# **Introduction:**

The repeated measures analysis of variance (ANOVA) is an extension of the paired-samples t-test and is used to determine whether there are any statistically significant differences between the population means of three or more related groups. The groups are related as they contain the same cases (e.g., participants) in each group and each group represents a repeated measurement on the same dependent variable. This test is also referred to as a within-subjects ANOVA or ANOVA with repeated measures.

In order to run a repeated measures ANOVA you require the following:

* One independent variable that is **categorical with three or more related groups** (e.g., time: pre-, 1-month, post-intervention).
* One dependent variable that is **continuous** (e.g., satisfaction score).

A repeated measures ANOVA is most often used for three types of study design:

**1. Determine if there are differences between three or more time points**

If you have a study design where you are measuring how a particular variable changes over time in the same participants and you want to compare three or more time points, a repeated measures ANOVA might be appropriate. It does not matter what occurs between the time points, so you could have initiated an intervention, such as a training program, or alternatively, simply measured the passage of time, as long as you are measuring the same variable at all times points.

**2. Determine if there are differences between conditions**

If you have a study design where the same participants are being measured on the same variable, but under three or more different conditions, a repeated measures ANOVA might be appropriate. In other words, participants are performing a cross-over design by receiving all conditions. These can either be short-term conditions, such as reaction times in a 10-second period under three different lighting conditions (e.g., blue *vs.* red *vs.* green light), or longer-term conditions, such as a six week control, exercise-training or dietary program with cholesterol concentration measured at the end of each trial.

**3. Determine if there are differences in change scores**

If you have a study design where the same participants have performed three or more different interventions (e.g., control/intervention 1/intervention 2), the same continuous dependent variable is measured at the beginning and end of each intervention in all groups, and a change score calculated (i.e., post-values minus pre-values), a repeated measures ANOVA might be appropriate.

**4. Determine if there are differences between measurements**

If you have a study design where the same participants are being measured on a different variable, but using the same measurement scale, a repeated measures ANOVA might be appropriate.

# **Assumptions of the Repeated Measures ANOVA:**

For a repeated measures ANOVA to be able to provide a valid result, the following three assumptions must hold about the data in each group:

* Assumption #1: You have **one dependent variable that is measured at the continuous (i.e., ratio or interval) level.**
* **Assumption #2:** Your independent variable is categorical with three or more separate measurements of the same participants.
* Assumption #3: There should be no significant outliers in any of the measurements of the participants, meaning each measurement should be assessed separately.
* Assumption #4: Your dependent variable should be approximately normally distributed for each measurement of the independent variable.
* Assumption #5: The variances of the differences between related groups are equal (the assumption of sphericity). This assumption is similar to the homogeneity of variances for separate groups that you tested for in the between subjects ANOVA. However, this assumption investigates if the variances of the *difference scores* between pairs of levels are the same. Therefore, if you had three measurements (levels), the variance of the difference between measurement 1 and measurement 2 should be the same as the variance of the difference between measurement 1 and 3 and measurement 2 and 3.

These assumptions need to be tested before you can run a repeated measures ANOVA. Fortunately, the repeated measures ANOVA is fairly "robust" to violations of normality. "Robust", in this case, means that the assumption can be violated (a little) and still provide valid results. Therefore, you will often hear of this test only requiring approximately normal data and some argue that data can even be fairly skewed as long as the number of cases (e.g. participants) in each group is similar.

## **Null and Alternative Hypotheses:**

The null hypothesis is:

H0: all related group means are equal (i.e. µ1 = µ2 = µ3 = ... = µk). There are no differences between TIME1/CONDITION1, TIME2/CONDITION2, and TIME3/CONDITION3 on the dependent variable.

The alternative hypothesis is:

