# **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").

# **Assumptions of the Paired Sample 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.**
* 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 Hypotheses:**

The null hypothesis for a paired-samples t-test is:

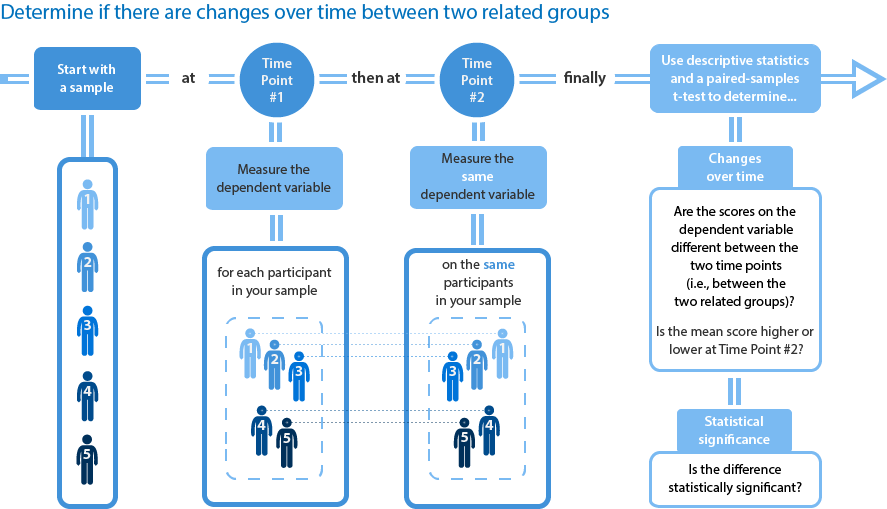
H0: The population mean difference between the paired values is equal to zero (i.e. µdiff = 0).

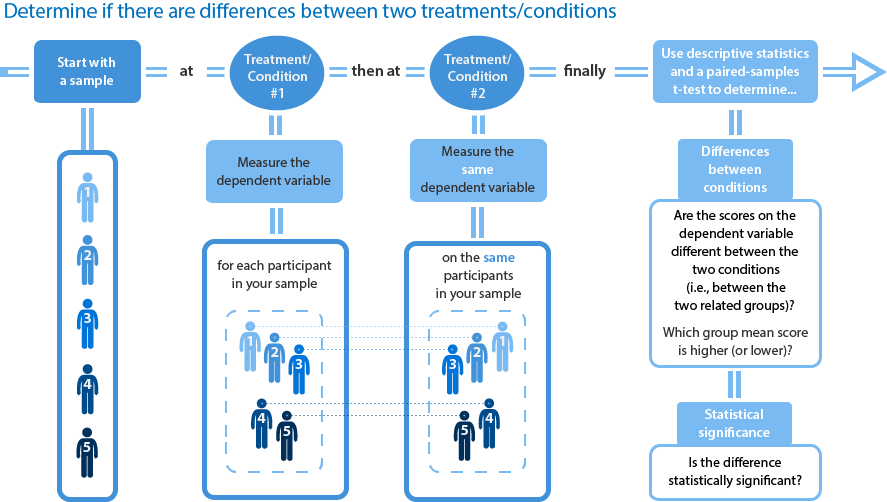
And the alternative hypothesis is:

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

## **Example:**

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





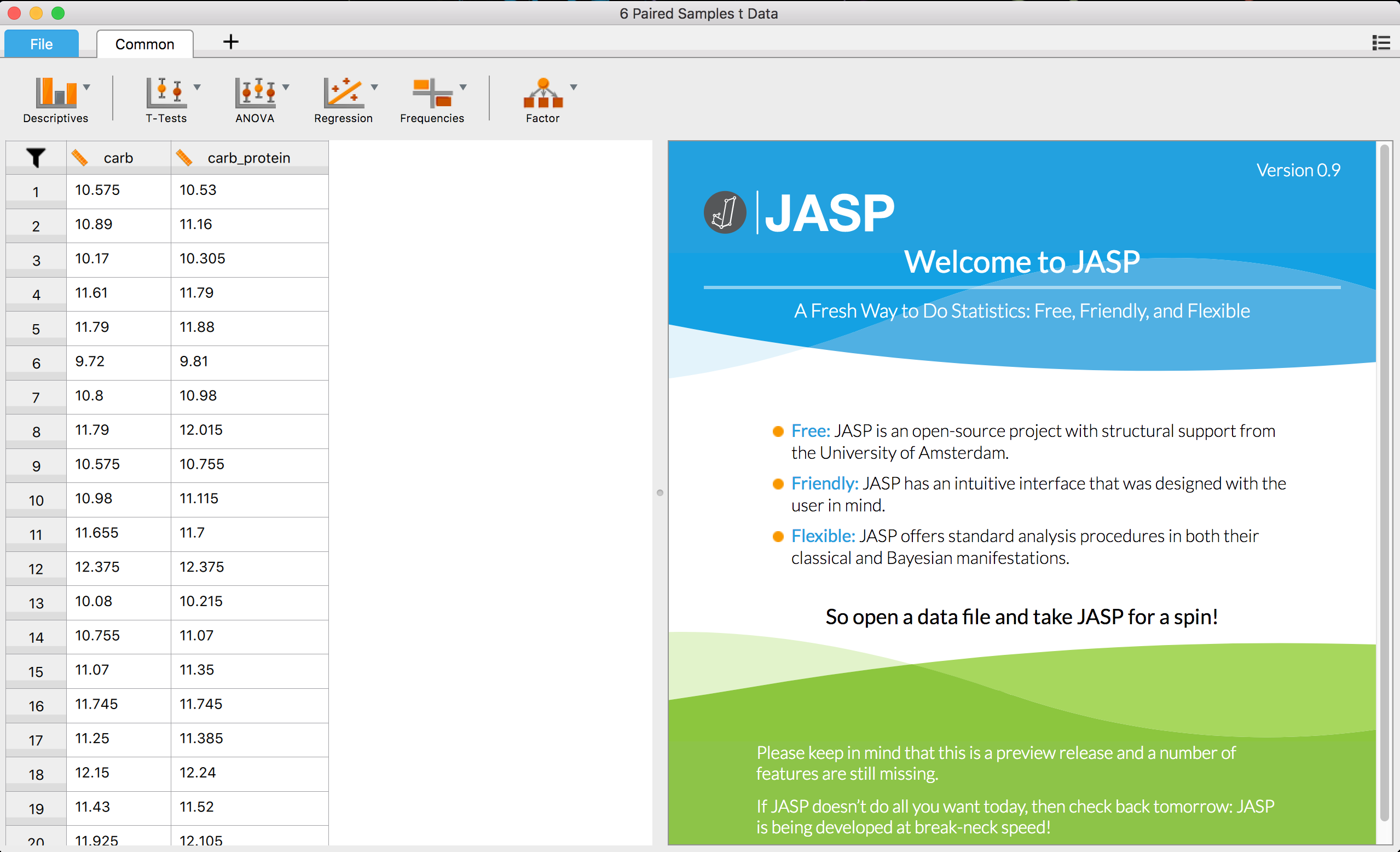
A researcher wants to test a new formula for a sports drink that improves running performance. Instead of a regular, carbohydrate-only drink, this new sports drink contains a new carbohydrate-protein mixture. The researcher would like to know whether this new carbohydrate-protein drink leads to a difference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

