Confidence Intervals

Lesson objectives

- I can distinguish among the meanings of common confidence levels such as 90%, 95%, and 99%
- I can determine the margin of error for a population mean estimated using a sample
- I can determine the upper and lower limits of the confidence interval

1.1

Lesson objectives | Teachers' notes | Lesson notes

MHR Page 359 #s 1 - 5, 7 & 9

Warm up

Fishing is a popular activity in parks such as Algonquin Provincial Park. Many of the lakes in the park contain brook trout as well as lake trout. It is important to collect data on the fish populations and watch for any significant or sudden changes that may indicate problems with the health of the marine ecosystem.

- How can wildlife researchers determine reasonable estimates of the number and species distributions in a large lake?
- Why is it not possible to obtain exact numbers?
- How confident can researchers be that their estimates are close to the correct values?

A researcher might summarize the findings as "Lake trout make up 20.4% of the fish population in Lake Lavieille. This estimate is considered correct within ±3.0%, 19 times out of 20."

What does this statement mean?

Suppose the researcher samples the fish in the lake 20 times. The researcher is confident that the percent of lake trout found in the samples will be within 3% of 20.4% in 19 out of the 20 samples. The **margin of error** is 3% above or below the mean.

The **confidence interval** range is

$20.4\% - 3\% = 17.4\%$ to $20.4\% + 3\% = 23.4\%$

The **confidence level** is 19 out of 20, or 95%. How does this percent connect to the standard deviation of a normal distribution?

Definitions

Margin of Error

- The range of values that a particular measurement is said to be within
- The smaller the margin of error, the greater the accuracy of the measurement

Confidence Interval

- The range of possible values of the measured statistic at a particular confidence level

Confidence Level

- The probability that a particular statistic is within the range indicated by the margin of error
- Commonly used confidence levels are 90%, 95%, and 99%

Investigate on Page 353

Example 1

Lake Trout in Lake Lavieille

Lake Lavieille is one of the largest lakes in Algonquin Park. In 2009, 234 lake trout were caught out of a total catch of 911. In 2012, 141 lake trout were caught out of a catch of 689.

a) Determine the percent of lake trout caught for each year.
b) Determine the margin of error for each year. Use a 95% confidence level.
c) Determine the confidence interval for each year.
d) Do the two confidence intervals overlap? Give a specific example.
e) Is it reasonable to conclude that the percent of lake trout in Lake Lavieille decreased from 2009 to 2012? Give a reason for your answer.

a) For 2009

$$p = \frac{234}{911} \approx 0.257$$

2012

$$p = \frac{141}{689} \approx 0.205$$

The percent of lake trout caught in 2009 was about 25.7% and in 2012 was about 20.5%

b) For 2009

$$E = z\sqrt{\frac{p(1-p)}{n}} \approx 1.96\sqrt{\frac{0.257(1-0.257)}{911}} \approx 0.028$$

2012

$$E = z\sqrt{\frac{p(1-p)}{n}} \approx 1.96\sqrt{\frac{0.205(1-0.205)}{689}} \approx 0.030$$

Confidence Level	z-Score
90%	1.645
95%	1.960
99%	2.576

The margin of error is about 2.8% for 2009 and about 3.0% for 2012.

c) For 2009

Lower limit = 25.7% - 2.8%

= 22.9%

Upper limit = 25.7% + 2.8%

= 28.5%

2012

Lower limit = 20.5% - 3.0%

= 17.5%

Upper limit = 20.5% + 3.0%

= 23.5%

The confidence interval is from 22.9% to 28.5% for 2009 and from 17.5% to 23.5% for 2012.

d) The confidence intervals do overlap. They both include from 22.9% to 23.5%.

e) A 95% confidence level is used. Despite the estimation of the number of trout being lower for 2012, we cannot say for certain as there is some overlap of the confidence intervals.

Your Turn

An opinion poll surveyed 100 households who were watching television at a particular time. Of these, 75% were watching *Hockey Night in Canada*.

a) Determine the margin of error at a 99% confidence level.

b) Determine the confidence interval for this situation.

c) How would a news source state the results?

