Confidence Intervals

I'll continue with the example of estimating the mean from some population using the mean from a random sample of *n* observations, using the following example .

Confidence Interval Construction

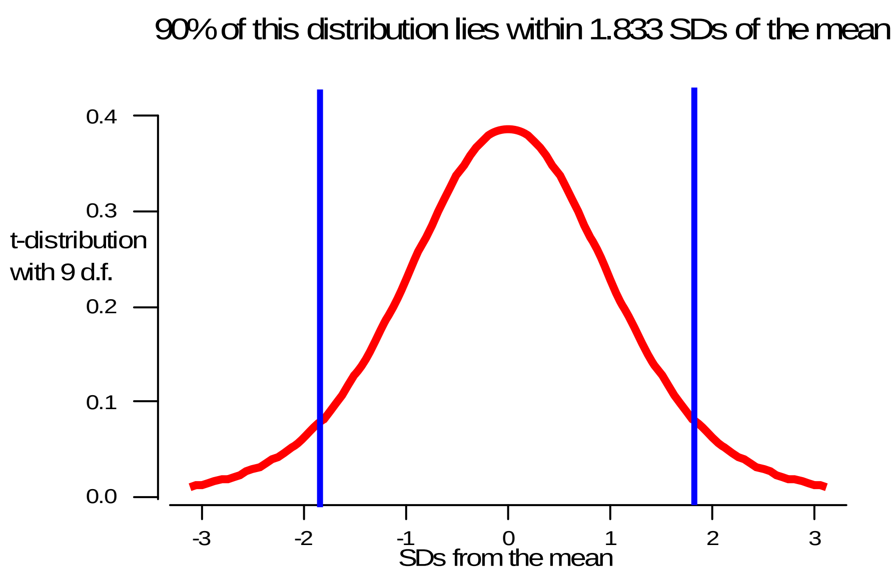
Classically constructed confidence intervals for a mean take the form: mean minus and mean plus margin of error. You might be familiar with the term margin of error from newspaper reports of polls. Results are typically stated something like, "52% preferred blablabla, with a margin of error of 2%, accurate 19 times out of 20." This leaves you with the feeling that the reports authors are pretty sure the truth is somewhere between 50% and 54%. There is an intuitive "plus/minus" interpretation to the phrase, "margin of error." In fact, the foregoing statement is the statement of a 95% confidence interval. I know that it was a 95% inyterval by the phrase, "19 times out of 20", written for readers who may not have taken a statistics class.

The margin of error is explicitly part of the construction of a classical confidence interval. To wit (sticking with our fish-in-nets example), mean minus margin of error gives the lower confidence limit; mean plus margin of error yields the upper limit. The margin of error is determined by several factors, including your choice of confidence level, sample size, and amount of innate variability in the data. For a mean from a random sample, the margin of error formula is

$$m.e. = t_{df,cl}\frac{SD}{\sqrt{n}}.$$

The logic behind this formula hinges on the imaginary act of repeating our study *ad infinitum*. In that imaginary act, we would generate all possible sample means that could arise from studies *exactly* like ours. We imagine, then, that there is some distribution of those means. The two pieces of the margin of error formula play a role that references that distribution, as follows.

The first term in the margin of error. $\left(t_{df,cl}\right)$ essentially answers the question, “How many standard deviations (of the mean) wide does the interval need to be?” The key to that answer is an assumption that the distribution of the mean is approximately Normal . For my purposes now, I will assume it to be so, and use it. The *df* in the subscript reminds us that the correct *t*–distribution to use is the one with *df* degrees of freedom. The other part connects your choice *cl* of confidence level to that *t*–distribution. This is perhaps better explained visually, so take a look at the following graph, illustrated with $cl = 0.90$, a 90% confidence level, and $df = 9$, coinciding with the sample size from our fish example. There are several equivalent statements that can be made about the vertical blue lines, connecting them conceptually to the choice of confidence level.



90% of the distribution lies within 1.833 SDs. This is perhaps the most intuitively natural statement, and could lead to notation $\left(t_{df,cl}\right)$. Note that this statement is a feature of the *t*–distribution (with 9 degrees of freedom, which is the *t*–distribution associated with a random sample of $n = 10$ observations). The units of measurement for the *t* are SDs (or SEs) of the mean; like a Normal distribution, approximately all of the area under the curve is encompassed within 3 SDs from the mean. It turns out for this particular case that the central 90% is within 1.833 SDs. Different sample sizes (or different choices of confidence levels) will change that number .

The shape of the Normal distribution determines the relationship between the width (as measured in SDs) of an interval and the confidence level Your choice of confidence level determines how many SDs (of the distribution of the mean) wide the margin of error must be (the *cl* subscript on the *t* is a reminder of the dependence on confidence level choice). For example, for 95% confidence intervals, the margin of error is usually about 2 SDs wide. A 90% interval requires a margin of error of about 1.6 SDs, while a 99% confidence interval needs a margin of error that is about 3 SDs wide. The exact number of SDs required depends on sample size, hence my use of the word, “about”. For further explanation: .

It turns out that we can’t use a perfect Normal distribution in our estimation of the margin of error. The reason is that we don’t know the SD of the distribution of the mean; we only have an estimate. The consequence is that we need to use a fudged version of the Normal. How much fudging depends on sample size: larger sample size requires less fudging. The distribution we use is called the *t* distribution . It’s actually a family of distributions; the formula for a given *t* distribution is driven by the degrees of freedom (hence the connection to sample size and the *df* in the subscript on the *t* as a reminder).

