The great importance of the \( t \) distribution in data analysis lies in the existence of numerous tests based upon it, such as the 1-sample \( t \), unpaired \( t \), and paired \( t \), as well as the use in calculating confidence intervals.

1 Definitions

Consider two independent random variables, \( Z \) which has a normal distribution with \( \mu = 0 \), \( \sigma^2 = 1 \), and \( C \) which has a chi-square distribution with \( k \) degrees of freedom. Then the ratio

\[
t_k = \frac{Z}{\sqrt{C/k}}
\]

is described as a \( t \) variable with \( k \) degrees of freedom. It should be noted incidentally that \( t_k^2 \) is distributed as \( F(1, k) \).

A special case arises when analyzing a sample of size \( n \) from a normal distribution with population mean \( \mu \) and population variance \( \sigma^2 \), because the sample mean

\[
\bar{x} = \frac{1}{n} \sum_{i=1}^{n} x_i
\]

is normally distributed with mean \( \mu \) and variance \( \sigma^2/n \), while \( nS^2/\sigma^2 \) using

\[
S^2 = \frac{1}{n} \sum_{i=1}^{n} (x_i - \bar{x})^2
\]

has a \( \chi^2 \) distribution with \( n - 1 \) degrees of freedom. Hence the statistic

\[
t_{n-1} = \frac{\bar{x} - \mu}{S/\sqrt{n-1}}
\]

has a \( t \) distribution with \( n - 1 \) degrees of freedom. Note that this \( t \) variable only has one unknown parameter, the population mean \( \mu \).

2 Simfit program ttest

Choose [A/Z] from the main Simfit menu and open program ttest when the following options will be available.
3 Degrees of freedom

An important use of the \( t \) distribution is when calculating confidence limits, for instance with a sample mean, or parameter estimate. The main thing to realize in such circumstances is that, although the mean value for \( t_n \) is zero irrespective of \( n \), the variance is heavily dependent on \( n \). This is why the confidence limits shrink as the sample size increases. Actually the \( t_n \) distribution is asymptotic to a standardized normal distribution as \( n \) increases, as shown by the next graph created from ttest.

Note how the area under the tails decreases rapidly as \( n \) increases from 2 to 6 but less slowly.
thereafter. A more detailed inspection of this will be clear from this table copied from the `ttest` results log file for a 95% confidence interval.

\[
\begin{align*}
P(t =< 4.303E+00) &= 0.975 *** P(t >= 4.303E+00) = 0.025, N = 2 \\
P(t =< 2.776E+00) &= 0.975 *** P(t >= 2.776E+00) = 0.025, N = 4 \\
P(t =< 2.447E+00) &= 0.975 *** P(t >= 2.447E+00) = 0.025, N = 6 \\
P(t =< 2.306E+00) &= 0.975 *** P(t >= 2.306E+00) = 0.025, N = 8 \\
P(t =< 2.228E+00) &= 0.975 *** P(t >= 2.228E+00) = 0.025, N = 10
\end{align*}
\]

4 Confidence range for the sample mean

Given \( \bar{x} \) and \( S^2 \) from a sample of size \( n \), then a symmetrical 100\((1 - \alpha)\)% confidence range for the population mean \( \mu \) can be constructed using the upper tail critical value \( t_{\alpha/2,n-1} \). We have that

\[
P \left( \frac{\bar{x} - \mu}{S/\sqrt{n-1}} \geq t_{\alpha/2,n-1} \right) = \alpha/2
\]

and

\[
P \left( \frac{\bar{x} - \mu}{S/\sqrt{n-1}} \leq -t_{\alpha/2,n-1} \right) = \alpha/2,
\]

so that

\[
P \left( \bar{x} - t_{\alpha/2,n-1}S/\sqrt{n-1} \leq \mu \leq \bar{x} + t_{\alpha/2,n-1}S/\sqrt{n-1} \right) = 1 - \alpha.
\]

Alternatively, note that it often causes confusion because an unbiased estimate of the population variance is not \( S^2 \) but the sample variance

\[
s^2 = \frac{1}{n-1} \sum_{i=1}^{n} (x_i - \bar{x})^2,
\]

so that an equivalent expression for \( t_{n-1} \) would then be

\[
t_{n-1} = \frac{\bar{x} - \mu}{s/\sqrt{n}}.
\]

whereupon

\[
P \left( \bar{x} - t_{\alpha/2,n-1}s/\sqrt{n} \leq \mu \leq \bar{x} + t_{\alpha/2,n-1}s/\sqrt{n} \right) = 1 - \alpha.
\]

using \( s^2 \) instead of \( S^2 \).

We see from the above table that the multipliers of the sample standard error required for a 95% confidence interval with sample sizes of \( n = 3, 5, 7, 9, \) and 11 would be 4.303, 2.776, 2.447, 2.306, and 2.228. Clearly, using the sample mean plus or minus twice the standard error as an approximate 95% confidence range will always underestimate the actual 95% confidence range unless the sample size exceeds 10, say.