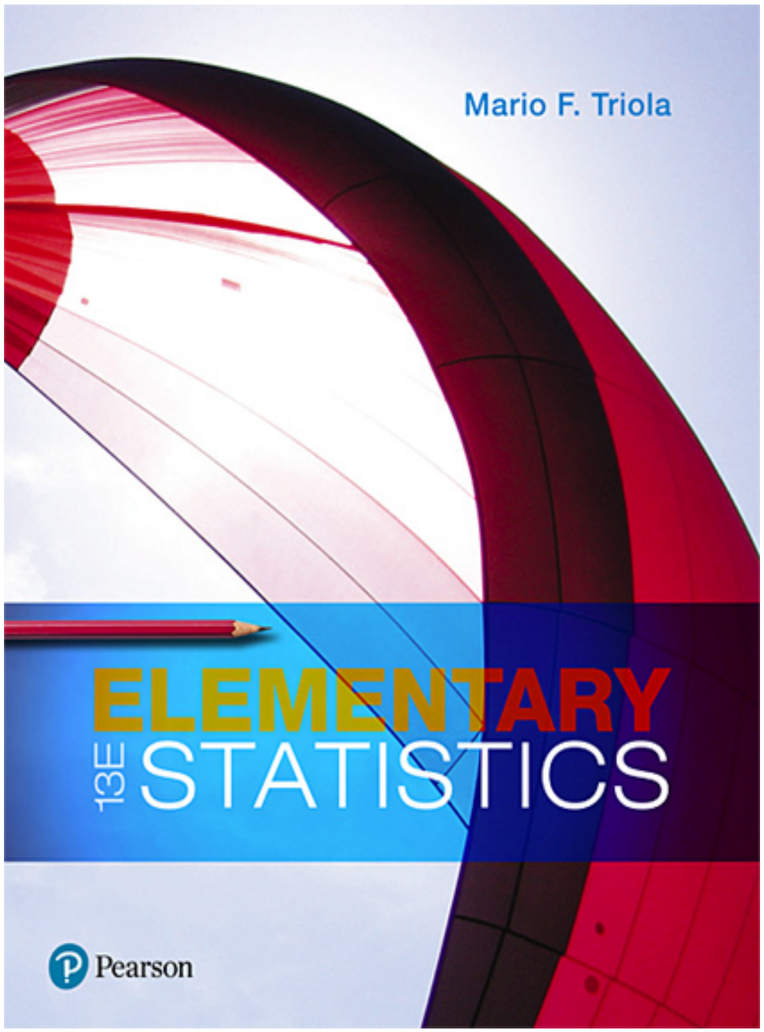
Elementary Statistics

Thirteenth Edition



Chapter 7

Estimating Parameters and Determining Sample Sizes

Estimating Parameters and Determining Sample Sizes

Key Concept (1 of 2)

The main goal of this section is to present methods for using a sample mean $\bar{x}$ to make an inference about the value of the corresponding population mean μ.

Key Concept (2 of 2)

There are three main concepts included in this section:

- **Point Estimate:** The sample mean $\bar{x}$ is the best **point estimate** (or single value estimate) of the population mean μ.
- **Confidence Interval:** Use sample data to construct and interpret a **confidence interval** estimate of the true value of a population mean μ.
- **Sample Size:** Find the sample size necessary to estimate a population mean.

Confidence Interval for Estimating a Population Mean with σ Not Known: Objective

Construct a confidence interval used to estimate a population mean.

Confidence Interval for Estimating a Population Mean with σ Not Known: Notation

μ = population mean

n = number of sample values

$\bar{x}$ = sample mean

E = margin of error

s = sample standard deviation

Confidence Interval for Estimating a Population Mean with σ Not Known: Requirements

1. The sample is a simple random sample.
2. Either or both of these conditions are satisfied: The population is normally distributed or $n > 30$.

Confidence Interval for Estimating a Population Mean with σ Not Known: Confidence Interval (1 of 2)

Margin of Error:

$$E = t_{\frac{\alpha}{2}} \cdot \frac{s}{\sqrt{n}} \quad \text{(Use df} = n - 1\text{)}$$

Confidence Interval: The confidence interval is associated with a confidence level, such as 0.95 (or 95%), and α is the complement of the confidence level. For a 0.95 (or 95%) confidence level, $\alpha = 0.05$.

Confidence Interval for Estimating a Population Mean with σ Not Known: Confidence Interval (2 of 2)

Critical Value: $t_{\frac{\alpha}{2}}$ is the critical t value separating an area of $\frac{\alpha}{2}$ in the right tail of the Student t distribution.

