

## 6.4: A Confidence Interval for A Population Proportion

During an election year, we see articles in the newspaper that state **confidence intervals** in terms of proportions or percentages. For example, a poll for a particular candidate running for president might show that the candidate has 40% of the vote within three percentage points (if the sample is large enough). Often, election polls are calculated with 95% confidence, so, the pollsters would be 95% confident that the true proportion of voters who favored the candidate would be between 0.37 and 0.43.

Investors in the stock market are interested in the true proportion of stocks that go up and down each week. Businesses that sell personal computers are interested in the proportion of households in the United States that own personal computers. Confidence intervals can be calculated for the true proportion of stocks that go up or down each week and for the true proportion of households in the United States that own personal computers.

The procedure to find the confidence interval for a population proportion is similar to that for the population mean, but the formulas are a bit different although conceptually identical. While the formulas are different, they are based upon the same mathematical foundation given to us by the Central Limit Theorem. Because of this we will see the same basic format using the same three pieces of information: the sample value of the parameter in question, the standard deviation of the relevant sampling distribution, and the number of standard deviations we need to have the confidence in our estimate that we desire.

How do you know you are dealing with a proportion problem? First, the underlying distribution has a **binary random variable**. (There is no mention of a mean or average.) To form a sample proportion, take  $x$ , the number of successes (or other observations of interest) and divide it by  $n$ , the number of trials (or the sample size). The random variable  $P'$  (read "P prime") is the sample proportion,

$$P' = \frac{x}{n}$$

$x$  = the **number** of successes in the sample

$n$  = the size of the sample

$P'$  = the **estimated proportion** of successes or sample proportion of successes ( $P'$  is a **point estimate** for  $P$ , the true population proportion, and thus  $1 - P$  is the probability of a "failure" in any one trial.)

The population standard deviation of this estimate is equal to:

$$\sigma_{P'} = \sqrt{\frac{P(1-P)}{n}}$$

The confidence interval for a population proportion, therefore, becomes:

$$P' \pm \left[ z_{\left(\frac{\alpha}{2}\right)} \sqrt{\frac{P'(1-P')}{n}} \right]$$

$z_{\left(\frac{\alpha}{2}\right)}$  is set according to our desired degree of confidence and  $\sqrt{\frac{P'(1-P')}{n}}$  is the estimated standard deviation of the sampling distribution (using the sample proportion as the point estimate for the population one).

### Example 6.4.1

Suppose that a market research firm is hired to estimate the percent of adults living in a large city who have cell phones. Five hundred randomly selected adult residents in this city are surveyed to determine whether they have cell phones. Of the 500 people sampled, 421 responded yes - they own cell phones. Using a 95% confidence level, compute a confidence interval estimate for the true proportion of adult residents of this city who have cell phones.

#### Answer

Let  $x$  = the number of people in the sample who have cell phones.  $x$  is binomial: the random variable is binary, people either have a cell phone or they do not.

To calculate the confidence interval, we must find  $P'$ .

$$n = 500$$

$x$  = the number of people who own cell phones in the sample = 421

$$P' = \frac{x}{n} = \frac{421}{500} = 0.842$$

Since the requested confidence level is  $CL = 0.95$ , then  $\alpha = 1 - CL = 1 - 0.95 = 0.05$ , and  $\left(\frac{\alpha}{2}\right) = 0.025$ .

Therefore,  $z_{\frac{\alpha}{2}} = z_{0.025} = 1.96$ . This can be found using the  $z$  table in Appendix A. This can also be found in the Student's  $t$ -table at the 0.025 column and the infinity degrees of freedom row, because at infinite degrees of freedom the Student's  $t$ -distribution becomes identical to the standard normal distribution,  $Z$ .

The confidence interval for the true population proportion is

$$P' - z_{\alpha} \sqrt{\frac{P'(1 - P')}{n}} \leq P \leq P' + z_{\alpha} \sqrt{\frac{P'(1 - P')}{n}}$$

Substituting in the values from above we find the confidence interval is:  $0.810 \leq P \leq 0.874$

### Interpretation

We estimate with 95% confidence that between 81% and 87.4% of all adult residents of this city have cell phones.

