Confidence Interval

A **population parameter** is a number that describes a fixed, but unknown characteristic of an entire group or population. For example, the mean height of all adult males in a country is a population parameter.

An **interval estimate** for a **population parameter** is called a **confidence interval**.

We cannot be certain that the interval contains the true, unknown population parameter— as we only use a sample from the full population to compute the point estimate and the interval.

However, the confidence interval is constructed so that we have high confidence that it does contain the unknown population parameter.

Confidence intervals are widely used in engineering and the sciences.

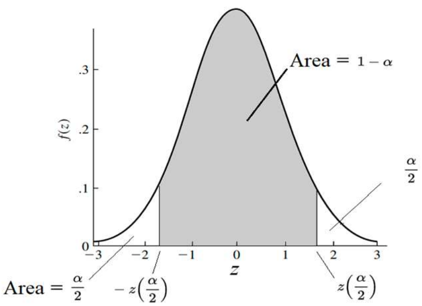
The basic ideas of a confidence interval (CI) are most easily understood by initially considering a simple situation. Suppose that we have a normal population with unknown mean μ and known variance σ^2.

This is a somewhat unrealistic scenario because typically we know the distribution mean before we know the variance. However, later in the lecture, we will present confidence intervals for more general situations.

Suppose that $X_1, X_2, \ldots, X_n$ is a random sample from a normal distribution with unknown mean μ and known variance σ^2. From the results of Chapter 5 we know that the sample mean $\overline{X}$ is normally distributed with mean μ and variance σ^2/n. We may standardize $\bar{X}$ by subtracting the mean and dividing by the standard deviation, which results in the variable

$$Z = \frac{\overline{X} - \mu}{\sigma/\sqrt{n}}$$

Now Z has a standard normal distribution.



$$Z = \frac{\overline{X} - \mu}{\sigma/\sqrt{n}}$$

A confidence interval estimate for μ is an interval of the form $l \le \mu \le u$, where the endpoints l and u are computed from the sample data. Because different samples will produce different values of l and u, these end-points are values of random variables L and U, respectively. Suppose that we can determine values of L and U such that the following probability statement is true:

$$P[L \le \mu \le U] = 1 - \alpha$$

where $0 \le \alpha \le 1$.

So, there is a probability of $1 - \alpha$ of selecting a sample for which the CI will contain the true value of μ.

Once we have selected the sample, so that $X_1 = x_1, X_2 = x_2, \ldots, X_n = x_n$, and computed l and u, the resulting confidence interval for μ is

$$l \leq \mu \leq u$$

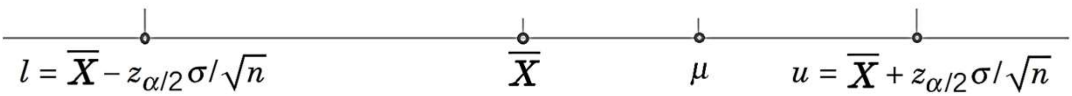
The end-points or bounds l and u are called the lower- and upper-confidence limits, respectively, and $1 - \alpha$ is called the confidence coefficient.

In our problem situation, because $Z = (\overline{X} - \mu)/(\sigma/\sqrt{n})$ has a standard normal distribution, we may write

$$P\left[-z_{\alpha/2} \leq \frac{\bar{X} - \mu}{\sigma/\sqrt{n}} \leq z_{\alpha/2}\right] = 1 - \alpha$$

Now manipulate the quantities inside the brackets by multiplying through by $\sigma/\sqrt{n}$, subtracting $\overline{X}$ from each term, and multiplying through by -1. This results in