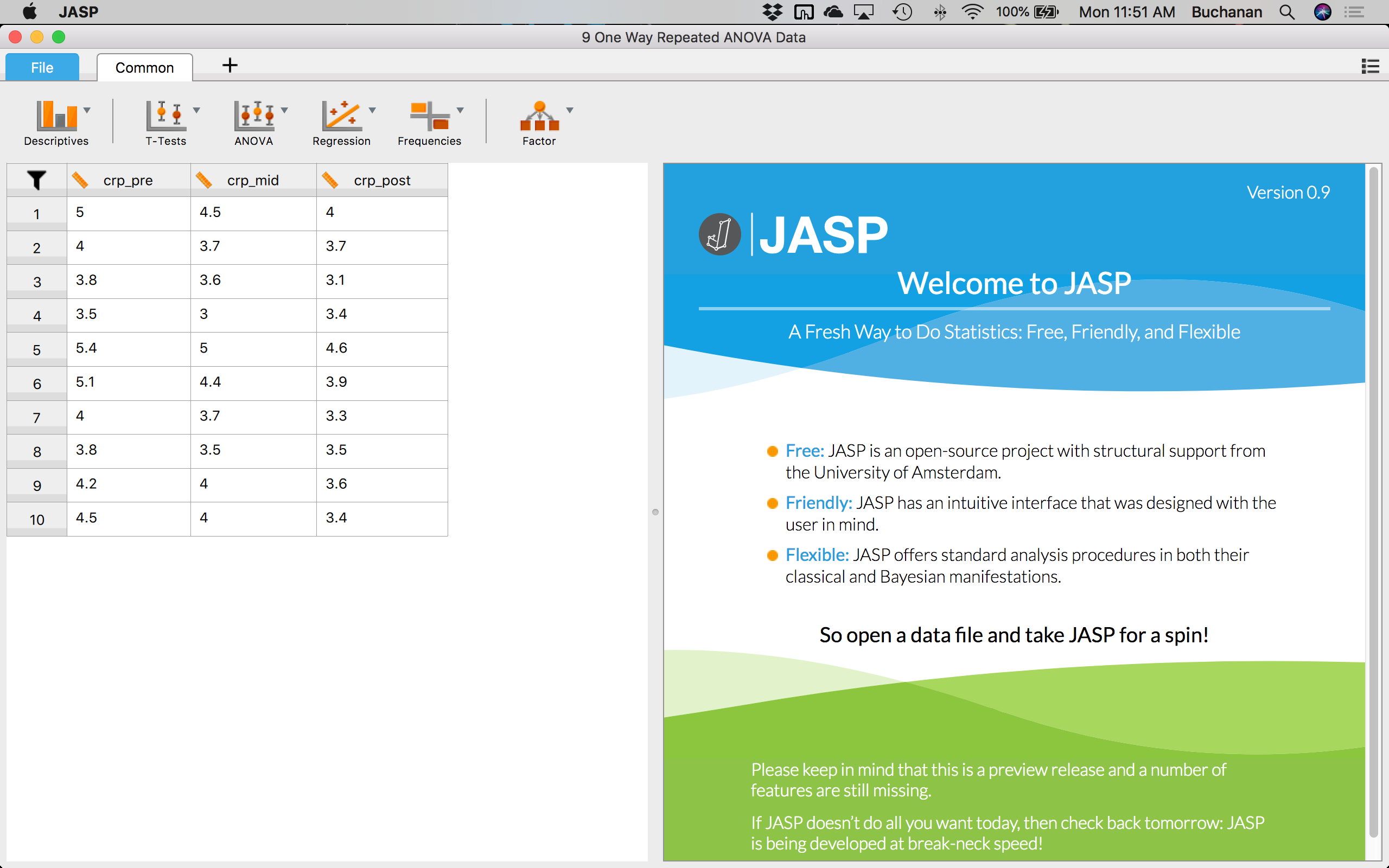
HA: at least one related group mean is different (i.e. they are not all the same). There are differences between TIME1/CONDITION1, TIME2/CONDITION2, and TIME3/CONDITION3 on the dependent variable.

## **Example:**

A researcher wishes to understand how exercise might reduce heart disease. They wish to concentrate on a protein called C-Reactive Protein (CRP) that is a marker of chronic inflammation in the body and associated with heart disease: the greater the concentration of CRP, the greater the risk of heart disease. Regular exercise reduces the risk of heart disease. The researcher would, therefore, like to know whether exercise has an effect on CRP concentration because this might indicate that exercise has an anti-inflammatory effect. To test this theory out, the researchers recruit 10 subjects to undergo a 6-month exercise-training program and CRP concentration is measured pre-, mid-way (3-months) and immediately post-intervention. The CRP concentrations pre-intervention were recorded in the crp\_pre variable, the CRP concentrations mid-way in the crp\_mid variable, and the post-intervention CRP concentrations in the crp\_post variable. The researcher would like to know whether there are changes in CRP concentration over time. In variable terms, the researcher would like to know if there are differences between the three variables, crp\_pre, crp\_mid and crp\_post.

