Business Analytics

UNIVERSITETI I EJL
ЈИЕ УНИВЕРЗИТЕТ
SEE UNIVERSITY

Confidence Intervals

Faton Berisha

Chapter 7

Confidence Intervals

Confidence Intervals

7.1 z-Based Confidence Intervals for a Population Mean: σ Known

7.2 t-Based Confidence Intervals for a Population Mean: σ Unknown

7.3 Sample Size Determination

7.4 Confidence Intervals for a Population Proportion

7.5 A Comparison of Confidence Intervals and Tolerance Intervals (Optional)

z-Based Confidence Intervals

for a Mean: σ Known

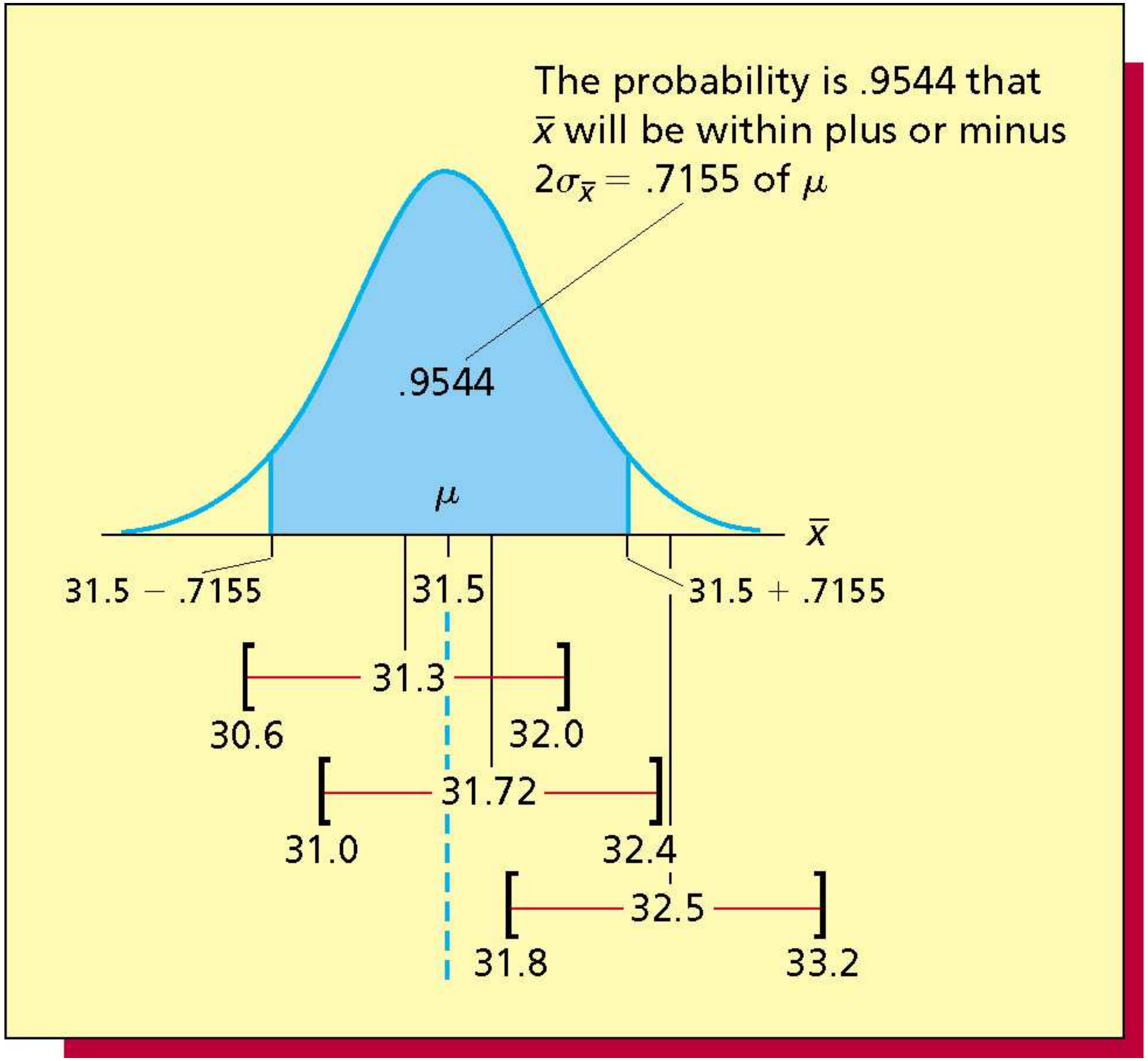
- The starting point is the sampling distribution of the sample mean
  - Recall from Chapter 6 that if a population is normally distributed with mean μ and standard deviation σ, then the sampling distribution of $\bar{x}$ is normal with mean $\mu_{\bar{x}} = \mu$ and standard deviation $\sigma_{\bar{x}} = \sigma / \sqrt{n}$
  - 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 samples

