Chapter 7
Estimates and Sample Sizes

7-1 Review and Preview

7-2 Estimating a Population Proportion

7-3 Estimating a Population Mean: σ Known

7-4 Estimating a Population Mean: σ Not Known

7-5 Estimating a Population Variance

Section 7-1
Review and Preview

Review

❖ **Chapters 2 & 3 we used "descriptive statistics" when we summarized data using tools such as graphs, and statistics such as the mean and standard deviation.**

❖ **Chapter 6 we introduced critical values: z_α denotes the z score with an area of α to its right.**
If α = 0.025, the critical value is $z_{0.025}$ = 1.96. That is, the critical value $z_{0.025}$ = 1.96 has an area of 0.025 to its right.

Preview

This chapter presents the beginning of inferential statistics.

❖ **The two major activities of inferential statistics are (1) to use sample data to estimate values of a population parameters, and (2) to test hypotheses or claims made about population parameters.**

❖ **We introduce methods for estimating values of these important population parameters: proportions, means, and variances.**

❖ **We also present methods for determining sample sizes necessary to estimate those parameters.**

Section 7-2
Estimating a Population Proportion

Key Concept

In this section we present methods for using a sample proportion to estimate the value of a population proportion.

- The sample proportion is the best point estimate of the population proportion.

- We can use a sample proportion to construct a confidence interval to estimate the true value of a population proportion, and we should know how to interpret such confidence intervals.

- We should know how to find the sample size necessary to estimate a population proportion.

Definition

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

Definition

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

Example:

In the Chapter Problem we noted that in a Pew Research Center poll, 70% of 1501 randomly selected adults in the United States believe in global warming, so the sample proportion is $\hat{p} = 0.70$. Find the best point estimate of the proportion of all adults in the United States who believe in global warming.

Because the sample proportion is the best point estimate of the population proportion, we conclude that the best point estimate of p is 0.70. When using the sample results to estimate the percentage of all adults in the United States who believe in global warming, the best estimate is 70%.

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. A confidence interval is sometimes abbreviated as CI.

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, assuming that the estimation process is repeated a large number of times. (The confidence level is also called **degree of confidence**, or the **confidence coefficient**.)

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

($\alpha = 10\%$), ($\alpha = 5\%$), ($\alpha = 1\%$)

Interpreting a Confidence Interval

We must be careful to interpret confidence intervals correctly. There is a correct interpretation and many different and creative incorrect interpretations of the confidence interval $0.677 < p < 0.723$.

"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, 95% of them would actually contain the value of the population proportion p.

(Note that in this correct interpretation, the level of 95% refers to the success rate of the process being used to estimate the proportion.)

Caution

Know the correct interpretation of a confidence interval.

Caution

Confidence intervals can be used informally to compare different data sets, *but the overlapping of confidence intervals should not be used for making formal and final conclusions about equality of proportions*.
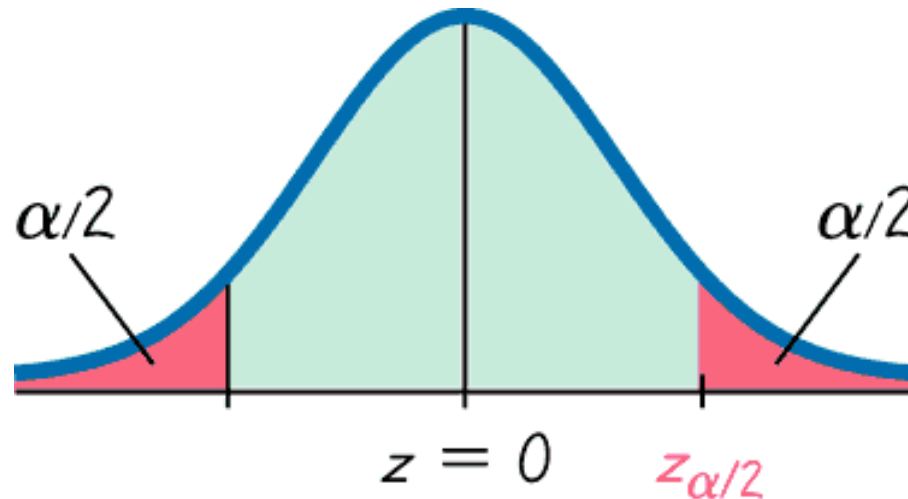
Critical Values

A standard *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. Critical values are based on the following observations:

