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 \( p \).

**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%. \((\alpha = 0.10), (\alpha = 0.05), (\alpha = 0.01)\)
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 0.828 < \( p \) < 0.872.

“We are 95% confident that the interval from 0.828 to 0.872 actually does contain the true value of the population proportion \( p \).”

This means that if we were to select many different samples of size 1007 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.

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.

Using Confidence Intervals for Hypothesis Tests

A confidence interval can be used to test some claim made about a population proportion \( p \).

For now, we do not yet use a formal method of hypothesis testing, so we simply generate a confidence interval and make an informal judgment based on the result.

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

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.

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.

![Diagram of the standard normal distribution with critical value and shaded area](image-url)
Finding $z_{\alpha/2}$ for a 95% Confidence Level

Common Critical Values

<table>
<thead>
<tr>
<th>Confidence Level</th>
<th>$\alpha$</th>
<th>Critical Value, $z_{\alpha/2}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>90%</td>
<td>0.10</td>
<td>1.645</td>
</tr>
<tr>
<td>95%</td>
<td>0.05</td>
<td>1.96</td>
</tr>
<tr>
<td>99%</td>
<td>0.01</td>
<td>2.575</td>
</tr>
</tbody>
</table>

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 $p$.

Margin of Error for Proportions

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:

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

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

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

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

Example

In the Chapter Problem we noted that a Pew Research Center poll of 1007 randomly selected adults showed that 85% of respondents know what Twitter is. The sample results are $n = 1007$ 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 more than 75% of adults know what Twitter is?

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 - Continued**

**Requirement check**: simple random sample; fixed number of trials, 1007; trials are independent; two outcomes per trial; probability remains constant. Note: number of successes and failures are both at least 5.

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

\[E = z_{\alpha/2} \sqrt{\frac{\hat{p} \hat{q}}{n}} = 1.96 \sqrt{\frac{(0.85)(0.15)}{1007}}\]

\[E = 0.0220545\]

**Example - Continued**

b) The 95% confidence interval:

\[\hat{p} - E < p < \hat{p} + E\]

\[0.85 - 0.0220545 < p < 0.85 + 0.0220545\]

\[0.828 < p < 0.872\]

**Example - Continued**

c) Based on the confidence interval obtained in part (b), it does appear that more than 75% of adults know what Twitter is.

Because the limits of 0.828 and 0.872 are likely to contain the true population proportion, it appears that the population proportion is a value greater than 0.75.

**Example - Continued**

d) Here is one statement that summarizes the results:

85% of U.S. adults know what Twitter is. That percentage is based on a Pew Research Center poll of 1007 randomly selected adults.

In theory, in 95% of such polls, the percentage should differ by no more than 2.2 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

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

(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.

Example

Many companies are interested in knowing the percentage of adults who buy clothing online. 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) Use a recent result from the Census Bureau: 66% of adults buy clothing online.

b) Assume that we have no prior information suggesting a possible value of the proportion.

Example - Continued

a) Use \( \hat{p} = 0.66 \) and \( \hat{q} = 1 - \hat{p} = 0.34 \)
\( \alpha = 0.05 \) so \( z_{\alpha/2} = 1.96 \)
\( E = 0.03 \)

\[ n = \frac{(z_{\alpha/2})^2 \hat{p} \hat{q}}{E^2} \]
\[ = \frac{(1.96)^2 (0.66)(0.34)}{(0.03)^2} \]
\[ = 957.839 \]
\[ = 958 \]

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 958 adults.
Example - Continued

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

$E = 0.03$

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.

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

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

$= 1067.1111$

$= 1068$

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:

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