

8.4: Proportions and Confidence Intervals for Proportions

We will now make use of the approximation of the binomial distribution by the z -distribution given in [Section 7.1: Using the Normal Distribution to Approximate the Binomial Distribution](#). As usual, the confidence interval will switch the roles of population and sample quantities. The recipe will be laid out first, then we will connect it to what you know about the binomial distribution.

First some definitions. Let X be the number of items in a population of size N that have a given quality. (e.g. the number of females in a population; or the number of people at the U of S wearing yellow sweaters). Then the proportion of the population having the given quality is

$$p = \frac{X}{N} \quad (8.4.1)$$

Given a sample from the population of size n , the best estimate for p is:

$$\hat{p} = \frac{x}{n} \quad (8.4.2)$$

where x is the number of items in the sample having the given quality. To go along with \hat{p} we also have

$$\hat{q} = 1 - \hat{p} \quad (8.4.3)$$

which is the proportion of items in the sample without the given quality.

To compute an C confidence interval for a proportion p we need to compute

$$E = z_c \sqrt{\frac{\hat{p}\hat{q}}{n}} \quad (8.4.4)$$

and it must be true that both $n\hat{p} \geq 5$ and $n\hat{q} \geq 5$ (otherwise we need to use the binomial distribution directly).

With E , the C confidence interval for a proportion is given by

$$\hat{p} - E < p < \hat{p} + E. \quad (8.4.5)$$

To derive the proportions confidence interval formula we'll begin with the sampling theory given by the binomial distribution and the corresponding z -approximation. Then we'll switch the roles of p and \hat{p} . Let

$$x_{\text{pop}} = \frac{n}{N} X = np \quad (8.4.6)$$

be the mean, the expected value, of x that you expect to find in a sample of size n randomly selected from the population with a proportion p of items of interest. This is true because p is also the probability of randomly selecting an item of interest (the probability of success) from the population as per what we did in Chapter 4. The binomial distribution tells you the probability of getting different numbers x of items of interest in your sample given p . The binomial distribution that describes our situation is shown in Figure 8.7; it has a standard deviation of $\sigma = \sqrt{npq}$.

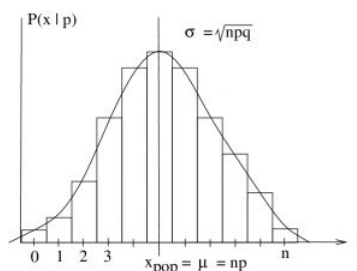


Figure 8.7 : The binomial distribution relevant to forming a sample of size n with x items of interest from a population with a proportion p of items of interest. The normal distribution with the same μ and σ is shown.

Moving to the normal approximation, we have the picture of Figure 8.8.

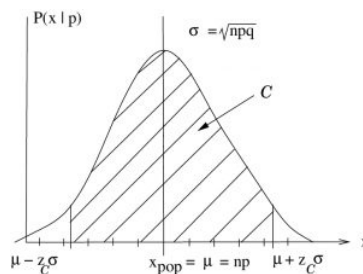


Figure 8.8 : The normal distribution relevant to forming a sample of size n with x items of interest from a population with a proportion p of items of interest. The boundaries of the area \mathcal{C} follow from an inverse z -transform of the z -distribution to a normal distribution of mean μ and standard deviation σ , $x = z\sigma + \mu$.

Figure 8.8 says :

$$\begin{aligned} \mu - z_C\sigma &< x < \mu + z_C\sigma \\ np - z_C\sqrt{npq} &< x < np + z_C\sqrt{npq} \end{aligned}$$

with a (frequentist) probability of \mathcal{C} . This is our sampling theory. Divide by n :

$$\begin{aligned} p - z_C\sqrt{\frac{pq}{n}} &< \frac{x}{n} < p + z_C\sqrt{\frac{pq}{n}} \\ p - z_C\sqrt{\frac{pq}{n}} &< \hat{p} < p + z_C\sqrt{\frac{pq}{n}} \end{aligned}$$

Swapping the roles of the population and sample, we arrive at the confidence interval formula :

$$\hat{p} - z_C\sqrt{\frac{\hat{p}\hat{q}}{n}} < p < \hat{p} + z_C\sqrt{\frac{\hat{p}\hat{q}}{n}}. \quad (8.4.7)$$

Time for a worked example.