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 Repeated Measures ANOVA data.



## **Check your assumptions:**

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

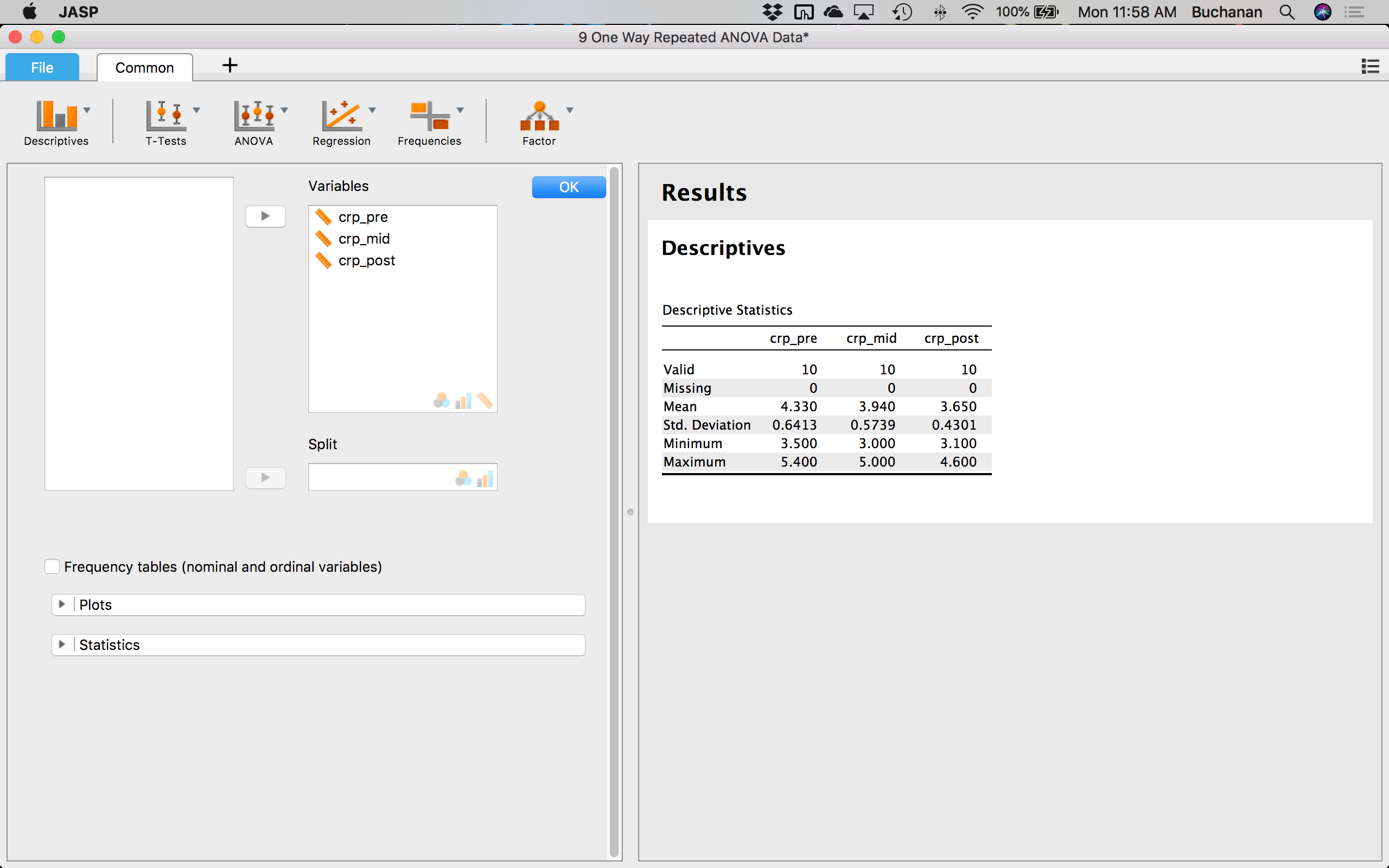
Yes, we are using ratio style data.

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

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 each of the measurements 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.