To get started, open the dataset for this example in JASP. Remember, you can always use the previous help guides for greater detail in case you do not remember how to do something.

File 🡪 Open 🡪 Computer 🡪 Browse 🡪 Pick the Paired Samples t Data.



## **Check your assumptions:**

**Is the dependent variable at least scale (ratio or interval)?**

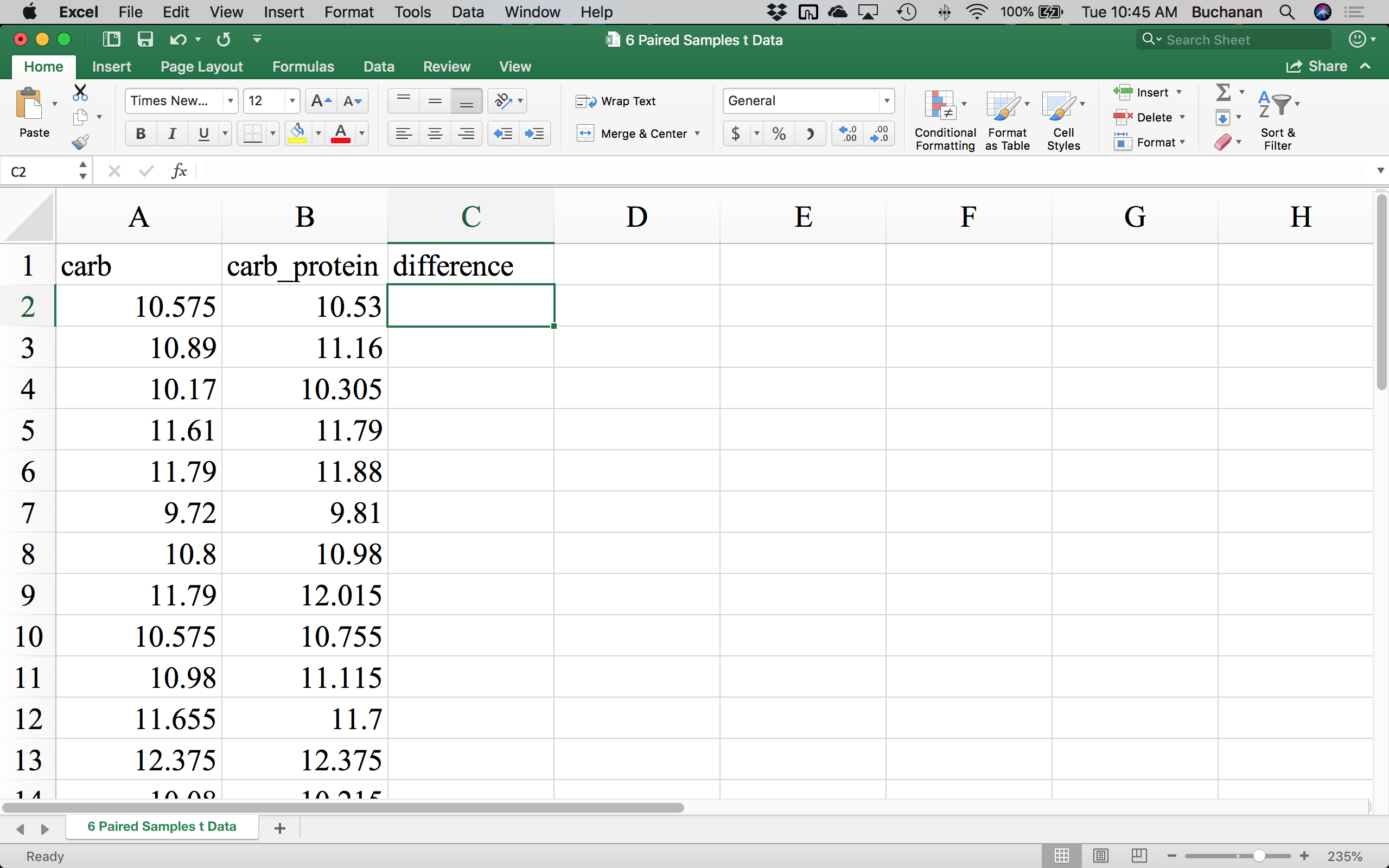
Yes, we are using ratio style data.

