Confidence Interval for Population Proportion

by Sophia Tutorial

WHAT'S COVERED

This tutorial will cover how to calculate confidence intervals for a population proportion. Our discussion breaks down as follows:

1. Calcuating a Confidence Interval for Population Proportion
2. Constructing a Confidence Interval for Population Proportion

1. Calcuating a Confidence Interval for Population Proportion

A **confidence interval for population proportions** is very similar to a confidence interval for population means. In general, a confidence interval is an estimate found by using a sample statistic and adding and subtracting an amount corresponding to how confident we are that the interval created captures the population parameter.

For a confidence interval for population proportions, the statistic is the sample proportion and population parameter is the population proportion. The following can be used to calculate the confidence interval:

FORMULA

Confidence Interval of Population Proportion

$$CI = \hat{p} \pm z^* \sqrt{\frac{\hat{p}\hat{q}}{n}}$$

HINT

We will use p-hat and q-hat because we do not have an assumed population proportion.

TERM TO KNOW

Confidence Interval for a Population Proportion
A confidence interval that gives a likely range for the value of a population proportion. It is the sample

proportion, plus and minus the margin of error from the normal distribution.

2. Constructing a Confidence Interval for Population Proportion

To construct a confidence interval for population proportions, the following steps must be followed:

⬡ STEP BY STEP

Step 1: Verify the conditions necessary for inference.

Step 2: Calculate the confidence interval.

Step 3: Interpret the confidence interval.

> ↪ EXAMPLE Obecalp is a popular prescription drug but is thought to cause headaches as a side effect. In a random sample of 206 patients taking Obecalp, 23 experienced headaches.
>
> Construct a 95% confidence interval for the proportion of *all* Obecalp users that would experience headaches.

Step 1: Verify the conditions necessary for inference.

Stating the conditions isn't enough, and it's not just a formality--you must verify them. Recall the conditions needed:

Condition	Description
Randomness	How was the sample obtained?
Independence	Population ≥ 10n
Normality	np ≥ 10 and nq ≥ 10

Let's go back to our example to check the requirements of randomness, independence, and normality.

- Randomness: The sample of Obecalp users was a random sample, so that is verified.
- Independence: The sample of Obecalp users taken was a small fraction of the population of Obecalp users. There's no way to verify that empirically unless you had the whole list of people taking the drug. You're going to have to assume there are at least ten times the sample size, or 2,060 people taking this drug.
- Normality: This "np is greater than or equal to 10" equation is a little harder to figure out. You don't know p, the true proportion of people who will get headaches, and you don't have a best guess for it from a null hypothesis. There is no null hypothesis in this problem. What you do have, as a point estimate for p, is p-hat. Verify normality by using p-hat instead of p. Say n times p-hat has to be at least 10. In this case, 206 times p, 23 out of 206, is 23, which is bigger than 10; n times q-hat is 183, which is also bigger than 10.

$$(206)(\frac{23}{206}) = 23;\ 23 > 10$$

$$(206)(\frac{183}{206}) = 183;\ 183 > 10$$

🚩 **HINT**

Recall that you need to use sample statistic, p-hat, to verify the normality condition because you don't know population parameter, p.

Step 2: Calculate the confidence interval.

To do this, we will take the point estimate, p-hat, plus or minus the z* critical value times the standard error of p-hat, which is the square root of p-hat times q hat, over n. The population proportion is not known, so you'll use p-hat for the standard error.

First, let's find the corresponding z* critical value for a 95% confidence interval by using a z-table. For a confidence interval, we can follow the same steps as a two-sided test. If we have a 95% confidence interval, this is actually the same as a 5% significance level. However, this is split between two tails, the lower and upper part of the distribution. Each tail will have 2.5%.
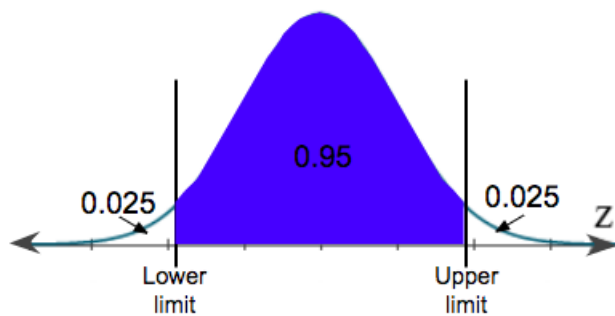
We can use the upper limit to find the critical z-score. Remember, a distribution is 100%, so to find the upper limit, we can subtract 0.025 from 1, which gives us 0.975. Now, we can use a z-table.

