Confidence Intervals for Proportions

Confidence Intervals for Proportions

A confidence interval estimates the population proportion, p, lies between two values. For a 95% level of confidence, 95% of all possible confidence intervals will contain the population proportion.

Determination of Confidence Intervals for Proportions

Level of confidence $= 1 - \alpha$, $\quad \mathrm{E}(\hat{P}) = p$, $\quad \mathrm{sd}(\hat{P}) = \sqrt{p(1-p)/n}$

For a sufficiently large n, the sample proportions are approximately normally distributed. Therefore:

$\Pr(-z < Z < z) = 1 - \alpha$, $\quad z = \mathrm{invNorm}\left(1 - \frac{\alpha}{2}\right)$, $\quad \mathrm{N}(0,1)$

$$\Pr\left(-z < \frac{\hat{P} - p}{\sqrt{p(1-p)/n}} < z\right) = 1 - \alpha \Rightarrow \Pr\left(-z\sqrt{p(1-p)/n} < \hat{P} - p < z\sqrt{p(1-p)/n}\right) = 1 - \alpha$$

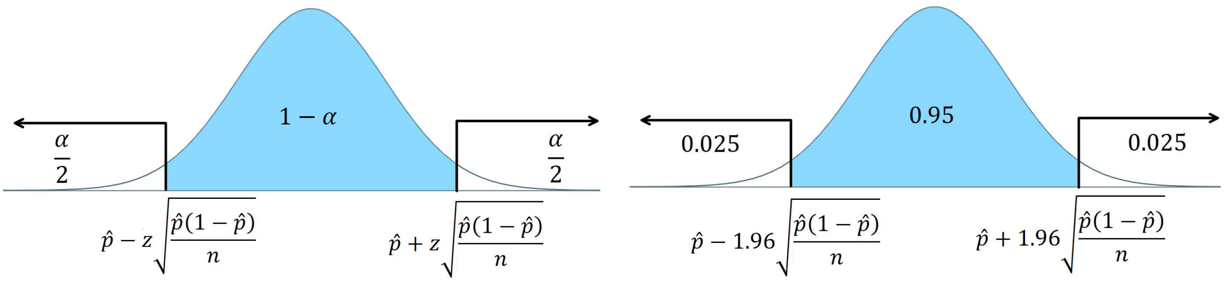
$$\Rightarrow \Pr\left(\hat{P} - z\sqrt{p(1-p)/n} < p < \hat{P} + z\sqrt{p(1-p)/n}\right) = 1 - \alpha$$

The $1 - \alpha$ confidence interval can be expressed as $\hat{P} - z\sqrt{\frac{p(1-p)}{n}} < p < \hat{P} + z\sqrt{\frac{p(1-p)}{n}}$

or more commonly as $\left(\hat{P} - z\sqrt{\frac{p(1-p)}{n}}, \hat{P} + z\sqrt{\frac{p(1-p)}{n}}\right)$

Construction of an Approximate Confidence Interval

Given a sample proportion $\hat{p}$ that is a sufficiently accurate estimate of the population proportion p, the $1 - \alpha$ confidence interval can be expressed as $\left(\hat{p} - z\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}, \hat{p} + z\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}\right)$.



For a 95% level of confidence, the appropriate quantile is $z \approx 1.96$. An integer approximation is $z = 2$.

Therefore, the 95% confidence interval is $\left(\hat{p} - 1.96\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}, \hat{p} + 1.96\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}\right)$

90%: $z = 1.64$, $\quad 99\%$: $z = 2.58$

Example VCAA 2016 Sample Exam 2 Question 2d

From a random sample of 100 members, it was found that the sample proportion of people who spent more than two hours per week in the gym was 0.6. An approximate 95% confidence interval for the population proportion corresponding to this sample proportion is

$$\left(0.6 - 1.96\sqrt{\frac{0.6(1-0.6)}{100}}, 0.6 + 1.96\sqrt{\frac{0.6(1-0.6)}{100}}\right) \approx (0.504, 0.696)$$

Determining the Sample Proportion from the Confidence Interval

We can use the fact that $\hat{p}$ lies halfway between the ends of the confidence interval to determine the sample proportion used to construct the confidence interval.

Example VCAA 2017 Exam 2 Question 5

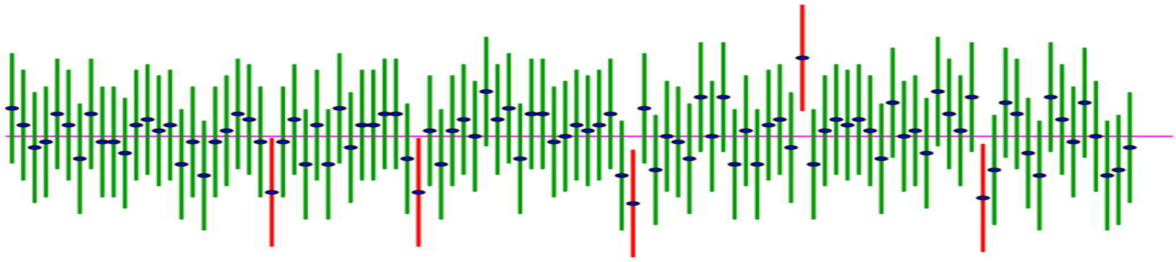
The 95% confidence interval for the proportion of ferry tickets that are cancelled on the intended departure day is calculated from a large sample to be $(0.039, 0.121)$.

The sample proportion from which this interval was constructed is $\dfrac{0.039 + 0.121}{2} = 0.080$

Level of Confidence

The percentage of all confidence intervals that will contain the population proportion. For a 95% level of confidence, 95% of all possible confidence intervals will contain the population proportion.

That means, 5% of the time, due to random chance alone, the confidence interval will not contain the population proportion.



Example VCAA 2018 Exam 2 Question 4dii

Doctors are studying the resting heart rate of adults in two neighbouring towns: Mathsland and Statsville. The probability that a randomly chosen Mathsland adult has a slow heart rate is 0.1587

The doctors took a large random sample of adults from the population of Statsville and calculated an approximate 95% confidence interval for the proportion of Statsville adults who have a slow heart rate. The confidence interval they obtained was $(0.102, 0.145)$.

Explain why this confidence interval suggests that the proportion of adults with a slow heart rate in Statsville could be different from the proportion in Mathsland.

The proportion of Mathsland adults that have a slow heart rate (0.1587) is not within the 95% confidence interval, $(0.102, 0.145)$. There is only a 5% chance that it was not in the confidence interval due to random chance alone. Therefore, it is possible that the proportion of adults with a slow heart rate in Statsville is different to the proportion of adults with slow heart rate in Mathsland.