# Chapter 8 Confidence Intervals

# Review and Outlook

- Sample mean is the point estimate of the population mean and the sample proportion is the point estimate of the population proportion
- In general, although a point estimate is a reasonable one-number estimate of a population parameter (mean, proportion, or the like), the point estimate will not—unless we are extremely lucky—equal the true value of the population parameter
- Use a confidence interval to estimate a population parameter
- A confidence interval for a population parameter is an interval, or range of numbers, constructed around the point estimate so that we are very sure, or confident, that the true value of the population parameter is inside the interval

# Outline

- 8.1 $z$-Based Confidence Intervals for a Population Mean: $\sigma$ Known
- 8.2 $t$-Based Confidence Intervals for a Population Mean: $\sigma$ Unknown
- 8.3 Sample Size Determination
- 8.4 Confidence Intervals for a Population Proportion

# Outline

- 8.1 *z*-Based Confidence Intervals for a Population Mean: $\sigma$ Known

# Section 8.1 z-Based Confidence Intervals for a Population Mean: $\sigma$ Known

- The starting point is the sampling distribution of the sample mean
  - Recall from Chapter 7 that if a population is normally distributed with mean $\mu$ and standard deviation $\sigma$, then the sampling distribution of $\bar{x}$ is normal with mean $\mu_{\bar{x}} = \mu$ and standard deviation

$$\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}}$$

  - Use a normal curve as a model of the sampling distribution of the sample mean
    - Exactly, because the population is normal
    - Approximately, by the Central Limit Theorem for large-size samples

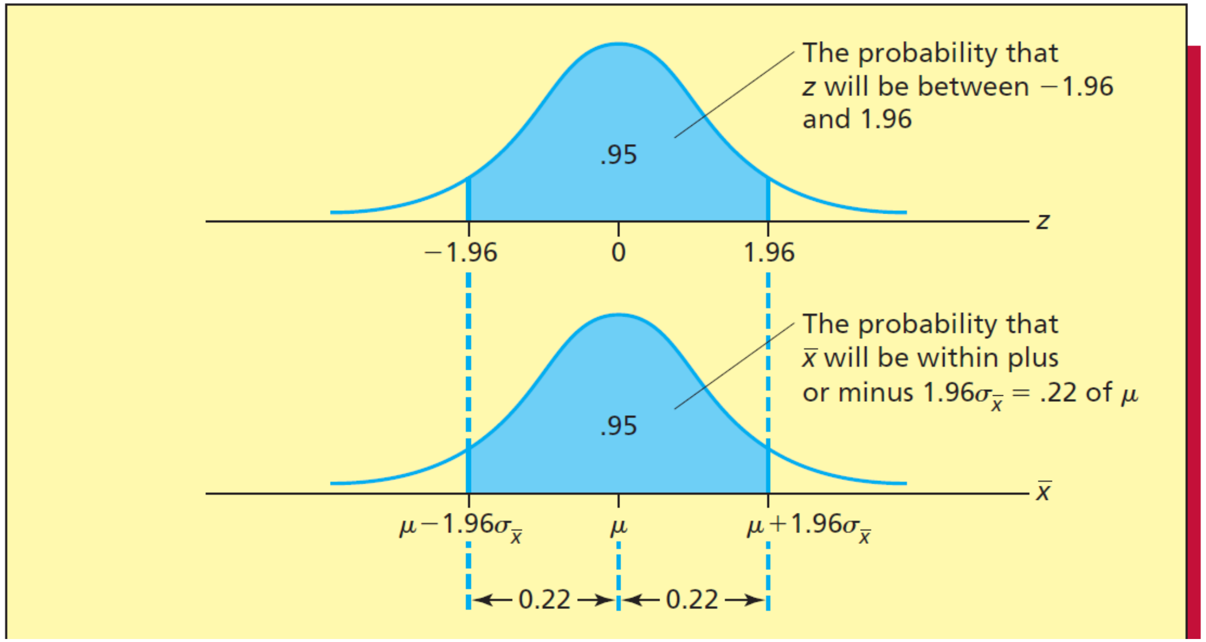
# Section 8.1 Confidence interval

- Recall the empirical rule, so…
  - *68.26%* of all possible sample means are within one standard deviation of the population mean
  - *95.44%* of all possible sample means are within two standard deviations of the population mean
  - *99.73%* of all possible sample means are within three standard deviations of the population mean

# Section 8.1 Confidence interval

A **confidence interval** for a population mean is an interval constructed around the sample mean so that we are reasonably sure, or confident, that this interval contains the population mean. Any confidence interval for a population mean is based on what is called a **confidence level.** This confidence level is a percentage (for example, 95 percent or 99 percent) that expresses how confident we are that the confidence interval contains the population mean.

FIGURE The Sampling Distribution of the Sample Mean



Example 8.1

# The Car Mileage Case

- Recall that the population of car mileages is normally distributed with mean μ and standard deviation σ = 0.8 mpg
  - Note that μ is unknown and is to be estimated
- Taking samples of size $n$ = 5, the sampling distribution of sample mean mileages $\bar{x}$ is normal with mean $\mu_{\bar{x}} = \mu$ (which is also unknown) and standard deviation

$$\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}} = \frac{0.8}{\sqrt{5}} = 0.35777$$

- The probability is 0.9544 that $\bar{x}$ will be within plus or minus $2\sigma_{\bar{x}}$ = 2 • 0.35777 = 0.7155 of μ

# The Car Mileage Case #2

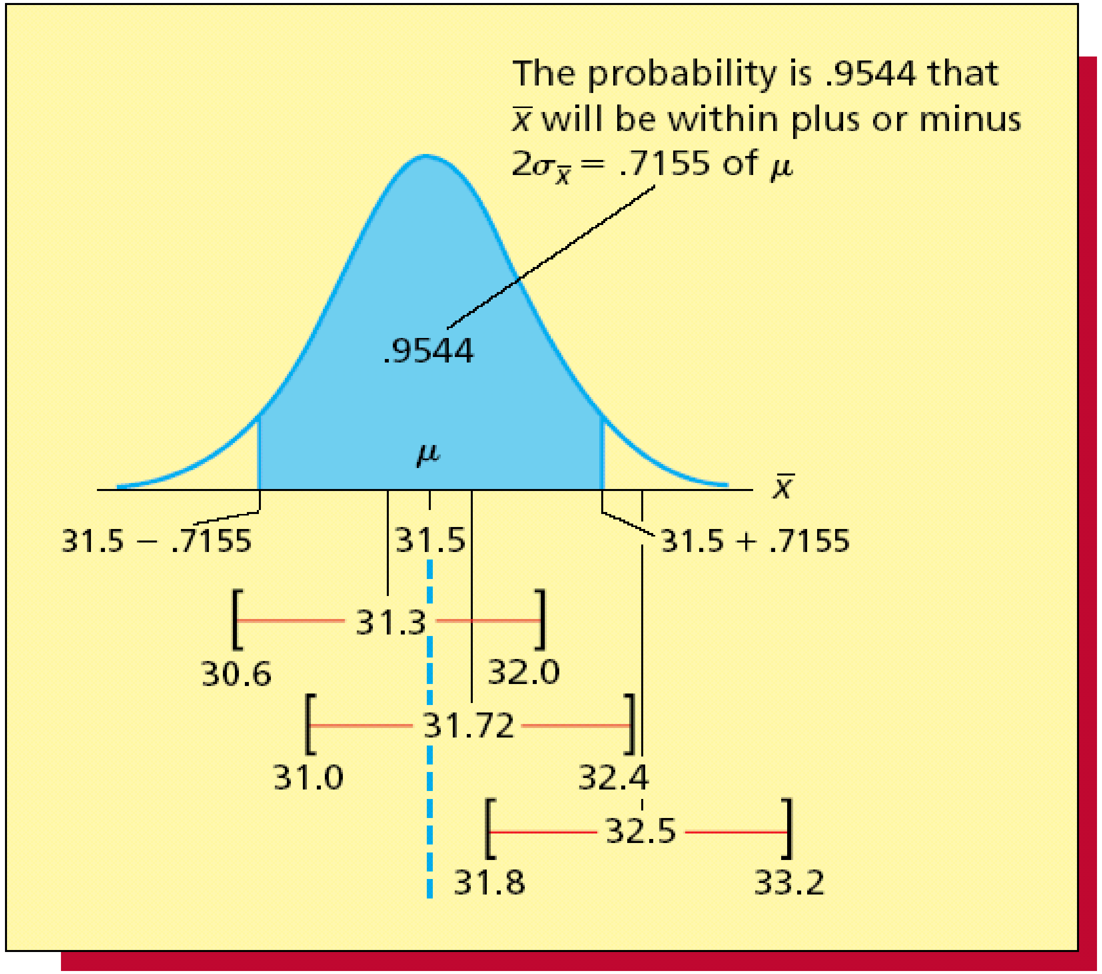
- The sample mean is within $\pm 0.7155$ of μ

$$P(\mu - 0.7155 \leq \bar{x} \leq \mu + 0.7155) = 0.9544$$

- Then there is a 0.9544 probability that $\bar{x}$ will be a value so that interval $[\bar{x} \pm 0.7115]$ contains μ
  - The interval $[\bar{x} \pm 0.7115]$ is referred to as the 95.44% **confidence interval** for μ
  - 0.7115 is called the **margin of error when estimating μ by** $\bar{x}$

# The Car Mileage Case #3

95.44% Confidence Intervals for $\mu$




- Three intervals with different sample means
- Two contain $\mu$
- One does not

- According to the 95.44% confidence interval, we know…

There is 95.44% probability that the sample mean that is calculated is such that the interval

[$\bar{x} \pm$ 0.7155] will contain the actual (but unknown) population mean μ

- In other words, of all possible sample means, 95.44% of all the corresponding intervals will contain the population mean μ

- Note that there is a 4.56% probability that the interval does not contain μ

# Generalizing

- In the example, we found the probability that μ is contained in an interval of integer multiples of $\sigma_{\bar{x}}$
- More usual to specify the (integer) probability and find the corresponding number of $\sigma_{\bar{x}}$
- The probability that the confidence interval will ***not*** contain the population mean μ is denoted by α
  - In the example, α = 0.0456