The Empirical Rule

- 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 observed values of *x* are within two standard deviations of the population mean
  - 99.73% of all possible observed values of *x* are within three standard deviations of the population mean

Three 95.44% Confidence Intervals for μ

- Car mileage case
  - Probability is also 0.9544 that μ will be inside plus ose minus $2\sigma_{\bar{x}}$ of $\bar{x}$.



Intervals of Confidence and the Empirical Rule

- Based on sampling distribution, we found the probability for $\bar{x}$ to be in the interval $[\mu \pm 2\sigma_{\bar{x}}]$
- The claim
  - $\bar{x}$ is in the inteval $[\mu \pm 2\sigma_{\bar{x}}]$

  is same as (equivalent) as the claim
  - μ will be in the interval $[\bar{x} \pm 2\sigma_{\bar{x}}]$

Generalizing

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

Generalizing Continued

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

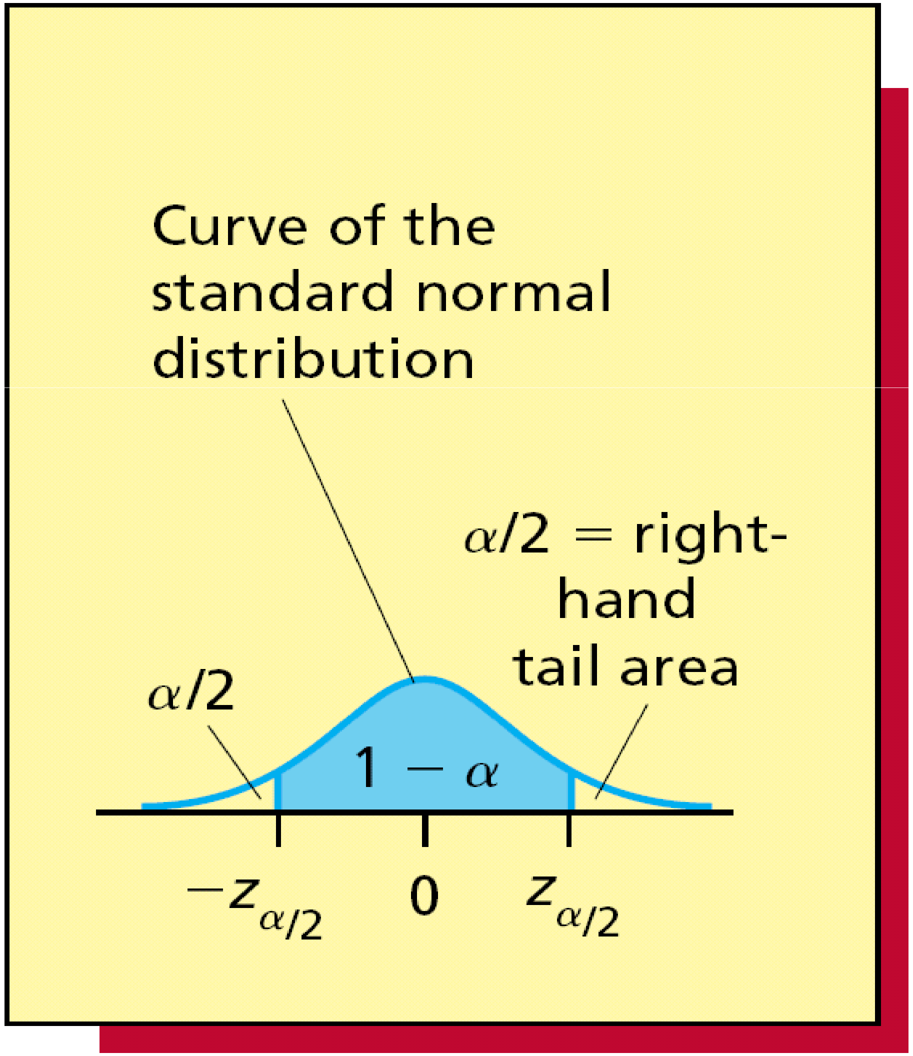
General Confidence Interval

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

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