**Are there any outliers in the sample?**

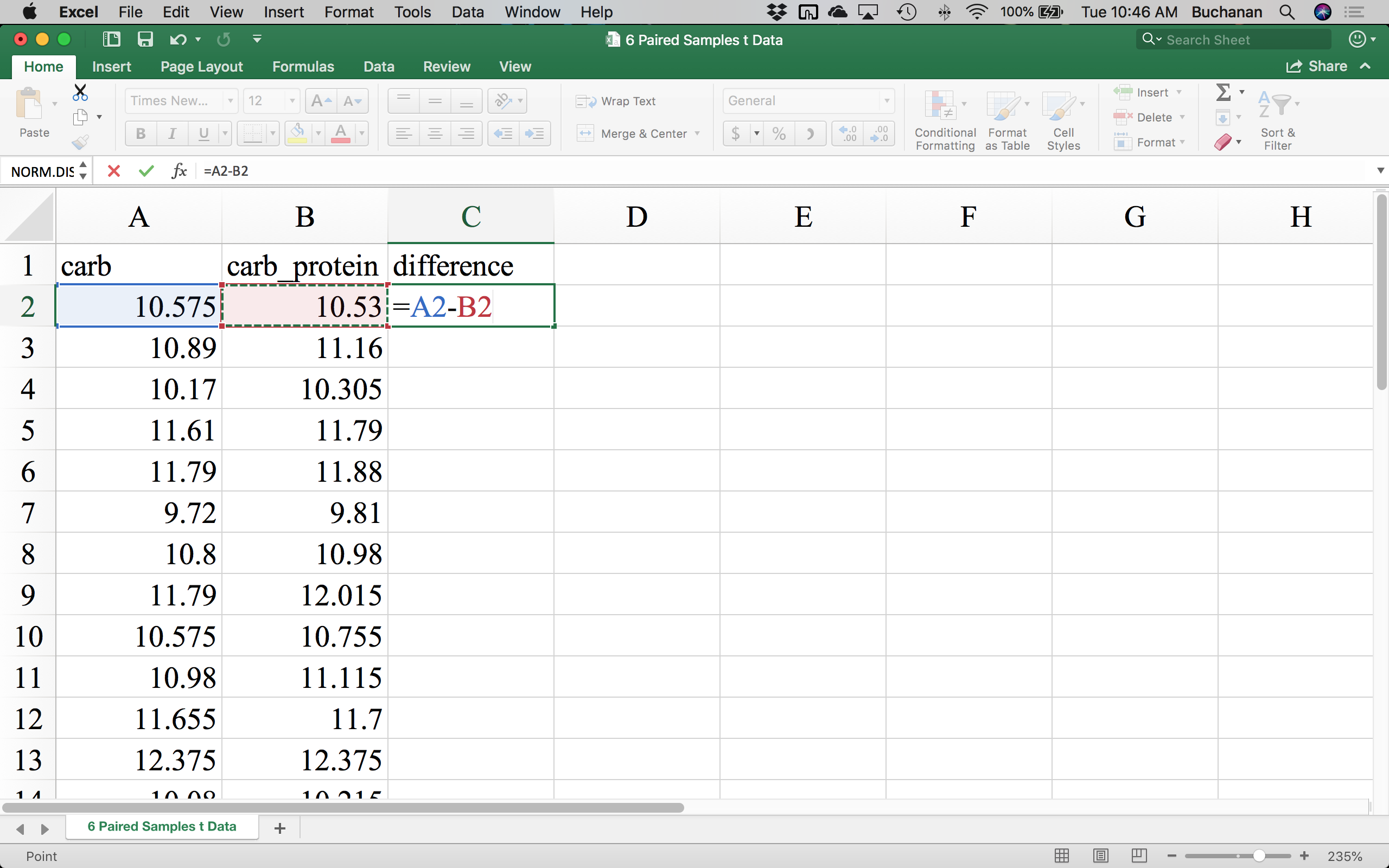
Calculating a difference score

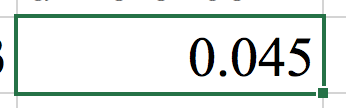
For a paired-samples t-test, the assumption of normality and no outliers is tested on the differences between the paired-values, not the values of the paired groups themselves. To accomplish this, you first need to compute the difference scores for paired-values and store them in a new variable. You 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:

