{X} = 44.1$, $\overline{Y} = 32.3$, $s_X = 10.09$, $s_Y = 8.56$, $n_X = 10$, and $n_Y = 10$. Under H_0, we take $\mu_X - \mu_Y = 2$. The test statistic is given by expression (6.5) to be

$$t = \frac{(\overline{X} - \overline{Y}) - 2}{\sqrt{s_X^2/n_X + s_Y^2/n_Y}}$$

Under H_0, the test statistic has a Student's t distribution with 17 degrees of freedom. Note that the number of degrees of freedom is calculated in the same way as in Example 6.12. The value of the test statistic is $t = 2.342$. This is a one-tailed test. The P-value is between 0.01 and 0.025. We conclude that the mean number of items identified with the new structured design exceeds that of the conventional design by more than 2.

Summary

Let $X_1, \ldots, X_{n_X}$ and $Y_1, \ldots, Y_{n_Y}$ be samples from *normal* populations with means μ_X and μ_Y and standard deviations σ_X and σ_Y, respectively. Assume the samples are drawn independently of each other.

If σ_X and σ_Y are not known to be equal, then, to test a null hypothesis of the form $H_0 : \mu_X - \mu_Y \leq \Delta_0$, $H_0 : \mu_X - \mu_Y \geq \Delta_0$, or $H_0 : \mu_X - \mu_Y = \Delta_0$:

- Compute $\nu = \dfrac{[(s_X^2/n_X) + (s_Y^2/n_Y)]^2}{[(s_X^2/n_X)^2/(n_X - 1)] + [(s_Y^2/n_Y)^2/(n_Y - 1)]}$, rounded down to the nearest integer.

- Compute the test statistic $t = \dfrac{(\overline{X} - \overline{Y}) - \Delta_0}{\sqrt{s_X^2/n_X + s_Y^2/n_Y}}$.

- Compute the P-value. The P-value is an area under the Student's t curve with ν degrees of freedom, which depends on the alternate hypothesis as follows:

Alternate Hypothesis	P-value
$H_1 : \mu_X - \mu_Y > \Delta_0$	Area to the right of t
$H_1 : \mu_X - \mu_Y < \Delta_0$	Area to the left of t
$H_1 : \mu_X - \mu_Y \neq \Delta_0$	Sum of the areas in the tails cut off by t and $-t$

When the Populations Have Equal Variances

When the population variances are known to be nearly equal, the pooled variance (see Section 5.6) may be used. The pooled variance is given by

$$s_p^2 = \frac{(n_X - 1)s_X^2 + (n_Y - 1)s_Y^2}{n_X + n_Y - 2}$$

The test statistic for testing any of the null hypotheses $H_0 : \mu_X - \mu_Y = 0$, $H_0 : \mu_X - \mu_Y \leq 0$, or $H_0 : \mu_X - \mu_Y \geq 0$ is

$$t = \frac{\overline{X} - \overline{Y}}{s_p\sqrt{1/n_X + 1/n_Y}}$$

Under H_0, the test statistic has a Student's t distribution with $n_X + n_Y - 2$ degrees of freedom.

Example
6.14

Two methods have been developed to determine the nickel content of steel. In a sample of five replications of the first method on a certain kind of steel, the average measurement (in percent) was $\overline{X} = 3.16$ and the standard deviation was $s_X = 0.042$. The average of seven replications of the second method was $\overline{Y} = 3.24$ and the standard deviation was $s_Y = 0.048$. Assume that it is known that the population variances are nearly equal. Can we conclude that there is a difference in the mean measurements between the two methods?

Solution
Substituting the sample sizes $n_X = 5$ and $n_Y = 7$ along with the sample standard deviations $s_X = 0.042$ and $s_Y = 0.048$, we compute the pooled standard deviation, obtaining $s_p = 0.0457$.

