Overview

This chapter presents:

- methods for estimating population means, proportions, and variances
- methods for determining sample sizes

Assumptions

- **n > 30**

  The sample must have more than 30 values.

- **Simple Random Sample**

  All samples of the same size have an equal chance of being selected.

Assumptions

- **n > 30**

  The sample must have more than 30 values.

- **Simple Random Sample**

  All samples of the same size have an equal chance of being selected.

Data collected carelessly can be absolutely worthless, even if the sample is quite large.

Example:

In a recent poll, 70% of 1501 randomly selected adults said they believed in global warming.

Q: What is the **proportion** of the adult **population** that believe in global warming?

Notation: p is the **population proportion** (an unknown parameter).

$\hat{p}$ is the **sample proportion** (computed).
From the poll data $\hat{p}$ = 0.70.

Apparently, 0.70 will be the best estimate of the proportion of all adults who believe in global warming.

Definition

A **point estimate** is a single value (or point) used to approximate a population parameter.

Definition

The sample proportion $\hat{p}$ is the <u>best point estimate</u> of the population proportion p.

Example (continued)

- We say that 0.70, or 70% is be the best point estimate of the proportion of all adults who believe in global warming.
- But how reliable (accurate) is this estimate?
- We will see that its margin of error is 2.3%. This means the true proportion of adults who believe in global warming is between 67.7% and 72.3%. This gives an interval (from 67.7% to 72.3%) containing the true (but unknown) value of the population proportion.

Definition

A **confidence interval** (or **interval estimate**) is a range (or an interval) of values used to estimate the true value of a population parameter.

Lower # < population parameter < Upper #

A confidence interval is sometimes abbreviated as **CI**.

