

8.4: 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 solar panels are interested in the proportion of households in Zimbabwe that have solar panels. 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 Zimbabwe that own solar panels.

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 and therefore is a binomial distribution**. (There is no mention of a mean or average.) If X is a binomial random variable, then $X \sim \text{binomial}(n, p)$ where n is the number of trials and p is the probability of a success. To form a sample proportion, take X , the random variable for the number of successes and divide it by n , the number of trials (or the sample size). The random variable \hat{p} (read "p hat") is the sample proportion,

$$\hat{p} = \frac{X}{n}$$

(Sometimes the random variable is denoted as p' , read "p prime".)

- \hat{p} = the **estimated proportion** of successes or sample proportion of successes. (\hat{p} is a **point estimate** for p , the true population proportion, and thus $q = 1 - p$ is the probability of a failure in any one trial.)
- X = the **number** of successes in the sample
- n = the size of the sample

The formula for the confidence interval for a population proportion follows the same format as that for an estimate of a population mean. Remembering the sampling distribution for the proportion from Chapter 5, the standard deviation was found to be:

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

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

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

$z_{\alpha/2}$ is set according to our desired degree of confidence and $\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$ is the standard deviation of the sampling distribution.

The **sample proportions \hat{p} and \hat{q} are estimates of the unknown population proportions p and q** . The estimated proportions \hat{p} and \hat{q} are used because p and q are not known.

Remember that as p moves further from 0.5 the binomial distribution becomes less symmetrical. Because we are estimating the binomial with the symmetrical normal distribution the further away from symmetrical the binomial becomes the less confidence we have in the estimate.

This conclusion can be demonstrated through the following analysis. Proportions are based upon the binomial probability distribution. The possible outcomes are binary, either "success" or "failure". This gives rise to a proportion, meaning the percentage of the outcomes that are "successes". It was shown that the binomial distribution could be fully understood if we knew only the probability of a success in any one trial, called p . The mean and the standard deviation of the binomial were found to be:

$$\mu = np$$

$$\sigma = \sqrt{npq}$$

It was also shown that the binomial could be estimated by the normal distribution if BOTH np AND nq were greater than 5. From the discussion above, it was found that the standardizing formula for the binomial distribution is:

$$Z = \frac{\hat{p} - p}{\sqrt{\left(\frac{pq}{n}\right)}}$$

which is nothing more than a restatement of the general standardizing formula with appropriate substitutions for μ and σ from the binomial. We can use the standard normal distribution, the reason Z is in the equation, because the normal distribution is the limiting distribution of the binomial. This is another example of the Central Limit Theorem. We have already seen that the sampling distribution of the sample mean is normally distributed. Recall the extended discussion in Chapter 5 concerning the sampling distribution of proportions and the conclusions of the Central Limit Theorem.

We can now manipulate this formula in just the same way we did for finding the confidence intervals for a mean, but to find the confidence interval for the binomial population parameter, p .

$$\hat{p} - z_{\alpha/2} \sqrt{\frac{\hat{p}\hat{q}}{n}} \leq p \leq \hat{p} + z_{\alpha/2} \sqrt{\frac{\hat{p}\hat{q}}{n}}$$

Where $\hat{p} = \frac{x}{n}$, the point estimate of p taken from the sample. Notice that \hat{p} has replaced p in the formula. This is because we do not know p , indeed, this is just what we are trying to estimate.

Unfortunately, there is no correction factor for cases where the sample size is small so $n\hat{p}$ and $n\hat{q}$ must always be greater than 5 to develop an interval estimate for p .

? Example 8.4.1

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 \hat{p} , \hat{q} .

$$n = 500$$

$$x = \text{the number of successes in the sample} = 421$$

$$\hat{p} = \frac{x}{n} = \frac{421}{500} = 0.842$$

$\hat{p} = 0.842$ is the sample proportion; this is the point estimate of the population proportion.

$$\hat{q} = 1 - \hat{p} = 1 - 0.842 = 0.158$$

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

$$\text{Then } z_{\alpha/2} = z_{0.025} = 1.96$$

This can be found using the Standard Normal probability table.

The confidence interval for the true binomial population proportion is

$$\hat{p} - z_{\alpha/2} \sqrt{\frac{\hat{p}\hat{q}}{n}} \leq p \leq \hat{p} + z_{\alpha/2} \sqrt{\frac{\hat{p}\hat{q}}{n}}$$

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 8.4.1

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 8.4.2

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 do not.

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

Next we find the sample proportion:

$$\hat{p} = \frac{0.08 + 0.16}{2} = 0.12$$

The \pm that makes up the confidence interval is thus 0.04 $0.12 + 0.04 = 0.16$ and $0.12 - 0.04 = 0.08$ the boundaries of the confidence interval. Finally, we solve for $z_{\alpha/2}$.

$$z_{\alpha/2} \sqrt{\frac{0.12(1-0.12)}{150}} = 0.04, \text{ therefore } z_{\alpha/2} = 1.51$$

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

$$P(0 < Z < 1.51) = 0.4345 \Rightarrow 0.4345 \cdot 2 = 0.8690 \text{ or } 86.90\%$$

? Example 8.4.3

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

$$x = 300 \text{ and } n = 500$$

$$\hat{p} = \frac{x}{n} = \frac{300}{500} = 0.600$$

$$\hat{q} = 1 - \hat{p} = 1 - 0.600 = 0.400$$

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

$$z_{\alpha/2} = z_{0.05} = 1.645$$

This z-value can be found using a standard normal probability table. We use this formula for a confidence interval for a proportion:

$$\hat{p} - z_{\alpha/2} \sqrt{\frac{\hat{p}\hat{q}}{n}} \leq p \leq \hat{p} + z_{\alpha/2} \sqrt{\frac{\hat{p}\hat{q}}{n}}$$

Substituting in the values from above we find the confidence interval for the true binomial 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 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 overdue 30 days.

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 8.4.2

- A student polls her school to see if students in the school district are for or against the 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.
- In a sample of 300 students, 68% said they own an iPod and a smart phone. Compute a 97% confidence interval for the true percent of students who own an iPod and a smartphone.

This page titled [8.4: A Confidence Interval for A Population Proportion](#) is shared under a [CC BY 4.0](#) license and was authored, remixed, and/or curated by [OpenStax](#).

- [8.3: A Confidence Interval for A Population Proportion](#) by [OpenStax](#) is licensed [CC BY 4.0](#). Original source: <https://openstax.org/details/books/introductory-business-statistics>.