This is “Large Sample Estimation of a Population Proportion”, section 7.3 from the book Beginning Statistics (v. 1.0). For details on it (including licensing), click here.

For more information on the source of this book, or why it is available for free, please see the project's home page. You can browse or download additional books there.

Has this book helped you? Consider passing it on:

Help Creative Commons

Creative Commons supports free culture from music to education. Their licenses helped make this book available to you.

Help a Public School

DonorsChoose.org helps people like you help teachers fund their classroom projects, from art supplies to books to calculators.

Previous Section | Table of Contents | Next Section

7.3 Large Sample Estimation of a Population Proportion

LEARNING OBJECTIVE

1. To understand how to apply the formula for a confidence interval for a population proportion.

Since from Section 6.3 "The Sample Proportion" in Chapter 6 "Sampling Distributions" we know the mean, standard deviation, and sampling distribution of the sample proportion $\hat{p}$, the ideas of the previous two sections can be applied to produce a confidence interval for a population proportion. Here is the formula.

Large Sample $100(1-\alpha)\%$ Confidence Interval for a Population Proportion

$$\widehat{p} \pm z_{\alpha/2}\sqrt{\frac{\widehat{p}(1-\widehat{p})}{n}}$$

A sample is large if the interval $[p-3\,\sigma_{\widehat{P}},\ p+3\,\sigma_{\widehat{P}}]$ lies wholly within the interval $[0,1]$.

In actual practice the value of *p* is not known, hence neither is $\sigma_{\widehat{P}}$. In that case we substitute the known quantity $\widehat{p}$ for *p* in making the check; this means checking that the interval

$$\widehat{p}-3\sqrt{\frac{\widehat{p}(1-\widehat{p})}{n}},\ \widehat{p}+3\sqrt{\frac{\widehat{p}(1-\widehat{p})}{n}}$$

lies wholly within the interval $[0,1]$.

EXAMPLE 7

To estimate the proportion of students at a large college who are female, a random sample of 120 students is selected. There are 69 female students in the sample. Construct a 90% confidence interval for the proportion of all students at the college who are female.

Solution:

The proportion of students in the sample who are female is $\widehat{p} = 69/120 = 0.575$.

Confidence level 90% means that $\alpha = 1-0.90 = 0.10$ so $\alpha/2 = 0.05$. From the last line of Figure 12.3 "Critical Values of " we obtain $z_{0.05} = 1.645$.

Thus

$$\widehat{p} \pm z_{\alpha/2}\sqrt{\frac{\widehat{p}(1-\widehat{p})}{n}} = 0.575 \pm 1.645\sqrt{\frac{(0.575)(0.425)}{120}} = 0.575 \pm 0.074$$

One may be 90% confident that the true proportion of all students at the college who are female is contained in the interval $(0.575 - 0.074, 0.575 + 0.074) = (0.501, 0.649)$.

KEY TAKEAWAYS

- We have a single formula for a confidence interval for a population proportion, which is valid when the sample is large.
- The condition that a sample be large is not that its size *n* be at least 30, but that the density function fit inside the interval $[0,1]$.

EXERCISES

BASIC

1. Information about a random sample is given. Verify that the sample is large enough to use it to construct a confidence interval for the population proportion. Then construct a 90% confidence interval for the population proportion.

   a. *n* = 25, $\widehat{p} = 0.7$

   b. *n* = 50, $\widehat{p} = 0.7$

2. Information about a random sample is given. Verify that the sample is large enough to use it to construct a confidence interval for the population proportion. Then construct a 95% confidence interval for the population proportion.

   a. *n* = 2500, $\widehat{p} = 0.22$

   b. *n* = 1200, $\widehat{p} = 0.22$

3. Information about a random sample is given. Verify that the sample is large enough to use it to construct a confidence interval for the population proportion. Then construct a

98% confidence interval for the population proportion.

a. n = 80, $\widehat{p} = 0.4$
b. n = 325, $\widehat{p} = 0.4$

4. Information about a random sample is given. Verify that the sample is large enough to use it to construct a confidence interval for the population proportion. Then construct a 99.5% confidence interval for the population proportion.

a. n = 200, $\widehat{p} = 0.85$
b. n = 75, $\widehat{p} = 0.85$

5. In a random sample of size 1,100, 338 have the characteristic of interest.

a. Compute the sample proportion $\widehat{p}$ with the characteristic of interest.
b. Verify that the sample is large enough to use it to construct a confidence interval for the population proportion.
c. Construct an 80% confidence interval for the population proportion *p*.
d. Construct a 90% confidence interval for the population proportion *p*.
e. Comment on why one interval is longer than the other.