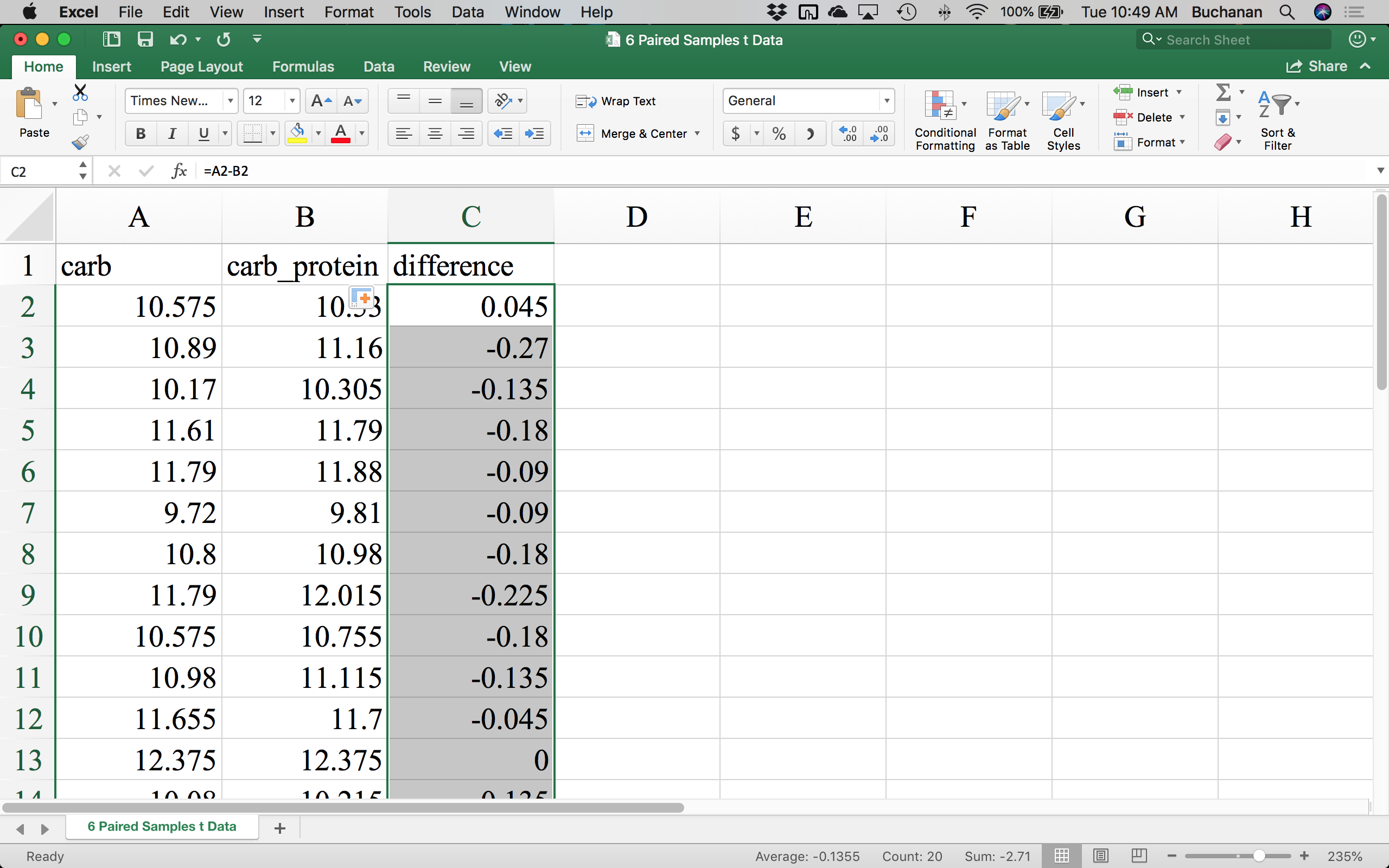
Double click on one of the numbers in the dataset and wait for Excel to open. Type a name in column C or any new open column. Here, we’ve called it difference to be able to remember what it is.



