

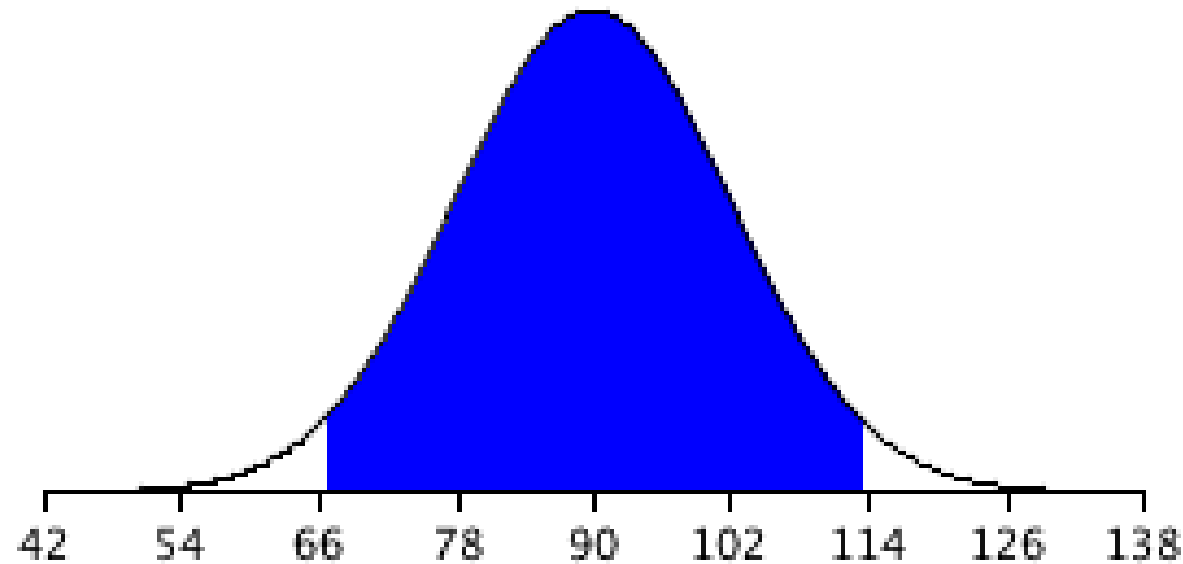
Confidence limits

Normal distribution

The Gaussian function

95%-confidence limit

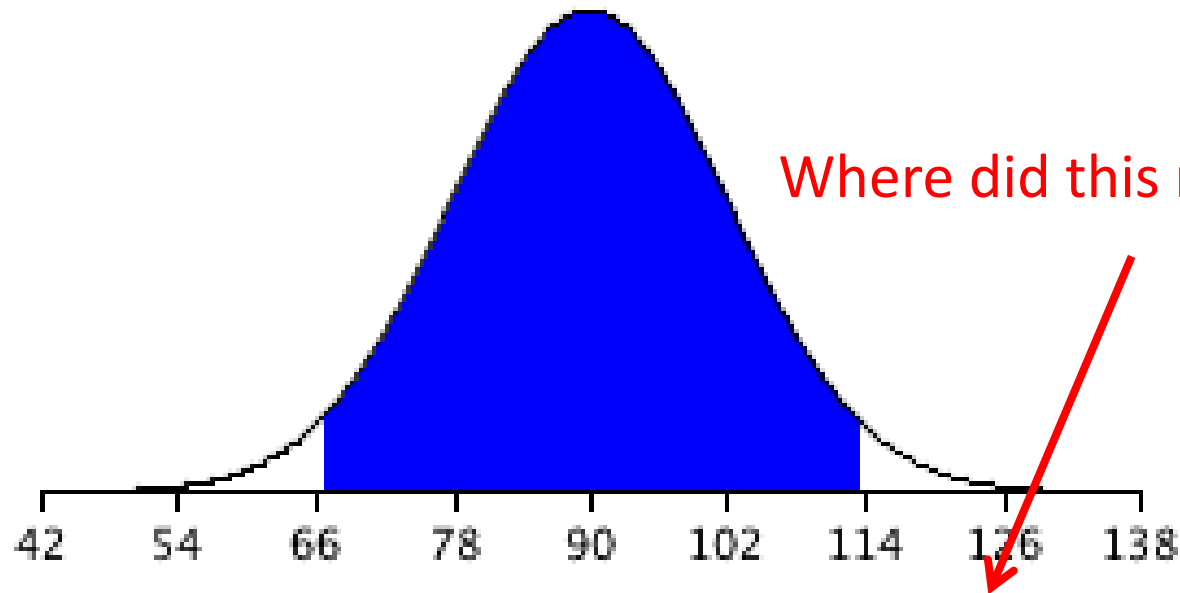
The 95% confidence limit is defined in terms of the area underneath the Gaussian (normal) distribution function. 95% of the population will be observed within the limits set by this definition.