# Generalizing Continued

- The probability that the confidence interval will contain the population mean $\mu$ is denoted by $1 - \alpha$
  - $1 - \alpha$ is referred to as the confidence coefficient
  - $(1 - \alpha) \times 100\%$ is called the **confidence level**
  - In the example, $1 - \alpha = 0.9544$
- Usual to use two decimal point probabilities for $1 - \alpha$
  - Here, focus on $1 - \alpha = 0.95$ *or* $0.99$

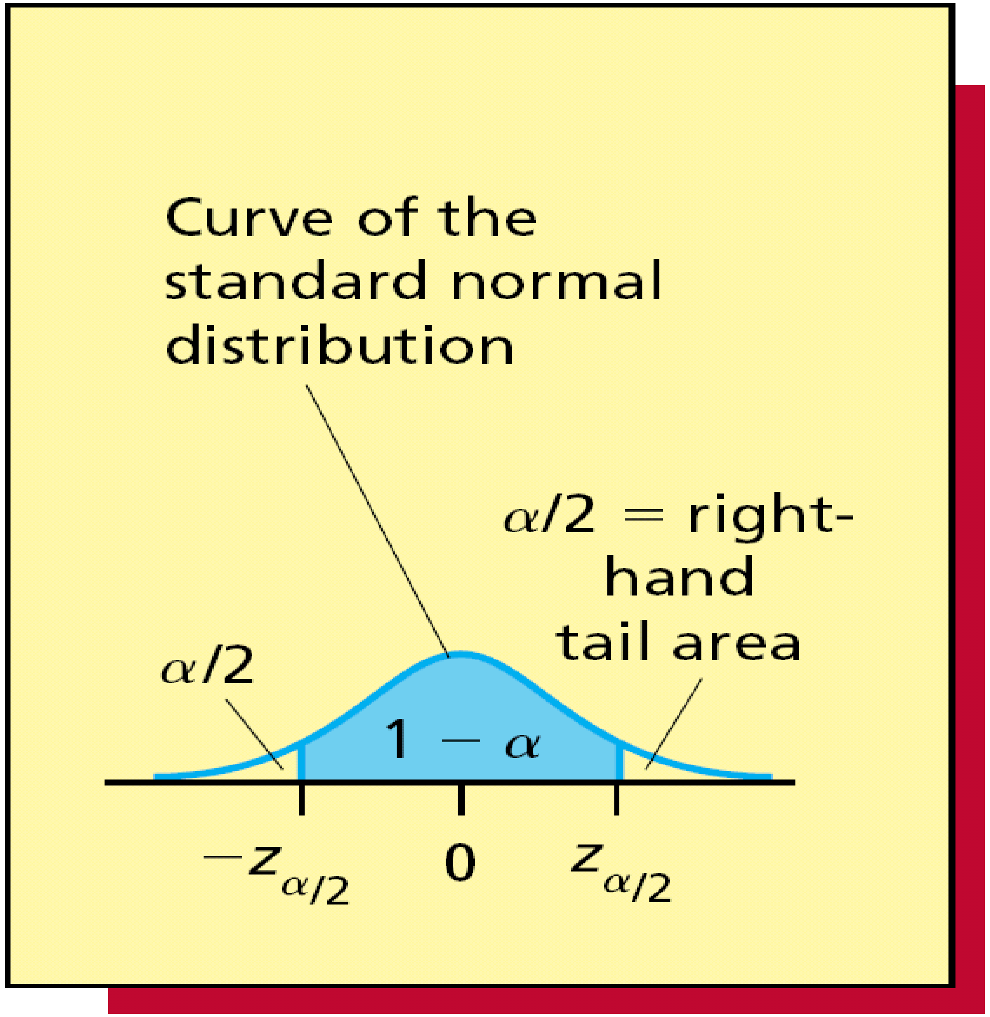
# General Confidence Interval

- In general, the probability is 1 – $\alpha$ that the population mean $\mu$ is contained in the interval

$$\left[\bar{x} \pm z_{\alpha/2}\sigma_{\bar{x}}\right] = \left[\bar{x} \pm z_{\alpha/2}\frac{\sigma}{\sqrt{n}}\right] = \left[\bar{x} - z_{\alpha/2}\frac{\sigma}{\sqrt{n}}, \bar{x} + z_{\alpha/2}\frac{\sigma}{\sqrt{n}}\right]$$

  - The normal point $z_{\alpha/2}$ gives a right hand tail area under the standard normal curve equal to $\alpha/2$
  - The normal point - $z_{\alpha/2}$ gives a left hand tail area under the standard normal curve equal to $\alpha/2$
  - The area under the standard normal curve between $-z_{\alpha/2}$ and $z_{\alpha/2}$ is 1 – $\alpha$
  - $z_{\alpha/2}\sigma_{\bar{x}}$ is the margin of error.

# A Confidence Interval for a Population Mean μ: σ Known




$$\left[\bar{x} \pm z_{\alpha/2}\sigma_{\bar{x}}\right] = \left[\bar{x} \pm z_{\alpha/2}\frac{\sigma}{\sqrt{n}}\right] =$$

$$\left[\bar{x} - z_{\alpha/2}\frac{\sigma}{\sqrt{n}}, \bar{x} + z_{\alpha/2}\frac{\sigma}{\sqrt{n}}\right]$$

## z-Based Confidence Intervals for a Mean with σ Known

- If a population has standard deviation σ (known),
- and if the population is normal or if sample size is large ($n \geq 30$), then …
- … a (1–α)*100%* **confidence interval for μ** is

$$\bar{x} \pm z_{\alpha/2}\frac{\sigma}{\sqrt{n}} = \left[\bar{x} - z_{\alpha/2}\frac{\sigma}{\sqrt{n}}, \bar{x} + z_{\alpha/2}\frac{\sigma}{\sqrt{n}}\right]$$

# 95% Confidence Level

- For a 95% confidence level,

  $1 - \alpha = 0.95 \qquad \alpha = 0.05 \quad \alpha/2 = 0.025$

- For 95% confidence, need the normal point $z_{0.025}$
  - The area under the standard normal curve between $-z_{0.025}$ and $z_{0.025}$ is 0.95
  - Then the area under the standard normal curve between 0 and $z_{0.025}$ is 0.475
  - From the standard normal table, the area is 0.475 for $z = 1.96$
  - Then $z_{0.025} = 1.96$

# 95% Confidence Interval

The 95% confidence interval is

$$[\bar{x} \pm z_{0.025}\sigma_{\bar{x}}] = \left[\bar{x} \pm 1.96\frac{\sigma}{\sqrt{n}}\right]$$
$$= \left[\bar{x} - 1.96\frac{\sigma}{\sqrt{n}}, \bar{x} + 1.96\frac{\sigma}{\sqrt{n}}\right]$$

# 99% Confidence Interval

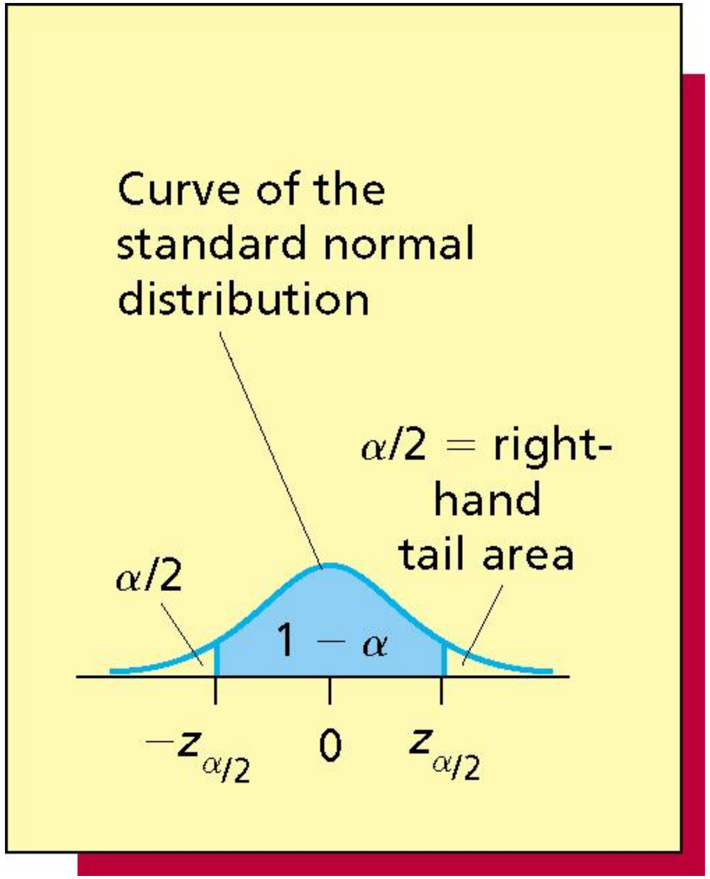
- For *99%* confidence, need the normal point $z_{0.005}$
  - Reading between table entries in the standard normal table, the area is 1-0.005=0.995 for $z_{0.005} = 2.575$
- The *99%* confidence interval is