Confidence Intervals from 20 Different Samples

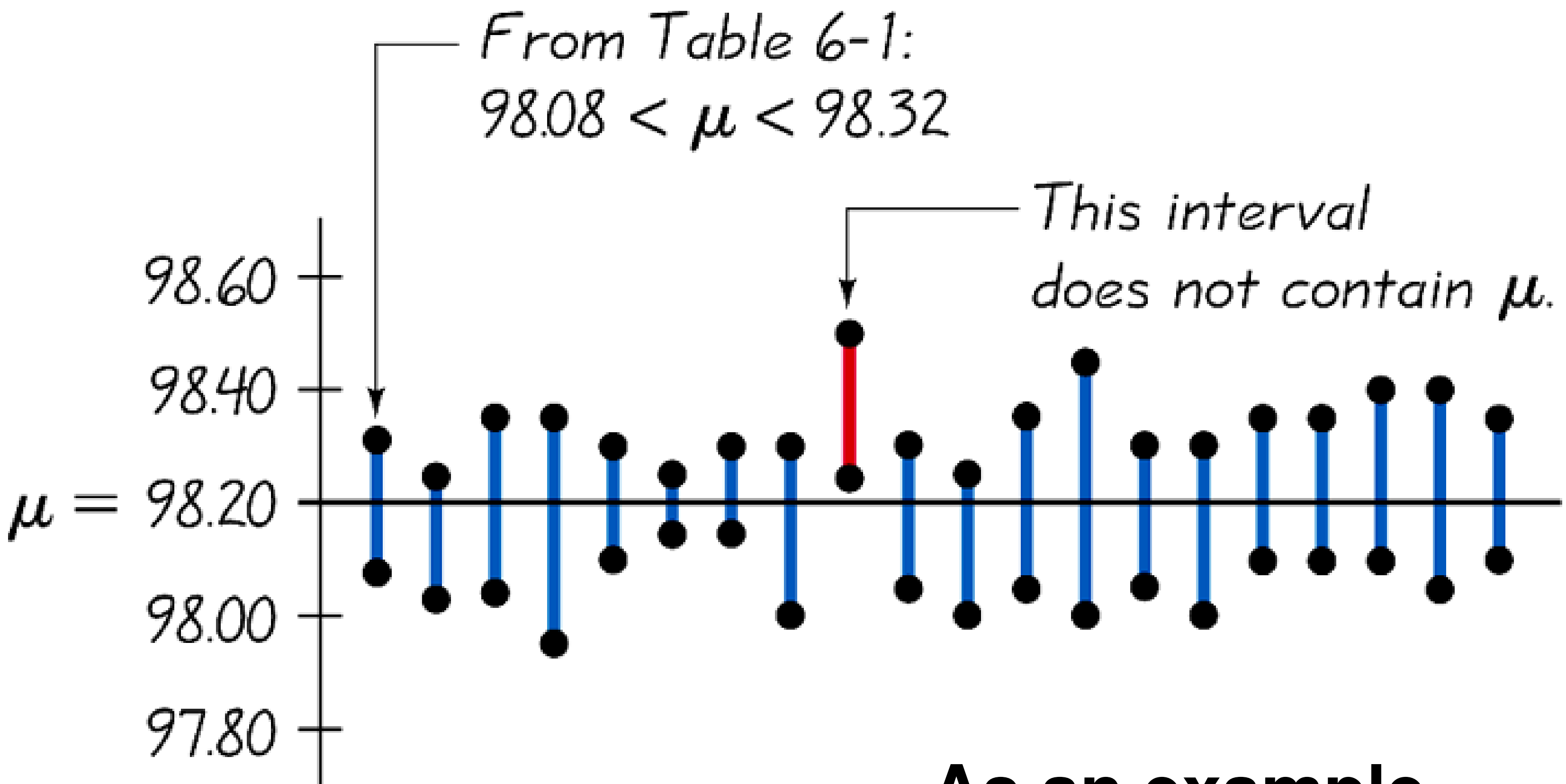


Figure 6-1

As an example
Lower # < μ < Upper #

Definition

A **confidence level** is the probability $1 - \alpha$ (often expressed as the equivalent percentage value) that the confidence interval actually does contain the population parameter.

The confidence level is also called **degree of confidence**, or the **confidence coefficient**.

Most common choices are 90%, 95%, or 99%.

(α = 10%), (α = 5%), (α = 1%)

Example (continued)

In a recent poll, 70% of 1501 randomly selected adults said they believed in global warming.

The sample proportion $\hat{p} = 0.70$ is the best estimate of the population proportion p.

A 95% confidence interval for the unknown population parameter is

$$0.677 < p < 0.723$$

What does it mean, exactly?

Interpreting a Confidence Interval

We are 95% confident that the interval from 0.677 to 0.723 actually does contain the true value of the population proportion p.

This means that if we were to select many different samples of size 1501 and construct the corresponding confidence intervals, then 95% of them would actually contain the value of the population proportion p.

Caution

Know the correct interpretation of a confidence interval.

It is wrong to say "the probability that the population parameter belongs to the confidence interval is 95%"

because the population parameter is <u>not a random variable</u>, it does not change its value.

Caution

Do not confuse two percentages: the proportion may be represented by percents (like 70% in the example), and the confidence level may be represented by percents (like 95% in the example).

Proportion may be any number from 0% to 100%.

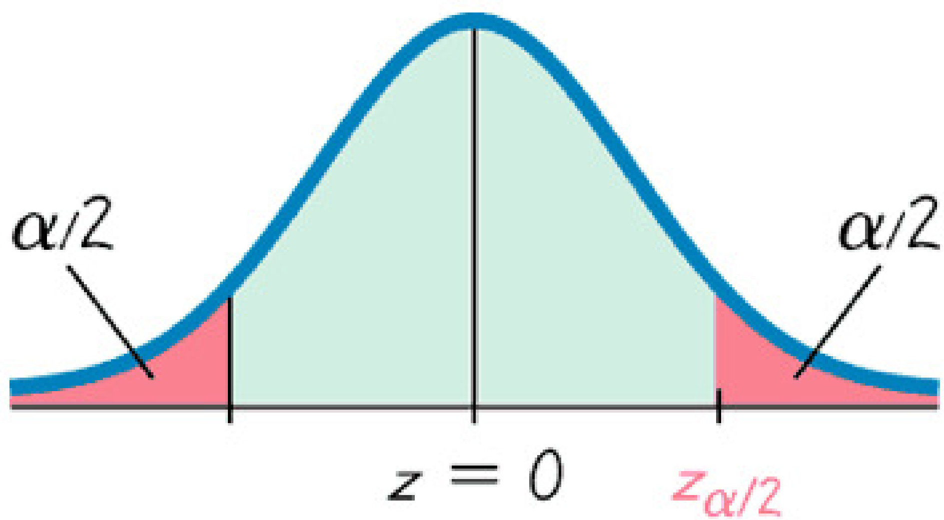
Confidence level is usually 90% or 95% or 99%.