### Explanation of 95% Confidence Level

Ninety-five percent of the confidence intervals constructed in this way would contain the true value for the population proportion of all adult residents of this city who have cell phones.

### Exercise 6.4.1

Suppose 250 randomly selected people are surveyed to determine if they own a tablet. Of the 250 surveyed, 98 reported owning a tablet. Using a 95% confidence level, compute a confidence interval estimate for the true proportion of people who own tablets.

### Example 6.4.2

The Dundee Dog Training School has a larger than average proportion of clients who compete in competitive professional events. A confidence interval for the population proportion of dogs that compete in professional events from 150 different training schools is constructed. The lower limit is determined to be 0.08 and the upper limit is determined to be 0.16. Determine the level of confidence used to construct the interval of the population proportion of dogs that compete in professional events.

### Answer

We begin with the formula for a confidence interval for a proportion because the random variable is binary; either the client competes in professional competitive dog events or they don't.

$$P' \pm \left[ z_{\left(\frac{\alpha}{2}\right)} \sqrt{\frac{P'(1 - P')}{n}} \right]$$

Next we find the sample proportion:

$$P' = \frac{0.08 + 0.16}{2} = 0.12$$

The  $\pm$  that makes up the confidence interval is thus 0.04

$0.12 + 0.04 = 0.16$  and  $0.12 - 0.04 = 0.08$  which are the boundaries of the confidence interval.

Finally, we solve for  $z$ .

$$\left[ z \cdot \sqrt{\frac{0.12(1-0.12)}{150}} \right] = 0.04, \text{ therefore } z = 1.51$$

And then look up the probability for 1.51 standard deviations on the standard normal table.

$$P(z = 1.51) = 0.4345, P(z) \cdot 2 = 0.8690 \text{ or } 86.90\%$$

### Example 6.4.3

A financial officer for a company wants to estimate the percent of accounts receivable that are more than 30 days overdue. He surveys 500 accounts and finds that 300 are more than 30 days overdue. Compute a 90% confidence interval for the true percent of accounts receivable that are more than 30 days overdue, and interpret the confidence interval.

#### Answer

$$x = 300 \text{ and } n = 500, \text{ so } P' = \frac{x}{n} = \frac{300}{500} = 0.60$$

Since confidence level = 0.90 then  $\alpha = 1 - \text{confidence level} = (1 - 0.90) = 0.10$ , and  $\left(\frac{\alpha}{2}\right) = 0.05$

Therefore,  $z_{\frac{\alpha}{2}} = z_{0.05} = 1.645$ . This z-value can be found using a standard normal probability table. The student's t-table can also be used by checking the table at the 0.05 tail probability column and reading at the line for infinite degrees of freedom. (The t-distribution is the normal distribution at infinite degrees of freedom. This is a handy trick to remember in finding z-values for commonly used levels of confidence.)

We use this formula for a confidence interval for a proportion:

$$P' - z_{\alpha/2} \sqrt{\frac{P'(1-P')}{n}} \leq P \leq P' + z_{\alpha/2} \sqrt{\frac{P'(1-P')}{n}}$$

Substituting in the values from above we find the confidence interval for the true population proportion is  $0.564 \leq P \leq 0.636$

#### Interpretation

We estimate with 90% confidence that the true percent of all accounts receivable overdue by 30+ days is between 56.4% and 63.6%. Alternate wording: We estimate with 90% confidence that between 56.4% and 63.6% of all accounts are 30+ days overdue.

#### Explanation of 90% Confidence Level

Ninety percent of all confidence intervals constructed in this way contain the true value for the population percent of accounts receivable that are overdue 30 days.

### Exercise 6.4.2

A student polls his school to see if students in the school district are for or against new legislation regarding school uniforms. She surveys 600 students and finds that 480 are against the new legislation.

Compute a 90% confidence interval for the true percent of students who are against the new legislation, and interpret the confidence interval.

This page titled [6.4: A Confidence Interval for A Population Proportion](#) is shared under a [CC BY](#) license and was authored, remixed, and/or curated by .