Confidence Level	z-Score
90%	1.645
95%	1.960
99%	2.576

p = 75 ÷ 100 = 0.75

For a 99% confidence level z = 2.576

Number of households polled was 100, so n = 100

a) $E = z\sqrt{\frac{p(1-p)}{n}}$

$= 2.576\sqrt{\frac{0.75(1-0.75)}{100}}$

≈ 0.112

The margin of error at a 99% confidence level is about 11.2%

b) Lower limit = 75% - 11.2%

= 63.8%

Upper limit = 75% + 11.2%

= 86.2%

The confidence interval is from 63.8% to 86.2%.

c) Hockey Night in Canada is watched by 75% of households. This estimate is considered correct within ± 11.2%, 99 times out of 100.

Example 2

Interpreting Survey Results

The first Major League Baseball franchise outside of the United States was the Montréal Expos, who played from 1969 to 2004. In 2005, the Expos moved to Washington, DC, and are now known as the Washington Nationals. A survey was conducted to determine whether a major league baseball team should come back to Montréal. Of the 1589 people surveyed, 69% were in favour of baseball coming back.

a) Determine the margin of error for this survey at a confidence level of 95%.

b) For what range of percents can you be 95% confident that people would be in favour of baseball returning to Montréal?

c) A second survey at a confidence level of 95% found that 56% were in favour, with a margin of error of 5.2%. Approximately how many people were surveyed?

Confidence Level	z-Score
90%	1.645
95%	1.960
99%	2.576

p = 69%

For a confidence level of 95% z = 1.96

n = 1589 (number of people surveyed)

a) $E = z\sqrt{\frac{p(1-p)}{n}}$

$E = 1.96\sqrt{\frac{0.69(1-0.69)}{1589}}$

$E \approx 0.023$

The margin of error is ± 2.3%

b) Lower limit = 69% - 2.3% = 66.7%

Upper limit = 69% + 2.3% = 71.3%

There is a 95% chance that the percent of people in favour of baseball returning to Montreal is between 66.7% and 71.3%.

c) $E = z\sqrt{\frac{p(1-p)}{n}}$

$E^2 = z^2\left(\frac{p(1-p)}{n}\right)$

$n = \frac{z^2(p(1-p))}{E^2}$

$n = \frac{(1.96)^2(0.56(1-0.56))}{0.052^2}$

$n \approx 350$

p = 0.56, z = 1.96, E = 0.052, solve for n.

The second survey had a sample of about 350 people.

Your Turn

A pharmaceutical manufacturer makes more than 500 000 pills of a certain drug each day. The company randomly samples 400 pills daily to check that they meet the proper weight and size standards. On a given day, 52 pills were found to be substandard.

a) What is the margin of error for this sample at a confidence level of 90%?

b) If the company would like to cut the margin of error in half, how would the sample size have to change?

Confidence Level	z-Score
90%	1.645
95%	1.960
99%	2.576

p = 0.13 (52 ÷ 400)

For a confidence level of 90% z = 1.645

n = 400 (number of pills sampled)

a) $E = z\sqrt{\frac{p(1-p)}{n}}$

$= 1.645\sqrt{\frac{0.13(1-0.13)}{400}}$

≈ 0.028

The margin of error at a 90% confidence level is ± 2.8%

b) Divide E by 2 = 0.014, now solve for n.

$n = \frac{z^2(p(1-p))}{E^2}$

$n = \frac{1.645^2(0.13(1-0.13))}{0.014^2}$

$n \approx 1561$

To cut the margin of error in half, the sample size would have to increase to about 1561 pills.

Repeated Sampling

Suppose that samples of the same size are repeatedly taken from a population that follows a normal distribution with a mean of μ and a standard deviation of σ. The means of the samples will be normally distributed with a mean of μ and a standard deviation of $\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}}$.

Consider the role of the sample size n. If $n = 1$, the standard deviation of the sample means is the same as the standard deviation of the normal distribution. However, as the value of n increases, the standard deviation of the sample means decreases. The usual rule is that a value of $n = 30$ produces a reasonable estimate of the population mean.

You can calculate the margin of error for a sample mean from the formula $E = z\frac{\sigma}{\sqrt{n}}$ or $E = z\sigma_{\bar{x}}$. This is also known as the standard error.

Example 3

Determining Confidence Levels for a Sample Mean

At an agricultural fair, the masses of 8 giant pumpkins entered in a contest were 11 kg, 13 kg, 15 kg, 18 kg, 12 kg, 14 kg, 10 kg, and 16 kg. Results from past fairs suggest that the masses are normally distributed with a mean of 14.2 kg and a standard deviation of 2.5 kg. Determine a 90% confidence interval for the sample mean.