1. Under certain conditions, the sampling distribution of sample proportions can be approximated by a normal distribution.
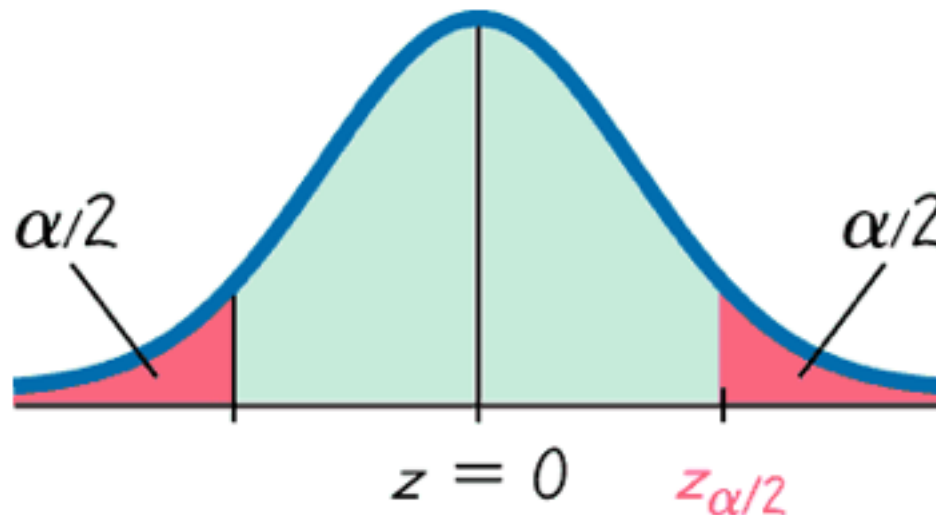
Critical Values

2. A *z* score associated with a sample proportion has a probability of $\alpha/2$ of falling in the right tail.

$\alpha/2$ $\alpha/2$

$z = 0$ $z_{\alpha/2}$

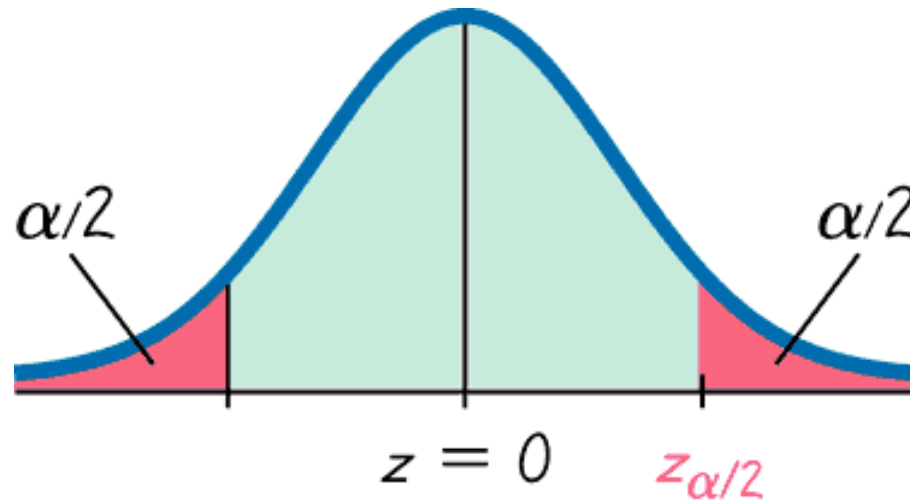Found from Table A-2 (corresponds to area of $1 - \alpha/2$)

Critical Values

3. The *z* score separating the right-tail region is commonly denoted by $z_{\alpha/2}$ and is referred to as a **critical value** because it is on the borderline separating *z* scores from sample proportions that are likely to occur from those that are unlikely to occur.

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. The number $z_{\alpha/2}$ is a critical value that is a *z* score with the property that it separates an area of $\alpha/2$ in the right tail of the standard normal distribution.

