14

Statistical inference

14.1 Kick off with CAS

Sample distributions

Data gathered from the last census showed that 40% of all households owned a dog. It would be nearly impossible to replicate this data, as to do this would mean surveying millions of people. However, samples of the population can be surveyed. The percentage of dog owners will vary from sample to sample, but the percentages from each set of samples can be graphed to reveal any patterns.

Whether or not a household owns a dog is a binomial random variable, so the percentages generated will be based on the binomial distribution. The proportion of successes in the population is 40%.

1 Generate data for 100 random samples of 5 people who were asked whether their household had a dog, and graph the data using a dot plot or histogram.

2 Generate data for 100 random samples of 50 people and graph the data using a dot plot or histogram.

3 Generate the data for 100 random samples of 200 people and graph the data using a dot plot or histogram.

4 a Do the data from each set of samples give a clear result as to the percentage of households that own a dog?

b Which set of samples gives more reliable information? Why?

c As the sample size increases, what happens to the shape of the graph?

d Is the graph symmetrical, and if so, about what value is it symmetrical?

Please refer to the Resources tab in the Prelims section of your **eBookPLUS** for a comprehensive step-by-step guide on how to use your CAS technology.

14.2 Population parameters and sample statistics

studyon

Units 3 & 4
AOS 4
Topic 5
Concept 1

Populations and samples
Concept summary
Practice questions

Suppose you were interested in the percentage of Year 12 graduates who plan to study Mathematics once they complete school. It is probably not practical to question every student. There must be a way that we can ask a smaller group and then use this information to make generalisations about the whole group.

Samples and populations

A **population** is a group that you want to know something about, and a **sample** is the group within the population that you collect the information from. Normally, a sample is smaller than the population; the exception is a census, where the whole population is the sample.

The number of members in a sample is called the **sample size** (symbol n), and the number of members of a population is called the **population size** (symbol N). Sometimes the population size is unknown.

WORKED EXAMPLE 1

Cameron has uploaded a popular YouTube video. He thinks that the 133 people in his year group at school have seen it, and he wants to know what they think. He decides to question 10 people. Identify the population and sample size.

THINK	WRITE
1 Cameron wants to know what the people in his year at school think. This is the population.	$N = 133$
2 He asks 10 people. This is the sample.	$n = 10$

WORKED EXAMPLE 2

A total of 137 people volunteer to take part in a medical trial. Of these, 57 are identified as suitable candidates and are given the medication. Identify the population and sample size.

THINK	WRITE
1 57 people are given the medication. This is the sample size.	$n = 57$
2 We are interested in the group of people who might receive the drug in the future. This is the population.	The population is unknown, as we don't know how many people may be given this drug in the future.

Statistics and parameters

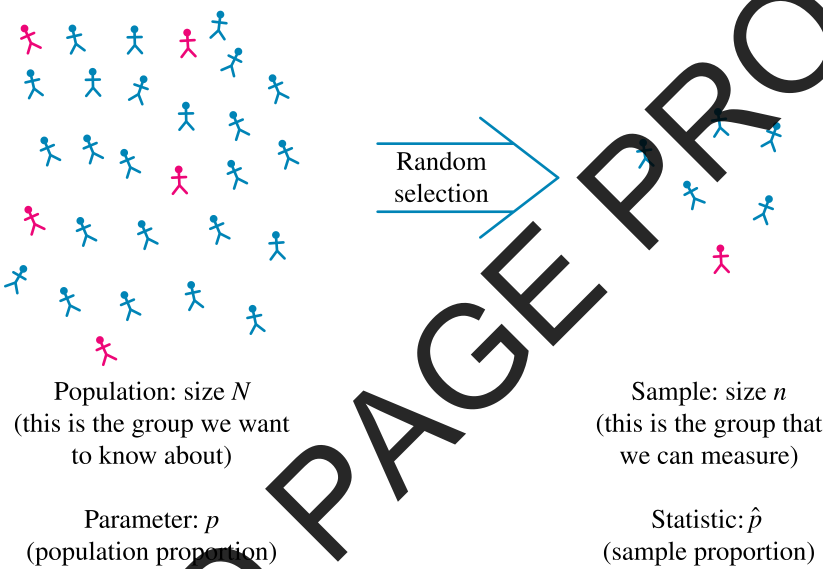
eBook*plus*

Interactivity
Population parameters and sample statistics
int-6442

A **parameter** is a characteristic of a population, whereas a **statistic** is a characteristic of a sample. This means that a statistic is always known exactly (because it is measured from the sample that has been selected). A parameter is usually estimated from a sample statistic. (The exception is if the sample is a census, in which case the parameter is known exactly.)

In this unit, we will study binomial data (that means that each data point is either yes/no or success/failure) with special regard to the proportion of successes.

The relationship between populations and samples



WORKED EXAMPLE 3

Identify the following as either sample statistics or population parameters.

a **Forty-three per cent of voters polled say that they are in favour of banning fast food.**

b **According to Australian Bureau of Statistics census data, the average family has 1.7 children.**

c **Between 18% and 23% of Australians skip breakfast regularly.**

d **Nine out of 10 kids prefer cereal for breakfast.**

THINK	WRITE
a 43% is an exact value that summarises the sample asked.	**a** Sample statistic
b The information comes from census data. The census questions the entire population.	**b** Population parameter
c 18%–23% is an estimate about the population.	**c** Population parameter
d Nine out of 10 is an exact value. It is unlikely that all kids could have been asked; therefore, it is from a sample.	**d** Sample statistic

Random samples

studyon

Units 3 & 4
AOS 4
Topic 5
Concept 2

Random sampling methods
Concept summary
Practice questions

eBook*plus*

Interactivity
Random samples
int-6443

A good sample should be representative of the population. If we consider our initial interest in the proportion of Year 12 graduates who intend to study Mathematics once they finished school, we could use a Mathematical Methods class as a sample. This would not be a good sample because it does not represent the population — it is a very specific group.