Confidence Level	z-Score
90%	1.645
95%	1.960
99%	2.576

First calculate the mean of these giant pumpkins:

$$\bar{x} = \frac{11 + 13 + 15 + 18 + 12 + 14 + 10 + 16}{8} \approx 13.6$$

as well as the standard deviation of the sample means:

$$\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}} = \frac{2.5}{\sqrt{8}} \approx 0.88$$

For a 90% confidence level z = 1.645

Lower limit $= \bar{x} - E = \bar{x} - z\sigma_{\bar{x}} \approx 13.6 - (1.645)(0.88) \approx 12.2$

Upper limit $= \bar{x} + E = \bar{x} + z\sigma_{\bar{x}} \approx 13.6 + (1.645)(0.88) \approx 15.0$

Therefore, the 90% confidence interval for the mean mass is from 12.2 kg to 15.0 kg. This means that with 90% confidence, you can predict that the mean mass of a pumpkin lies between 12.2 kg and 15.0 kg.

Your Turn

A consumers' group tested batches of light bulbs to see how long they lasted. The results, in hours, from one batch were 998, 1234, 1523, 1760, 937, 1193, 996, 1002, 986, 1285, 1163, and 1716. The manufacturer claims that the life of the light bulbs is normally distributed with a mean of 1200 h and a standard deviation of 420 h.

a) Calculate the mean of the sample and the standard deviation for the sample means.

b) Determine the 99% confidence interval for the sample mean.

Confidence Level	z-Score
90%	1.645
95%	1.960
99%	2.576

a) First calculate the mean life of the light bulbs:

$$\bar{x} = \frac{998+1234+1523+1760+937+1193+996+1002+986+1285+1163+1716}{12}$$
$$=1232.75$$

and then the standard deviation of the means:

$$\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}}$$
$$= \frac{420}{\sqrt{12}}$$
$$\approx 121.24$$

b) For a 99% confidence interval z = 2.576

Lower limit $= \bar{x} - E$
$= \bar{x} - z\sigma_{\bar{x}}$
$= 1232.75 - 2.576(121.24)$
≈ 920.44

Upper limit $= \bar{x} + E$
$= \bar{x} + z\sigma_{\bar{x}}$
$= 1232.75 + 2.576(121.24)$
≈ 1545.06

The 99% confidence interval for the mean life of a light bulb is from about 920.44 hours to 1545.06 hours.

Key Concepts

- The confidence level is the probability that a particular statistic is within the range indicated by the margin of error.
- Commonly used confidence levels are 90%, 95%, and 99%. These are related to the z-scores of the distribution.
- A margin of error is the range of values that a particular statistic is said to be within. For a statistic with probability p, the margin of error is $E = z\sqrt{\frac{p(1-p)}{n}}$.
- The greater the sample size, the smaller the margin of error. The smaller the margin of error, the greater the accuracy of the measurement.
- The confidence interval is the range of possible values of the measured statistic.
- For repeated samples of the same size taken from the same population with a normal distribution, the standard deviation of the sample means is $\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}}$ and the margin of error is $E = z\frac{\sigma}{\sqrt{n}}$.

R1. Can you use the terms "confidence level" and "margin of error" interchangeably? Give reasons for your answer.

No. The confidence level is the probability that a particular statistic is within the range indicated by the margin of error. The confidence level's related z-score is used to calculate the margin of error.

R2. A computer manufacturer found that a mean of 5.6% of a model of tablet computers were returned as defective within one year. The service department considered this number accurate within 1.4%, 9 times out of 10. What information does the confidence interval provide? What is the benefit of using the confidence interval?

Lower limit = 5.6 - 1.4 Upper limit = 5.6 + 1.4

= 4.2% = 7.0%

The confidence interval of 4.2% to 7% is the expected percentage of tablets to be defective. The benefit of knowing this allows the manufacturer to budget for these returns as well. They may also choose to investigate how to improve the manufacturing process.

R3. A 95% confidence interval can be stated as 19 times out of 20. Restate a 90% and a 99% confidence interval in a similar manner.

A 90% confidence interval can be stated as 9 times out of 10.
A 99% confidence interval can be stated as 99 times out of 100.