Chapter 7
Estimates and Sample Sizes

Overview
This chapter presents the beginning of inferential statistics using sample data to:

1. estimate a population parameter
2. test a claim (hypothesis) about a population

7-1 Overview
7-2 Estimating a Population Proportion
7-4 Estimating a Population Mean: \( \sigma \) Not Known
7-2 and 7-3: Determining Sample Size to Estimate \( p \) and \( \mu \)
Overview
This chapter presents methods for:

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

7-2
Estimating a Population Proportion

Assumptions
1. The sample is a simple random sample.
2. The conditions for the binomial distribution are satisfied (See Section 5-3.)
3. The normal distribution can be used to approximate the distribution of sample proportions because \( np \geq 5 \) and \( nq \geq 5 \) are both satisfied.
Data collected carelessly can be absolutely worthless, even if the sample is quite large.

Example: In the Chapter Problem, we noted that 829 adult Minnesotans were surveyed, and 51% of them are opposed to the use of the photo-cop for issuing traffic tickets. Use these survey results.

Find the 95% confidence interval for the population proportion \( p \).

\[ 0.476 < p < 0.544 \]

Notation for Proportions
Notation for Proportions

\[ p = \text{population proportion} \]

\[ \hat{p} = \frac{x}{n} \]

(sample proportion of \( x \) successes in a sample of size \( n \))

(pronounced 'p-hat')

\[ \hat{q} = 1 - \hat{p} = \text{sample proportion of } x \text{ failures in a sample size of } n \]
Definition
Point Estimate

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

Definition
Point Estimate

The sample proportion \( \hat{p} \) is the best point estimate of the population proportion \( p \).
Definition
Confidence Interval
(or Interval Estimate)

a range (or an interval) of values used to estimate the true value of the population parameter

Lower # < population parameter < Upper #
Definition
Confidence Interval
(or Interval Estimate)

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

Lower # < population parameter < Upper #

As an example
Lower # < p < Upper #

Definition
Confidence Level
(degree of confidence or confidence coefficient)

As an example
0.476 < p < 0.544
**Definition**

**Confidence Level**
(degree of confidence or confidence coefficient)

the probability $1 - \alpha$ (often expressed as the equivalent percentage value) that is the proportion of times the confidence interval actually does contain the population parameter, assuming that the estimation process is repeated a large number of times

usually 90%, 95%, or 99%
($\alpha = 10\%$), ($\alpha = 5\%$), ($\alpha = 1\%$)

**Interpreting a Confidence Interval**

$0.476 < p < 0.544$

Correct: We are 95% confident that the interval from 0.476 to 0.544 actually does contain the true value of $p$. This means that if we were to select many different samples of size 829 and construct the confidence intervals, 95% of them would actually contain the value of the population proportion $p$.

Wrong: There is a 95% chance that the true value of $p$ will fall between 0.476 and 0.544.
Finding Critical Values

Critical Value Observations

1. The sampling distribution of sample proportions can be approximated by a normal distribution.
2. Sample proportions have a relatively small chance (denoted by $\alpha$) of falling in one of the red tails.
3. Denoting the area of each shaded region by $\alpha/2$, there is a total probability of $\alpha$ that a sample proportion will fall in either of the two red tails.
4. By the rule of complements, there is a probability of 1 - \( \alpha \) that the sample proportion will fall within the green-shaded region.

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

(The value of \( -z_{\alpha/2} \) is at the vertical boundary for area \( \alpha/2 \) in the left tail.)
Definition

Critical Value

the number on the borderline separating sample statistics that are likely to occur from those that are unlikely to occur. The number $z_{a/2}$ is a critical value that is a $z$ score with the property that it separates an area $\alpha/2$ in the right tail of the standard normal distribution.