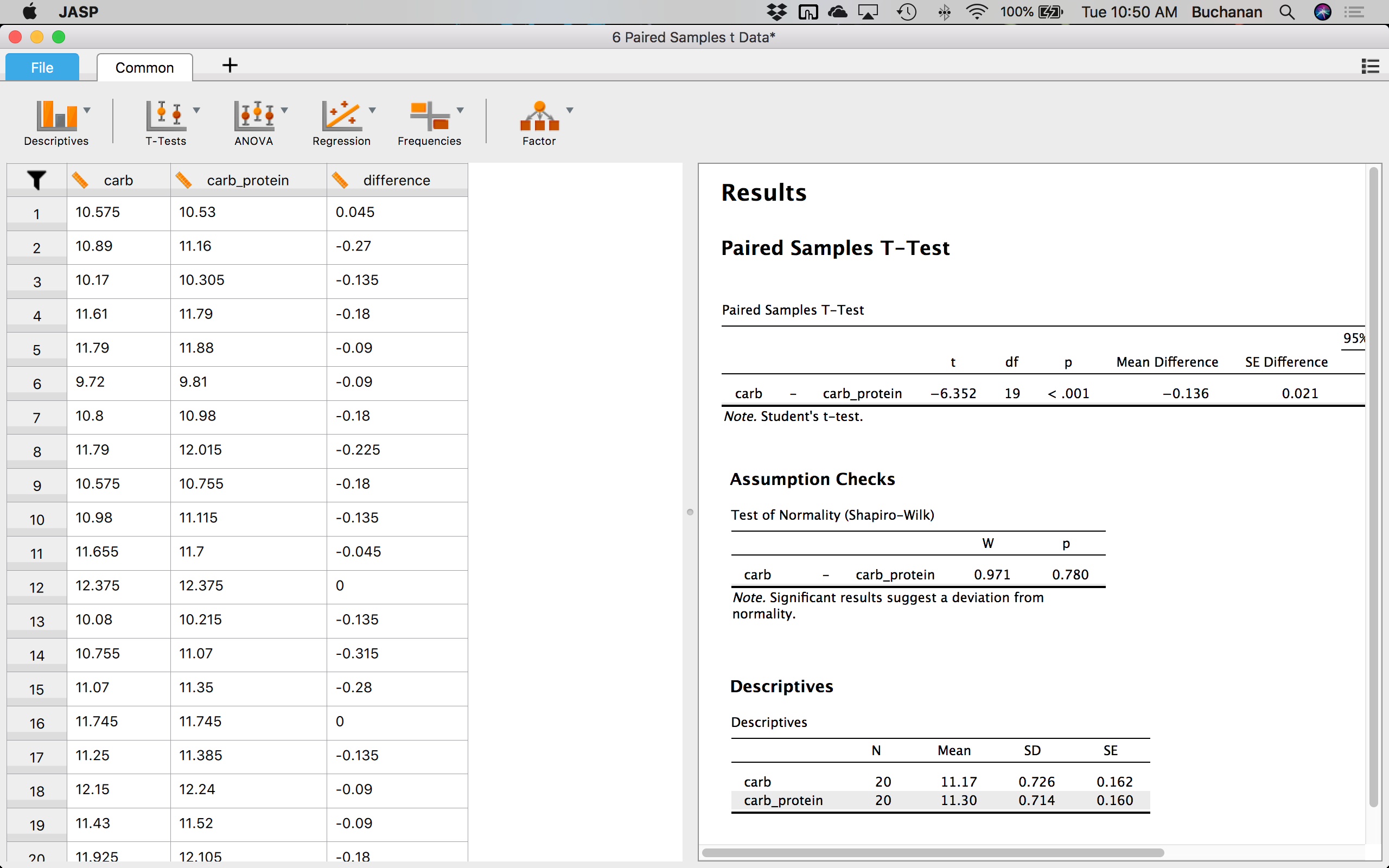
Type the equals (=) key in cell C2 (where the green box is). Then click on A2 and type the minus key (-) and click on B2. It should look like this when you are done:



Then hit enter so that Excel will calculate the difference score for you. Then you can fill in the rest of them by hovering over the bottom right corner of the cell that appears to have a small square on it. If you have your mouse over the small square, you will see the mouse change from white to black 🡪 double click on the square and it will fill in all the other scores.



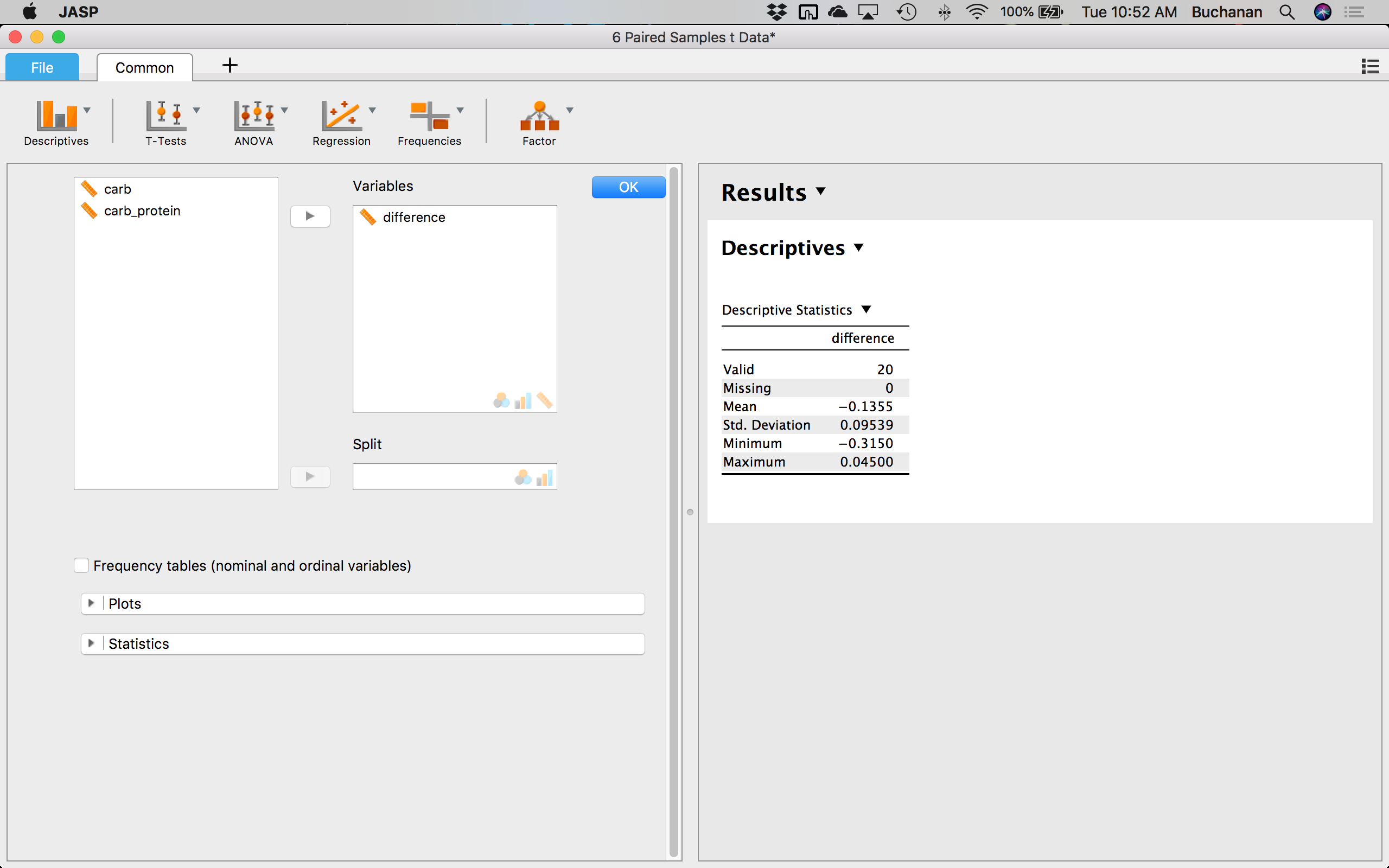
Now save the file by clicking the save icon , and the data should update in JASP. You can now close Excel.



