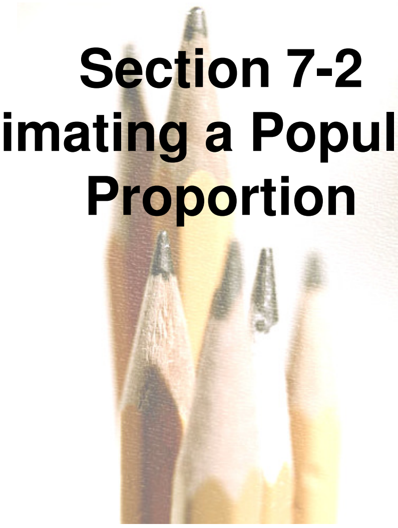
Section 7-2
Estimating a Population Proportion



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 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.

Example

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

$$0.677 < p < 0.723$$

"will know later where these numbers come from"

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%.

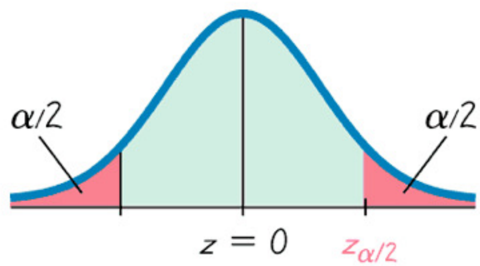
Confidence level is usually 90% or 95% or 99%.

Next we learn how to construct confidence intervals

Critical Values

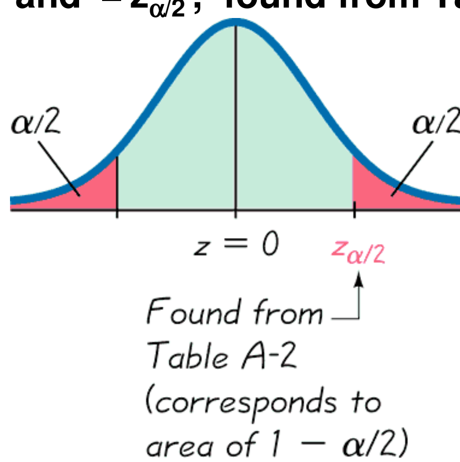
A z score can be used to distinguish between sample statistics that are likely to occur and those that are unlikely to occur. Such a z score is called a **critical value**.

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

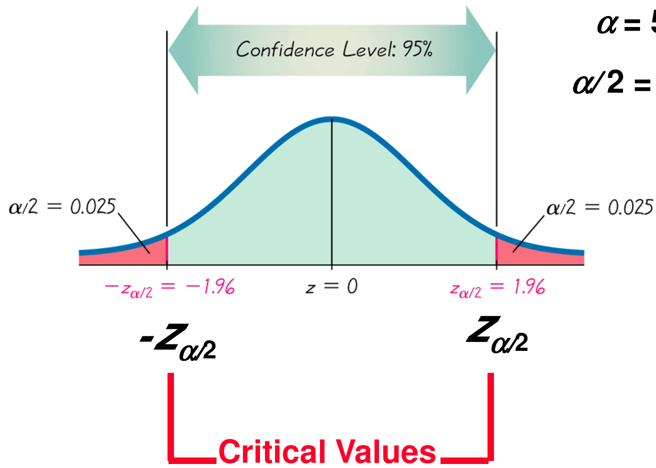
A critical value is the number on the borderline separating sample statistics that are likely to occur from those that are unlikely to occur.

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



α = 5%

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

Definition

Margin of error, denoted by E, is the maximum likely difference (with probability 1 – α, 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

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%. $0.677 < p < 0.723$ .This gives an interval (from 67.7% to 72.3%) containing the true (but unknown) value of the population proportion.

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 $z_{\alpha/2} = 1.96$

$E = 0.03$

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

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

$$= 841.3104$$

$$= 842$$

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 $z_{\alpha/2} = 1.96$

$E = 0.03$

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

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

$$= 1067.1111$$

$$= 1068$$

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

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

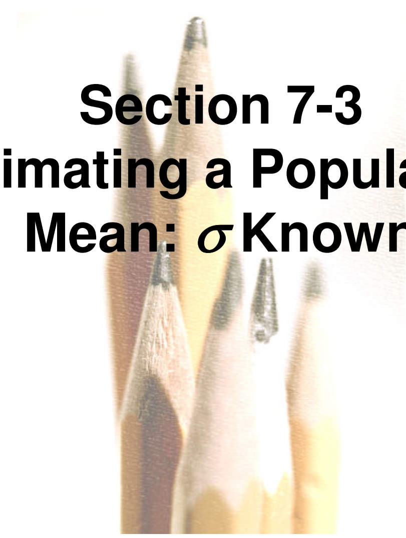
Never follow the common misconception that poll results are unreliable if the sample size is a small percentage of the population size. The population size is usually not a factor in determining the reliability of a poll.

Recap

In this section we have discussed:

- Point estimates.
- Confidence intervals.
- Confidence levels.
- Critical values.
- Margin of error.
- Determining sample sizes.

Section 7-3
Estimating a Population Mean: σ Known



Key Concept

This section presents methods for estimating a population mean. In addition to knowing the values of the sample data or statistics, we must also know the value of the population standard deviation, σ.