The Critical Value $Z_{\alpha/2}$

$\alpha/2$ $\alpha/2$

$z = 0$ $z_{\alpha/2}$

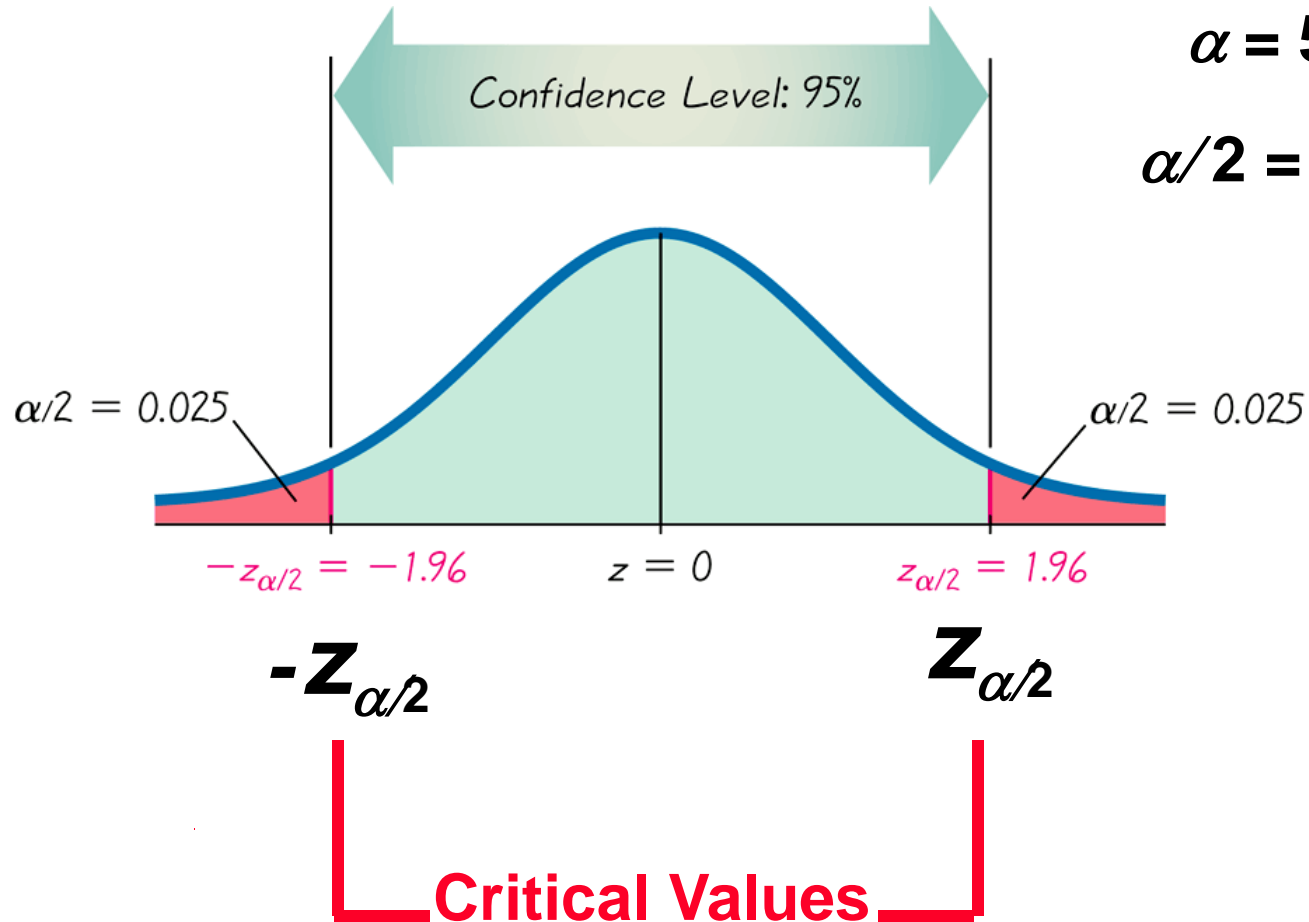Found from Table A-2 (corresponds to area of $1 - \alpha/2$)

