Shutterstock.com/Amy Johansson

10

INTERVAL ESTIMATES FOR PROPORTIONS

CONFIDENCE INTERVALS

TERMINOLOGY

central limit theorem
confidence interval
confidence level
estimator
interval estimate
margin of error
parameter
point estimate
population
probability
proportion
quantile
sample
sampling distribution
sample proportion
simulation
standard deviation
standard normal distribution
standard normal variable
statistic
variance
z-score

Prior learning

CONFIDENCE INTERVALS FOR PROPORTIONS

- the concept of an interval estimate for a parameter associated with a random variable (ACMMM177)
- the approximate confidence interval $\left(\hat{p}-z\sqrt{\hat{p}(1-\hat{p})/n}, \hat{p}+z\sqrt{\hat{p}(1-\hat{p})/n}\right)$, as an interval estimate for p, where z is the appropriate quantile for the standard normal distribution (ACMMM178)
- define the approximate margin of error $E = z\sqrt{\hat{p}(1-\hat{p})/n}$ and understand the trade-off between margin of error and level of confidence (ACMMM179)
- use simulation to illustrate variations in confidence intervals between samples and to show that most but not all confidence intervals contain p. (ACMMM180) AC

10.01 INTERVAL ESTIMATES FOR SAMPLE PROPORTIONS

You cannot predict the value of a random variable on any particular occasion. However, you do expect the typical value of a random variable to be close to the mean value. You can use a value from a sample as an estimate of the value for the population concerned.

Remember that a value obtained from a population is called a **parameter**, while a value obtained from a sample is called a **statistic**. You use statistics as estimates of parameters.

IMPORTANT

A **point estimate** of a parameter is a single value obtained from a sample. The statistic used is called an **estimator** of the parameter.

You have already seen that the expected value of a sample proportion is the probability of a property (characteristic) in a population; $E(\hat{p}) = p$. This means that sample proportion is the best estimator of that probability. You also expect that as the sample size is increased, the accuracy of the estimate will improve. In the extreme case where the whole population is used as the sample, the sample proportion must actually be the probability.

You can estimate the probability of Australians having blue eyes using the sample proportion, but if you said you thought it was between 20% and 40%, this would be an **interval estimate**.

IMPORTANT

An **interval estimate** of a parameter is an interval that is likely to include the value of the parameter.

9780170250290

Example 1

State whether each of the following is a point or interval estimate.

a The average height of Year 12 Australian boys is thought to be about 181 cm.

b Average adult human height is between 165 cm and 175 cm.

Solution

a 181 cm is a single value. This is a point estimate.

b This is a range. This is an interval estimate.

INVESTIGATION Point and interval estimates in the media

The use of estimates in the media is very common, both in reporting and in advertising. In many cases, the estimates are based on samples, but it may not be clear whether a point estimate or an interval estimate has been used. For example, a headline might read '60% of Australians support a republic'. This is apparently a point estimate of the probability of an Australian supporting such a change. However, you might find that details of the survey contain a phrase such as '60% with an error of 4%', which indicates an interval estimate of 0.56 £ *p* £ 0.64. Before you could decide whether such an estimate is likely to be worthwhile you should probably ask 'how big was the sample?'.

- Examine media reports to determine whether statistical estimates are point estimates or interval estimates.
- What proportion of media reports use interval estimates?
- What proportion of reports state the sample size?
- What proportion of reports state the nature of the sample?

The variance and standard deviation of sample proportions are given by $Var(\hat{p}) = \frac{pq}{n}$ and $SD(\hat{p}) = \sqrt{\frac{pq}{n}} = \sqrt{\frac{p(1-p)}{n}}$, where p is the probability of the characteristic and n is the number in the sample.

$\hat{p}$ is the best estimator of p, so you can *estimate* the variance and standard deviation of the sample proportion by replacing p by $\hat{p}$.

IMPORTANT

The **estimated variance** and **estimated standard deviation** of a sample proportion $\hat{p}$ with sample size n are $Var(\hat{p}) \approx \frac{\hat{p}\hat{q}}{n} = \frac{\hat{p}(1-\hat{p})}{n}$ and $SD(\hat{p}) \approx \sqrt{\frac{\hat{p}\hat{q}}{n}} = \sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$.

This means that you can estimate both mean and the standard deviation of the **sampling distribution** of proportions from a single sample. Remember that a sampling distribution is the distribution of a statistic obtained from multiple samples (of the same size) from the same population.

Example 2

From a sample of 60 Year 12 students, 45 said they like chocolate. Estimate the probability of Year 12 students liking chocolate and the variance of the sampling distribution.

Solution

Use the sample proportion to estimate p.

$$p \approx \hat{p} = \frac{45}{60} = 0.75$$

Write the formula for the estimated variance.

$$\text{Var}(\hat{p}) \approx \frac{\hat{p}(1-\hat{p})}{n}$$

Substitute $\hat{p}$ and n.

$$= \frac{0.75 \times (1-0.75)}{60}$$
$$= 0.003\ 125$$

Write the answer.

The probability of Year 12 students liking chocolate is about 0.75 and the variance of the sampling distribution is about 0.0031.

It doesn't make sense to show more than 2 significant figures in the estimates because the figures used in the calculation were only accurate to 2 figures.

You can use the standard deviation of the sampling distribution to work out an interval estimate of probabilities obtained from a sample proportion.

WS
Calculations with sample proportions

Example 3

In 2010, a survey of Victorians found 268 out of every 327 people aged 20–24 had completed Year 12. Use this to estimate the probability that a Victorian aged 20–24 has completed Year 12, within one standard deviation.

Solution

Estimate the probability.

$$\hat{p} \approx \frac{268}{327} = 0.8195\ldots$$

Write the formula for the standard deviation.

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

Substitute $\hat{p}$ and n.

$$= \sqrt{\frac{0.8195\ldots(1-0.8195\ldots)}{327}}$$
$$= 0.0212\ldots$$

Find the interval.

$$\hat{p} \approx 0.8195\ldots \pm 0.0212\ldots$$
$$= 0.7983\ldots \text{ to } 0.8408\ldots$$

Write the answer.

The probability that a Victorian aged 20–24 has completed Year 12 is between 0.798 and 0.841, within one standard deviation.

Is it reasonable to give the estimate in the previous example to 3 significant figures, or should it be given to greater or less accuracy?

9780170250290

EXERCISE 10.01 Interval estimates for sample proportions

WS Simple interval estimates

Concepts and techniques

1 Example 1 State whether each of the following is a point or interval estimate.
   a The mean income is about $47 000.
   b The median house price in suburbs is between $350 000 and $720 000.
   c The variance is between 800 and 1500.
   d The probability of dark hair is about 0.6.