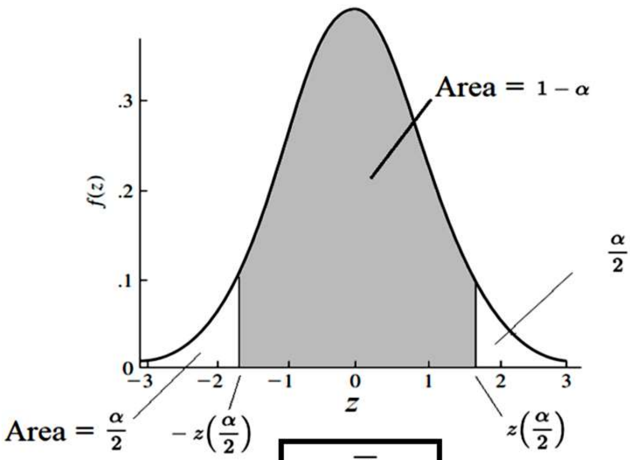
$$P\left[\overline{X} - z_{\alpha/2}\frac{\sigma}{\sqrt{n}} \leq \mu \leq \overline{X} + z_{\alpha/2}\frac{\sigma}{\sqrt{n}}\right] = 1 - \alpha$$



If $\overline{X}$ is the sample mean of a random sample of size n from a normal population with known variance σ^2, a $100(1 - \alpha)\%$CI on μ is given by

$$\overline{X} - z_{\alpha/2}\sigma/\sqrt{n} \leq \mu \leq \overline{X} + z_{\alpha/2}\sigma/\sqrt{n}$$

where $z_{\alpha/2}$ is the upper $100\alpha/2$ percentage point of the standard normal distribution.



$$\boxed{Z = \frac{\overline{X} - \mu}{\sigma/\sqrt{n}}}$$

Example 1

ASTM Standard E23 defines standard test methods for notched bar impact testing of metallic materials. The Charpy V-notch (CVN) technique measures impact energy and is often used to determine whether or not a material experiences a ductile-to-brittle transition with decreasing temperature. Ten measurements of impact energy (J) on specimens of A238 steel cut at 60°C are as follows: $64.1, 64.7, 64.5, 64.6, 64.5, 64.3, 64.6, 64.8, 64.2$, and 64.3. Assume that impact energy is normally distributed with $\sigma = 1$ J. We want to find a 95% CI for μ, the mean impact energy.

The required quantities are $z_{\alpha/2} = z_{0.025} = 1.96, n = 10, \sigma = 1$, and $\overline{X} = 64.46$.
The resulting 95%CI is found from Equation 8-7 as follows:

$$\overline{X} - z_{\alpha/2}\frac{\sigma}{\sqrt{n}} \leq \mu \leq \overline{X} + z_{\alpha/2}\frac{\sigma}{\sqrt{n}}$$

$$64.46 - 1.96\frac{1}{\sqrt{10}} \leq \mu \leq 64.46 + 1.96\frac{1}{\sqrt{10}}$$

$$63.84 \leq \mu \leq 65.08$$

Critical Values, $P(Z \geq z_\alpha) = \alpha$

α	.10	.05	.025
z_α	1.2816	1.6449	1.9600

That is, based on the sample data, a range of highly plausible vaules for mean impact energy for A238 steel at 60°C is $63.84 \text{ J} \leq \mu \leq 65.08$ J.

Confidence interval length = $65.08 - 63.84 = 1.24$.

Interpreting Confidence Interval

How does one interpret a confidence interval?

In the Example (last slide), the 95%CI is $63.84 \leq \mu \leq 65.08$, so it is tempting to conclude that μ is within this interval with probability 0.95.

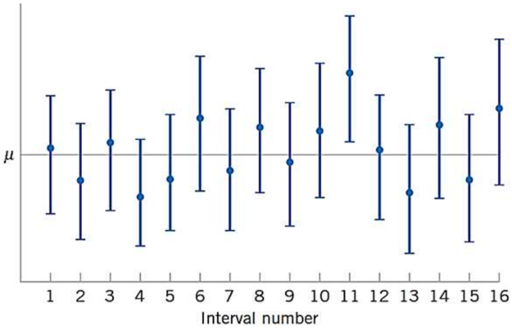
However, with a little reflection, it's easy to see that this cannot be correct; the true value of μ is unknown and...

the statement $63.84 \leq \mu \leq 65.08$ is either correct (true with probability 1)...

or incorrect (false with probability 1).