Notation for Critical Value

The critical value $z_{\alpha/2}$ is the positive z value that is at the vertical boundary separating an area of $\alpha/2$ in the right tail of the standard normal distribution. (The value of $-z_{\alpha/2}$ is at the vertical boundary for the 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 right tail of the standard normal distribution.
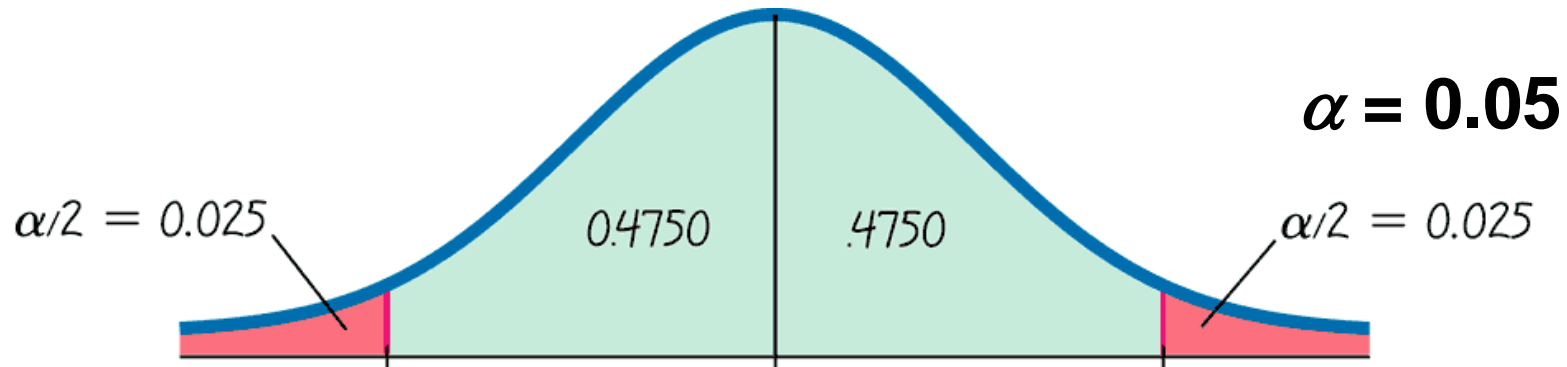
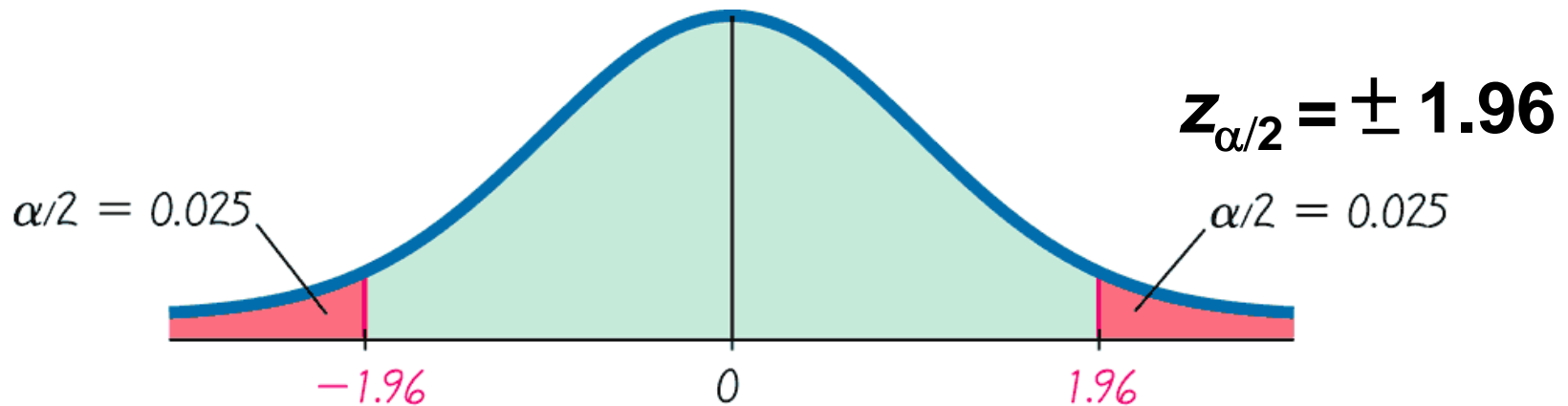# Finding $z_{\alpha/2}$ for a 95% Confidence Level

$\alpha = 5\%$

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

Confidence Level: 95%

$\alpha/2 = 0.025$

$\alpha/2 = 0.025$

$-z_{\alpha/2} = -1.96$

$z = 0$

$z_{\alpha/2} = 1.96$

$-z_{\alpha/2}$

$z_{\alpha/2}$

Critical Values

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

$\alpha = 0.05$

$\alpha/2 = 0.025$ 0.4750   .4750 $\alpha/2 = 0.025$

Use Table A-2 to find a z score of 1.96

$z_{\alpha/2} = \pm 1.96$

$\alpha/2 = 0.025$ $\alpha/2 = 0.025$

-1.96 0 1.96

Definition

When data from a simple random sample are used to estimate a population proportion p, the **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** and can be found by multiplying the critical value and the standard deviation of the sample proportions:

Margin of Error for Proportions

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

Confidence Interval for Estimating a Population Proportion *p*

p = **population proportion**

$\hat{p}$ = **sample proportion**

n = **number of sample values**

E = **margin of error**

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

Confidence Interval for Estimating a Population Proportion *p*

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.**

3. **There are at least 5 successes and 5 failures.**

Confidence Interval for Estimating 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 Estimating a Population Proportion *p*

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

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

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

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

Round the confidence interval limits for *p* to

three significant digits.

