

**GOAL**

Use the normal distribution to solve problems that involve confidence intervals.

**LEARN ABOUT the Math**

A telephone survey of 600 randomly selected people was conducted in an urban area. The survey determined that 76% of people, from 18 to 34 years of age, have a social networking account.

The results are accurate within plus or minus 4 percent points, 19 times out of 20.



**?** How can this result be interpreted, if the total population of 18- to 34-year-olds is 92 500?

**EXAMPLE 1** Analyzing and applying survey results

Calculate the range of people that have a social networking account, and determine the certainty of the results.

**Danica's Solution**

The **margin of error** for the data is  $\pm 4\%$ , so the **confidence interval** is  $76\% \pm 4\%$ , which is from  $(76 - 4)\%$  or 72% to  $(76 + 4)\%$  or 80%.

The **confidence level** of the survey is 95%. The probability of error for this result is 5%. If the survey were conducted 100 times, then 95 times out of 100, the percent of people in the population with a social networking account would be from 72% to 80%.

I interpreted the survey statement.

The results are accurate 19 times out of 20, which is 95% of the time.

**YOU WILL NEED**

- calculator
- z-score tables (pages 592 to 593)

**EXPLORE...**

- Often it is impractical to survey an entire population. For example, if a light bulb company wants to test the number of hours that a light bulb will burn before failing, it cannot test every bulb. Propose a method that the company could use to determine the longevity of its light bulbs.

**margin of error**

The possible difference between the estimate of the value you're trying to determine, as determined from a random sample, and the true value for the population; the margin of error is generally expressed as a plus or minus percent, such as  $\pm 5\%$ .

**confidence interval**

The interval in which the true value you're trying to determine is estimated to lie, with a stated degree of probability; the confidence interval may be expressed using  $\pm$  notation, such as  $54.0\% \pm 3.5\%$ , or ranging from 50.5% to 57.5%.

**confidence level**

The likelihood that the result for the "true" population lies within the range of the confidence interval; surveys and other studies usually use a confidence level of 95%, although 90% or 99% is sometimes used.

$92\,500 \times 76\% = 70\,300$   
 $92\,500 \times 4\% = 3\,700$   
The confidence interval for the population is  $70\,300 \pm 3\,700$ .

I used the confidence interval,  $76\% \pm 4\%$ , and the population, 92 500, to calculate the range of the number of people in the population who have a social networking account.

$$70\,300 - 3\,700 = 66\,600$$

$$70\,300 + 3\,700 = 74\,000$$

It can be said, with 95% confidence, that 66 600 to 74 000 people, in a population of 92 500 people from ages 18 to 34, have a social networking account.

## Reflecting

- A. Based on this survey, what is the range for 18- to 34-year-olds who do not have a social networking account?
- B. The same telephone survey was conducted by a different company, using a sample of 600 randomly selected people in both urban and rural areas. According to this survey, 76% of people, from 18 to 34 years of age, have a social networking account. (The results are accurate within plus or minus 5.3 percent points, 99 times out of 100.) How are the results of this survey different from those of the first survey? How are they the same?

## APPLY the Math

### EXAMPLE 2

### Analyzing the effect of sample size on margin of error and confidence intervals

Polling organizations in Canada frequently survey samples of the population to gauge voter preference prior to elections. People are asked:

1. “If an election were held today, which party would you vote for?”

If they say they don’t know, then they are asked:

2. “Which party are you leaning toward voting for?”

The results of three different polls taken during the first week of November, 2010, are shown on the next page. The results of each poll are considered accurate 19 times out of 20.



Polling Organization & Data	Conservative (%)	Liberal (%)	NDP (%)	Bloc Quebecois (%)	Green Party (%)	Undecided (%)
<b>Ekos</b>	29	29	19	9	11	12.6
sample size, 1815 margin of error, $\pm 2.3\%$						
<b>Nanos</b>	37	32	15	11	5	19.2
sample size, 844 margin of error, $\pm 3.4\%$						
<b>Ipsos</b>	35	29	12	11	12	n.a.
sample size, 1000 margin of error, $\pm 3.1\%$						

source: <http://www.sfu.ca/~aheard/elections/polls.html>

How does the sample size used in a poll affect:

- the margin of error in the reported results?
- the confidence interval in the reported results?

### Martin's Solution

a)  $\frac{19}{20} = 95\%$

The confidence level of each poll is 95%.

