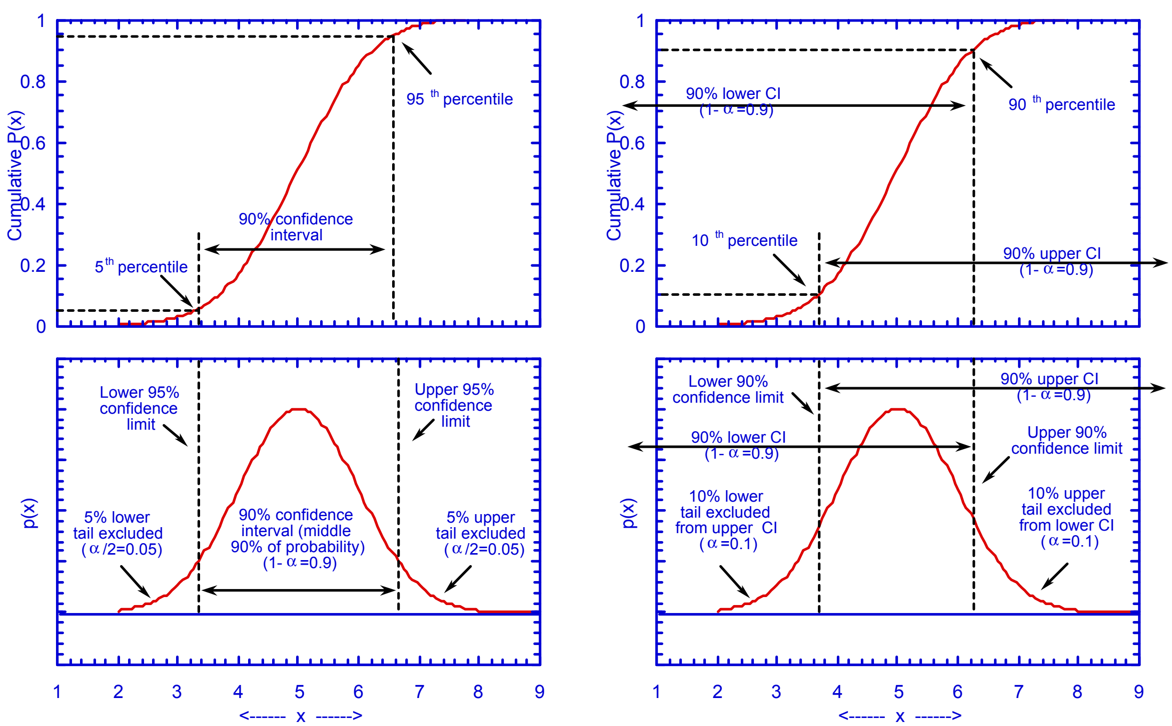
The *confidence interval* of any uncertain quantity *x* is *the range that x is expected to occupy with a specified confidence*. Anything that is uncertain can have a confidence interval. Confidence intervals are most often used to express the uncertainty in a sample mean, but confidence intervals can also be calculated for sample medians, variances, and so forth. If you can calculate a quantity, and if you know something about its probability distribution, you can estimate confidence intervals for it.

Nomenclature: *confidence intervals* are also sometimes called *confidence limits*. Used this way, these terms are equivalent; more precisely, however, the *confidence limits* are the values that mark the ends of the confidence interval. The *confidence level* is the likelihood associated with a given confidence interval; it is the level of confidence that the value of *x* falls within the stated confidence interval.

General Approach

If you know the theoretical distribution for a quantity, then you know any confidence interval for that quantity. For example, the 90% confidence interval is the range that encloses the middle 90% of the likelihood, and thus excludes the lowest 5% and the highest 5% of the possible values of *x*. That is, the 90% confidence interval for *x* is the range between the 5^{th} percentile of *x* and the 95^{th} percentile of *x* (see the two left-hand graphs below). This is an example of a *two-tailed* confidence interval, one that excludes an equal probability from both tails of the distribution. One also occasionally finds one-sided confidence intervals, which only exclude values from the upper tail or the lower tail. Examples of one-sided 90% confidence intervals are shown in the two right-hand graphs below.



By convention, the total fraction of probability that is excluded by the confidence interval is denoted α, and thus the probability that is included in the confidence interval is $1-\alpha$. In one-sided confidence intervals, all of the excluded probability α is in one tail; in the more common two-tailed confidence intervals, α is divided among the two tails, with the outermost $\alpha/2$ excluded from each.

The confidence interval depends on three things: 1) the *shape* of the distribution, 2) the *width* of the distribution, and 3) the desired *confidence level*, the probability to be encompassed within the interval. Thus the procedure for determining the confidence interval for a quantity includes three steps: 1) figure out--or assume--the shape of the distribution, 2) measure--or estimate--the width of the distribution, usually by calculating a standard error, and 3) decide on the confidence level desired. The tricky part is being able to calculate the percentiles of the distribution; in practice this means that one uses distributions whose mathematical properties are well understood (such as the normal distribution).