In a **random sample**, every member of the population has the same probability of being selected. The Mathematical Methods class is not a random sample because students who don't study Mathematical Methods have no chance of being selected; furthermore, students who don't attend that particular school have no chance of being selected.

A **systematic sample** is almost as good as a random sample. In a systematic sample, every kth member of the population is sampled. For example, if $k = 20$, a customs official might choose to sample every 20th person who passes through the arrivals gate. The reason that this is almost as good as random sample is that there is an assumption that the group passing the checkpoint during the time the sample is taken is representative of the population. This assumption may not always be true; for example, people flying for business may be more likely to arrive on an early morning flight. Depending on the information you are collecting, this may influence the quality of the data.

In a **stratified random sample**, care is taken so that subgroups within a population are represented in a similar proportion in the sample. For example, if you were collecting information about students in Years 9–12 in your school, the proportions of students in each year group should be the same in the sample and the population. Within each subgroup, each member has the same chance of being selected.

A **self-selected sample**, that is one where the participants choose to participate in the survey, is almost never representative of the population. This means, for example, that television phone polls, where people phone in to answer yes or no to a question, do not accurately reflect the opinion of the population.

WORKED EXAMPLE 4

A survey is to be conducted in a middle school that has the distribution detailed in the table below. It is believed that students in different year levels may respond differently, so the sample chosen should reflect the subgroups in the population (that is, it should be a stratified random sample). If a sample of 100 students is required, determine how many from each year group should be selected.

Year level	Number of students
7	174
8	123
9	147

THINK	WRITE
1 Find the total population size.	Total population $= 174 + 123 + 147 = 444$
2 Find the number of Year 7s to be surveyed.	Number of Year 7s $= \frac{174}{444} \times 100$ $= 39.1$ Survey 39 Year 7s.
3 Find the number of Year 8s to be surveyed.	Number of Year 8s $= \frac{123}{444} \times 100$ $= 27.7$ Survey 28 Year 8s.

4 Find the number of Year 9s to be surveyed.

$$\text{Number of Year 9s} = \frac{147}{444} \times 100$$
$$= 33.1$$

Survey 33 Year 9s.

5 There has been some rounding, so check that the overall sample size is still 100.

Sample size = 39 + 28 + 33 = 100

The sample should consist of 39 Year 7s, 28 Year 8s and 33 Year 9s.

Using technology to select a sample

If you know the population size, it should also be possible to produce a list of population members. Assign each population member a number (from 1 to N). Use the random number generator on your calculator to generate a random number between 1 and N. The population member who was allocated that number becomes the first member of the sample. Continue generating random numbers until the required number of members has been picked for the sample. If the same random number is generated more than once, ignore it and continue selecting members until the required number has been chosen.

EXERCISE 14.2 Population parameters and sample statistics

PRACTISE

Work without CAS

1 WE1 On average, Mr Parker teaches 120 students per day. He asks one class of 30 about the amount of homework they have that night. Identify the population and sample size.

2 Bruce is able to hem 100 shirts per day. Each day he checks 5 to make sure that they are suitable. Identify the population and sample size.

3 WE2 Ms Lane plans to begin her Statistics class each year by telling her students a joke. She tests her joke on this year's class (15 students). She plans to retire in 23 years time. Identify the population and sample size.

4 Lee-Yin is trying to perfect a recipe for cake pops. She tries 5 different versions before she settles on her favourite. She takes some samples to school and asks 9 friends what they think. Identify the population and sample size.

5 WE3 Identify the following as either sample statistics or population parameters.

a Studies have shown that between 85% and 95% of lung cancers are related to smoking.

b About 50% of children aged between 9 and 15 years eat the recommended daily amount of fruit.

6 Identify the following as either sample statistics or population parameters.

a According to the 2013 census, the ratio of male births per 100 female births is 106.3.

b About 55% of boys and 40% of girls reported drinking at least 2 quantities of 500 ml of soft drink every day.

7 WE4 A school has 523 boys and 621 girls. You are interested in finding out about their attitudes to sport and believe that boys and girls may respond differently. If a sample of 75 students is required, determine how many boys and how many girls should be selected.

8 In a school, 23% of the students are boarders. For this survey, it is believed that boarders and day students may respond differently. To select a sample of 90 students, how many boarders and day students should be selected?

CONSOLIDATE

Apply the most appropriate mathematical processes and tools

9 You are trying out a new chocolate pudding recipe. You found 40 volunteers to taste test your new recipe compared to your normal pudding. Half of the volunteers were given a serving the new pudding first, then a serving of the old pudding. The other half were given the old pudding first and then the new pudding. The taste testers did not know the order of the puddings they were trying. The results show that 31 people prefer the new pudding recipe.

a What is the population size?

b What is the sample size?

10 You want to test a new flu vaccine on people with a history of chronic asthma. You begin with 500 volunteers and end up with 247 suitable people to test the vaccine.

a What is the population size?

b What is the sample size?

11 In a recent survey, 1 in 5 students indicated that they ate potato crisps or other salty snacks at least four times per week. Is this a sample statistic or a population parameter?

12 Around 25 to 30% of children aged 0–15 years eat confectionary at least four times a week. Is this a sample statistic or a population parameter?

13 According to the Australian Bureau of Statistics, almost a quarter (24%) of internet users did not make an online purchase or order in 2012–13. The three most commonly reported main reasons for not making an online purchase or order were: 'Has no need' (33%); 'Prefers to shop in person/see the product' (24%); and 'Security concerns/concerned about providing credit card details online' (12%). Are these sample statistics or population parameters?

14 According to the 2011 census, there is an average of 2.6 people per household. Is this a sample statistic or a population parameter?

