Estimating parameters and Determining sampling sizes

Glossary

1. **Point Estimate:** a single value used to estimate population parameter.
2. **Confidence Interval**: A confidence interval (or interval estimate) is a range (or an interval) of values used to estimate the true value of a population parameter. A confidence interval is
sometimes abbreviated as CI.
3. **Confidence Level:** The confidence level is the probability 1 - a (such as 0.95, or 95%) that the confidence interval actually does contain the population parameter, assuming that the
estimation process is repeated a large number of times. (The confidence level is also called the degree of confidence, or the confidence coefficient.)

Most Common Confidence Levels	Corresponding Values of α
90% (or 0.90) confidence level:	$\alpha = 0.10$
95% (or 0.95) confidence level:	$\alpha = 0.05$
99% (or 0.99) confidence level:	$\alpha = 0.01$

Here's an example of a confidence interval found later in Example 3:

The 0.95 (or 95%) confidence interval estimate of the population proportion p is $0.405 < p < 0.455$.

4. **Critical value**: A critical value is the number on the borderline separating sample statistics that are significantly high or low from those that are not significant. The number za>2 is a
critical value that is a z score with the property that it is at the border that separates an area of a>2 in the right tail of the standard normal distribution

Find the critical value $z_{\alpha/2}$ corresponding to a 95% confidence level.

SOLUTION

A 95% confidence level corresponds to $\alpha = 0.05$, so $\alpha/2 = 0.025$. Figure 7-3 shows that the area in each of the green-shaded tails is $\alpha/2 = 0.025$. We find $z_{\alpha/2} = 1.96$ by noting that the cumulative area to its left must be $1 - 0.025$, or 0.975. We can use technology or refer to Table A-2 to find that the cumulative left area of 0.9750 corresponds to $z = 1.96$. For a 95% confidence level, the critical value is therefore $z_{\alpha/2} = 1.96$.

Note that when finding the critical z score for a 95% confidence level, we use a cumulative left area of 0.9750 (*not* 0.95). Think of it this way:

This is our confidence level:		The area in *both* tails is:		The area in the *right* tail is:		The cumulative area from the left, excluding the right tail, is:
95%	→	$\boldsymbol{\alpha = 0.05}$	→	$\boldsymbol{\alpha/2 = 0.025}$	→	$\mathbf{1 - 0.025 = 0.975}$

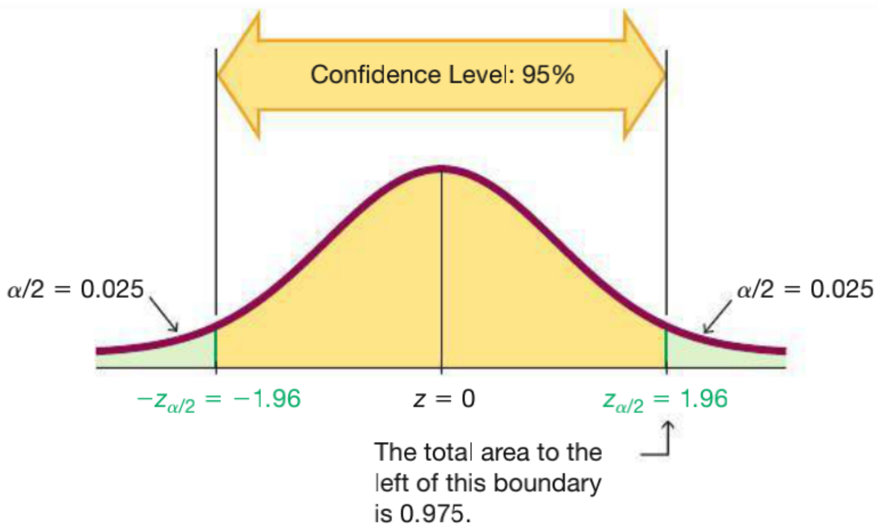

FIGURE 7-3 **Finding the Critical Value $z_{\alpha/2}$ for a 95% Confidence Level**

YOUR TURN Do Exercise 5 "Finding Critical Values."