The value of the test statistic is therefore

$$t = \frac{3.16 - 3.24}{0.0457\sqrt{1/5 + 1/7}} = -2.990$$

Under H_0, the test statistic has the Student's t distribution with 10 degrees of freedom. Consulting the Student's t table, we find that the area under the curve in each tail is between 0.01 and 0.005. Since the null hypothesis stated that the means are equal, this is a two-tailed test, so the P-value is the sum of the areas in both tails. We conclude that $0.01 < P < 0.02$ (see Figure 6.15). There does appear to be a difference in the mean measurements between the two methods.
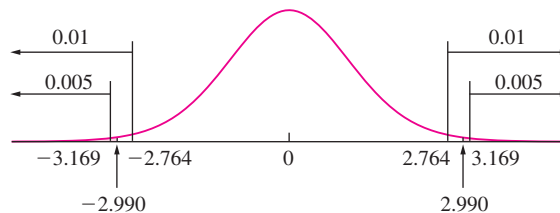

FIGURE 6.15 Solution to Example 6.14. The P-value is the sum of the areas in both tails, which is between 0.01 and 0.02.

Don't Assume the Population Variances Are Equal Just Because the Sample Variances Are Close

The major problem with using the pooled variance is that the assumption that the population variances are equal is very strict. The hypothesis test based on the pooled variance can be quite unreliable if it is used when the population variances are not equal. In practice, the population variances are almost always unknown, and it is usually impossible to be sure that they are equal.

In situations where the *sample* variances are nearly equal, it is tempting to assume that the population variances are nearly equal as well. This assumption is not justified, however, because it is possible for the sample variances to be nearly equal even when the population variances are quite different. Computer packages often offer a choice of assuming variances to be equal or unequal. The best practice is to assume the variances to be unequal unless it is quite certain that they are equal. See the discussion in Section 5.6.

Summary

Let $X_1, \ldots, X_{n_X}$ and $Y_1, \ldots, Y_{n_Y}$ be samples from *normal* populations with means μ_X and μ_Y and standard deviations σ_X and σ_Y, respectively. Assume the samples are drawn independently of each other.

If σ_X and σ_Y are known to be equal, then, to test a null hypothesis of the form $H_0 : \mu_X - \mu_Y \leq \Delta_0$, $H_0 : \mu_X - \mu_Y \geq \Delta_0$, or $H_0 : \mu_X - \mu_Y = \Delta_0$:

- Compute $s_p = \sqrt{\dfrac{(n_X - 1)s_X^2 + (n_Y - 1)s_Y^2}{n_X + n_Y - 2}}$.

- Compute the test statistic $t = \dfrac{(\overline{X} - \overline{Y}) - \Delta_0}{s_p\sqrt{1/n_X + 1/n_Y}}$.

- Compute the P-value. The P-value is an area under the Student's t curve with $n_X + n_Y - 2$ degrees of freedom, which depends on the alternate hypothesis as follows:

Alternate Hypothesis	P-value
$H_1 : \mu_X - \mu_Y > \Delta_0$	Area to the right of t
$H_1 : \mu_X - \mu_Y < \Delta_0$	Area to the left of t
$H_1 : \mu_X - \mu_Y \neq \Delta_0$	Sum of the areas in the tails cut off by t and $-t$

Exercises for Section 6.7

1. A crayon manufacturer is comparing the effects of two kinds of yellow dye on the brittleness of crayons. Dye B is more expensive than dye A, but it is thought that it might produce a stronger crayon. Four crayons are tested with each kind of dye, and the impact strength (in joules) is measured for each. The results are as follows:

Dye A: 1.0   2.0   1.2   3.0
Dye B: 3.0   3.2   2.6   3.4

  a. Can you conclude that the mean strength of crayons made with dye B is greater than that of crayons made with dye A?

  b. Can you conclude that the mean strength of crayons made with dye B exceeds that of crayons made with dye A by more than 1 J?

