Estimating a Population Proportion

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

See the following lesson in Course Documents of CourseCompass.
S4.D1.MAT 155 Chapter 7 Estimates and Sample Sizes
155 Chapter 7 Lesson (Package file)

These notes cover the following topics: point estimate; level of confidence; confidence interval for the population proportion; confidence interval for the population mean when the population standard deviation is known; confidence interval for the population mean when the population standard deviation is unknown; determine the sample size for attribute and variable sampling.

Test 2 (Chapters 4, 5, & 6) Results

Confidence intervals for a proportion (simulation)
http://www.mathxl.com/info/MediaPopup.aspx?origin=1&disciplineGroup=1
&type=Java%20Applets&loc=Pdf1@aw_mml_shared_1/statistics/West_Applets/propci.html&width=600&height=800
&autoh=yes&centerwin=yes

The Excel program will calculate critical values, confidence intervals, and sample size for proportions. It is available at http://cfcc.edu/faculty/cmoore/0702proportion.xls.

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:
  • z₀ denotes the z score with an area of α to its right.
  • If α = 0.025, the critical value is z₀.025 = 1.96.
  • That is, the critical value z₀.025 = 1.96 has an area of 0.025 to its right.

TI-83/84 Tutorials — http://cfcc.edu/faculty/cmoore/TI-STAT.htm
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.

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.

Definitions

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

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.

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. 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.

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

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

- $95\% = 1 - \alpha \Rightarrow \alpha = 5\% = 0.05$
- $\alpha/2 = 2.5\% = 0.025$

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

- Use Table A-2 to find a $z$ score of 1.96
- $95\% = 1 - \alpha \Rightarrow \alpha = 5\% = 0.05$
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:

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

Margin of Error for Proportions

Confidence Interval for Estimating a Population Proportion \( p \)

- \( p \) = population proportion
- \( \hat{p} \) = sample proportion
- \( n \) = number of sample
- \( E \) = 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 $E = z_{\alpha/2} \sqrt{\frac{\hat{p}\hat{q}}{n}}$
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:

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.

\[
E = z_{\alpha/2} \sqrt{\frac{\hat{p} \hat{q}}{n}} = 1.96 \sqrt{\frac{(0.70)(0.30)}{1501}} = 0.023183
\]

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. [Margin of error $E = 0.023$.]

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

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

(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:

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.

See Sample Size tab on the Excel program at http://cfcc.edu/faculty/cmoore/0702proportion.xls
Example:
a) Use \( \hat{p} = 0.73 \) and \( \hat{q} = 1 - \hat{p} = 0.27 \)
\[ \alpha = 0.05 \quad \text{so} \quad z_{\alpha/2} = 1.96 \]
\[ E = 0.03 \]
\[ n = \left( \frac{z_{\alpha/2}}{E} \right)^2 \hat{p} \hat{q} \]
\[ = \left( \frac{1.96}{2} \right)^2 (0.73)(0.27) \]
\[ = 0.03 \]
\[ = 841.3104 \]
\[ = 842 \]

Example:
b) Use \( \alpha = 0.05 \) so \( z_{\alpha/2} = 1.96 \)
\[ E = 0.03 \]
\[ n = \left( \frac{z_{\alpha/2}}{E} \right)^2 \]
\[ = \left( \frac{1.96}{0.25} \right)^2 \]
\[ = 1067.1111 \]
\[ = 1068 \]

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

Recap
In this section we have discussed:
- Point estimates.
- Confidence intervals.
- Confidence levels.
- Critical values.
- Margin of error.
- Determining sample sizes.

The Excel program "Proportion (xls)" will calculate critical z-values, confidence intervals for proportion, and minimum sample size. It is available through the Technology link at http://cfcc.edu/faculty/cmoore/indexExcHtm.htm

"Proportion (xls)" may be accessed directly at http://cfcc.edu/faculty/cmoore/0702proportion.xls
Section 7.2 Estimating a Population Proportion

346/6. Find the critical value $z_{\alpha/2}$ that corresponds to 99.5% confidence level.

Section 7.2 Estimating a Population Proportion

346/8. Find $z_{\alpha/2}$ for $\alpha = 0.02$.

Section 7.2 Estimating a Population Proportion

346/10. Express the confidence interval $0.720 < \hat{p} < 0.780$ in the form of $\hat{p} \pm E$.

Section 7.2 Estimating a Population Proportion

346/12. Express the confidence interval $0.222 \pm 0.044$ in the form $\hat{p} - E < \hat{p} < \hat{p} + E$. 
Section 7.2 Estimating a Population Proportion

346/14. Use the confidence interval limits $0.772 < p < 0.776$ to find the point estimate $\hat{p}$ and the margin of error $E$.

346/16. Use the given confidence interval limits to find the point estimate $\hat{p}$ and the margin of error $E$: $0.102 < p < 0.236$

346/18. Assume that a sample is used to estimate a population proportion $p$. Find the margin of error $E$ that corresponds to the given statistics and confidence level: $n = 500; x = 220$; 99% confidence

346/20. Assume that a sample is used to estimate a population proportion $p$. Find the margin of error $E$ that corresponds to the given statistics and confidence level: 90% confidence; sample size is 1780, of which 35% are successes.
346/22. Use the sample data and confidence level to construct the confidence interval estimate of the population proportion \( p \): \( n = 2000; \ x = 400; \ 95\% \) confidence

\[
\hat{p} = \frac{x}{n} = \frac{400}{2000} = 0.2
\]

The Excel program "Proportion (xls)" will calculate critical \( z \)-values, confidence intervals for proportion, and minimum sample size. It is available through the Technology link at http://cfcc.edu/faculty/cmoore/IndexExcHtm.htm

99% confidence interval for proportion is (0.182, 0.218).

