6.

Sample proportions

- Variation between samples
- Sample proportion distribution
- How would we determine the distribution of $\hat{p}$ if we don't know p?
- Why is it useful to know how the sample proportions are distributed?
- Confidence intervals
- Margin of error
- Sample size
- Increasing the level of confidence increases the margin of error
- Let's check
- Miscellaneous exercise six

Did you know that (at the time of writing):

Approximately 78% of the people living in Western Australia live in Perth.

Approximately 64% of the people who live in New South Wales live in Sydney.

Approximately 38% of Labor party MPs (federal) are women.

Approximately 10% of people in the world are left handed.

In Australia approximately 32% of human births are by caesarean section.

China won almost 13% of the gold medals awarded at the London Olympics.

Approximately 17% of the world's population live in India.

Approximately 71% of the Earth's surface is covered by water.

More than 80% of Australians live within 100 kilometres of the sea.

In discussion with others suggest proportions that you think appropriate for each of the following:

What proportion of the world's population are Chinese?

What proportion of Australian adults have never held a driver's licence?

What proportion of homes in Australia are double storey?

What proportion of Australians live in Victoria?

What proportion of Australians are female?

What proportion of Australian families own a dog?

What proportion of Australian married couples never have children?

What proportion of Australian families own more than one car?

What proportion of the world's adults are vegetarian?

What proportion of the world's population celebrate Christmas?

iStock.com/rzarek

As you may have realised from the items on the previous page, and indeed the title of the chapter, we are now considering proportions of a population.

Suppose we wanted to know what proportion of the Australian population engaged in a particular activity, for example following a vegetarian diet or being a regular attender at a gym, or what proportion possessed a particular characteristic, for example being aged over 60 or being left handed etc. We could include a suitable question in the census so that all of the population is checked for this activity or characteristic, or we could consider a sample that is representative of the whole population and use the proportion from the sample to suggest the proportion for the population. For example, if we found that in a sample of 200 Australians, 19 were left handed, i.e. 9.5% of our sample were left handed, we could suggest that 9.5% of Australians were left handed, or, as a proportion of the whole, 0.095. However, if we were to choose a different sample we could well obtain a different proportion. How different?

How different might the proportions be from one sample to another?

How might the proportions from a number of such samples be distributed?

How different might the proportion in a sample be from the proportion that exists in the population?

If we had the proportions from a number of samples we could combine the results to suggest the proportion of Australians who were left handed. (If the other samples also each involved 200 Australians this would simply be the mean of the proportions.)

The proportion of left handers in samples we choose is a random variable, capable of taking any value from 0 to 1, but probably close to the proportion that exists in the whole population, especially if the sample is large. The proportion in the population is the **population proportion**, sometimes simply referred to as p. The proportion from our random sample varies according to the sample we choose and is called the **sample proportion**, sometimes written as $\hat{p}$ (pronounced 'p hat').

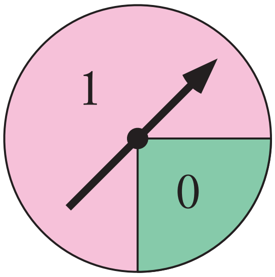
How close would our sample proportion, $\hat{p}$, be to the population proportion, p?

Is the size of our sample of any significance?

We have posed a lot of questions about the proportion of left handed people in our sample. Let us now consider this idea of 'sample proportions' further, and let us start by considering something for which we know the value of p, the population proportion, or long term proportion.

Variation between samples

Consider the spinner on the right.



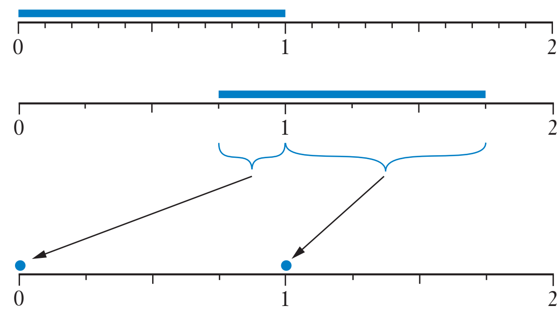
From the geometry of the spinner we know that in the long term, the proportions of spins that will result in a 1 will be 0.75.

Now suppose we simulate 20 spins of such a spinner, a number of times, and see how the proportion of 1s in our samples of 20 spins vary.

Question: How could we simulate the spinning of such a spinner?
Answer: We could use random numbers, as shown on the next page.

ISBN 9780170395205

To change from a random number with an output in the range zero to one, to an output of either zero or one, with P(1) = 0.75 we can proceed as follows:



Usual random number output is between 0 and 1:

Add 0.75 to obtain numbers between 0.75 and 1.75.

Display only the integer part of such numbers.

This would make 1 three times as likely as 0.

The display below left shows six ones or zeros generated in this way.

The display below right shows the generation of 20 such numbers (not all of which are displayed), summing them (to determine the number of 1s) and expressing the number of 1s as a proportion of the 20 spins. In this case, 80% of the 20 spins were 1s. For this sample $\hat{p} = 0.8$.

```
Int(Ran# + 0.75)
                         1
                         1
                         0
                         1
                         1
                         0
```

```
int(randList(20) + 0.75)
        {1,1,1,0,1,0,1,0,1,1,1,1 ▷
sum(ans) / 20
                       0.8
```

Alternatively we could generate zeros and ones according to a Binomial distribution, Bin (1, 0.75). With one trial involved we have either 0 successes or 1 success and with P(success) = 0.75 we will have

$$P(0) = 0.25 \text{ and } P(1) = 0.75$$

as required.

For the sample on the right $\hat{p} = 0.7$.

```
randBin(1,0.75,20)
        {0,1,1,1,1,1,0,0,1,1,1,1 ▷
sum(ans) / 20
                       0.7
```

Three further samples of twenty spins, together with the proportion of 1s obtained, are given below:

1 1 1 1 0 1 1 0 0 1 1 1 1 1 1 1 1 1 1 1

Proportion of 1s: 0.85, i.e., for this sample, $\hat{p} = 0.85$.

1 0 1 1 1 1 0 1 0 1 1 1 0 1 1 1 1 1 1 1

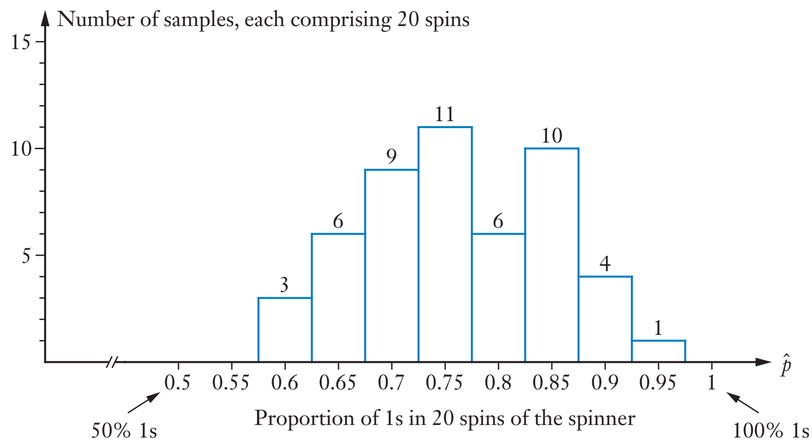
Proportion of 1s: 0.8, i.e., for this sample, $\hat{p} = 0.8$.

0 0 1 0 1 0 0 1 0 1 1 1 1 0 1 0 1 1 1 1

Proportion of 1s: 0.6, i.e., for this sample, $\hat{p} = 0.6$.

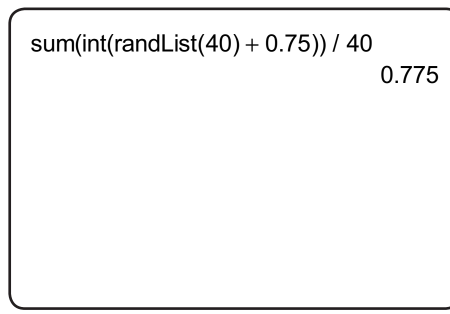
ISBN 9780170395205

The graph below shows the results of 50 such samples, each involving 20 simulated spins of the spinner, with the distribution of the proportion of 1s as shown.

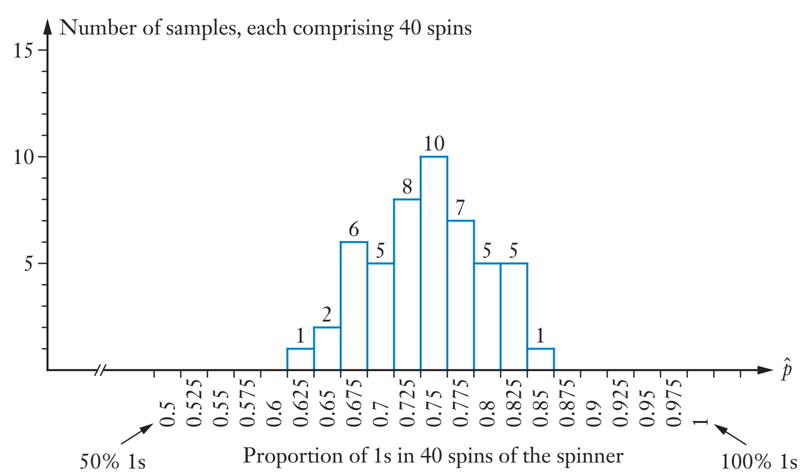


Suppose instead we again collected 50 samples but we now increase the number of spins in each sample to, say 40, rather than 20.