Here are three key concepts that should be learned in this section:

Key Concept

1. We should know that the sample mean $\bar{x}$ is the best point estimate of the population mean μ.
2. We should learn how to use sample data to construct a confidence interval for estimating the value of a population mean, and we should know how to interpret such confidence intervals.
3. We should develop the ability to determine the sample size necessary to estimate a population mean.

Point Estimate of the Population Mean

The sample mean $\bar{x}$ is the best point estimate of the population mean μ.

Confidence Interval for Estimating a Population Mean (with σ Known)

μ = population mean

σ = population standard deviation

$\bar{x}$ = sample mean

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 Mean (with σ Known)

1. The sample is a simple random sample. (All samples of the same size have an equal chance of being selected.)

2. The value of the population standard deviation σ is known.

3. Either or both of these conditions is satisfied: The population is normally distributed or $n > 30$.

Confidence Interval for Estimating a Population Mean (with σ Known)

$$\bar{x} - E < \mu < \bar{x} + E \quad \text{where} \quad E = z_{\alpha/2} \cdot \frac{\sigma}{\sqrt{n}}$$

or $\bar{x} \pm E$

or $(\bar{x} - E, \bar{x} + E)$

Definition

The two values $\bar{x} - E$ and $\bar{x} + E$ are called **confidence interval limits**.

Sample Mean

1. For all populations, the sample mean $\bar{x}$ is an **unbiased estimator** of the population mean μ, meaning that the distribution of sample means tends to center about the value of the population mean μ.

2. For many populations, the distribution of sample means $\bar{x}$ tends to be more consistent (with **less variation**) than the distributions of other sample statistics.

Procedure for Constructing a Confidence Interval for μ (with Known σ)

1. Verify that the requirements are satisfied.
2. Refer to Table A-2 to find the critical value $z_{\alpha/2}$ that corresponds to the desired confidence level.
3. Evaluate the margin of error $E = z_{\alpha/2} \cdot \sigma/\sqrt{n}$
4. Find the values of $\bar{x} - E$ and $\bar{x} + E$. Substitute those values in the general format of the confidence interval:

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

5. Round using the confidence intervals round-off rules.

Round-Off Rule for Confidence Intervals Used to Estimate μ

1. When using the original set of data, round the confidence interval limits to one more decimal place than used in original set of data.
2. When the original set of data is unknown and only the summary statistics (n, $\bar{x}$, s) are used, round the confidence interval limits to the same number of decimal places used for the sample mean.

Example: page 347

People have died in boat and aircraft accidents because an obsolete estimate of the mean weight of men was used. In recent decades, the mean weight of men has increased considerably, so we need to update our estimate of that mean so that boats, aircraft, elevators, and other such devices do not become dangerously overloaded.

Using the weights of men from Data Set 1 in Appendix B, we obtain these sample statistics for the simple random sample: $n = 40$ and $\bar{x}$ = 172.55 lb. Research from several other sources suggests that the population of weights of men has a standard deviation given by $\sigma = 26$ lb.

Example:

a. Find the best point estimate of the mean weight of the population of all men.

b. Construct a 95% confidence interval estimate of the mean weight of all men.

c. What do the results suggest about the mean weight of 166.3 lb that was used to determine the safe passenger capacity of water vessels in 1960 (as given in the National Transportation and Safety Board safety recommendation M-04-04)?

Example page 347

Since :

1- Sample is random

2- the value of σ is known

3 – n >30

i.e. Conditions are satisfied

Then:

a) The sample mean 172.55 is the best point estimate for the population mean

Example:

a. The sample mean of 172.55 lb is the best point estimate of the mean weight of the population of all men.

b. A 95% confidence interval or 0.95 implies σ = 0.05, so $z_{\alpha/2}$ = 1.96.
Calculate the margin of error.

$$E = z_{\alpha/2} \cdot \frac{\sigma}{\sqrt{n}} = 1.96 \cdot \frac{26}{\sqrt{40}} = 8.0574835$$

Construct the confidence interval.

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

$$172.55 - 8.0574835 < \mu < 172.55 + 8.0574835$$

$$164.49 < \mu < 180.61$$

Example:

c. Based on the confidence interval, it is possible that the mean weight of 166.3 lb used in 1960 could be the mean weight of men today.

However, the best point estimate of 172.55 lb suggests that the mean weight of men is now considerably greater than 166.3 lb.

Considering that an underestimate of the mean weight of men could result in lives lost through overloaded boats and aircraft, the mean weight of men has been increased.

Finding a Sample Size for Estimating a Population Mean

μ = population mean

σ = population standard deviation

$\bar{x}$ = population standard deviation

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

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

Round-Off Rule for Sample Size *n*

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

Example:

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?

$\alpha = 0.05$
$\alpha/2 = 0.025$
$z_{\alpha/2} = 1.96$
$E = 3$
$\sigma = 15$

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

$$n = \left[\frac{1.96 \cdot 15}{3}\right]^2 = 96.04 = 97$$

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

Recap

In this section we have discussed:

- Margin of error.
- Confidence interval estimate of the population mean with σ known.
- Round off rules.
- Sample size for estimating the mean μ.