$$95\% \text{ confidence} = \langle x \rangle \pm 1.96 \frac{\sigma}{\sqrt{N}}$$

95%-confidence limit

The 95% confidence limit is defined in terms of the area underneath the Gaussian (normal) distribution function. 95% of the population will be observed within the limits set by this definition.



$$95\% \text{ confidence} = \langle x \rangle \pm \boxed{1.96} \frac{\sigma}{\sqrt{N}}$$

The Gauss Error Function

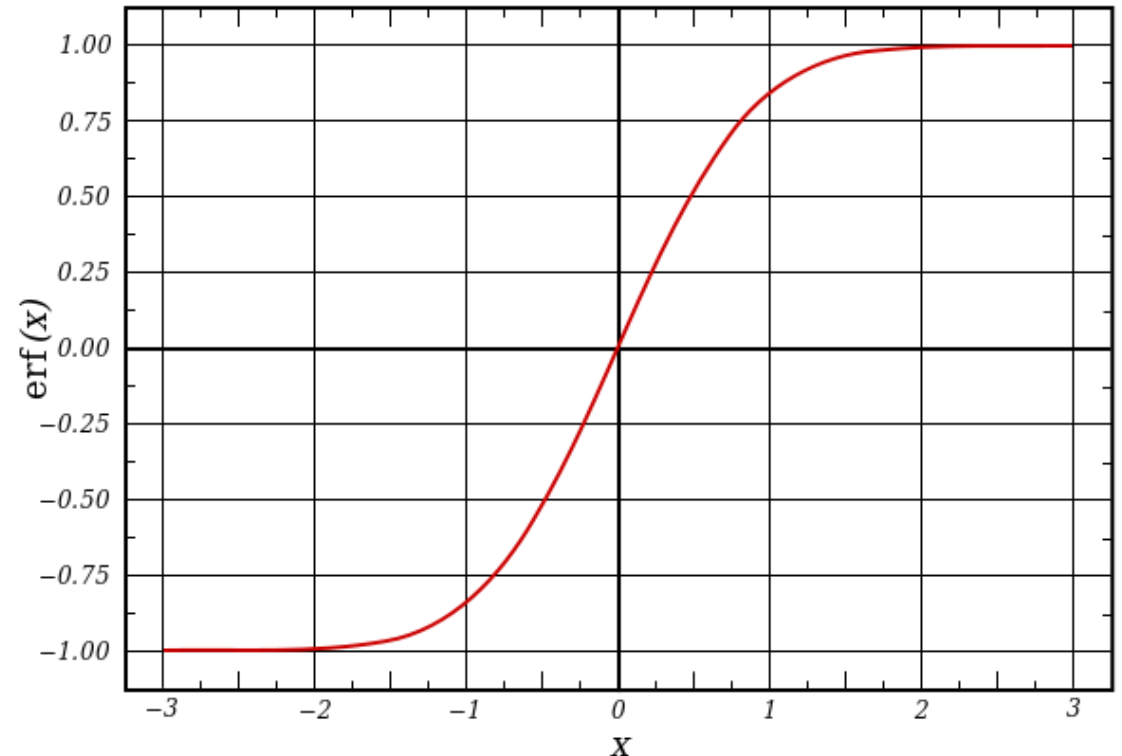
The Gauss error function is a special function that occurs in statistics (and partial differential equations describing diffusion). It is defined as:

$$\operatorname{erf}(x) = \frac{1}{\sqrt{\pi}} \int_{-x}^x e^{-t^2} dt$$

The Gauss error function gives the area under a curve with limits $+x$ and $-x$. We can use the $\operatorname{erf}(x)$ to calculate a 95% confidence limit by solving

$$\operatorname{erf}(x) = 0.95$$

This can be done numerically using a series expansion of the defined Gaussian.



Using the erf(x) as a reference for smaller samples

The Gauss error function assumes that the sample size is large enough that we can ignore an contributions from N (the number of samples in the set). The point of t-test is that we cannot ignore N when the sample size is small.

So for an ideal Gaussian we find $\text{erf}(1.96) = 0.95$, which explains why the width of the Gaussian is $+ \text{ or } - 1.96 \sigma/N^{1/2}$. When the sample size is small the Gaussian width will tend to be larger than the ideal. This reflects the greater uncertainty in the measurement. There are many cases one can study involving application of this method to more than one type of measurement, measurement of lines etc. However, if we simply consider a single measurement then we can ask the question:

Does the measured average differ from the assumed value (null hypothesis) with greater than 95% probability? Does it fall outside the 95% confidence limit. This corresponds to $p < 0.05$.

The Student's *t*-test

Student's *t*-test deals with the problems associated with inference based on "small" samples: the calculated mean ($\langle x \rangle$) and standard deviation (σ) may by chance deviate from the "real" mean and standard deviation (i.e., what you'd measure if you had many more data items: a "large" sample). For example, it is likely that the true mean size of maple leaves is "close" to the mean calculated from a sample of N randomly collected leaves.