	Nanos	Ipsos	Ekos
<b>sample size</b>	844	1000	1815
<b>margin of error</b>	$\pm 3.4\%$	$\pm 3.1\%$	$\pm 2.3\%$

If polls are assessed using the same confidence level, when the sample size increases, the margin of error decreases.

A larger sample size results in the possibility of a poll that more accurately represents the population.

In this case, the confidence level used by each polling organization is the same. This enables me to compare the effect that sample size has on the margin of error.

I created a table to compare the polls. I arranged the polls in increasing order of sample size, then looked for a trend in the margin of error.

My observation makes sense because a larger sample should be a better indicator of how the population might vote.



b) Let  $n$  represent the number of people polled.

**Nanos**

$$n = 844$$

$$37\% \pm 3.4\% \text{ or } 33.6\% \text{ to } 40.4\%$$

**Ipsos**

$$n = 1000$$

$$35\% \pm 3.1\% \text{ or } 31.9\% \text{ to } 38.1\%$$

**Ekos**

$$n = 1815$$

$$29\% \pm 2.3\% \text{ or } 26.7\% \text{ to } 31.3\%$$

The Nanos poll predicts that 33.6% to 40.4% of the population will vote for a Conservative. That is a range of 6.8%.

The Ipsos poll predicts that 31.9% to 38.1% of the population will vote for a Conservative. That is a range of 6.2%.

The Ekos poll predicts that 26.7% to 31.3% of the population will vote for a Conservative. That is a range of 4.6%.

If polls are conducted using the same confidence level, when the sample size increases, the range in the confidence interval decreases.

I decided to compare the confidence interval for the Conservative Party for each of the 3 different polls. I wrote the confidence interval for each poll in increasing order of sample size.

I interpreted each of the confidence intervals for these polls.

My observation makes sense because the confidence interval is determined by the margin of error. So, as the sample size increases, the margin of error decreases and in turn the range of the confidence interval decreases.

### Your Turn

Compare the confidence intervals for the Liberal Party for each of the three polls. Do your results reflect Martin's results above? Explain.

#### EXAMPLE 3

#### Analyzing the effect of confidence levels on sample size

To meet regulation standards, baseballs must have a mass from 142.0 g to 149.0 g. A manufacturing company has set its production equipment to create baseballs that have a mean mass of 145.0 g.

To ensure that the production equipment continues to operate as expected, the quality control engineer takes a random sample of baseballs each day and measures their mass to determine the mean mass. If the mean mass of the random sample is 144.7 g to 145.3 g, then the production equipment is running correctly. If the mean mass of the sample is outside the acceptable level, the production equipment is shut down and adjusted. The quality control engineer refers to the chart shown on the next page when conducting random sampling.



Confidence Level	Sample Size Needed
99%	110
95%	65
90%	45

- What is the confidence interval and margin of error the engineer is using for quality control tests?
- Interpret the table.
- What is the relationship between confidence level and sample size?

### Geoffrey's Solution

- The confidence interval is 144.7 g to 145.3 g.

Margin of error:

$$145.3 - 145.0 \text{ or } 0.3$$

$$144.7 - 145.0 \text{ or } -0.3$$

The margin of error is  $\pm 0.3$  g.

- In order to be confident that, 99 out of 100 times, the mean mass of the sample measures from 144.7 g to 145.3 g, the engineer needs to take a random sample of 110 baseballs from the production line.
  - In order to be confident that, 95 out of 100 times, the mean mass of the sample measures from 144.7 g to 145.3 g, the engineer needs to take a random sample of 65 baseballs from the production line.
  - In order to be confident that, 90 out of 100 times, the mean mass of the sample measures from 144.7 g to 145.3 g, the engineer needs to take a random sample of 55 baseballs from the production line.
- For a constant margin of error, as the confidence level increases, the size of the sample needed to attain that confidence level increases. To have greater confidence that the baseballs meet quality standards, the engineer must use a larger sample.

----- The confidence interval is the range that the mean mass of the random sample can fall in and be acceptable.

----- I subtracted the mean from the upper and lower limits of the confidence interval to determine the margin of error.

----- I interpreted each entry in the table.

----- I observed the trend in the table.

### Your Turn

After making adjustments in equipment, the quality control engineer decided that the mean mass of baseballs must lie in the range 144.2 g to 146.4 g.

- What margin of error is being used in the new sampling process?
- What is the mean mass of a baseball that the engineer is trying to achieve?
- Will the new baseballs meet regulation standards?

