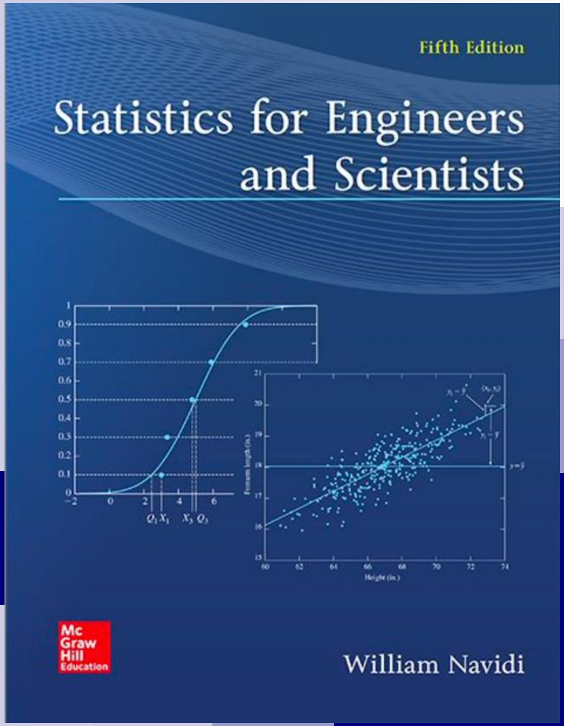


Chapter 5

Confidence Intervals (part 2)

Ch. 5 Overview (required sections)

- **5-1** Large-Sample Confidence Intervals for a Population Mean
- **5-2 Confidence Intervals for Proportions**
- **5-3 Small-Sample Confidence Intervals for a Population Mean**
- **5-4** Confidence Intervals for the Difference Between Two Means
- **5-7** Confidence Intervals with Paired Data

Confidence Intervals for Proportions

Example 1:

”A USA TODAY Snapshots feature stated that 12% of the pleasure boats in the US were named *Serenity*.”

- The parameter 12% is called a **proportion**.
- It means that of all the pleasure boats in the US, 12 out of every 100 are named Serenity.
- A proportion represents a part of a whole.
- It can be expressed as a fraction, decimal, or percentage.
- In this case, 12% = 0.12 = 12 /100 or 3/25.

Confidence Intervals for Proportions…

Symbols Used in Proportion Notation

$$p = \text{population proportion}$$

$$\hat{p} \text{ (read “}p\text{ hat”)} = \text{sample proportion}$$

For a sample proportion,

$$\boxed{\hat{p} = \frac{X}{n}} \quad \text{and} \quad \hat{q} = \frac{n - X}{n} \quad \text{or} \quad \hat{q} = 1 - \hat{p}$$

where X = number of sample units that possess the characteristics of interest and n = sample size.

Confidence Intervals for Proportions…

Example 2:

"In a study, 200 people were asked if they were satisfied with their jobs or professions; 162 said that they were."

In this case, n = 200, X = 162, and $\hat{p}$ = X/n = 162/200 = 0.81

It can be said that for this sample, 0.81, or 81%, of those surveyed were satisfied with their jobs or professions.

The sample **proportion is** $\hat{p}$ = 0.81.

Exercise: Air-Conditioned Households

In a recent survey of 150 households, 54 had central air conditioning. Find $\hat{p}$ and $\hat{q}$, where $\hat{p}$ is the proportion of households that have central air conditioning.

Exercise: Air-Conditioned Households

In a recent survey of 150 households, 54 had central air conditioning. Find $\hat{p}$ and $\hat{q}$, where $\hat{p}$ is the proportion of households that have central air conditioning.

Since X = *54 and* n = *150*

$$\hat{p} = \frac{X}{n} = \frac{54}{150} = 0.36 = 36\%$$

$$\hat{q} = 1 - \hat{p} = 1 - 0.36 = 0.64 = 64\%$$

Formula for a Specific Confidence Interval for a Proportion

$$\hat{p} - z_{\alpha/2}\sqrt{\frac{\hat{p}\hat{q}}{n}} < p < \hat{p} + z_{\alpha/2}\sqrt{\frac{\hat{p}\hat{q}}{n}}$$

when $n\hat{p} \geq 5$ and $n\hat{q} \geq 5$.

Rounding Rule: Round off to three decimal places.