Check that you understand the display on the right, which indicates that the proportion of 1s in the single simulated sample of 40 spins of our spinner was 0.775.



The graph below shows the distribution of the proportion of 1s for a simulation involving 50 samples, each sample comprising 40 spins. (The display above, with its sample proportion of 0.775, is just one of these 50 samples, there being 7 with sample proportion of 0.775.)



ISBN 9780170395205

Notice that both graphs of $\hat{p}$ for multiple samples appear somewhat 'bell shaped' and both peak around the theoretical long term proportion of 0.75.

Using the information given on each graph, and with the assistance of a calculator we can determine that the first graph has a mean of 0.762 and a standard deviation of 0.09,

	List 1	List 2	List 3	List 4
1	0.5	0		
2	0.55	0		
3	0.6	3		
4	0.65	6		

1VAR 2VAR SET

→

$\bar{x}$	= 0.762
Σx	= 38.1
Σx^2	= 29.42
$x\sigma_n$	= 0.08806815
$x\sigma_{n-1}$	= 0.08896227
n	= 50

whilst the second graph has a mean of 0.7435 and a standard deviation of 0.053.

	List 1	List 2	List 3	List 4
5	0.6	0		
6	0.625	1		
7	0.65	2		
8	0.675	6		

1VAR 2VAR SET

→

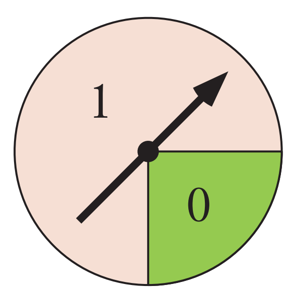
$\bar{x}$	= 0.7435
Σx	= 37.175
Σx^2	= 27.779375
$x\sigma_n$	= 0.05287012
$x\sigma_{n-1}$	= 0.05340689
n	= 50

Thus it would appear that as we increase the sample size, the mean of the sample proportions gets closer to the theoretical mean and the sample proportions are less spread out.

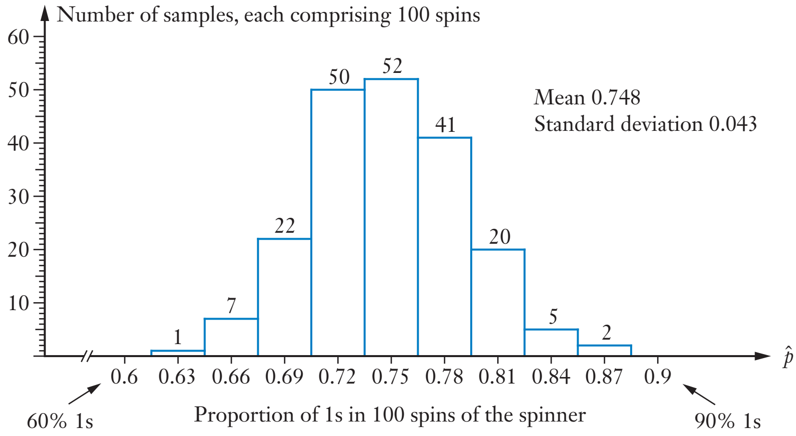
If we increase both the number of spins in each sample, i.e. we increase the sample size, and we increase the number of samples, the shape of the histogram becomes more like the characteristic bell shape of the normal distribution.

The graph on the next page shows the distribution of the proportion of 1s in two hundred samples with each sample involving one hundred spins of the spinner.

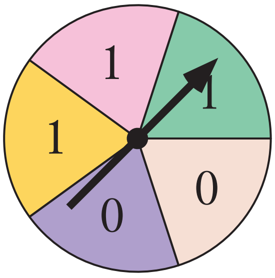
(The graph involves grouped data on the horizontal axis with the mean and the standard deviation calculated from the grouped data.)



ISBN 9780170395205



SIMULATION – YOUR TURN



If we made repeated spins of the spinner shown on the right we would expect, in the long run, that the proportion of spins giving a 1 would be 0.6, i.e. we would expect that approximately 60% of the spins would result in a 1.

Step 1: Using the random number generating facility of some calculators or computer spreadsheets, simulate twenty spins of this spinner and note the proportion of 1s.

Step 2: Repeat step 1 nine more times, to give ten samples in all, with each sample involving twenty spins of this spinner, and for each sample note the proportion of 1s.

Step 3: Find the mean and standard deviation of the ten proportion values you have from carrying out the previous steps.

Is your mean close to the expected long term value of 0.6?

Step 4: Simulate fifty spins of this spinner and note the proportion of 1s.

Is your proportion close to the expected long term value of 0.6?

Step 5: Repeat step 4 twenty-nine more times, to give thirty samples in all, with each sample involving fifty spins of this spinner, and for each sample note the proportion of 1s.

Step 6: Find the mean and standard deviation of the thirty proportion values you have from carrying out steps 4 and 5.

Is your mean close to the expected long term value of 0.6?

Is the standard deviation for the proportions of 1s with a sample size of 50 less than the standard deviation of the proportions with a sample size of 20?

Compare your response to this question with those of others in your class.

ISBN 9780170395205

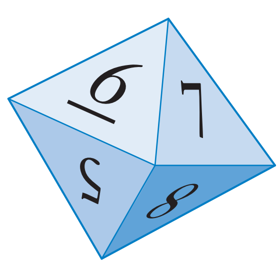
If we were to roll a normal fair octagonal die, ten times, what proportion of the ten rolls would you expect to give an odd number?

Well, with an equal number of odd {1, 3, 5, 7} and even {2, 4, 6, 8} numbers we would expect approximately half of the ten rolls to give an odd number. But what variation should we expect in this proportion of odd numbers? The results of twenty sample rolls of such a die, with each sample involving ten rolls are shown below, together with the proportion of odd numbers obtained in each sample.

Sample 1:	3	4	1	2	5	5	3	6	4	1	Proportion of odd numbers	0.6
Sample 2:	8	4	2	7	8	4	5	7	6	1	Proportion of odd numbers	0.4
Sample 3:	2	7	5	1	8	7	8	3	8	5	Proportion of odd numbers	0.6
Sample 4:	4	2	7	6	1	1	6	3	1	6	Proportion of odd numbers	0.5
Sample 5:	4	8	1	1	2	2	6	4	5	4	Proportion of odd numbers	0.3
Sample 6:	2	1	8	1	7	4	6	4	5	2	Proportion of odd numbers	0.4
Sample 7:	2	2	7	8	6	8	3	4	8	4	Proportion of odd numbers	0.2
Sample 8:	8	5	5	6	8	1	5	1	1	3	Proportion of odd numbers	0.7
Sample 9:	2	1	1	3	2	6	3	5	8	1	Proportion of odd numbers	0.6
Sample 10:	4	2	7	1	5	1	3	4	8	5	Proportion of odd numbers	0.6
Sample 11:	7	1	2	1	4	7	3	7	1	1	Proportion of odd numbers	0.8
Sample 12	3	3	7	8	2	2	7	7	8	5	Proportion of odd numbers	0.6
Sample 13:	5	3	1	8	6	7	7	1	6	1	Proportion of odd numbers	0.7
Sample 14:	7	2	4	3	6	2	5	1	8	3	Proportion of odd numbers	0.5
Sample 15:	5	7	7	2	2	3	1	8	8	1	Proportion of odd numbers	0.6
Sample 16:	1	3	2	1	1	8	1	6	7	6	Proportion of odd numbers	0.6
Sample 17:	7	7	4	4	7	8	5	4	5	1	Proportion of odd numbers	0.6
Sample 18:	7	2	4	2	2	5	2	6	6	7	Proportion of odd numbers	0.3
Sample 19:	8	7	8	7	5	2	2	6	2	3	Proportion of odd numbers	0.4
Sample 20:	6	5	7	1	3	1	4	3	1	2	Proportion of odd numbers	0.7

Our 20 sample proportions have a mean of 0.535
and a standard deviation, σ_n, of 0.153.

SIMULATION – YOUR TURN



What might the mean and standard deviation of the sample proportions be if we still have 20 samples but now the sample size is 20, or 50, or …?

Investigate.

However, before you do, read the next page for some suggestions that may help you create the simulated die rolling.

ISBN 9780170395205

Simulation method 1

Some calculators can produce a list of random integers between two values, for example 1 to 8 inclusive. Hence, to simulate the rolling of the octagonal die ten times we could instruct such a calculator to list ten integers in this way. We would then note what proportion of the numbers are odd numbers.

```
randList (10, 1, 8)
                {1, 4, 5, 2, 7, 3, 1, 1, 1, 4}
```

```
randInt (1, 8, 10)
                {8, 7, 8, 7, 5, 2, 2, 6, 2, 3}
```

(Alternatively we could use the ability of computer spreadsheets and calculators to generate random numbers between 0 and 1 but alter the output appropriately, as explained earlier in this book.)