Next we learn how to construct confidence intervals

Critical Values

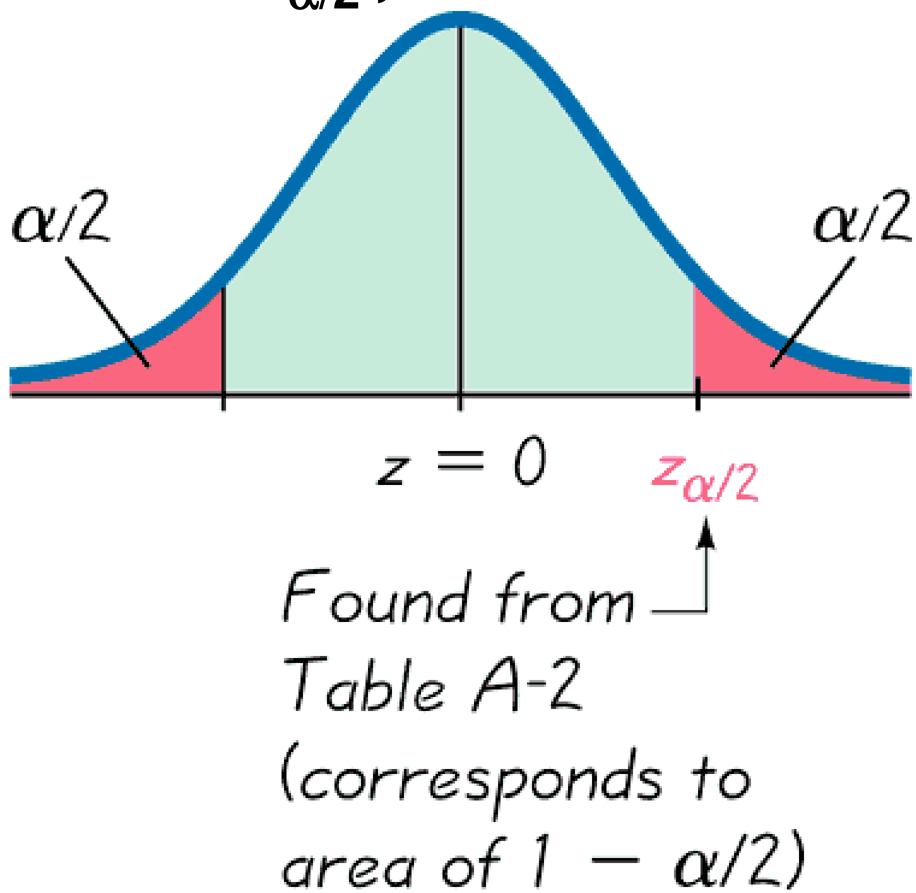
A z score can be used to distinguish between sample statistics that are likely to occur and those that are unlikely to occur. Such a z score is called a critical value.

The standard normal distribution is divided into three regions: middle part has area $1-\alpha$ and two tails (left and right) have area $\alpha/2$ each:



Critical Values

The *z* scores separate the middle interval (likely values) from the tails (unlikely values). They are $z_{\alpha/2}$ and $-z_{\alpha/2}$, found from Table A-2.



