Simfit

Tutorials and worked examples for simulation, curve fitting, statistical analysis, and plotting.
http://www.simfit.org.uk

The paired t test should not be viewed as an alternative to the unpaired t test. It is used to see if it is reasonable to conclude that the mean of the differences between two sets of dependent observations is zero in the populations, and the null hypothesis is based on the following assumptions.

- The pairwise differences are normally distributed
- The mean of the pairwise differences is zero
- The sample sizes are equal and greater than 1 (and preferably very much greater)
- The samples are pairwise dependent, i.e. highly correlated

For instance, the columns could be repeated observations of blood pressure on the same set of subjects, before, then after treatment.

The paired t test is equivalent to repeated analysis of variance (ANOVA) with just two columns if the samples are normally distributed with the same variance. Alteratively, as the test is only examining the hypothesis that a single sample of differences is normally distributed with zero mean, and variance estimated from the sample, it can be viewed as a one sample t test.

To be precise, the user has two samples (i.e. vectors X and Y) with n observations

$$X = (x_1, x_2, \ldots, x_n)$$
$$Y = (y_1, y_2, \ldots, y_n)$$

and wishes to test the null hypothesis that the samples have the same zero differences, against the alternative hypothesis that they are not equal, or possibly the one-sided alternatives. That is

$$H_0 : \ x_i = y_i$$
$$H_1 : \ x_i \neq y_i$$
$$H_2 : \ x_i > y_i$$
$$H_3 : \ x_i < y_i$$

and SIMFIT provides all the information that is required to perform such tests.

It is important to emphasize the difference between a paired and an unpaired t test, so that it will not be mistakenly assumed that the choice between them is arbitrary. For instance, given two random variables X and Y, then the expectation and variance of the difference is as follows.

$$E(X - Y) = E(X) - E(Y)$$
$$V(X - Y) = V(X) + V(Y) - 2CV(X, Y)$$

When the samples are uncorrelated the covariance $CV(X, Y)$ would be zero, so that the variance of the difference only depends on the variance of X and Y. However, the paired hypothesis under consideration would be consistent with a strong positive correlation between X and Y, so that the covariance term would make the variance of the difference smaller.

In the case of the unpaired t test, the test statistic is

$$U = \frac{\bar{x} - \bar{y}}{\sqrt{s_p^2\left(\frac{1}{m} + \frac{1}{n}\right)}},$$

which is t distributed with $m + n - 2$ degrees of freedom under the unpaired t test null hypothesis. Here $\bar{x}$ and $\bar{y}$ are sample means, and the pooled estimate of variance is expressed in terms of the independent sample variance estimates s_x^2 and s_y^2 as

$$s_p^2 = \frac{(m-1)s_x^2 + (n-1)s_y^2}{m+n-2}.$$

The paired t test is quite different. It uses the paired differences d_i, the mean difference $\hat{d}$, and the variance estimate for the differences s_d^2 to define the test statistic t_d defined as follows

$$\begin{aligned} d_i &= x_i - y_i \\ \bar{d} &= \sum_{i=1}^{n} d_i/n \\ s_d^2 &= \sum_{i=1}^{n} (d_i - \bar{d})^2/(n-1) \\ t_d &= \frac{\bar{d}}{\sqrt{s_d^2/n}} \end{aligned}$$

which is t distributed with $n - 1$ degrees of freedom under the paired t test null hypothesis. It therefore makes allowances for the strong correlation between the two samples.

From the main SimFIT menu select [A/Z], then choose to open program **ttest**, and perform a paired t test to obtain the following results.

Paired *t* test

Number of paired comparisons	10
Number of degrees of freedom	9
Paired t test statistic S	−0.9040
$P(t \geq S)$	0.8052
$P(t \leq S)$	0.1948
p for two tailed t test	0.3895
Mean of differences MD	−1.300
Lower 95% confidence limit for MD	−4.553
Upper 95% confidence limit for MD	1.953
Conclusion: *Consider accepting* $H_0 : MD = 0$	

Note that the paired t test only requires that the differences $x_i - y_i$ are normally distributed with zero mean, and the requirement for X and Y to be both normally distributed with the same variance, is not so strictly required. Nevertheless, as discussed for the unpaired t test, the Shapiro-Wilks and variance ratio tests are provided to explore the distribution of the observations if that is thought necessary, but they should only be used routinely for large samples, say $n > 50$.