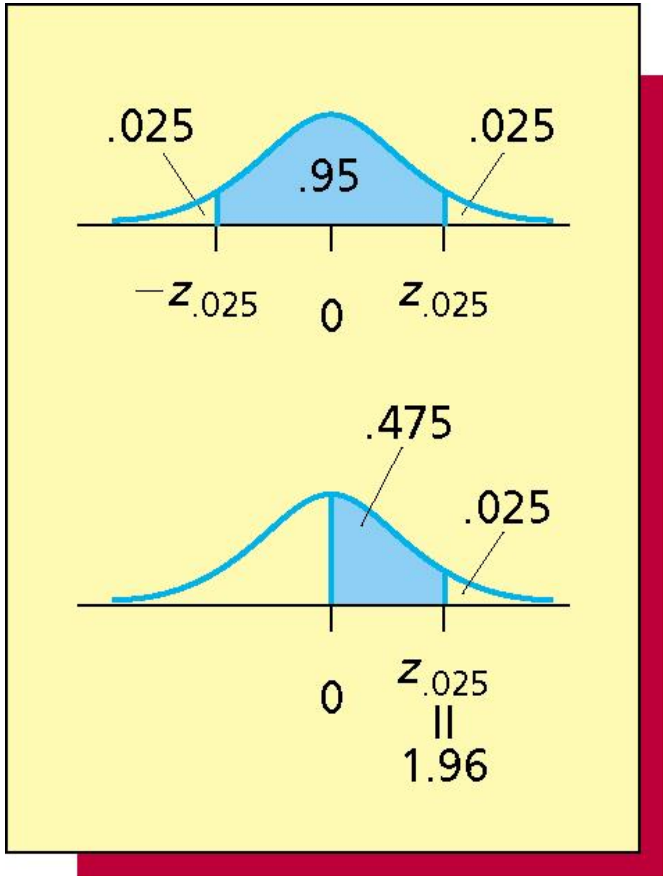
$$
[\bar{x} \pm z_{0.025}\sigma_{\bar{x}}] = \left[\bar{x} \pm 2.575\frac{\sigma}{\sqrt{n}}\right]
$$

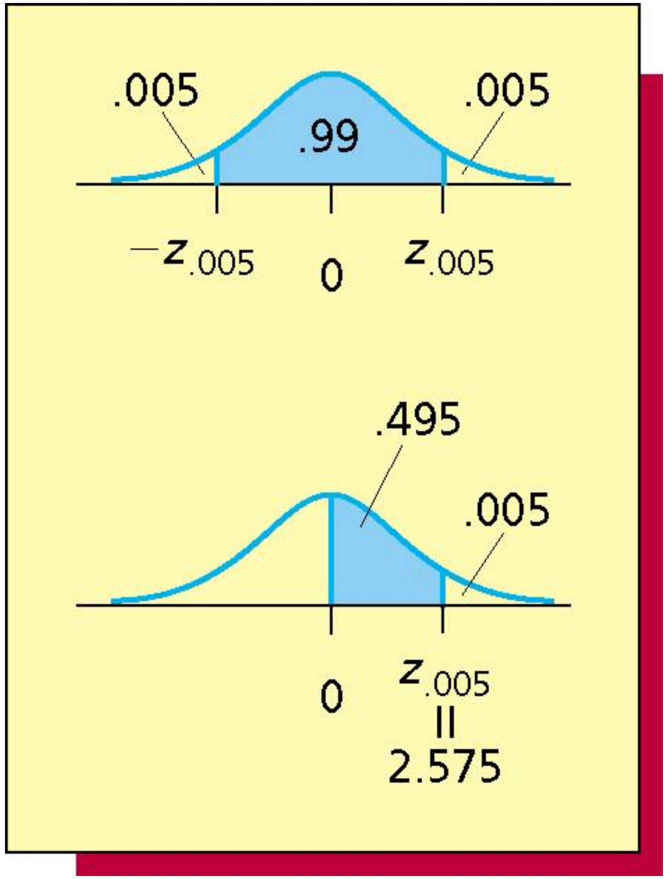
$$
= \left[\bar{x} - 2.575\frac{\sigma}{\sqrt{n}}, \bar{x} + 2.575\frac{\sigma}{\sqrt{n}}\right]
$$

# The Effect of α on Confidence Interval Width

$z_{\alpha/2} = z_{0.025} = 1.96$ $z_{\alpha/2} = z_{0.005} = 2.575$



**Step 1:** As illustrated in Figure 8.3, place a symmetrical area of .95 under the standard normal curve and find the area in the normal curve tails beyond the .95 area. Because the entire area under the standard normal curve is 1, the area in both normal curve tails is $1 - .95 = .05$, and the area in each tail is .025.

**Step 2:** Find the normal point $z_{.025}$ that gives a right-hand tail area under the standard normal curve equal to .025, and find the normal point $-z_{.025}$ that gives a left-hand tail area under the curve equal to .025. As shown in Figure 8.3, the area under the standard normal curve between $-z_{.025}$ and $z_{.025}$ is .95, and the area under this curve to the left of $z_{.025}$ is .975. Looking up a cumulative area of .975 in Table A.3 (see page 790) or in Table 8.1 (which shows a portion of Table A.3), we find that $z_{.025} = 1.96$.

**Step 3:** Form the following 95 percent confidence interval for the population mean:

$$[\bar{x} \pm z_{.025}\sigma_{\bar{x}}] = \left[\bar{x} \pm 1.96\frac{\sigma}{\sqrt{n}}\right]$$

We next start with a confidence coefficient of .99 and find the corresponding 99 percent confidence interval for the population mean:

**Step 1:** As illustrated in Figure 8.4, place a symmetrical area of .99 under the standard normal curve, and find the area in the normal curve tails beyond the .99 area. Because the entire area under the standard normal curve is 1, the area in both normal curve tails is $1 - .99 = .01$, and the area in each tail is .005.

**Step 2:** Find the normal point $z_{.005}$ that gives a right-hand tail area under the standard normal curve equal to .005, and find the normal point $-z_{.005}$ that gives a left-hand tail area under the curve equal to .005. As shown in Figure 8.4, the area under the standard normal curve between $-z_{.005}$ and $z_{.005}$ is .99, and the area under this curve to the left of $z_{.005}$ is .995. Looking up a cumulative area of .995 in Table A.3 (see page 790) or in Table 8.1, we find that $z_{.005} = 2.575$.

**Step 3:** Form the following 99 percent confidence interval for the population mean.

$$[\bar{x} \pm z_{.005}\sigma_{\bar{x}}] = \left[\bar{x} \pm 2.575 \frac{\sigma}{\sqrt{n}}\right]$$

In general, we let $\alpha$ denote the probability that a confidence interval for a population mean will *not* contain the population mean. This implies that $1 - \alpha$ is the probability that the confidence interval will contain the population mean. In order to find a confidence interval for a population mean that is based on a confidence coefficient of $1 - \alpha$ (that is, a **100(1 − $\alpha$) percent confidence interval** for the population mean), we do the following:

**Step 1:** As illustrated in Figure 8.5, place a symmetrical area of $1 - \alpha$ under the standard normal curve, and find the area in the normal curve tails beyond the $1 - \alpha$ area. Because the entire area under the standard normal curve is 1, the combined areas in the normal curve tails are $\alpha$, and the area in each tail is $\alpha/2$.

**Step 2:** Find the normal point $z_{\alpha/2}$ that gives a right-hand tail area under the standard normal curve equal to $\alpha/2$, and find the normal point $-z_{\alpha/2}$ that gives a left-hand tail area under this curve equal to $\alpha/2$. As shown in Figure 8.5, the area under the standard normal curve between $-z_{\alpha/2}$ and $z_{\alpha/2}$ is $(1-\alpha)$, and the area under this curve to the left of $z_{\alpha/2}$ is $1-\alpha/2$. **This implies that we can find $z_{\alpha/2}$ by looking up a cumulative area of $1-\alpha/2$ in Table A.3** (page 790).

**Step 3:** Form the following $100(1 - \alpha)$ percent confidence interval for the population mean.

$$[\bar{x} \pm z_{\alpha/2}\sigma_{\bar{x}}] = \left[\bar{x} \pm z_{\alpha/2}\frac{\sigma}{\sqrt{n}}\right]$$

If all possible samples were used to calculate this interval, then $100(1 - \alpha)$ percent of the resulting intervals would contain the population mean. Moreover, we call **100(1 − $\alpha$) percent** the **confidence level** associated with the confidence interval.

Example 8.2

# The Car Mileage Case

Given: $\bar{x}$ = 31.5531 mpg

σ = 0.8 mpg

$n$ = 49

95% Confidence Interval:

$$\bar{x} \pm z_{0.025}\frac{\sigma}{\sqrt{n}} = 31.5531 \pm 1.96\frac{0.8}{\sqrt{49}}$$
$$= 31.5531 \pm 0.224$$
$$= [31.33,\ 31.78]$$

99% Confidence Interval:

$$\bar{x} \pm z_{0.005}\frac{\sigma}{\sqrt{n}} = 31.5531 \pm 2.575\frac{0.8}{\sqrt{49}}$$
$$= 31.5531 \pm 0.294$$
$$= [31.26,\ 31.85]$$

# The Car Mileage Case

- The 99% confidence interval is slightly wider than the 95% confidence interval
  - The higher the confidence level, the wider the interval
  - According to the 95% confidence interval, we can be 95% confident that the mean mileage is between 31.33 and 31.78 mpg
  - So, we can be 95% confident that, on average, the mean mileage exceeds 31 by at least 0.33 mpg and at most 0.78 mpg

## EXAMPLE 8.1 The e-billing Case: Reducing Mean Bill Payment Time

Recall that a management consulting firm has installed a new computer-based electronic billing system in a Hamilton, Ohio, trucking company. The population mean payment time using the trucking company's old billing system was approximately equal to, but no less than, 39 days. In order to assess whether the population mean payment time, $\mu$, using the new billing system is substantially less than 39 days, the consulting firm will use the sample of $n = 65$ payment times in Table 2.4 to find a 99 percent confidence interval for $\mu$. The mean of the 65 payment times is $\bar{x} = 18.1077$ days, and we will assume that the true value of the population standard deviation $\sigma$ for the new billing system is 4.2 days (as discussed on page 280 of Chapter 7). Then, because we previously showed that the normal point corresponding to 99 percent confidence is $z_{\alpha/2} = z_{.005} = 2.575$, a 99 percent confidence interval for $\mu$ is

99% Confidence Interval:

$$
\begin{aligned}
\bar{x} \pm z_{0.005} \frac{\sigma}{\sqrt{n}} &= 18.1077 \pm 2.575 \frac{4.2}{\sqrt{65}} \\
&= 18.1077 \pm 1.3414 \\
&= [16.8,\ 19.4]
\end{aligned}
$$

Recalling that the mean payment time using the old billing system is 39 days, this interval says that we are 99 percent confident that the population mean payment time using the new billing system is between 16.8 days and 19.4 days. Therefore, we are 99 percent confident that the new billing system reduces the mean payment time by at most 39-16.8=22.2 days and by at least 39-19.4=19.6 days.

