**Introduction**

The paired-samples t-test is used to determine whether the mean difference between paired observations is significantly different from zero. The participants are either the same individuals tested on two occasions, or under two different conditions, on the same dependent variable. Alternatively, you could have two groups of participants that have been matched (paired) on one or more characteristics (e.g., IQ, age, gender, etc.) and tested on one dependent variable. The paired-samples t-test is also often referred to as a dependent t-test, repeated measures t-test, or simply abbreviated to paired t-test.

For example, you could use a paired-samples t-test to understand whether there was a mean difference in smokers' daily cigarette consumption before and after a 6 week hypnotherapy program (i.e., your dependent variable would be "daily cigarette consumption", and your two related groups would be the cigarette consumption values "before" and "after" the hypnotherapy program). You could also use a paired-samples t-test to determine whether there was a mean difference in reaction times under two different lighting conditions (i.e., your dependent variable would be "reaction time", measured in milliseconds, and your two related groups would be reaction times in a room using "blue light" versus "red light").

## Basic requirements of the paired-samples t-test

In order to run a paired-samples t-test, there are four assumptions that need to be considered. The first two relate to your choice of study design, whilst the second two reflect the nature of your data:

* Assumption #1: You have **one dependent variable** that is measured at the **continuous** (i.e., **ratio** or **interval**) level. Examples of **continuous variables** include revision time (measured in hours), intelligence (measured using IQ score), exam performance (measured from 0 to 100), weight (measured in kg), and so forth.
* Assumption #2: You have **one independent variable** that consists of **two categorical**, **related groups** or **matched pairs** (i.e., a **dichotomous variable**). "Related groups" indicates that the two groups are not independent. The primary reason for having related groups is having the same participants in each group. It is possible to have the same participants in each group when each participant has been measured on two occasions on the same dependent variable. For example, you might have measured 10 individuals' performance in a spelling test (the dependent variable) before and after they underwent a new form of computerized teaching method to improve spelling. You would like to know if the computer training improved their spelling performance. The first related group consists of the participants at the beginning (prior to) the computerized spelling training and the second related group consists of the same participants, but now at the end of the computerized training. The paired-samples t-test can also be used to compare different participants (e.g., matched pairs), but this does not happen as often.
* Assumptions #3 and #4: A paired-samples t-test must also meet two assumptions that relate to the nature of your data in order to provide a valid result. These are that: (a) there should be no significant outliers in the differences between the two related groups; and (b) the distribution of the dependent variable should be approximately normally distributed.

## Null and alternative hypothesis

Sometimes, you will be required in your work to explicitly state the null and alternative hypotheses for a paired-samples t-test, and to then state which was accepted or rejected at the end of the experiment. The null hypothesis for a paired-samples t-test is:

H0: The population mean difference between the paired values is equal to zero (i.e. µdiff = 0).

And the alternative hypothesis is:

HA: The population mean difference between the paired values is not equal to zero (i.e. µdiff ≠ 0).

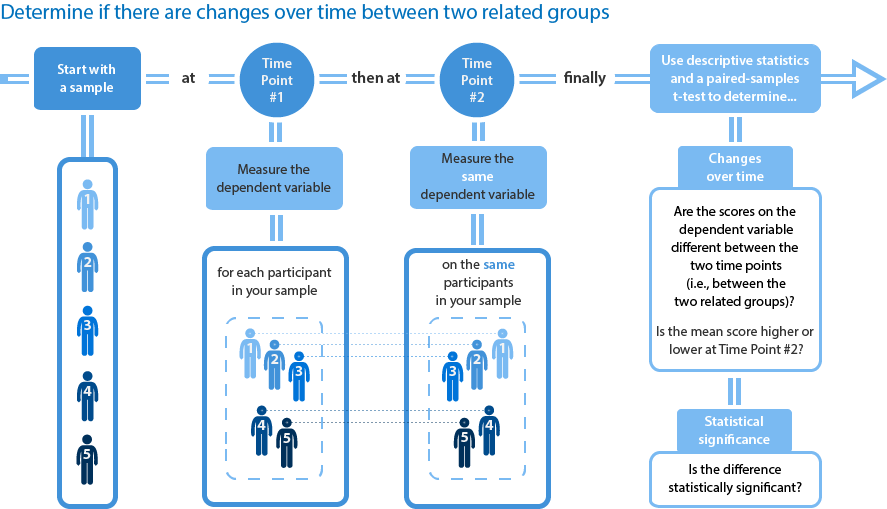
Depending on your results, you can then state whether to reject, or fail to reject, the null hypothesis, and whether to reject or accept the alternative hypothesis. You will be shown how to do this later in this guide.