The correct interpretation lies in the realization that a CI is a random interval because in the probability statement defining the end-points of the interval, L and U are random variables.

Consequently, the correct interpretation of a $100(1-\alpha)\%$ CI depends on the relative frequency view of probability. Specifically, if an infinite number of random set of samples are collected and a $100(1-\alpha)\%$ confidence interval for μ is computed from each set, $100(1-\alpha)\%$ of these intervals will contain the true value of μ.



Repeated construction of a confidence interval of μ

The situation is illustrated in the figure, which shows several $100(1-\alpha)\%$ confidence intervals for the mean μ of a normal distribution. The dots at the center of the intervals indicate the point estimate of μ (that is, $\overline{X}$).

Notice that one of the intervals **fails** to contain the **true** value of μ. If this were a 95% confidence interval, in the long run only 5% of the intervals would fail to contain μ.

In practice, we obtain only one random sample and calculate one confidence interval.

Since this interval either will or will not contain the true value of μ, it is not reasonable to attach a probability level to this specific event. The appropriate statement is the observed interval $[l, u]$ brackets the true value of μ with confidence $100(1-\alpha)$.

This statement has a frequency interpretation; that is, we don't know if the statement is true for this specific sample, but the method used to obtain the interval $[l, u]$ yields correct statements $100(1-\alpha)\%$ of the time.

Confidence Level and Precision of Estimation

Notice in the Example that our choice of the 95% level of confidence was essentially arbitrary. What would have happened if we had chosen a higher level of confidence, say, 99% ?

In fact, doesn't it seem reasonable that we would want the higher level of confidence?

At $\alpha = 0.01$, we find $z_{\alpha/2} = z_{0.01/2} = z_{0.005} = 2.58$, while for $\alpha = 0.05$, $z_{0.025} = 1.96$.
Thus, the length of the 95% confidence interval is

$$2(1.96\sigma/\sqrt{n}) = 3.92\sigma/\sqrt{n}$$

whereas the length of the 99%CI is

$$2(2.58\sigma/\sqrt{n}) = 5.16\sigma/\sqrt{n}$$

Thus, the 99% CI is longer than the 95% CI.

This is why we have a higher level of confidence in the 99% confidence interval. Generally, for a fixed sample size n and standard deviation σ, the higher the confidence level, the longer the resulting CI.

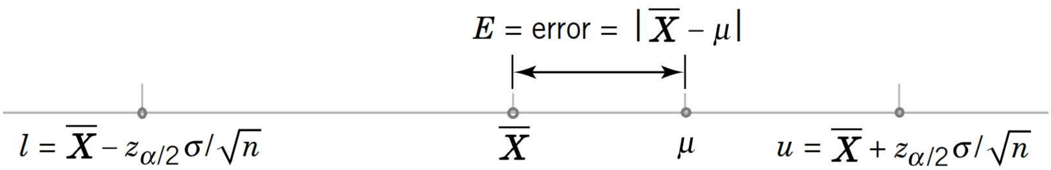
Confidence Level and Precision of Estimation

The length of a confidence interval is a measure of the **precision** of estimation. **Precision** tells how close is the esimate $\overline{X}$ to the true value μ.

From the preceeding discussion, we see that precision is **inversely** related to the confidence level.

It is desirable to obtain a confidence interval that is **short** enough for decision-making purposes and that also has adequate confidence. One way to achieve this is by choosing the sample size n to be large enough to give a CI of specified length or precision with prescribed confidence.

Choice of Sample Size



The confidence interval is $2z_{\alpha/2}\sigma/\sqrt{n}$. This means that in using $\overline{X}$ to estimate μ, the error $E = \left|\overline{X} - \mu\right|$ is less than or equal to $z_{\alpha/2}\sigma/\sqrt{n}$ with confidence $100(1-\alpha)$.

This is shown graphically in the figure. In situations where the sample size can be controlled, we can choose n so that we are $100(1-\alpha)\%$ confident that the error in estimating μ is less than a specified bound on the error E. The appropriate sample size is found by choosing n such that $z_{\alpha/2}\sigma/\sqrt{n} = E$.