To examine if any data might be considered an outlier, we can use the Descriptives  options you learned about previously. Click Descriptives 🡪 Descriptive Statistics.

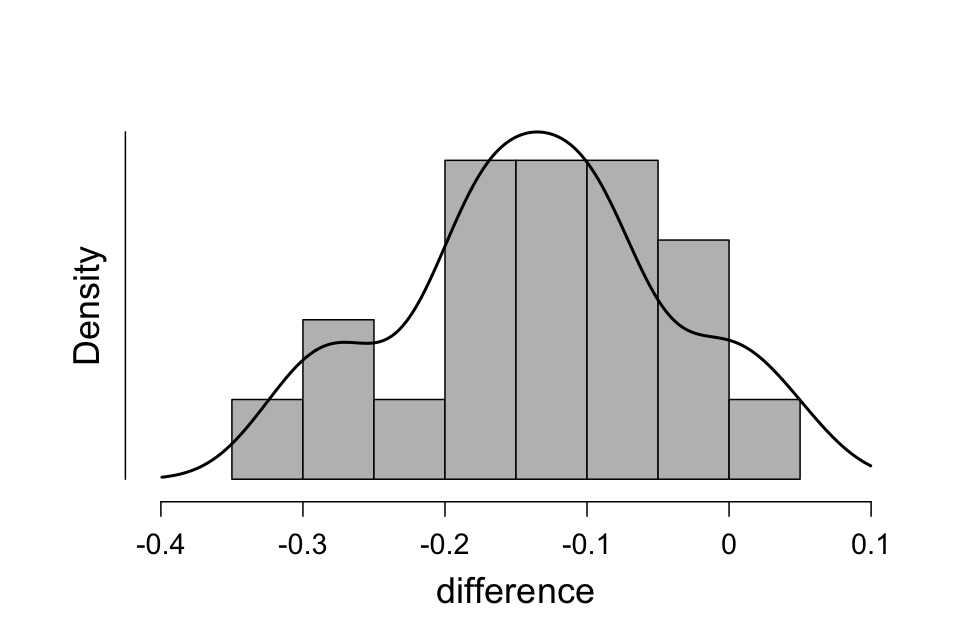


In this window, we want to click on difference and click the arrow  to move it over to the right hand side under Variables. Click on the plots options:  to see more available options.



Here we can look at two different options to see if any participants scores are very different from other participants scores. First, click on Distribution plots. 

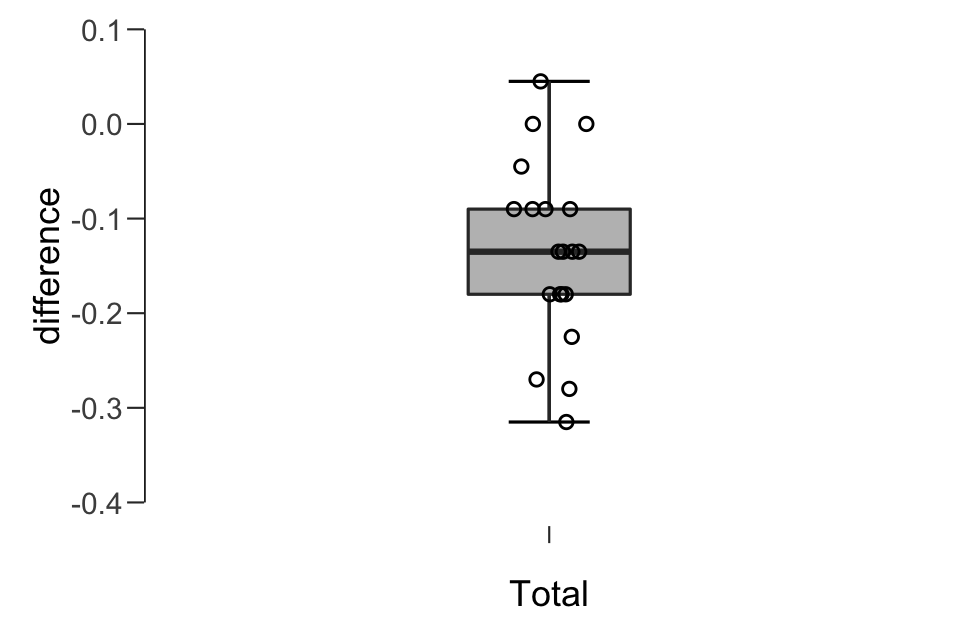
##### difference



A histogram will appear. We can tell that most people are fairly close together, so probably not any outliers.

