14 Statistics V: Confidence Intervals and Hypothesis Testing with More Accurate Margins of Error; *p*-Values

aims

- To know and understand the difference between the *approximate* formula for margin of error, $\frac{1}{\sqrt{n}}$, and the more accurate standard error formula from the booklet of formulae and tables.
- To know how to calculate and apply the more accurate formulae for standard error when constructing a 95% confidence interval.
- To apply the more accurate confidence intervals when carrying out hypothesis tests.
- To calculate and apply the *p*-value for a test statistic as an alternative approach to hypothesis testing.

A more accurate standard error formula

The background

The **central limit theorem** in the previous chapter has introduced us to the standard error of the mean, σ_x (sometimes written $\sigma_{\bar{x}}$). When we take one sample of size n elements and calculate its proportion, $\hat{p}$, we are only finding the proportion from one of many samples of size n. If we were to consider all such proportions, we would have the **sampling distribution of the proportion**.

The standard deviation of the sampling distribution of the proportion is called the **standard error (SE) of the proportion**, written as $\sigma_{\hat{p}}$. The formula for $\sigma_{\hat{p}}$ is given in the booklet of formulae and tables as

$$\sigma_{\hat{p}} = \sqrt{\frac{p(1-p)}{n}}$$ where n is the sample size and p is the population proportion.

Hence, if the true population proportion, p, is known, we can use either the empirical rule or the standard normal tables to estimate the probability that a particular sample proportion will lie within a certain distance of the population proportion.

key point

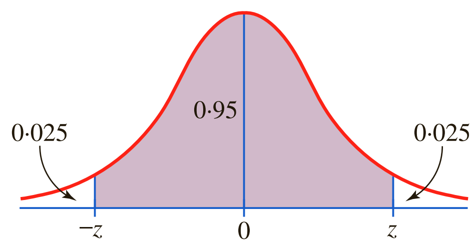
From the previous chapter we know that the empirical rule gives a rough guide, while the standard normal tables give a more accurate estimate.

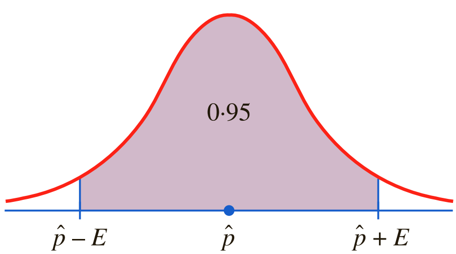
However, in most cases we will not know the true population proportion. This is what we are required to estimate. We reverse the process: instead of writing that there is a 95% chance that a sample proportion, $\hat{p}$, lies in the interval $[p - E, p + E]$, we write that there is a 95% chance that p, the population proportion, lies in the interval $[\hat{p} - E, \hat{p} + E]$. Hence, we use the standard error (SE) of the proportion:

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

The 95% confidence interval using the standard error

At the 95% level of confidence the value of z is: 2 using the empirical rule or 1·96 from the standard normal tables.





z is the number of standard deviation (SE) that the margin of error is from the mean. Hence, the margin of error, E, is given by:

$$E = (z)(\sigma_{\hat{p}})$$

At the 95% confidence interval, $E = 1{\cdot}96\,\sigma_{\hat{p}}$.

We then state that the 95% confidence interval for the true proportion, p, is:

$$\hat{p} - E \leq p \leq \hat{p} + E$$

$$\hat{p} - 1{\cdot}96\sqrt{\frac{\hat{p}(1-\hat{p})}{n}} \leq p \leq \hat{p} + 1{\cdot}96\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$$

Example 1

During the making of a movie, a survey found that out of 724 extras, only 181 were suitable for parts in a major movie.

(i) Find the 95% confidence interval for the proportion of all film extras that may be suitable for parts in a major movie.

(ii) With 95% confidence, what is the highest proportion of extras who would be suitable for parts in a major movie?

(iii) If only 400 extras were surveyed.

(a) What effect would this have on the margin of error?

(b) Are there any implications of taking this action?