EXAMPLE 7–9 Covering College Costs

A survey conducted by Sallie Mae and Gallup of 1404 respondents found that 323 students paid for their education by student loans. Find the 90% confidence interval of the true proportion of students who paid for their education by student loans.

SOLUTION

Step 1 Determine $\hat{p}$ and $\hat{q}$.

$$\hat{p} = \frac{X}{n} = \frac{323}{1404} = 0.23$$

$$\hat{q} = 1 - \hat{p} = 1.00 - 0.23 = 0.77$$

Step 2 Determine the critical value.

$$\alpha = 1 - 0.90 = 0.10$$

$$\frac{\alpha}{2} = \frac{0.10}{2} = 0.05$$

$$z_{\alpha/2} = 1.65$$

Step 3 Substitute in the formula

$$\hat{p} - z_{\alpha/2}\sqrt{\frac{\hat{p}\hat{q}}{n}} < p < \hat{p} + z_{\alpha/2}\sqrt{\frac{\hat{p}\hat{q}}{n}}$$

$$0.23 - 1.65\sqrt{\frac{(0.23)(0.77)}{1404}} < p < 0.23 + 1.65\sqrt{\frac{(0.23)(0.77)}{1404}}$$

$$0.23 - 0.019 < p < 0.23 + 0.019$$

$$0.211 < p < 0.249$$

or

$$21.1\% < p < 24.9\%$$

Hence, you can be 90% confident that the percentage of students who pay for their college education by student loans is between 21.1 and 24.9%.

- The method described in the previous slides is often called *'traditional method'* (textbook - p.342).

Summary

The Traditional Method for Computing Confidence Intervals for a Proportion (widely used but not recommended)

Let $\widehat{p}$ be the proportion of successes in a *large* number n of independent Bernoulli trials with success probability p. Then the traditional level $100(1 - \alpha)\%$ confidence interval for p is

$$\widehat{p} \pm z_{\alpha/2}\sqrt{\frac{\widehat{p}(1-\widehat{p})}{n}} \qquad (5.8)$$

The method cannot be used unless the sample contains at least 10 successes and 10 failures.

- This method is **REQUIRED** in this course

The ‘Modern Method’ (p.341)

- ✓ Recent research has shown that the effect of a small sample size can be compensated for by modifying both *n* and *p* slightly.
- ✓ Specifically, one should add 4 to the number of trials, and 2 to the number of successes.

Summary

Let X be the number of successes in n independent Bernoulli trials with success probability p, so that $X \sim \text{Bin}(n, p)$.

Define $\tilde{n} = n + 4$, and $\tilde{p} = \dfrac{X+2}{\tilde{n}}$. Then a level $100(1-\alpha)\%$ confidence interval for p is

$$\tilde{p} \pm z_{\alpha/2}\sqrt{\frac{\tilde{p}(1-\tilde{p})}{\tilde{n}}} \tag{5.5}$$

If the lower limit is less than 0, replace it with 0. If the upper limit is greater than 1, replace it with 1.

‘Modern Method’ vs. ‘Traditional Method’

✓For very large sample sizes, the results of the traditional method are almost identical to those of the modern method.

✓ For small or moderately large sample sizes, the modern approach is better.

Formula for Minimum Sample Size Needed for Interval Estimate of a Population Proportion

$$n = \hat{p}\hat{q}\left(\frac{z_{\alpha/2}}{E}\right)^2$$

If necessary, round up to the next whole number.

Minimum Sample Size for Population Proportion

There are 2 situations to consider:

- Some approximation of $\hat{p}$ is known (for example from a previous study)
- No approximation of $\hat{p}$ is known, use $\hat{p}$ = 0.5. This will yield the largest possible value of *n*.

$$n = \hat{p}\hat{q}\left(\frac{z_{\alpha/2}}{E}\right)^2$$

$\hat{p}$	$\hat{q}$	$\hat{p}\hat{q}$
0.1	0.9	0.09
0.2	0.8	0.16
0.3	0.7	0.21
0.4	0.6	0.24
0.5	**0.5**	**0.25**
0.6	0.4	0.24
0.7	0.3	0.21
0.8	0.2	0.16
0.9	0.1	0.09

Example: Land Line Phones

A researcher wishes to estimate, with 95% confidence, the proportion of people who did not have a land line phone. A previous study shows that 40% of those interviewed did not have a land line phone. The researcher wishes to be accurate within 2% of the true proportion. Find the minimum sample size necessary.