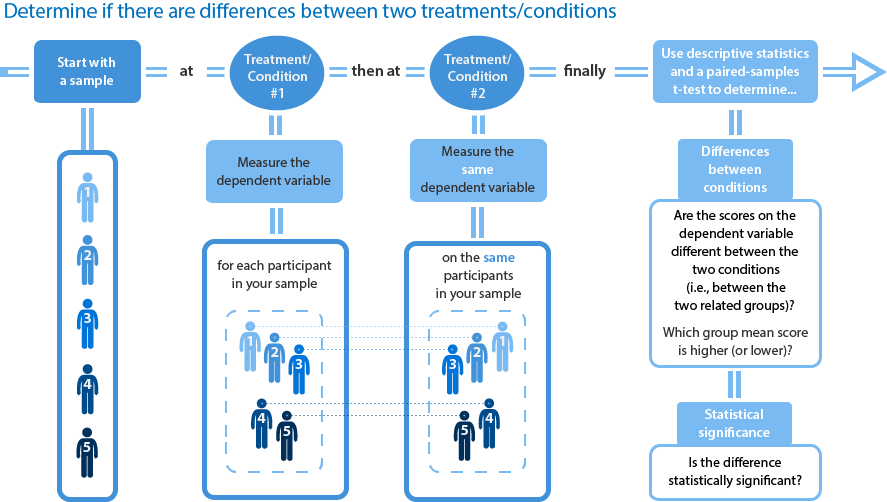
## What you will calculate

1. A point estimate and confidence intervals of the mean difference between related groups.  
2. The statistical significance of the difference (null hypothesis significance testing).  
3. A measure of effect size.

## RESEARCH DESIGNS Problems you can solve using a paired-samples t-test

There are two common types of study design where a paired-samples t-test is used, they are illustrated in the diagrams below.





## Example

A researcher wants to test a new formula for a sports drink that improves running performance. Instead of a regular, carbohydrate-only drink, this new sports drink contains a new carbohydrate-protein mixture. The researcher would like to know whether this new carbohydrate-protein drink leads to a difference in performance compared to the carbohydrate-only sports drink.

To do this, the researcher recruited 20 participants who each performed two trials in which they had to run as far as possible in 2 hours on a treadmill. In one of the trials they drank the carbohydrate-only drink and in the other trial they drank the carbohydrate-protein drink. The order of the trials was counterbalanced and the distance they ran in both trials was recorded.

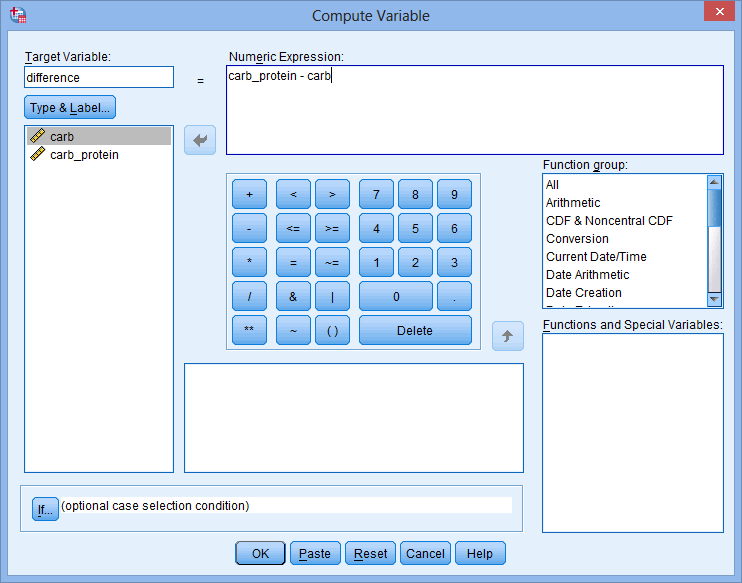
The distance the participants ran in the carbohydrate-only drink trial was recorded in the carb variable and the distance ran in the carbohydrate-protein drink trial in the carb\_protein variable. The researcher would like to determine whether there was a difference in the distance run between the two trials, and therefore, if there is a performance difference between the two different sports drinks. In variable terms, the researcher wants to know if there is a mean difference between the carb and carb\_protein scores.

## Calculating a difference score

For a paired-samples t-test, the assumption of normality and no outliers is tested on the differences between the paired-values, not the values of the paired groups themselves. To accomplish this, you first need to compute the difference scores for paired-values and store them in a new variable. You can then instruct SPSS to perform the tests of assumptions on this newly created variable, which contains the difference scores. In this example, the new variable you will create will be called difference. So, to create this new variable, follow these instructions:

1. Click **Transform > Compute Variable...** on the main menu. You will be presented with the **Compute Variable** dialogue box.