2 Example 2 From 80 people interviewed at Spencer St Station, 57 said they were competent in using chopsticks. Estimate the probability that Melbourne residents are competent in using chopsticks and the variance and standard deviation of the sampling distribution.

3 From 40 tosses, a coin lands heads up 18 times. Estimate the probability of this coin landing with heads up and the variance and standard deviation of the sampling distribution.

4 From a sample of 87 Australian men over the age of 30, 52 were overweight or obese. Estimate the probability of Australian men over 30 being overweight or obese and the variance and standard deviation of the sampling distribution.

5 The protractors of 50 students were checked for accuracy, and 16 were found to be inaccurate in measuring an angle of 60° (more than 1° out). Estimate the probability of students' protractors being inaccurate and the variance and standard deviation of the sampling distribution.

6 Example 3 From a group of 50 computer professionals, 32 said they used the Linux operating system on their home computers. Estimate the probability of a computer professional using Linux on their home computer to within one standard deviation.

7 From a class of 28 Year 12 students, only 19 had completed all of their homework the previous night. Use this information to estimate the probability of a Year 12 student completing all of their homework to within 1.5 standard deviations.

8 A group of 32 parents waiting for interviews at a high school Parent-teacher evening were discussing the habits of teenagers and 20 of them agreed that 'teenagers take too long in the bathroom'. What is the probability that parents of high school students think they take too long in the bathroom, to within 1.6 standard deviations?

Reasoning and communication

9 A poll of 500 voters found that 44% supported a particular political party. What is the standard deviation of the sampling distribution?

10 A sample showed 30% support for a proposal and the estimated variance of the sampling distribution was 0.005. How many people were surveyed?

10.02 CONFIDENCE LEVELS AND MARGIN OF ERROR

The **Central limit theorem** guarantees that for sufficiently large samples, the distribution of the sample means is approximately normal. This means that you can use the normal distribution to work out the size an interval must be to contain any particular proportion of the sample means.

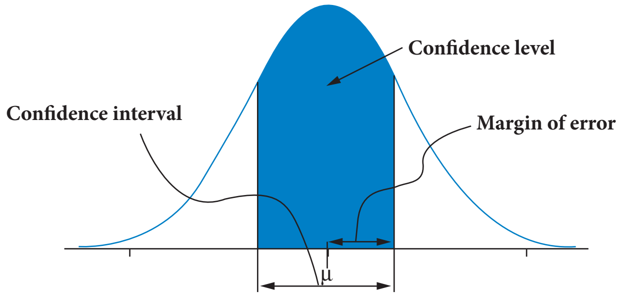
If you choose an interval of a normal distribution from 2 standard deviations below the mean to 2 standard deviations above the mean, then about 95% of the values will be in the interval. The standard normal distribution has the mean $\mu = 0$ and standard deviation $\sigma = 1$. This means that for the standard normal variable Z, $P(-2 \leq z \leq 2) \approx 0.95$. You say that the 95% **confidence level** for the standard normal variable has a **margin of error** of ±2.

IMPORTANT

For a **confidence interval** symmetric around the mean in a statistical distribution:

- the **confidence level** is the proportion of values that lie within the interval
- the **margin of error** is the distance of the ends of the interval from the mean.

For a normal distribution, you can picture this as shown below.



The confidence level is the area under the curve, which is both the proportion of values and the probability of lying within the interval.

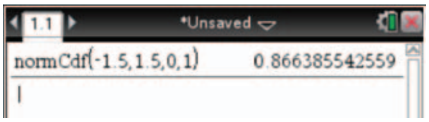
Example 4

CAS What confidence level is produced by a margin of error of 1.5 in a standard normal variable?

Solution

TI-Nspire CAS

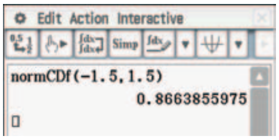
Use a calculator page.
Press menu, 6: Statistics, 5: Distributions and 2: Normal Cdf. Choose boundaries of −1.5 and 1.5; or type normCdf(−1.5, 1.5, 0, 1).



ClassPad

Use the Main menu.

Tap Action, Distribution/Inv. Dist, Continuous and choose normCDf. Choose boundaries of −1.5 and 1.5; or type normCDf(−1.5, 1.5, 1, 0).

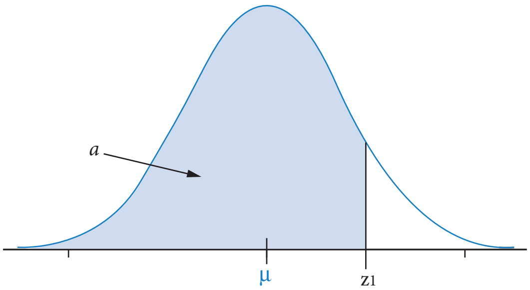


Round and write the answer.

The confidence level is about 87%.

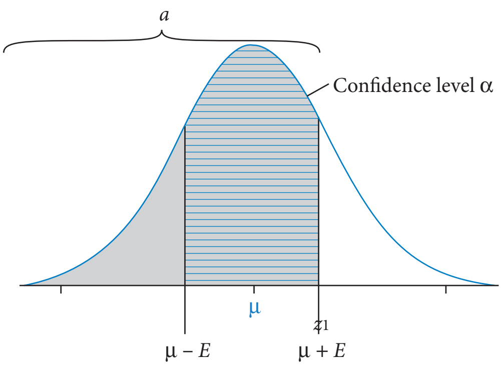
For both CAS calculators, you can just use normCdf(−1.5, 1.5) and the standard normal distribution with $\mu = 0$ and $\sigma = 1$ will be assumed.

You need to use the inverse normal distribution function to work out a margin of error for a given confidence level. The inverse function finds the value z_1 so the given proportion of values lies below z_1. In symbols, they find the value z_1 such that $P(-\infty \le z \le z_1) = a$, where a is the given area. You can see this on the diagram below.



For the confidence level a, you want the margin of error E such that $P(\mu - E \le \mu + E) = \alpha$.

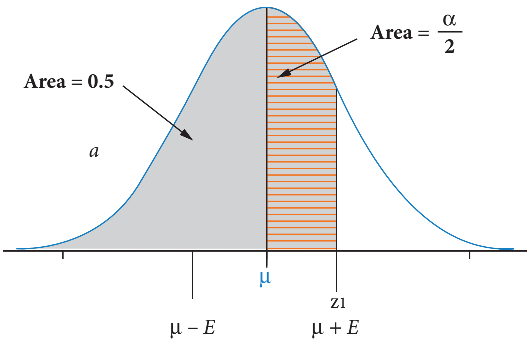
If you choose $z_1 = \mu + E$, then you get the following diagram.



You can divide the area differently to give the diagram below.