  - 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 $1 - \alpha$

Sampling Distribution Of All
Possible Sample Means



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 $(1-\alpha)$**100% confidence interval for** μ is

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

The Car Mileage Case

- Example 7.1.
  - $n = 49$, $\bar{x} = 31.5531$, $s = 0.7992$
  - Find 95% confidence interval for the mean mileage μ of population of all cars.
  - $1 - \alpha = 0.95$
    $\alpha = 0.05$
  - $z_{\alpha/2} = z_{0.025}$

95% Confidence Level

- ❖ For a 95% confidence level,

  $1 - \alpha = 0.95$

  $\alpha = 0.05$

  $\alpha/2 = 0.025$

- 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

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

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

99% Confidence Interval

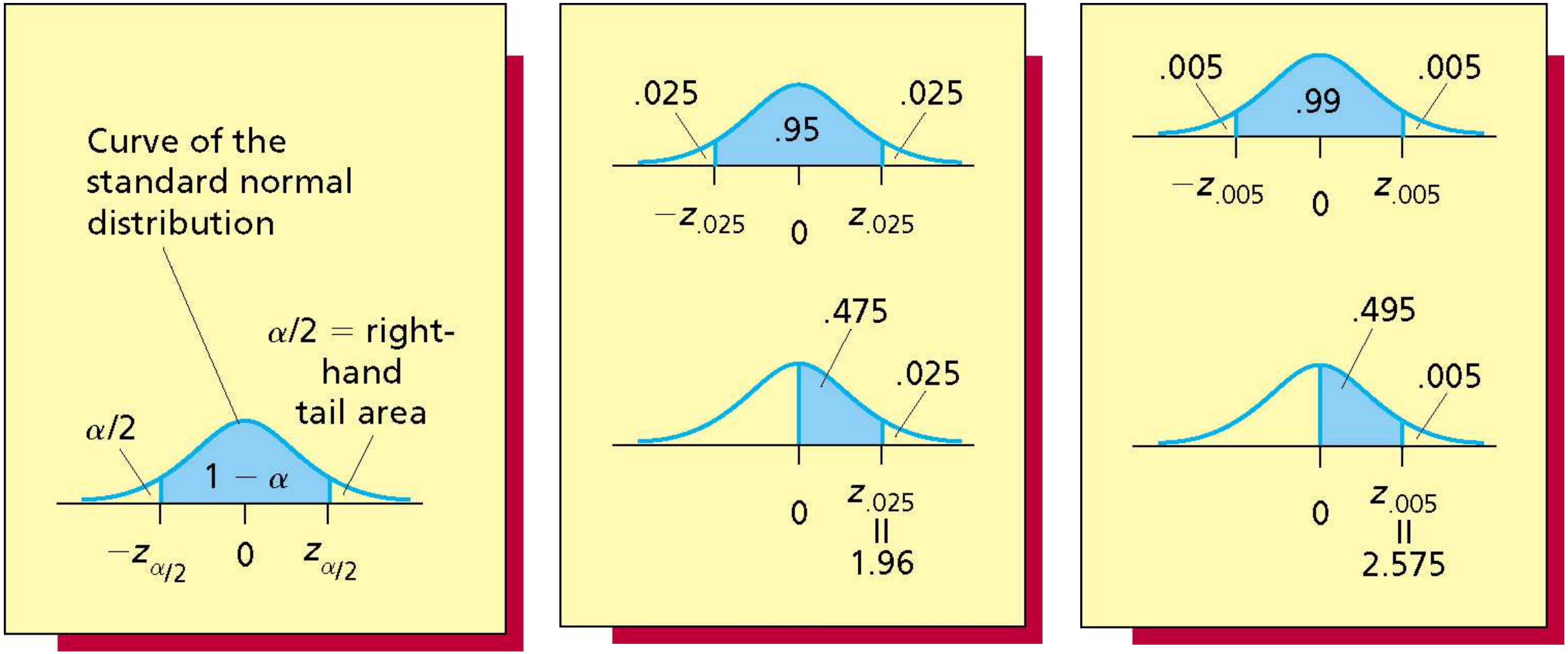
- For 99% confidence, need the normal point $z_{0.005}$
  - Reading between table entries in the standard normal table, the area is 0.495 for $z_{0.005} = 2.575$
- The 99% confidence interval is