Solution

(i) $$\hat{p} = \frac{181}{724} = \frac{1}{4} = 0{\cdot}25\ (= 25\%)$$

$$1 - \hat{p} = 1 - \frac{1}{4} = \frac{3}{4} = 0{\cdot}75$$

$$n = 724$$

Then $$\sigma_{\hat{p}} = \sqrt{\frac{\hat{p}(1-\hat{p})}{n}} = \sqrt{\frac{0{\cdot}25 \times 0{\cdot}75}{724}} = 0{\cdot}01609279 = 0{\cdot}016$$

Hence, the standard error $= \text{SE} = \sigma_{\hat{p}} = 1{\cdot}6\%$.

Remember, the margin of error, E, at the 95% level of confidence $= (1{\cdot}96)\,\sigma_{\hat{p}} = (1{\cdot}96)\,(1{\cdot}6) = 3{\cdot}136\% = 3{\cdot}1\%$.

The 95% confidence interval for the true proportion, p, is then:

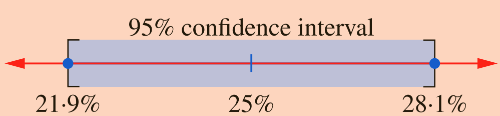
$$\hat{p} - E \leq p \leq \hat{p} + E$$

$$0{\cdot}25 - 0{\cdot}031 \leq p \leq 0{\cdot}25 + 0{\cdot}031$$

$$25\% - 3{\cdot}1\% \leq p \leq 25\% + 3{\cdot}1\%$$

$$21{\cdot}9\% \leq p \leq 28{\cdot}1\%$$

That is, the true proportion, with 95% confidence, lies between 21·9% and 28·1%.



(ii) From part (i) we are 95% confident that the highest proportion of extras who would be suitable for parts in a major movie is 28·1%.

(iii) (a) With fewer extras surveyed, the standard error would be greater,

$$\text{i.e. } \sigma_{\hat{p}} = \sqrt{\frac{\hat{p}(1-\hat{p})}{n}} = \sqrt{\frac{(0{\cdot}25)(0{\cdot}75)}{400}} = 0{\cdot}021650635 = 0{\cdot}022 = 2{\cdot}2\%$$

and $2{\cdot}2\% > 1{\cdot}6\%$. Hence, the margin of error would be greater.

(b) The new survey would be less accurate. It would be quicker and less expensive to carry out.

Example 2

A poll shows that the government's approval rating is at 70%.
The poll is based on a random sample of 896 voters with a margin of error of 3%.
Show that the poll used a 95% level of confidence.

Solution

$$\begin{aligned}
\text{Confidence limits} &= \pm(z)(\sigma_{\hat{p}}) \\
0{\cdot}03 &= \pm(z)\sqrt{\frac{\hat{p}(1-\hat{p})}{n}} \\
0{\cdot}03 &= \pm(z)\sqrt{\frac{0{\cdot}7(1-0{\cdot}7)}{896}} \\
0{\cdot}03 &= \pm(z)(0{\cdot}153) \\
\frac{0{\cdot}03}{0{\cdot}153} &= \pm z \\
\pm 1{\cdot}96 &= z
\end{aligned}$$

Hence, the poll is using the 95% level of confidence.

95% confidence interval for the population mean

The central limit theorem may also be applied to form a confidence interval for the mean of a population, given the mean of a large enough sample, and a standard deviation.

A population has a mean of μ and a standard deviation of σ. Suppose a sample of size $n \geq 30$ has a mean of $\bar{x}$. Then the 95% confidence interval for the population mean, μ, is:

$$\bar{x} - 1{\cdot}96\frac{\sigma}{\sqrt{n}} \leq \mu \leq \bar{x} + 1{\cdot}96\frac{\sigma}{\sqrt{n}}$$

In practice, if the standard deviation of the population, σ, is not known, then we use the standard deviation, s, of the sample in its place.

exam Q

A survey was carried out to find the weekly rental costs of holiday apartments in a certain country. A random sample of 400 apartments was taken. The mean of the sample was €320 and the standard deviation was €50.