Degrees of Freedom: df = $n - 1$ is the number of degrees of freedom used when finding the critical value.

Confidence Interval for Estimating a Population Mean with σ Not Known: Round-Off Rule

1. **Original Data:** When using an **original set of data** values, round the confidence interval limits to one more decimal place than is used for the original set of data.
2. **Summary Statistics:** When using the **summary statistics** of n, x, and s, round the confidence interval limits to the same number of decimal places used for the sample mean.

Key Points about the Student *t* Distribution (1 of 5)

- **Student *t* Distribution** If a population has a normal distribution, then the distribution of

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

is a **Student *t* distribution** for all samples of size *n*. A Student *t* distribution is commonly referred to as a ***t* distribution.**

Key Points about the Student *t* Distribution (2 of 5)

- **Degrees of Freedom** Finding a critical value $t_{\frac{\alpha}{2}}$ requires a value for the **degrees of freedom** (or **df**).

  In general, the number of degrees of freedom for a collection of sample data is the number of sample values that can vary after certain restrictions have been imposed on all data values. For the methods of this section, the number of degrees of freedom is the sample size minus 1.

  Degrees of freedom = $n - 1$

Key Points about the Student *t* Distribution (3 of 5)

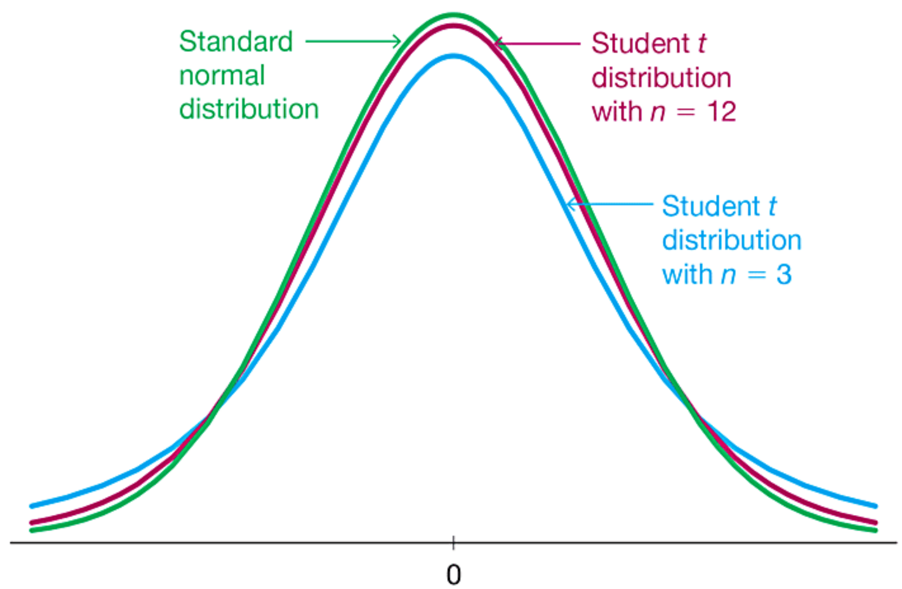
Finding Critical Value $t_{\frac{\alpha}{2}}$

A critical value $t_{\frac{\alpha}{2}}$ can be found using technology or Table A-3. Technology can be used with any number of degrees of freedom, but Table A-3 can be used for select numbers of degrees of freedom only. If using Table A-3 to find a critical value of $t_{\frac{\alpha}{2}}$, but the table does not include the exact number of degrees of freedom, you could use the closest value, or you could be conservative by using the next lower number of degrees of freedom found in the table, or you could interpolate.

Key Points about the Student *t* Distribution (4 of 5)

- The Student *t* distribution is different for different sample sizes. See the figure for the cases $n = 3$ and $n = 12$.

The Student *t* distribution has the same general shape and symmetry as the standard normal distribution, but it has the greater variability that is expected with small samples.



Key Points about the Student *t* Distribution (5 of 5)

- The Student *t* distribution has the same general symmetric bell shape as the standard normal distribution, but has more variability (with wider distributions), as we expect with small samples.
- The Student *t* distribution has a mean of $t = 0$ (just as the standard normal distribution has a mean of $z = 0$).
- The standard deviation of the Student *t* distribution varies with the sample size, but it is greater than 1 (unlike the standard normal distribution, which has $s = 1$).
- As the sample size *n* gets larger, the Student *t* distribution gets closer to the standard normal distribution.