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

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

The Effect of α on Confidence
Interval Width

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



t-Based Confidence Intervals for a Mean: σ Unknown

- If σ is unknown (which is usually the case), we can construct a confidence interval for μ based on the sampling distribution of

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

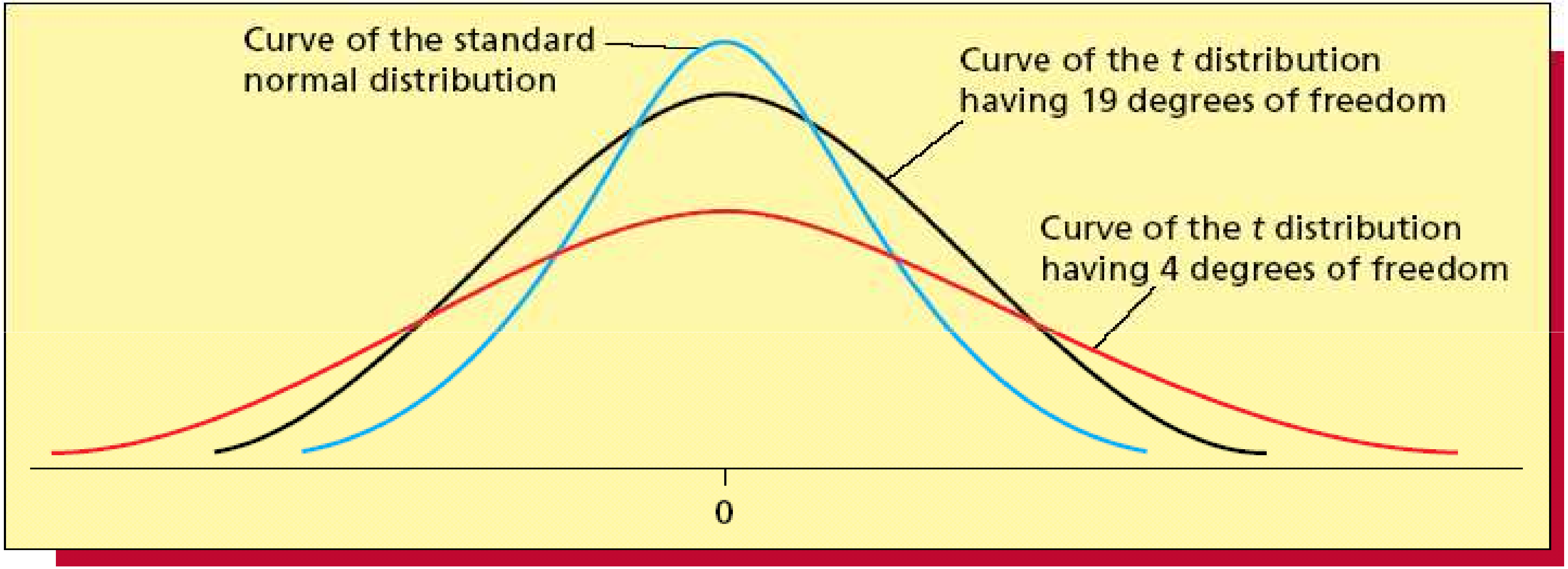
- If the population is normal, then for any sample size n, this sampling distribution is called the ***t* distribution**