2. Type "**difference**" (without the quotation marks) into the Target Variable: box. This will mean that the newly-created variable is called difference. Then, to calculate the difference between related groups (i.e., paired values), in the Numeric Expression: box, type in "**DV1 – DV2**" (without the quotation marks), as shown below:



3. Click the https://statistics.laerd.com/premium/pstt/img/ok-button.pngbutton. You will be returned to the **Data View** window where you will see that your new variable, difference, has been created and populated with the difference scores.

Now that you have calculated these difference scores, tests for normality and outliers can be conducted on them, as explained in the Assumptions section that follows.

## Assumptions of the paired-samples t-test

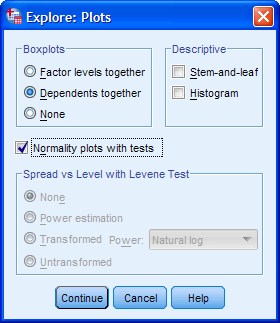
The following instructions show you how to run tests to detect outliers and check if your data is normally distributed:

1. Click **Analyze > Descriptive Statistics > Explore.**

**2.** Transfer your difference variable into the Dependent List: box by clicking on it while holding down the shift-key, and then clicking the top https://statistics.laerd.com/premium/rma/img/right-arrow-button.pngbutton.

**3.** Click the https://statistics.laerd.com/premium/rma/img/plots-button.pngbutton and you will be presented with the **Explore: Plots** dialogue box.

**4.** Select Dependents together in the -Boxplots- area, but deselect Stem-and-leaf in the -Descriptive- area and select Normality plots with tests, so that you end up with the following screen:

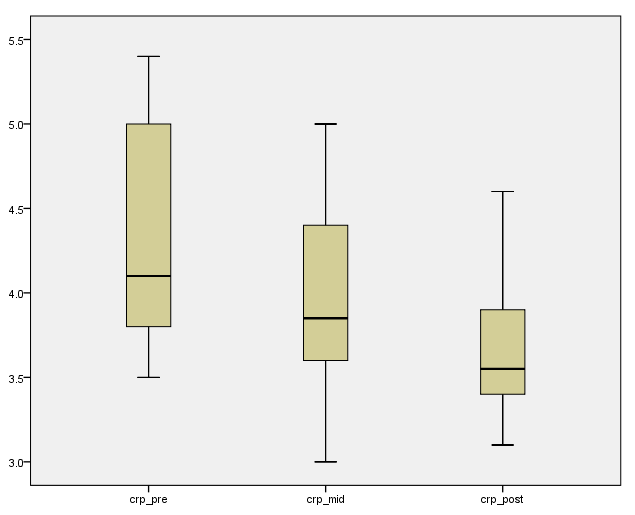


Note: The Shapiro-Wilk is used for sample sizes <50 and the Kolmogorov-Smirnov is used for sample sizes >50. Both test for data normality.

5. Click the https://statistics.laerd.com/premium/rma/img/continue-button.pngbutton. You will be returned to the **Explore** dialogue box.

6. Click the https://statistics.laerd.com/premium/rma/img/ok-button.pngbutton to generate the output.

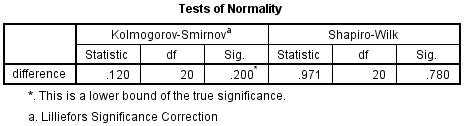
**1. Are there any outliers in our data?**



Note: Any data points that are more than 1.5 box-lengths from the edge of their box are classified by SPSS as outliers. These data points are illustrated as circular dots and labeled with their case number (i.e., their row number in the **Data View** window). If any data points are more than 3 box-lengths away from the edge of their box, they are classified as extreme points (i.e., extreme outliers) and illustrated as an asterisk (\*) with their case number labeled.

You can see that there are no outliers in our data, as evidenced by the lack of any circular points or asterisks.