6. In a random sample of size 2,400, 420 have the characteristic of interest.

a. Compute the sample proportion $\widehat{p}$ with the characteristic of interest.
b. Verify that the sample is large enough to use it to construct a confidence interval for the population proportion.
c. Construct a 90% confidence interval for the population proportion *p*.
d. Construct a 99% confidence interval for the population proportion *p*.
e. Comment on why one interval is longer than the other.

APPLICATIONS

7. A security feature on some web pages is graphic representations of words that are readable by human beings but not machines. When a certain design format was tested on 450 subjects, by having them attempt to read ten disguised words, 448 subjects could read all the words.

a. Give a point estimate of the proportion *p* of all people who could read words

disguised in this way.

b. Show that the sample is not sufficiently large to construct a confidence interval for the proportion of all people who could read words disguised in this way.

8. In a random sample of 900 adults, 42 defined themselves as vegetarians.

   a. Give a point estimate of the proportion of all adults who would define themselves as vegetarians.
   b. Verify that the sample is sufficiently large to use it to construct a confidence interval for that proportion.
   c. Construct an 80% confidence interval for the proportion of all adults who would define themselves as vegetarians.

9. In a random sample of 250 employed people, 61 said that they bring work home with them at least occasionally.

   a. Give a point estimate of the proportion of all employed people who bring work home with them at least occasionally.
   b. Construct a 99% confidence interval for that proportion.

10. In a random sample of 1,250 household moves, 822 were moves to a location within the same county as the original residence.

    a. Give a point estimate of the proportion of all household moves that are to a location within the same county as the original residence.
    b. Construct a 98% confidence interval for that proportion.

11. In a random sample of 12,447 hip replacement or revision surgery procedures nationwide, 162 patients developed a surgical site infection.

    a. Give a point estimate of the proportion of all patients undergoing a hip surgery procedure who develop a surgical site infection.
    b. Verify that the sample is sufficiently large to use it to construct a confidence interval for that proportion.
    c. Construct a 95% confidence interval for the proportion of all patients undergoing a hip surgery procedure who develop a surgical site infection.

12. In a certain region prepackaged products labeled 500 g must contain on average at

least 500 grams of the product, and at least 90% of all packages must weigh at least 490 grams. In a random sample of 300 packages, 288 weighed at least 490 grams.

a. Give a point estimate of the proportion of all packages that weigh at least 490 grams.
b. Verify that the sample is sufficiently large to use it to construct a confidence interval for that proportion.
c. Construct a 99.8% confidence interval for the proportion of all packages that weigh at least 490 grams.

13. A survey of 50 randomly selected adults in a small town asked them if their opinion on a proposed “no cruising” restriction late at night. Responses were coded 1 for in favor, 0 for indifferent, and 2 for opposed, with the results shown in the table.

$$
\begin{array}{cccccccccc}
1 & 0 & 2 & 0 & 1 & 0 & 0 & 1 & 1 & 2 \\
0 & 2 & 0 & 0 & 0 & 1 & 0 & 2 & 0 & 0 \\
0 & 2 & 1 & 2 & 0 & 0 & 0 & 2 & 0 & 1 \\
0 & 2 & 0 & 2 & 0 & 1 & 0 & 0 & 2 & 0 \\
1 & 0 & 0 & 1 & 2 & 0 & 0 & 2 & 1 & 2
\end{array}
$$

a. Give a point estimate of the proportion of all adults in the community who are indifferent concerning the proposed restriction.
b. Assuming that the sample is sufficiently large, construct a 90% confidence interval for the proportion of all adults in the community who are indifferent concerning the proposed restriction.