Solving this equation gives the following formula for n as follows:

$$n = \left(\frac{z_{\alpha/2}\sigma}{E}\right)^2$$

Example 2

To illustrate the use of this procedure, consider the Example 1, and suppose that we wanted to determine how many specimens must be tested to ensure that the 95%CI on μ for A238 steel cut at 60°C has a length of at most 1.0 J.

Since the bound on error in estimation E is one-half of the length of the CI, to determine n we use Equation 8-8 with $E = 0.5, \sigma = 1$, and $z_{\alpha/2} = 0.025$. The required sample size is 16

$$n = \left(\frac{z_{\alpha/2}\sigma}{E}\right)^2 = \left(\frac{(1.96)1}{0.5}\right)^2 = 15.37$$

and because n must be an integer, the required sample size is $n = 16$.

$$\text{Confidence level} = 1 - \alpha = P\left[\overline{X} - z_{\alpha/2}\frac{\sigma}{\sqrt{n}} \leq \mu \leq \overline{X} + z_{\alpha/2}\frac{\sigma}{\sqrt{n}}\right]$$

$$\text{Confidence interval} = 2E = 2z_{\alpha/2}\frac{\sigma}{\sqrt{n}}$$

Notice the general relationship between sample size n, desired length of the confidence interval $2E$, confidence level $100(1 - \alpha)$, and standard deviation σ:

- As the desired length of the interval $2E$ decreases, the required sample size n increases for a fixed value of σ and specified confidence.

- As σ increases, the required sample size n increases for a fixed desired length $2E$ and specified confidence.

- As the level of confidence increases, the required sample size n increases for fixed desired length $2E$ and standard deviation σ.

One-Sided Confidence Bounds

The confidence interval, as discussed earlier, gives both a lower confidence bound and an upper confidence bound for μ. Thus it provides a two-sided CI. It is also possible to obtain one-sided confidence bounds for μ by setting either $l = -\infty$ or $u = \infty$ and replacing $z_{\alpha/2}$ by z_α.

A $100(1-\alpha)\%$ **upper-confidence bound** for μ is

$$\mu \leq u = \overline{X} + z_\alpha \, \sigma/\sqrt{n}$$

and a $100(1-\alpha)\%$ **lower-confidence bound** for μ is

$$\overline{X} - z_\alpha \, \sigma/\sqrt{n} = l \leq \mu$$

A Large-Sample Confidence Interval for $\boldsymbol{\mu}$ and Unknown Variance

Earlier, we have assumed that the population distribution is normal with unknown mean and known standard deviation σ (or variance σ^2). We now present a large-sample **CI** and μ that does not require these assumptions.

Let $X_1, X_2, \ldots, X_n$ be a random sample from a population with unknown mean μ and variance σ^2. Now if the sample size n is large ($n \geq 40$), the central limit theorem implies that $\widehat{\mu} = \overline{X}$ has approximately a normal distribution with mean μ and variance σ^2/n.

Therefore $Z = (\overline{X} - \mu)/(\sigma/\sqrt{n})$ has approximately a standard normal distribution. This ratio could be used as a pivotal quantity and manipulated as in Section 8-2.1 to produce an approximate CI for μ.

However, the standard deviation σ is unknown. It turns out that when n is large, replacing σ by the sample standard deviation S has little effect on the distribution of Z. This leads to the following useful result.

When n is large ($n \geq 40$), the quantity

$$\frac{\overline{X} - \mu}{S/\sqrt{n}}$$

has an approximate standard normal distribution. Consequently,

$$\overline{X} - z_{\alpha/2}\frac{S}{\sqrt{n}} \leq \mu \leq \overline{X} + z_{\alpha/2}\frac{S}{\sqrt{n}}$$

is a **large sample confidence interval** for μ, with confidence level of approximately $100(1 - \alpha)\%$

For large samples, this holds regardless of the shape of the population distribution. Generally n should be at least 40 to use this result reliably. The central limit theorem generally holds for $n \geq 30$, but the larger sample size is usually recommended.