**2. Are our data normally distributed?**



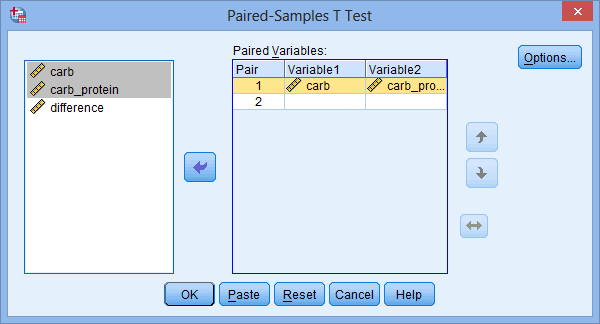
## If your data is normally distributed (i.e., the assumption of normality is met), the significance level (value in the "**Sig.**" column) should be more than .05 (i.e., p > .05). If your data is not normally distributed (i.e., the assumption of normality is violated), the significance level ("**Sig.**") will be less than .05 (i.e., p < .05). The null hypothesis of the Shapiro-Wilk test is that your data's distribution is equal to a normal distribution and the alternative hypothesis is that your data's distribution is not equal to a normal distribution. Thus, if you reject the null hypothesis (p < .05), this means that your data's distribution is not equal to a normal distribution and if you fail to reject the null hypothesis, your data is normally distributed.

## Paired-samples t-test procedure

The following procedure shows you how to run a paired-samples t-test in SPSS. To run this procedure, you will no longer need to refer to the difference variable you calculated.

Step One: Click **Analyze > Compare Means > Paired-Samples T Test...** on the main menu. You will be presented with the **Paired-Samples T Test** dialogue box.

Step Two: Transfer the dependent variables into the Paired Variables: box. To do this, highlight both variables by clicking on one (e.g., carb), then holding down the shift-key, click the other variable (e.g., carb\_protein). Now, click the https://statistics.laerd.com/premium/pstt/img/right-arrow-button.pngbutton. You will end up with a screen similar to the one below:



Step Three: Click the https://statistics.laerd.com/premium/pstt/img/continue-button.pngbutton and you will be returned to the **Paired-Samples T Test** dialogue box.

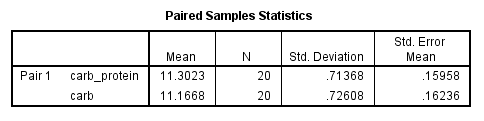
Step Four: Click the https://statistics.laerd.com/premium/pstt/img/ok-button.pngbutton to generate the output.

## Interpreting and reporting the output

**Descriptive statistics**

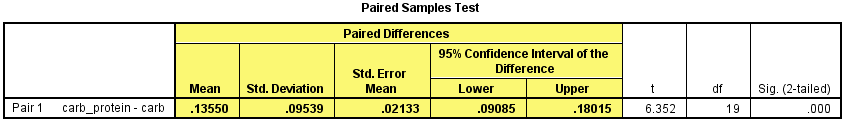
You can make an initial interpretation of the data using the **Paired Samples Statistics** table, which presents relevant descriptive statistics, as shown below:

Reporting in APA Format: Participants ran further when imbibing the carbohydrate-protein drink (M = 11.302 km, SD = 0.717) as opposed to the carbohydrate-only drink (M = 11.167 km, SD = 0.726).



**Differences between trials**

The **Paired Samples Test** table presents the mean difference between the two variables (i.e., carb\_protein minus carb), as well as different measures of variability in the first section of the table (columns below the "**Paired Differences**" column), as highlighted below:

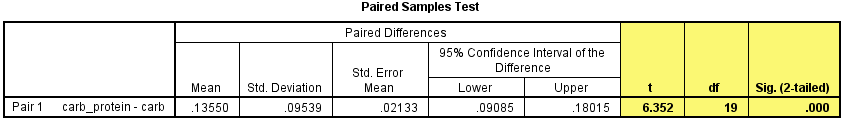


You can see here that there is a mean difference between the two trials of 0.1355 km (the "**Mean**" column) with a standard deviation of 0.09539 km (the "**Std. Deviation**" column), a standard error of the mean of 0.02133 km (the "**Std. Error Mean**" column), and 95% confidence intervals of 0.09085 to 0.18015 km (the "**Lower**" and "**Upper**" columns, respectively). Notice that because you re-ordered the variables, the difference is positive if the mean carbohydrate-protein score (carb\_protein) is larger than the mean carbohydrate-only score (carb). You are reminded of this in the far-left of the table as "**Pair 1 carb\_protein - carb**". You could report these results as follows:

Reporting in APA Format: The carbohydrate-protein drink elicited a mean increase of 0.136 km, 95% CI [0.091, 0.180] in the distance run in two hours compared to a carbohydrate-only drink.