Example 8.3 : A sample of 500 nursing applications included 60 men. Find the 90% confidence interval of the true proportion of men who applied to the nursing program.

Solution : From the **t Distribution Table**, look up

$$z_C = 1.65 \quad (8.4.8)$$

and compute

$$\hat{p} = \frac{x}{n} = \frac{60}{500} = 0.12 \quad (8.4.9)$$

$$\hat{q} = 1 - \hat{p} = 1 - 0.12 = 0.88 \quad (8.4.10)$$

Rendered by QuickLaTeX.com

Then

$$\hat{p} + E < p < \hat{p} - E \quad (8.4.11)$$

$$0.12 - 0.024 < p < 0.12 + 0.024 \quad (8.4.12)$$

$$0.096 < p < 0.144 \quad (8.4.13)$$

is the confidence interval with 90% confidence.

□

Sample size need for a poll

Measuring proportions is what pollsters do. For example in an election you might want to know how many people will vote for liberals (items of interest) and how many will vote for conservatives (items not of interest)^[1] In a news paper you might see: “The poll says that 72% of the voters will vote liberal. The poll is considered accurate to 2 percentage points 19 time out of 20.” This means that the 95% confidence interval ($19/20 = 0.95$) of the proportion of liberal voters is 0.72 ± 0.02 (note how proportions are

presented as percentages in the newspaper). The error here is $E = 0.02$. Before the pollster starts telephoning people, she must know how many people to phone to arrive at that goal error of 2%. She needs to know what the sample size n needed is. In general, the minimum sample size needed to attain a goal error E on a confidence interval of \mathcal{C} is

$$n = \hat{p}\hat{q} \left(\frac{z_{\mathcal{C}}}{E} \right)^2. \quad (8.4.14)$$

Here \hat{p} and \hat{q} could come from a previous survey if available. If there is no such survey or if you want to be sure of ending up with an error equal to or less than a goal E , then use $\hat{p} = \hat{q} = 0.5$, see Figure 8.9.

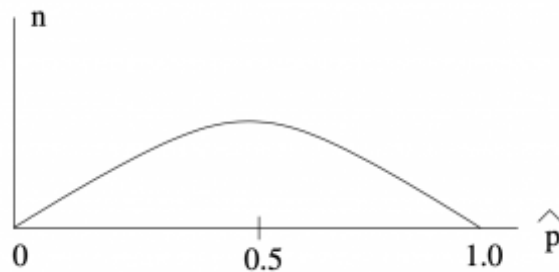


Figure 8.9 : The formula $n = \hat{p}\hat{q} \left(\frac{z_{\mathcal{C}}}{E} \right)^2$ is a quadratic formula. Substitute $\hat{q} = 1 - \hat{p}$ to get $n = \hat{p}(1 - \hat{p}) \left(\frac{z_{\mathcal{C}}}{E} \right)^2$ or $n = (\hat{p} - \hat{p}^2) \left(\frac{z_{\mathcal{C}}}{E} \right)^2$. The maximum of $n_{\max} = \frac{1}{4} \left(\frac{z_{\mathcal{C}}}{E} \right)^2$ is at $\hat{p} = 0.5$.

Example 8.4 : We want to estimate, with 95% confidence, the proportion of people who own a home computer. A previous study gave an answer of 40%. For a new study we want an error of 2%. How many people should we poll?

Solution : From the question we have :

$$\hat{p} = 0.40, \quad \hat{q} = 0.60 \quad (8.4.15)$$

$$E = 0.02, \quad \alpha = 0.95 \quad (8.4.16)$$

From the **t Distribution Table** (or the **Standard Normal Distribution Table** if you think about the areas correctly) we find

$$z_{\mathcal{C}} = z_{95\%} = 1.960. \quad (8.4.17)$$

Therefore

Rendered by QuickLaTeX.com

Which we round up to a sample size of 2305 to ensure that $E < 0.02$.

□

1. We assume here that there are only two parties. For the real life situation of more than two parties we need the multinomial distribution and to approximate it with a multivariate normal distribution. That is a topic for multivariate statistics but the principles are the same as what we cover here. [←](#)

This page titled [8.4: Proportions and Confidence Intervals for Proportions](#) is shared under a [CC BY-NC-SA 4.0](#) license and was authored, remixed, and/or curated by [Gordon E. Sarty](#) via [source content](#) that was edited to the style and standards of the LibreTexts platform.