**EXAMPLE 4****Analyzing statistical data to support a position**

A poll was conducted to ask voters the following question: If an election were held today, whom would you vote for? The results indicated that 53% would vote for Smith and 47% would vote for Jones. The results were stated as being accurate within 3.8 percent points, 19 times out of 20. Who will win the election?

**Kylie's Solution**

Smith would have 53% of the votes, and Jones would have only 47% of the votes. Based on these numbers, Smith should win.

I examined the mean percent of votes that each candidate would receive, based on the poll.

Percent of votes for Jones in the population:

$$47\% \pm 3.8\%$$

Confidence interval:

$$43.2\% \text{ to } 50.8\%$$

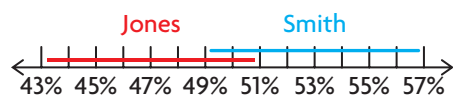
Percent of votes for Smith in the population:

$$53\% \pm 3.8\%$$

Confidence interval:

$$49.2\% \text{ to } 56.8\%$$

The margin of error is 3.8%. I used this value to determine the confidence interval for both candidates.



The two confidence intervals overlap from 49.2% to 50.8%.

I graphed the confidence intervals on a number line.

If the poll is accurate, Smith is more likely to win. However, there also is a chance that Jones will win, since the confidence intervals overlap by 1.6% of the votes.

If voters' opinions are the same on election day, Smith may receive only 49.5% of the votes and Jones could receive 50.5% of the votes.

**Your Turn**

Is it possible that Smith could receive more than 56.8% of the votes, according to this survey? Explain why or why not.

## In Summary

### Key Ideas

- It is often impractical, if not impossible, to obtain data for a complete population. Instead, random samples of the population are taken, and the mean and standard deviation of the data are determined. This information is then used to make predictions about the population.
- When data approximates a normal distribution, a confidence interval indicates the range in which the mean of any sample of data of a given size would be expected to lie, with a stated level of confidence. This confidence interval can then be used to estimate the range of the mean for the population.
- Sample size, confidence level, and population size determine the size of the confidence interval for a given confidence level.

### Need to Know

- A confidence interval is expressed as the survey or poll result, plus or minus the margin of error.
- The margin of error increases as the confidence level increases (with a constant sample size). The sample size that is needed also increases as the confidence level increases (with a constant margin of error).
- The sample size affects the margin of error. A larger sample results in a smaller margin of error, assuming that the same confidence level is required.

For example:

- A sample of 1000 is considered to be accurate to within  $\pm 3.1\%$ , 19 times out of 20.
- A sample of 2000 is considered to be accurate to within  $\pm 2.2\%$ , 19 times out of 20.
- A sample of 3000 is considered to be accurate to within  $\pm 1.8\%$ , 19 times out of 20.

## CHECK Your Understanding

In the following questions, assume that the data approximates a normal distribution.

1. A poll determined that 81% of people who live in Canada know that climate change is affecting Inuit more than the rest of Canadians. The results of the survey are considered accurate within  $\pm 3.1$  percent points, 19 times out of 20.
  - a) State the confidence level.
  - b) Determine the confidence interval.
  - c) The population of Canada was 33.5 million at the time of the survey. State the range of the number of people who knew that climate change is affecting Inuit more than the rest of Canadians.
2. A cereal company takes a random sample to check the masses of boxes of cereal. For a sample of 200 boxes, the mean mass is 542 g, with a margin of error of  $\pm 1.9$  g. The result is considered accurate 95% of the time.
  - a) State the confidence interval for the mean mass of the cereal boxes.
  - b) Three other samples of different sizes were taken using the same confidence level, as shown at left, but the margin of error for each sample was mixed up. Match the correct margin of error with each sample size.

Sample Size	Margin of Error (g)
50	1.2
100	3.9
500	2.7

## PRACTISING

3. An advertisement for a new toothpaste states that 64% of users reported better dental checkups. The results of the poll are accurate within 3.4 percent points, 9 times out of 10.
  - a) State the confidence level.
  - b) Determine the confidence interval.
  - c) If all 32 students in a mathematics class used this toothpaste, determine the range of the mean number of classmates who could expect better dental checkups.
4. In a 2006 Centre de recherche sur l'opinion publique (CROP) poll, 81% of Canadians indicated that they support bilingualism in Canada and that they want Canada to remain a bilingual country. This poll was reported accurate  $\pm 2.2\%$ , 19 times out of 20.
  - a) Interpret the poll.
  - b) Mark claims that this poll must be flawed because if the majority of Canadians felt this way, then most people would speak both French and English, but they don't. Do you agree with Mark? Justify your decision.