**Reporting statistical significance**

The results of the hypothesis test are included at the end of the **Paired Samples Test** table, as highlighted below:



Moving from left-to-right columns, you are presented with the obtained *t*-value (the "**t**" column), the degrees of freedom (the "**df**" column), and the statistical significance (*p*-value) (the "**Sig. (2-tailed)**" column) of the paired-samples t-test. If *p* < .05, this means that the mean difference between the two related groups is statistically significant. Alternatively, if *p* > .05, you do not have a statistically significant mean difference between the two related groups. In this example, the statistical significance level is stated as **.000**, which means *p* = .0005 (i.e. *p* < .05).

As p < .05 in this example, it can be concluded that there is a statistically significant mean difference between the mean of our two related groups (i.e., carb and carb\_protein) in the population. Stated another way, the mean difference between the two run distances is not equal to zero in the population.

It is important to remember that the level of significance (p-value) ***does not*** indicate the strength or importance of the mean difference between the related groups, only the likelihood of a mean difference as large or larger as the one you observed, given that the null hypothesis is true. For example, if this example had produced a p-value of 0.012 (p = .012), this does not mean it is twice as 'strong' or 'important' as p = .024. In layman's terms, the p-value is simply trying to inform you whether the mean difference in the two groups you studied is not a 'fluke' and it really is likely that you would expect to see a mean difference like the one in your study in the population (not just in your sample). In that sense, a lower p-value simply indicates how confident you can be that your result is a 'real' one. Whether that makes it important or large, it cannot tell you.

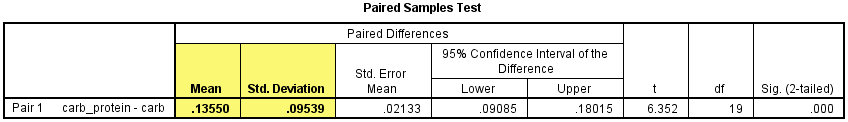
Note: There is a link between the 95% confidence intervals of the mean difference and the statistical significance of the mean difference. If the confidence intervals do not contain the number zero, you have a statistically significant mean difference (p < .05). If they do contain zero, you do not have a statistically significant mean difference (p > .05). In this example, you discovered that the 95% confidence intervals were from 0.091 to 0.180, thus not including zero and indicating a statistically significant result.

## Calculating and reporting an effect size

You will be required to report effect sizes less often than statistical significance, particularly at the undergraduate level, but they are becoming a more common method of analysis with each passing year and some journals now insist on their inclusion wherever possible. To calculate an effect size, called *d* or Cohen's *d*, for a paired-samples t-test you need to divide the mean difference by the standard deviation of the difference (Cohen, 1988), as shown below:

https://statistics.laerd.com/premium/pstt/img/effect-size-calculation-1.png

Where *M* is the mean difference between the two related groups and *SD* is the standard deviation of this difference. These two values have already been calculated for you and are are highlighted in the table below:



Therefore, the effect size, *d*, can be calculated as:

https://statistics.laerd.com/premium/pstt/img/effect-size-calculation-2.png

An effect size is an attempt to provide a measure of the practical significance of the result. The importance of the value of Cohen's *d* (as reported by Cohen (1998)) is as follows:

|  | **Effect Size** | **Strength** |
| --- | --- | --- |
|  | .2 | small |
|  | .5 | medium |
|  | .8 | large |
| Table: Interpretation of values of Cohen's *d*. | | | |

As the effect size, *d*, is 1.42 you can conclude that there is a large effect. However, one of the major weaknesses of using effect sizes is that guidelines indicating the importance of the effect size are subject-specific and there are not, as yet, comprehensive guidelines for interpretation of the strength of an effect size. Moreover, there is usually more than one measure of effect size (i.e., different types and calculations of effect size). You should report the value of the effect size with the result of the hypothesis test (i.e., the paired-samples t-test), which you could do as follows:

Reporting in APA Format: The carbohydrate-protein drink elicited a statistically significant mean increase in distance run in two hours compared to a carbohydrate-only drink, t(19) = 6.352, p < .0005, d = 1.42.

Putting it all together in APA format: A paired-samples t-test was used to determine whether there was a statistically significant mean difference between the distance ran when participants imbibed a carbohydrate-protein drink compared to a carbohydrate-only drink. Two outliers were detected that were more than 1.5 box-lengths from the edge of the box in a boxplot. Inspection of their values did not reveal them to be extreme and they were kept in the analysis. The assumption of normality was not violated, as assessed by Shapiro-Wilk's test (p = .780). Participants ran further when imbibing the carbohydrate-protein drink (M = 11.302, SD = 0.717 km) as opposed to the carbohydrate only drink (M = 11.167, SD = 0.726 km), a statistically significant mean increase of 0.136 km, 95% CI [0.091, 0.180], t(19) = 6.352, p < .001, d = 1.42.