Simulation method 2

With each roll of the die equally likely to produce an odd as it is an even, we could simply output 10 numbers that are either 0 (for even) or 1 (for odd), sum the list (to give the total number of odd numbers) and then divide by 10 to give the proportion of odd numbers.

```
sum(randList (10, 0, 1))/10
                                0.6
```

Simulation method 3

Suppose you wish to keep the authentic feel to the simulation by actually simulating the ten rolls and then checking for each number being even or odd, which method 2 lacks. In that case first note that if we were to divide a randomly generated integer by 2 the remainder is 0 when the integer is even, and 1 when the integer is odd. Some calculators and computer spreadsheets have the facility to return a zero if a number is even and a 1 if the number is odd. Applying this facility to a list of numbers, and then summing the 0s and 1s, gives a count of the odd numbers.

Such a function is sometimes called *remain* (as in *remain*der) *mod* (as in *mod*ulo arithmetic) or perhaps *iMod*.

So if we

- generate an appropriate list of randomly generated integers,
- apply the remain, mod, or iMod function, to return a 1 for an odd number and a 0 for an even number,
- sum the 0s and 1s produced,
- divide by the sample size,

we can quickly generate appropriate proportions.

See the displays at the top of the next page for the sort of calculator language needed for this task. The display top left shows the step by step process for one sample of ten numbers and top right shows two 'all at once' formulae each generating the proportion of odd numbers in a sample of ten random integers between 1 and 8 inclusive.

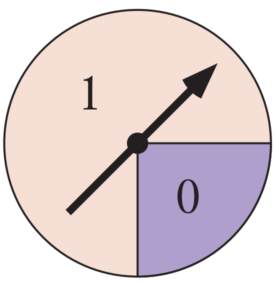
ISBN 9780170395205

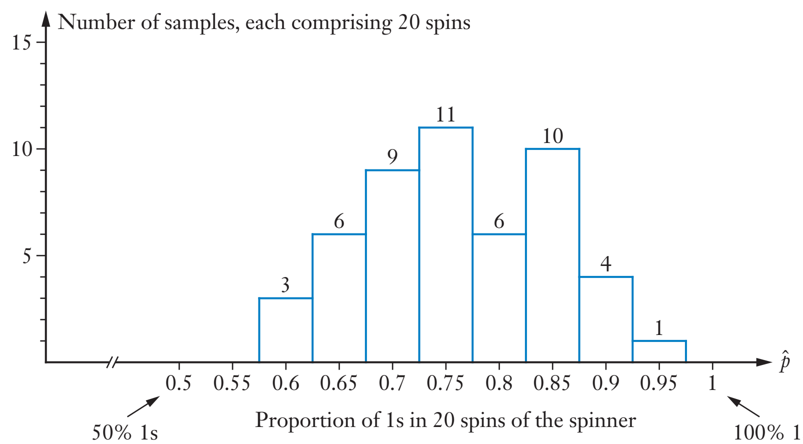
```
randList (10, 1, 8)
                    {6, 4, 4, 2, 3, 5, 3, 7, 4, 1}
iMod (ans, 2)
                    {0, 0, 0, 0, 1, 1, 1, 1, 0, 1}
sum (ans)/10
                                               0.5
```

```
sum(mod(randInt (1, 8, 10), 2))
              10
                                               0.6
sum(remain(randInt (1, 8, 10), 2))
              10
                                               0.5
```

Sample proportion distribution

Earlier in this chapter we considered 50 samples each involving 20 spins of the spinner shown below. For each sample we determined the proportion of the 20 spins resulting in a 1.





Now we know from the geometry of the spinner that in the long term the proportion of 1s will be 0.75. Our sample values estimated this long term proportion and gave us a distribution with a mean of 0.762 and a standard deviation of 0.09.

Counting the number of 'successes' in n trials involves a binomial distribution Bin(n, p), where p is the probability of success. This has mean np, standard deviation $\sqrt{np(1-p)}$.

The *proportion* of successes simply divides the number of successes by n. Hence the distribution of sample proportions will have mean p, standard deviation $\sqrt{p(1-p)/n}$.

In fact, according to a statistical rule called the Central Limit Theorem:

As the sample size, n, increases, the distribution of the sample proportions will approach that of a normal distribution, mean p, standard deviation $\sqrt{p(1-p)/n}$.

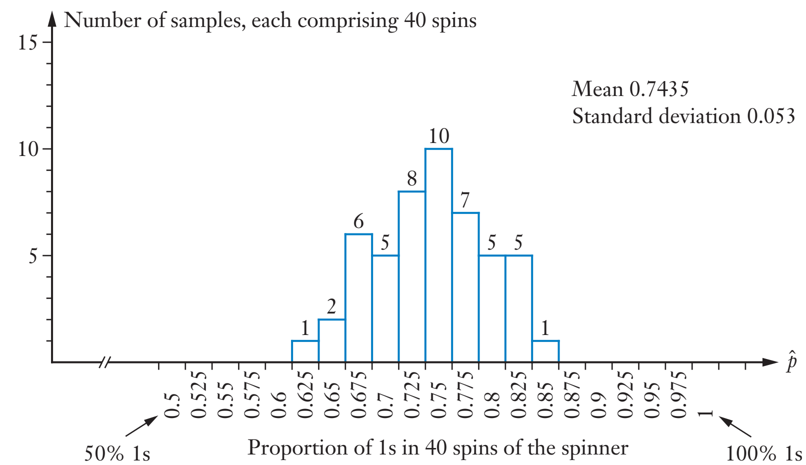
Thus in our situation with $p = 0.75$, and $n = 20$, the theory suggests that the proportion of 1s would approach a normal distribution with mean 0.75 and standard deviation 0.097. Our values of 0.762 and 0.09 are reasonably close to the theoretical model. However, the graph does not look especially normal. This is because the approximation to the normal distribution improves as the sample size increases. To be confident that the sample distributions will approximate to the stated normal distribution we need both $np \geq 10$ and $n(1-p) \geq 10$. (Indeed some statistical textbooks even suggest ≥ 15.)

For this same spinner we also considered sample sizes of 40 and 100:

With $p = 0.75$ and $n = 40$ the rule suggests that our distribution of sample proportions should be approximately normal with mean 0.75 and standard deviation

$$\sqrt{\frac{0.75(1-0.75)}{40}} = 0.068$$

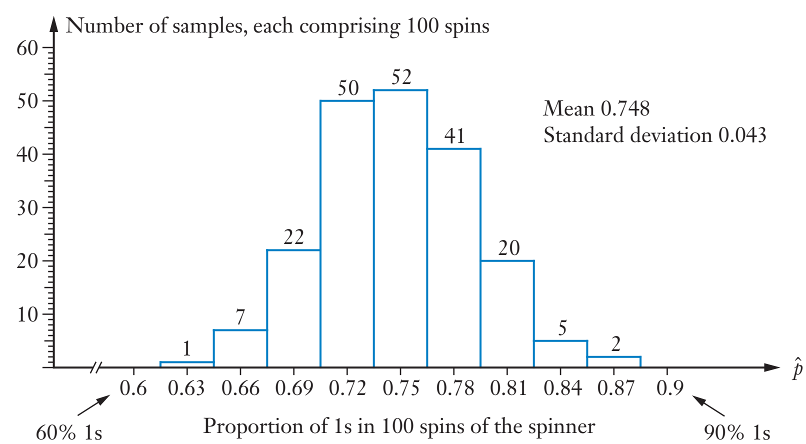
We obtained:



With $p = 0.75$ and $n = 100$ the rule suggests that our distribution of sample proportions should be approximately normal with mean 0.75 and standard deviation

$$\sqrt{\frac{0.75(1-0.75)}{100}} = 0.0433.$$

We obtained:



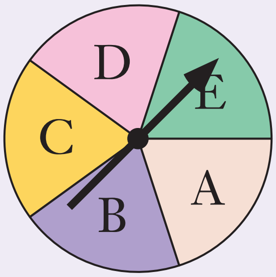
For increasing sample size, and with both np and $n(1 - p) \geq 10$, the graphs above do appear to be getting more normally distributed and the means and standard deviations are roughly as the rule would have us expect.

ISBN 9780170395205

How do your results for the simulation of the proportion of odd numbers obtained when rolling an octagonal die compare with the normal distribution that we would expect?
Does the distribution of sample proportions approximate towards the expected normal distribution as the sample size gets larger, with np and $n(1-p) \geq 10$?

EXAMPLE 1

When the spinner on the right was spun 200 times an A occurred on 43 occasions.



a What is the value of p, the population proportion of As?

b What is the value of $\hat{p}$, the sample proportion of As?

c Calculate the mean and standard deviation of $\hat{p}$ for such samples of 200 spins.