5. The responses to another question in the poll from question 1 were summarized as follows: 58% of people living in Canada know that the cost of living for the average Inuit is 50% higher than the cost of living for other Canadians.
- Determine the confidence interval for this question.
  - Predict the range of the mean number of people in your city or town who could have answered this question correctly.
6. Toxic materials, such as arsenic, lead, and mercury, can be released into the air if a discarded cellphone is incinerated. Toxins can be released into groundwater if a discarded cellphone ends up in a landfill. In a recent survey, 89% of those surveyed answered yes to the following question: Would you recycle your cellphone if it were convenient? The survey is considered accurate to within 4.3 percent points, 99 times out of 100.
- Determine the confidence level and the confidence interval.
  - If 23 500 000 people in Canada own cellphones, state the range of the number of people who would indicate that they would recycle their cellphone if it were convenient.
7.
  - Look in print or electronic media to find an example of a poll or survey that used a confidence level to report the results.
  - Determine the confidence interval.
  - Do you agree or disagree with any concluding statements that were made about the data from the poll or survey? Explain.
8. A company produces regulation ultimate discs. The discs have a mean mass of 175.0 g, with a standard deviation of 0.9 g. To ensure that few discs are rejected, the quality control manager must ensure that the mean mass of the discs lies in the acceptable range of 174.8 g to 175.2 g. During each shift, a random sample of discs is selected and the mass of each disc in the sample is measured. The table shown helps guide the sampling process.
- What is the confidence interval and margin of error this company is using for its quality control tests?
  - Approximately how many discs should be measured to ensure the mean mass is within  $\pm 0.2$  g, 99% of the time?
  - The manager wants to save on labour costs by using a smaller sample. She knows that any discs that do not meet the regulation standards can be sold as recreational discs. Approximately how many discs should be measured to ensure that the mean mass is within  $\pm 0.2$  g, 90% of the time?
  - Estimate the number of discs the company should measure to be confident that the mean mass of the ultimate discs lies in the acceptable range 19 times out of 20.



Confidence Level	Sample Size Needed
90%	55
95%	78
99%	135

## Math in Action

### Finding and Interpreting Data in the Media

Research and polling firms regularly conduct surveys of samples of specific populations in areas such as agribusiness, food and animal health, energy, financial services, health care, higher education, lotteries and gaming, media and entertainment, retail, technology and communications, travel and tourism, and public opinion on political issues. Many companies and organizations rely on the data that is collected to make business and marketing decisions.

- With a partner or in a small group, look in print or electronic media for examples that cite confidence intervals or confidence levels. List the important statistical information that is given.
- Do you agree or disagree with the statements about the survey results in your examples? Use your knowledge of statistics to support your reasoning.
- What could be done to change the outcome of a future survey about one of these subjects? (For example, many different approaches have been taken to curb smoking addictions in Canada.)
- How much of a change would there have to be in the survey responses to make the outcome noticeably different?

9. Use confidence intervals to interpret each of the following statements.
  - a) In a recent survey, 54% of post-secondary graduates indicated that they expected to earn at least \$100 000/year by the time they retire. The survey is considered accurate within  $\pm 4.5\%$ , 9 times in 10.
  - b) A market research firm found that among online shoppers, 63% search for online coupons or deals when they purchase something on the Internet. The survey is considered accurate within  $\pm 2.1$  percent points, 99% of the time.
  - c) A recent report indicated that Canadians spend an average of 18.1 h per week online, compared with 16.9 h per week watching television. The results are considered accurate with a margin of error of  $\pm 3.38\%$ , 19 times out of 20.
  - d) A survey conducted at the expense of the political party that holds office indicated that 39% of decided voters said they would not vote for candidates of that party in the next election. The result is considered accurate within  $\pm 3\%$ , 95% of the time.

### Closing

10. Explain why, for a given confidence level,
  - a) the margin of error decreases as the sample size increases
  - b) the margin of error increases as the confidence level increases

### Extending

11. As sample size increases, the margin of error, expressed as a percent, decreases. Consider the table below.

Sample Size	Margin of Error (%)
100	9.80
400	4.90
900	3.27
1600	2.45
2500	1.96
3600	1.63

- a) What mathematical relationship exists between increased sample size and the margin of error?
- b) What would be the margin of error for a sample size of
  - i) 4900
  - ii) 2000
- c) Use your results from parts a) and b) to explain why a relatively small sample will give a fairly accurate indication of the trend for an entire population.