Procedure for Constructing a Confidence Interval for μ (1 of 2)

1. Verify that the two requirements are satisfied: The sample is a simple random sample and the population is normally distributed or $n > 30$.
2. With σ unknown (as is usually the case), use $n - 1$ degrees of freedom and use technology or a t distribution table (such as Table A-3) to find the critical value $t_{\frac{\alpha}{2}}$ that corresponds to the desired confidence level.
3. Evaluate the margin of error using $E = t_{\frac{\alpha}{2}} \cdot \frac{s}{\sqrt{n}}$.

Procedure for Constructing a Confidence Interval for μ (2 of 2)

4. Using the value of the calculated margin of error E and the sample mean $\bar{x}$, substitute the values in one of the formats for CI:

$$\bar{x} - E < \mu < \bar{x} + E \text{ or } \bar{x} \pm E \text{ or } (\bar{x} - E, \bar{x} + E).$$

5. With an **original set of data** values, round the confidence interval limits to one more decimal place than used for the original set of data, but when using the **summary statistics** of n, $\bar{x}$, and s, round the confidence interval limits to the same number of decimal places used for the sample mean.

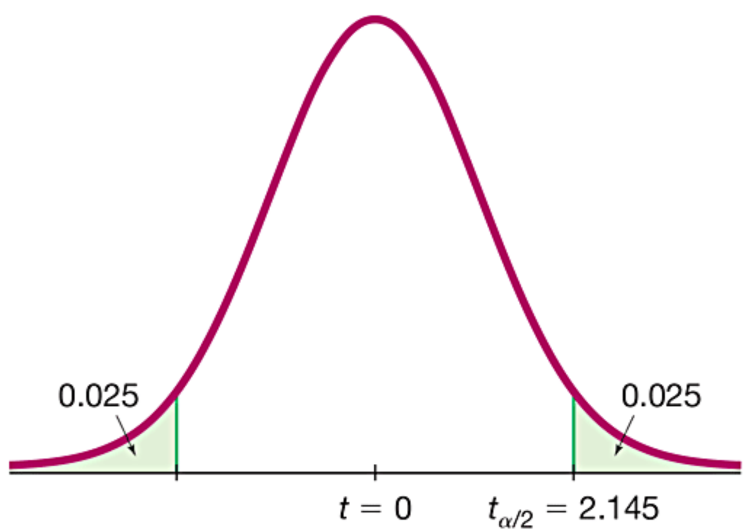
Example: Finding a Critical Value $t_{\frac{\alpha}{2}}$ (1 of 4)

Find the critical value $t_{\frac{\alpha}{2}}$ corresponding to a 95% confidence level, given that the sample has size $n = 15$.

Example: Finding a Critical Value $t_{\alpha/2}$ (2 of 4)

Solution

Because n = 15, the number of degrees of freedom is $n - 1 = 14$. The 95% confidence level corresponds to σ = 0.05, so there is an area of 0.025 in each of the two tails of the t distribution, as shown.

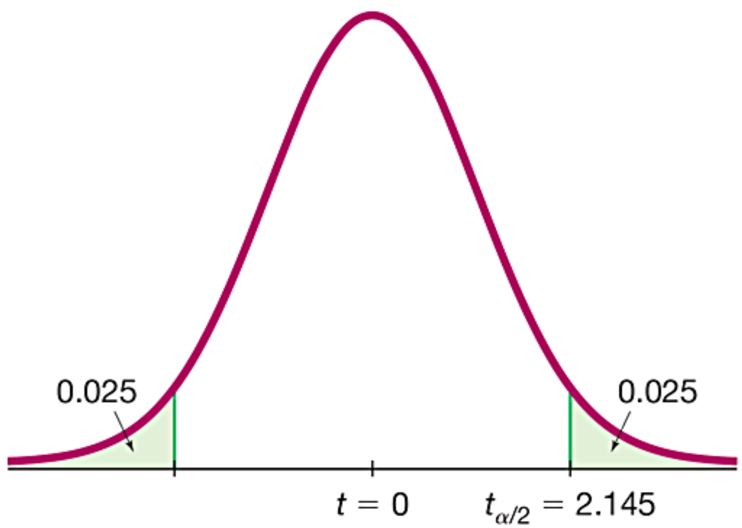


Pearson

Example: Finding a Critical Value t alpha by 2 (3 of 4)

Solution

Using Technology Technology can be used to find that for 14 degrees of freedom and an area of 0.025 in each tail, the critical value is $t_{\frac{\alpha}{2}} = t_{0.025} = 2.145$.



