

⚠ Lagunita is retiring and will shut down at 12 noon Pacific Time on March 31, 2020. A few courses may be open for self-enrollment for a limited time. We will continue to offer courses on other online learning platforms; visit <http://online.stanford.edu>.

Course > Inference: Estimation > Estimation: Population Mean >
Confidence Intervals for the Population Mean: The General Case

🔖 Bookmark this page

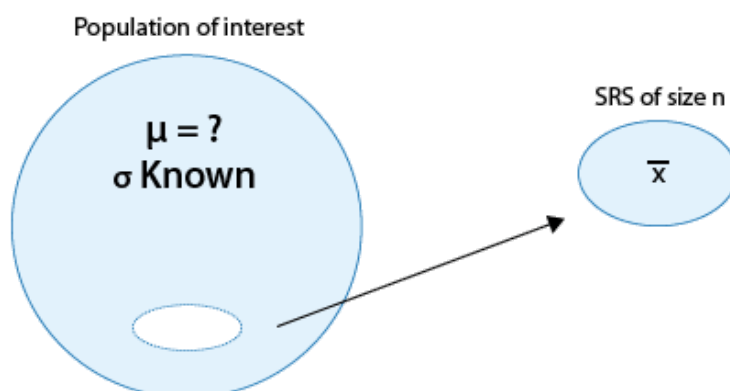
Confidence Intervals for the Population Mean: The General Case

Learning Objective: Explain what a confidence interval represents and determine how changes in sample size and confidence level affect the precision of the confidence interval.

Learning Objective: Find confidence intervals for the population mean and the population proportion (when certain conditions are met), and perform sample size calculations.

The General Case

Let's generalize the IQ example. Suppose that we are interested in estimating the unknown population mean (μ) based on a random sample of size n . Further, we assume that the population standard deviation (σ) is known.



The values of \bar{x} follow a normal distribution with (unknown) mean μ and standard deviation $\frac{\sigma}{\sqrt{n}}$ (known, since both σ and n are known). By the (second part of the) Standard Deviation Rule, this means that:

There is a 95% chance that our sample mean (\bar{x}) will fall within $2 * \frac{\sigma}{\sqrt{n}}$ of μ ,

which means that:

We are 95% confident that μ falls within $2 * \frac{\sigma}{\sqrt{n}}$ of our sample mean (\bar{x}).

Or, in other words, a 95% confidence interval for the population mean μ is:

$$\left(\bar{x} - 2 * \frac{\sigma}{\sqrt{n}}, \bar{x} + 2 * \frac{\sigma}{\sqrt{n}} \right)$$

Here, then, is the **general result**:

Suppose a random sample of size n is taken from a normal population of values for a quantitative variable whose mean (μ) is unknown, when the standard deviation (σ) is given. A 95% confidence interval (CI) for μ is:

$$\bar{x} \pm 2 * \frac{\sigma}{\sqrt{n}}$$

Comment

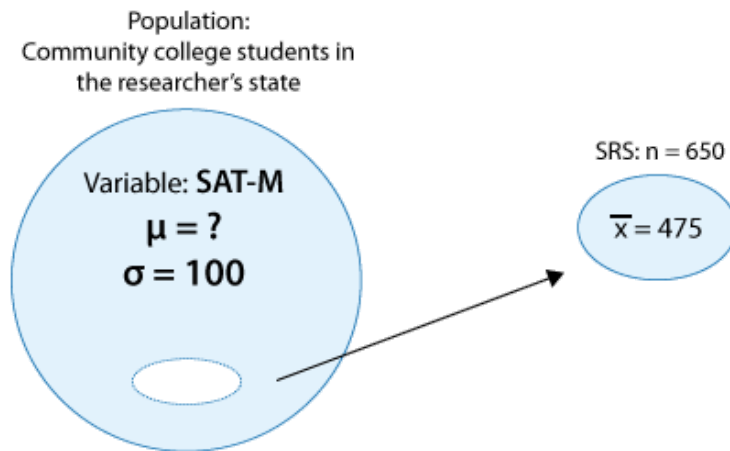
Note that for now we require the population standard deviation (σ) to be known. Practically, σ is rarely known, but for some cases, especially when a lot of research has been done on the quantitative variable whose mean we are estimating (such as IQ, height, weight, scores on standardized tests), it is reasonable to assume that σ is known. Eventually, we will see how to proceed when σ is unknown, and must be estimated with sample standard deviation (s).

Let's look at another example.

Example

An educational researcher was interested in estimating μ , the mean score on the math part of the SAT (SAT-M) of all community college students in his state. To this end, the researcher has chosen a random sample of 650 community college students from his state, and found that their average SAT-M score is 475. Based on a large body of research that was done on the SAT, it is known that the scores roughly follow a normal distribution with the standard deviation $\sigma = 100$.

Here is a visual representation of this story, which summarizes the information provided:



Based on this information, let's estimate μ with a 95% confidence interval.

Using the formula we developed before, $\bar{x} \pm 2 * \frac{\sigma}{\sqrt{n}}$, a 95% confidence interval for μ is:

$\left(475 - 2 * \frac{100}{\sqrt{650}}, 475 + 2 * \frac{100}{\sqrt{650}}\right)$, which is $(475 - 7.8, 475 + 7.8) = (467.2, 482.8)$. In this case, it makes sense to round, since SAT scores can be only whole numbers, and say that the 95% confidence interval is (467,483).