						Standard Normal Distribution				
						Z-Table				
z	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
0.0	0.5000	0.5040	0.5080	0.5120	0.5160	0.5199	0.5239	0.5279	0.5319	0.5359
0.1	0.5398	0.5438	0.5478	0.5517	0.5557	0.5596	0.5636	0.5675	0.5714	0.5753
0.2	0.5793	0.5832	0.5871	0.5910	0.5948	0.5987	0.6026	0.6064	0.6103	0.6141
0.3	0.6179	0.6217	0.6255	0.6293	0.6331	0.6368	0.6406	0.6443	0.6480	0.6517
0.4	0.6554	0.6591	0.6628	0.6664	0.6700	0.6736	0.6772	0.6808	0.6844	0.6879
0.5	0.6915	0.6950	0.6985	0.7019	0.7054	0.7088	0.7123	0.7157	0.7190	0.7224
0.6	0.7257	0.7291	0.7324	0.7357	0.7389	0.7422	0.7454	0.7486	0.7517	0.7549
0.7	0.7580	0.7611	0.7642	0.7673	0.7704	0.7734	0.7764	0.7794	0.7823	0.7852
0.8	0.7881	0.7910	0.7939	0.7967	0.7995	0.8023	0.8051	0.8078	0.8106	0.8133

0.9	0.8159	0.8186	0.8212	0.8238	0.8264	0.8289	0.8315	0.8340	0.8365	0.8389
1.0	0.8413	0.8438	0.8461	0.8485	0.8508	0.8531	0.8554	0.8577	0.8599	0.8621
1.1	0.8643	0.8665	0.8686	0.8708	0.8729	0.8749	0.8770	0.8790	0.8810	0.8830
1.2	0.8849	0.8869	0.8888	0.8907	0.8925	0.8944	0.8962	0.8980	0.8997	0.9015
1.3	0.9032	0.9049	0.9066	0.9082	0.9099	0.9115	0.9131	0.9147	0.9162	0.9177
1.4	0.9192	0.9207	0.9222	0.9236	0.9251	0.9265	0.9279	0.9292	0.9306	0.9319
1.5	0.9332	0.9345	0.9357	0.9370	0.9382	0.9394	0.9406	0.9418	0.9429	0.9441
1.6	0.9452	0.9463	0.9474	0.9484	0.9495	0.9505	0.9515	0.9525	0.9535	0.9545
1.7	0.9554	0.9564	0.9573	0.9582	0.9591	0.9599	0.9608	0.9616	0.9625	0.9633
1.8	0.9641	0.9649	0.9656	0.9664	0.9671	0.9678	0.9686	0.9693	0.9699	0.9706
1.9	0.9713	0.9719	0.9726	0.9732	0.9738	0.9744	**0.9750**	0.9756	0.9761	0.9767
2.0	0.9772	0.9778	0.9783	0.9788	0.9793	0.9798	0.9803	0.9808	0.9812	0.9817
2.1	0.9821	0.9826	0.9830	0.9834	0.9838	0.9842	0.9846	0.9850	0.9854	0.9857
2.2	0.9861	0.9864	0.9868	0.9871	0.9875	0.9878	0.9881	0.9884	0.9887	0.9890

In a z-table, the value 0.975 corresponds with a 1.9 in the left column and 0.06 in the top row. This tells us that the z-score is 1.96.

Another way is to use a t-table, which you will learn more about in a later lesson but is available to view at the end of this tutorial. We don't use the t-distribution for proportions; however, we can use the last row in this table to find the confidence levels. Z confidence level, critical values, are found in the last row of this t table, under the infinity value, or ">1000". Essentially, the normal distribution is the t distribution with infinite degrees of freedom. We're going to look in this row to find the z critical value that we should use, which is the same as the 1.96 we got from before.

Now that we have the corresonding z* critical value, we need to use p-hat, which is 23 out of 206, q-hat, which is the complement of p-hat, and the sample size, n, which is 206 and put all this information in the formula:

$$\hat{p} = \frac{23}{206} = 0.112$$

$$\hat{q} = 1 - \hat{p} = 1 - 0.112 = 0.888$$

$$n = 206$$

$$z^* \text{ for } 95\% \; CI = 1.96$$

$$CI = \hat{p} \pm z^* \sqrt{\frac{\hat{p}\hat{q}}{n}}$$

$$CI = 0.112 \pm 1.96 \sqrt{\frac{(0.112)(0.888)}{206}}$$

$$CI = 0.112 \pm 1.96 \sqrt{\frac{0.099456}{206}}$$

$$CI = 0.112 \pm 1.96 \sqrt{0.00048}$$

$$CI = 0.112 \pm 1.96 \, (0.022)$$

$$CI = 0.112 \pm 0.043$$

lower limit: $0.112 - 0.043 = 0.069$
upper limit: $0.112 + 0.043 = 0.155$

From this formula, we obtain 0.112, which was our p-hat, plus or minus 0.043, which is the margin of error. When we evaluate the interval, it's going to be from 0.069 all the way up to 0.155.

Step 3: Interpret the confidence interval.

The confidence interval of 0.069 to 0.155 means we're 95% certain that if everyone who was taking Obecalp was in the study, the true proportion of all Obecalp users who would experience headaches is somewhere between 6.9% and 15.5%. We don't know exactly where in that range, but the true proportion is probably somewhere in this range.

🗹 SUMMARY

You can create point estimates for population proportions, which is your sample proportion, and then use that sample proportion to determine the margin of error for a confidence interval. First, verify the conditions for inference are met, then construct and interpret a confidence interval based on the data that you've gathered and the statistics that you've calculated.

Good luck!

Source: Adapted from Sophia tutorial by Jonathan Osters.

📄 TERMS TO KNOW

Confidence Interval for a Population Proportion
A confidence interval that gives a likely range for the value of a population proportion. It is the sample proportion, plus and minus the margin of error from the normal distribution.

FORMULAS TO KNOW

Confidence Interval of Population Proportion

$$CI = \hat{p} \pm z^* \sqrt{\frac{\hat{p}\hat{q}}{n}}$$