Solution

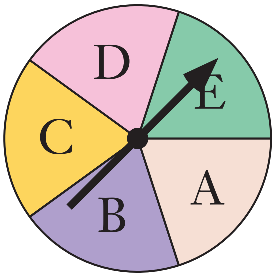
a In the long term we would expect an A to occur 20% of the time.
Thus $p = 0.2$

b In our sample of 200 spins an A occurred on 43 occasions.
Thus $\hat{p} = 0.215$.

c $\hat{p}$ will have a mean equal to p and a standard deviation of $\sqrt{\frac{p(1-p)}{n}}$, with $p = 0.2$, and $n = 200$.

Thus $\hat{p}$ has a mean of 0.2 and a standard deviation of 0.0283, correct to 4 dp.

How would we determine the distribution of $\hat{p}$ if we don't know p?



It is all very well considering the proportion of 1s that might occur when a normal die is rolled, or the proportion of As that will occur when the spinner on the right is spun a number of times, because for these sort of random events the long term proportion, or population proportion, p, is known. However it will often be the case that the population proportion, p, will not be known and we will be sampling from the population in order to estimate p. If we need to give a single value estimate of the population proportion we would use the best estimate we have available, i.e our sample proportion. (Though we will see later in this chapter that we can instead give an *interval* in which the population proportion is likely to lie.) Similarly, to estimate the standard deviation of the distribution of sample proportions we can again use our best estimate for p, i.e. $\hat{p}$, in the formula for standard deviation. For large n this use of $\hat{p}$ for p, is still likely to give a good approximation for the standard deviation (and after all, if we only had one sample, and not knowing p, it would be the best we can do).

ISBN 9780170395205

EXAMPLE 2

A survey was carried out to investigate the number of fifty to sixty year old males who had suffered back problems. The survey found that in a sample of 221 fifty to sixty year old males, 124 had suffered back problems.

a Calculate the sample proportion, $\hat{p}$, of those surveyed who had suffered back problems.

b Estimate the standard deviation of the random variable $\hat{p}$ for such samples of size 221.

Solution

a In the sample of 221 fifty to sixty year old males, 124 had suffered back problems.

Thus $\hat{p} = \frac{124}{221} \approx 0.561.$

b The random variable $\hat{p}$ will have standard deviation $\approx \sqrt{\dfrac{\frac{124}{221}\left(1-\frac{124}{221}\right)}{221}} \approx 0.0334.$

Why is it useful to know how the sample proportions are distributed?

Knowing that sample proportions are normally distributed with mean p, the population proportion, and standard deviation $\sqrt{p(1-p)/n}$, where n is the sample size, allows us to apply our understanding of the normal distribution to sample proportions, as the following examples show.

EXAMPLE 3

If we classify as 'extremely unlikely' the likelihood that something that is normally distributed is more than three standard deviations from the mean explain why it is extremely unlikely that, when sampling 100 light bulbs from a batch that is thought to have 20% of the bulbs defective, we would find more than 35 of our sample defective.

Solution

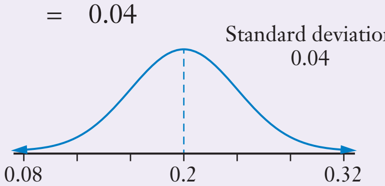
The sample proportions would be normally distributed with mean $= 0.2$

and standard deviation $= \sqrt{\dfrac{0.2(1-0.2)}{100}}$

$= 0.04$

For 35 defective globes in 100 we have a sample proportion of 0.35.

Now $\dfrac{0.35-0.2}{0.04} = 3.75.$



Thus a sample proportion of 0.35 is 3.75 standard deviations above the mean. To find more than 35 defective bulbs in a sample of 100 is therefore extremely unlikely.

If we did get more than 35 defective light bulbs in our sample of 100 we would question the statement that 20% were defective, or we would question whether our sample was truly representative of the batch.

ISBN 9780170395205

Shutterstock.com/Ambient Idea

Remember:

- The actual proportion (p) in a population, for example the proportion of people who are left handed, is fixed for a given population.
- The proportion who are left handed in a sample, $\hat{p}$, will vary according to the sample. Hence $\hat{p}$ forms a random variable.
- These sample proportions are normally distributed with mean p and standard deviation $\sqrt{\frac{p(1-p)}{n}}$ where n is the sample size.

EXAMPLE 4

For this question use the rule that for a normal distribution we would expect approximately 95% of the 'scores' to be within 2 standard deviations of the mean.

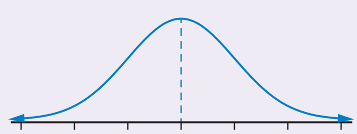
If we were to roll a normal six sided die 80 times we would expect to get an even number approximately 50% of the time. Between what two values, situated symmetrically either side of the 50% long term average, would we expect the proportion of even numbers to lie for approximately 95% of the time?

Solution

The sample proportions would be normally distributed with mean $= 0.5$

and standard deviation $= \sqrt{\frac{0.5(1-0.5)}{80}}$

≈ 0.0559



Standard deviation 0.0559

If $\hat{p}$ is within two standard deviations of the mean then

$$0.5 - 2 \times 0.0559 < \hat{p} < 0.5 + 2 \times 0.0559$$

i.e. $$0.3882 < \hat{p} < 0.6118$$

On approximately 95% of the occasions that we roll a normal six sided die 80 times we would expect the proportion of even numbers to be between 39% and 61%.

ISBN 9780170395205

The answer for the previous example stated that:

On approximately 95% of the occasions that we roll a normal six sided die 80 times we would expect the proportion of even numbers to be between 39% and 61%.

The 95% assumes that rolling the die 80 times is carried out numerous times. However even if we were to carry out the 80 rolls just 100 times we would still expect to find that the number of times the proportion of even numbers fell between 0.39 and 0.61 was pretty close to 95.

Simulating 80 rolls of a normal die, noting the proportion of times an even number was achieved and then repeating this simulation a further 99 times gave the following 100 proportions.

0.5125	0.5	0.6125	0.55	0.425	0.4125	0.5625	0.4	0.45	0.5125
0.45	0.55	0.5625	0.4625	0.5	0.5875	0.5375	0.525	0.525	0.575
0.4875	0.4	0.625	0.5375	0.55	0.4375	0.5375	0.5625	0.4625	0.4625
0.375	0.5	0.5	0.5375	0.55	0.525	0.5125	0.475	0.525	0.4
0.4375	0.5	0.475	0.4375	0.4	0.55	0.5	0.5375	0.5875	0.6
0.45	0.5	0.6	0.4875	0.5125	0.4625	0.5625	0.45	0.525	0.45
0.5375	0.4	0.4375	0.6625	0.5875	0.5125	0.525	0.4375	0.5625	0.4875
0.4875	0.425	0.5375	0.525	0.475	0.425	0.5375	0.5875	0.525	0.525
0.5375	0.4375	0.5375	0.475	0.525	0.5375	0.5125	0.4875	0.5	0.4625
0.5125	0.525	0.5125	0.5375	0.4875	0.5	0.525	0.475	0.5	0.475

Repeating this process again gave the following 100 proportions.

0.525	0.6	0.4125	0.5375	0.55	0.475	0.5875	0.575	0.575	0.625
0.5	0.475	0.525	0.5625	0.4875	0.5125	0.525	0.475	0.5625	0.5625
0.575	0.4875	0.5125	0.475	0.5375	0.6125	0.55	0.525	0.5	0.45
0.45	0.35	0.625	0.4375	0.475	0.6125	0.3375	0.525	0.55	0.45
0.4875	0.4875	0.4375	0.575	0.475	0.5375	0.525	0.5375	0.475	0.4625
0.5	0.5	0.4625	0.5375	0.4875	0.5875	0.475	0.45	0.5	0.4375
0.4625	0.45	0.525	0.45	0.5	0.475	0.4625	0.45	0.5375	0.6
0.5125	0.5	0.4625	0.5	0.5625	0.55	0.5125	0.5875	0.525	0.55
0.575	0.5625	0.4875	0.5	0.45	0.55	0.5875	0.4625	0.525	0.525
0.4375	0.4625	0.5125	0.475	0.5125	0.3625	0.6	0.5125	0.525	0.45

Are the above results consistent with the statement given earlier? i.e.:

On approximately 95% of the occasions that we roll a normal six sided die 80 times we would expect the proportion of even numbers to be between 39% and 61%.

ISBN 9780170395205

EXAMPLE 5

Let us suppose that 54% of a community are in favour of a particular development occurring.

A sample of 500 people from the community are to be surveyed to see if they are in favour of the development occurring.