Example: Land Line Phones

A researcher wishes to estimate, with 95% confidence, the proportion of people who did not have a land line phone. A previous study shows that 40% of those interviewed did not have a land line phone. The researcher wishes to be accurate within 2% of the true proportion. Find the minimum sample size necessary.

$$\boxed{n = \hat{p}\hat{q}\left(\frac{z_{\alpha/2}}{E}\right)^2} = (0.40)(0.60)\left(\frac{1.96}{0.02}\right)^2 = 2304.96$$

The researcher should interview a sample of at least **2305** people.

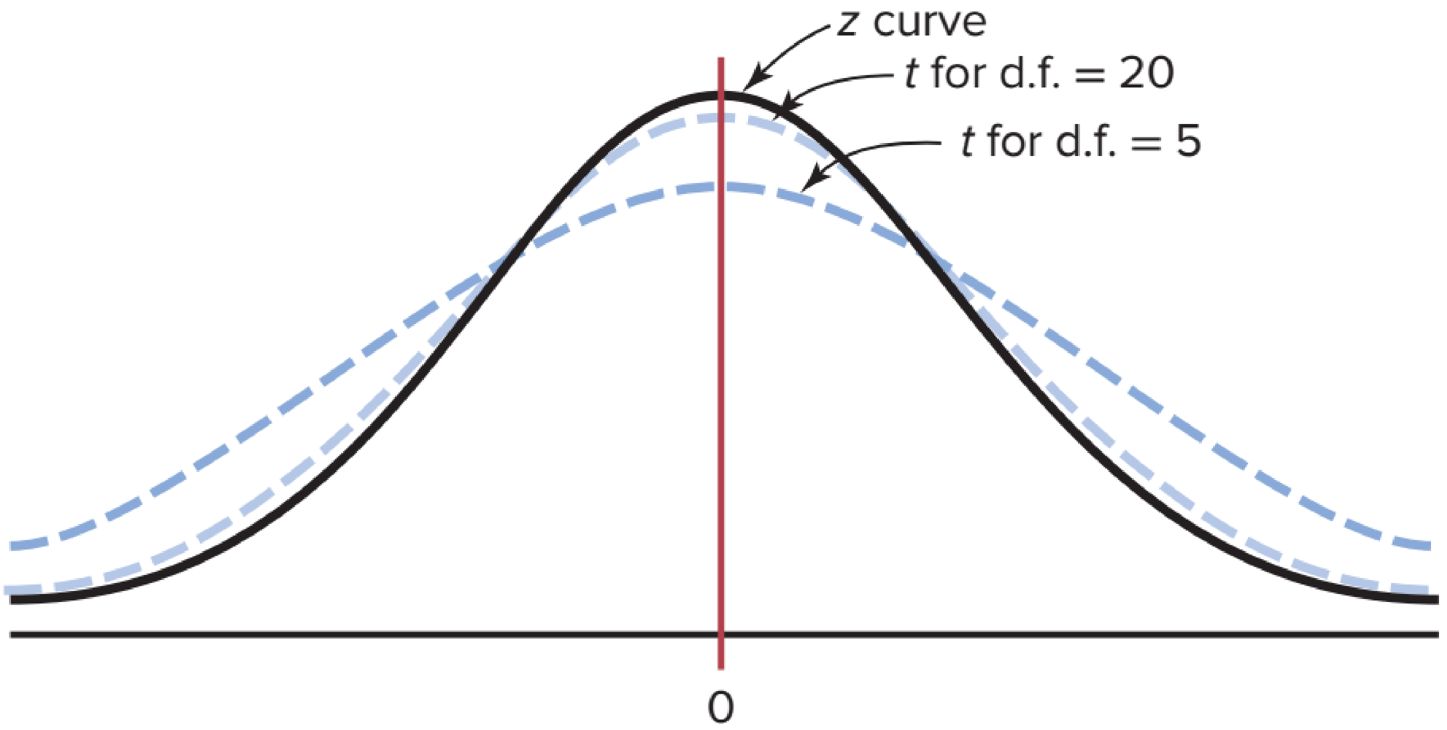
Small-Sample Confidence Intervals for a Population Mean (p.346)

When the value of σ **is not known**, it must be estimated by using *s*, the standard deviation of the sample.