Confidence interval for the population mean

The key to estimating confidence intervals for population means is Student's t, which is the number of standard errors that encompass a specified probability for the population mean around a sample mean. Student's t is a function of α, the probability to be excluded in the tails, and ν, the number of degrees of freedom. Here's how to use it:

1. Calculate the sample standard deviation, s_x.
2. From this, calculate the standard error of the mean, $s_{\bar{x}} = s_x / \sqrt{n}$.
3. Decide on the confidence level, $1-\alpha$, to be included in the confidence interval.
4. Calculate the degrees of freedom, ν. The standard error of the mean has $\nu=n-1$ degrees of freedom.
5. For a two-tailed confidence interval, look up t for $\alpha/2$ and ν degrees of freedom. One can have a confidence of $1-\alpha$ that the true population mean μ lies in the interval,

$$\bar{x} - t_{\alpha/2,\nu} s_{\bar{x}} < \mu < \bar{x} + t_{\alpha/2,\nu} s_{\bar{x}}$$

6. For a one-sided confidence interval, look up t for α and ν degrees of freedom. One can have a confidence of $1-\alpha$ that the true population mean μ lies either in the interval,

$$\bar{x} - t_{\alpha,\nu} s_{\bar{x}} < \mu < \infty$$

or in the interval,

$$\infty < \mu < \bar{x} + t_{\alpha,\nu} s_{\bar{x}}$$

Assumptions: either (a) the individual measurements x are normally distributed, or (b) the number of measurements is large enough that the mean of x is normally distributed (via the central limit theorem) even if the individual x values are not. For distributions typically encountered in environmental data, $n>20$ or so is usually sufficient.

Similar approaches are available for variables with non-normal distributions. Helsel and Hirsch (see below) give formulas for log-normal distributions, and references to related literature.

Interpretation of confidence intervals

Strictly speaking, one should not say that the *probability* is $1-\alpha$ that the confidence interval encloses the true mean μ, although this is often the intuitive inference that is drawn. In classical statistics, the *parameters* of the *population* (such as its mean μ) are not probabilistic--that is, the *true* mean is whatever it is, and it does not have a probability distribution. Instead it is the *statistics* of the *sample* that are probabilistic. Thus the correct statement is that if one takes many sets of measurements on a population, and calculates a sample mean for each of them, and calculates a confidence interval around each sample mean, then a fraction $1-\alpha$ of those confidence intervals will include the true mean, and a fraction α of them will not--but one cannot say anything about the *probability* that any one interval encloses the mean. Either it does or it doesn't--one may be *uncertain* about whether the mean is enclosed in any one case, but that makes it a question of one's *confidence* that this is so, rather than the *probability* of it (hence the name *confidence interval*).

So strictly speaking, one should say that one has a *confidence* of $1-\alpha$ that the stated interval encloses the mean. More generally, it is important to recognize that *one can only make probability statements about samples or data, not about populations or hypotheses*. The right way to make probability statements about populations and hypotheses (or rather, to make confidence statements about them in a rigorous way) is to use Bayesian inference. But in this case the common parlance (attributing probability to the value of the mean) is relatively harmless. Even Zar uses it.

Sample size required for a given confidence interval and confidence level

If the half-width of the confidence interval is *d*, that is,

$$\bar{x} - d < \mu < \bar{x} + d$$

then (by rearranging the equations above),

$$n = \left(\frac{t_{\alpha/2,\nu} s_x}{d} \right)^2$$

Because *t* depends on the number of degrees of freedom (which in turn depends on n), this equation must be solved iteratively, by updating ν based on the value of *n* from the previous iteration. Note that the sample size increases--rapidly!--as the desired confidence level increases (α decreases and thus *t* increases), or as the confidence interval becomes narrower (d decreases).

Note also that *n* is predicted from an estimate of s_x, which is presumably derived from a smaller "pilot" study. Once you have actually collected *n* measurements, there is a 50% chance that the standard deviation for the whole *n* will be larger or smaller than the s_x from your pilot study; thus there's a 50% chance that the actual confidence interval will turn out to be wider than *d*, and a 50% chance that it will be narrower. In order to reduce the chance of underestimating the *n* required to achieve a specified confidence interval (a 50% chance, if you use the equation above), you will need to account for statistical *power*; see the Toolkit on Hypothesis Testing, Significance, and Power.

Prediction intervals for individual measurements

Prediction intervals are different from confidence intervals, though related to them. Prediction intervals specify the range that a new individual measurement is expected to fall within, if it comes from the same distribution as the previous measurements. The uncertainty in any individual measurement comes from two sources: 1) its variability about the true mean (estimated by the standard deviation s_x) and 2) the uncertainty in the location of the mean itself (estimated by the standard error, $s_{\bar{x}}$). When these two sources of uncertainty are combined, the probability is 1-α that a new measurement x will fall within the range,

$$\bar{x} - t_{\alpha/2,\nu} \sqrt{{s_x}^2 + {s_{\bar{x}}}^2} < \mu < \bar{x} + t_{\alpha/2,\nu} \sqrt{{s_x}^2 + {s_{\bar{x}}}^2}$$

This can be simplified to:

$$\bar{x} - t_{\alpha/2,\nu} s_x \sqrt{1 + \frac{1}{n}} < \mu < \bar{x} + t_{\alpha/2,\nu} s_x \sqrt{1 + \frac{1}{n}}$$

Assumptions: that *x* is normally distributed. Note that because prediction intervals are concerned with the distribution of individual *x* values rather than means, the Central Limit Theorem is no help here; the individual *x* values themselves must be approximately normal.

Reference: Helsel, D. R. and R. M. Hirsch, *Statistical Methods in Water Resources*, 522 pp., Elsevier, 1992 (Chapter 3).