# Outline

- **8.2 *t*-Based Confidence Intervals for a Population Mean: $\sigma$ Unknown**

# 8.2 t-Based Confidence Intervals for a Population Mean: σ unknown

If σ is unknown (which is usually the case), we use the sample standard deviation *s* to help construct a confidence interval for μ

- we can construct a confidence interval for μ based on the sampling distribution of

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

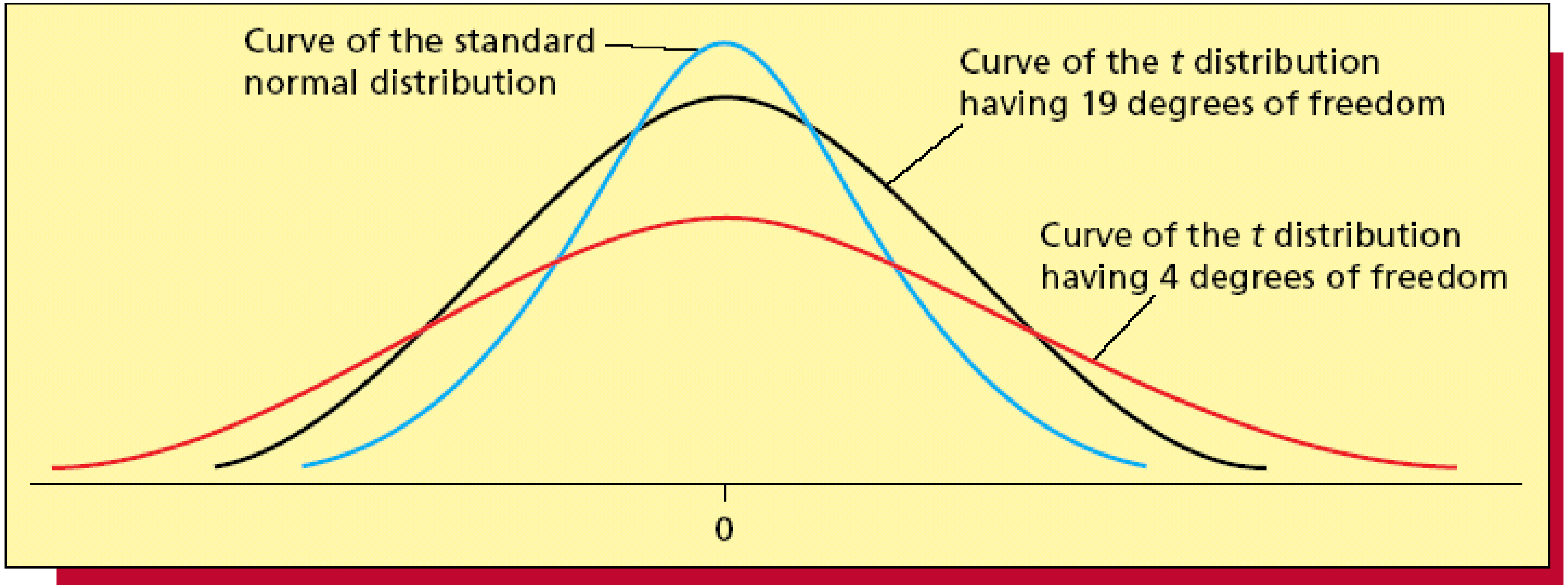
- If the population is normal, then for any sample size *n*, this sampling distribution is called the ***t* distribution**

# The *t* Distribution

- The curve of the *t* distribution is similar to that of the standard normal curve
  - Symmetrical and bell-shaped
  - The *t* distribution is more spread out than the standard normal distribution
  - The spread of the *t* is given by the **number of degrees of freedom**
    - Denoted by $df$
    - For a sample of size *n*, there are one fewer degrees of freedom, that is,

$$df = n - 1$$

# Degrees of Freedom and the t-Distribution




As the number of degrees of freedom increases, the spread of the *t* distribution decreases and the *t* curve approaches the standard normal curve

# The t Distribution and Degrees of Freedom

- For a $t$ distribution with $n - 1$ degrees of freedom,
  - As the sample size $n$ increases, the degrees of freedom also increases
  - As the degrees of freedom increase, the spread of the $t$ curve decreases
  - As the degrees of freedom increases indefinitely, the $t$ curve approaches the standard normal curve
    - If $n \geq 30$, so $df = n - 1 \geq 29$, the $t$ curve is very similar to the standard normal curve

# $t$ and Right Hand Tail Areas

- Use a $t$ point denoted by $t_\alpha$
  - $t_\alpha$ is the point on the horizontal axis under the t curve that gives a right hand tail equal to $\alpha$
  - So the value of $t_\alpha$ in a particular situation depends on the right hand tail area $\alpha$ and the number of degrees of freedom
    - $df = n - 1$

**FIGURE 8.7** An Example of a $t$ Point Giving a Specified Right-Hand Tail Area (This $t$ Point Gives a Right-Hand Tail Area Equal to $\alpha$).

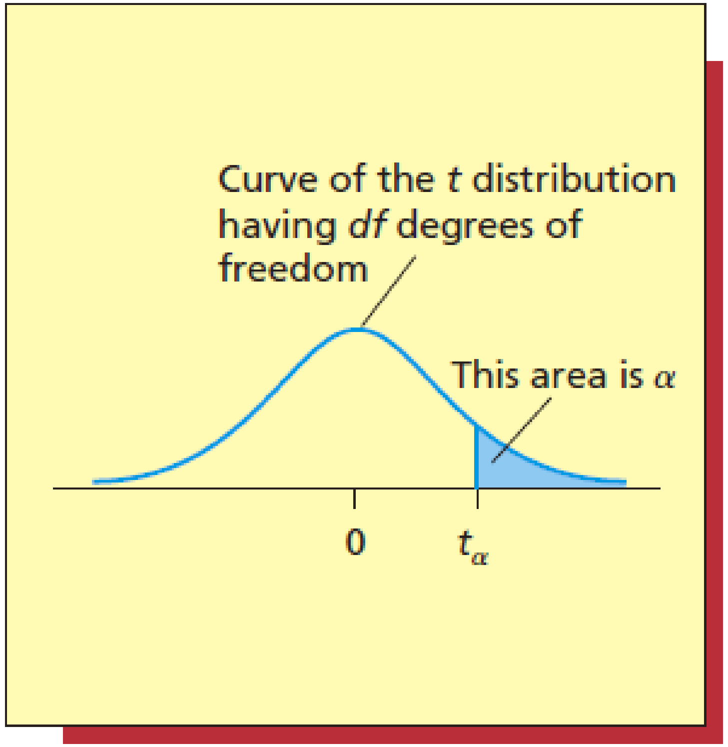
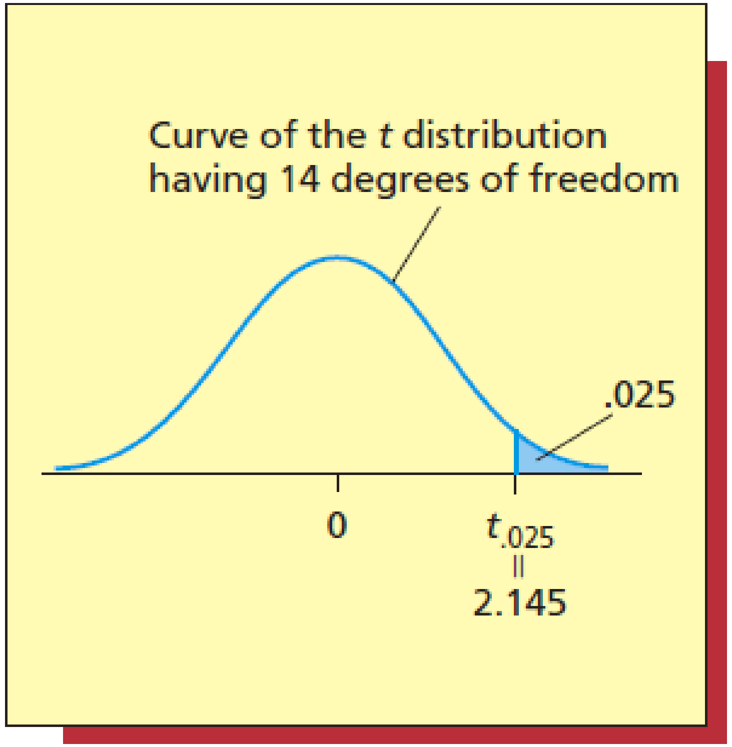



**FIGURE 8.8** The $t$ Point Giving a Right-Hand Tail Area of .025 under the $t$ Curve Having 14 Degrees of Freedom: $t_{.025} = 2.145$



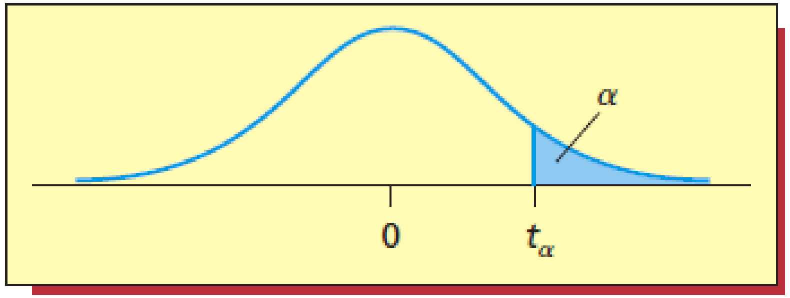
# Using the $t$ Distribution Table

- Rows correspond to the different values of $df=n-1$
  Columns correspond to different values of $\alpha$

- See *t* Distribution Table in Appendix
  The table gives $t$ points for $df$ from 1 to 100, then for $df$ = 120, and $\infty$
  - On the row for $\infty$, the $t$ points are the $z$ points