14. To try to understand the reason for returned goods, the manager of a store examines the records on 40 products that were returned in the last year. Reasons were coded by 1 for “defective,” 2 for “unsatisfactory,” and 0 for all other reasons, with the results shown in the table.

$$
\begin{array}{cccccccccc}
0 & 2 & 0 & 0 & 0 & 0 & 0 & 2 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 2 \\
0 & 0 & 2 & 0 & 0 & 0 & 0 & 2 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0
\end{array}
$$

a. Give a point estimate of the proportion of all returns that are because of something wrong with the product, that is, either defective or performed unsatisfactorily.
b. Assuming that the sample is sufficiently large, construct an 80% confidence interval

for the proportion of all returns that are because of something wrong with the product.

15. In order to estimate the proportion of entering students who graduate within six years, the administration at a state university examined the records of 600 randomly selected students who entered the university six years ago, and found that 312 had graduated.

    a. Give a point estimate of the six-year graduation rate, the proportion of entering students who graduate within six years.
    b. Assuming that the sample is sufficiently large, construct a 98% confidence interval for the six-year graduation rate.

16. In a random sample of 2,300 mortgages taken out in a certain region last year, 187 were adjustable-rate mortgages.

    a. Give a point estimate of the proportion of all mortgages taken out in this region last year that were adjustable-rate mortgages.
    b. Assuming that the sample is sufficiently large, construct a 99.9% confidence interval for the proportion of all mortgages taken out in this region last year that were adjustable-rate mortgages.

17. In a research study in cattle breeding, 159 of 273 cows in several herds that were in estrus were detected by means of an intensive once a day, one-hour observation of the herds in early morning.

    a. Give a point estimate of the proportion of all cattle in estrus who are detected by this method.
    b. Assuming that the sample is sufficiently large, construct a 90% confidence interval for the proportion of all cattle in estrus who are detected by this method.

18. A survey of 21,250 households concerning telephone service gave the results shown in the table.

	Landline	No Landline
Cell phone	12,474	5,844
No cell phone	2,529	403

a. Give a point estimate for the proportion of all households in which there is a cell phone but no landline.
b. Assuming the sample is sufficiently large, construct a 99.9% confidence interval for the proportion of all households in which there is a cell phone but no landline.
c. Give a point estimate for the proportion of all households in which there is no telephone service of either kind.
d. Assuming the sample is sufficiently large, construct a 99.9% confidence interval for the proportion of all all households in which there is no telephone service of either kind.

ADDITIONAL EXERCISES

19. In a random sample of 900 adults, 42 defined themselves as vegetarians. Of these 42, 29 were women.

    a. Give a point estimate of the proportion of all self-described vegetarians who are women.
    b. Verify that the sample is sufficiently large to use it to construct a confidence interval for that proportion.
    c. Construct a 90% confidence interval for the proportion of all all self-described vegetarians who are women.

20. A random sample of 185 college soccer players who had suffered injuries that resulted in loss of playing time was made with the results shown in the table. Injuries are classified according to severity of the injury and the condition under which it was sustained.

	Minor	Moderate	Serious
Practice	48	20	6
Game	62	32	17

    a. Give a point estimate for the proportion p of all injuries to college soccer players that are sustained in practice.
    b. Construct a 95% confidence interval for the proportion p of all injuries to college

soccer players that are sustained in practice.

c. Give a point estimate for the proportion *p* of all injuries to college soccer players that are either moderate or serious.

d. Construct a 95% confidence interval for the proportion *p* of all injuries to college soccer players that are either moderate or serious.

21. The body mass index (BMI) was measured in 1,200 randomly selected adults, with the results shown in the table.

	BMI		
	Under 18.5	**18.5–25**	**Over 25**
Men	36	165	315
Women	75	274	335

a. Give a point estimate for the proportion of all men whose BMI is over 25.

b. Assuming the sample is sufficiently large, construct a 99% confidence interval for the proportion of all men whose BMI is over 25.

c. Give a point estimate for the proportion of all adults, regardless of gender, whose BMI is over 25.

d. Assuming the sample is sufficiently large, construct a 99% confidence interval for the proportion of all adults, regardless of gender, whose BMI is over 25.