Procedure for Constructing a Confidence Interval for *p*

1. **Verify that the required assumptions are satisfied. (The sample is a simple random sample, the conditions for the binomial distribution are satisfied, and the normal distribution can be used to approximate the distribution of sample proportions because $np \geq 5$, and $nq \geq 5$ are both satisfied.)**

2. **Refer to Table A-2 and find the critical value $z_{\alpha/2}$ that corresponds to the desired confidence level.**

3. **Evaluate the margin of error** $E = z_{\alpha/2}\sqrt{\hat{p}\hat{q}/n}$

Procedure for Constructing a Confidence Interval for *p* - cont

4. Using the value of the calculated margin of error, *E* and the value of the sample proportion, $\hat{p}$, find the values of $\hat{p} - E$ and $\hat{p} + E$. Substitute those values in the general format for the confidence interval:

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

5. Round the resulting confidence interval limits to three significant digits.

Example:

In the Chapter Problem we noted that a Pew Research Center poll of 1501 randomly selected U.S. adults showed that 70% of the respondents believe in global warming. The sample results are *n* = 1501, and $\hat{p} = 0.70$

a. Find the margin of error E that corresponds to a 95% confidence level.

b. Find the 95% confidence interval estimate of the population proportion *p*.

c. Based on the results, can we safely conclude that the majority of adults believe in global warming?

d. Assuming that you are a newspaper reporter, write a brief statement that accurately describes the results and includes all of the relevant information.

Example:

Requirement check: simple random sample; fixed number of trials, 1501; trials are independent; two categories of outcomes (believes or does not); probability remains constant. Note: number of successes and failures are both at least 5.

a) Use the formula to find the margin of error.

$$E = z_{\alpha/2}\sqrt{\frac{\hat{p}\hat{q}}{n}} = 1.96\sqrt{\frac{(0.70)(0.30)}{1501}}$$

$$E = 0.023183$$

Example:

b) The 95% confidence interval:

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

$$0.70 - 0.023183 < p < 0.70 + 0.023183$$

$$0.677 < p < 0.723$$

Example:

c) Based on the confidence interval obtained in part (b), it does appear that the proportion of adults who believe in global warming is greater than 0.5 (or 50%), so we can safely conclude that the majority of adults believe in global warming. Because the limits of 0.677 and 0.723 are likely to contain the true population proportion, it appears that the population proportion is a value greater than 0.5.

Example:

d) Here is one statement that summarizes the results: 70% of United States adults believe that the earth is getting warmer. That percentage is based on a Pew Research Center poll of 1501 randomly selected adults in the United States. In theory, in 95% of such polls, the percentage should differ by no more than 2.3 percentage points in either direction from the percentage that would be found by interviewing all adults in the United States.

Analyzing Polls

When analyzing polls consider:

1. **The sample should be a simple random sample, not an inappropriate sample (such as a voluntary response sample).**

2. **The confidence level should be provided. (It is often 95%, but media reports often neglect to identify it.)**

3. **The sample size should be provided. (It is usually provided by the media, but not always.)**

4. **Except for relatively rare cases, the quality of the poll results depends on the sampling method and the size of the sample, but the size of the population is usually not a factor.**

Caution

Never follow the common misconception that poll results are unreliable if the sample size is a small percentage of the population size. The population size is usually not a factor in determining the reliability of a poll.

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

$$E = Z_{\alpha/2} \sqrt{\dfrac{\hat{p}\,\hat{q}}{n}}$$
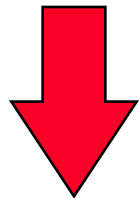
⬇ **(solve for *n* by algebra)**

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

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.

Example:

The Internet is affecting us all in many different ways, so there are many reasons for estimating the proportion of adults who use it. Assume that 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 simple random sample of 842 adults.

Example:

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

$$E = 0.03$$

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

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

$$= 1067.1111$$

$$= 1068$$

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

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:

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

Recap

In this section we have discussed:

- ❖ **Point estimates.**

- ❖ **Confidence intervals.**

- ❖ **Confidence levels.**

- ❖ **Critical values.**

- ❖ **Margin of error.**

- ❖ **Determining sample sizes.**