The 95% confidence interval is:

If $N=5$, $\langle x \rangle \pm 2.776 \sigma/N^{1/2}$
if $N=10$: $\langle x \rangle \pm 2.262 \sigma/N^{1/2}$
if $N=20$: $\langle x \rangle \pm 2.093 \sigma/N^{1/2}$
if $N=40$; $\langle x \rangle \pm 2.023 \sigma/N^{1/2}$
for "large" N : $\langle x \rangle \pm 1.960 \sigma/N^{1/2}$.

There are tables you can use to look up the *t*-value appropriate for your problem.

The null hypothesis

Statistics is a branch of science that is concerned with reliability or verifiability of information. Modern statistics is attributable to Ronald Fischer who studied genetics in the early 1900s. Fischer was able to use statistics to show that Gregor Mendel cherry picked his data. The basic conclusions are correct, but Mendel made simplifying assumptions about how alleles are expressed. Then he threw out data that did not agree with those assumptions so he would get a “perfect” result. Mendel was a monk. We can imagine many reasons that he would think that this was justified. But, statistics proves that what he did was not justified from a scientific point of view. Fischer wanted to put science on a firm basis so he came up with the idea of a null hypothesis.

The null hypothesis is the assumption that an experiment will not give a new result. A result is “new” if it differs significantly from the null hypothesis.

An example of a null hypothesis

A medical researcher wants to compare the effectiveness of two medications. The null and alternative hypotheses are:

Null hypothesis (H_0): $\mu_1 = \mu_2$ The two medications are equally effective.

Alternative hypothesis (H_1): $\mu_1 \neq \mu_2$ The two medications are not equally effective.

In a clinical study we need to show that a result is statistically meaningful. When patients are given a new medication they often will feel better simply believing that the medication is helping them (the placebo effect). Thus, the method used is to divide the patients into two groups: the clinical group and the placebo (null) group. Then test for effects and use statistics to see whether the clinical group (receiving the new medication) is actually better off.

p-test = significance testing

Significance testing is an important aspect of statistical analysis. The idea is to make an initial hypothesis (the null hypothesis), then to use statistical observations to test the hypothesis. The p-test is a measure of how far one is off the peak of a normal distribution. If the statistical value is near the peak of the normal distribution (i.e. near the average) then $p = 1$ as one moves further away from the average the p value decreases until it approaches 0. You could think of the p-value as the fractional probability that the null hypothesis is valid. If $p < 0.05$ then we can say that there is less than a 5% chance that the null hypothesis is valid. At this level of significance we could say that the null hypothesis is invalid. This is related to the confidence limit

$$p = 1 - \text{fractional confidence limit}$$

Specific example: treating cancer tumors

In cancer treatments there is an objective measure of success for medical treatment, which is whether a tumor grows or shrinks. If two groups patients have similar tumors and they are given medication and placebo the sizes of their tumors can be compared over time. In our example the average tumor has a initial radius of 1.0 cm (± 0.36 cm) where the 0.4 is the standard deviation. Under normal treatment (with placebo) the tumors shrink to an average of 0.80 cm (± 0.36 cm) . The researcher can compare this value (and distribution!) to the treatment plus the new medication to see whether the tumors are significantly smaller. If, for example, the tumors treated with the medication are 0.64 (± 0.36 cm) would we conclude that this is significant (i.e. that $p < 0.05$)? We need to know the sample size. This will make a large difference in our conclusion.

Specific example: treating 16 patients

If the sample size is $N = 16$ patients then the standard error is

$$\sigma_{se} = \frac{\sigma}{\sqrt{N}} = \frac{\sigma}{4}$$

Under normal treatment the tumors have 0.80 cm (± 0.09 cm) .

The experimental treatment group has 0.64 cm (± 0.09 cm).

The tumors in the experimental group are 0.16 cm smaller.

The standard error should be multiplied by p-value of 2.12 (obtained from a table using the entry at $N = 16$). Thus, if the tumor radius differs by more than $2.12(0.09) = 0.19$ cm then we will consider the result significant. But, it does not differ by that much since the tumors are only smaller by 0.16 cm. If you believe that the drug is working then you should do the study again.

Specific example: treating 32 patients

If the sample size is $N = 32$ patients then the standard error is

$$\sigma_{se} = \frac{\sigma}{\sqrt{N}} = \frac{\sigma}{5.65}$$

Under normal treatment the tumors have 0.80 cm (± 0.063 cm) .

The experimental treatment group has 0.64 cm (± 0.063 cm).

The tumors in the experimental group are 0.16 cm smaller.

The standard error should be multiplied by p-value of 2.03 (obtained from a table using the entry at $N = 32$). Thus, if the tumor radius differs by more than $2.03(0.063) = 0.13$ cm then we will consider the result significant. In this case the tumors shrank by 0.16 cm, which is greater than the confidence limit and therefore we now conclude that $p < 0.05$.