If $X_1, X_2, \ldots, X_n$ is a sample of n observations, the sample variance is

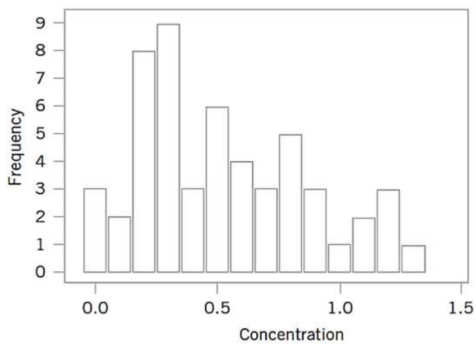
$$S^2 = \frac{\sum_{i=1}^{n} \left(X_i - \overline{X}\right)^2}{n - 1}$$

The sample standard deviation, S, is the positive square root of the sample variance.

Example 3

An article in the 1993 volume of the *Transactions of the American Fisheries Society* reports the results of a study to investigate the mercury contamination in largemouth bass. A sample of fish was selected from 53 Florida lakes and mercury concentration in the muscle tissue was measured (ppm). The mercury concentration values are

1.230	0.490	0.490	1.080	0.590	0.280	0.180	0.100	0.940
1.330	0.190	1.160	0.980	0.340	0.340	0.190	0.210	0.400
0.040	0.830	0.050	0.630	0.340	0.750	0.040	0.860	0.430
0.044	0.810	0.150	0.560	0.840	0.870	0.490	0.520	0.250
1.200	0.710	0.190	0.410	0.500	0.560	1.100	0.650	0.270
0.270	0.500	0.770	0.730	0.340	0.170	0.160	0.270	



Example 3

We want to find an approximate 95% CI on μ. Note $n > 40$.

The required quantities are $n = 53, \overline{X} = 0.5250, S = 0.3486$, and $z_{0.025} = 1.96$. The approximate 95%CI on μ is

$$\overline{X} - z_{0.025}\frac{S}{\sqrt{n}} \leq \mu \leq \overline{X} + z_{0.025}\frac{S}{\sqrt{n}}$$

$$0.5250 - 1.96\frac{0.3486}{\sqrt{53}} \leq \mu \leq 0.5250 + 1.96\frac{0.3486}{\sqrt{53}}$$

$$0.4311 \leq \mu \leq 0.6189$$

This interval is fairly wide because there is a lot of variability in the mercury concentration measurements.

Example 4

A civil engineer is analyzing the compressive strength of concrete. Compressive strength is normally distributed with $\sigma^2 = 1000(\text{psi})^2$. A random sample of 12 specimens has a mean compressive strength of $\overline{X} = 3250\text{psi}$.

(a) Construct a 95% two-sided confidence interval on mean compressive strength.

(b) Construct a 99% two-sided confidence interval on mean compressive strength. Compare the width of this confidence interval with the width of the one found in part (a).

$3232.11 \leq \mu \leq 3267.89$

$3226.5 \leq \mu \leq 3273.5$

A Small-Sample Confidence Interval for $\boldsymbol{\mu}$ and Unknown Variance

Suppose that the population of interest has a normal distribution with unknown mean μ and unknown variance σ^2.

Assume that a random sample of size n, say $X_1, X_2, \ldots, X_n$, is available, and let $\overline{X}$ and S^2 be the sample mean and variance, respectively.

We wish to construct a two-sided CI on μ. If the variance σ^2 is known, we know that $Z = (\overline{X} - \mu)/(\sigma/\sqrt{n})$ has a standard normal distribution; we use Z-statistics.

When σ^2 is unknown, and the sample size is large, i.e., $n \geq 40$, we may still use Z-statistics. But if $n < 40$, we need to do something else.

We first realize that $T = (\overline{X} - \mu)/(S/\sqrt{n})$ is a t-distributed random variable with $n - 1$ degrees of freedom.