- Always look at the accompanying figure for guidance on how to use the table

# A $t$ Table

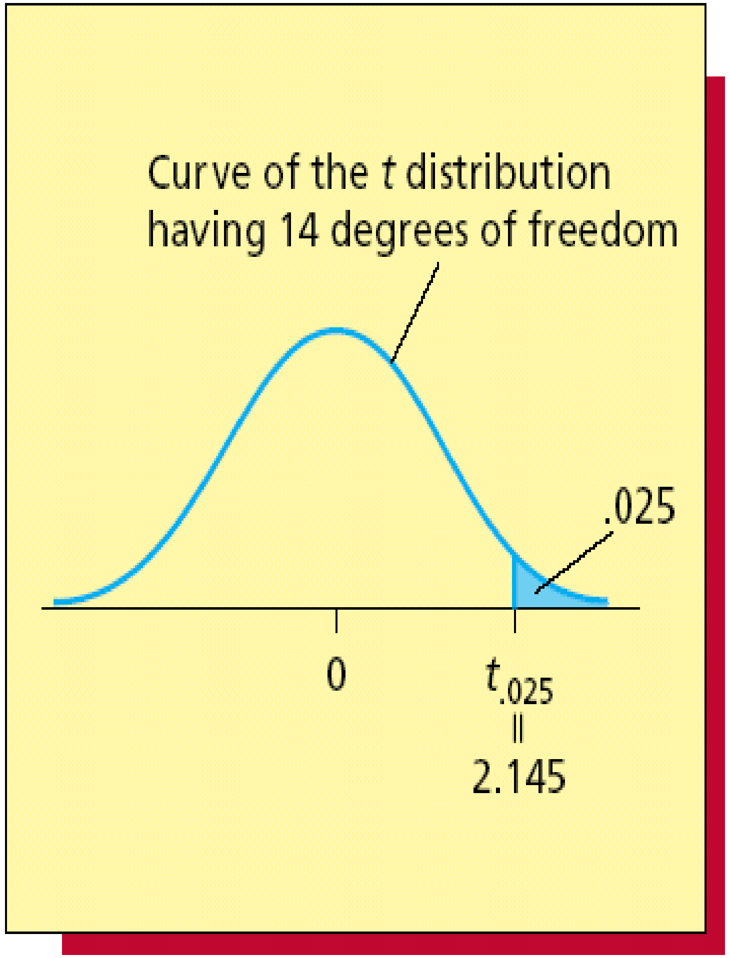



| df | $t_{.100}$ | $t_{.050}$ | $t_{.025}$ | $t_{.01}$ | $t_{.005}$ | $t_{.001}$ | $t_{.0005}$ |
|---|---|---|---|---|---|---|---|
| 1 | 3.078 | 6.314 | 12.706 | 31.821 | 63.657 | 318.31 | 636.62 |
| 2 | 1.886 | 2.920 | 4.303 | 6.965 | 9.925 | 22.326 | 31.598 |
| 3 | 1.638 | 2.353 | 3.182 | 4.541 | 5.841 | 10.213 | 12.924 |
| 4 | 1.533 | 2.132 | 2.776 | 3.747 | 4.604 | 7.173 | 8.610 |
| 5 | 1.476 | 2.015 | 2.571 | 3.365 | 4.032 | 5.893 | 6.869 |
| 6 | 1.440 | 1.943 | 2.447 | 3.143 | 3.707 | 5.208 | 5.959 |
| 7 | 1.415 | 1.895 | 2.365 | 2.998 | 3.499 | 4.785 | 5.408 |
| 8 | 1.397 | 1.860 | 2.306 | 2.896 | 3.355 | 4.501 | 5.041 |
| 9 | 1.383 | 1.833 | 2.262 | 2.821 | 3.250 | 4.297 | 4.781 |
| 10 | 1.372 | 1.812 | 2.228 | 2.764 | 3.169 | 4.144 | 4.587 |
| 11 | 1.363 | 1.796 | 2.201 | 2.718 | 3.106 | 4.025 | 4.437 |
| 12 | 1.356 | 1.782 | 2.179 | 2.681 | 3.055 | 3.930 | 4.318 |
| 13 | 1.350 | 1.771 | 2.160 | 2.650 | 3.012 | 3.852 | 4.221 |
| 14 | 1.345 | 1.761 | 2.145 | 2.624 | 2.977 | 3.787 | 4.140 |
| 15 | 1.341 | 1.753 | 2.131 | 2.602 | 2.947 | 3.733 | 4.073 |
| 16 | 1.337 | 1.746 | 2.120 | 2.583 | 2.921 | 3.686 | 4.015 |
| 17 | 1.333 | 1.740 | 2.110 | 2.567 | 2.898 | 3.646 | 3.965 |
| 18 | 1.330 | 1.734 | 2.101 | 2.552 | 2.878 | 3.610 | 3.922 |
| 19 | 1.328 | 1.729 | 2.093 | 2.539 | 2.861 | 3.579 | 3.883 |
| 20 | 1.325 | 1.725 | 2.086 | 2.528 | 2.845 | 3.552 | 3.850 |

## Example:

Find $t_\alpha$ for a sample of size $n$ = 15 and right hand tail area of 0.025

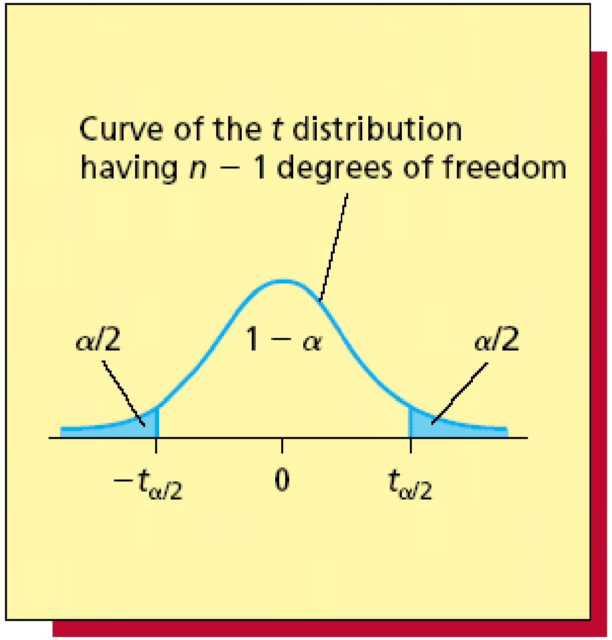
- For $n$ = 15, $df$ = n-1= 14
- $\alpha = 0.025$
  - Note that $\alpha = 0.025$ corresponds to a confidence level of 0.95
- In the $t$ table, along row labeled 14 and under column labeled 0.025, read a table entry of 2.145
- So $t_\alpha = 2.145$



# t-Based Confidence Intervals for a Mean: σ Unknown

If the sampled population is normally distributed with mean $\mu$, then a **(1–α)100% confidence interval for $\mu$** is

$$\bar{x} \pm t_{\alpha/2} \frac{s}{\sqrt{n}}$$




$t_{\alpha/2}$ is the $t$ point giving a right-hand tail area of $\alpha/2$ under the $t$ curve having $n - 1$ degrees of freedom

Example 8.4

# Debt-to-Equity Ratio

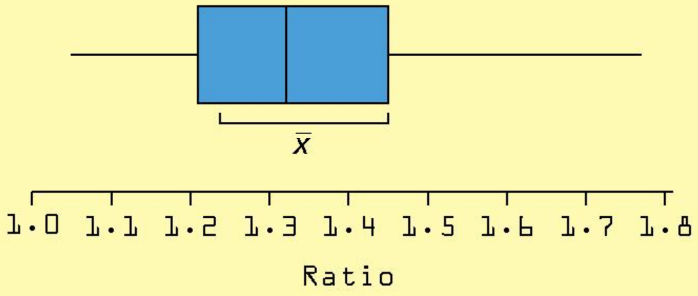
One measure of a company's financial health is its *debt-to-equity ratio*. This quantity is defined to be the ratio of the company's corporate debt to the company's equity. If this ratio is too high, it is one indication of financial instability. For obvious reasons, banks often monitor the financial health of companies to which they have extended commercial loans. Suppose that, in order to reduce risk, a large bank has decided to initiate a policy limiting the mean debt-to-equity ratio for its portfolio of commercial loans to being less than 1.5. In order to estimate the mean debt-to-equity ratio of its (current) commercial loan portfolio, the bank randomly selects a sample of 15 of its commercial loan accounts. Audits of these companies result in the following debt-to-equity ratios:

| | | | | |
|---|---|---|---|---|
| 1.31 | 1.05 | 1.45 | 1.21 | 1.19 |
| 1.78 | 1.37 | 1.41 | 1.22 | 1.11 |
| 1.46 | 1.33 | 1.29 | 1.32 | 1.65 |

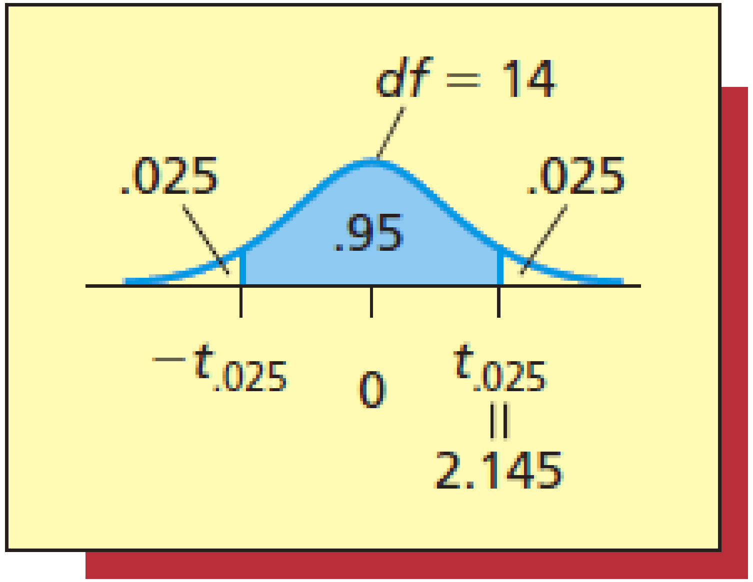
# Example 8.4 Debt-to-Equity Ratio

- Estimate the mean debt-to-equity ratio of the loan portfolio of a bank
- Select a random sample of 15 commercial loan accounts
  - Box plot is given in figure below
- Know: $\bar{x} = 1.34$  $s = 0.192$  $n = 15$
- Want a 95% confidence interval for the ratio
- Assume all ratios are normally distributed but $\sigma$ unknown

| Variable | N | Mean | StDev | SE Mean | 95.0 % CI |
|---|---|---|---|---|---|
| Ratio | 15 | 1.3433 | 0.1921 | 0.0496 | (1.2370, 1.4497) |