Half of a normal distribution is below the mean, so $a = 0.5 + \frac{\alpha}{2}$.

In Chapter 8, you saw that the value t_a for which $P(-\infty < t_a) = a$ is a **quantile**. For the proportion a of the standard normal variable Z, the quantile is written as z_a.



Example 5

CAS Find the margin of error needed to give a confidence level of 90% for a standard normal variable.

Solution

Find the area above the mean.

$$\text{Area above the mean} = \frac{0.9}{2} = 0.45$$

Find the area required for the inverse function.

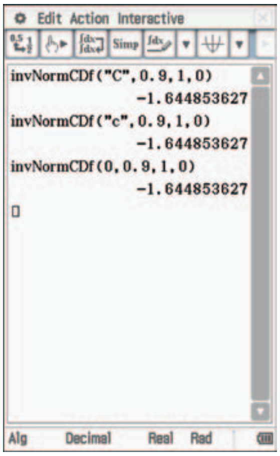
$$\text{Inverse function area } a = 0.5 + 0.45 = 0.95$$

TI-Nspire CAS

Use a calculator page.
Press menu, 6: Statistics, 5: Distributions and 3: Inverse Normal to find the quantile $z_{0.95}$; or type invNormCdf(0.95, 0, 1).

ClassPad

Use the Main menu.
Tap Action, Distribution/Inv. Dist, Inverse and invNormCDf.
Enter 0 or "C" or "c" (for central), 0.9 (confidence limit), 1 (s.d.) and 0 (mean).
Only the lower boundary is given. The confidence level of 0.9 is approximately contained in $-1.64 \le Z \le 1.64$.



Write the answer.

The margin of error is about 1.64.

Margins of error for standard normal variables

In the TI, you can omit the mean and standard deviation and $\mu = 0$ and $\sigma = 1$ will be assumed.

Example 6

Find the confidence interval needed to give a confidence level of 80% for a standard normal variable.

Solution

Find the area above the mean.

$$\text{Area above the mean} = \frac{0.8}{2} = 0.4$$

Find the area required for the inverse function.

$$\text{Inverse function area } a = 0.5 + 0.4 = 0.9$$

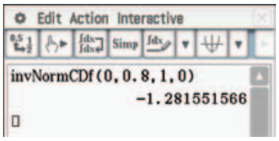
9780170250290

TI-Nspire CAS

Use a calculator page.
Press [menu], 6: Statistics, 5: Distributions and 3: Inverse Normal to find the quantile $z_{0.9}$; or type invNorm(0.9).

ClassPad

Use the Main menu.
Tap Action, Distribution/Inv. Dist, Inverse and invNormCDf.
Enter 0 (central), 0.8, 1 and 0.



Write the margin of error. $E \approx 1.28$

Write the answer. The confidence interval is $-1.28 < z < 1.28$.

It doesn't matter whether you write the confidence interval as $-1.28 < z < 1.28$ or as $-1.28 \leq z \leq 1.28$, because the probability of getting the exact boundary is 0 for a continuous distribution.

EXERCISE 10.02 Confidence levels and margin of error

Concepts and techniques

1 Example 4 CAS Find the confidence levels corresponding to the following margins of error in standard normal variables.

a $E = 2.5$　b $E = 3$　c a margin error of 1　d $E = 2.7$

2 Example 5 CAS Find the margins of error needed in standard normal variables to give the following confidence levels.

a $\alpha = 0.98$　b 85% confidence　c 99%　d $\alpha = 0.75$

3 Example 6 CAS Find the confidence intervals in standard normal variables needed to give the following confidence levels.

a $\alpha = 0.9$　b $\alpha = 0.995$　c 95% confidence　d 99% confidence

4 CAS Find the confidence levels corresponding to the following confidence intervals in standard normal variables.

a $-2.5 \leq z \leq 2.5$　b $-1.5 \leq z \leq 1.5$　c $-1 \leq z \leq 1$　d $-3 \leq z \leq 3$

Reasoning and communication

5 Can you have a small margin of error and a high confidence level for a standard normal variable? Explain your answer.

9780170250290

10.03 CONFIDENCE INTERVALS AND CONFIDENCE LEVELS FOR SAMPLE PROPORTIONS

When you use a sample to work out the probability of something, how do you know whether the results are reasonable? Intuitively, a large sample would give better results than a small sample. You can use a confidence interval to give a reasonable range for the answer.

Example 7

CAS 12 students from a sample of 30 Year 12 students said they liked the 'Pirates of the Caribbean' movies.

a Estimate the probability of Year 12 students liking these movies.

b Estimate the standard deviation of the sampling distribution.

c Estimate the 90% confidence interval for the probability of liking these movies.

d What does the confidence interval mean in this case?

Solution

a Use the sample proportion.

$p \approx \hat{p} = \frac{12}{30} = 0.4$

Write the answer.

The probability is about 0.4.

b Write the formula for the standard deviation.

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

Substitute $\hat{p}$ and n.

$= \sqrt{\frac{0.4(1-0.4)}{30}}$

$= 0.089\,44\ldots$

Write the answer.

The standard deviation is about 0.089.

c Find the area above the mean.

Area above the mean $= \frac{0.9}{2}$

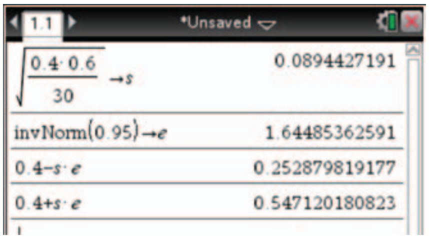
$= 0.45$

Find the area required for the inverse function.

Inverse function area $a = 0.5 + 0.45 = 0.95$

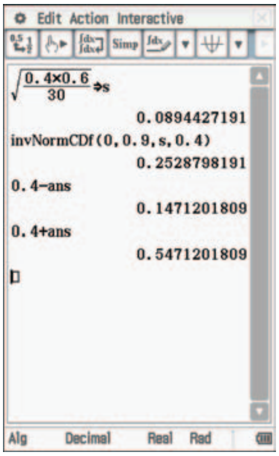
TI-Nspire CAS

Use a calculator page to find the quantile $z_{0.95}$ for the standard normal distribution.
Type invNorm(0.95)



ClassPad

Find the lower bound for 'central' values.
invNormCDf(0, 0.9, s, 0.4) gives about 0.25.
Find the difference from the mean.
0.4 – ans (0.25) ≈ 0.15
Add this value (ans) to the mean.
The upper bound ≈ 0.4 + 0.15 = 0.55



Work out the margin of error for $\sigma \approx 0.089$.

$E = 0.089\,44\ldots \times 1.64\ldots$
$= 0.1471\ldots$

Find the lower boundary of the interval.