With A and B symmetrically placed either side of the 54% population proportion copy and complete the following statement:

> *There is an 80% chance that in a sample of 500 people from this community, the sample proportion in favour of the development occurring will be between _A%_ and _B%_.*

Solution

The sample proportions would be normally distributed with mean $= 0.54$

and standard deviation $= \sqrt{\dfrac{0.54(1-0.54)}{500}}$

≈ 0.0223

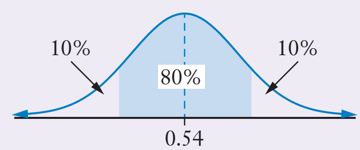
For $X \sim N(0.54, 0.0223^2)$

If $P(X < k) = 0.1$

$k = 0.5114$

If $P(X < k) = 0.9$

$k = 0.5686$



Standard deviation 0.0223

There is an 80% chance that in a sample of 500 people from this community, the sample proportion in favour of the development occurring will be between 51.1% and 56.9%.

```
invNorm (0.1, 0.54, 0.0223)
                    0.5114214
invNorm (0.9, 0.54, 0.0223)
                    0.5685786
```

Some calculators give both values 'in one go' as suggested below (and as you may have discovered when working through chapter four):

Tail setting	Centre ▼
prob	0.8
σ	0.0223
μ	0.54

x_1InvN	0.5114214
x_2InvN	0.5685786
prob	0.8
σ	0.0223
μ	0.54

ISBN 9780170395205

Exercise 6A

1 A survey of 123 randomly selected Australians found that 49 of them were aged over 50.

A second survey involving 2348 randomly selected Australians found that 761 of them were aged over 50.

Determine the sample proportion of people aged over 50 for each survey.

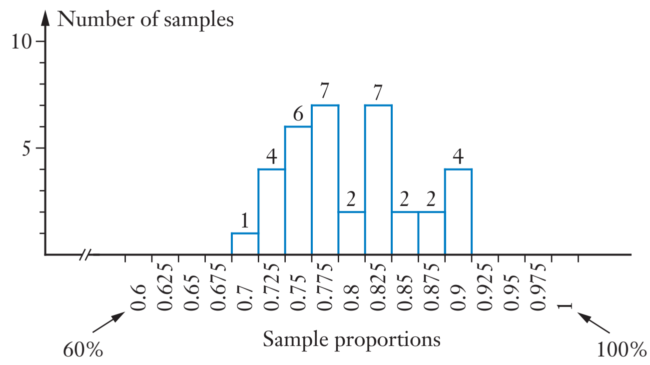
Which sample proportion is likely to be the better estimate of the proportion of Australians aged over 50?

2 During the course of one day ten samples of 'shoppers' were surveyed. Each sample involved 25 shoppers who had purchased at least one item from a particular supermarket, and for each sample the proportion of the 25 shoppers purchasing milk from the supermarket was noted. The proportions for the ten samples were as follows:

Sample	1	2	3	4	5	6	7	8	9	10
Proportion purchasing milk	0.72	0.68	0.88	0.56	0.60	0.76	0.64	0.84	0.72	0.80

For shoppers buying at least one item from this supermarket, estimate the population proportion of the shoppers buying milk from this supermarket and explain your answer.

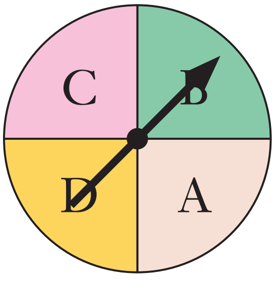
3 A number of sample proportions, for samples of the same size, gave rise to the graph shown below.



a How many samples were involved?

b Estimate p, the population proportion.

4 When the spinner on the right was spun 320 times, an A occurred on 72 occasions.



a What is the value of $\hat{p}$, the sample proportion of As for the 320 spin sample mentioned?

b What is the value of p, the population proportion of As?

c Calculate the mean and standard deviation of the random variable, $\hat{p}$, for samples involving 320 spins.

ISBN 9780170395205

5 Two normal fair six sided dice are rolled and the two numbers this process gives are added together.

This is repeated a further 99 times and the 100 trials produce 'a total less than 9' on 67 of the 100 occasions.

a What is the value of p, the population proportion of obtaining a total less than 9 when two normal fair dice are rolled?

b What is the value of $\hat{p}$, the sample proportion of a total less than 9 for the 100 trials mentioned?

c Calculate the mean and standard deviation of $\hat{p}$, the sample proportions for samples of 100 rolls of the two dice.

d How many standard deviations from p is our value for $\hat{p}$?

6 The census for a particular country showed that 84% of the households in that country had internet access.

At about the same time as the census data was collected, a sample of 240 households in a particular region of the country showed that 147 of the households had internet access.

a Determine p, the population proportion of households with internet access.

b Determine $\hat{p}$, the sample proportion of households with internet access.

c Comment on your results from **a** and **b**.

7 The abstract for a paper by Curtis Hardyck and Lewis Petrinovich, on left handedness and published by the American Psychological Association in 1977 states that *left-handedness, ranging from moderate through strongly left-handed, is found in approximately 10% of the population*.

Comment on each of the following:

a A survey of 1000 school students classified 112 of the students as being left handed.

b According to an article by L. J. Harris in the Psychology Press publication 'Laterality':

In a survey of participants in the 1981 World Fencing Championship, 35% of the athletes in the foil competition were left-handed.

8 A survey was carried out to investigate the gender of teachers in Australian schools. The survey found that in a sample of 1247 teachers, 461 were male.

a Calculate the sample proportion, $\hat{p}$, of those teachers surveyed who were male.

b Estimate the standard deviation of the random variable $\hat{p}$, for such samples of size 1247.

iStock.com/ monkeybusinessimages

9 A survey involving 248 midwives found that 143 of them were aged between 46 and 60.

a Calculate the sample proportion, $\hat{p}$, of those midwives surveyed who were aged between 46 and 60.

b Estimate the standard deviation of the random variable $\hat{p}$ for such samples of size 248.

ISBN 9780170395205

10 Twenty samples, not all involving the same number of people, involved males of 18 years and over and considered the proportion of males in the sample who had high blood pressure, according to a particular classification of high blood pressure. The proportion having high blood pressure in each sample, and the number of people in the sample are given below.

Sample	Number in sample	Sample proportion having high blood pressure
1	8	0.5
2	10	0.1
3	50	0.28
4	25	0.24
5	10	0.2
6	80	0.2375
7	56	0.286
8	10	0.1
9	50	0.2
10	180	0.261
11	20	0.35
12	10	0.1
13	8	0.375
14	25	0.2
15	8	0.5
16	150	0.2
17	20	0.3
18	25	0.32
19	10	0.8
20	1	1

a Why would it be unwise to estimate the population proportion of males aged 18 years and over who have high blood pressure, according to the classification, simply by finding the mean of the above sample proportions?

b Explain a better way to use the above figures to determine an estimate for the population proportion of males aged 18 years and over who have high blood pressure, according to the classification, and determine that estimate.

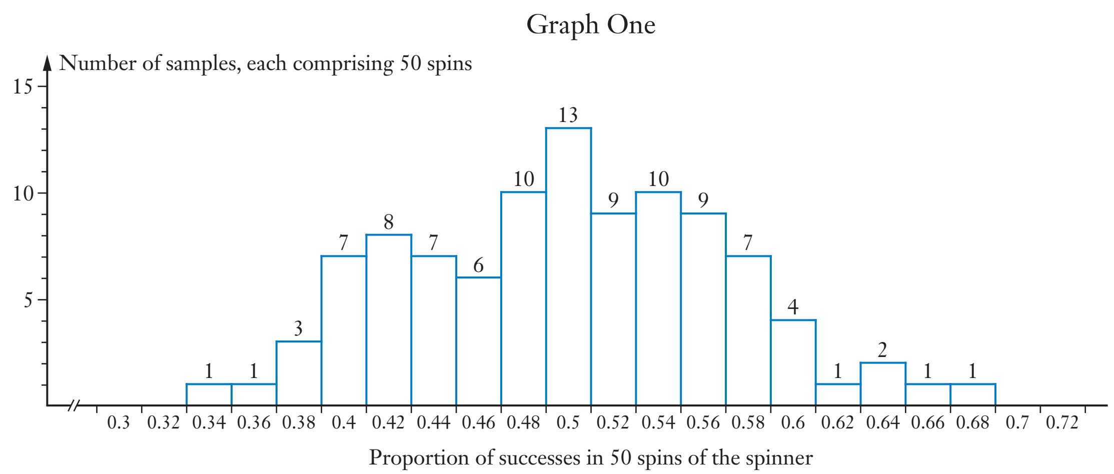
11 If we classify as 'extremely unlikely' the likelihood that something that is normally distributed is more than three standard deviations from the mean, explain why it is extremely unlikely that, when sampling 200 seeds from a batch that is thought to have 10% of the seeds unable to germinate, we would find 35 or more of our sample unable to germinate.

ISBN 9780170395205

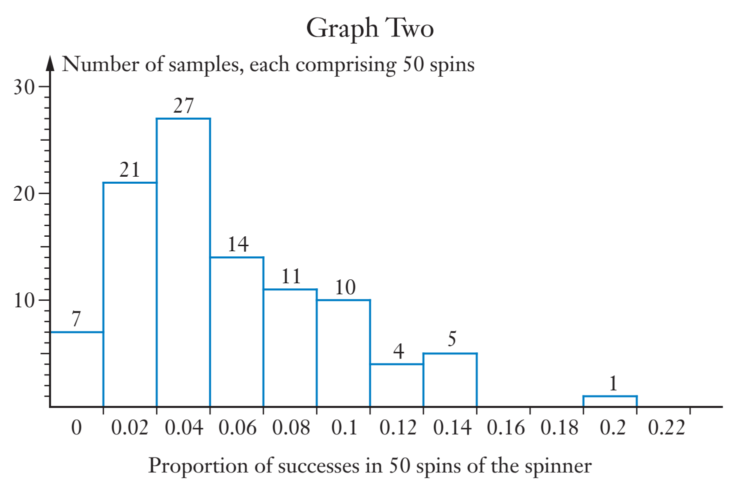
12 A random process for which P(success) = 0.5 was carried out 50 times and the proportion of successes recorded.