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

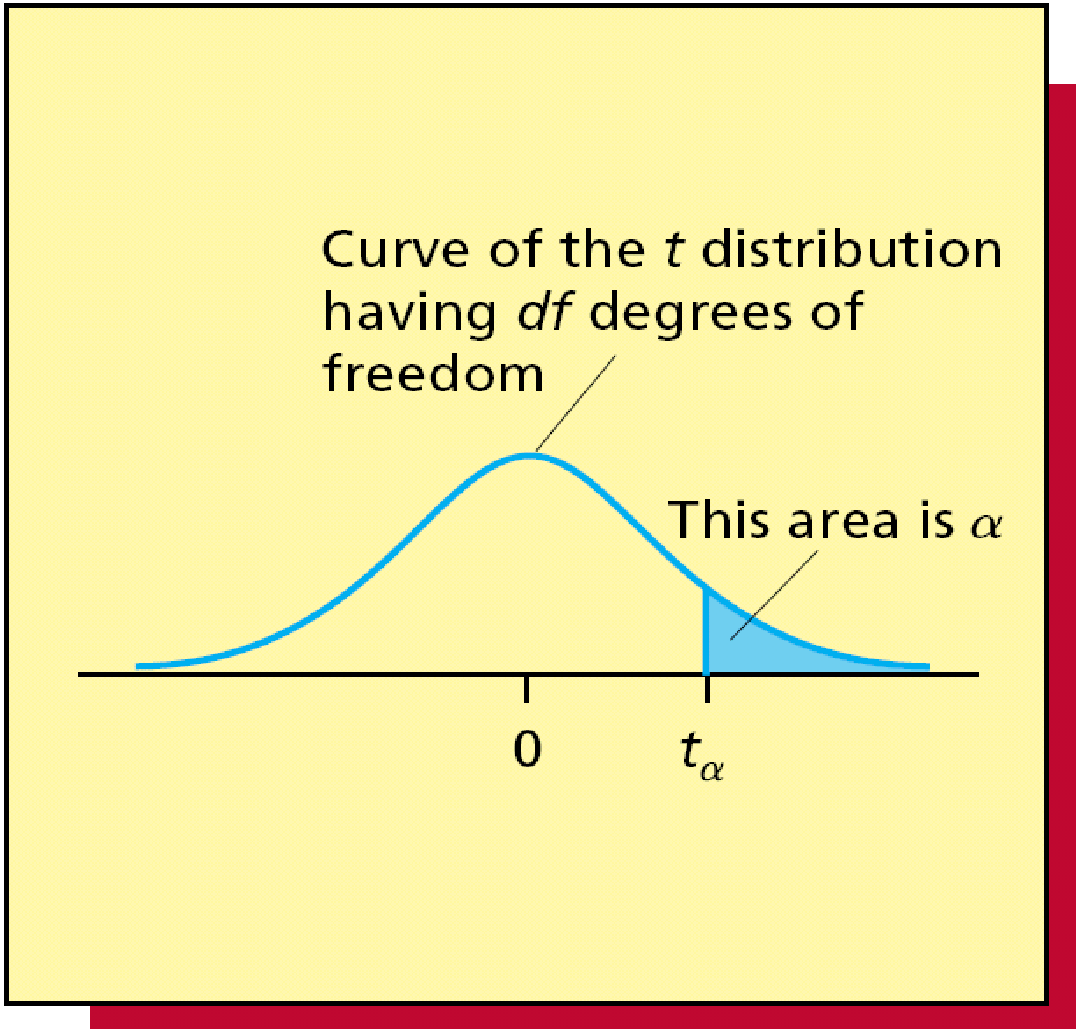
The *t* Distribution and Degrees of Freedom