We are not done yet. An equally important part is to **interpret what this means in the context of the problem.**

We are 95% confident that the mean SAT-M score of all community college students in the researcher's state is covered by the interval (467,483). Note that the confidence interval was obtained by taking **475 ± 8** (rounded). This means that we are 95% confident that by using the sample mean ($\bar{x} = 475$) to estimate μ , our error is no more than 8.

Scenario: Smoking During Pregnancy

A study was done on pregnant women who smoked during their pregnancies. In particular, the researchers wanted to study the effect that smoking has on pregnancy length. A sample of 114 pregnant women who were smokers participated in the study and were followed until the birth of their child. At the end of the study, the collected data were analyzed and it was found that the average pregnancy length of the 114 women was 260 days. From a large body of research, it is known that length of human pregnancy has a standard deviation of 16 days.

Learn By Doing (1/1 point)

Based on this study, find a 95% confidence interval for μ , the mean pregnancy length of women who smoke during their pregnancy, and interpret your interval in context.

Your Answer:

mean \pm std. 260 \pm 2*16/sqrt(114) = interval from 257.00 to 263.00

This means we are 95% confident that the mean pregnancy length of women who smoke is within 257 to 263.

Our Answer:

The problem provides the following information: • Our sample size is $n = 114$ • This sample gives a sample mean of • Human pregnancy is known to have a standard deviation of $\sigma = 16$. The 95% confidence interval for μ is therefore: or (257,263). Interpretation: We are 95% confident that the mean pregnancy length of women who smoke during pregnancy is covered by the interval (257, 263).

Resubmit

Reset

Learn By Doing (1/1 point)

Note that the confidence interval was obtained by calculating 260 ± 3 . Give an interpretation of the number 3.

Your Answer:

It's based on the formula $2*16/\sqrt{114}$. This means that our estimate is off by at most 3 days.

Our Answer:

We are 95% confident that by using 260 days as the estimate for μ , our estimation error is no more than 3 days.

Resubmit

Reset

You just gained practice computing and interpreting a confidence interval for a population mean. Note that the way a confidence interval is used is that we hope the interval contains the population mean μ . This is why we call it an "interval **for the population mean**."

We just saw that one interpretation of a 95% confidence interval is that we are 95% confident that the population mean (μ) is contained in the interval. Another useful interpretation in practice is that, given the data, the confidence interval represents the set of plausible values for the population mean μ .

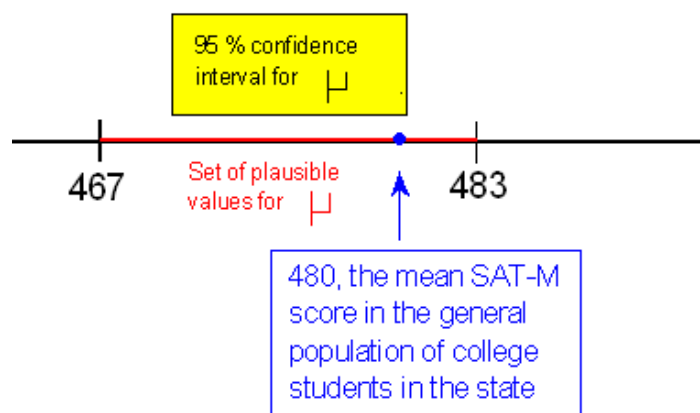
Example

As an illustration, let's return to the example of mean SAT-Math score of community college students. Recall that we had constructed the confidence interval (467,483) for the unknown mean SAT-M score for all community college students.

Here is a way that we can use the confidence interval:

Do the results of this study provide evidence that μ , the mean SAT-M score of community college students, is lower than the mean SAT-M score in the general population of college students in that state (which is 480)?

The 95% confidence interval for μ was found to be (467,483). Note that 480, the mean SAT-M score in the general population of college students in that state, falls inside the interval, which means that it is one of the plausible values for μ .



This means that μ could be 480 (or even higher, up to 483), and therefore we cannot conclude that the mean SAT-M score among community college students in the state is lower than the mean in the general population of college students in that state. (Note that the fact that most of the plausible values for μ fall below 480 is not a consideration here.)

Comment

Recall that in the formula for the 95% confidence interval for μ , $\bar{x} \pm 2 * \frac{\sigma}{\sqrt{n}}$, the 2 comes from the Standard Deviation Rule, which says that any normal random variable (in our case \bar{X}), has a 95% chance (or probability of .95) of taking a value that is within 2 standard deviations of its mean.

As you recall from the discussion about the normal random variable, this is only an approximation, and to be more accurate, there is a 95% chance that a normal random variable will take a value within 1.96 standard deviations of its mean. Therefore, a more accurate formula for the 95% confidence interval for

μ is $\bar{x} \pm 1.96 * \frac{\sigma}{\sqrt{n}}$, which you'll find in most introductory statistics books. In this course, we'll use 2 (and not 1.96), which is close enough for our purposes.

Open Learning Initiative [↗](#)



[↗](#) Unless otherwise noted this work is licensed under a Creative Commons Attribution-NonCommercial-ShareAlike 4.0 International License [↗](#).

© All Rights Reserved