Another option would be to select Box Plots , Label Outliers , and Jitter Element . You will see outliers labeled with a special symbol, and Jitter Element allows you to see all the participants scores as dots on the plot.

##### difference



First Quartile

Third Quartile

Median

Quartile - 1.5 \* IQR

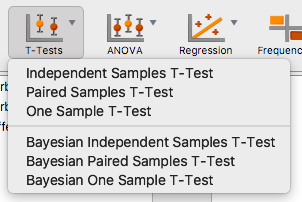
Quartile + 1.5 \* IQR

A boxplot indicates the following parts (labeled on top of the diagram above, which you will not see in JASP). Remember that IQR is the interquartile range or the third quartile minus the first quartile. If we had outliers, they would be outside the top and bottom lines and would be marked with a star.

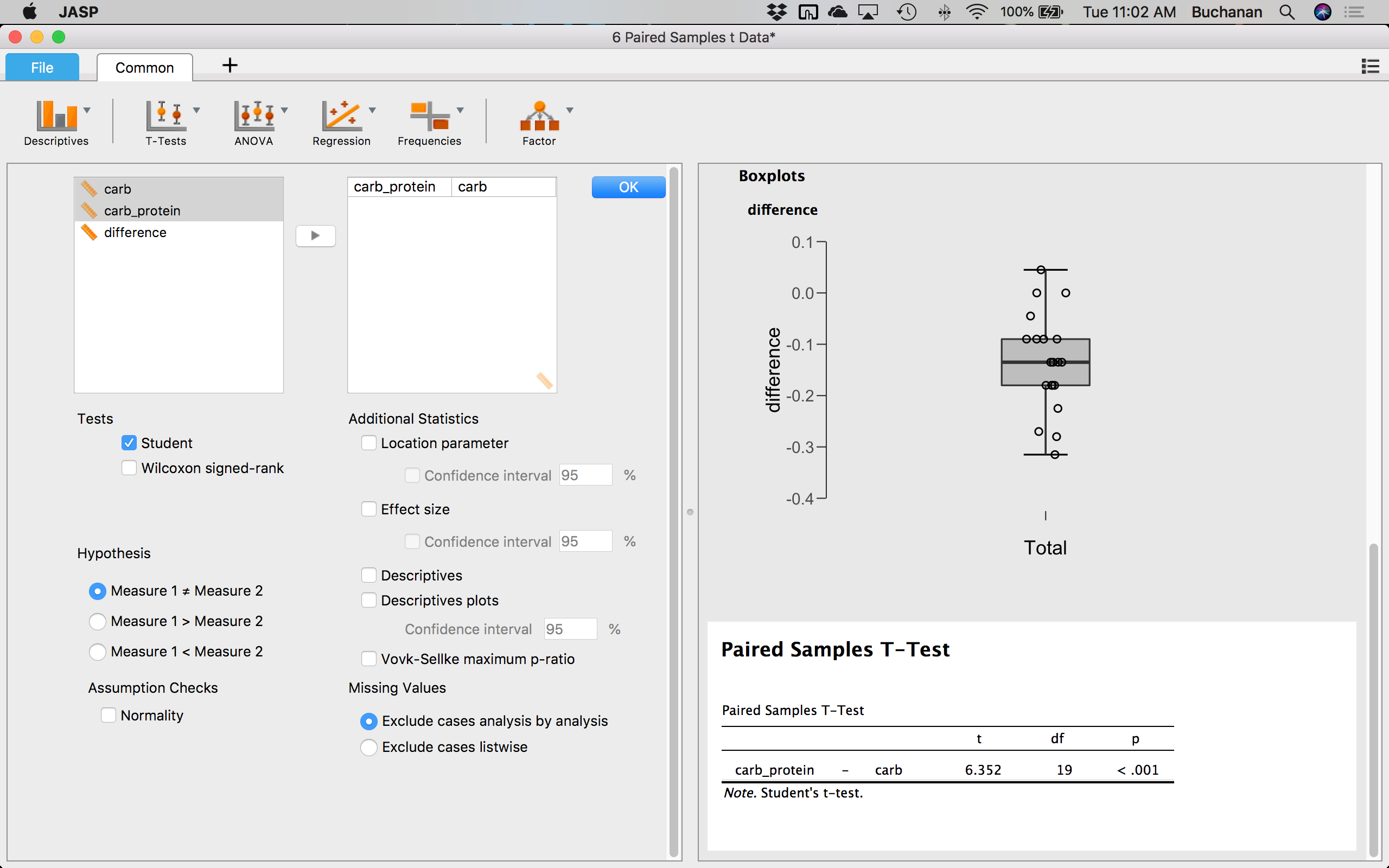
**Is the dependent variable normally distributed?**

We can view the histogram created earlier to look at if the data appears normal, but we might also consider using the Shapiro-Wilk test to determine if the data is normal. To get this test, we need to run the actual t-test to see that output.

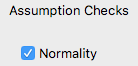
Click on t-tests  🡪 Paired Samples t Test



In this window, we want to click on carb and carb\_protein and click the arrow  to move it over to the right hand side under Variables.