15 A doctor is undertaking a study about sleeping habits. She decides to ask every 10th patient about their sleeping habits.

a What type of sample is this?

b Is this a valid sampling method?

16 A morning television show conducts a viewer phone-in poll and announces that 95% of listeners believe that Australia should become a republic. Comment on the validity of this type of sample.

17 Tony took a survey by walking around the playground at lunch and asking fellow students questions. Why is this not the best sampling method?

18 A company has 1500 staff members, of whom 60% are male; 95% of the male staff work full time, and 78% of the female staff work full time. If a sample of

80 staff is to be selected, identify the numbers of full-time male staff, part-time male staff, full-time female staff and part-time female staff that should be included in the sample.

MASTER

19 Use CAS technology to produce a list of 10 random numbers between 1 and 100.

20 Use CAS technology to select a random sample from students in your Mathematical Methods class.

14.3 The distribution of $\hat{p}$

studyon

Units 3 & 4
AOS 4
Topic 5
Concept 4

Sample proportion
Concept summary
Practice questions

eBook*plus*

Interactivity
Distribution of $\hat{p}$
int-6444

Let us say that we are interested in the following collection of balls. As you can see in Figure 1, there are 20 balls, and $\frac{1}{4}$ of them are red. This means that the population parameter, p, is $\frac{1}{4}$ and the population size, N, is 20.

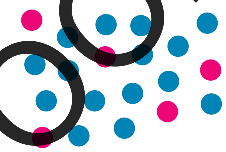


Figure 1

Normally we wouldn't know the population parameter, so we would choose a sample from the population and find the sample statistic. In this case, we are going to use a sample size of 5, that is, $n = 5$.

If our sample is the group shown in Figure 2, then as there is 1 red ball, the **sample proportion** would be $\hat{p} = \frac{1}{5}$.

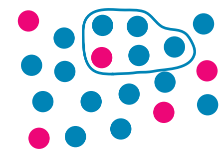


Figure 2

A different sample could have a different sample proportion. In the case shown in Figure 3, $\hat{p} = \frac{2}{5}$.

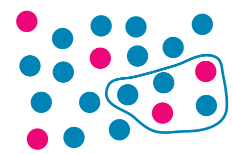


Figure 3

In the case shown in Figure 4, $\hat{p} = 0$.

Figure 4

It would also be possible to have samples for which $\hat{p} = \frac{3}{5}$, $\hat{p} = \frac{4}{5}$ or $\hat{p} = 1$, although these samples are less likely to occur.

In summary,

$$\hat{p} = \frac{\textbf{number of successful outcomes in the sample}}{\textbf{sample size}}$$

It might seem that using a sample does not give a good estimate about the population. However, the larger the sample size, the more likely that the sample proportions will be close to the population proportion.

WORKED EXAMPLE 5

You are trying out a new chocolate tart recipe. You found 40 volunteers to taste test your new recipe compared to your normal one. Half the volunteers were given a serving of the new tart first, then a serving of the old tart. The other half were given the old tart first and then the new one. The taste testers did not know the order of the tarts they were trying. The results show that 31 people prefer the new tart recipe.

What is the sample proportion, $\hat{p}$?

THINK	WRITE
1 There are 40 volunteers. This is the sample size.	$n = 40$
2 31 people prefer the new recipe.	Number of successes $= 31$
3 Calculate the sample proportion.	$\hat{p} = \frac{31}{40}$

Revision of binomial distributions

In a set of binomial data, each member of the population can have one of two possible values. We define one value as a success and the other value as a failure. (A success isn't necessarily a good thing, it is simply the name for the condition we are counting. For example, a success may be having a particular disease and a failure may be not having the disease).

The proportion of successes in a population is called p and is a constant value.

$$p = \frac{\text{number in the population with the favourable attribute}}{\text{population size}}$$

The proportion of failures in a population is called q, where $q = 1 - p$.

The sample size is called n.

The number of successes in the sample is called X.

The proportion of successes in the sample, $\hat{p}$, will vary from one sample to another.

$$\begin{aligned}\hat{p} &= \frac{\text{number in the sample with the favourable attribute}}{\text{sample size}}\\ &= \frac{X}{n}\end{aligned}$$

Sampling distribution of $\hat{p}$

eBook*plus*

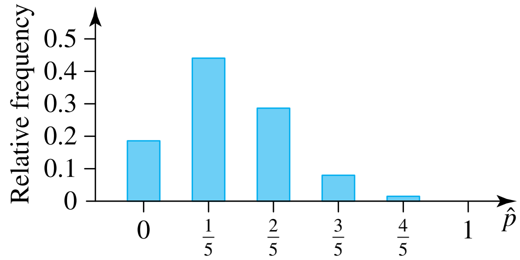
Interactivity
Sampling distribution of $\hat{p}$
int-6445

Normally, you would take one sample from a population and make some inferences about the population from that sample. In this section, we are going to explore what would happen if you took lots of samples of the same size. (Assume you return each sample back to the population before selecting again.)

Consider our population of 20 balls (5 red and 15 blue). There are ${}^{20}C_5 = 15\,504$ possible samples that could be chosen. That is, there are 15 504 possible ways of choosing 5 balls from a population of 20 balls. A breakdown of the different samples is shown in the table, where X is the number of red balls in the sample.

X	$\hat{p}$	Number of samples	Relative frequency
0	0	${}^5C_0\ {}^{15}C_5 = 3003$	0.194
1	$\frac{1}{5}$	${}^5C_1\ {}^{15}C_4 = 6825$	0.440
2	$\frac{2}{5}$	${}^5C_2\ {}^{15}C_3 = 4550$	0.293
3	$\frac{3}{5}$	${}^5C_3\ {}^{15}C_2 = 1050$	0.068
4	$\frac{4}{5}$	${}^5C_4\ {}^{15}C_1 = 75$	0.005
5	1	${}^5C_5\ {}^{15}C_0 = 1$	6.450×10^{-5}
	Total number of samples	**15 504**	

Graphing the distribution of $\hat{p}$ against the relative frequency of $\hat{p}$ results in the following.