2. In a study of the relationship of the shape of a tablet to its dissolution time, 6 disk-shaped ibuprofen tablets and 8 oval-shaped ibuprofen tablets were dissolved in water. The dissolve times, in seconds, were as follows:

Disk: 269.0 249.3 255.2 252.7 247.0 261.6

Oval: 268.8 260.0 273.5 253.9 278.5 289.4
       261.6 280.2

Can you conclude that the mean dissolve times differ between the two shapes?

3. The article "Modeling Resilient Modulus and Temperature Correction for Saudi Roads" (H. Wahhab, I. Asi, and R. Ramadhan, *Journal of Materials in Civil Engineering*, 2001:298–305) describes a study

designed to predict the resilient modulus of pavement from physical properties. One of the questions addressed was whether the modulus differs between rutted and nonrutted pavement. Measurements of the resilient modulus at 40°C (in 10^6 kPa) are presented below for 5 sections of rutted pavement and 9 sections of nonrutted pavement.

Rutted: 1.48 1.88 1.90 1.29 1.00

Nonrutted: 3.06 2.58 1.70 2.44 2.03
                   1.76 2.86 2.82 1.04

Perform a hypothesis test to determine whether it is plausible that the mean resilient modulus is the same for rutted and nonrutted pavement. Compute the P-value. What do you conclude?

4. The article "Time Series Analysis for Construction Productivity Experiments" (T. Abdelhamid and J. Everett, *Journal of Construction Engineering and Management*, 1999:87–95) presents a study comparing the effectiveness of a video system that allows a crane operator to see the lifting point while operating the crane with the old system in which the operator relies on hand signals from a tagman. Three different lifts, A, B, and C, were studied. Lift A was of little difficulty, lift B was of moderate difficulty, and lift C was of high difficulty. Each lift was performed several times, both with the new video system and with the old tagman system. The time (in seconds) required to perform each lift was recorded. The following tables present the means, standard deviations, and sample sizes.

| | **Low Difficulty** | | |
	Mean	**Standard Deviation**	**Sample Size**
Tagman	47.79	2.19	14
Video	47.15	2.65	40

| | **Moderate Difficulty** | | |
	Mean	**Standard Deviation**	**Sample Size**
Tagman	69.33	6.26	12
Video	58.50	5.59	24

| | **High Difficulty** | | |
	Mean	**Standard Deviation**	**Sample Size**
Tagman	109.71	17.02	17
Video	84.52	13.51	29

a. Can you conclude that the mean time to perform a lift of low difficulty is less when using the video system than when using the tagman system? Explain.

b. Can you conclude that the mean time to perform a lift of moderate difficulty is less when using the video system than when using the tagman system? Explain.

c. Can you conclude that the mean time to perform a lift of high difficulty is less when using the video system than when using the tagman system? Explain.

5. The article "Calibration of an FTIR Spectrometer" (P. Pankratz, *Statistical Case Studies for Industrial and Process Improvement*, SIAM-ASA, 1997:19–38) describes the use of a spectrometer to make five measurements of the carbon content (in ppm) of a certain silicon wafer on each of two successive days. The results were as follows:

Day 1:   2.1321 2.1385 2.0985 2.0941 2.0680

Day 2:   2.0853 2.1476 2.0733 2.1194 2.0717

Can you conclude that the calibration of the spectrometer has changed from the first day to the second day?

6. Two weights, each labeled as weighing 100 g, are each weighed several times on the same scale. The results, in units of μg above 100 g, are as follows:

First weight: 53 88 89 62 39 66
Second weight: 23 39 28   2 49

Since the same scale was used for both weights, and since both weights are similar, it is reasonable to assume that the variance of the weighing does not depend on the object being weighed. Can you conclude that the weights differ?

