THE UNIVERSITY OF QUEENSLAND AUSTRALIA | CREATE CHANGE

Statistics in Mathematical Methods

Dr Michael Bulmer
School of Mathematics and Physics
The University of Queensland

Making Maths Matter 2026

Examine the approximate confidence interval

$$\left(\overline{x} - z\frac{s}{\sqrt{n}}, \overline{x} + z\frac{s}{\sqrt{n}}\right)$$

as an interval estimate for μ, the population mean.

Use the approximate confidence interval

$$\left(\hat{p} - z\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}, \hat{p} + z\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}\right)$$

as an interval estimate for p.

Normal Distributions

We say a continuous random variable X has a *Normal distribution* if its density curve is of the form

$$p(x) = \frac{1}{\sqrt{2\pi}\boldsymbol{\sigma}} e^{-\frac{1}{2}\left(\frac{x-\boldsymbol{\mu}}{\boldsymbol{\sigma}}\right)^2}$$

If X has this density curve then $\mathsf{E}(X) = \mu$ and $\mathsf{sd}(X) = \sigma$.

What is the probability that an outcome of X is within 1 standard deviation of the mean? Within 2 standard deviations?

Sample Proportion

The sample proportion of a particular dataset is just **one** outcome of the random variable, $\hat{P}$, obtained by finding the proportion of successes in a random sample.

If we were to repeat this random process again and again, we would get a different outcome, $\hat{p}$, each time.

It can be shown that

$$\mathsf{E}(\hat{P}) = p$$

and

$$\mathsf{sd}(\hat{P}) = \sqrt{\frac{p(1-p)}{n}}.$$

Handed

For example, consider a 'population' of 829 students at UQ.

```
survey = read.csv("survey.csv")
```

$p = \frac{88}{829} = 0.106$ of these students are left handed.

We can simulate sampling from this population to test our theory.

Samples and Proportions

```
sample(survey$Handed, 10)
sample(survey$Handed, 10) == "Left"
sum(sample(survey$Handed, 10) == "Left")
sum(sample(survey$Handed, 10) == "Left")/10

replicate(100, sum(sample(survey$Handed, 10) == "Left")/10)
```

Summary

The spread gets smaller as n increases. This implies that the sample proportion is a more *precise* estimator of the population proportion for larger samples.

As sample size increases, the distribution of the sample proportion becomes closer to a Normal distribution.

The Normal distribution is our friend because we can use it to

1. Describe the distribution of observations
2. Describe the distribution of *statistics*, such as the sample mean and sample proportion

Confidence Intervals

Consider the following sequence of reasoning:

- The sample proportion has roughly a Normal distribution with mean p and standard deviation $\sqrt{\frac{p(1-p)}{n}}$.
- In a Normal distribution about 95% of observations occur within 1.96 standard deviations of the mean.
- So in 95% of samples the sample proportion will be within $1.96\sqrt{\frac{p(1-p)}{n}}$ of p.
- Reversing this, in 95% of samples p will be within $1.96\sqrt{\frac{p(1-p)}{n}}$ of $\hat{p}$.

Confidence Interval

We say we are 95% *confident* that the population proportion is

$$\hat{p} \pm 1.96\sqrt{\frac{p(1-p)}{n}},$$

or that it is in the interval

$$\left(\hat{p} - 1.96\sqrt{\frac{p(1-p)}{n}},\ \hat{p} + 1.96\sqrt{\frac{p(1-p)}{n}}\right).$$

We call this range a 95% *confidence interval* for the population proportion.

However we cannot actually calculate this interval... ☹

Use the approximate confidence interval

$$\left(\hat{p} - z\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}, \hat{p} + z\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}\right)$$

as an interval estimate for p.

Notation

z is the number of standard deviations needed in a Normal distribution to obtain the desired level of confidence (e.g. $z = 1.96$ for 95%, $z = 1.645$ for 90%, etc.)

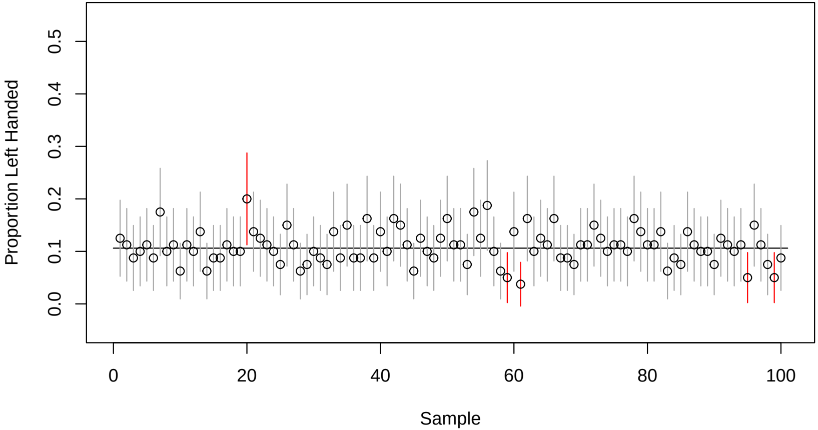
$\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$ is the *standard error of the sample proportion*.

The combined value

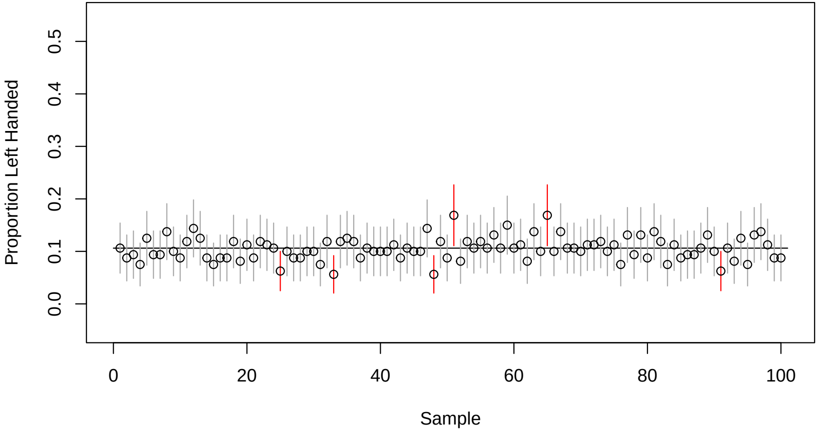
$$z\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$$

is the *margin of error* for the confidence interval.

95% Confidence Intervals ($n = 80$)



95% Confidence Intervals ($n = 160$)



The Islands

The Islands provides an online population of virtual human subjects for use in statistical investigations.

Schools are welcome to use this to support student projects.

islands.smp.uq.edu.au/register

Choosing Sample Size

Suppose we wanted to estimate the proportion of people in a population who belong to a club with a margin of error of 1%.

What sample size should we survey?