Example 2 showed that a 95% confidence level results in a critical value of $z_{\alpha/2} = 1.96$. This is the most common critical value, and it is listed with two other common values in the table that follows.

Confidence Level	α	Critical Value, $z_{\alpha/2}$
90%	0.10	1.645
95%	0.05	1.96
99%	0.01	2.575

5. Margin of Error: When data from a simple random sample are used to estimate a population proportion p, the difference between the sample proportion pn and the population proportion
p is an error. The maximum likely amount of that error is the margin of error, denoted
by E. The margin of error E is also called the maximum error of the estimate and can be found by multiplying the critical value and the estimated standard deviation of sample proportions.

FORMULA 7-1

$$\underbrace{E = z_{\alpha/2}}_{\text{Critical value}}\underbrace{\sqrt{\frac{\hat{p}\hat{q}}{n}}}_{\text{Estimated standard deviation of sample proportions}} \quad \text{margin of error for proportions}$$

6. **Confidence Interval for Estimating a Population Proportion p**

Requirements

1. The sample is a simple random sample.
2. The conditions for the binomial distribution are satisfied: There is a fixed number of trials, the trials are independent, there are two categories of outcomes, and the probabilities remain constant for each trial (as in Section 5-2).
3. There are at least 5 successes and at least 5 failures. (This requirement is a way to verify that $np \geq 5$ and $nq \geq 5$, so the normal distribution is a suitable approximation to the binomial distribution.)

Confidence Interval Estimate of *p*

$$\hat{p} - E < p < \hat{p} + E \qquad \text{where} \qquad E = z_{\alpha/2}\sqrt{\frac{\hat{p}\hat{q}}{n}}$$

The confidence interval is often expressed in the following two equivalent formats:

$$\hat{p} \pm E \quad \text{or} \quad (\hat{p} - E, \hat{p} + E)$$

Round-Off Rule for Confidence Interval Estimates of *p*

Round the confidence interval limits for p to three significant digits.

in Internet
eys?

Capitalizing on the wide-spread use of technology and social media, there is a growing trend duct surveys using only

Procedure for Constructing a Confidence Interval for *p*

1. Verify that the requirements in the preceding Key Elements box are satisfied.
2. Use technology or Table A-2 to find the critical value $z_{\alpha/2}$ that corresponds to the desired confidence level.
3. Evaluate the margin of error $E = z_{\alpha/2}\sqrt{\hat{p}\hat{q}/n}$.
4. Using the value of the calculated margin of error E and the value of the sample proportion $\hat{p}$, find the values of the *confidence interval limits* $\hat{p} - E$ and $\hat{p} + E$. Substitute those values in the general format for the confidence interval.
5. Round the resulting confidence interval limits to three significant digits.

7. Finding the Point Estimate and E from a Confidence Interval

Point estimate of p:

$$\hat{p} = \frac{(\text{upper confidence interval limit}) + (\text{lower confidence interval limit})}{2}$$

Margin of error:

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

8. Finding the Sample Size Required to Estimate a Population Proportion

Finding the Sample Size Required to Estimate a Population Proportion

Objective

Determine how large the sample size n should be in order to estimate the population proportion p.

Notation

p = population proportion

$\hat{p}$ = sample proportion

n = number of sample values

E = desired margin of error

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

Requirements

The sample must be a simple random sample of independent sample units.

When an estimate $\hat{p}$ is known: **Formula 7-2** $n = \frac{[z_{\alpha/2}]^2 \hat{p}\hat{q}}{E^2}$

When no estimate $\hat{p}$ is known: **Formula 7-3** $n = \frac{[z_{\alpha/2}]^2 0.25}{E^2}$

If a reasonable estimate of $\hat{p}$ can be made by using previous samples, a pilot study, or someone's expert knowledge, use Formula 7-2. If nothing is known about the value of $\hat{p}$, use Formula 7-3.

Round-Off Rule for Determining Sample Size

If the computed sample size n is not a whole number, round the value of n up to the next *larger* whole number, so the sample size is sufficient instead of being slightly insufficient. For example, round 1067.11 to 1068.

