M1Maths

M1Maths.com – Explanations and Practice for all school maths topics

P6-7 Confidence Intervals

- confidence intervals for proportions
- confidence intervals for means

Summary Learn Solve Revise Answers

Summary

Proportions

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 (1 minus half the confidence level required) and then the same with a right tail of 0.025.

Means

At other times, we might need to estimate the mean of some property of a population by surveying a sample of the population. The most likely mean for the population will be the same as the mean found in the sample.

The range in which the population mean has a 95% probability of falling is called the 95% confidence interval. To find the confidence interval we divide the standard deviation of the sample by the square root of the sample size. Then we enter this as the standard deviation in the Inverse Normal function along with the sample mean, and left tail area of 0.025, and then the same with a right tail of 0025.

Confidence Intervals for Proportions

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 0.544.

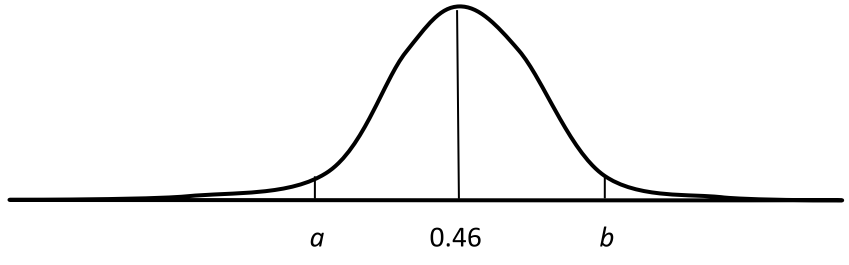
Now, how do we get those figures? We reason as follows.

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 we assume that the possible population actual proportions are distributed around the measured sample proportion. In actual fact, the sample proportions are distributed around the population proportion. So, 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.

Confidence Intervals for Means

Suppose we wanted to know the mean mature height of a certain breed of tomato plants. We plant a sample of 40 seeds and measure them when they mature.

Suppose that we find that the mean height is 47.2 cm and that the standard deviation is 4.8 cm.

Our best estimate for the mean is of course 47.2 cm. But, again, the mean height for all possible tomatoes of that breed might be a bit different from that, maybe 46 cm or 49.2 cm. We can find the confidence interval, the range of heights for which the probability that the true mean lies within that range is say 95%. In this case the confidence interval would be from 45.7 cm to 48.7 cm.

This is how we do it.

Firstly, we need to know that when many random samples of size n are chosen from a large population with standard deviation σ the means of the samples are distributed with standard deviation ${}^{\sigma}/_{\sqrt{n}}$. [The larger the samples, the less variation there will be in the sample means. This should make sense if you think about it.]

So the sample means will have a mean of 47.2 and a standard deviation of $4.8 \div \sqrt{40} =$ 0.759.

If the population has a normal distribution, the sample means will also be distributed normally. For reasonably large samples, the distribution of sample means will be approximately normal, even if the population distribution isn't normal. So we assume that the distribution of sample means is normal.

We then use the Inverse Normal distribution function on the calculator with

Tail	:Left
Area	:0.025
σ	:0.759
μ	:47.2

and we get the result 45.7, which is the lower limit of the confidence interval. Then we repeat with the right tail to get 48.7, which is the upper limit.

So the 95% confidence interval is 457 to 48.7 and there is a 95% probability that the population mean lies within this range.

Practice

Q5 A researcher wants to know the mean height of the adult males on Wotapi Island. She selects 122 men at random and measures them, finding that their mean height is 174.3 cm with a standard deviation of 5.1 cm. Find the likely mean height of the men on the island and the 95% confidence interval.

Q6 35 sample were taken of the ore being loaded onto a ship. The mean copper content was 12.47% and the standard deviation was 1.18%. Find the most likely copper content of the shipment and the 90% confidence interval.

Q7 An organisation wants to know the average quarterly electricity bill for Australian households. They survey 300 households chosen at random and finds that their bills average $432.85 with a standard deviation of $44.58. Find the likely mean bill for the nation and the 99% confidence interval.

Solve

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

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

Explain how this formula is derived.

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.

Q62 Bottles of soft drink are labelled as containing 600 mL. 80 bottles are sampled and found to have a mean of 598.4 mL and a standard deviation of 7.2 mL. Find the likely mean contents of that type of bottle and the 90% confidence interval. Comment on the result.

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 high 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%

Q5 174.3 cm, 173.4 to 175.2 cm

Q6 12.47%, 12.14% to 12.80%

Q7 $432.85, $426.37 to $439.33

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 zscore 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%

Q62 598.4 mL, 597.1 mL to 599.7 mL.
It seems very likely that the bottles are under-filled on average.