This same sampling process was carried out a further 99 times to give 100 sample proportions, each with sample size of 50.

The 100 sample proportions are shown as Graph One below.



Graph Two below also shows 100 sample proportions of successes for samples each of size 50 but this time for a random process for which P(success) = 0.05.



Graph One seems to have much more of the characteristic Normal Distribution 'bell shape' about it than Graph Two.

Why should this not really be a surprise?

13 For this question use the rule that for a normal distribution we would expect approximately 68% of the 'scores' to be within 1 standard deviation of the mean.

If we were to roll a normal six sided die 100 times we would expect to get an even number approximately 50% of the time. Between what two values, situated symmetrically either side of the 50% long term average, would we expect the proportion of even numbers to lie for approximately 68% of samples of size 100?

14 A politician claimed that he expected to win 52% of the votes in the forthcoming election for the seat of Dasha.

Wishing to check this, a newspaper carried out a survey and found that out of 200 people who intended to vote in the election for the seat of Dasha, 81 said they would be voting for the politician.

Comment.

15 Let us suppose that 24% of cars produced by the XYZ car manufacturing company are blue.

A random sample of 800 cars produced by this company are surveyed and the proportion of blue cars in the sample noted.

With A and B symmetrically placed either side of the 24% population proportion copy and complete the following statement:

There is a 90% chance that in a sample of 800 cars produced by this company the sample proportion that are blue will be between _A%_ *and* _B%_.

iStock.com/rawpixel

Confidence intervals

In statistics, samples are taken in an attempt to estimate information regarding the population of which the sample forms a part. Thus if we took a sample from a population and found that 24% of our sample possessed some characteristic, be it left handedness, being over a particular age, supporting a particular political party, or whatever, we could estimate the proportion of the population possessing this characteristic as being 24%. This is a **point estimate** of the population proportion because it gives a 'one value estimate', 0.24. However, we know that the sample proportions over many samples are normally distributed with mean p and standard deviation $\sqrt{p(1-p)/n}$, where p is the population proportion and n is the sample size.

This allows us, with a particular level of confidence, to give a range of values, or *interval*, that we can expect the population proportion to lie within. This is called an **interval estimate**.

First let us revisit the **standard normal distribution**, i.e. $Z \sim N(0, 1^2)$, with its associated 'z scores' and establish some important numbers (sometimes called **critical scores**) relating to what we will call the

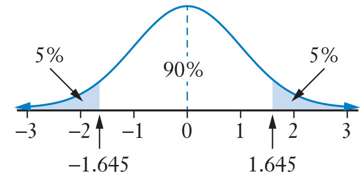
90%, 95% and 99% **confidence intervals**.

 ISBN 9780170395205

For a *90% confidence interval*

Solving	$P(Z < k)$	=	0.95
gives	k	=	1.645

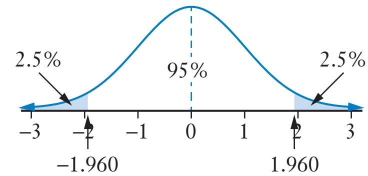
90% of the scores from a normal distribution lie within 1.645 standard deviations of the mean.



For a *95% confidence interval*

Solving	$P(Z < k)$	=	0.975
gives	k	=	1.960

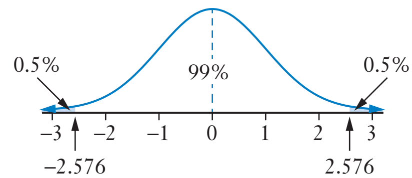
95% of the scores from a normal distribution lie within 1.960 standard deviations of the mean.



For a *99% confidence interval*

Solving	$P(Z < k)$	=	0.995
gives	k	=	2.576

99% of the scores from a normal distribution lie within 2.576 standard deviations of the mean.



When we take one sample of a sufficiently large size, we know that the proportion of our single sample comes from a distribution of sample proportions that approximate to a normal distribution with a mean equal to the population proportion. Hence we can be 90% confident that our sample proportion is within 1.645 standard deviations of the population proportion, (1.645 being the critical score for the 90% confidence interval). Now if A is within 1.645 units of some fixed value B then B is within 1.645 units of A.

Therefore:

We can be 90% confident that the population proportion is within 1.645 standard deviations of the sample proportion.

Similarly:

We can be 95% confident that the population proportion is within 1.960 standard deviations of the sample proportion.

And:

We can be 99% confident that the population proportion is within 2.576 standard deviations of the sample proportion.

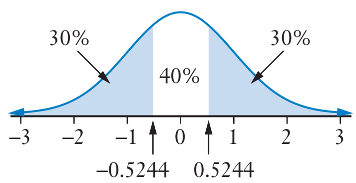
Thus, to infer **an interval estimate for the population proportion**, p, from a sample proportion, $\hat{p}$:

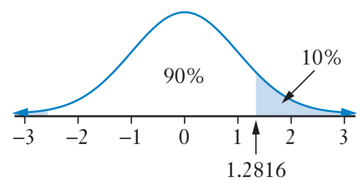
- Assume that the sample proportions are normally distributed with mean p and standard deviation $\sqrt{\hat{p}(1-\hat{p})/n}$. (This standard deviation is the *best we can do*).
- The interval estimate for p is then $\hat{p} - k\sqrt{\dfrac{\hat{p}(1-\hat{p})}{n}} \leq p \leq \hat{p} + k\sqrt{\dfrac{\hat{p}(1-\hat{p})}{n}}$

  where k is the appropriate number of standard deviations for the required confidence interval.

i.e., to 3 dp,	k	=	1.645 for a 90% confidence interval
	k	=	1.960 for a 95% confidence interval
and	k	=	2.576 for a 99% confidence interval.

- **If we were to construct many 95% (or 90% or 99%) confidence intervals we would expect approximately 95% (or 90% or 99%) of them to contain the population proportion.**
- Whilst we have concentrated here on the 90%, 95% and 99% confidence intervals we could have intervals involving other percentages, say 40% as shown below left, or intervals not centred on the mean value, as shown below right. However we will tend to concentrate on the 90%, 95% and 99% intervals centred on the mean.





EXAMPLE 6

A survey of 1000 people found that 143 said they were satisfied with the way the government was running the country. Within what range of values (centred on the sample proportion) can we be 95% confident that the population proportion lies?

Solution

Using a standard deviation of 0.011, ($\sqrt{0.143(1-0.143)/1000} \approx 0.011$), the sample proportion of 0.143, and the critical value for the 95% confidence interval of 1.960 we have:

$$\begin{aligned} 0.143 - 1.960 \times 0.011 &= 0.121 \\ 0.143 + 1.960 \times 0.011 &= 0.165 \end{aligned}$$

We can be 95% confident that the population proportion lies between 0.121 and 0.165.

Some calculators, given the sample proportion and sample size, can give confidence intervals for population proportions, as the displays shown below suggest.

zInterval_1Prop 143,1000,0.95

"Title"	"1-Prop z Interval"
"CLower"	0.121303
"CUpper"	0.164697
"$\hat{p}$"	0.143
"ME"	0.021697
"n"	1000

C-Level	.95
x	143
n	1000

→

Lower	0.1213027
Upper	0.1646973
$\hat{p}$	0.143
n	1000

ISBN 9780170395205

Note: The 95% confidence interval means that for such samples, 95% of such confidence intervals would contain the population proportion. We tend to express this by saying that we are 95% confident that our interval contains the population proportion, as in the previous example. However, if asked to interpret or explain the confidence interval it is perhaps best to include the fuller interpretation, as in the next example.

EXAMPLE 7

A survey of 500 drivers asks each person if they think the current penalties for using a mobile phone when driving are too harsh.

184 of the 500 say they do think the penalties are too harsh.

Find the 90% confidence interval for the population proportion and interpret your answer.

Solution

Using a standard deviation of 0.0216, ($\sqrt{0.368(1-0.368)/500} \approx 0.0216$), the sample proportion of 0.368, and the critical value for the 90% confidence interval of 1.645:

$$0.368 - 1.645 \times 0.0216 = 0.3325$$
$$0.368 + 1.645 \times 0.0216 = 0.4035$$

The 90% confidence interval for the population proportion is 0.3325 to 0.4035.

Interpretation: We could expect 90% of the 90% confidence intervals constructed in this way to contain the population proportion. Thus, with 90% confidence we estimate that between 33.25% and 40.35% of all drivers think that the current penalties for using a mobile phone when driving are too harsh.

The reader should confirm that these same values can be obtained from a calculator capable of giving confidence intervals.