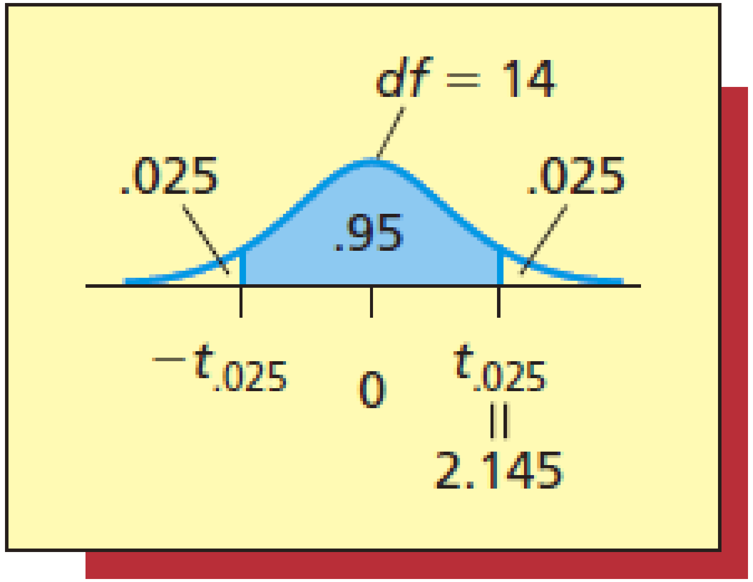
- Have to use the $t$ distribution
- At 95% confidence,
  - $1 - \alpha = 0.95$ so $\alpha = 0.05$ and $\alpha/2 = 0.025$
- For $n = 15$,
  - $df = 15 - 1 = 14$
- Use the $t$ table to find $t_{\alpha/2}$ for $df = 14$
  - $t_{\alpha/2} = t_{0.025} = 2.145$ for $df = 14$
- The 95% confidence interval:




$$\bar{x} \pm t_{0.025}\frac{s}{\sqrt{n}} = 1.343 \pm 2.145\frac{0.192}{\sqrt{15}}$$
$$= 1.343 \pm 0.106$$
$$= [1.237, 1.449]$$

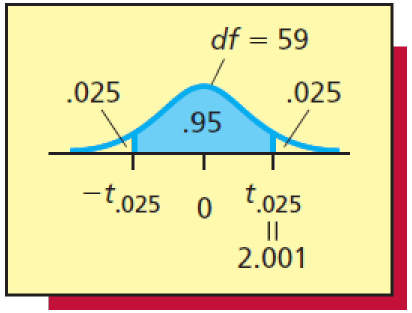
The 95% confidence interval:

$$\bar{x} \pm t_{0.025} \frac{s}{\sqrt{n}} = 1.343 \pm 2.145 \frac{0.192}{\sqrt{15}}$$
$$= 1.343 \pm 0.106$$
$$= [1.237, 1.449]$$




This interval says the bank is 95 percent confident that the mean debt-to-equity ratio for its portfolio of commercial loan accounts is between 1.2369 and 1.4497. Based on this interval, the bank has strong evidence that the portfolio's mean ratio is less than 1.5 (or that the bank is in compliance with its new policy).

# Example 8.5 The Marketing Research Case




Recall that a brand group is considering a new bottle design for a popular soft drink and that Table 1.5 (page 10) gives a random sample of $n = 60$ consumer ratings of this new bottle design. Let $\mu$ denote the mean rating of the new bottle design that would be given by all consumers. In order to assess whether $\mu$ exceeds the minimum standard composite score of 25 for a successful bottle design, the brand group will calculate a 95 percent confidence interval for $\mu$. The mean and the standard deviation of the 60 bottle design ratings are $\bar{x} = 30.35$ and $s = 3.1073$. It follows that a 95 percent confidence interval for $\mu$ is

$$\left[\bar{x} \pm t_{.025} \frac{s}{\sqrt{n}}\right] = \left[30.35 \pm 2.001 \frac{3.1073}{\sqrt{60}}\right] = [29.5, 31.2]$$

where $t_{.025} = 2.001$ is based on $n - 1 = 60 - 1 = 59$ degrees of freedom—see Table A.4 (page 792). Because the interval says we are 95 percent confident that the population mean rating of the new bottle design is between 29.5 and 31.2, we are 95 percent confident that this mean rating exceeds the minimum standard of 25 by at least 4.5 points and by at most 6.2 points.

# Outline

- **8.3 Sample Size Determination**

# 8.3 Sample Size Determination

In Section 8.1 we used a sample of 50 mileages to construct a 95 percent confidence interval for the midsize model's mean mileage $\mu$

$$\bar{x} \pm z_{0.025}\frac{\sigma}{\sqrt{n}} = 31.3 \pm 1.96\frac{0.8}{\sqrt{5}}$$
$$= 31.3 \pm 0.701$$
$$= [30.6,\ 32.0]$$

Interval's margin of error $z_{0.025}\frac{\sigma}{\sqrt{n}} = 0.701$ is not small enough

We can attempt to make the margin of error in the interval smaller by increasing the sample size *n*.

# 8.3 Interval's margin of error

Letting $E$ denote the desired margin of error

$$z_{\alpha/2}\frac{\sigma}{\sqrt{n}} = E$$

$$\sqrt{n} = z_{\alpha/2}\frac{\sigma}{E}$$

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

# 8.3 Sample Size Determination

## Determining the Sample Size for a Confidence Interval for $\mu$: $\sigma$ Known

A sample of size

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

makes the margin of error in a $100(1 - \alpha)$ percent confidence interval for $\mu$ equal to $E$. That is, this sample size makes us $100(1 - \alpha)$ percent confident that $\bar{x}$ is within $E$ units of $\mu$. If the calculated value of $n$ is not a whole number, round this value up to the next whole number (so that the margin of error is at least as small as desired).

If $\sigma$ is unknown and is estimated from $s$, then the sample size can be determined as

$$n = \left(\frac{t_{\alpha/2}s}{E}\right)^2$$

so that $\bar{x}$ is within $E$ units of $\mu$, with $100(1-\alpha)\%$ confidence. The number of degrees of freedom for the $t_{\alpha/2}$ point is the size of the preliminary sample minus 1

Example 8.6

# Car Mileage Case

Suppose in the car mileage situation we wish to find the sample size that is needed to make the margin of error in a 95 percent confidence interval for μ equal to 0.3. Assuming that σ is known to equal 0.8.

$z_{0.025} = 1.96$

$$n = \left(\frac{z_{\alpha/2}\sigma}{E}\right)^2 = \left(\frac{z_{0.025}\sigma}{E}\right)^2 = \left(\frac{1.96 \times 0.8}{0.3}\right)^2 = 27.32$$

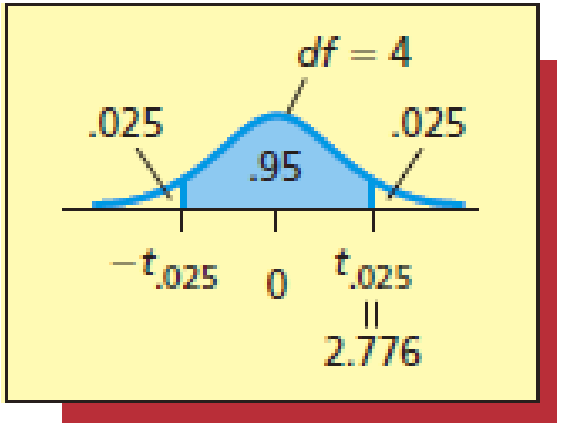
Rounding up, we would employ a sample of size 28

Example 8.6

# Car Mileage Case

- Suppose in the car mileage situation we wish to find the sample size that is needed to make the margin of error in a 95 percent confidence interval for $\mu$ equal to 0.3. Assuming that $\sigma$ is unknown.
- We regard the sample of five mileages (n=5) as a preliminary sample with standard deviation s=0.7583.
- $t_{\alpha/2} = t_{0.025} = 2.776$ for $df = 4$

$$n = \left(\frac{t_{\alpha/2}s}{E}\right)^2 = \left(\frac{2.776 \times 0.7583}{0.3}\right)^2 = 49.24$$




Rounding up, we employ a sample of size 50.

95 percent of all possible sample means based on 50 mileages are within 0.3 of $\mu$

# Estimate σ when σ unknown and no preliminary sample

Finally, sometimes we do not know $\sigma$ and we do not have a preliminary sample that can be used to estimate $\sigma$. In this case it can be shown that, if we can make a reasonable guess of the range of the population being studied, then a conservatively large estimate of $\sigma$ is this estimated range divided by 4. For example, if the automaker's design engineers feel that almost all of its midsize cars should get mileages within a range of 5 mpg, then a conservatively large estimate of $\sigma$ is $5/4 = 1.25$ mpg. When employing such an estimate of $\sigma$, it is sufficient to use the $z$-based sample size formula $n = (z_{\alpha/2}\sigma/E)^2$, because a conservatively large estimate of $\sigma$ will give us a conservatively large sample size.