When *s* is used, especially when the sample size is small (**less than 30**), critical values greater than the values for $z_{\alpha/2}$ are used in confidence intervals in order to keep the interval at a given level, such as the 95%.

These values are taken from the **Student *t* distribution**, most often called the ***t* distribution**.

Characteristics of the *t* Distribution



Characteristics of the *t* Distribution…

The *t* distribution is **similar** to the standard normal distribution in these ways:

1. It is bell-shaped.
2. It is symmetric about the mean.
3. The mean, median, and mode are equal to 0 and are located at the center of the distribution.
4. The curve never touches the *x* axis.

Characteristics of the *t* Distribution…

The *t* distribution **differs** from the standard normal distribution in the following ways:

1. The variance is greater than 1.
2. The *t* distribution is actually a family of curves based on the concept of degrees of freedom, which is related to sample size.
3. As the sample size increases, the *t* distribution approaches the standard normal distribution.

Degrees of Freedom

- The symbol d.f. will be used for **degrees of freedom**.
- The degrees of freedom for a confidence interval for the mean are found by subtracting 1 from the sample size. That is, **d.f. = *n* - *1*.**
- *Note:* For some statistical tests used later in this course, the degrees of freedom are not equal to *n - 1.*

Degrees of Freedom (p.347)

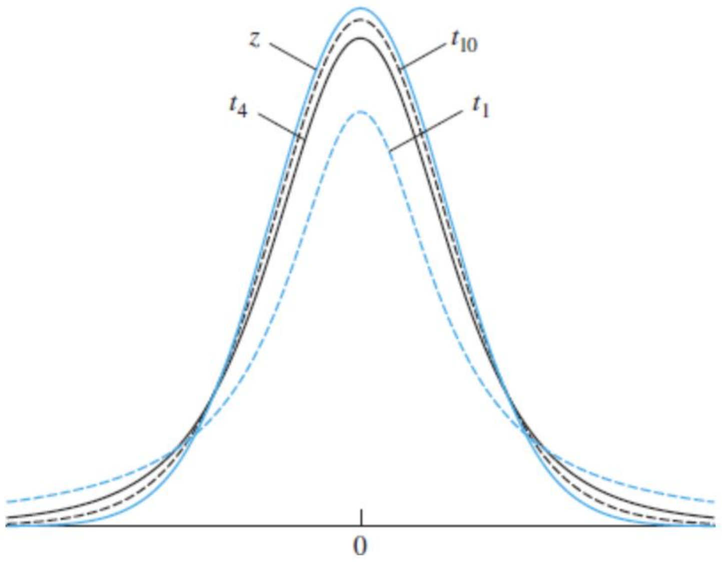
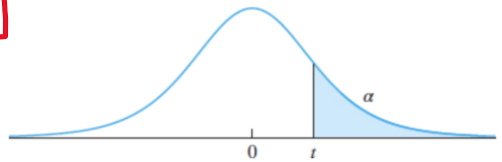


FIGURE 5.9 Plots of the probability density function of the Student's t curve for various degrees of freedom. The normal curve with mean 0 and variance 1 (z curve) is plotted for comparison. The t curves are more spread out than the normal, but the amount of extra spread decreases as the number of degrees of freedom increases.

