Section 4.4 Margin of error – Confidence intervals – Hypothesis testing

When dealing with sampling in *Statistics 1*, it was stated that the purpose of sampling is to gain information about the whole population by surveying a small part of the population. If data from a sample is collected in a proper way, then the sample survey can give an accurate indication of the population characteristic that is being studied.

Before a general election, a national newspaper generally requests a market research company to survey a sample of the electorate regarding their voting intentions in the election. The number surveyed is generally about 1000.

The result of the survey might appear in the daily newspaper as follows:

40% support for **The Democratic Right**.

The **40%** support is called the **sample proportion**, that is, the part or portion of the sample who indicated that they would vote for **The Democratic Right**.
A sample proportion is used to give an estimate of the **population proportion** who intend to vote for **The Democratic Right**.

The notation $\hat{p}$ is used to denote **sample proportion**.
The notation p is used to represent **population proportion**.
Since p is generally not known, $\hat{p}$ is used as an **estimor** for the true population proportion, p.

Of course everybody knows that sample surveys are not always 100% accurate. There is generally some 'element of chance' or **error** involved.

The newspaper might add to their headline the following sentence:

The margin of error is 3%.

The **margin of error** of 3% is a way of saying that the result of the survey is 40% **± 3%**. That means that the research company is quite 'confident' that the proportion of the whole electorate who intend to vote for **The Democratic Right** could be anywhere between 37% and 43%.

How does the research company calculate 'the margin of error'?

The margin of error in opinion polls is generally calculated using the formula,

$$E = \frac{1}{\sqrt{n}}, \text{ where } n \text{ is the sample size.}$$

Margin of error

$$E = \frac{1}{\sqrt{n}}$$

If the sample size is 1000, then $E = \frac{1}{\sqrt{1000}} \approx 3\%$.

If the sample size is increased, the margin of error will be reduced.

Confidence interval

The result of the opinion poll above was given as 40% $\pm$ 3%.
That could be written as $37\% < p < 43\%$, where p is the population proportion.
$37\% < p < 43\%$ is called the **confidence interval**.

The 'confidence' level is pitched at 95%.
The 95% confidence implies that the interval was obtained by a method which 'works 95% of the time'.

The confidence interval, $37\% < p < 43\%$, is a way of stating that if you surveyed many samples of 1000 people on the same day, the results would be in the interval 37% to 43% in 95% of the samples.

In our course, the confidence level is always at 95%.

At the 95% confidence level, the confidence interval for a population proportion is given on the right.

Confidence interval is

$$\hat{p} - \frac{1}{\sqrt{n}} < p < \hat{p} + \frac{1}{\sqrt{n}}$$

The confidence interval above may be also expressed as $\hat{p} \pm \frac{1}{\sqrt{n}}$.

Example 1

What sample size would be required to have a margin of error of

(i) 0.05 (ii) $2\frac{1}{2}\%$?

$$E = \frac{1}{\sqrt{n}} \Rightarrow n = \frac{1}{E^2}$$

(i) $n = \dfrac{1}{(0.05)^2} = \dfrac{1}{20} = 5\%$

(ii) $n = \dfrac{1}{(0.025)^2} = 1{,}600$

Example 2

A random sample of 400 persons are given a flu vaccine and 136 of them experienced some discomfort.
Construct a 95% confidence interval, p, for the population proportion who might experience discomfort.

$$E = \frac{1}{\sqrt{n}} = \frac{1}{\sqrt{400}} = 5\%$$

$$p = \frac{136 \times 100}{400} = 34\%$$

Confidence interval of 95%

$$29\% < p < 39\%$$

Example 3

A survey of 100 residents of a Dublin suburb were asked if they remembered seeing an advertisement for McCain's chips on television. 60 respondents said that they had.

(i) Calculate the sample proportion, $\hat{p}$.

(ii) Find the margin of error, E.

(iii) Construct a 95% confidence interval for p.

(i) $\hat{p} = \frac{60}{100} = 60\%$

(ii) $E = \frac{1}{\sqrt{n}} = \frac{1}{\sqrt{100}} = 10\%$

(iii) 95% confidence interval

$$50\% < p < 70\%$$

Hypothesis testing

A **hypothesis** is a statement or conjecture made about some statistic or characteristic of a population.

Here is an example of a hypothesis:

'A football team is most likely to concede a goal just after it has scored a goal'.

A **hypothesis test** is a statistical method of proving the truth or otherwise of the statement or claim.

A local council reduced the speed limit on a dangerous 8 km stretch of country road from 80 km/hr to 60 km/hr. The number of accidents on the stretch was reduced from 5 per mon to 3 per month. The council claimed that the speed reduction was effective. Is the council correct in its claim?

In cases like this, a hypothesis test is set up to prove or disprove the claim.

Procedure for carrying out a hypothesis test

The procedure for carrying out a hypothesis test will involve the following steps:

1. Write down $\mathbf{H_0}$, the **null hypothesis**, and $\mathbf{H_1}$, the **alternative hypothesis**
   For example, to test if a coin is biased if we get 7 heads in 10 tosses, we could formulate the following hypothesis:
   H_0 : The coin is not biased.
   H_1 : The coin is biased.

2. Write down or calculate the sample proportion, $\hat{p}$.

3. Find the margin of error.

4. Write down the confidence interval for p, using

$$\hat{p} - \frac{1}{\sqrt{n}} < p < \hat{p} + \frac{1}{\sqrt{n}}$$

5. (i) If the value of the population proportion stated is within the confidence interval, accept the null hypothesis H_0 and reject H_1.
   (ii) If the value of the population proportion is outside the confidence interval, reject the null hypothesis H_0 and accept H_1.

Example 4

A drugs company produced a new pain-relieving drug for migraine sufferers and claimed that the drug had a 90% success rate. A group of doctors doubted the company's claim. They prescribed the drug for a group of 150 patients. After six months, 120 of these patients said that their migraine symptoms had been relieved by the drug.
At the 95% level of confidence, can the company's claim be upheld?

$$\hat{p} = \frac{120 \times 100}{150} = 80\%$$

$$E = \frac{1}{\sqrt{n}} = \frac{1}{\sqrt{150}} = 8\%$$

95% Confidence interval

$$72\% < \hat{p} < 88\%$$

H_0 = drug is not 90% Successful
H_1 = drug is 90% Successful

$\therefore$ H_0 is Supported

Example 5

A coin is tossed 1000 times and heads occur 550 times.
At the 95% confidence level, does the result indicate that the coin is biased?

$$E = \frac{1}{\sqrt{n}} = \frac{1}{\sqrt{1000}} = 3\%$$

H_0 = coin not biased : H_1 = coin is biased.

Expected outcome for H_0 = 50% or 500 times out of 1000.

$\hat{p} = 550$ or 55%

Confidence level of 95% $\rightarrow$ $52\% < p < 58\%$

$\Rightarrow$ this suggests that H_1 is true