# Outline

- 8.4 Confidence Intervals for a Population Proportion

# 8.4 Confidence Intervals for a Population Proportion

In Chapter 7, the soft cheese spread producer decided to replace its current spout with the new spout if $p$, the true proportion of all current purchasers who would stop buying the cheese spread if the new spout were used, is less than .10. Suppose that when 1,000 current purchasers are randomly selected and are asked to try the new spout, 63 say they would stop buying the spread if the new spout were used. The point estimate of the population proportion $p$ is the sample proportion $\hat{p} = 63/1{,}000 = .063$. This sample proportion says we estimate that 6.3 percent of all current purchasers would stop buying the cheese spread if the new spout were used. Since $\hat{p}$ equals .063, we have some evidence that $p$ is less than .10.

# 8.4 Confidence Intervals for a Population Proportion

In order to see if there is strong evidence that $p$ is less than .10, we can calculate a confidence interval for $p$. As explained in Chapter 7, if the sample size $n$ is large, then the sampling distribution of the sample proportion $\hat{p}$ is approximately a normal distribution with mean $\mu_{\hat{p}} = p$, and standard deviation $\sigma_{\hat{p}} = \sqrt{\frac{p(1-p)}{n}}$

If the sample size $n$ is large*, a **(1–α)100% confidence interval for *p*** is

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

* Here $n$ should be considered large if both $n \cdot \hat{p} \geq 5$ and $n \cdot (1-\hat{p}) \geq 5$

## EXAMPLE 8.6 The Cheese Spread Case: Improving Profitability

Suppose that the cheese spread producer wishes to calculate a 99 percent confidence interval for $p$, the population proportion of purchasers who would stop buying the cheese spread if the new spout were used. To determine whether the sample size $n = 1{,}000$ is large enough to enable us to use the confidence interval formula just given, recall that the point estimate of $p$ is $\hat{p} = 63/1{,}000 = .063$. Therefore, because $n\hat{p} = 1{,}000(.063) = 63$ and $n(1 - \hat{p}) = 1{,}000(.937) = 937$ are both greater than 5, we can use the confidence interval formula. It follows that the 99 percent confidence interval for $p$ is

# Example: The Cheese Spread Case

$$[\widehat{p} \pm z_{\alpha/2}\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}] = [\hat{p} \pm z_{0.005}\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}]$$

$$= [0.063 \pm 2.575\sqrt{\frac{0.063 \times (0.937)}{1000}}]$$

$$= [0.0432, 0.0828]$$

This interval says that we are 99 percent confident that between 4.32 percent and 8.28 percent of all current purchasers would stop buying the cheese spread if the new spout were used. Moreover, because the upper limit of the 99 percent confidence interval is less than 0.10, we have very strong evidence that the true proportion *p* of all current purchasers who would stop buying the cheese spread is less than 0.10. Based on this result, it seems reasonable to use the new spout.

# Example 8.9

## The Phe-Mycin Case: Drug Side Effects

Antibiotics occasionally cause nausea as a side effect. Scientists working for a major drug company have developed a new antibiotic called Phe-Mycin. The company wishes to estimate $p$, the proportion of all patients who would experience nausea as a side effect when being treated with Phe-Mycin. Suppose that a sample of 200 patients is randomly selected. When these patients are treated with Phe-Mycin, 35 patients experience nausea. The point estimate of the population proportion $p$ is the sample proportion $\hat{p} = 35/200 = .175$. This sample proportion says that we estimate that 17.5 percent of all patients would experience nausea as a side effect of taking Phe-Mycin. Furthermore, because $n\hat{p} = 200(.175) = 35$ and $n(1 - \hat{p}) = 200(.825) = 165$ are

Example 8.9

# Phe-Mycin Side Effects

Given: $n$ = 200, 35 patients experience nausea.

$$\hat{p} = \frac{35}{200} = 0.175$$

Note:

$$n\hat{p} = 200 \times 0.175 = 35$$
$$n(1-\hat{p}) = 200 \times 0.825 = 165$$

so both quantities are > 5

For 95% confidence, $z_{\alpha/2} = z_{0.025} = 1.96$ and

$$\hat{p} \pm z_{\alpha/2}\sqrt{\frac{\hat{p}(1-\hat{p})}{n}} = 0.175 \pm 1.96\sqrt{\frac{0.175 \times 0.825}{200}}$$
$$= 0.175 \pm 0.053$$
$$= [0.122,\ 0.228]$$

# Example 8.10

## The Marketing Ethics Case: Confidentiality

C

In the book *Essentials of Marketing Research,* William R. Dillon, Thomas J. Madden, and Neil H. Firtle discuss a survey of marketing professionals, the results of which were originally published by Ishmael P. Akoah and Edward A. Riordan in the *Journal of Marketing Research*. In the study, randomly selected marketing researchers were presented with various scenarios involving ethical issues such as confidentiality, conflict of interest, and social acceptability. The marketing researchers were asked to indicate whether they approved or disapproved of the actions described in each scenario. For instance, one scenario that involved the issue of confidentiality was described as follows:

**Use of ultraviolet ink** A project director went to the marketing research director's office and requested permission to use an ultraviolet ink to precode a questionnaire for a mail survey. The project director pointed out that although the cover letter promised confidentiality, respondent identification was needed to permit adequate cross-tabulations of the data. The marketing research director gave approval.

Of the 205 marketing researchers who participated in the survey, 117 said they disapproved of the actions taken in the scenario. It follows that a point estimate of $p$, the proportion of all marketing researchers who disapprove of the actions taken in the scenario, is $\hat{p} = 117/205 = .5707$. Furthermore, because $n\hat{p} = 205(.5707) = 117$ and $n(1 - \hat{p}) = 205(.4293) = 88$ are both at least 5, a 95 percent confidence interval for $p$ is

$$\left[\hat{p} \pm z_{.025}\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}\right] = \left[.5707 \pm 1.96\sqrt{\frac{(.5707)(.4293)}{205}}\right]$$
$$= [.5707 \pm .0678]$$
$$= [.5029, .6385]$$

This interval says we are 95 percent confident that between 50.29 percent and 63.85 percent of all marketing researchers disapprove of the actions taken in the ultraviolet ink scenario. Notice that because the margin of error (.0678) in this interval is rather large, this interval does not provide a very precise estimate of $p$. Below we show the MINITAB output of this interval.

**CI for One Proportion**

| X | N | Sample p | 95% CI |
|---|---|---|---|
| 117 | 205 | 0.570732 | (0.502975, 0.638488) |

## Determining the Sample Size for a Confidence Interval for $p$

A sample of size

$$n = p(1 - p)\left(\frac{z_{\alpha/2}}{E}\right)^2$$

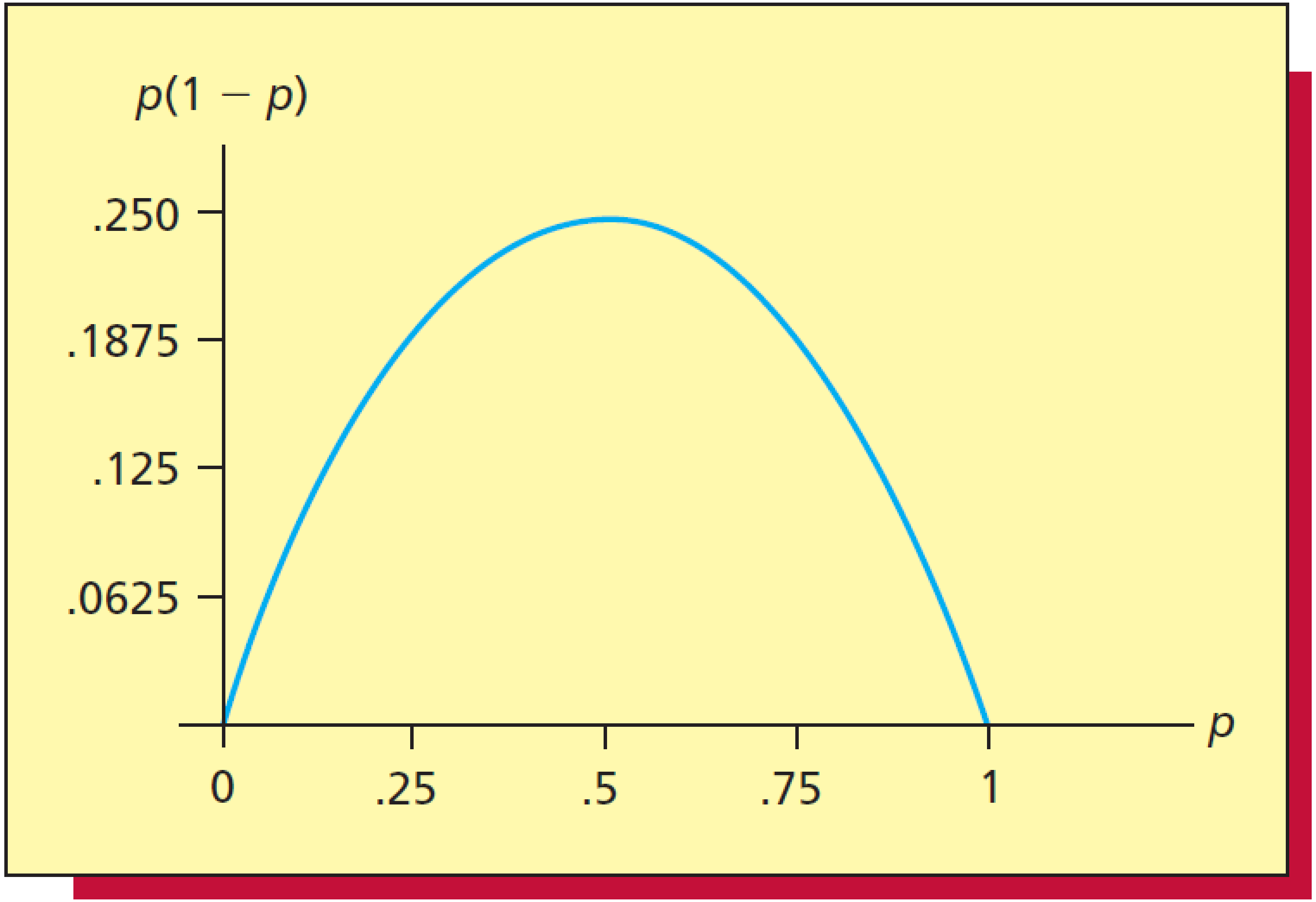
makes the margin of error in a $100(1 - \alpha)$ percent confidence interval for $p$ equal to $E$. That is, this sample size makes us $100(1 - \alpha)$ percent confident that $\hat{p}$ is within $E$ units of $p$. If the calculated value of $n$ is not a whole number, round this value up to the next whole number.