Form a 95% confidence interval for the mean weekly rental costs of holiday apartments in that country.

Solution

Let μ = the mean weekly rental of all holiday apartments. For the sample, $\bar{x} = 320$, the standard deviation $s = 50$ and $n = 400$.

We use $s = 50$ because the standard deviation, σ, of the population is not available.

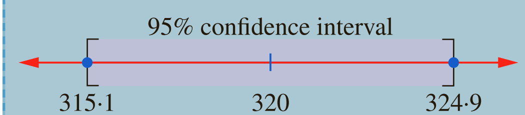
SE of the mean $= \dfrac{s}{\sqrt{n}} = \dfrac{50}{\sqrt{400}} = 2{\cdot}5.$

Then the 95% confidence interval for the mean weekly rental of all apartments is given by:

$$\bar{x} - 1{\cdot}96\left(\frac{s}{\sqrt{n}}\right) \leq \mu \leq \bar{x} + 1{\cdot}96\left(\frac{s}{\sqrt{n}}\right)$$

$$320 - 1{\cdot}96(2{\cdot}5) \leq \mu \leq 320 + 1{\cdot}96(2{\cdot}5)$$

$$315{\cdot}1 \leq \mu \leq 324{\cdot}9$$



Hence, the 95% confidence interval for the mean weekly rental cost is from €315·10 to €324·90.

Hypothesis testing of the population proportion using the more accurate standard error

Hypothesis testing is a technique used in statistics to test whether a claim that is made is consistent with the data obtained.

You could refer back to the material on hypothesis testing in the previous chapter before tackling the next two examples.

Example 1

National data in 1970 showed that 58% of the adult population had never smoked cigarettes. In 2010, a national health survey interviewed a random sample of 880 adults and found that 52% had never smoked cigarettes.

(i) Construct a 95% confidence interval for the proportion of adults in 2010 who had never smoked cigarettes.

(ii) Does this provide evidence of a change in behaviour among the Irish?

Write appropriate hypotheses.

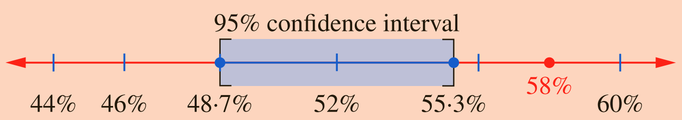
Using your confidence interval, test an appropriate hypothesis and state your conclusion.

Solution

(i) Use $$\hat{p} - E \leq p \leq \hat{p} + E$$

where
$$\hat{p} = 52\% = 0{\cdot}52$$
$$1 - \hat{p} = 1 - 0{\cdot}52 = 0{\cdot}48$$
$$n = 880$$
$$z = 1{\cdot}96$$

Then $$0{\cdot}52 - 1{\cdot}96\sqrt{\frac{(0{\cdot}52)(0{\cdot}48)}{880}} \leq p \leq 0{\cdot}52 + 1{\cdot}96\sqrt{\frac{(0{\cdot}52)(0{\cdot}48)}{880}}$$
$$0{\cdot}52 - 0{\cdot}033 \leq p \leq 0{\cdot}52 + 0{\cdot}033$$
$$48{\cdot}7\% \leq p \leq 55{\cdot}3\%$$



Based on these data, we are 95% confident that the proportion of adults in 2010 who had never smoked cigarettes is between 48·7% and 55·3%.

(ii) $H_0: p = 58\%$

$H_A: p \neq 58\%$

Since 58% is not in the confidence interval (see the diagram above), we reject H_0.

We conclude that the proportion of adults in 2010 who had never smoked was less than in 1970.

Example 2

A study addressed the issue of whether pregnant women can correctly guess the sex of their baby.

Among a random sample of 312 pregnant women, 171 correctly guessed the sex of the baby.

(i) Construct a 95% confidence interval from the given data.

(ii) Use these sample data to test the claim, at the 5% level of significance, that the success rate of such guesses is no different from the 50% success rate expected with random chance guesses.

