UNIVERSITY OF LONDON

Probability and Statistics: To p, or not to p?

Module Leader: Dr James Abdey

5.6 Proportions: confidence intervals and hypothesis testing

Recall the (approximate) sampling distribution of the sample proportion from Section 5.5:

$$\bar{X} = P \to N\left(\mu, \frac{\sigma^2}{n}\right) = N\left(\pi, \frac{\pi(1-\pi)}{n}\right)$$

as $n \to \infty$.

We will now use this result to conduct **statistical inference for proportions**.

Confidence intervals

In Section 4.6 we viewed a confidence interval for a mean as:

best guess ± margin of error.

As the sample proportion is a special case of the sample mean, this construct continues to hold. Here, the:

$$\textbf{point estimate} = p$$

where p is the observed sample proportion, and the:

margin of error = confidence coefficient × standard error.

The confidence coefficient continues to be a z-value such that:

- for 90% confidence, use the confidence coefficient $z = 1.645$
- for 95% confidence, use the confidence coefficient $z = 1.960$
- for 99% confidence, use the confidence coefficient $z = 2.576$.

while the (estimated) standard error[1] is:

$$\sqrt{\frac{\boldsymbol{p\,(1-p)}}{\boldsymbol{n}}}$$

Therefore, a confidence interval for a proportion is given by:

$$\boldsymbol{p \pm z \times \sqrt{\frac{p\,(1-p)}{n}}}.$$

Example

In opinion polling, sample sizes of about 1000 are used as this leads to a margin of error of approximately three percentage points – deemed an **acceptable tolerance on the estimation error** by most political scientists. Suppose 630 out of 1000 voters in a random sample said they would vote 'Yes' in a binary referendum. The sample proprotion is:

$$p = \frac{630}{1000} = 0.63$$

and a 95% confidence interval for π, the true proportion who would vote 'Yes' in the electoral population, is:

$$0.63 \pm 1.96 \times \sqrt{\frac{0.63 \times 0.37}{1000}} = 0.63 \pm 0.03 \quad \Rightarrow \quad (0.60,\ 0.66) \quad \text{or} \quad (60\%,\ 66\%)$$

demonstrating the three percentage-point margin of error.

Hypothesis testing

Suppose we wish to test:

$$\mathrm{H}_0 : \pi = 0.4 \quad \text{vs.} \quad \mathrm{H}_1 : \pi \neq 0.4$$

and a random sample of $n = 1000$ returned a sample proportion of $p = 0.44$. To undertake this test, we follow a similar approach to that outlined in Section 5.4.

We proceed by **standardising $\boldsymbol{P}$** such that:

$$\boldsymbol{Z = \frac{P - \pi}{\sqrt{\pi\,(1-\pi)/n}} \sim N(0,\ 1)}$$

approximately for large n, which is satisfied here since $n = 1000$. Note the test statistic includes the effect size, $P - \pi$, as well as the sample size, n.

[1]The true standard error of the sample proportion is $\pi\,(1-\pi)/n$. However, π is unknown (which is why we are estimating it via a confidence interval), hence we intuitively use the estimated standard error which estimates π with p.

Using our sample data, we now obtain the test statistic value (noting the **influence of both the effect size and the sample size**, and hence ultimately the influence on the p-value):

$$\frac{\mathbf{0.44 - 0.4}}{\sqrt{\mathbf{0.4 \times (1 - 0.4)/1000}}} = \mathbf{2.58.}$$

The p-value is the probability of our test statistic value or a more extreme value conditional on H_0. Noting that $H_1 : \pi \neq 0.4$, 'more extreme' here means a z-score > 2.58 and < -2.58. Due to the symmetry of the standard normal distribution about zero, this can be expressed as:

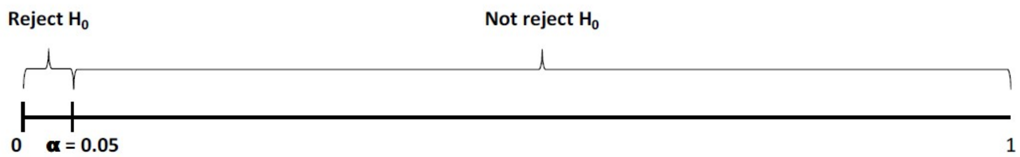
$$\boldsymbol{p}\textbf{-value} = \boldsymbol{P(Z \geq |2.58|) = 0.0099.}$$

Note this value can easily be obtained using Microsoft Excel, say, as:

`=NORM.S.DIST(-2.58)*2` **or** `=(1-NORM.S.DIST(2.58))*2`

where the function `NORM.S.DIST(z)` returns $P(Z \leq z)$ for $Z \sim N(0, 1)$.

Recall the ***p*-value decision rule**, shown below for $\alpha = 0.05$:



Therefore, since $0.0099 < 0.05$ we reject H_0 and conclude that the result is 'statistically significant' at the 5% significance level (and also, just, at the 1% significance level). Hence there is (strong) evidence that $\pi \neq 0.4$. Since $p > \pi$ we might go further and suppose that $\pi > 0.4$.

Finally, recall the possible **decision space**:

		Decision made	
		H_0 not rejected	**H_0 rejected**
True state of nature	**H_0 true**	**Correct decision**	**Type I error**
	H_1 true	**Type II error**	**Correct decision**

As we have rejected H_0 this means one of two things:

- **we have correctly rejected H_0**
- **we have committed a Type I error**.

Although the p-value is very small, indicating it is *highly unlikely* that this is a Type I error, unfortunately we cannot be *certain* which outcome has actually occurred!