As the value of $\hat{p}$, the sample proportion, varies depending on the sample, these values can be considered as the values of the random variable, $\hat{P}$.

The graph of the distribution of $\hat{p}$ can also be represented in a probability distribution table. This distribution is called a **sampling distribution**.

$\hat{p}$	0	$\frac{1}{5}$	$\frac{2}{5}$	$\frac{3}{5}$	$\frac{4}{5}$	1
$\Pr(\hat{P} = \hat{p})$	0.194	0.440	0.293	0.068	0.005	6.450×10^{-5}

We can find the average value of $\hat{p}$ as shown.

$\hat{p}$	Frequency, f	$f \cdot \hat{p}$
0	3003	0
$\frac{1}{5}$	6825	1365
$\frac{2}{5}$	4550	1820
$\frac{3}{5}$	1050	630
$\frac{4}{5}$	75	60
1	1	1
Totals	**15 504**	**3876**

$$\text{The average value of } \hat{p} = \frac{3876}{15\,504}$$
$$= 0.25$$

For this distribution, the average value for $\hat{p}$ is equal to the population proportion, p.

Sampling where the population is large

It was mentioned earlier that larger samples give better estimates of the population.

studyon

Units 3 & 4
AOS 4
Topic 5
Concept 6

Normal approximation to sampling distribution of proportion
Concept summary
Practice questions

Expected value

The proportion of $\hat{p}$ in a large sample conforms to $\hat{P} = \frac{X}{n}$. As the sample is from a large population, X can be assumed to be a binomial variable.

$$\therefore \text{E}(\hat{P}) = \text{E}\left(\frac{X}{n}\right)$$
$$= \frac{1}{n}\text{E}(X) \qquad \left(\text{because } \frac{1}{n} \text{ is a constant}\right)$$
$$= \frac{1}{n} \times np$$
$$= p$$

Variance and standard deviation

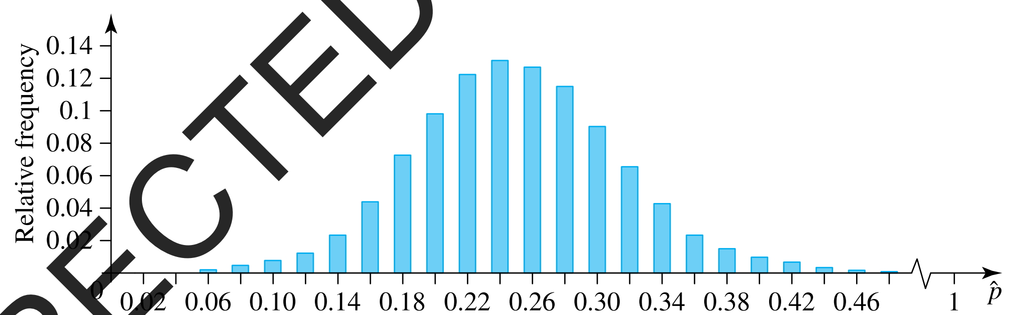
The variance and standard deviation can be found as follows.

$$\begin{aligned}\operatorname{Var}(\hat{P}) &= \operatorname{Var}\left(\frac{X}{n}\right)\\ &= \left(\frac{1}{n}\right)^2 \operatorname{Var}(X)\\ &= \frac{1}{n^2} \times npq\\ &= \frac{pq}{n}\\ &= \frac{p(1-p)}{n}\\ \therefore \operatorname{SD}(\hat{P}) &= \sqrt{\frac{p(1-p)}{n}}\end{aligned}$$

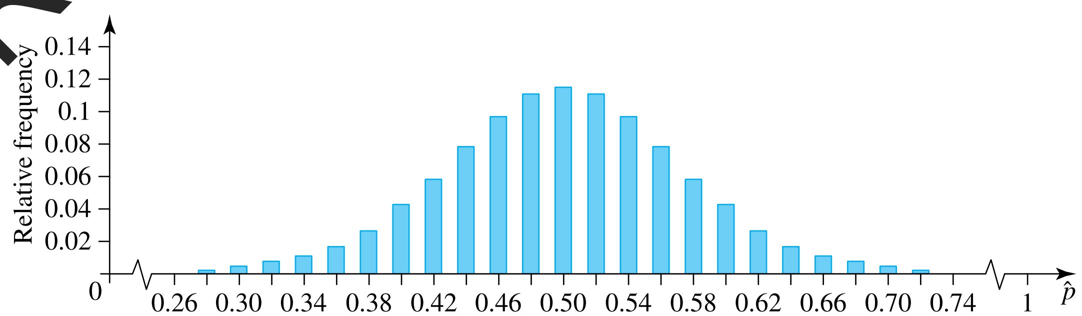
For large samples, the distribution of $\hat{p}$ is approximately normal with a mean or expected value of $\mu_{\hat{p}} = p$ and a standard deviation of $\sigma_{\hat{p}} = \sqrt{\frac{p(1-p)}{n}}$.

There are a number of different ways to decide if a sample is large. One generally accepted method that we will adopt for this section is that if $np \geq 10$, $nq \geq 10$ and $10n \leq N$, then the sample can be called large.

Consider the distribution of $\hat{p}$ when $N = 1000$, $n = 50$ and $p = 0.25$.



And consider this distribution of $\hat{p}$ when $N = 1000$, $n = 50$ and $p = 0.5$.



As these graphs show, the value of p doesn't matter. The distribution of $\hat{p}$ is symmetrical about p.

WORKED EXAMPLE 6

Consider a population size of 1000 and a sample size of 50. If $p = 0.1$, would this still be a large sample? If not, how big would the sample need to be?

