Confidence intervals

- A plausible range of values for the population parameter is called a *confidence interval*.
- Using only a sample statistic to estimate a parameter is like fishing in a murky lake with a spear, and using a confidence interval is like fishing with a net.

We can throw a spear where we saw a fish but we will probably miss. If we toss a net in that area, we have a good chance of catching the fish.

- If we report a point estimate, we probably won't hit the exact population parameter. If we report a range of plausible values we have a good shot at capturing the parameter.

Photos by Mark Fischer (http://www.flickr.com/photos/fischerfotos/7439791462) and Chris Penny (http://www.flickr.com/photos/clearlydived/7029109617) on Flickr.

Average number of exclusive relationships

A random sample of 50 college students were asked how many exclusive relationships they have been in so far. This sample yielded a mean of 3.2 and a standard deviation of 1.74. Estimate the true average number of exclusive relationships using this sample.

Average number of exclusive relationships

A random sample of 50 college students were asked how many exclusive relationships they have been in so far. This sample yielded a mean of 3.2 and a standard deviation of 1.74. Estimate the true average number of exclusive relationships using this sample.

$$\bar{x} = 3.2 \qquad s = 1.74$$

Average number of exclusive relationships

A random sample of 50 college students were asked how many exclusive relationships they have been in so far. This sample yielded a mean of 3.2 and a standard deviation of 1.74. Estimate the true average number of exclusive relationships using this sample.

$$\bar{x} = 3.2 \qquad s = 1.74$$

The approximate 95% confidence interval is defined as

$$point\ estimate \pm 2 \times SE$$

Average number of exclusive relationships

A random sample of 50 college students were asked how many exclusive relationships they have been in so far. This sample yielded a mean of 3.2 and a standard deviation of 1.74. Estimate the true average number of exclusive relationships using this sample.

$$\bar{x} = 3.2 \qquad s = 1.74$$

The approximate 95% confidence interval is defined as

$$point\ estimate \pm 2 \times SE$$

$$SE = \frac{s}{\sqrt{n}} = \frac{1.74}{\sqrt{50}} \approx 0.25$$

Average number of exclusive relationships

A random sample of 50 college students were asked how many exclusive relationships they have been in so far. This sample yielded a mean of 3.2 and a standard deviation of 1.74. Estimate the true average number of exclusive relationships using this sample.

$$\bar{x} = 3.2 \qquad s = 1.74$$

The approximate 95% confidence interval is defined as

$$point\ estimate \pm 2 \times SE$$

$$SE = \frac{s}{\sqrt{n}} = \frac{1.74}{\sqrt{50}} \approx 0.25$$

$$\bar{x} \pm 2 \times SE \quad = \quad 3.2 \pm 2 \times 0.25$$

Average number of exclusive relationships

A random sample of 50 college students were asked how many exclusive relationships they have been in so far. This sample yielded a mean of 3.2 and a standard deviation of 1.74. Estimate the true average number of exclusive relationships using this sample.

$$\bar{x} = 3.2 \qquad s = 1.74$$

The approximate 95% confidence interval is defined as

$$point\ estimate \pm 2 \times SE$$

$$SE = \frac{s}{\sqrt{n}} = \frac{1.74}{\sqrt{50}} \approx 0.25$$

$$\begin{aligned} \bar{x} \pm 2 \times SE &= 3.2 \pm 2 \times 0.25 \\ &= (3.2 - 0.5, 3.2 + 0.5) \end{aligned}$$

Average number of exclusive relationships

A random sample of 50 college students were asked how many exclusive relationships they have been in so far. This sample yielded a mean of 3.2 and a standard deviation of 1.74. Estimate the true average number of exclusive relationships using this sample.

$$\bar{x} = 3.2 \qquad s = 1.74$$

The approximate 95% confidence interval is defined as

$$point\ estimate \pm 2 \times SE$$

$$SE = \frac{s}{\sqrt{n}} = \frac{1.74}{\sqrt{50}} \approx 0.25$$

$$\begin{aligned} \bar{x} \pm 2 \times SE &= 3.2 \pm 2 \times 0.25 \\ &= (3.2 - 0.5, 3.2 + 0.5) \\ &= (2.7, 3.7) \end{aligned}$$

Which of the following is the correct interpretation of this confidence interval?

We are 95% confident that

(a) the average number of exclusive relationships college students in this sample have been in is between 2.7 and 3.7.

(b) college students on average have been in between 2.7 and 3.7 exclusive relationships.

(c) a randomly chosen college student has been in 2.7 to 3.7 exclusive relationships.

(d) 95% of college students have been in 2.7 to 3.7 exclusive relationships.