Solution

(i) $\hat{p} = \dfrac{171}{312} = 0{\cdot}548 \ (= 54{\cdot}8\%)$

$1 - \hat{p} = 1 - 0{\cdot}548 = 0{\cdot}452$

$n = 312$

$z = 1{\cdot}96$

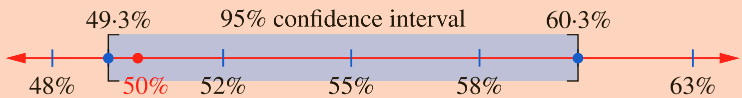
Use $\qquad \hat{p} - E \le p \le \hat{p} + E$

$$0{\cdot}548 - 1{\cdot}96\sqrt{\frac{(0{\cdot}548)(0{\cdot}452)}{312}} \le p \le 0{\cdot}548 + 1{\cdot}96\sqrt{\frac{(0{\cdot}548)(0{\cdot}452)}{312}}$$

$$0{\cdot}548 - 0{\cdot}055 \le p \le 0{\cdot}548 + 0{\cdot}055$$

$$0{\cdot}493 \le p \le 0{\cdot}603$$

$$49{\cdot}3\% \le p \le 60{\cdot}3\%$$



Based on the given data, we are 95% confident that the proportion of pregnant women who correctly predicted the sex of their baby was between 49·3% and 60·3%.

(ii) H_0, the null hypothesis: the success rate is no different from 50%.

$H_0: p = 0{\cdot}5 = 50\%$

$H_A: p \neq 0{\cdot}5$

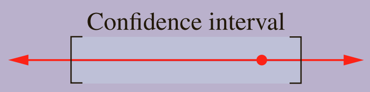
As 50% is included in the confidence interval above we fail to reject the null hypothesis.

Hence, we conclude that there is not sufficient evidence to warrant rejection of the claim that women can correctly guess the sex of their baby.

exam focus

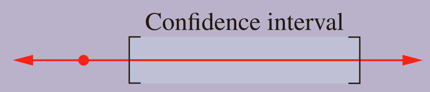
In the final analysis, testing the null hypothesis, H_0, simply involves a confidence interval and a red dot.

Either



If the red dot is inside the confidence interval, we fail to reject H_0.

Or



If the red dot is outside the confidence interval, we reject H_0.

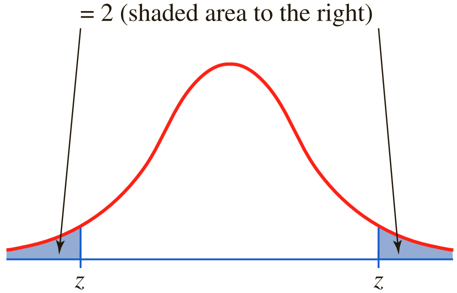
p-Values: An alternative approach to hypothesis testing

A *p*-value is:

- a probability
- used to make a decision on the null hypothesis, H_0
- a measure of the strength of evidence to reject or fail to reject the null hypothesis.

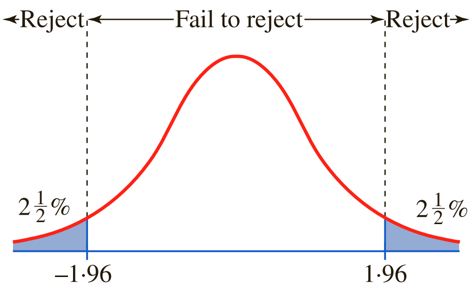
***p*-value**

p-value = Sum of two equal shaded regions
= 2 (shaded area to the right)



Critical values for *R*

(at the 5% level of significance)



The decision to reject, or fail to reject, H_0 is based on the comparison of the *p*-value with the level of significance. On our course we only use a two-tailed test at the 5% level of significance.

exam focus

It is vital to know that the critical *p*-value = 0·05 at the 5% significance level.

If $p \leq 0{\cdot}05$, there is strong evidence to reject H_0.
If $p > 0{\cdot}05$, there is strong evidence to fail to reject H_0.