THINK	WRITE
1 Is $10n \le N$?	$n = 50$ and $N = 1000$ $10n = 500$ Therefore, $10n \le N$.
2 Is $np \ge 10$?	$p = 0.1$ $np = 0.1 \times 50$ $= 5$ $5 \ngeq 10$ The sample is not large.
3 Find a value for n to make a large sample by solving $np = 10$.	$np = 10$ $0.1n = 10$ $n = 100$
4 Check the other conditions.	$10n = 10 \times 100$ $= 1000$ $= N$ $nq = 100 \times 0.9$ $= 90$ $nq \ge 10$ A sample size of 100 would be needed for a large sample.

WORKED EXAMPLE 7

If $N = 600$, $n = 60$ and $p = 0.3$:

a find the mean of the distribution

b find the standard deviation of the distribution, correct to 2 decimal places.

THINK	WRITE
a The mean is p.	a $\mu_{\hat{p}} = p$ $= 0.3$
b 1 Write the rule for the standard deviation.	b $\sigma_{\hat{p}} = \sqrt{\dfrac{p(1-p)}{n}}$
2 Substitute the appropriate values and simplify.	$= \sqrt{\dfrac{0.3 \times (1 - 0.3)}{60}}$ $= 0.06$

EXERCISE 14.3 The distribution of $\hat{p}$

PRACTISE

1 WE5 In a 99-g bag of lollies, there were 6 green lollies out of the 15 that were counted. What is the sample proportion, $\hat{p}$?

2 Hang is interested in seedlings that can grow to more than 5 cm tall in the month of her study period. She begins with 20 seedlings and finds that 6 of them are more than 5 cm tall after the month. What is the sample proportion, $\hat{p}$?

3 WE6 Consider a population size of 1000 and a sample size of 50. If $p = 0.85$, is this a large sample? If not, how big does the sample need to be?

4 If the population size was 10 000 and $p = 0.05$, what would be a large sample size?

5 WE7 If $N = 500$, $n = 50$ and $p = 0.5$:

a find the mean of the distribution
b find the standard deviation of the distribution, correct to 2 decimal places.

6 If $N = 1000$, $n = 100$ and $p = 0.8$:

a find the mean of the distribution
b find the standard deviation of the distribution, correct to 2 decimal places.

CONSOLIDATE

Apply the most appropriate mathematical processes and tools

7 A car manufacturer has developed a new type of bumper that is supposed to absorb impact and result in less damage than previous bumpers. The cars are tested at 25 km/h. If 30 cars are tested and only 3 are damaged, what is the proportion of undamaged cars in the sample?

8 A standard warranty lasts for 1 year. It is possible to buy an extended warranty for an additional 2 years. The insurer decides to use the sales figures from Tuesday to estimate the proportion of extended warranties sold. If 537 units were sold and 147 of them included extended warranties, estimate the proportion of sales that will include extended warranties.

9 A Year 12 Mathematical Methods class consists of 12 girls and 9 boys. A group of 4 students is to be selected at random to represent the school at an inter-school Mathematics competition.

a What is the value of p, the proportion of girls in the class?
b What could be the possible values of the sample proportion, $\hat{p}$, of girls?
c Use this information to construct a probability distribution table to represent the sampling distribution of the sample proportion of girls in the small group.
d Find $\Pr(\hat{P} > 0.6)$. That is, find the probability that the proportion of girls in the small group is greater than 0.6.
e Find $\Pr(\hat{P} > 0.5 \mid \hat{P} > 0.3)$.

10 In a particular country town, the proportion of employment in the farming industry is 0.62. Five people aged 15 years and older are selected at random from the town.

a What are the possible values of the sample proportion, $\hat{p}$, of workers in the farming industry?
b Use this information to construct a probability distribution table to represent the sampling distribution of the sample proportion of workers in the farming industry.
c Find the probability that the proportion of workers in the farming industry in the sample is greater than 0.5.

11 In a population of 1.2 million, it is believed that $p = 0.01$. What would be the smallest sample size that could be considered large?

12 If $N = 1500$, $n = 150$ and $p = 0.15$, find the mean and standard deviation for the distribution of $\hat{p}$. Give your answers correct to 3 decimal places where appropriate.

13 If $N = 1200$, $n = 100$ and $p = 0.75$, find the mean and standard deviation for the distribution of $\hat{p}$. Give your answers correct to 3 decimal places where appropriate.

14 A distribution for $\hat{p}$ has a mean of 0.12 and a standard deviation of 0.0285. Find the population proportion and the sample size.

15 A distribution for $\hat{p}$ has a mean of 0.81 and a standard deviation of 0.0253. Find the population proportion and the sample size.

16 If $N = 1500$, $n = 150$ and $p = 0.15$, use CAS technology to graph the distribution for $\hat{p}$.

MASTER

17 A distribution for $\hat{p}$ has a standard deviation of 0.015. If the sample size was 510 and $\hat{p} > 0.5$, what was the population proportion, correct to 2 decimal places?

18 A distribution for $\hat{p}$ has a standard deviation of 0.0255. If the sample size was 350, what was the population proportion, correct to 2 decimal places?

14.4 Confidence intervals

studyon

Units 3 & 4
AOS 4
Topic 5
Concept 7

Confidence limits for the population proportion
Concept summary
Practice questions

We have just learned that different samples can have different values for $\hat{p}$. So what can one sample tell us about a population?

Let us say that you are interested in the proportion of the school that buys their lunch. You decide that your class is a reasonable sample and find out that 25% of the class will buy their lunch today. What can you say about the proportion of the whole school that will buy their lunch today? Assuming that your class is in fact a representative sample, you may say that around 25% of the school will buy their lunch. Is it possible to be more specific? By using **confidence intervals**, it is possible to say how confident you are that a population parameter will lie in a particular interval.

