Confidence Intervals

LET Mathematics & Statistics

In this workshop you will perform some calculations to create confidence intervals for the mean. You will also explore what factors influence the size of a confidence interval.

Researchers often want to estimate a measure, such as the mean, of a population. Measures that describe a population are known as parameters. If everyone in the population can be measured (i.e. a census) then there is no need to estimate parameters. However, samples from a population are more often used and statistics are calculated from the sample. These statistics are then used to estimate the population parameters.

Point and Interval Estimates

A statistic means something that is calculated from a sample. For example, when we calculate the mean score from a sample, this score is known as a statistic

A parameter is a measure from the population, and is often unknown. When you calculate statistics from a sample you are often trying to estimate the value of a population parameter

The mean is often a useful piece of information to know about a population. A sample mean can be used to estimate the population mean. So if a sample of 80 people in a particular population of interest gives a mean score of 52, you would estimate the population mean to be 52. This is known as a point estimate

Point estimates are good for giving readers a precise answer, but you have no confidence that this is the true value in the population. This is because:

$$\text{true score} = \text{sample estimate} \pm \text{error}$$

Researchers will often also give interval estimates of the population parameter, so a range of possible values it can take. It will also be associated with a level of confidence in the interval. A commonly used interval is a 95% confidence interval. So a 95% confidence interval for a population mean

may be given as:

$$52 \pm 6$$

This means that we can be 95% confident that the true mean lies between $52-6 = 46$ and $52+6 = 58$, so the confidence interval may also be written $(46, 58)$.

It's 95%, or whatever percent is given, because when a researcher calculates the 95% confidence interval from their sample, 95

1. Let's say that a researcher is interested in the mean length of time people had symptoms of a particular virus. They collect a sample from the population of people who had the virus and calculate the mean length to be 6.4 days. Using other statistics collected from the sample, the researcher also calculates a 95% confidence interval for the population mean. It is calculated as (4.6, 8.2).

    (a) With the information given, what would be the researcher's estimate of the population mean number of days people were ill with the virus?

    > **Solution**
    >
    > The best estimate for the population mean would be the sample mean: 6.4 days

    (b) What is the interpretation of this confidence interval?

    > **Solution**
    >
    > We are 95% confident that the true mean is between 4.6 and 8.2 days

What factors influence the width of a confidence interval?

The size of the confidence interval (the difference between the two values in the interval) is determined by:

- the amount of spread there is in the sample data

- the size of the sample

- the level of confidence

Below are some confidence intervals for the mean where:

Sample mean = 70, sample standard deviation = 5, sample size = 50.

80% Confidence Interval (69.1 to 70.9)

90% Confidence Interval (68.8 to 71.2)

95% Confidence Interval (68.6 to 71.4)

99% Confidence Interval (68.2 to 71.8)

Each of these confidence intervals can be written as mean±margin of error.

So, for example, the 80% confidence interval can be written as:

$$70 \pm 0.9$$

So, the margin of error here was 0.9.

2. For the confidence intervals above

    (a) What is the margin of error for the 90% confidence interval?

> **Solution**
> 1.2 -: 71.1 is 1.2 above the mean and 68.8 is 1.2 below the mean

    (b) What is the margin of error for the 95% confidence interval?

> **Solution**
> 1.4 -: 71.4 is 1.4 above the mean and 68.6 is 1.4 below the mean

    (c) What is the margin of error for the 99% confidence interval?

> **Solution**
> 1.8 -: 71.8 is 1.8 above the mean and 68.2 is 1.8 below the mean

3. What do you notice between the relationship of confidence level and margin of error?

○ There is no relationship between the confidence level and the margin of error

○ As the level of confidence increases, the margin of error decreases

⊗ **As the level of confidence increases, the margin of error increases**

Below are 95% confidence intervals for different samples that have a mean of 70 and a standard deviation of 5. The sample size is different for each of the samples:

Sample size = 50; 95% Confidence Interval (68.6 to 71.4)

Sample size = 10; 95% Confidence Interval (66.9 to 73.1)

Sample size = 2000; 95% Confidence Interval (69.8 to 70.2)

4. What do you notice between the relationship of sample size and margin of error for the 95% confidence intervals?

   ○ There is no relationship between the sample size and the margin of error

   ⊗ **As the sample size increases, the margin of error decreases**

   ○ As the sample increases, the margin of error increases

> **Solution**
>
> Sample size 50, margin of error is 1.4, sample size of 10, margin of error is 3.1, sample size of 2000, margin of error is 0.2

Calculating a 95% confidence interval for the mean, with a known population standard deviation

The margin of error in a confidence interval is determined by the level of confidence, the sample size, and the standard deviation in the sample. When the population standard deviation is known you can use the formula, for a 95% confidence interval:

$$\bar{x} \pm 1.96 \times \frac{\sigma}{\sqrt{n}}$$

Where $\bar{x}$ is the sample mean, σ is the population standard deviation, n is the sample size.

If the sample size is large, then this formula is a close approximation when using the sample standard deviation rather than the population.

5. Calculate the 95% confidence intervals for the following

    (a) sample mean = 40, population standard deviation = 10, sample size = 100

    (b) sample mean = 100, population standard deviation = 15, sample size = 25

    (c) sample mean = 100, population standard deviation = 15, sample size = 225

    (d) sample mean = 100, population standard deviation = 75, sample size = 225

    (e) sample mean = 100, population standard deviation = 75, sample size = 225

Solution

a. The margin of error is $1.96 \times \frac{10}{\sqrt{10}} = 1.96$ so
$40 - 1.96 = 38.04$,
$40 + 1.96 = 41.96$: **(38.04, 41.96)**
b. The margin of error is $1.96 \times \frac{15}{\sqrt{25}} = 5.88$ so
$100 - 5.88 = 94.12$,
$100 + 5.88 = 105.88$: **(94.12, 105.88)**
c. The margin of error is $1.96 \times \frac{15}{\sqrt{225}} = 1.96$ so
$100 - 1.96 = 98.04$,
$100 + 1.96 = 101.96$: **(98.04, 101.96)**
d. The margin of error is $1.96 \times \frac{75}{\sqrt{225}} = 9.2$ so
$100 - 9.2 = 90.8$
$100 + 9.2 = 109.2$: **(90.2, 109.2)**