22. Confidence intervals constructed using the formula in this section often do not do as well as expected unless *n* is quite large, especially when the true population proportion is close to either 0 or 1. In such cases a better result is obtained by adding two successes and two failures to the actual data and then computing the confidence interval. This is the same as using the formula

$$\tilde{p} \pm z_{\alpha/2}\sqrt{\frac{\tilde{p}(1-\tilde{p})}{\tilde{n}}}$$

where

$$\tilde{p} = \frac{x+2}{n+4} \text{ and } \tilde{n} = n+4$$

Suppose that in a random sample of 600 households, 12 had no telephone service of any kind. Use the adjusted confidence interval procedure just described to form a 99.9% confidence interval for the proportion of all households that have no telephone service of

any kind.

LARGE DATA SET EXERCISES

23. Large Data Sets 4 and 4A list the results of 500 tosses of a die. Let *p* denote the proportion of all tosses of this die that would result in a four. Use the sample data to construct a 90% confidence interval for *p*.

    http://www.flatworldknowledge.com/sites/all/files/data4.xls

    http://www.flatworldknowledge.com/sites/all/files/data4A.xls

24. Large Data Set 6 records results of a random survey of 200 voters in each of two regions, in which they were asked to express whether they prefer Candidate *A* for a U.S. Senate seat or prefer some other candidate. Use the full data set (400 observations) to construct a 98% confidence interval for the proportion *p* of all voters who prefer Candidate *A*.

    http://www.flatworldknowledge.com/sites/all/files/data6.xls

25. Lines 2 through 536 in Large Data Set 11 is a sample of 535 real estate sales in a certain region in 2008. Those that were foreclosure sales are identified with a 1 in the second column.

    http://www.flatworldknowledge.com/sites/all/files/data11.xls

    a. Use these data to construct a point estimate $\hat{p}$ of the proportion *p* of all real estate sales in this region in 2008 that were foreclosure sales.
    b. Use these data to construct a 90% confidence for *p*.

26. Lines 537 through 1106 in Large Data Set 11 is a sample of 570 real estate sales in a certain region in 2010. Those that were foreclosure sales are identified with a 1 in the second column.

    http://www.flatworldknowledge.com/sites/all/files/data11.xls

a. Use these data to construct a point estimate $\hat{p}$ of the proportion *p* of all real estate sales in this region in 2010 that were foreclosure sales.

b. Use these data to construct a 90% confidence for *p*.

ANSWERS

1. a. (0.5492, 0.8508)
   b. (0.5934, 0.8066)

3. a. (0.2726, 0.5274)
   b. (0.3368, 0.4632)

5. a. 0.3073
   b. $\hat{p} \pm 3\sqrt{\frac{\hat{p}\hat{q}}{n}} = 0.31 \pm 0.04$

   and

   $[0.27, 0.35] \subset [0,1]$

   c. (0.2895, 0.3251)
   d. (0.2844, 0.3302)
   e. Asking for greater confidence requires a longer interval.

7. a. 0.9956
   b. (0.9862, 1.005)

9. a. 0.244
   b. (0.1740, 0.3140)

11. a. 0.013
    b. (0.01, 0.016)
    c. (0.011, 0.015)

13. a. 0.52
    b. (0.4038, 0.6362)

15. a. 0.52

    b. (0.4726, 0.5674)

17. a. 0.5824

    b. (0.5333, 0.6315)

19. a. 0.69

    b. $\widehat{p} \pm 3\sqrt{\frac{\widehat{p}\widehat{q}}{n}} = 0.69 \pm 0.21$

    and

    $[0.48, 0.90] \subset [0,1]$

    c. 0.69 ± 0.12

21. a. 0.6105

    b. (0.5552, 0.6658)

    c. 0.5583

    d. (0.5214, 0.5952)

23. $(0.1368, 0.1912)$

25. a. $\widehat{p} = 0.2280$

    b. $(0.1982, 0.2579)$