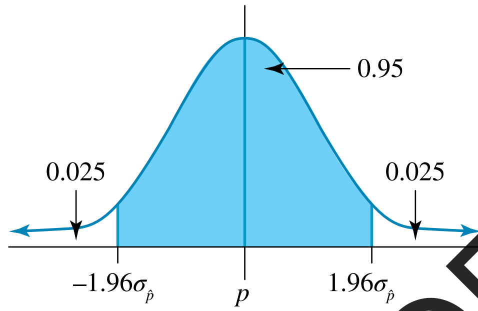
Normal approximation to the distribution of $\hat{p}$

We have learned that when we consider the distributions of $\hat{p}$, they are normally distributed with a $\mu_{\hat{p}} = p$ and $\sigma_{\hat{p}} = \sqrt{\dfrac{p(1-p)}{n}}$. As we don't know the exact value for p, the best estimate is $\hat{p}$. This means that the best estimate of the standard deviation is $\sqrt{\dfrac{\hat{p}(1-\hat{p})}{n}}$.

We know that for normal distributions, $z = \dfrac{x-\mu}{\sigma}$. This means that, to find the upper and lower values of z, we can use $z = \dfrac{\hat{p} \pm p}{\sigma_{\hat{p}}}$. Rearranging gives us $p = \hat{p} \pm z\sigma_{\hat{p}}$.

An approximate confidence interval for a population proportion is given by $(\hat{p} - z\sigma_{\hat{p}}, \hat{p} + z\sigma_{\hat{p}})$, where $\sigma_{\hat{p}} = \sqrt{\dfrac{\hat{p}(1-\hat{p})}{n}}$.

A 95% confidence interval means that 95% of the distribution is in the middle area of the distribution. This means that the tails combined contain 5% of the distribution (2.5% on each end). The z-score for this distribution is 1.96.



The confidence interval for this distribution can be expressed as

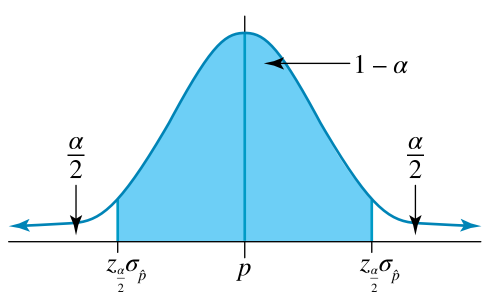
$$\left(\hat{p} - 1.96\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}, \hat{p} + 1.96\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}\right).$$

WORKED EXAMPLE 8

There are 20 people in your class and 25% are planning on buying their lunch. Estimate the proportion of the school population that will purchase their lunch today. Find a 95% confidence interval for your estimate, given $z = 1.96$.

THINK	WRITE
1 There are 20 people in the class. This is the sample size. 25% are buying their lunch. This is the sample proportion.	$n = 20$ $\hat{p} = 0.25$
2 For a 95% confidence interval, $z = 1.96$.	$z = 1.96$
3 The confidence interval is $\left(\hat{p} - z\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}, \hat{p} + z\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}\right)$. Find $z\sigma_{\hat{p}}$.	$z\sigma_{\hat{p}} = z\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$ $= 1.96\sqrt{\frac{0.25 \times 0.75}{20}}$ $= 0.1898$
4 Identify the 95% confidence interval by finding the upper and lower values.	$\hat{p} - z\sigma_{\hat{p}} = 0.25 - 0.1898$ $= 0.0602$ $\hat{p} + z\sigma_{\hat{p}} = 0.25 + 0.1898$ $= 0.4398$ We can be 95% confident that between 6% and 44% of the population will buy their lunch today.

To find other confidence intervals, we can talk in general about a $1 - \alpha$ confidence interval. In this case, the tails combined will have an area of α (or $\frac{\alpha}{2}$ in each tail). In this case, the z-score that has a tail area of $\frac{\alpha}{2}$ is used.



WORKED EXAMPLE 9

Paul samples 102 people and finds that 18 of them like drinking coconut milk. Estimate the proportion of the population that likes drinking coconut milk. Find a 99% confidence interval for your estimate, correct to 1 decimal place.

THINK	WRITE
1 There are 102 people in the sample. This is the sample size. 18 like drinking coconut milk.	$n = 102$ $\hat{p} = \frac{18}{102}$ $= 0.18$
2 For a 99% confidence interval, find the z score using the inverse standard normal distribution.	For the 99% confidence interval, 1% will be in the tails, so 0.5% in each tail. Therefore, the area under the normal distribution curve to the left of z is 0.995. $z = 2.58$
3 The confidence interval is $(\hat{p} - z\sigma_{\hat{p}}, \hat{p} + z\sigma_{\hat{p}})$. Find $z\sigma_{\hat{p}}$.	$z\sigma_{\hat{p}} = z\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$ $= 2.58\sqrt{\frac{0.18 \times 0.82}{102}}$ $= 0.098$
4 Identify the 99% confidence interval by finding the upper and lower values, correct to 1 decimal place.	$\hat{p} - z\sigma_{\hat{p}} = 0.18 - 0.098$ $= 0.082$ $\hat{p} + z\sigma_{\hat{p}} = 0.18 + 0.098$ $= 0.278$ We can be 99% confident that between 8.2% and 27.8% of the population like drinking coconut milk.

WORKED EXAMPLE 10

Grow Well are 95% sure that 30% to 40% of shoppers prefer their mulch. What sample size was needed for this level of confidence?