Lower boundary = 0.4 – 0.1471…
≈ 0.25

Find the upper boundary.

Upper boundary = 0.4 + 0.1471…
≈ 0.55

Write the answer.

The confidence interval is about (0.25, 0.55).

d 90% of the values of the sampling distribution lie in the 90% confidence interval .

The probability of a Year 12 student liking these movies has a 90% chance of lying in the interval from 0.25 to 0.55.

You can use the same method as shown in Example **7** to find any confidence interval for a sample proportion.

For the confidence level α, you need the quantile z_a for $a = 0.5 + \frac{\alpha}{2}$.

The estimated standard deviation for the sample proportion $SD(\hat{p}) \approx \sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$, so the margin of error is about $z_a \times \sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$. You can use this to find the confidence level.

IMPORTANT

The area for the **standard normal quantile** z_a of the z-score for the confidence level α is given by $a = 0.5 + \frac{\alpha}{2}$.

The corresponding margin of error for the sample proportion $\hat{p}$ is given by $E \approx z_a\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$.

and the confidence interval is approximately $\left(\hat{p} - z_a\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}, \hat{p} + z_a\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}\right)$.

The area a is often omitted in the formulas for the margin of error and confidence interval.

Example **8** shows the use of the formula.

Example 8

CAS A sample of 80 has a sample proportion of 0.6. Find each of the following.

a The margin of error for a confidence level of 80% for the corresponding probability.

b The 80% confidence interval for the probability.

Solution

a Find the area of the required quantile.

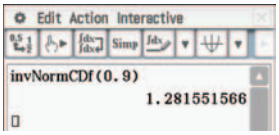
$a = 0.5 + 0.8 \div 2$
$= 0.9$

TI-Nspire CAS

Use the calculator page to find the quantile.

ClassPad

Find the main menu to find the quantile.



Write the margin of error formula.

$E \approx z\sqrt{\dfrac{\hat{p}(1-\hat{p})}{n}}$

Substitute in the values.

$= 1.28\ldots \times \sqrt{\dfrac{0.6(1-0.6)}{80}}$

Calculate the value.

$= 0.0701\ldots$

Write the answer.

The margin of error is about 0.07.

b Find the confidence interval.

Interval = (0.6 – 0.07, 0.6 + 0.07)

Write the answer.

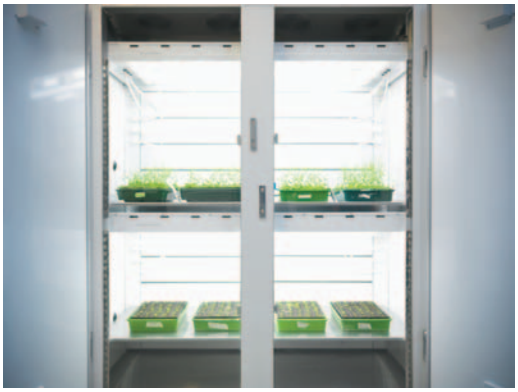
The 80% confidence interval is (0.53, 0.67).

It doesn't matter whether you write the confidence interval as the open interval (0.53, 0.67) or the closed interval [0.53, 0.67] because it is continuous.

9780170250290

Example 9

CAS As a test of the germination rate, 300 carrot seeds were moistened and placed in an incubator. When they were checked 5 days later, 250 were found to have germinated. Estimate the 95% confidence interval for the germination rate.



Science Photo Library/Monty Rakusen/Cultura

Solution

Estimate the germination rate. $p \approx \hat{p} = \frac{250}{300} = 0.833\ldots$

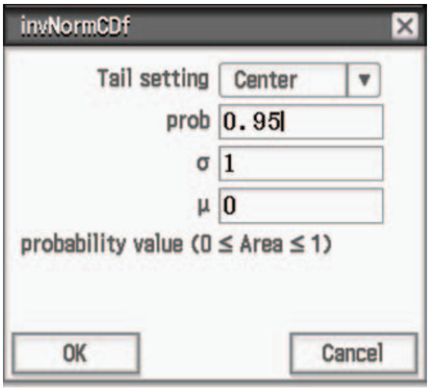
Find the area of the required quantile. $a = 0.5 + 0.95 \div 2$
$= 0.975$

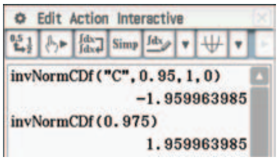
TI-Nspire CAS

Use the calculator page to find the quantile.

ClassPad

Use the main menu to find the quantile. Tap Interactive, Distribution/Inv. Dist, Inverse and invNormCDf. Tail setting is Center and prob (area) is 0.95.
The answer is the lower bound, −1.95...
Since the mean is 0, the upper bound is 1.95...
Alternatively, use a left bound with an area of 0.975.





Write the margin of error formula.	$E \approx z\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$
Substitute in the values.	$= 1.95\ldots \times \sqrt{\frac{0.833...(1-0.833...)}{300}}$
Calculate the value.	≈ 0.0422
Work out the lower boundary.	$0.8333 - 0.0422 = 0.7911$
Work out the upper boundary.	$0.8333 + 0.0422 = 0.8744$
Write the answer.	The 95% confidence interval for the germination rate is about (0.791, 0.874).

You could work out the error margin directly on your CAS calculator by including the standard deviation of the sample proportion in the inverse normal function, but it is probably safer to do it as shown in the examples.

WS Sample proportion confidence intervals

EXERCISE 10.03 Confidence intervals and confidence levels for sample proportions

Concepts and techniques

1 Example 7 CAS 35 out of 50 people crossing the intersection of Collins and Swanston Streets in July were wearing dark-coloured clothing.

a Estimate the probability of people wearing dark clothing in the city in winter.

b Estimate the standard deviation of the sampling distribution.

c Estimate the 95% confidence interval for the probability of wearing dark clothing in the city.

2 Example 8 CAS Find the approximate margin of error for each of the following confidence levels for the stated values of $\hat{p}$ and n.

a $\hat{p} = 0.8, n = 40, \alpha = 90$

b $\hat{p} = 0.5, n = 60, \alpha = 95$

c $\hat{p} = 0.45, n = 140, \alpha = 99$

d $\hat{p} = 0.65, n = 100, \alpha = 80$

3 CAS Find the confidence intervals for each of the stated values of $\hat{p}$, n and a.

a $\hat{p} = 0.1, n = 60, \alpha = 85$

b $\hat{p} = 0.75, n = 40, \alpha = 95$

c $\hat{p} = 0.32, n = 55, \alpha = 98$

d $\hat{p} = 0.61, n = 70, \alpha = 90$

Reasoning and communication

4 Example 9 CAS The songs on popular radio stations in 'drive-time' were checked over several weeks, and from a total of 140 songs, 125 were of less than 3 minutes in duration. Estimate the 90% confidence interval for songs on 'drive-time' radio being less than 3 minutes long.