How to perform a hypothesis test using *p*-value

1. State H_0 and H_A.
2. Calculate the z score (this is often called the test statistic, T).
3. Determine the p-value (a diagram is useful).
4. If $p \leq 0{\cdot}05$, reject H_0. If $p > 0{\cdot}05$, fail to reject H_0.
5. State the conclusion in words.

Example 1

A machine produces metal rods which have a mean length of 500 cm with a standard deviation of 4 cm. After a service to the machine, it is claimed that the machine now produces rods with lengths that are not equal to 500 cm. To test the claim, a random sample of 100 rods from the serviced machine are measured and found to have a mean length of 500·5 cm.

(i) Write down H_0 and H_A.

(ii) Calculate the test statistic for this sample mean.

(iii) Calculate a p-value for this sample mean.

(iv) At the 5% level of significance, is there evidence to show that the mean length of the metal rod from the serviced machine is not 500 cm? Justify your answer.

Solution

(i) The null hypothesis, H_0: $\mu = 500$ cm.

The alternative hypothesis, H_A: $\mu \neq 500$ cm.

(ii) $\bar{x} = 500{\cdot}5$, $\mu = 500$, $\sigma = 4$ and $n = 100$

The test statistic is given by:

$$T = \frac{\bar{x} - \mu}{\frac{\sigma}{\sqrt{n}}} = \frac{500{\cdot}5 - 500}{\frac{4}{\sqrt{100}}} = 1{\cdot}25$$

(iii) $p(T > 1{\cdot}25)$

$= 1 - P(z \leq 1{\cdot}25)$

$= 1 - 0{\cdot}8944$ (from tables)

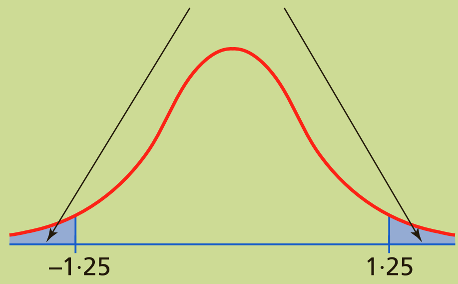
$= 0{\cdot}1056$

The p-value $= 2\,(0{\cdot}1056)$

$= 0{\cdot}2112$

key point

The p-value = sum of the two shaded regions.



(iv) Since $0{\cdot}2112 > 0{\cdot}05$ (or $21{\cdot}12\% > 5\%$), we conclude there is strong evidence not to reject the null hypothesis

We state that we fail to reject the claim that the mean length of the metal rods from the serviced machines is not 500 cm.

Example 2

A company claims that the average weight of a packet of cereal it produces is 400 g with a standard deviation of 12 g. To test this claim, a random sample of 64 of these packets were weighed and found to have a mean value of 403 g.

(i) Write down H_0 and H_A.

(ii) Calculate the test statistic for this sample mean.

(iii) Calculate a p-value for this sample mean.

(iv) At the 5% level of significance, is there evidence to show that the mean weight of the packets of cereal is not 400 g? Justify your answer.

Solution

(i) The null hypothesis, H_0: $\mu = 400$ g

The alternative hypothesis, H_A: $\mu \neq 400$ g.

(ii) $\bar{x} = 403$, $\mu = 400$, $\sigma = 12$ and $n = 64$

The test statistic is given by:

$$T = \frac{\bar{x} - \mu}{\frac{\sigma}{\sqrt{n}}} = \frac{403 - 400}{\frac{12}{\sqrt{64}}} = 2$$

(iii) $P(T > 2)$

$= 1 - P(z \leq 2)$

$= 1 - 0{\cdot}9772$ (from tables)

$= 0{\cdot}0228$

The p-value $= 2\,(0{\cdot}0228) = 0{\cdot}0456$

(iv) Since $0{\cdot}0456 < 0{\cdot}05$ (or $4{\cdot}56\% < 5\%$), we conclude there is strong evidence to reject the null hypothesis, H_0.

We state that there is strong evidence to reject the claim by the company that the average weight of a packet of cereal is 400 g.

exam focus

- If p is low, H_0 must go.
- The lower the p-value, the stronger the evidence against H_0.
- The larger the sample size, the more precise the estimate.