THINK	WRITE
1 The confidence interval is symmetric about $\hat{p}$: $(\hat{p} - z\sigma_{\hat{p}}, \hat{p} + z\sigma_{\hat{p}})$, so the value of $\hat{p}$ must be halfway between the upper and lower values of the confidence interval.	$\hat{p} = \frac{30 + 40}{2}$ $= 35\%$ $= 0.35$
2 State the z-score related to the 95% confidence interval.	$z = 1.96$
3 The lower value of the confidence interval, 30%, is equivalent to $\hat{p} - z\sigma_{\hat{p}}$. Substitute the appropriate values. *Note:* The equation $0.4 = \hat{p} + z\sigma_{\hat{p}}$ could also have been used.	$0.3 = \hat{p} - z\sigma_{\hat{p}}$ $= \hat{p} - z\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$ $= 0.35 - 1.96\sqrt{\frac{0.35(1-0.35)}{n}}$
4 Solve for n.	$n = 349.586$
5 Write the answer.	The sample size needed was 350 people.

Margin of error

The distance between the endpoints of the confidence interval and the sample estimate is called the **margin of error**, M.

Worked example 10 considered a 95% confidence interval, $(\hat{p} - z\sigma_{\hat{p}}, \hat{p} + z\sigma_{\hat{p}})$. In this case the margin of error would be $M = z\sigma_{\hat{p}}$.

For a 95% level of confidence, $M = 1.96\sqrt{\dfrac{\hat{p}\,(1 - \hat{p})}{n}}$.

Note that the larger the sample size, the smaller the value of M will be. This means that one way to reduce the size of a confidence interval without changing the level of confidence is to increase the sample size.

EXERCISE 14.4 Confidence intervals

PRACTISE

1 WE8 Of 30 people surveyed, 78% said that they like breakfast in bed. Estimate the proportion of the population that like breakfast in bed. Find a 95% confidence interval for the estimate.

2 Of the 53 people at swimming training today, 82% said that their favourite stroke is freestyle. Estimate the proportion of the population whose favourite stroke is freestyle. Find a 95% confidence interval for the estimate.

3 WE9 Jenny samples 116 people and finds that 86% of them plan to go swimming over the summer holidays. Estimate the proportion of the population that plan to go swimming over the summer holidays. Find a 99% confidence interval for your estimate.

4 Yuki samples 95 people and finds that 30% of them eat chocolate daily. Estimate the proportion of the population that eats chocolate daily. Find a 90% confidence interval for your estimate.

5 WE10 In a country town, the owners of Edie's Eatery are 95% sure that 35% to 45% of their customers love their homemade apple pie. What sample size was needed for this level of confidence?

6 If Parkers want to be 90% confident that between 75% and 85% of their customers will shop in their store for more than 2 hours, what sample size will be needed?

CONSOLIDATE

Apply the most appropriate mathematical processes and tools

The following information relates to question 7 and 8.

Teleco is being criticised for its slow response time when handling complaints. The company claims that it will respond within 1 day. Of the 3760 complaints in a given week, a random sample of 250 was selected. Of these, it was found that 20 of them had not been responded to within 1 day.

7 Find the 95% confidence interval for the proportion of claims that take more than 1 day to resolve.

8 What is the 99% confidence interval for the proportion of claims that take less than 1 day to resolve?

9 A sample of 250 blood donors have their blood types recorded. Of this sample, 92 have Type A blood. What is the 90% confidence interval for the proportion of Australians who have Type A blood?

10 It is believed that 65% of people have brown hair. A random selection of 250 people were asked the colour of their hair. Applying the normal approximation, find the probability that less than 60% of the people in the sample have brown hair.

11 Nidya is a top goal shooter. The probability of her getting a goal is 0.8. To keep her skills up, each night she has 200 shots on goal. Applying the normal approximation, find the probability that on Monday the proportion of times she scores a goal is between 0.8 and 0.9, given that it is more than 0.65.

12 Smooth Writing are 95% sure that 25% to 35% of shoppers prefer their pen. What sample size was needed for this level of confidence?

13 An online tutoring company is 99% sure that 20% to 30% of students prefer to use their company. What sample size was needed for this level of confidence?

14 Teleco want to be 95% sure that less than 5% of their complaints take more than 1 day to resolve. What sample proportion do they need and how large does the sample need to be to support this claim?

15 Barton's Dentistry want to be able to claim that 90% to 98% of people floss daily. They would like 99% confidence about their claim. How many people do they need to survey?

16 Tatiana is conducting a survey to estimate the proportion of Year 12 students who will take a gap year after they complete their VCE. Previous surveys have shown the proportion to be approximately 15%. Determine the required size of the sample so that the margin of error for the survey is 3% in a confidence interval of approximately 95% for this proportion.

MASTER

17 Bentons claim that between 85% and 95% of their customers stay for more than 2 hours when they shop. If they surveyed 100 people, how confident are they about that claim?

18 The Hawthorn Football Club claim that between 75% and 80% of their members remain members for at least 10 years. If they surveyed 250 people, how confident are they about that claim? Give your answer to the nearest whole number.

eBook*plus* ONLINE ONLY

14.5 Review

www.jacplus.com.au

The Maths Quest Review is available in a customisable format for you to demonstrate your knowledge of this topic.

The Review contains:

- **Multiple-choice** questions — providing you with the opportunity to practise answering questions using CAS technology
- **Short-answer** questions — providing you with the opportunity to demonstrate the skills you have developed to efficiently answer questions without the use of CAS technology
- **Extended-response** questions — providing you with the opportunity to practise exam-style questions.

A summary of the key points covered in this topic is also available as a digital document.