5 CAS A survey of 178 Victorian students found that 87 travelled to school by car. Estimate the 95% confidence interval for the probability of Victorian students travelling to school by car.

9780170250290

6 CAS A flour mill produced 2 kg packets of flour. From 100 packets checked at the end of the run, 6 were found to be underweight. What is the 95% confidence interval for the proportion of underweight packets?

7 CAS A survey of 1200 voters commissioned by a political party found that 48.5% on a two-party preferred basis intended to vote for the party. What is the 90% margin of error for this survey?

8 CAS Estimate the 90% error margins for $n = 100$ for $\hat{p} = 0.1, 0.3, 0.5, 0.7$ and 0.9 and comment on the differences.

10.04 VARIATION OF CONFIDENCE INTERVALS

You can use a CAS calculator simulation to check whether or not the population probability for a sample proportion is within a confidence interval.

Example 10

CAS A property has a probability of occurrence of $p = 0.6$.

a Simulate a sample of 40 items and estimate the 80% error margin from the sample proportion.

b Repeat the simulation and calculation another 5 times.

c Comment on the position of p in the confidence intervals.

Solution

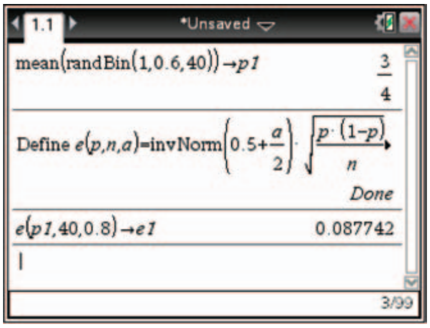
TI-Nspire CAS

a Use the calculator page.
Generate 40 Bernoulli trials with $p = 0.6$ using randBin(1,0.6,40) and find the sample proportion $p1$ using mean(randBin(1,0.6,40)).

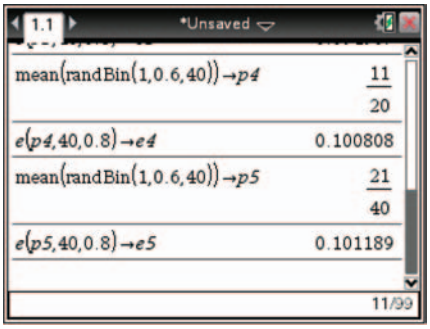
Define $e(p,n,a) =$

$\text{invNorm}\left(0.5+\frac{a}{2}\right)\times\sqrt{\frac{p\times(1-p)}{n}}$.

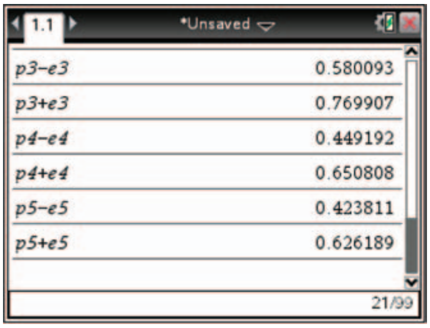
Find the 80% sample error $e1$ using $e(p1,40,0.8)$.



b Generate 40 new Bernoulli trials and find the sample proportion $p2$. Find the new 80% sample error $e2$.
Do this another 3 times.



c Find the confidence intervals using ($p1-e1$, $p1+e1$) and so on for your calculator.



Write the results.

The confidence intervals were about (0.66, 0.84), (0.55, 0.75), (0.58, 0.77), (0.45, 0.65) and (0.42, 0.63).

Comment on the positions of p.

$p = 0.6$ was in all except one of the confidence intervals.

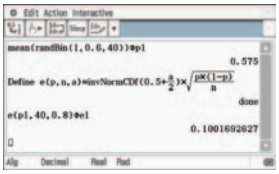
ClassPad

Use the main menu.

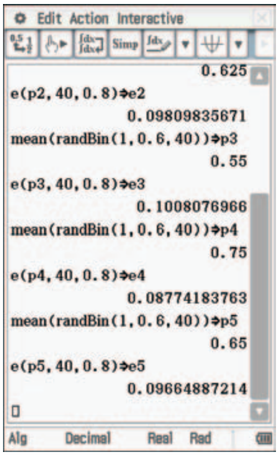
a You can generate 40 Bernoulli trials with $p = 0.6$ using randBin(1,0.6,40) and find the sample proportion $p1$ using mean(randBin(1,0.6,40)).
Define $e(p,n,a) =$

$$\text{invNormCDf}\,(0.5 + \frac{a}{2}) \times \sqrt{\frac{p \times (1-p)}{n}}$$

The multiplication signs are essential!
Find the 80% sample error $e1$ using $e(p1,40,0.8)$.

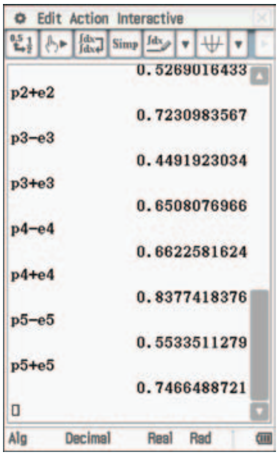


b Generate 40 new Bernoulli trials and find the sample proportion $p2$. Find the new 80% sample error $e2$.
Do this another 3 times.



9780170250290

c Find the confidence intervals using $p1-e1$, $p1+e1$, and so on for your calculator.



Write the results.

The confidence intervals were about (0.58, 0.78), (0.53, 0.72), (0.45, 0.65), (0.66, 0.84) and (0.55, 0.74)

Comment on the positions of p.

$p = 0.6$ was in all of the confidence intervals.

Once you understand what you are doing, you can use the spreadsheet facility to do this. The spreadsheet method is much quicker, but you need to be sure you know how and why it works. This can be done using CAS calculators, or a computer spreadsheet such as Excel.

TI-Nspire CAS

First calculate $a = 0.5 + (0.8 \div 2) = 0.9$

Enter the following formulas for $p1$ and $e1$ into A1 and B1, using $p = 0.6$, $a = 0.9$ and sample size $n = 40$.

A1: =mean(randBin(1, 0.6, 40))

B1: =invNorm(0.9)×(A1×(1–A1)/40)^0.5

Since A1 is storing $p1$ and B1 is storing $e1$, make C1 and D1 the confidence limits by entering

C1: = A1–B1 $(p1 - e1)$

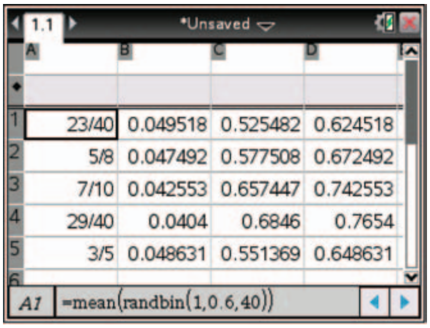
D1: =A1+B1 $(p1 + e1)$

In this case, the confidence limits are (0.61, 0.79).

