

P6-7 Confidence Intervals for Proportions

- confidence intervals for proportions

[Summary](#) [Learn](#) [Solve](#) [Revise](#) [Answers](#)

Summary

Often we need to estimate the proportion (i.e. fraction or percentage) of a population having a certain property by surveying a sample of the population. The most likely proportion for the population will be the same as the proportion found in the sample.

The range in which the population proportion has a 95% probability of falling is called the 95% confidence interval. As the situation is binomial, the confidence interval is calculated by finding the standard deviation of the number of successes (\sqrt{npq}), then dividing by the number of trials to get the standard deviation of the proportions. This standard deviation along with the mean (most likely proportion) and sample size are then entered into the Inverse Normal distribution function of a calculator along with a left tail area of 0.025 and then the same with a right tail of 0.025.

Learn

Suppose we wanted to know the proportion of voters who plan to vote Labour in the upcoming election. (Proportion is another word for fraction or percentage.) We conduct a survey of 500 voters picked randomly and we find that 230 of those plan to vote Labour. We can conclude from that that the proportion of the voters plan to vote Labour is 0.46 (or 46%).

But that may not be exactly right. If we picked another 500 voters to survey, it is quite possible that 245 of those will plan to vote labour, giving a proportion of 0.49. It is even possible that 290 of our surveyed people would plan to vote Labour, giving a proportion of 0.58. Though the probability of this would be very low.

We can say that the 95% confidence interval for the true proportion is from 0.416 to 0.504. This means that there is a 95% probability that the true proportion will lie between 0.416 and 0.504.

Now, how do we get those figures? We reason as follows.

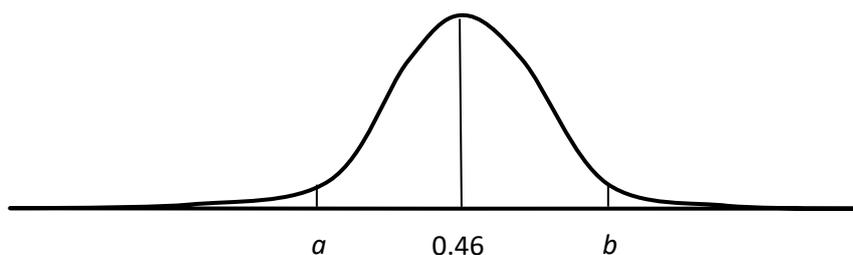
As each time we ask someone whether they plan to vote labour, there is a fixed probability of success independent of what others answered and we are interested in

the number of successes from a certain number of trials. This is a binomial situation where the probability of success (finding someone plans to vote Labour), p , is 0.46, the probability of failure, q , is 0.54, and the number of trials, n , is 500.

The mean number of successes will be $np = 230$ and the standard deviation will be $\sqrt{npq} = \sqrt{500 \times 0.46 \times 0.54} = 11.14$

The proportion will be $230 \div 500 = 0.46$ and the standard deviation of the proportion will be $11.14 \div 500 = 0.02229$.

Now, a binomial distribution for reasonably large n (say > 30), p (say > 0.1) and q (say > 0.1) approximates a normal distribution. So the probability distribution for the actual proportion will be normal with mean 0.46 and standard deviation 0.02229.



a and b are chosen such that 95% of the distribution lies between them, i.e. such that 2.5% lies below a and 2.5% lies above b .

We can find the values of a and b using the Inverse Normal distribution function on a calculator. We enter these values:

| | |
|----------|----------|
| Tail | :Left |
| Area | :0.025 |
| σ | :0.02229 |
| μ | :0.46 |

and we should get the result 0.416, which is the lower limit of the confidence interval.

Then we repeat the calculation with Tail :Right and we should get 0.504, which is the upper limit.

Our conclusion from the survey is then that about 46% of voters plan to vote for the Labour party and that we are 95% confident that the proportion is between 41.6% and 50.4%.

Practice

Q1 In a survey of 400 people chosen at random, it was found that 284 believed in a god. Find the likely proportion of the whole population with this belief and the 95% confidence limits.

- Q2 In a survey of 200 nurses chosen at random, it was found that 156 believed that nurses are under-paid. Find (a) the likely proportion of the nurses with this belief and
- (b) the 95% confidence limits
 - (c) the 90% confidence limits
 - (d) the 99% confidence limits

Comment on how the degree of confidence required affects the width of the confidence interval.

- Q3 In an attempt to determine the percentage of defective parts produced by a machine, 40 parts were tested and it was found that 5 were defective. Find the likely proportion of defective parts and the 90% confidence limits. Why is this confidence interval wider than those found in the previous questions?
- Q4 A slightly bent coin came down heads 79 times out of 180 tosses. Find the likely probability that it will come down heads on the next toss and the 95% confidence limits.

One thing you might have noticed in the above method is that, in reality, the possible population actual proportions would be distributed around the measured sample proportion. In actual fact, the sample proportions are distributed around the population proportion. Also, we should really use the population standard deviation, but, as we don't know this, we have to use the sample proportion.

Because of this our results may be very slightly inaccurate. But they are the best we can do. So we don't worry about these things.

Solve

- Q51 It is possible to write a formula for the confidence interval for a proportion. It is

$$P\left(p - z_{\alpha/2} \times \sqrt{\frac{p(1-p)}{n}} < \theta < p + z_{\alpha/2} \times \sqrt{\frac{p(1-p)}{n}}\right) = 1 - \alpha$$

Try to make sense of this.

Revise

Revision Set 1

- Q61 A slightly misshapen die gave a 6 on 48 rolls out of 200. Find the likely probability that it will come down 6 on the next roll and the 95% confidence

limits.

Answers

Q1 71%, 66.6% to 75.4%

Q2 (a) 78% (b) 72.3% to 83.7% (c) 73.2% to 82.8% (d) 70.5% to 85.5%

The higher the degree of confidence, the wider the confidence interval.

Q3 12.5%, 3.9% to 21.1%

The confidence interval is wide because the sample size is small.

Q4 43.9%, 36.7% to 51.1%

Q51 p is the proportion calculated from the sample, θ is the population proportion, n is the sample size, $1 - \alpha$ is the degree of confidence (e.g. for a 95% confidence interval, $\alpha = 0.05$), $z_{\alpha/2}$ is the z-score which leaves an area of $\alpha/2$ to the right.

$\frac{\sqrt{npq}}{n}$ is written as $\sqrt{\frac{p(1-p)}{n}}$.

Q61 24%, 18.1% to 29.9%