346/24. Use the sample data and confidence level to construct the confidence interval estimate of the population proportion \( p \): \( n = 5200; \ x = 4821; \ 99\% \) confidence

\[
\hat{p} = \frac{x}{n} = \frac{4821}{5200} = 0.923
\]

The Excel program "Proportion (xls)" will calculate critical \( z \)-values, confidence intervals for proportion, and minimum sample size. It is available through the Technology link at http://cfcc.edu/faculty/cmoore/IndexExcHtm.htm

99% confidence interval for proportion is (0.918, 0.936).

346/26. Use the given data to find the minimum sample size required to estimate a population proportion or percentage: margin of error 0.005; 99% confidence level; \( \hat{p} \) and \( \hat{q} \) unknown.

\[
n = \frac{E^2 \hat{p} \hat{q}}{0.005^2} = \frac{(0.005)^2 \cdot 0.5 \cdot 0.5}{0.005^2} = 66348.97, \text{ rounded up to } n = 66349.
\]

The Excel program "Proportion (xls)" will calculate critical \( z \)-values, confidence intervals for proportion, and minimum sample size. It is available through the Technology link at http://cfcc.edu/faculty/cmoore/IndexExcHtm.htm

99% = 0.99 yields area of 0.005 on each side of normal curve or 0.995 area to the left of critical value; so, \( c.v. \) from Table A-2 is \( z_{0.995} = 2.58 \).

347/28. Use the given data to find the minimum sample size required to estimate a population proportion or percentage: margin of error is three percentage points; confidence level is 95%; from a prior study, \( \hat{p} \) is estimated by the decimal equivalent of 87%.

\[
n = \frac{E^2 \hat{p} \hat{q}}{0.03^2} = \frac{(0.03)^2 \cdot 0.87 \cdot 0.13}{0.03^2} = 482.743, \text{ rounded up to } n = 483.
\]

The Excel program "Proportion (xls)" will calculate critical \( z \)-values, confidence intervals for proportion, and minimum sample size. It is available through the Technology link at http://cfcc.edu/faculty/cmoore/IndexExcHtm.htm

95% = 0.95 yields area of 0.025 on each side of normal curve or 0.975 area to the left of critical value; so, \( c.v. \) from Table A-2 is \( z_{0.975} = 1.96 \).

p-hat = 0.87 and E = 0.03

n = 482.743, rounded up to n = 483.
347/30. Gender Selection The Genetics and IVF Institute conducted a clinical trial of the YSORT method designed to increase the probability of conceiving a boy. As of this writing, 152 babies were born to parents using the YSORT method, and 127 of them were boys.

a. What is the best point estimate of the population proportion of boys born to parents using the YSORT method?

b. Use the sample data to construct a 99% confidence interval estimate of the percentage of boys born to parents using the YSORT method.

c. Based on the results, does the YSORT method appear to be effective? Why or why not?

(a) Best point estimate is 0.836.

(b) 99% confidence interval is (0.758, 0.913).

(c) Yes, it appears that the YSORT method of gender selection is effective because the lower value of the interval is larger than 0.5 or 50%.

347/32. Medical Malpractice An important issue facing Americans is the large number of medical malpractice lawsuits and the expenses that they generate. In a study of 1228 randomly selected medical malpractice lawsuits, it is found that 856 of them were later dropped or dismissed (based on data from the Physician Insurers Association of America).

a. What is the best point estimate of the proportion of medical malpractice lawsuits that are dropped or dismissed?

b. Construct a 99% confidence interval estimate of the proportion of medical malpractice lawsuits that are dropped or dismissed.

c. Does it appear that the majority of such suits are dropped or dismissed? Why or why not?

(a) Best point estimate is 0.697.

(b) 99% confidence interval is (0.663, 0.731).

(c) Yes, it appears that the majority of the law suits are dropped or dismissed because the lower value of the interval is larger than 0.5 or 50%.

349/42. Cell Phones As the newly hired manager of a company that provides cell phone service, you want to determine the percentage of adults in your state who live in a household with cell phones and no land-line phones. How many adults must you survey? Assume that you want to be 90% confident that the sample percentage is within four percentage points of the true population percentage.

a. Assume that nothing is known about the percentage of adults who live in a household with cell phones and no land-line phones.

b. Assume that a recent survey suggests that about 8% of adults live in a household with cell phones and no land-line phones (based on data from the National Health Interview Survey).

(a) Use sample proportion as 0.5; 90% yields $z_{0.05/2} = \text{invnorm}(0.05) = 1.645$ (Table A-2 value); $E = 0.04$.

Round up to next whole number to get $n = 423$.

(b) Use sample proportion as 0.08; 90% yields $z_{0.05/2} = \text{invnorm}(0.05) = 1.645$ (Table A-2 value); $E = 0.04$.

Round up to next whole number to get $n = 125$.

349/44. Name Recognition As this book was being written, former New York City mayor Rudolph Giuliani announced that he was a candidate for the presidency of the United States. If you were a campaign worker and needed to determine the percentage of people that recognized his name, how many people should you have surveyed to estimate that percentage? Assume that you wanted to be 95% confident that the sample percentage was in error by no more than two percentage points, and also assume that a recent survey indicated that Giuliani’s name is recognized by 10% of all adults (based on data from a Gallup poll).

95% = 0.95 yields area of 0.025 on each side of normal curve or 0.975 area to the left of critical value; so, c.v. from Table A-2 is $z_{0.025} = 1.96$.

$n = 864.328$, rounded up to $n = 865$. 

15