Fill each column down (Edit, Fill, Fill Range) and you can generate as many sets of results as you like.

A new set can be generated by filling column A down to row 6. The other columns will adjust automatically, but row 1 will be unchanged. Take rows 2 to 6 as the next set of five results. This can be repeated as often as required.

Alternatively, all columns can be filled down to row 10 for the second set of results.



Write the answer.

The confidence limits are (0.53, 0.62), (0.58, 0.67), (0.66, 0.74), (0.68, 0.76) and (0.55, 0.65).

Comment on the positions of p.

$p = 0.6$ was in all except two of the confidence intervals.

ClassPad

First calculate $a = 0.5 + (0.8 \div 2) = 0.9$

Enter the following formulas for $p1$ and $e1$ into A1 and B1, using $p = 0.6$, $a = 0.9$ and sample size $n = 40$.

A1: =mean(randBin(1, 0.6, 40))

B1: =invNormCDf(0.9)×(A1×(1–A1)/40)^0.5

Since A1 is storing $p1$ and B1 is storing $e1$, make C1 and D1 the confidence limits by entering

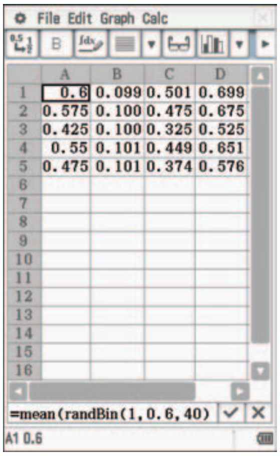
C1: = A1–B1 $(p1 - e1)$

D1: =A1+B1 $(p1 + e1)$

In this case, the confidence limits are (0.61, 0.79).

Fill each column down (Edit, Fill, Fill Range) and you can generate as many sets of results as you like.

A new set can be generated by filling column A down again. New data will appear in A1 to A5 and the other columns will adjust automatically. Alternatively, all columns can be filled down to row 10 for the second set of results.



Note that the columns have been narrowed to display all the data on one screen. This is not necessary.

Write the answer.

The confidence limits are (0.61, 0.79), (0.53, 0.72), (0.61, 0.79), (0.40, 0.60) and (0.37, 0.58).

Comment on the positions of p.

$p = 0.6$ was in all except one of the confidence intervals.

Your teacher will probably want everybody to work through Example **10** separately and compare results. You might find that all of your confidence intervals include the probability of 0.6, or you could find that 2 or 3 of the 5 do not include the probability. You are dealing with random variables, so should expect your results to vary. Across the whole class, you are likely to find that on average, about 20% of the class confidence intervals do not include the probability of 0.6.

For sample proportions, a margin error of 90% for a confidence interval means that if you take multiple samples, then you can expect to find the population probability within the approximate confidence interval calculated from the sample proportions about 90% of the time.

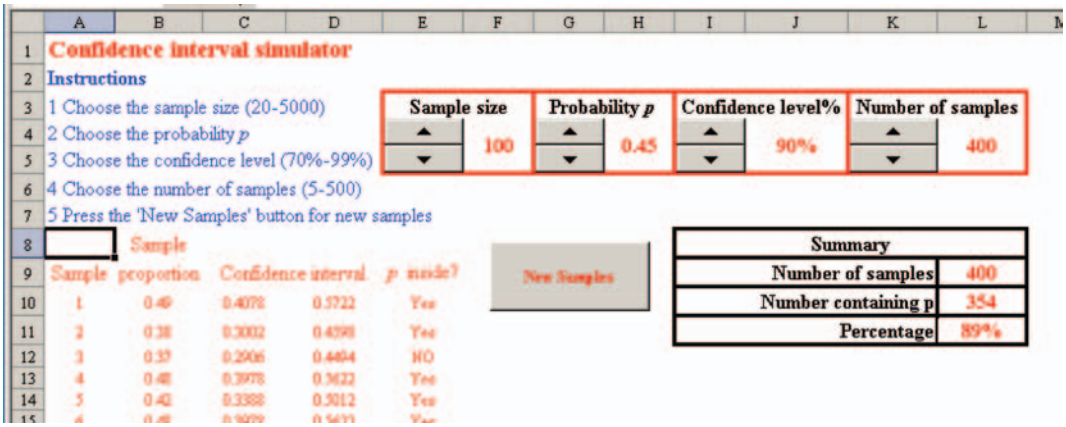
INVESTIGATION Dice simulation of confidence intervals

In this investigation you will pool results from the whole class to simulate sample proportions and confidence intervals. The probability of the event for the simulation is $p = \frac{1}{3}$ and $n = 20$.

- Roll a normal die and note when it lands with 5 or 6 uppermost
- Repeat the roll another 19 times and find the sample proportion of high numbers (5 or 6)
- Calculate the approximate 90% confidence interval for your 20 rolls
- Do another 20 rolls and find the new sample proportion and 90% confidence interval
- Do this whole procedure another 3 times
- How many of your 5 simulations of the confidence interval contained $p = \frac{1}{3}$?
- Put the results of the whole class together and compare the total number of simulations and the number of times that the confidence intervals contained the probability
- Is this what you would have expected?

The Excel spreadsheet program 'confidence interval simulator' allows you to simulate many samples and calculations of $\hat{p}$ and confidence intervals for various values of p to find the proportion of values of p that lie in different confidence intervals.

You teacher will tell you what simulations you should perform.



Confidence interval simulator

EXERCISE 10.04 Variation of confidence intervals

Reasoning and communication

Your teacher will tell you which questions each person in the class should do and probably want you to pool results with groups who have done the same questions. You may have to give a report to the class as a whole.

1 Example 10 CAS A property has a probability of occurrence of $p = 0.3$.
   a Simulate a sample of 20 items and estimate the 75% error margin from the sample proportion.
   b Repeat the simulation and calculation another 9 times.
   c Comment on the position of p in the confidence intervals

2 Example 10 CAS A characteristic has a probability of occurrence of $p = 0.8$.
   a Simulate a sample of 20 items and estimate the 90% error margin from the sample proportion.
   b Repeat the simulation and calculation another 9 times.
   c Comment on the position of p in the confidence intervals.

3 Example 10 CAS A property has a probability of occurrence of $p = 0.4$.
   a Simulate a sample of 20 items and estimate the 95% error margin from the sample proportion.
   b Repeat the simulation and calculation another 9 times.
   c Comment on the position of p in the confidence intervals.