Definition

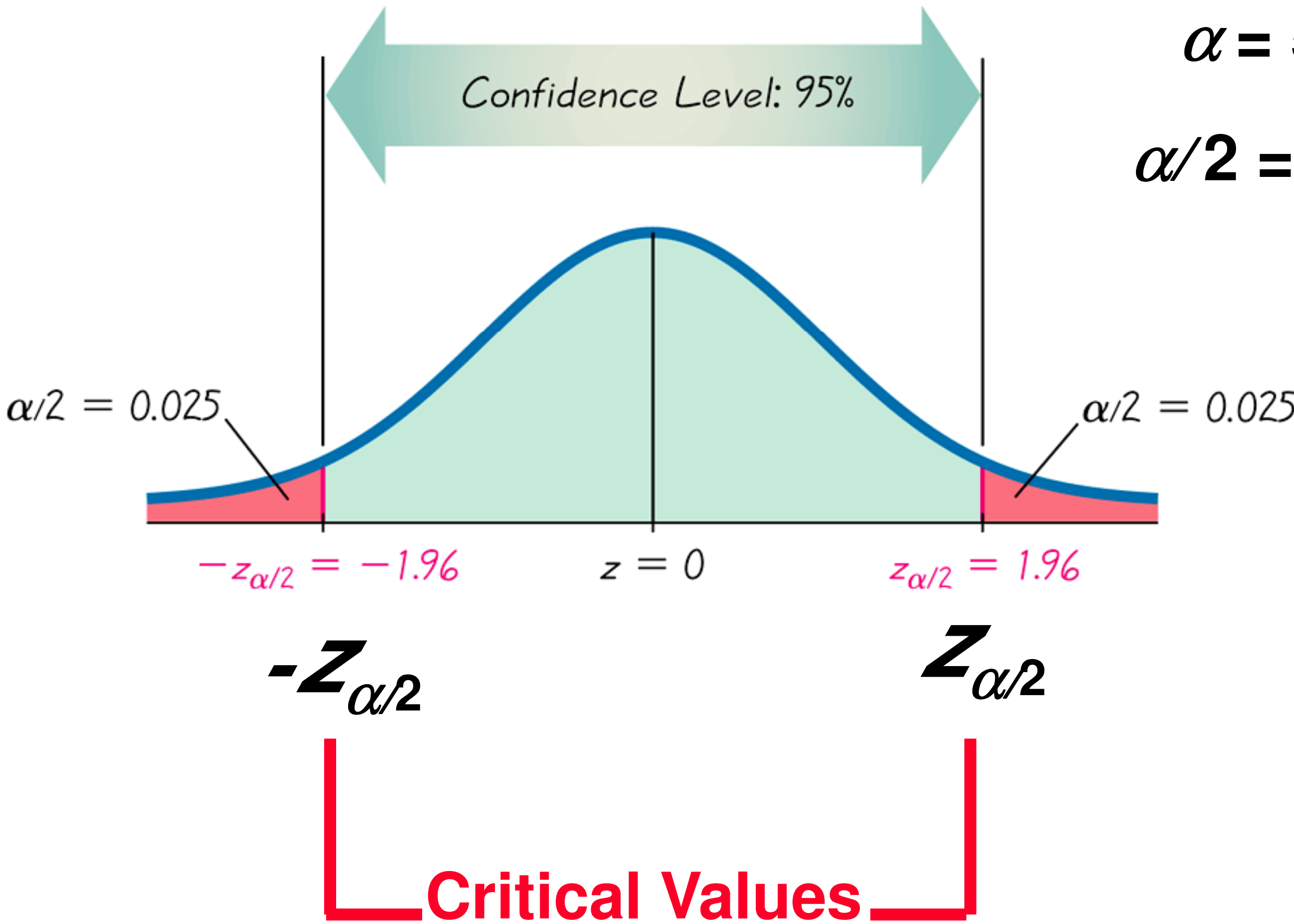
A critical value is the number on the borderline separating sample statistics that are likely to occur from those that are unlikely to occur.