The t probability density function is

$$f(x) = \frac{\Gamma[(k+1)/2]}{\sqrt{\pi k}\,\Gamma(k/2)} \cdot \frac{1}{\left[x^2/k + 1\right]^{(k+1)/2}} \qquad -\infty < x < \infty$$

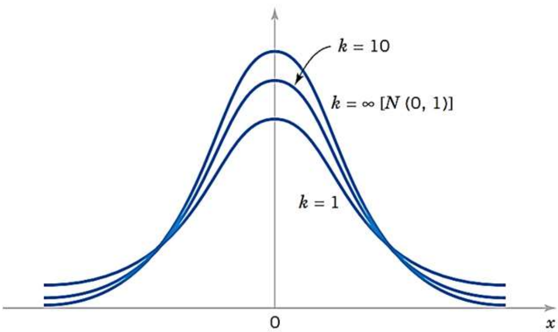
where k is the number of degrees of freedom.

The mean and variance of the t distribution are zero and $k/(k-2)$, respectively, for $k > 2$.

Several t distributions are shown in the figure. The general appearance of the t distribution is similar to the standard normal distribution in that both distributions are symmetric and unimodal.

However, the t distribution has heavier tails than the normal; that is, it has more probability in the tails than the normal distribution.

As the number of degrees of freedom $k \to \infty$, the limiting form of the t distribution is the standard normal distribution.



We know that the distribution of $T = (\overline{X} - \mu)/(S/\sqrt{n})$ is t with $n - 1$ degrees of freedom. Letting $t_{\alpha/2,n-1}$ be the upper $100\alpha/2$ percentage point of the t distribution with $n - 1$ degrees of freedom, we may write:

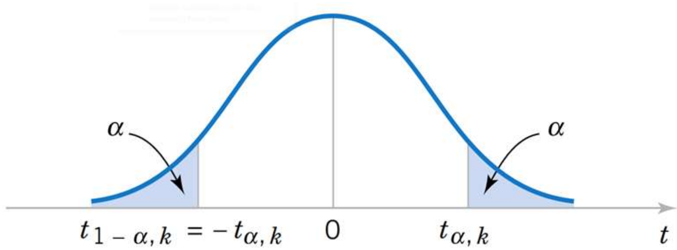
$$P(-t_{\alpha/2,n-1} \le T \le t_{\alpha/2,n-1}) = 1 - \alpha$$

or

$$P\left(-t_{\alpha/2,n-1} \le \frac{\overline{X} - \mu}{S/\sqrt{n}} \le t_{\alpha/2,n-1}\right) = 1 - \alpha$$

Rearranging this last equation yields

$$P\left(\overline{X} - t_{\alpha/2,n-1}S/\sqrt{n} \le \mu \le \overline{X} + t_{\alpha/2,n-1}S/\sqrt{n}\right) = 1 - \alpha$$



Example

An article in the journal *Materials Engineering* (1989, Vol. II, No. 4, pp. 275–281) describes the results of tensile adhesion tests on 22 U-700 alloy specimens. The load at specimen failure is as follows (in megapascals):

19.8	10.1	14.9	7.5	15.4	15.4
15.4	18.5	7.9	12.7	11.9	11.4
11.4	14.1	17.6	16.7	15.8	
19.5	8.8	13.6	11.9	11.4	

The sample mean is $\bar{x} = 13.71$, and the sample standard deviation is $s = 3.55$. We want to find a 95% CI on μ. Since $n = 22$, we have $n - 1 = 21$ degrees of freedom for t, so $t_{0.025,21} = 2.080$. The resulting CI is

$$\overline{X} - t_{\alpha/2,n-1}s/\sqrt{n} \le \mu \le \overline{X} + t_{\alpha/2,n-1}s/\sqrt{n}$$
$$13.71 - 2.080(3.55)/\sqrt{22} \le \mu \le 13.71 + 2.080(3.55)/\sqrt{22}$$
$$13.71 - 1.57 \le \mu \le 13.71 + 1.57$$
$$12.14 \le \mu \le 15.28$$