⚠ Lagunita is retiring and will shut down at 12 noon Pacific Time on March 31, 2020. A few courses may be open for self-enrollment for a limited time. We will continue to offer courses on other online learning platforms; visit http://online.stanford.edu.

Course > Inference: Estimation > Estimation: Population Proportion >
Confidence Intervals for Population Proportion p: Sample Size Calculations

🔖 **Bookmark this page**

Confidence Intervals for Population Proportion p: Sample Size Calculations

Learning Objective: Explain what a confidence interval represents and determine how changes in sample size and confidence level affect the precision of the confidence interval.

Learning Objective: Find confidence intervals for the population mean and the population proportion (when certain conditions are met), and perform sample size calculations.

Determining Sample Size for a Given Margin of Error in Estimating Proportions

Just as we did for means, when we have some level of flexibility in determining the sample size, we can set a desired margin of error for estimating the population proportion and find the sample size that will achieve that.

For example, a final poll on the day before an election would want the margin of error to be quite small (with a high level of confidence) in order to be able to predict the election results with the most precision. This is particularly relevant when it is a close race between the candidates. The polling company needs to figure out how many eligible voters it needs to include in their sample in order to achieve that.

Let's see how we do that.

(**Comment:** For our discussion here we will focus on a 95% confidence level (z* = 2), since this is the most commonly used level of confidence.)

The 95% confidence interval for p is

$$\hat{p} \pm z^* \cdot \sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$$
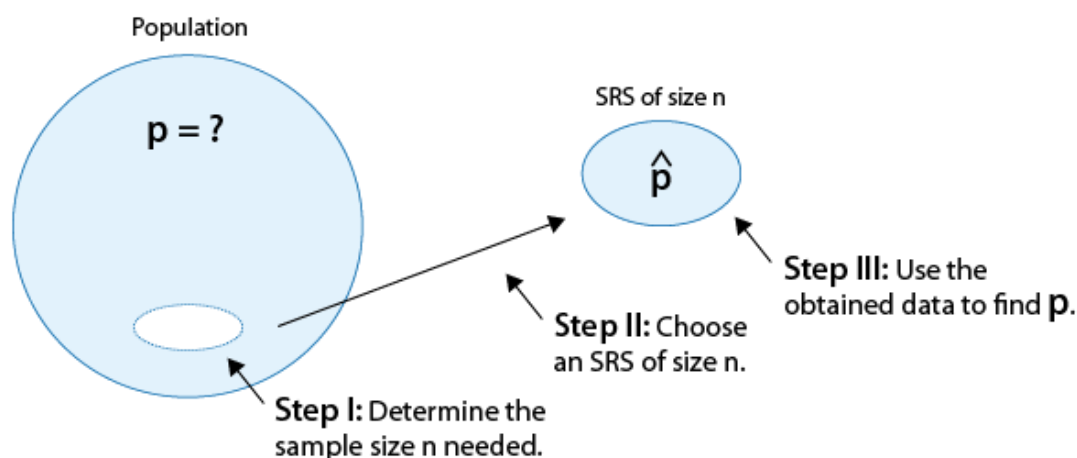
The margin of error, then, is

$$m = 2\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$$

Now we isolate n (i.e., express it as a function of m).

$$n = \frac{4\hat{p}(1-\hat{p})}{m^2}$$

There is a practical problem with this expression that we need to overcome.

Practically, you first determine the sample size, then you choose a random sample of that size, and then use the collected data to find $\hat{p}$.

So the fact that the expression above for determining the sample size depends on $\hat{p}$ is problematic.

The way to overcome this problem is to take the conservative approach by setting $\hat{p} = \frac{1}{2}$.

Why do we call this approach conservative?

It is conservative because the expression that appears in the numerator, $4\hat{p}(1-\hat{p})$ is maximized when $\hat{p} = \frac{1}{2}$.