7. It is thought that a new process for producing a certain chemical may be cheaper than the currently used process. Each process was run 6 times, and the cost of

producing 100 L of the chemical was determined each time. The results, in dollars, were as follows:

New Process: 51 52 55 53 54 53
Old Process: 50 54 59 56 50 58

Can you conclude that the mean cost of the new method is less than that of the old method?

8. The article "Effects of Aerosol Species on Atmospheric Visibility in Kaohsiung City, Taiwan" (C. Lee, C. Yuan, and J. Chang, *Journal of Air and Waste Management*, 2005:1031–1041) reported that for a sample of 20 days in the winter, the mass ratio of fine to coarse particles averaged 0.51 with a standard deviation of 0.09, and for a sample of 14 days in the spring the mass ratio averaged 0.62 with a standard deviation of 0.09. Let μ_1 represent the mean mass ratio during the winter and let μ_2 represent the mean mass ratio during the summer. It is desired to test $H_0: \mu_2 - \mu_1 = 0$ versus $H_1: \mu_2 - \mu_1 \neq 0$.

   a. Someone suggests that since the sample standard deviations are equal, the pooled variance should be used. Do you agree? Explain.

   b. Using an appropriate method, perform the test.

9. The article "Mechanical Grading of Oak Timbers" (D. Kretschmann and D. Green, *Journal of Materials in Civil Engineering*, 1999:91–97) presents measurements of the ultimate compressive stress, in MPa, for green mixed oak 7 by 9 timbers from West Virginia and Pennsylvania. For 11 specimens of no. 1 grade lumber, the average compressive stress was 22.1 with a standard deviation of 4.09. For 7 specimens of no. 2 grade lumber, the average compressive stress was 20.4 with a standard deviation of 3.08. Can you conclude that the mean compressive stress is greater for no. 1 grade lumber than for no. 2 grade?

10. The article "Magma Interaction Processes Inferred from Fe-Ti Oxide Compositions in the Dölek and Sariçiçek Plutons, Eastern Turkey" (O. Karsli, F. Aydin, et al., *Turkish Journal of Earth Sciences*, 2008:297–315) presents chemical compositions (in weight-percent) for several rock specimens. Fourteen specimens (two outliers were removed) of limenite grain had an average iron oxide (Fe_2O_3) content of 9.30 with a standard deviation of 2.71, and seven specimens of limenite lamella had an average iron oxide content of 9.47 with a standard deviation of 2.22. Can you

conclude that the mean iron oxide content differs between limenite grain and limenite lamella?

11. The article "Structural Performance of Rounded Dovetail Connections Under Different Loading Conditions" (T. Tannert, H. Prion, and F. Lam, *Can J Civ Eng*, 2007:1600–1605) describes a study of the deformation properties of dovetail joints. In one experiment, 10 rounded dovetail connections and 10 double rounded dovetail connections were loaded until failure. The rounded connections had an average load at failure of 8.27 kN with a standard deviation of 0.62 kN. The double-rounded connections had an average load at failure of 6.11 kN with a standard deviation of 1.31 kN. Can you conclude that the mean load at failure is greater for rounded connections than for double-rounded connections?

12. The article "Variance Reduction Techniques: Experimental Comparison and Analysis for Single Systems" (I. Sabuncuoglu, M. Fadiloglu, and S. Celik, *IIE Transactions*, 2008:538–551) describes a study of the effectiveness of the method of Latin Hypercube Sampling in reducing the variance of estimators of the mean time-in-system for queueing models. For the M/M/1 queueing model, ten replications of the experiment yielded an average reduction of 6.1 with a standard deviation of 4.1. For the serial line model, ten replications yielded an average reduction of 6.6 with a standard deviation of 4.3. Can you conclude that the mean reductions differ between the two models?

13. In an experiment to test the effectiveness of a new sleeping aid, a sample of 12 patients took the new drug, and a sample of 14 patients took a commonly used drug. Of the patients taking the new drug, the average time to fall asleep was 27.3 minutes with a standard deviation of 5.2 minutes, and for the patients taking the commonly used drug the average time was 32.7 minutes with a standard deviation of 4.1 minutes. Can you conclude that the mean time to sleep is less for the new drug?