Finding $\pm z_{a/2}$ for 95% Degree of Confidence

Critical Values

$\alpha = 5\%$
$\alpha/2 = 2.5\% = .025$

NEGATIVE Z Scores

<table>
<thead>
<tr>
<th>Table A-2</th>
<th>Standard Normal (z) Distribution: Cumulative area from the LFT</th>
</tr>
</thead>
<tbody>
<tr>
<td>$z$</td>
<td>$00$</td>
</tr>
<tr>
<td>$-3.0$</td>
<td>.001</td>
</tr>
<tr>
<td>$-2.9$</td>
<td>.001</td>
</tr>
<tr>
<td>$-2.8$</td>
<td>.001</td>
</tr>
<tr>
<td>$-2.7$</td>
<td>.001</td>
</tr>
<tr>
<td>$-2.6$</td>
<td>.001</td>
</tr>
<tr>
<td>$-2.5$</td>
<td>.001</td>
</tr>
<tr>
<td>$-2.4$</td>
<td>.001</td>
</tr>
<tr>
<td>$-2.3$</td>
<td>.001</td>
</tr>
<tr>
<td>$-2.2$</td>
<td>.001</td>
</tr>
<tr>
<td>$-2.1$</td>
<td>.001</td>
</tr>
<tr>
<td>$-2.0$</td>
<td>.001</td>
</tr>
<tr>
<td>$-1.9$</td>
<td>.001</td>
</tr>
<tr>
<td>$-1.8$</td>
<td>.001</td>
</tr>
<tr>
<td>$-1.7$</td>
<td>.001</td>
</tr>
<tr>
<td>$-1.6$</td>
<td>.001</td>
</tr>
<tr>
<td>$-1.5$</td>
<td>.001</td>
</tr>
</tbody>
</table>
Finding $\pm z_{\alpha/2}$ for 95% Degree of Confidence

$\alpha = 5\%$

$\alpha/2 = 2.5\% = 0.025$

Critical Values

Area $= 1 - 0.025 = 0.975$

$z = 1.96$
Definition

Margin of Error

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 (with probability \( 1 - \alpha \)) difference between the observed proportion \( \hat{p} \) and the true value of the population proportion \( p \).

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

lower limit upper limit

Margin of Error of the Estimate of \( p \)

Formula 7-1

\[
E = z_{\alpha/2} \sqrt{\frac{\hat{p} \cdot q}{n}}
\]
Confidence Interval for Population Proportion

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

where

\[ E = z_{\alpha/2} \sqrt{\frac{\hat{p}(1 - \hat{p})}{n}} \]

Confidence Interval for Population Proportion

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

\[ p = \hat{p} \pm E \]

Confidence Interval for Population Proportion

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

\[ p = \hat{p} \pm E \]

\[ (\hat{p} - E, \hat{p} + E) \]
Round-Off Rule for Confidence Interval Estimates of $p$

Round the confidence interval limits 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 $E = \sqrt{\hat{p}(1-\hat{p})/n}$.

4. Using 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.

Procedure for Constructing a Confidence Interval for $p$
Example: Given the example that 829 adult Minnesotans were surveyed, and 51% of them are opposed to the use of the photo-cop for issuing traffic tickets. Use these survey results.

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 adult Minnesotans oppose the use of the photo-cop?

First, we check for assumptions. We note that $np = 422.79 \geq 5$, and $nq = 406.21 \geq 5$.

Next, we calculate the margin of error. We have found that $\hat{p} = 0.51$, $q = 1 - 0.51 = 0.49$, $z_{0.025} = 1.96$, and $n = 829$.

$$E = z_{0.025} \sqrt{\frac{\hat{p} \cdot q}{n}}$$

$$E = 1.96 \sqrt{\frac{0.51 \cdot 0.49}{829}} = 0.03403$$
Example: Given the example that 829 adult Minnesotans were surveyed, and 51% of them are opposed to the use of the photo-cop for issuing traffic tickets. Use these survey results.

b) Find the 95% confidence interval for the population proportion $p$.

We substitute our values from Part a into:

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

Based on the results, can we safely conclude that the majority of adult Minnesotans oppose use of the photo-cop?

Based on the survey results, we are 95% confident that the limits of 47.6% and 54.4% contain the true percentage of adult Minnesotans opposed to the photo-cop. The percentage of opposed adult Minnesotans is likely to be any value between 47.6% and 54.4%. However, a majority requires a percentage greater than 50%, so we cannot safely conclude that the majority is opposed (because the entire confidence interval is not greater than 50%).
Example: In a survey of 1002 people, 701 said that they voted in a recent presidential election (based on data from ICR Research Group). Voting records show that 61% of eligible voters actually did vote.

a) Find a 99% confidence interval estimate of the proportion of people who say they voted.

b) Are the survey results consistent with the actual voter turnout or 61%? Why or why not?