That way, the n we get will work in giving us the desired margin of error regardless of what the value of $\hat{p}$ is. This is a "worst case scenario" approach. So when we do that we get:

$$n = \frac{(4)\frac{1}{2}\left(1-\frac{1}{2}\right)}{m^2} = \frac{1}{m^2}$$

Example

It seems like media polls usually use a sample size of 1,000 to 1,200. This could be puzzling.

How could the results obtained from, say, 1,100 U.S. adults give us information about the entire population of U.S. adults? 1,100 is such a tiny fraction of the actual population. Here is the answer:

What sample size n is needed if a margin of error m = 0.03 is desired?

$n = \frac{1}{.03^2} = 1111.11 \rightarrow n = 1112$ (remember, always round up). In fact, .03 is a very commonly used margin of error, especially for media polls. For this reason, most media polls work with a sample of around 1,100 people.

Example

A few days before an election, a media outlet would like to estimate p, the proportion of eligible voters who support the Democratic candidate. The media outlet would like the estimate to be within 1% (that is, 0.01) of the true proportion. What is the sample size needed to achieve this in a poll? Set

$$n = \frac{1}{m^2} = \frac{1}{.01^2} = 10000.$$

Note that if we take the same conservative approach for the margin of error:

$$m = 2\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$$

and use $\hat{p} = \frac{1}{2}$, we'll get

$$m = \sqrt{\frac{(4)\frac{1}{2}\left(1-\frac{1}{2}\right)}{n}} = \frac{1}{\sqrt{n}},$$

a conservative estimate for the margin of error, which is useful when we want to get a rough idea of its size without taking the trouble to make detailed calculations.

Also, typically, there are several questions in polls, each yielding a different $\hat{p}$. Rather than reporting the separate margin of error for each question using $m = 2\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$, polls report just one, the conservative margin of error $m = \frac{1}{\sqrt{n}}$

as the margin of error of the poll, which is guaranteed to work for all the questions regardless what the value of $\hat{p}$ ends up being.

> ### Example
>
> A random sample of 2,500 U.S. adults was chosen to participate in a public opinion survey about different issues related to crime. What is the margin of error of this survey?
>
> We'll simply use $m = \dfrac{1}{\sqrt{n}} = \dfrac{1}{\sqrt{2500}} = .02$. The survey has a margin of error of 2%. This means that for each of the questions asked, the obtained sample proportion will be within 2% of the proportion among all U.S. adults.

Did I Get This

1/1 point (graded)

What sample size of U.S. adults do you need, if you would like to estimate the proportion of U.S. adults who are "pro-choice" with a 2.5% margin of error (at the 95% level)?

- ○ 40

- ○ 16

- ○ 1,000

- ● 1,600 ✔

Answer
Correct:
Indeed, the sample size needed for a margin of error of 2.5% (at the 95% level) is $1 / (0.025)^2 = 1,600$.

[Submit]

Did I Get This

1/1 point (graded)

Your answer to the above question indicates that if you take a sample of that size, the sample proportion of adults who are pro-choice is:

- ○ more than 2.5% away from the proportion who are pro-choice among all U.S. adults.

⬤ within 2.5% of the proportion who are pro-choice among all U.S. adults. ✔

◯ exactly equal to the proportion who are pro-choice among all U.S. adults.

Answer
Correct:
Indeed, a sample size of n = 1,600 guarantees that m = 2.5%, which means that the sample proportion that we get will be no more than 2.5% away from the population proportion it estimates (at the 95% level).

Submit

Did I Get This

1/1 point (graded)
If you were to use a random sample of size n = 640 U.S. adults (instead of what you found in question 1), what would the margin of error roughly be?

⬤ 4% ✔

◯ 0.156%

◯ 3%

◯ 5%

Answer
Correct:
Indeed, using the conservative approach, the margin of error would roughly be 1 / sqrt(640) = 0.0395, which equals approximately 0.04

Submit