14. Refer to Exercise 11 in Section 5.6. Can you conclude that the mean breaking strength is greater for hockey sticks made from composite B?

15. Refer to Exercise 12 in Section 5.6. Can you conclude that the mean permeability coefficient at 60°C differs from that at 61°C?

16. The following MINITAB output presents the results of a hypothesis test for the difference $\mu_X - \mu_Y$ between two population means.

```
Two-sample T for X vs Y

        N     Mean    StDev   SE Mean
X      10    39.31     8.71      2.8
Y      10    29.12     4.79      1.5

Difference = mu (X) − mu (Y)
Estimate for difference: 10.1974
95% lower bound for difference: 4.6333
T-Test of difference = 0 (vs >): T-Value = 3.25  P-Value = 0.003  DF = 13
```

a. Is this a one-tailed or two-tailed test?
b. What is the null hypothesis?
c. Can H_0 be rejected at the 1% level? How can you tell?

17. The following MINITAB output presents the results of a hypothesis test for the difference $\mu_X - \mu_Y$ between two population means. Some of the numbers are missing. Fill in the numbers for (a) through (d).

```
Two-sample T for X vs Y

        N     Mean    StDev   SE Mean
X       6    1.755    0.482     (a)
Y      13    3.239     (b)     0.094
Difference = mu (X) − mu (Y)
Estimate for difference: (c)
95% CI for difference: (−1.99996, −0.96791)
T-Test of difference = 0 (vs not =): T-Value = (d)  P-Value = 0.000  DF = 7
```

6.8 Tests with Paired Data

We saw in Section 5.7 that it is sometimes better to design a two-sample experiment so that each item in one sample is paired with an item in the other. In this section, we present a method for testing hypotheses involving the difference between two population means on the basis of such paired data. We begin with an example.

Particulate matter (PM) emissions from automobiles are a serious environmental concern. Eight vehicles were chosen at random from a fleet, and their emissions were measured under both highway driving and stop-and-go driving conditions. The differences (stop-and-go emission − highway emission) were computed as well. The results, in milligrams of particulates per gallon of fuel, were as follows:

				Vehicle				
	1	2	3	4	5	6	7	8
Stop-and-go	1500	870	1120	1250	3460	1110	1120	880
Highway	941	456	893	1060	3107	1339	1346	644
Difference	559	414	227	190	353	−229	−226	236

Can we conclude that the mean level of emissions is less for highway driving than for stop-and-go driving?

The basic idea behind the construction of the hypothesis test in this example is the same as the idea behind the construction of confidence intervals for paired data presented in Section 5.7. We treat the collection of differences as a single random sample from a population of differences. Denote the population mean of the differences by μ_D and the standard deviation by σ_D. There are only eight differences, which is a small sample. If we assume that the population of differences is approximately normal, we can use the Student's t test, as presented in Section 6.4.

The observed value of the sample mean of the differences is $\overline{D} = 190.5$. The sample standard deviation is $s_D = 284.1$. The null and alternate hypotheses are

$$H_0 : \mu_D \leq 0 \quad \text{versus} \quad H_1 : \mu_D > 0$$

The test statistic is

$$t = \frac{\overline{D} - 0}{s_D / \sqrt{n}} = \frac{190.5 - 0}{284.1 / \sqrt{8}} = 1.897$$