The actual value of $\left(t_{df,cl}\right)$, then, depends primarily on choice of confidence level; it gets tweaked a little depending on sample size (actually, it gets tweaked a lot for very small sample sizes). The following table illustrates how these two factors affect $\left(t_{df,cl}\right)$.

Values of $\left(t_{df,cl}\right)$ for a variety of confidence levels and sample sizes			
Sample Size	90% confidence	95% confidence	99% confidence
5	2.13	2.78	4.60
10	1.83	2.26	3.25
20	1.73	2.09	2.86
40	1.69	2.02	2.71
100	1.66	1.98	2.63

For further exploration of confidence properties:  and details of construction:  .

In our example, having chosen a 95% confidence level, the multiplier is $\left(t_{9,\,0.95}\right) = 2.2622$. The estimated SD of the mean is $4.72/\sqrt{10} = 1.49$, yielding a margin of error of 2.2622×1.49 = 3.37. Thus the confidence interval is (12.03, 18.77).

Interpretation

There are several commonly used interpretations of a confidence interval (at least among folks just learning about them), not all of them correct. Consider each of the following statements (about a 95% interval), and select an answer.

There is a 95% chance that the sample mean is contained in our interval.

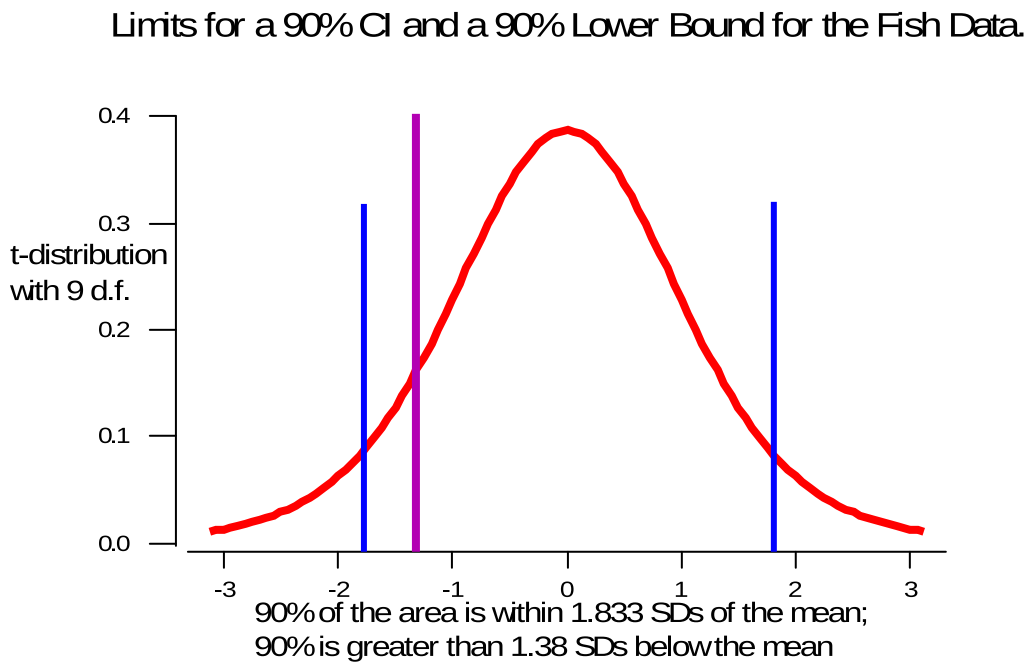
If we were to repeat our experiment *ad infinitum* (not possible except as a thought experiment), 95% of the confidence intervals we construct would contain the population mean.

There is a 95% chance that the population mean is in our interval. .

We are 95% confident that the population mean is in the interval.

Confidence Bounds

It is sometimes of interest to construct a so-called one-sided confidence interval (or a confidence bound). Just as people conventionally choose complementary confidence levels and alpha levels (i.e. an alpha of 0.05 usually coincides with the choice of 95% confidence intervals), one can choose to use a one-sided confidence interval to coincide with a one-tailed hypothesis test. The construction parallels the confidence interval, except that one chooses "statistic minus margin of error," (to make a lower confidence bound) *or* "statistic plus margin of error," to make an upper confidence bound The following graph illustrates the difference between the two constructions.



If one were to make a 90% C.I. here, you would make it as "estimate plus and minus 1.833 SEs (of the estimate). If you were interested in a 90% lower bound, it would be constructed as "estimate minus 1.38 SDs". Look closely at the picture, and you'll agree (I hope) that about 9/10ths of the area under the curve is between the two blue lines (the choice for a 90% interval) and, also, abut 9/10ths is to the right of the purple line (this would be used for a lower bound).

I will denote the *t*-multiplier by $\left(t_{df,cl}\right)$, with the understanding that the exact value will be determined by whether one is doing a confidenc interval or a confidence bound.

For our fish case, a one-sided interval would be a reasonable choice (consonant with the one-tailed hypothesis test). Keeping with the usual complementarity between alpha and confidence levels, I'm going to use a confidence level of 90% (recall that the biologists chose alpha = 0.10 in the example. Here, $t_{9,0.90} = 1.38$; the margin of error is 1.38×1.33 (the SD of the mean), which equals 1.84. Thus we would say that we are 90% confident that the true mean is 15.4 - 1.84 = 13.56 or higher. In this case, since we are making a one-sided interval (in this case leading to a so-called lower bound), we use the number of SDs indicated by the purple line in the figure above (1.38), not the value indicated by the blue lines (1.83).