- 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_α
  - t_α is the point on the horizontal axis under the t curve that gives a right hand tail equal to α
  - So the value of t_α in a particular situation depends on the right hand tail area α and the number of degrees of freedom
    - $df = n - 1$
    - $\alpha = 1 - \alpha$, where $1 - \alpha$ is the specified confidence coefficient

t and Right Hand Tail Areas



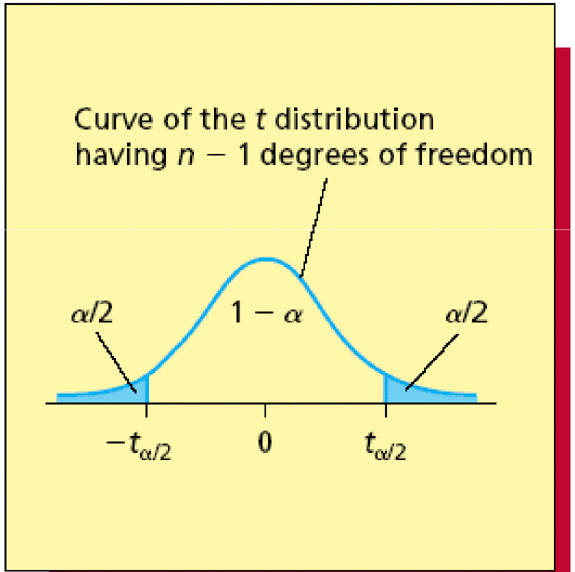
Using the *t* Distribution Table

- Rows correspond to the different values of *df*
- Columns correspond to different values of α
- See Table 7.3, Tables A.4 and A.20 in Appendix A and the table on the inside cover
  - Table 7.3 and A.4 gives *t* points for *df* 1 to 30, then for *df* = 40, 60, 120 and ∞
    - On the row for ∞, the *t* points are the *z* points
  - Table A.20 gives *t* points for *df* from 1 to 100
    - For *df* greater than 100, *t* points can be approximated by the corresponding *z* points on the bottom row for *df* = ∞
  - Always look at the accompanying figure for guidance on how to use the table

t-Based Confidence Intervals for a Mean: σ Unknown

If the sampled population is normally distributed with mean μ, then a $(1-\alpha)100\%$ confidence interval for μ 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

Sample Size Determination (z)

If σ is known, then a sample of size

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

so that ⊠ is within *E* units of μ, with 100(1-α)% confidence

Sample Size Determination (t)

If σ is unknown and is estimated from *s*, then a sample of size

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

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

Confidence Intervals for a
Population Proportion

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

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

* Here n should be considered large if both

$$n \cdot \hat{p} \geq 5 \quad \textit{and} \quad n \cdot \left(1-\hat{p}\right) \geq 5$$

Determining Sample Size for
Confidence Interval for p

A sample size

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

will yield an estimate $\hat{p}$, precisely within E units of p, with 100(1-α)% confidence.

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.

A Comparison of Confidence
Intervals and Tolerance Intervals

A **tolerance interval** contains a specified percentage of individual population measurements

- Often 68.26%, 95.44%, 99.73%

A **confidence interval** is an interval that contains the population mean μ, and the confidence level expresses how sure we are that this interval contains μ

- Often confidence level is set high (e.g., 95% or 99%)
  - Because such a level is considered high enough to provide convincing evidence about the value of μ

Summary: Selecting an Appropriate
Confidence Interval for a Population Mean