Chapter 7
Estimates and Sample Sizes

7-1 Overview
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

Overview
This chapter presents the beginning of inferential statistics.

- The two major applications of inferential statistics involve the use of sample data to (1) estimate the value of a population parameter, and (2) test some claim (or hypothesis) about a population.
- 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 important methods for using a sample proportion to estimate the value of a population proportion with a confidence interval. We also present methods for finding the size of the sample needed to estimate a population proportion.

Requirements for Estimating a Population Proportion

1. The sample is a simple random sample.
2. The conditions for the binomial distribution are satisfied. (See Section 5-3.)
3. There are at least 5 successes and 5 failures.

Notation for Proportions

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

\[ \hat{p} = \frac{x}{n} \]  
\[ \text{(sample proportion of } x \text{ successes in a sample of size } n) \]

\[ \hat{q} = 1 - \hat{p} \]  
\[ \text{(sample proportion of failures in a sample size of } n) \]

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: 829 adult Minnesotans were surveyed, and 51% of them are opposed to the use of the photo-cop for issuing traffic tickets. Using these survey results, find the best point estimate of the proportion of all adult Minnesotans opposed to photo-cop use.

Because the sample proportion is the best point estimate of the population proportion, we conclude that the best point estimate of \( p \) is 0.51. When using the survey results to estimate the percentage of all adult Minnesotans that are opposed to photo-cop use, our best estimate is 51%.

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 is the proportion of times 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%. 
\( (\alpha = 10\%), (\alpha = 5\%), (\alpha = 1\%) \)

Example: 829 adult Minnesotans were surveyed, and 51% of them are opposed to the use of the photo-cop for issuing traffic tickets. Using these survey results, find the 95% confidence interval of the proportion of all adult Minnesotans opposed to photo-cop use.

“We are 95% confident that the interval from 0.476 to 0.544 does contain the true value of \( p \).”

This means if we were to select many different samples of size 829 and construct the corresponding confidence intervals, 95% of them would actually contain the value of the population proportion \( p \).

Critical Values

1. We know from Section 6-6 that under certain conditions, the sampling distribution of sample proportions can be approximated by a normal distribution, as in Figure 7-2, following.

2. Sample proportions have a relatively small chance (with probability denoted by \( \alpha \)) of falling in one of the red tails of Figure 7-2, following.

3. Denoting the area of each shaded tail by \( \alpha/2 \), we see that there is a total probability of \( \alpha \) that a sample proportion will fall in either of the two red tails.
Critical Values

4. By the rule of complements (from Chapter 4), there is a probability of $1 - \alpha$ that a sample proportion will fall within the inner region of Figure 7-2, following.

5. The $z$ score separating the right-tail is commonly 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.

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.

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. (See Figure 7-2.)

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

$\alpha = 5$

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

$z_{\alpha/2} = 1.96$

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

$\alpha = .05$

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

Use Table A-2 to find a $z$ score of 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 (with probability \( 1 - \alpha \)) difference 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, as shown in Formula 7-1, following.

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

Formula 7-1

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

Confidence Interval for Population Proportion

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

where

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

Confidence Interval for Population Proportion - cont

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

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

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

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 \( E = \sqrt{\frac{\hat{p} \cdot \hat{q}}{n}} \).
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: 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.

First, we check for assumptions. We note that $n\hat{p} = 422.79 \geq 5$, and $n\hat{q} = 406.21 \geq 5$.

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

$$E = 1.96 \sqrt{\frac{\hat{p}(1 - \hat{p})}{n}}$$

$$E = 0.03403$$

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

We substitute our values from Part a to obtain:

$$0.51 - 0.03403 < p < 0.51 + 0.03403$$

$$0.476 < p < 0.544$$

Example: 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.

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

Sample Size

Suppose we want to collect sample data with the objective of estimating some population. 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)

\[ n = \left( \frac{Z_{\alpha/2}}{E} \right)^2 \hat{p} \hat{q} \]

Sample Size for Estimating Proportion \( p \)

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

\[ n = \left( \frac{Z_{\alpha/2}}{E} \right)^2 \hat{p} \hat{q} \]

Formula 7-2

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

\[ n = \left( \frac{Z_{\alpha/2}}{E} \right)^2 0.25 \]

Formula 7-3

Example: Suppose a sociologist wants to determine the current percentage of U.S. households using e-mail. How many households must be surveyed in order to be 95% confident that the sample percentage is in error by no more than four percentage points?

a) Use this result from an earlier study: In 1997, 16.9% of U.S. households used e-mail (based on data from The World Almanac and Book of Facts).

b) Assume that we have no prior information suggesting a possible value of \( \hat{p} \).

Example: Suppose a sociologist wants to determine the current percentage of U.S. households using e-mail. How many households must be surveyed in order to be 95% confident that the sample percentage is in error by no more than four percentage points?

To be 95% confident that our sample percentage is within four percentage points of the true percentage for all households, we should randomly select and survey 338 households.

Example: Suppose a sociologist wants to determine the current percentage of U.S. households using e-mail. How many households must be surveyed in order to be 95% confident that the sample percentage is in error by no more than four percentage points?

b) Assume that we have no prior information suggesting a possible value of \( \hat{p} \).

\[ n = \frac{(1.96)^2 \cdot 0.25}{0.04^2} \]

With no prior information, we need a larger sample to achieve the same results with 95% confidence and an error of no more than 4%.

= 601 households

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:

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