Note that the formula requires a preliminary estimate of $p$. The conservative value of $p = 0.5$ is generally used when there is no prior information on $p$

## FIGURE 8.16 The Graph of $p(1-p)$ versus $p$

When $p$=0.5, $p(1-p)$ is the largest (=0.25).

Therefore, if you don't know $p$, assume it equal 0.5

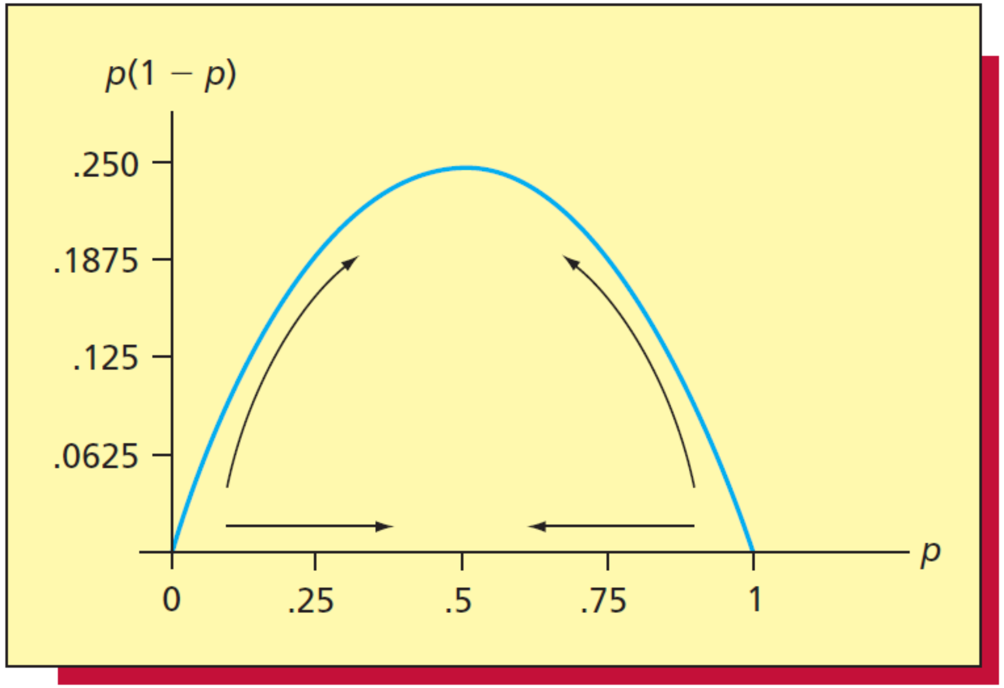


For example, suppose we wish to estimate the proportion $p$ of all registered voters who currently favor a particular candidate for president of the United States. If this candidate is the nominee of a major political party, or if the candidate enjoys broad popularity for some other reason, then $p$ could be near .5. Furthermore, suppose we wish to make the margin of error in a 95 percent confidence interval for $p$ equal to .02. If the sample to be taken is random, it should consist of $n$ registered voters

$$n = p(1-p)\left(\frac{z_{\alpha/2}}{E}\right)^2 = 0.25 \times \left(\frac{1.96}{0.02}\right)^2 = 2401$$

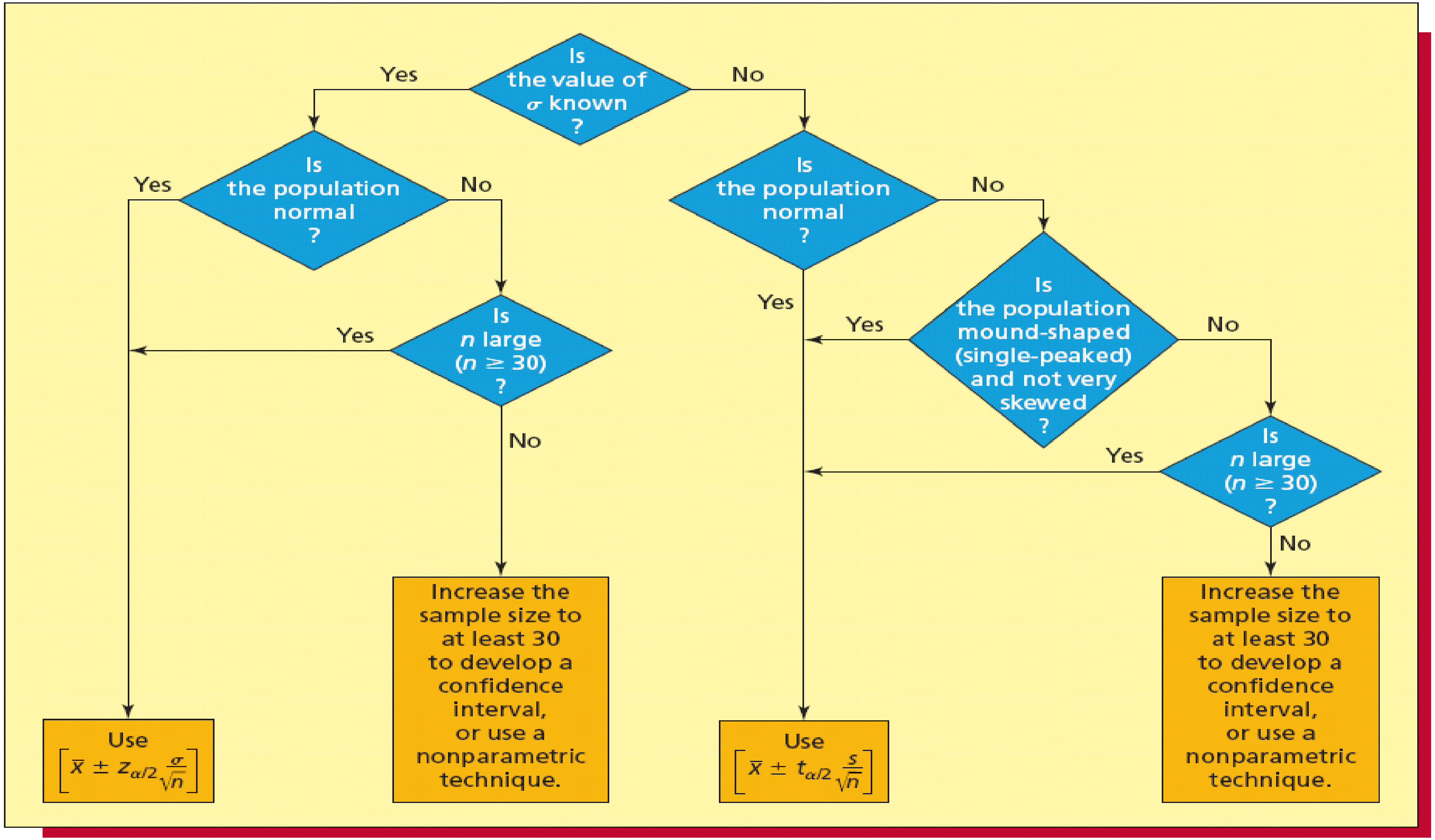
For instance, if the sample proportion $\hat{p}$ equals .53, we are 95 percent confident that the proportion of all registered voters who favor the candidate is between .51 and .55.

**FIGURE 8.17** As $p$ Gets Closer to .5, $p(1 - p)$ Increases




If the population proportion we are estimating is substantially different from .5, setting $p$ equal to .5 will give a sample size that is much larger than is needed. In this case, we should use our intuition or previous sample information (along with Figure 8.17) to determine the largest reasonable value for $p(1 - p)$. Figure 8.17 implies that as $p$ gets closer to .5, $p(1 - p)$ increases. It follows that $p(1 - p)$ is maximized by the reasonable value of $p$ that is closest to .5. Therefore, **when we are estimating a proportion that is substantially different from .5, we use the reasonable value of $p$ that is closest to .5 to calculate the sample size needed to obtain a specified margin of error.**

# Summary: Selecting an Appropriate Confidence Interval for a Population Mean



**Example:** Suppose an insurance company studies repair costs after rear collisions, and finds the mean repair cost to be $2300 based on a sample of 40 accidents. Suppose the standard deviation of population is $1025.

Find the 95% Confidence Interval.

A sample of 500 executives who own their own home revealed 175 planned to sell their homes and retire to Arizona. Develop a 98% confidence interval for the proportion of executives that plan to sell and move to Arizona.

The American Kennel Club wanted to estimate the proportion of children that have a dog as a pet. If the club wanted the estimate to be within 3% of the population proportion, how many children would they need to contact?

Assume a 95% level of confidence and that the club estimated that 30% of the children have a dog as a pet.

# Chapter Summary

- In this chapter we discussed **confidence intervals** for population **means** and **proportions.**
- First, we studied how to compute a confidence interval for a **population mean.** We saw that when the population standard deviation $\sigma$ is known, we can use the **normal distribution** to compute a confidence interval for a population mean.
- When $\sigma$ is not known, if the population is normally distributed (or at least mound-shaped) or if the sample size $n$ is large, we use the ***t* distribution** to compute this interval.
- We also studied how to find the size of the sample needed if we wish to compute a confidence interval for a mean with a prespecified *confidence level* and with a prespecified *margin of error.*
- Next we saw that we are often interested in estimating the proportion of population units falling into a category of interest. We showed how to compute a large sample confidence interval for a **population proportion,** and we saw how to find the sample size needed to estimate a population proportion.

Thank you!