TABLE A.3 Upper percentage points for the Student's t distribution



v	α 0.40	0.25	0.10	0.05	0.025	0.01	0.005	0.001	0.0005
1	0.325	1.000	3.078	6.314	12.706	31.821	63.657	318.309	636.619
2	0.289	0.816	1.886	2.920	4.303	6.965	9.925	22.327	31.599
3	0.277	0.765	1.638	2.353	3.182	4.541	5.841	10.215	12.924
4	0.271	0.741	1.533	2.132	2.776	3.747	4.604	7.173	8.610
5	0.267	0.727	1.476	2.015	2.571	3.365	4.032	5.893	6.869
6	0.265	0.718	1.440	1.943	2.447	3.143	3.707	5.208	5.959
7	0.263	0.711	1.415	1.895	2.365	2.998	3.499	4.785	5.408
8	0.262	0.706	1.397	1.860	2.306	2.896	3.355	4.501	5.041
9	0.261	0.703	1.383	1.833	2.262	2.821	3.250	4.297	4.781
10	0.260	0.700	1.372	1.812	2.228	2.764	3.169	4.144	4.587
11	0.260	0.697	1.363	1.796	2.201	2.718	3.106	4.025	4.437
12	0.259	0.695	1.356	1.782	2.179	2.681	3.055	3.930	4.318
13	0.259	0.694	1.350	1.771	2.160	2.650	3.012	3.852	4.221
14	0.258	0.692	1.345	1.761	2.145	2.624	2.977	3.787	4.140
15	0.258	0.691	1.341	1.753	2.131	2.602	2.947	3.733	4.073
16	0.258	0.690	1.337	1.746	2.120	2.583	2.921	3.686	4.015
17	0.257	0.689	1.333	1.740	2.110	2.567	2.898	3.646	3.965
18	0.257	0.688	1.330	1.734	2.101	2.552	2.878	3.610	3.922
19	0.257	0.688	1.328	1.729	2.093	2.539	2.861	3.579	3.883
20	0.257	0.687	1.325	1.725	2.086	2.528	2.845	3.552	3.850
21	0.257	0.686	1.323	1.721	2.080	2.518	2.831	3.527	3.819
22	0.256	0.686	1.321	1.717	2.074	2.508	2.819	3.505	3.792
23	0.256	0.685	1.319	1.714	2.069	2.500	2.807	3.485	3.768
24	0.256	0.685	1.318	1.711	2.064	2.492	2.797	3.467	3.745
25	0.256	0.684	1.316	1.708	2.060	2.485	2.787	3.450	3.725
26	0.256	0.684	1.315	1.706	2.056	2.479	2.779	3.435	3.707
27	0.256	0.684	1.314	1.703	2.052	2.473	2.771	3.421	3.690
28	0.256	0.683	1.313	1.701	2.048	2.467	2.763	3.408	3.674
29	0.256	0.683	1.311	1.699	2.045	2.462	2.756	3.396	3.659
30	0.256	0.683	1.310	1.697	2.042	2.457	2.750	3.385	3.646
35	0.255	0.682	1.306	1.690	2.030	2.438	2.724	3.340	3.591
40	0.255	0.681	1.303	1.684	2.021	2.423	2.704	3.307	3.551
60	0.254	0.679	1.296	1.671	2.000	2.390	2.660	3.232	3.460
120	0.254	0.677	1.289	1.658	1.980	2.358	2.617	3.160	3.373
∞	0.253	0.674	1.282	1.645	1.960	2.326	2.576	3.090	3.291

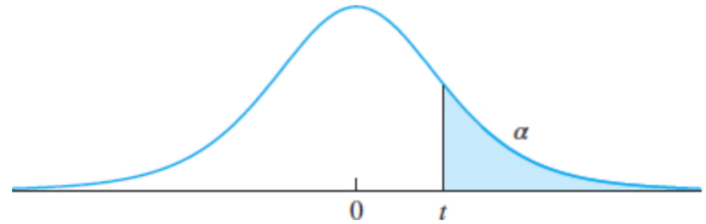
Example: Using Table A.3

Find the $t_{\alpha/2}$ value for a 95% confidence interval when the sample size is 22.

Degrees of freedom are **v = 21**.

$\boldsymbol{t_{0.025} = 2.080}$

TABLE A.3 Upper percentage points for the Student's *t* distribution



	α						
v	**0.40**	**0.25**	**0.10**	**0.05**	**0.025**	**0.01**	**0.005**
1	0.325	1.000	3.078	6.314	12.706	31.821	63.657
2	0.289	0.816	1.886	2.920	4.303	6.965	9.925
3	0.277	0.765	1.638	2.353	3.182	4.541	5.841
4	0.271	0.741	1.533	2.132	2.776	3.747	4.604
5	0.267	0.727	1.476	2.015	2.571	3.365	4.032
6	0.265	0.718	1.440	1.943	2.447	3.143	3.707
7	0.263	0.711	1.415	1.895	2.365	2.998	3.499
8	0.262	0.706	1.397	1.860	2.306	2.896	3.355
9	0.261	0.703	1.383	1.833	2.262	2.821	3.250
10	0.260	0.700	1.372	1.812	2.228	2.764	3.169
11	0.260	0.697	1.363	1.796	2.201	2.718	3.106
12	0.259	0.695	1.356	1.782	2.179	2.681	3.055
13	0.259	0.694	1.350	1.771	2.160	2.650	3.012
14	0.258	0.692	1.345	1.761	2.145	2.624	2.977
15	0.258	0.691	1.341	1.753	2.131	2.602	2.947
16	0.258	0.690	1.337	1.746	2.120	2.583	2.921
17	0.257	0.689	1.333	1.740	2.110	2.567	2.898
18	0.257	0.688	1.330	1.734	2.101	2.552	2.878
19	0.257	0.688	1.328	1.729	2.093	2.539	2.861
20	0.257	0.687	1.325	1.725	2.086	2.528	2.845
21	0.257	0.686	1.323	1.721	2.080	2.518	2.831
22	0.256	0.686	1.321	1.717	2.074	2.508	2.819
23	0.256	0.685	1.319	1.714	2.069	2.500	2.807