4 Example 10 CAS A characteristic has a probability of occurrence of $p = 0.7$.
   a Simulate a sample of 20 items and estimate the 85% error margin from the sample proportion.
   b Repeat the simulation and calculation another 9 times.
   c Comment on the position of p in the confidence intervals.

5 Example 10 CAS A property has a probability of occurrence of $p = 0.2$.
   a Simulate a sample of 20 items and estimate the 80% error margin from the sample proportion.
   b Repeat the simulation and calculation another 9 times.
   c Comment on the position of p in the confidence intervals

10.05 APPLICATIONS OF CONFIDENCE INTERVALS

The only way to be *certain* about a parameter is by using a census. Any sample has a level of confidence, but it is never 100%. When you design a survey, you need to decide what level of confidence is acceptable. Since larger surveys cost more to conduct, the level of confidence will be traded off with the cost of the survey. If you have some idea of the likely value of the parameter, then you can calculate the size of the survey needed for any particular level of confidence.

IMPORTANT

The table shows the most commonly used levels of confidence and corresponding z-scores for the standard normal distribution.

Level of confidence α	90%	95%	98%	99%
***z*-score**	1.645	1.960	2.326	2.576

9780170250290

Example 11

The probability of an Australian Year 12 student driving to school is thought to be about 6%. How large a sample of students should be used to establish the probability to a 95% confidence level within 0.5% of the true probability?

Solution

Write the formula for E for a sample proportion. $E \approx z\sqrt{\dfrac{\hat{p}(1-\hat{p})}{n}}$

Write the known values. $E = 0.005,\ z = 1.96,\ \hat{p} \approx 0.06$

Substitute in the formula. $0.005 = 1.96 \times \sqrt{\dfrac{0.06(1-0.06)}{n}}$

Square both sides. $(0.005)^2 = (1.96)^2 \times \dfrac{0.06 \times 0.94}{n}$

Solve for n.
$$n = \frac{1.96^2 \times 0.06 \times 0.94}{0.005^2} \approx 8667$$

Write the answer. About 9000 Year 12 students would need to be asked for a margin error of 0.5% to a level of confidence of 95%.

It would clearly be impractical to perform a survey to the level of accuracy requested in Example **11**. The example shows that the most critical influence on the sample size is the margin of error. To decrease the size of the sample, a larger margin of error must be accepted.

Example 12

It is thought that about 60% of Year 12 students obtain their driver's licence before they complete Year 12. How large a sample would be needed to establish this to within a margin of error of 5% at the 90% confidence level?

Solution

Write the formula for E for a sample proportion. $E \approx z\sqrt{\dfrac{\hat{p}(1-\hat{p})}{n}}$

Write the known values. $E = 0.05,\ z = 1.645,\ \hat{p} \approx 0.6$

Substitute in the formula. $0.05 = 1.645 \times \sqrt{\dfrac{0.6(1-0.6)}{n}}$

Square both sides. $(0.05)^2 = (1.645)^2 \times \dfrac{0.6 \times 0.4}{n}$

Solve for n.
$$n = \frac{1.645^2 \times 0.6 \times 0.4}{0.05^2} \approx 260$$

Write the answer. About 260 Year 12 students at the end of the year would need to be asked to get this accuracy.

What should you do if you have no real idea of the likely value of $\hat{p}$? In this case, you use the value that produces the largest margin of error to make sure that whatever value found, the margin of error will be within the accuracy required.

WS Sample sizes

This is easily shown using calculus as demonstrated below.

If everything else is the same, then the size of the error margin is determined by $\sqrt{\hat{p}(1-\hat{p})}$, as this is the only part of the equation for E that changes. The largest value of the square root is produced by the largest value of $f(\hat{p}) = \hat{p}(1 - \hat{p}) = \hat{p} - \hat{p}^2$.

IMPORTANT

For the same sample size and z-score, the largest margin of error is produced by $\hat{p} = 0.5$.

Now $f'(\hat{p}) = 1 - 2\hat{p}$ and $f''(\hat{p}) = -2$, so there is a maximum at $\left(\frac{1}{2}, \frac{1}{4}\right)$.

You may find it easier in problems of this kind to transform the formula before substituting values.

Example 13

The transport department wants to determine the proportion of cars on the road that have faulty brakes to within 10% of the correct value to a level of confidence of 98% to work out whether they should request a police crackdown. How many vehicles should have their brakes tested at a random spot check?

Solution

What value of $\hat{p}$ should be used?	Use $\hat{p} = 0.5$ as the worst case.
Write the formula for E for a sample proportion.	$E = z\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$
Square both sides.	$E^2 = z^2 \times \frac{\hat{p}(1-\hat{p})}{n}$
Transform to get n.	$n = \frac{z^2\hat{p}(1-\hat{p})}{E^2}$
Write the values.	$z = 2.326$, $\hat{p} = 0.5$, $E = 0.1$
Substitute in the values.	$n = \frac{2.326^2 \times 0.5 \times 0.5}{0.1^2}$ ≈ 135
Write the answer.	Testing the brakes of 135 vehicles at random should be enough to determine the proportion of cars with faulty brakes to within 10% at a level of confidence of 98%.

Example **13** shows that you can obtain a small size by accepting a low margin of error. In that particular case, it is unlikely that the transport department would accept such a low margin of error.

If you want to use levels of confidence other than the common ones listed in the table, you will need to use the inverse normal function to find the value of z needed.

EXERCISE 10.05 Applications of confidence intervals

Concepts and techniques

1 Examples 11, 12 What sample size would be needed to establish an error margin of 0.03 at a 98% confidence level for the probability of a characteristic occurring in a proportion of about 0.3 in a population?

2 Example 13 The probability of a particular property in a population is unknown. What sample size should be used to establish the probability with an error margin of 5% at a confidence level of 90%?

Reasoning and communication

3 It is generally assumed that the chances of a baby being a girl or boy are 50-50. How many births would need to be checked to establish the true proportion in Australia to within 0.1% at a confidence level of 95%?

4 Suppose you were organising the venue for a Year 12 formal. How many students should you survey to establish the preferred venue from a list of 3 at an error margin of 10% and confidence level of 90%?

5 The tourist information centre counter staff in Federation square said that about 2 in every 5 of the people coming in asked about local attractions. How many enquiries would need to be noted to establish the proportion to within 2% at a level of confidence of 99%?

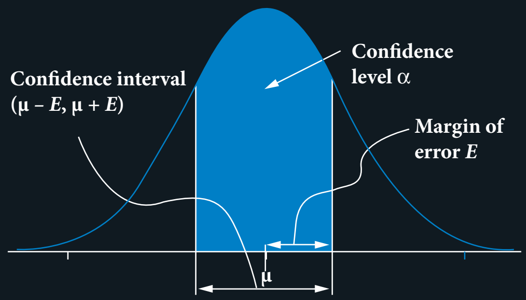
6 An advertising company wanted to conduct a small survey of consumers to establish a baseline for the proportion who were aware of a particular brand of ice-cream before a marketing campaign. How many consumers should they target to get a result accurate to within 3% at a confidence level of 95%?