To get the normality assumption test, click on Normality, under Assumptions:



### Assumption Checks

| **Test of Normality (Shapiro-Wilk)** | | | | | | | | | |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | |  | |  | | **W** | | **p** | |
| carb\_protein |  | - |  | carb |  | 0.971 |  | 0.780 |  |
|  | | | | | | | | | |
| Note.  Significant results suggest a deviation from normality. | | | | | | | | | |

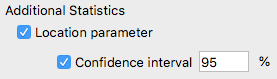
If your data is normally distributed (i.e., the assumption of normality is met), the significance level (value in the *p* column) should be more than .05 (i.e., *p* > .05). If your data is not normally distributed (i.e., the assumption of normality is violated), the significance level will be less than .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.

Here we see that our data is normally distributed because *p* > .05. Remember, t-tests are robust to violations of normality, especially with larger sample sizes.

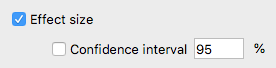
## **The t-test and effect size:**

Now, we can finish out running the t-test by clicking on a few more options.

On the right hand side, you will want to add the location parameter which will show you the average difference score between your two measurements. You can also add the confidence interval.



You will want to add effect size (*d*):



You can also add the descriptive statistics of each measurement individually by clicking on descriptives: .

## **Interpretation and Reporting:**

**Paired Samples T-Test**

| **Paired Samples T-Test** | | | | | | | | | | | | | | | | | | | | | |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | | | | | | | | | | | | | | | | **95% CI for Mean Difference** | | | |  | |
|  | |  | |  | | **t** | | **df** | | **p** | | **Mean Difference** | | **SE Difference** | | **Lower** | | **Upper** | | **Cohen's d** | |
| carb\_protein |  | - |  | carb |  | 6.352 |  | 19 |  | < .001 |  | 0.136 |  | 0.021 |  | 0.091 |  | 0.180 |  | 1.420 |  |
|  | | | | | | | | | | | | | | | | | | | | | |
| Note.  Student's t-test. | | | | | | | | | | | | | | | | | | | | | |

In the first box, you will find your *t* value that tells you the test statistic for the difference between your two measurements carb\_protein and carb (remember that t-values are the difference between mean and population mean divided by the standard error 🡪 it’s the difference on the top and the error on the bottom).

Second are the degrees of freedom (df, here sample N-1), followed by the *p* value. Let’s say you set your alpha value (or Type I error rate) at *p* < .05. You can also use other values, such as *p* < .10 or *p* < .01, but *p* < .05 is a very popular value. If *p* < .05 in our results, this means that the sample and the population are different and this test is statistically significant. Alternatively, if *p* > .05, you do not have a statistically significant mean difference between the sample and population mean. In this example, the statistical significance level is stated as < .001, which means *p* < .001 (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.

Next, we find the Mean Difference score, which is the average score of the differences between carb\_protein and carb. The SE difference score is the standard error of those difference scores. The confidence interval is presented next to the mean difference. Confidence intervals represent the range around the mean that you might expect to find the true population mean if you repeated this experiment 95% of the time (note: we can change to 90% or 99%, usually this value is set to 1 – alpha or your Type 1 error rate).

There is a link between the 95% confidence intervals of the mean difference and the statistical significance of the mean difference. If the 95% 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.

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. It is the last value presented here on the right. As the effect size, *d*, is 1.42 you can conclude that there is a large effect.

You might report all these numbers like this:

A statistically significant mean increase of 0.14 km, 95% CI [0.09, 0.18], t(19) = 6.35, p < .001, d = 1.42 was found in our study.

Remember, you want two decimals normally, except for *p* values which get three decimals. We have also included the 95% confidence interval by using square brackets [] around the lower and upper end reported in the table.

**Assumption Checks**

| **Test of Normality (Shapiro-Wilk)** | | | | | | | | | |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | |  | |  | | **W** | | **p** | |
| carb\_protein |  | - |  | carb |  | 0.971 |  | 0.780 |  |
|  | | | | | | | | | |
| *Note.*  Significant results suggest a deviation from normality. | | | | | | | | | |

As described earlier, we would note here that we did meet the assumption of normality because *p* > .05. Remember that assumptions tests are backwards – you do not want them to be significant. You might report this test like this:

The assumption of normality was not violated, as assessed by Shapiro-Wilk's test (p = .780)

**Descriptives**

| **Descriptives** | | | | | | | | | |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | | **N** | | **Mean** | | **SD** | | **SE** | |
| carb\_protein |  | 20 |  | 11.30 |  | 0.714 |  | 0.160 |  |
| carb |  | 20 |  | 11.17 |  | 0.726 |  | 0.162 |  |
|  | | | | | | | | | |

This box contains the means, standard deviations, and standard errors for each individual measurement. We can report these values to help the reader understand what the mean difference score reported earlier actually tells you. You might do that like this:

Participants ran further when imbibing the carbohydrate-protein drink (M = 11.30, SD = 0.72 km) as opposed to the carbohydrate only drink (M = 11.17, SD = 0.73 km).

## **Reporting All Together:**

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. No outliers were detected that were more than 1.5 box-lengths from the edge of the box in a boxplot. 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 = 11.30, SD = 0.72 km) as opposed to the carbohydrate only drink (M = 11.17, SD = 0.73 km), a statistically significant mean increase of 0.14 km, 95% CI [0.09, 0.18], t(19) = 6.35, p < .001, d = 1.42.