Margin of error

In the previous example we could have expressed the 90% confidence interval as:

$$0.368 \pm 0.0355.$$

Using the minus gives us the lower bound: $0.368 - 0.0355 = 0.3325$
Using the plus gives us the upper bound: $0.368 + 0.0355 = 0.4035$

The value 0.0355 is the **margin of error**. It is the maximum amount that p can differ from $\hat{p}$ whilst still being in the confidence interval.

ISBN 9780170395205

With the confidence interval given by:

$$\hat{p} - k\sqrt{\frac{\hat{p}(1-\hat{p})}{n}} \quad \le \quad p \quad \le \quad \hat{p} + k\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$$

with a suitably chosen value of k, it follows that the margin of error is given by

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

i.e.: The appropriate critical value from the standard normal distribution multiplied by the standard deviation of the sample proportions.

Sample size

Suppose that, in the previous example, instead of having the 90% confidence interval as

$$0.368 \quad \pm \quad 0.0355$$

we wanted it to be $\quad 0.368 \quad \pm \quad 0.025.$

I.e., we wanted the margin of error to be 0.025.

Then $\quad k\sqrt{\frac{\hat{p}(1-\hat{p})}{n}} \quad = \quad 0.025$

For the 90% confidence interval we have $k = 1.645$, and the example had $\hat{p} = 0.368$.

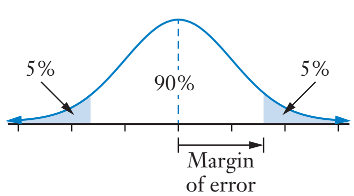
Thus $\quad 1.645 \times \sqrt{\frac{0.368(1-0.368)}{n}} \quad = \quad 0.025$

Solving gives $\quad n \quad = \quad 1006.97$

Remembering that n must be an integer, $\quad n \quad = \quad 1007$

The sample size should be 1007.

Check:



$$\text{Margin of error} \quad = \quad 1.645 \times \sqrt{\frac{0.368(1-0.368)}{1007}}$$

$$\approx \quad 0.025 \text{ as required.}$$

In this way, if we are planning a survey and already have some idea of what level of confidence we want, and to within what margin of error, we can choose the sample size appropriately. This would also require us to have some idea of the likely sample proportion but, to get started we could choose

$$\hat{p} \quad = \quad 0.5$$

as this will give the maximum value for $\hat{p}(1-\hat{p})$. This will make sure that whatever the value of $\hat{p}$, n will be such that the margin of error is at or within the desired level.

[Can you prove that $x(1-x)$ is maximised for $x = 0.5$?]

ISBN 9780170395205

EXAMPLE 8

A survey is to be carried out with the aim of having the 95% confidence interval on the population proportion with margin of error just 0.022. Taking the sample proportion as 0.5, find the sample size.

Solution

For the 95% confidence interval we have $k = 1.960$, and we are told that $\hat{p} = 0.5$.

Thus $$1.960\sqrt{\frac{0.5(1-0.5)}{n}} = 0.022.$$

Solving gives $n \approx 1984$,

The sample size needs to be 1984.

Sample size

```
zInterval_1Prop 992,1984,0.95

["Title"     "1-Prop z Interval"
 "CLower"    0.477999
 "CUpper"    0.522001
 "p̂"         0.5
 "ME"        0.022001
 "n"         1984              ]
```

Margin of error approx 0.022

- If it is important that our margin of error does not exceed 0.022 we would round n to the next integer **up**, i.e. 1985.

Increasing the level of confidence increases the margin of error

Suppose a survey of 500 people in Australia finds that 122 reply that they have lost confidence with the democratic system.

Using the 90% confidence interval we could say that:

We are 90% confident that the proportion of people in the whole Australian population who would say that they have lost confidence with the democratic system would lie in the interval 0.2124 to 0.2756. (Check that you can obtain these figures.)

I.e., with 90% confidence our estimate for the population proportion is 0.244 ± 0.0316.

If we want to increase our level of confidence, to say 95% or even 99%, still using the 122 out of 500 from our sample, we have to accept the increase in the margin of error. In other words, by increasing the size of our interval we can be more confident that it includes the population proportion.

Check that you agree with the following statements for this situation and notice how the margin of error increases as the level of confidence increases (for fixed n and $\hat{p}$).

The 95% confidence interval is 0.2064 to 0.2816.
I.e., with 95% confidence our estimate for the population proportion is 0.244 ± 0.0376.

The 99% confidence interval is 0.1945 to 0.2935.
I.e., with 99% confidence our estimate for the population proportion is 0.244 ± 0.0495.

ISBN 9780170395205

Let's check

```
zInterval_1Prop 502,1000,0.95
 "Title"     "1-Prop z Interval"
 "CLower"        0.47101
 "CUpper"        0.53299
 "p̂"              0.502
 "ME"            0.03099
 "n"              1000
```

Suppose we toss a fair coin 1000 times and obtain a head on 502 occasions, i.e. the proportion of heads is 0.502, or 50.2%. Using this proportion, the 95% confidence interval would be 0.471 to 0.533, i.e. 0.502 ± 0.031.

If we took many such samples of 1000 tosses of the coin we would expect approximately 95% of the *95% confidence intervals* so produced to contain the population proportion.

```
sum(randInt(0,1,1000))
        1000
                              0.502
```

Of course, in most situations, we would probably be carrying out the sampling to estimate the population proportion. However, in this fair coin situation we have the luxury of knowing the long term proportion so let us simulate the situation a number of times and see if we do indeed find that most, but not all of the 95% confidence intervals so produced do indeed contain 0.5, the population proportion.

Thirty such samples each of size 1000 are shown below, and on the next two pages. As can be seen, at least for these 30 samples, it is indeed the case that most, but not all of the 95% confidence intervals do contain 0.5, the population proportion, p.

Sample	Sample proportion	95% confidence interval	Interval contains 0.5
1	0.502	0.471 to 0.533	✓
2	0.499	0.468 to 0.530	✓
3	0.506	0.475 to 0.537	✓
4	0.493	0.462 to 0.524	✓
5	0.497	0.466 to 0.528	✓
6	0.498	0.467 to 0.529	✓
7	0.526	0.495 to 0.557	✓
8	0.492	0.461 to 0.523	✓
9	0.490	0.459 to 0.521	✓

ISBN 9780170395205

10	0.492	0.461 to 0.523	✓
11	0.490	0.459 to 0.521	✓
12	0.506	0.475 to 0.537	✓
13	0.486	0.455 to 0.517	✓
14	0.490	0.459 to 0.521	✓
15	0.497	0.466 to 0.528	✓
16	0.504	0.473 to 0.535	✓
17	0.510	0.479 to 0.541	✓
18	0.517	0.486 to 0.548	✓
19	0.455	0.424 to 0.486	✗
20	0.475	0.444 to 0.506	✓
21	0.506	0.475 to 0.537	✓
22	0.511	0.480 to 0.542	✓
23	0.490	0.459 to 0.521	✓
24	0.497	0.466 to 0.528	✓
25	0.499	0.468 to 0.530	✓
26	0.532	0.501 to 0.563	✗

ISBN 9780170395205

27	0.493	0.462 to 0.524 (0.400 0.450 0.500 0.550 0.600)	✓
28	0.480	0.449 to 0.511 (0.400 0.450 0.500 0.550 0.600)	✓
29	0.498	0.467 to 0.529 (0.400 0.450 0.500 0.550 0.600)	✓
30	0.513	0.482 to 0.544 (0.400 0.450 0.500 0.550 0.600)	✓

Carry out some similar simulations yourself, perhaps for coin tossing or die rolling and for 95%, 90% or even 80% confidence intervals, and comment on your findings.

Exercise 6B

(Note: Survey results referred to in this exercise are for the purpose of this exercise only and do not necessarily reflect any real situation.)

1 A survey of 1200 people living in Australia found that 450 were in favour of the idea of introducing compulsory national service. Find the 95% confidence interval for the population proportion and interpret your answer.

2 A survey involving 800 people living in Australia, and who had recently contacted their bank for online help, found that 680 of the 800 were either satisfied or very satisfied with the service they received. Find the 90% confidence interval for the population proportion and interpret your answer.

3 A survey of 250 people who all regularly played a particular sport found that 190 of the 250 agreed that the recent rule changes were a good idea. Find the 99% confidence interval for the population proportion and interpret your answer.

4 With a sample size of 880 and a sample proportion of 70% state the 95% confidence interval for the population proportion in the form $a\% \pm b\%$, with b given correct to the nearest integer.

5 A national opinion poll surveyed 2000 people and found that 45% wanted to see changes to the current daylight saving rules.

Find the 90% confidence interval for the population proportion in the form $a\%$ to $b\%$, with a and b each given correct one decimal place. Interpret your answer.

When all 200 people in a particular community were surveyed it was found that 140 of the 200 wanted to see changes to the daylight saving rules. Comment.

6 The testing of 1000 seeds from a particular batch of millions of such seeds found that 28% failed to germinate. Use this information to write a statement about the batch involving the idea of being 95% confident.