Pearson

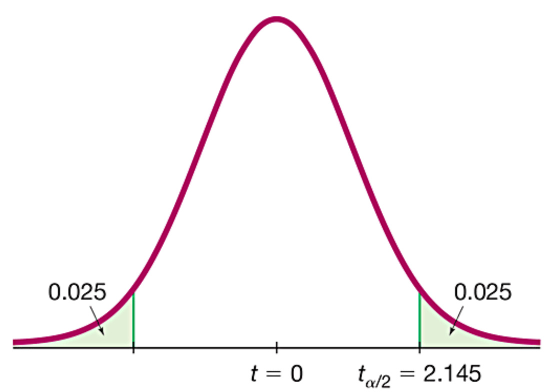
Example: Finding a Critical Value *t* alpha by 2 (4 of 4)

Solution

Using Table A-3 To find the critical value using Table A-3, use the column with 0.05 for the "Area in Two Tails" (or use the same column with 0.025 for the "Area in One Tail").

The number of degrees of freedom is df = $n - 1 = 14$.

We get $t_{\frac{\alpha}{2}} = t_{0.025} = 2.145$.



Finding a Point Estimate and Margin of Error *E* from a Confidence Interval

Point estimate of μ:

$$\bar{x} = \frac{(\text{upper confidence limit}) + (\text{lower confidence limit})}{2}$$

Margin of error:

$$E = \frac{(\text{upper confidence limit}) - (\text{lower confidence limit})}{2}$$

Finding the Sample Size Required to Estimate a Population Mean: Objective

Determine the sample size *n* required to estimate the value of a population mean μ.

Finding the Sample Size Required to Estimate a Population Mean: Notation

μ = population mean

σ = population standard deviation

$\bar{x}$ = sample mean

E = desired margin of error

$z_{\frac{\alpha}{2}}$ = z score separating an area of $\frac{\alpha}{2}$ in the right tail of the standard normal distribution

Finding the Sample Size Required to Estimate a Population Mean: Requirement

The sample must be a simple random sample.

Finding the Sample Size Required to Estimate a Population Mean: Sample Size

The required sample size is found by

$$n = \left[\frac{z_{\frac{\alpha}{2}} \sigma}{E} \right]^2$$

Finding the Sample Size Required to Estimate a Population Mean: Round-Off Rule

If the computed sample size n is not a whole number, round the value of n up to the next **larger** whole number.

Dealing with Unknown σ When Finding Sample Size

1. Use the range rule of thumb to estimate the standard deviation as follows: $\sigma \approx \text{range}/4$, where the range is determined from sample data.
2. Start the sample collection process without knowing σ and, using the first several values, calculate the sample standard deviation s and use it in place of σ. The estimated value of σ can then be improved as more sample data are obtained, and the sample size can be refined accordingly.
3. Estimate the value of σ by using the results of some other earlier study.

Example: IQ Scores of Statistics Students (1 of 3)

Assume that we want to estimate the mean IQ score for the population of statistics students. How many statistics students must be randomly selected for IQ tests if we want 95% confidence that the sample mean is within 3 IQ points of the population mean?

Example: IQ Scores of Statistics Students (2 of 3)

Solution

For a 95% confidence interval, we have $\alpha = 0.05$, so $z_{\frac{\alpha}{2}} = 1.96$. Because we want the sample mean to be within 3 IQ points of μ, the margin of error is $E = 3$. Also, we can assume that $\sigma = 15$ (see the discussion that immediately precedes this example). We get

$$n = \left[\frac{z_{\frac{\alpha}{2}}\sigma}{E}\right]^2 = \left[\frac{1.96 \cdot 15}{3}\right]^2 = 96.04 = 97. \text{ (rounded } \textbf{up}\text{)}$$

Example: IQ Scores of Statistics Students (3 of 3)

Interpretation

Among the thousands of statistics students, we need to obtain a simple random sample of at least 97 of their IQ scores.

With a simple random sample of only 97 statistics students, we will be 95% confident that the sample mean $\bar{x}$ is within 3 IQ points of the true population mean μ.