REVIEW QUESTIONS

Download the Review questions document from the links found in the Resources section of your eBookPLUS.

eBook*plus* ONLINE ONLY

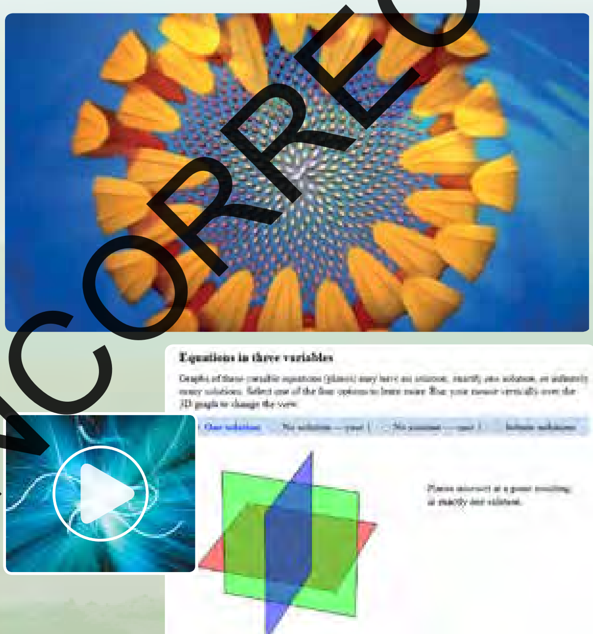
Activities

To access eBookPLUS activities, log on to

www.jacplus.com.au

Interactivities

A comprehensive set of relevant interactivities to bring difficult mathematical concepts to life can be found in the Resources section of your eBookPLUS.



studyon

studyON is an interactive and highly visual online tool that helps you to clearly identify strengths and weaknesses prior to your exams. You can then confidently target areas of greatest need, enabling you to achieve your best results.

14 Answers

EXERCISE 14.2

1 $N = 120$, $n = 30$

2 $N = 100$, $n = 5$

3 $n = 15$, population size is unknown

4 $n = 9$, population size is unknown

5 a Population parameter b Sample statistic

6 a Population parameter b Sample statistic

7 34 boys and 41 girls

8 21 boarders, 69 day students

9 a The population size is unknown.
  b $n = 40$

10 a The population is people who will receive the vaccine in the future. The size is unknown.
  b $n = 247$

11 Sample statistic

12 Population parameter

13 Sample statistics

14 Population parameter

15 a A systematic sample with $k = 10$
  b Yes, assuming that the order of patients is random

16 The sample is not random; therefore, the results are not likely to be random.

17 It is probably not random. Tony is likely to ask people who he knows or people who approach him.

18 Full-time male staff: 46
Part-time male staff: 2
Full-time female staff: 25
Part-time female staff: 7

19 Use the random number generator on your calculator to produce numbers from 1 to 100. Keep generating numbers until you have 10 different numbers.

20 First, assign every person in your class a number, e.g. 1 to 25 if there are 25 students in your class. Decide how many students will be in your sample, e.g. 5. Then use the random number generator on your calculator to produce numbers from 1 to 25. Keep generating numbers until you have 5 different numbers. The students that were assigned these numbers are the 5 students in your random sample.

EXERCISE 14.3

1 $\hat{p} = \frac{2}{5}$

2 $\hat{p} = \frac{3}{10}$

3 This is not a large sample; $n = 67$ would be a large sample.

4 $n = 200$

5 a $\mu_{\hat{p}} = 0.5$ b $\sigma_{\hat{p}} = 0.07$

6 a $\mu_{\hat{p}} = 0.8$ b $\sigma_{\hat{p}} = 0.04$

7 $\hat{p} = \frac{9}{10}$

8 $\hat{p} = \frac{147}{537}$

9 a $p = \frac{4}{7}$
  b $0, \frac{1}{4}, \frac{1}{2}, \frac{3}{4}, 1$
  c

$\hat{p}$	0	$\frac{1}{4}$	$\frac{1}{2}$	$\frac{3}{4}$	1
$\Pr(\hat{P} = \hat{p})$	0.021	0.168	0.397	0.331	0.083

  d 0.414
  e 0.510

10 a $0, \frac{1}{5}, \frac{2}{5}, \frac{3}{5}, \frac{4}{5}, 1$
  b

$\hat{p}$	0	$\frac{1}{5}$	$\frac{2}{5}$	$\frac{3}{5}$	$\frac{4}{5}$	1
$\Pr(\hat{P} = \hat{p})$	0.008	0.064	0.211	0.344	0.281	0.092

  c 0.717

11 $n = 1000$

12 $\mu_{\hat{p}} = 0.15$, $\sigma_{\hat{p}} = 0.029$

13 $\mu_{\hat{p}} = 0.75$, $\sigma_{\hat{p}} = 0.043$

14 $p = 0.12$, $n = 130$

15 $p = 0.81$, $n = 240$

16 See the figure at the foot of the page.*

17 $p = 0.87$

18 $p = 0.35$ or $p = 0.65$

EXERCISE 14.4

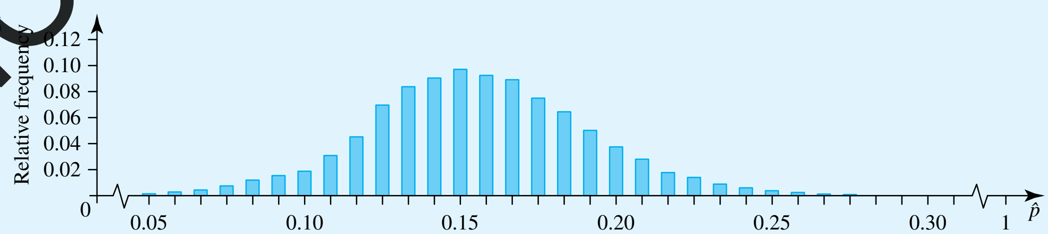
1 63%–93%

2 72%–92%

3 78%–94%

4 22%–38%

*16



5 $n = 369$

6 $n = 173$

7 4.6%–11.4% of complaints take more than 1 day to resolve.

8 87.6%–96.4% of complaints are resolved within 1 day.

9 31.8%–41.8% of Australians have Type A blood.

10 0.0487

11 0.4998

12 $n = 323$

13 $n = 498$

14 $\hat{p} = 2.5\%$, $n = 150$

15 $n = 235$

16 544 people

17 90% sure

18 66%