Which of the following is the correct interpretation of this confidence interval?

We are 95% confident that

(a) the average number of exclusive relationships college students in this sample have been in is between 2.7 and 3.7.

(b) *college students on average have been in between 2.7 and 3.7 exclusive relationships.*

(c) a randomly chosen college student has been in 2.7 to 3.7 exclusive relationships.

(d) 95% of college students have been in 2.7 to 3.7 exclusive relationships.

A more accurate interval

Confidence interval, a general formula

$$\textit{point estimate} \pm z^{\star} \times SE$$

A more accurate interval

Confidence interval, a general formula

$$point\ estimate \pm z^{\star} \times SE$$

Conditions when the point estimate = $\bar{x}$:

1. *Independence:* Observations in the sample must be independent
    - random sample/assignment
    - if sampling without replacement, $n <$ 10% of population
2. *Sample size / skew:* $n \geq 30$ and population distribution should not be extremely skewed

A more accurate interval

Confidence interval, a general formula

$$\textit{point estimate} \pm z^{\star} \times SE$$

Conditions when the point estimate = $\bar{x}$:

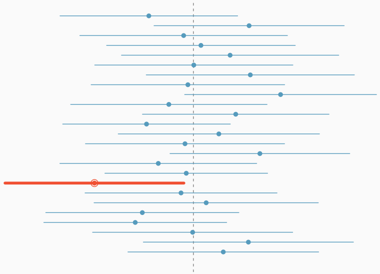
1. *Independence:* Observations in the sample must be independent
    - random sample/assignment
    - if sampling without replacement, $n <$ 10% of population
2. *Sample size / skew:* $n \geq 30$ and population distribution should not be extremely skewed

Note: We will discuss working with samples where $n < 30$ in the next chapter

What does 95% confident mean?

- Suppose we took many samples and built a confidence interval from each sample using the equation $point\ estimate \pm 2 \times SE$.
- Then about 95% of those intervals would contain the true population mean (μ).

- The figure shows this process with 25 samples, where 24 of the resulting confidence intervals contain the true average number of exclusive relationships, and one does not.



Width of an interval

If we want to be more certain that we capture the population parameter, i.e. increase our confidence level, should we use a wider interval or a smaller interval?

Width of an interval

If we want to be more certain that we capture the population parameter, i.e. increase our confidence level, should we use a wider interval or a smaller interval?

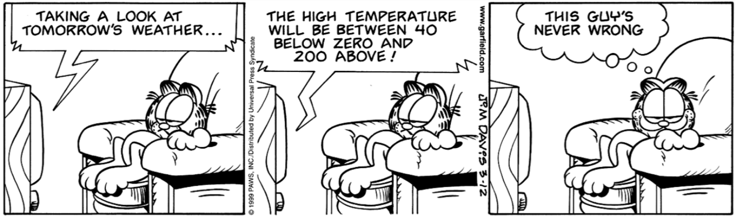
A wider interval.

Width of an interval

If we want to be more certain that we capture the population parameter, i.e. increase our confidence level, should we use a wider interval or a smaller interval?

A wider interval.

Can you see any drawbacks to using a wider interval?

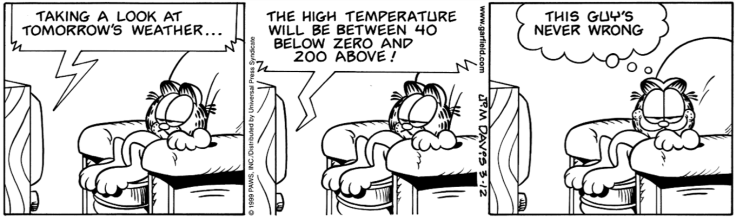


Width of an interval

If we want to be more certain that we capture the population parameter, i.e. increase our confidence level, should we use a wider interval or a smaller interval?

A wider interval.

Can you see any drawbacks to using a wider interval?



If the interval is too wide it may not be very informative.

Image source: http://web.as.uky.edu/statistics/users/earo227/misc/garfield_weather.gif

Changing the confidence level

$$point\ estimate \pm z^\star \times SE$$

- In a confidence interval, $z^\star \times SE$ is called the *margin of error*, and for a given sample, the margin of error changes as the confidence level changes.
- In order to change the confidence level we need to adjust $z^\star$ in the above formula.
- Commonly used confidence levels in practice are 90%, 95%, 98%, and 99%.
- For a 95% confidence interval, $z^\star = 1.96$.
- However, using the standard normal (z) distribution, it is possible to find the appropriate $z^\star$ for any confidence level.