7 CAS A survey of 50 people claimed that 71% ± 5% of the population could roll their tongues. What is the confidence level of this result?

8 CAS A poll of voting intentions among 300 voters claimed that the vote for a new political party would be 15% ± 3%. What level of confidence has been used in the survey?

10 CHAPTER SUMMARY
CONFIDENCE INTERVALS

- A value obtained from a population is called a **parameter**, while a value obtained from a sample is called a **statistic**.
- A **point estimate** of a parameter is a single value obtained from a sample. The statistic used is called an **estimator** of the parameter.
- An **interval estimate** of a parameter is an interval that is likely to include the value of the parameter.
- The **estimated variance** and **estimated standard deviation** of a sample proportion $\hat{p}$ with sample size n are $Var(\hat{p}) \approx \frac{\hat{p}\hat{q}}{n} = \frac{\hat{p}(1-\hat{p})}{n}$ and $SD(\hat{p}) \approx \sqrt{\frac{\hat{p}\hat{q}}{n}} = \sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$.
- For a **confidence interval** symmetric around the mean in a statistical distribution, the **confidence level** α is the proportion of values that lie within the interval and the **margin of error** E is the distance of the ends of the interval from the mean.



- For the sample proportion $\hat{p}$, the margin of error is given by $E \approx z\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$ and the confidence interval is approximately $\left(\hat{p} - z\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}, \hat{p} + z\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}\right)$, where z is the **standard normal quantile** for the value $0.5 + \frac{\alpha}{2}$, so the proportion of the standard normal distribution below z is $P(-\infty < x < z) = 0.5 + \frac{\alpha}{2}$.
- The most commonly used levels of confidence and z values are:

Level of confidence α	90%	95%	98%	99%
***z*-score**	1.645	1.960	2.326	2.576

- The **largest margin of error** for a sample proportion occurs when $\hat{p} = 0.5$, assuming that the sample size n and level of confidence are set.

9780170250290

CHAPTER REVIEW CONFIDENCE INTERVALS 10

Multiple choice

1 Example 1 The following estimations were contained in a document about human development.

P 2-3 Months: Can lift head, follows objects with eyes, opens and closes hands

Q 6 Months: Enjoys 'peekaboo', rolls over, remains sitting without support, recognises others

R 10 ± 2 months: Moves around by crawling or scooting, babbles, first words, eats finger food

S 18 months: Walks, talks in simple words or phrases, shy with strangers, becoming independent

Which of **P**, **Q**, **R** and **S** are interval estimates?

A P only B P and Q C P and R
D Q and S E All of them

2 Example 2 From 40 Maths Methods students, 16 were also studying Chemistry. The estimated probability and variance of the sampling distribution of students who do Maths Methods and Chemistry are:

A 0.4 and 0.06 B 0.6 and 0.04 C 0.4 and about 0.24
D 0.16 and about 0.024 E 0.4 and 0.006

3 Example 4 A margin of error of 2 in a standard normal variable corresponds to a confidence level of about:

A 85% B 90% C 95%
D 98% E 99%

4 Examples 5, 6 What is the confidence interval of a standard normal variable for a level of confidence of 85%?

A (−1.44, 1.44) B (−1.04, 1.04) C (−1.96, 1.96)
D (−0.45, 0.45) E (−1.645, 1.645)

5 Examples 8, 9 30% of a group of 80 Year 12 students were strong swimmers. Estimate the 90% margin of error for the probability of a Year 12 student being a strong swimmer.

A 0.0026 B 0.029 C 0.051
D 0.084 E 0.17

6 Examples 11, 12 In Australia, the probability of someone voting number 1 for one of the major parties is generally about 0.4. How large a sample would be needed to determine the level of support for a major party to within 3% at a 99% confidence level?

A About 50 B About 260 C About 1100
D About 1800 E About 2100

7 Example 13 What sample size should be used to determine the proportion of students who swim at least once a week to an accuracy of 10% at a 95% confidence level?

A About 50 B About 100 C About 200
D About 400 E About 800

Short answer

8 Examples 2, 3 A survey of 90 Australians found that 62 owned their own homes, with some having mortgages. Estimate the probability of Australians owning their own home to within 1.5 standard deviations.

9 Example 4 CAS Find the confidence level that corresponds to an error margin of 2.5 standard deviations of a standard normal variable.

10 Examples 5, 6 CAS What margin of error corresponds to a confidence level of 92% for a standard normal variable?

11 Example 7 CAS 11 out of 40 Year 12 students said that they had part-time jobs.
- a What is the sample proportion for Year 12 students having part-time jobs?
- b Estimate the standard deviation of the sampling distribution.
- c Estimate the 95% confidence interval for the probability of Year 12 students having part-time jobs.
- d What does the 95% confidence interval mean in this case?

12 Examples 8, 9 CAS A sample of 120 items has a sample proportion of 0.75.
- a Estimate the 90% margin of error for the corresponding probability.
- b What is the 90% confidence interval for this probability?

13 Example 10 CAS a Simulate a sample of 30 for a property that has a probability of occurrence of 0.67.
- b Find the 95% confidence interval for the sample proportion.
- c Comment on the position of the probability within the confidence interval.

14 Examples 11, 12 CAS What sample size should be used to establish a 98% confidence interval with a margin of error of 5% for a characteristic that occurs about 70% of the time?

15 Example 13 CAS A survey is to be conducted to determine the proportion of a population having a certain property.
- a What should be assumed about the occurrence to find the sample size required?
- b What sample size is needed to establish the occurrence within 8% at the 90% confidence level?

Alamy/JS Callahan/tropicalpix

9780170250290

Application

16 The estimated probability and 95% margin of error from a survey were given as 0.31 and 0.058. What was the sample size?

17 A survey of 500 Australians renting houses in outer suburbs found that 32% of their income was paid in rent.

a Estimate the standard deviation of the sampling distribution.

b What z-score should be used for an 85% confidence interval?

c Estimate the 85% confidence interval for the proportion of income used for rent in outer suburbs in Australia.

18 A surfer who regularly surfs at Bell's beach says that you get the best rides from about 1 in 5 waves. Another surfer standing at the lookout watching the waves sees that they are coming about every 11 seconds.

a Assuming that the first surfer is close to being correct, how many waves would you have to watch to determine the proportion of good waves to an accuracy of 10% at a 95% level of confidence?

b How long would you need to watch?

Practice quiz