Notation for Critical Value

The critical value $z_{\alpha/2}$ separates an area of $\alpha/2$ in the right tail of the standard normal distribution. The value of $-z_{\alpha/2}$ separates an area of $\alpha/2$ in the left tail.

The subscript $\alpha/2$ is simply a reminder that the z score separates an area of $\alpha/2$ in the tail.

Finding $z_{\alpha/2}$ for a 95% Confidence Level



$\alpha = 5\%$

$\alpha/2 = 2.5\% = .025$

Definition

Margin of error, denoted by E, is the maximum likely difference (with probability $1 - \alpha$, such as 0.95) between the observed proportion $\hat{p}$ and the true value of the population proportion p.

The margin of error E is also called the **maximum error of the estimate.**

Margin of Error for Proportions

$$E = z_{\alpha/2}\sqrt{\frac{\hat{p}\hat{q}}{n}}$$

Notation

E = margin of error

$\hat{p}$ = sample proportion

$\hat{q} = 1 - \hat{p}$

n = number of sample values

Confidence Interval for a Population Proportion *p*

$$\hat{p} - E < p < \hat{p} + E$$

where

$$E = z_{\alpha/2}\sqrt{\frac{\hat{p}\hat{q}}{n}}$$

Confidence Interval for a Population Proportion p

$$\hat{p} - E < p < \hat{p} + E$$

$$\hat{p} \pm E$$

$$(\hat{p} - E, \hat{p} + E)$$

Finding the Point Estimate and E from a Confidence Interval

Point estimate of $\hat{p}$:

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

Margin of Error:

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

Round-Off Rule for Confidence Interval Estimates of p

Round the confidence interval limits for p to

three significant digits.

Sample Size

Suppose we want to collect sample data in order to estimate some population proportion. The question is how many sample items must be obtained?

Determining Sample Size

$$E = z_{\alpha/2}\sqrt{\frac{\hat{p}\hat{q}}{n}}$$

(solve for *n* by algebra)

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

Sample Size for Estimating Proportion p

When an estimate of $\hat{p}$ is known:

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

When no estimate of $\hat{p}$ is known:

$$n = \frac{(z_{\alpha/2})^2 \, 0.25}{E^2}$$

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.

Examples:

n=310.67 round up to 311
n=310.23 round up to 311
n=310.01 round up to 311

Example:

A manager for E-Bay wants to determine the current percentage of U.S. adults who now use the Internet.

How many adults must be surveyed in order to be 95% confident that the sample percentage is in error by no more than three percentage points?

a) In 2006, 73% of adults used the Internet.
b) No known possible value of the proportion.

Example:

a) Use $\hat{p} = 0.73$ and $\hat{q} = 1 - \hat{p} = 0.27$

$\alpha = 0.05$ so $z_{\alpha/2} = 1.96$

$E = 0.03$

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

$$= \frac{(1.96)^2 \cdot (0.73)(0.27)}{(0.03)^2}$$

$$= 841.3104$$

$$= 842$$

To be 95% confident that our sample percentage is within three percentage points of the true percentage for all adults, we should obtain a random sample of 842 adults.

Example:

b) Use $\alpha = 0.05$ so $z_{\alpha/2} = 1.96$

$E = 0.03$

$$n = \frac{\left(z_{\alpha/2}\right)^2 \square 0.25}{E^2}$$

$$= \frac{(1.96)^2 \square 0.25}{(0.03)^2}$$

$$= 1067.1111$$

$$= 1068$$

To be 95% confident that our sample percentage is within three percentage points of the true percentage for all adults, we should obtain a random sample of 1068 adults.

Example:

We want to estimate the standard deviation σ.
We want to be 95% confident that our estimate is within 20% of the true value of σ.
How large should the sample be?
Assume that the population is normally distributed.

From Table 7-2, we can see that 95% confidence and an error of 20% for σ correspond to a sample of size 48.
We should obtain a sample of 48 values.