The null distribution of the test statistic is Student's t with seven degrees of freedom. Figure 6.16 presents the null distribution and indicates the location of the test statistic. This is a one-tailed test. The t table indicates that 5% of the area in the tail is cut off by a t value of 1.895, very close to the observed value of 1.897. The P-value is approximately 0.05. The following computer output (from MINITAB) presents this result.

```
Paired T-Test and CI: StopGo, Highway

Paired T for StopGo - Highway

                N        Mean        StDev       SE Mean
StopGo          8        1413.75     850.780     300.796
Highway         8        1223.25     820.850     290.214
Difference      8         190.50     284.104     100.446

95% lower bound for mean difference: 0.197215
T-Test of mean difference = 0 (vs > 0):
T-Value = 1.90  P-Value = 0.050
```
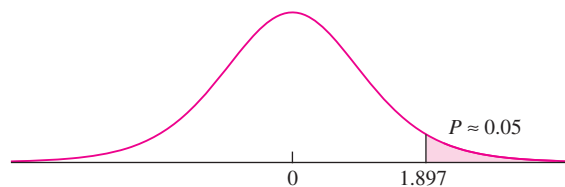

FIGURE 6.16 The null distribution of $t = (\overline{D} - 0)/(s_D/\sqrt{8})$ is t_7. The observed value of t, corresponding to the observed values $\overline{D} = 190.5$ and $s_p = 284.1$, is 1.897. If H_0 is true, the probability that t takes on a value as extreme as or more extreme than that observed is very close to 0.05.

Note that the 95% lower bound is just barely consistent with the alternate hypothesis. This indicates that the P-value is just barely less than 0.05 (although it is given by 0.050 to two significant digits).

Summary

Let $(X_1, Y_1), \ldots, (X_n, Y_n)$ be a sample of ordered pairs whose differences $D_1, \ldots, D_n$ are a sample from a *normal* population with mean μ_D. Let s_D be the sample standard deviation of $D_1, \ldots, D_n$.

To test a null hypothesis of the form $H_0 : \mu_D \leq \mu_0$, $H_0 : \mu_D \geq \mu_0$, or $H_0 : \mu_D = \mu_0$:

■ Compute the test statistic $t = \dfrac{\overline{D} - \mu_0}{s_D/\sqrt{n}}$.

■ Compute the P-value. The P-value is an area under the Student's t curve with $n - 1$ degrees of freedom, which depends on the alternate hypothesis as follows:

Alternate Hypothesis	*P*-value
$H_1 : \mu_D > \mu_0$	Area to the right of t
$H_1 : \mu_D < \mu_0$	Area to the left of t
$H_1 : \mu_D \neq \mu_0$	Sum of the areas in the tails cut off by t and $-t$

■ If the sample is large, the D_i need not be normally distributed, the test statistic is $z = \dfrac{\overline{D} - \mu_0}{s_D/\sqrt{n}}$, and a z test should be performed.

Exercises for Section 6.8

1. Muscles flex when stimulated through electric impulses either to motor points (points on the muscle) or to nerves. The article "Force Assessment of the Stimulated Arm Flexors: Quantification of Contractile Properties" (J. Hong and P. Iaizzo, *Journal of Medical Engineering and Technology*, 2002:28–35) reports a study in which both methods were applied to the upper-arm regions of each of several subjects. The latency time (delay between stimulus and contraction) was measured (in milliseconds) for each subject. The results for seven subjects are presented in the following table (one outlier has been deleted).

	Subject						
	1	**2**	**3**	**4**	**5**	**6**	**7**
Nerve	59	57	58	38	53	47	51
Motor point	56	52	56	32	47	42	48
Difference	3	5	2	6	6	5	3

Can you conclude that there is a difference in latency period between motor point and nerve stimulation?

2. The Valsalva maneuver involves blowing into a closed tube in order to create pressure in respiratory airways. Impedance cardiography is used during this maneuver to assess cardiac function. The article "Impedance Cardiographic Measurement of the Physiological Response to the Valsalva Manoeuvre" (R. Patterson and J. Zhang, *Medical and Biological Engineering and Computing*, 2003: 40–43) presents a study in which the impedance ratio was measured for each of 11 subjects in both a standing and a reclining position. The results at an airway pressure of 10 mmHg are presented in the following table.