7 For a sample size of 2000 and sample proportion 0.45, find the margin of error at the 95% confidence level.

8 For a sample size of n and sample proportion b, which out of the 90% confidence interval and the 99% confidence interval has the larger margin of error?

ISBN 9780170395205

9 A sample of 200 machine components made by a particular machine finds that 36 of the components are deemed to be of an unacceptable standard. (A component can be judged to be either acceptable or unacceptable.)

- **a** What is the proportion of acceptable components in this sample?
- **b** Write a statement about the percentage of acceptable components in the population of components made by this machine including in your statement the idea of being 90% confident.
- **c** Write a statement about the percentage of acceptable components in the population of components made by this machine including in your statement the idea of being 99% confident.

10 A survey is to be carried out with the aim of having the 95% confidence interval on the population proportion with margin of error just 0.065. Taking the sample proportion as 0.5 find the sample size.

11 A survey is to be carried out with the aim of having the 90% confidence interval on the population proportion with margin of error just 0.03. Taking the sample proportion as 0.5 find the sample size.

12 A survey is to be carried out with the aim of having the 95% confidence interval for the population proportion equal to, or within, 0.60 to 0.70. Taking the sample proportion as 0.65 find the sample size.

13 Let us suppose that in a survey of 500 Australian males aged between 20 and 30 it was found that 76% of the males were taller than their father.

According to the results of this survey copy and complete the following statement (give percentages to the nearest whole percent):

We can be 95% confident that of all Australian males between the ages of 20 and 30,
between __% and __% are taller than their father.

If we wanted to be 99% confident would our interval be larger or smaller than the 95% interval? Explain why.

14 A government inquiry wanted to estimate the proportion of Australians who possessed a particular attribute. The results from a random sample of Australians led to the calculation of the 90% confidence interval for the population proportion who have the particular attribute as 0.241 ± 0.060.

Copy and complete the following statement, giving percentages to the nearest whole percent.

We can be 90% confident that the proportion of Australians having
the particular attribute lies between __% and __%.

Determine the sample size and the number in the sample possessing the attribute.

15 Let us suppose that a random sample of 480 year twelve Australian school students were surveyed and 168 of the students said that they intended proceeding to University the following year.

- **a** Calculate the sample proportion of these year 12s intending to proceed to University the following year.
- **b** We would expect all such sample proportions for samples of this size to approximate to a normal distribution. Calculate the standard deviation of this normal distribution (rounded correct to four decimal places).
- **c** Calculate the 95% confidence interval for the population proportion and interpret your answer.
- **d** A second random sample is planned but this time the organisers would like the 95% confidence interval to involve a margin of error of, at most, 3%.
  Calculate the sample size.

Miscellaneous exercise six

This miscellaneous exercise may include questions involving the work of this chapter, the work of any previous chapters, and the ideas mentioned in the Preliminary work section at the beginning of the book.

Answer questions **1**, **2**, **3** and **4** *without* the assistance of your calculator.

1 Determine the value of x in each of the following:

a $\log_x 64 = 6$

b $\log_2 x = 3$

c $\log x = 2$

d $\ln e = x$

2 Differentiate

a $x\sqrt{x}$

b $4x^5 + \log_e x$

c $7\ln x$

d $\ln(5x^3 - 6x)$

3 Find $\frac{dy}{dx}$ for each of the following

a $y = \log_e(5x - 1)$

b $y = \ln(x^4 + 1)$

c $y = \ln[(x + 1)(x - 1)]$

4 Determine the equation of the tangent to $y = 3 - \ln x$ at the point $(e, 2)$.

5 Two fair coins are flipped 160 times and come down with both heads facing upwards on 46 of the 160 occasions.

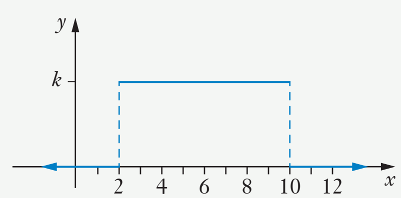
a What is the value of p, the population proportion of obtaining two heads?

b What is the value of $\hat{p}$, the sample proportion of two heads for the 160 flips?

c Calculate the mean and standard deviation of the random variable $\hat{p}$, for samples involving 160 flips of two coins.

d How many standard deviations from p is our value for $\hat{p}$?

6 The continuous random variable X has the probability density function shown on the right.



Determine

a k

b $P(X \geq 4)$

c $P(X < 8)$

d $P(X > 3 \mid X < 7)$

7 A parent suggests that when four year olds are given a new fibre-tip pen the number of minutes until they manage to lose the cap of the pen forms a random variable with probability density function:

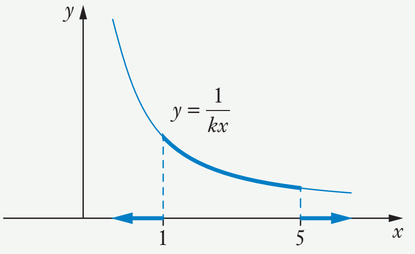
$$f(x) = \begin{cases} 0.5e^{-0.5x} & \text{for } x > 0 \\ 0 & \text{elsewhere.} \end{cases}$$

According to this suggested rule, find the probability of a four year old losing the cap within 5 minutes of being given the pen.

ISBN 9780170395205

8 If $f'(x) = \dfrac{x+5}{x}$ and $f(1) = 5$ determine expressions for $f''(x)$ and $f(x)$.

9 A continuous random variable, X, has the probability density function



$$f(x) = \begin{cases} \dfrac{1}{kx} & \text{for } 1 \le x \le 5 \\ 0 & \text{for all other values of } x. \end{cases}$$

Determine each of the following, giving your answers as exact values.

a The value of k,

b $P(2 \le X \le 4)$.

10 $X \sim N(24, 64)$, i.e. X, is normally distributed with a mean of 24 and a standard deviation of 8. Given that $P(X > k) = 0.2266$ determine k.

11 The response times on a psychological test were found to be approximately normally distributed with mean 2.32 seconds and standard deviation 0.48 seconds.

a Using this normal distribution as the basis for prediction, what is the probability that a randomly chosen individual would achieve a time of less that 1 second in this test?

b Given that a randomly chosen individual achieved a time that was less than the mean, what is the probability that they achieved a time that is less than 1 second.

12 A random variable X is normally distributed with mean μ and standard deviation σ. If $P(X < 25)$ is 0.1082 and $P(X > 42)$ is 0.1303 determine

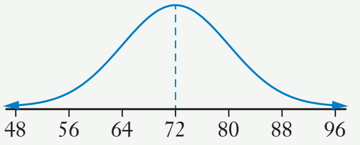
a $P(25 < X < 42)$, giving your answer correct to two decimal places,

b the values of μ and σ, giving answers correct to 1 decimal place,

c $P(X > 22)$, giving your answer correct to two decimal places.

13 A coin is tossed 1000 times and a head is obtained on 521 occasions. Does this suggest the coin is unfair? Comment.

14 In an examination for which the marks are approximately normally distributed with a mean score of 72 and a standard deviation of 8, 85% of the candidates pass. What is the pass mark? (Round your answer to the nearest 0.5%.)



15 If we roll five normal fair six sided dice and add the numbers on the uppermost faces our answer could be from a low of 5 to a high of 30.

Use a calculator or computer to simulate approximately 100 such rolls of 5 dice and record your answers grouped as follows:

Score	≤ 9	$10 \rightarrow 13$	$14 \rightarrow 17$	$18 \rightarrow 21$	$22 \rightarrow 25$	≥ 26
Frequency						

The probability distribution of possible total scores can be modelled using a normal distribution with mean 17.5 and standard deviation 3.8.

Compare your distribution with that given theoretically by the suggested normal distribution model, i.e. $N(17.5, 3.8^2)$, with the necessary adjustment for continuity being made. (See page 96 for an explanation of adjusting for continuity.)

16 Let us suppose that a survey of mobile phone use amongst Australian adults found that the 95% confidence interval for the proportion of adult Australians who were 'mobile only', i.e. they did not have land line phone access in their place of residence but relied totally on their mobile phone, was 0.19 ± 0.02.

Explain what this 0.19 ± 0.02 confidence interval means.

In view of the above survey comment on the following:

- A survey of 500 Australians aged 65 or over found that 20 were 'mobile only'.
- A survey of 800 Australians aged in their twenties found that 336 were 'mobile only'.

17 At an election one of the minor parties polled 22% of the vote. Six months later a survey of 400 people in the electorate showed that just 20% supported the party. Could we conclude that the party has lost popularity or could the change be explained by expected variation in sample proportions? Explain.

Suppose the 20% figure had instead come from a survey involving 4000 people.

18 A company makes a batch of batteries and rejects them on the basis that tests show that about 15% of the batteries don't work.

A second company buys the batteries and sells them cheaply with the statements:

> Batteries for sale at half price:
> 15% of these batteries don't work but at half price that
> means the 85% you are getting that do work are cheap.

Joe complains to a consumer protection organisation claiming that he bought 100 of the batteries and found that 17 were faulty.

Comment on the above.

ISBN 9780170395205