9. Estimating Population Mean

Confidence Interval for Estimating a Population Mean with σ Not Known

Objective

Construct a confidence interval used to estimate a population mean.

Notation

μ = population mean

$\bar{x}$ = sample mean

s = sample standard deviation

n = number of sample values

E = margin of error

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

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

- **Margin of Error:** $E = t_{\alpha/2} \cdot \frac{s}{\sqrt{n}}$ (Use df $= n - 1$.)
- **Confidence Level:** 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, $\alpha = 0.05$.
- **Critical Value:** $t_{\alpha/2}$ is the critical t value separating an area of $\alpha/2$ 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.

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, $\bar{x}$, and s, round the confidence interval limits to the same number of decimal places used for the sample mean.

Procedure for Constructing a Confidence Interval for μ

Confidence intervals can be easily constructed with technology or they can be manually constructed by using the following procedure.

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 $t_{\alpha/2}$ that corresponds to the desired confidence level.
3. Evaluate the margin of error using $E = t_{\alpha/2} \cdot s/\sqrt{n}$.
4. Using the value of the calculated margin of error E and the value of the sample mean $\bar{x}$, substitute those values in one of the formats for the confidence interval: $\bar{x} - E < \mu < \bar{x} + E$ or $\bar{x} \pm E$ or $(\bar{x} - E, \bar{x} + E)$.
5. Round the resulting confidence interval limits as follows: With 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, 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.

10. **Sample Size required to estimate a population mean**

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.

Notation

μ = population mean

σ = population standard deviation

$\bar{x}$ = sample mean

E = desired margin of error

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

Requirement

The sample must be a simple random sample.

Sample Size

The required sample size is found by using Formula 7-4.

Formula 7-4
$$n = \left[\frac{z_{\alpha/2}\sigma}{E}\right]^2$$

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.

11.

TABLE 7-1 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.

12. Confidence Interval for Estimating a Population Standard Deviation or Variance

Confidence Interval for Estimating a Population Standard Deviation or Variance

Objective

Construct a confidence interval estimate of a population standard deviation or variance.

Notation

σ = population standard deviation

s = sample standard deviation

n = number of sample values

χ^2_L = left-tailed critical value of χ^2

σ^2 = population variance

s^2 = sample variance

E = margin of error

χ^2_R = right-tailed critical value of χ^2

Requirements

1. The sample is a simple random sample.
2. The population must have normally distributed values (even if the sample is large). The requirement of a normal distribution is much stricter here than in earlier sections, so large departures from normal distributions can result in large errors. (If the normality requirement is not satisfied, use the bootstrap method described in Section 7-4.)

Confidence Interval for the Population Variance σ^2

$$\frac{(n-1)\,s^2}{\chi^2_R} < \sigma^2 < \frac{(n-1)s^2}{\chi^2_L}$$

Confidence Interval for the Population Standard Deviation σ

$$\sqrt{\frac{(n-1)\,s^2}{\chi^2_R}} < \sigma < \sqrt{\frac{(n-1)\,s^2}{\chi^2_L}}$$

Round-Off Rule

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

Procedure for Constructing a Confidence Interval for σ or σ^2

Confidence intervals can be easily constructed with technology or they can be constructed by using Table A-4 with the following procedure.

1. Verify that the two requirements are satisfied: The sample is a random sample from a normally distributed population.

continued

2. Using $n - 1$ degrees of freedom, find the critical values χ^2_R and χ^2_L that correspond to the desired confidence level (as in Example 1).
3. To get a confidence interval estimate of σ^2, use the following:

$$\frac{(n-1)s^2}{\chi^2_R} < \sigma^2 < \frac{(n-1)s^2}{\chi^2_L}$$

4. To get a confidence interval estimate of σ, take the square root of each component of the above confidence interval.
5. Round the confidence interval limits using the round-off rule given in the preceding Key Elements box.

HEAVY USE OF STATDISK IN THIS CHAPTER