Estimating a Population Mean When σ Is Known

If we somehow do know the value of σ, the confidence interval is constructed using the standard normal distribution instead of the Student *t* distribution, so the same procedure can be used with this margin of error:

$$\text{Margin of Error: } E = z_{\frac{\alpha}{2}} \cdot \frac{\sigma}{\sqrt{n}} \quad \text{(used with known } \sigma\text{).}$$

Example: Confidence Interval Estimate of μ with Known σ (1 of 6)

Use the 15 birth weights of girls given below, for which n = 15 and $\bar{x}$ = 30.9 hg. Construct a 95% confidence interval estimate of the mean birth weight of all girls by assuming that σ is known to be 2.9 hg.

33 28 33 37 31 32 31
28 34 28 33 26 30 31 28

Example: Confidence Interval Estimate of μ with Known σ (2 of 6)

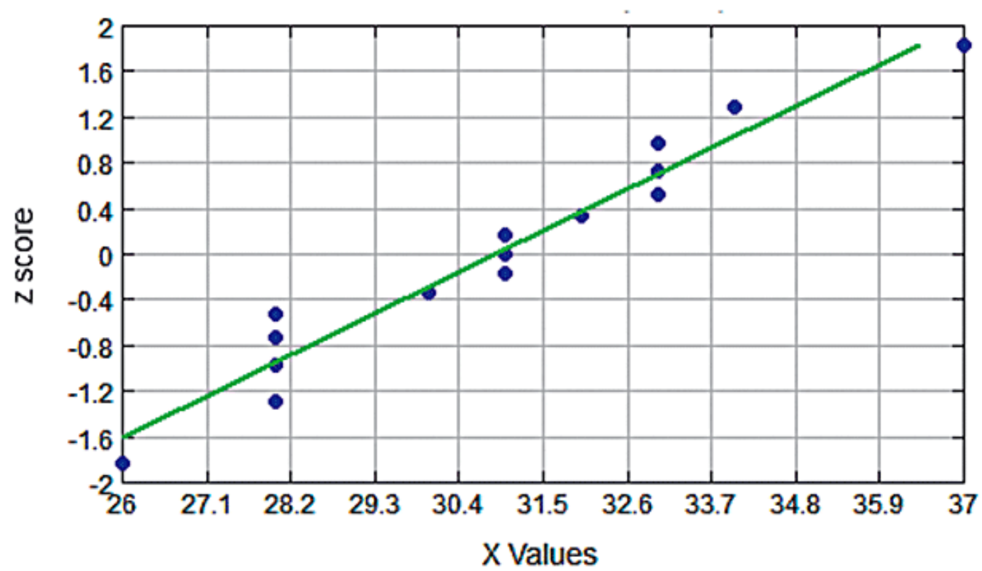
Solution

Requirement Check: (1) The sample is a simple random sample. (2) Because the sample size is $n = 15$, the requirement that "the population is normally distributed or the sample size is greater than 30" can be satisfied only if the sample data appear to be from a normally distributed population, so we need to investigate normality.

Example: Confidence Interval Estimate of μ with Known σ (3 of 6)

Solution

Sample data appear to be from a normally distributed population.



Example: Confidence Interval Estimate of μ with Known σ (4 of 6)

Solution

With a 95% confidence level, we have $\alpha = 0.5$, and we get $z_{\frac{\alpha}{2}} = 1.96$, $\sigma = 2.9$ hg, and $n = 15$, we find the value of the margin of error E:

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

$$= 1.96 \cdot \frac{2.9}{\sqrt{15}} = 1.46760$$

Example: Confidence Interval Estimate of μ with Known σ (5 of 6)

Solution

With $\bar{x} = 30.9$ hg and $E = 1.46760$, we find the 95% confidence interval as follows:

$$\bar{x} - E < \mu < \bar{x} + E$$

$$30.9 - 1.46760 < \mu < 30.9 + 1.46760$$

$$29.4 \text{ hg} < \mu < 32.4 \text{ hg}$$

(rounded to one decimal place)

Example: Confidence Interval Estimate of μ with Known σ (6 of 6)

Solution

Remember, this example illustrates the situation in which the population standard deviation σ is known, which is rare. The more realistic situation with σ unknown is considered in Part 1 of this section.

Choosing an Appropriate Distribution

Choosing between Student *t* and *z* (Normal) Distributions

Conditions	Method
σ not known and normally distributed population **or** σ not known and $n > 30$	Use student *t* distribution
σ known and normally distributed population **or** σ known and $n > 30$ (In reality, σ is rarely known.)	Use normal (*z*) distribution.
Population is not normally distributed and $n \leq 30$.	Use the bootstrapping method (Section 7-4) or a nonparametric method.