Example 5.15

A random sample of size 10 is to be drawn from a normal distribution with mean 4. The Student's t statistic $t = (\overline{X} - 4)/(s/\sqrt{10})$ is to be computed. What is the probability that $t > 1.833$?

Solution

This t statistic has $10 - 1 = 9$ degrees of freedom. From the t table, $P(t > 1.833) = 0.05$. See Figure 5.10.

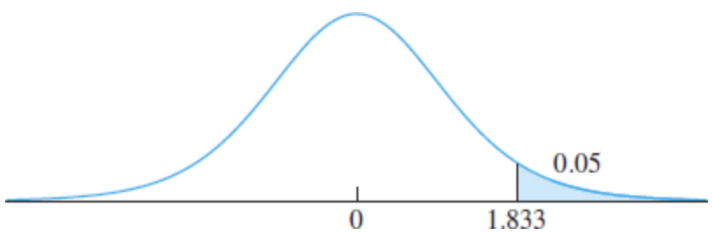
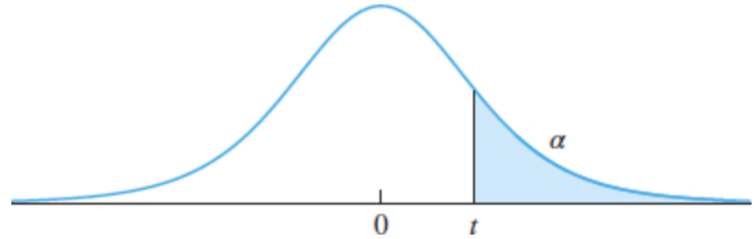


FIGURE 5.10 Solution to Example 5.15.

TABLE A.3 Upper percentage points for the Student's *t* distribution



v \ α	0.40	0.25	0.10	0.05	0.025	0.01	0.005
1	0.325	1.000	3.078	6.314	12.706	31.821	63.657
2	0.289	0.816	1.886	2.920	4.303	6.965	9.925
3	0.277	0.765	1.638	2.353	3.182	4.541	5.841
4	0.271	0.741	1.533	2.132	2.776	3.747	4.604
5	0.267	0.727	1.476	2.015	2.571	3.365	4.032
6	0.265	0.718	1.440	1.943	2.447	3.143	3.707
7	0.263	0.711	1.415	1.895	2.365	2.998	3.499
8	0.262	0.706	1.397	1.860	2.306	2.896	3.355
9	0.261	0.703	1.383	1.833	2.262	2.821	3.250
10	0.260	0.700	1.372	1.812	2.228	2.764	3.169
11	0.260	0.697	1.363	1.796	2.201	2.718	3.106
12	0.259	0.695	1.356	1.782	2.179	2.681	3.055
13	0.259	0.694	1.350	1.771	2.160	2.650	3.012
14	0.258	0.692	1.345	1.761	2.145	2.624	2.977
15	0.258	0.691	1.341	1.753	2.131	2.602	2.947

Example 5.16

Refer to Example 5.15. Find $P(t > 1.5)$.

Solution

Looking across the row corresponding to 9 degrees of freedom, we see that the t table does not list the value 1.5. We find that $P(t > 1.383) = 0.10$ and $P(t > 1.833) = 0.05$. We conclude that $0.05 < P(t > 1.5) < 0.10$. See Figure 5.11. If a more precise result were required, linear interpolation could be used as follows:

	α						
v	**0.40**	**0.25**	**0.10**	**0.05**	**0.025**	**0.01**	**0.005**
1	0.325	1.000	3.078	6.314	12.706	31.821	63.657
2	0.289	0.816	1.886	2.920	4.303	6.965	9.925
3	0.277	0.765	1.638	2.353	3.182	4.541	5.841
4	0.271	0.741	1.533	2.132	2.776	3.747	4.604
5	0.267	0.727	1.476	2.015	2.571	3.365	4.032
6	0.265	0.718	1.440	1.943	2.447	3.143	3.707
7	0.263	0.711	1.415	1.895	2.365	2.998	3.499
8	0.262	0.706	1.397	1.860	2.306	2.896	3.355
9	0.261	0.703	1.383	1.833	2.262	2.821	3.250
10	0.260	0.700	1.372	1.812	2.228	2.764	3.169
11	0.260	0.697	1.363	1.796	2.201	2.718	3.106
12	0.259	0.695	1.356	1.782	2.179	2.681	3.055
13	0.259	0.694	1.350	1.771	2.160	2.650	3.012
14	0.258	0.692	1.345	1.761	2.145	2.624	2.977
15	0.258	0.691	1.341	1.753	2.131	2.602	2.947

Example 5.16

Refer to Example 5.15. Find $P(t > 1.5)$.

$$P(t > 1.5) \approx 0.10 - \frac{1.5 - 1.383}{1.833 - 1.383}(0.10 - 0.05) = 0.0870$$

A computer package gives the answer correct to three significant digits as 0.0839.

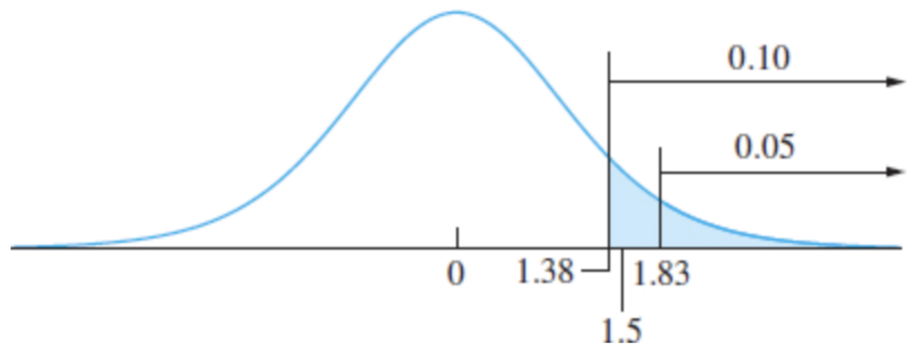


FIGURE 5.11 Solution to Example 5.16.

Formula for a Specific Confidence Interval for the Mean When σ Is Unknown and ***n* < 30**

$$\bar{X} - t_{\alpha/2}\left(\frac{s}{\sqrt{n}}\right) < \mu < \bar{X} + t_{\alpha/2}\left(\frac{s}{\sqrt{n}}\right)$$

The degrees of freedom are ***n* - 1**.

Small-Sample Confidence Intervals for a Population Mean (p.346)

Summary

Let $X_1, \ldots, X_n$ be a *small* (e.g., $n < 30$) sample from a *normal* population with mean μ. Then the quantity

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

has a Student's t distribution with $n - 1$ degrees of freedom, denoted t_{n-1}.

When n is large, the distribution of the quantity $(\overline{X} - \mu)/(s/\sqrt{n})$ is very close to normal, so the normal curve can be used, rather than the Student's t.

Example: Home Fires by Candles

The data represent a random sample of the number of home fires started by candles for the past several years. Find the 99% confidence interval for the mean number of home fires started by candles each year.

5460 5900 6090 6310 7160 8440 9930

Step 1: Find the mean and standard deviation. The mean $\bar{X}$ = 7041.4 and standard deviation s = 1610.3

Step 2: Find $t_{\alpha/2}$ in Table A.3. The confidence level is 99%, and the degrees of freedom equal to 6

$t_{0.005} = 3.707$

Example: Home Fires by Candles…

	α						
v	**0.40**	**0.25**	**0.10**	**0.05**	**0.025**	**0.01**	**0.005**
1	0.325	1.000	3.078	6.314	12.706	31.821	63.657
2	0.289	0.816	1.886	2.920	4.303	6.965	9.925
3	0.277	0.765	1.638	2.353	3.182	4.541	5.841
4	0.271	0.741	1.533	2.132	2.776	3.747	4.604
5	0.267	0.727	1.476	2.015	2.571	3.365	4.032
6	0.265	0.718	1.440	1.943	2.447	3.143	3.707
7	0.263	0.711	1.415	1.895	2.365	2.998	3.499
8	0.262	0.706	1.397	1.860	2.306	2.896	3.355
9	0.261	0.703	1.383	1.833	2.262	2.821	3.250
10	0.260	0.700	1.372	1.812	2.228	2.764	3.169
11	0.260	0.697	1.363	1.796	2.201	2.718	3.106
12	0.259	0.695	1.356	1.782	2.179	2.681	3.055
13	0.259	0.694	1.350	1.771	2.160	2.650	3.012
14	0.258	0.692	1.345	1.761	2.145	2.624	2.977
15	0.258	0.691	1.341	1.753	2.131	2.602	2.947

Example: Home Fires by Candles…

Step 3: Substitute in the formula.

$$\bar{X} - t_{\alpha/2}\left(\frac{s}{\sqrt{n}}\right) < \mu < \bar{X} + t_{\alpha/2}\left(\frac{s}{\sqrt{n}}\right)$$

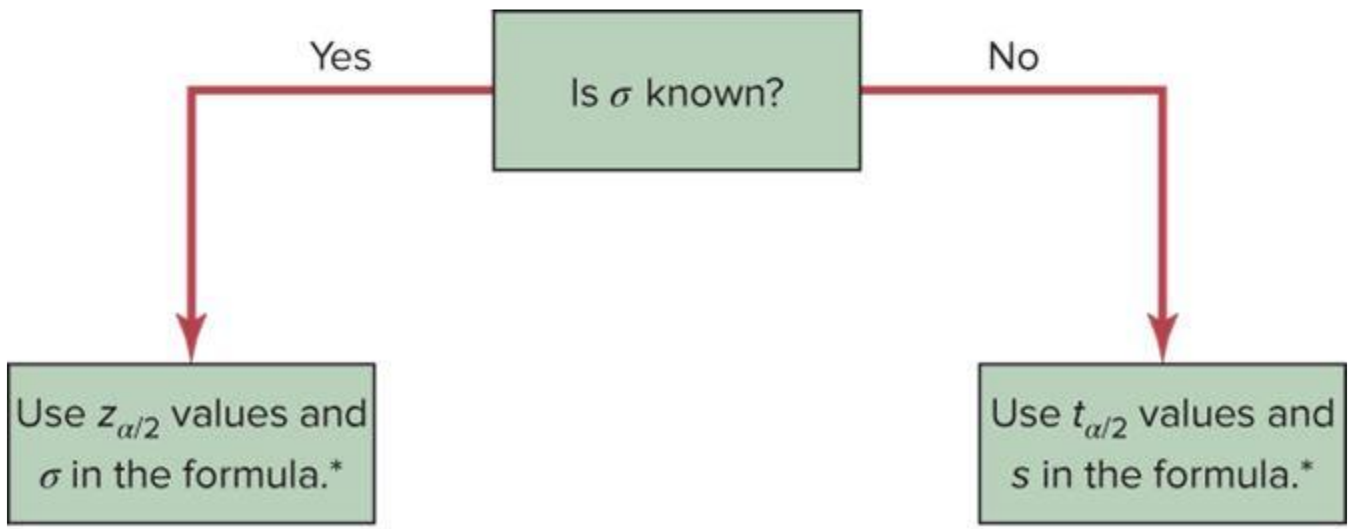
$$7041.4 - 3.707\left(\frac{1610.3}{\sqrt{7}}\right) < \mu < 7041.4 - 3.707\left(\frac{1610.3}{\sqrt{7}}\right)$$

$$7041.4 - 2256.2 < \mu < 7041.4 + 2256.2$$

$$4785.2 < \mu < 9297.6$$

One can be 99% confident that the population mean number of home fires started by candles each year is between 4785.2 and 9297.6, based on a sample of home fires occurring over a period of 7 years.

When to use the *z* or *t* distribution?



*If $n < 30$, the variable must be normally distributed.

When *n* is large (generally n>30), *t*-distribution is very close to *z*-distribution, so it is acceptable to use z-distribution can be used, even if σ is unknown.