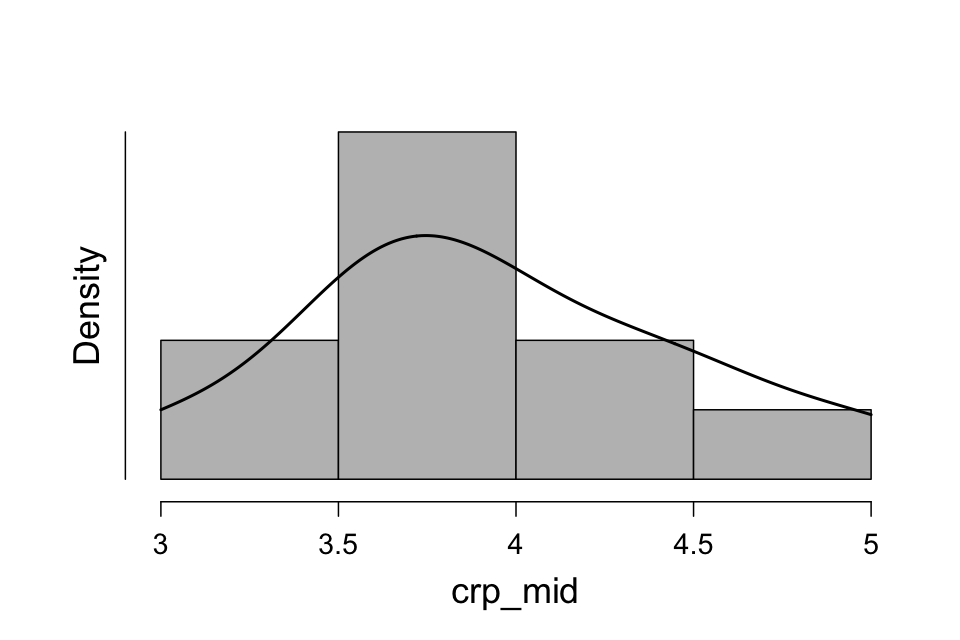
Here we can look at two different options to see if any participants scores are very different from other participants scores. First, click on Distribution plots.  You can also add display density A close up of a logo

Description automatically generated if you want to see the lines on the plot.

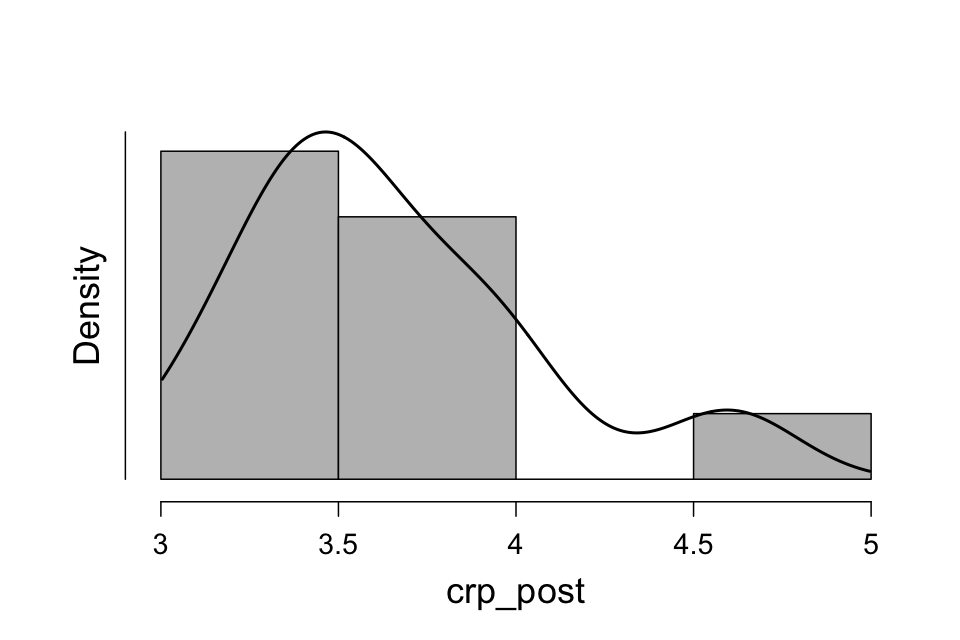
Another option would be to select Box Plots , Label Outliers , and Jitter Element . You will see outliers labeled with a special symbol, and Jitter Element allows you to see all the participants scores as dots on the plot.

#### Distribution plots

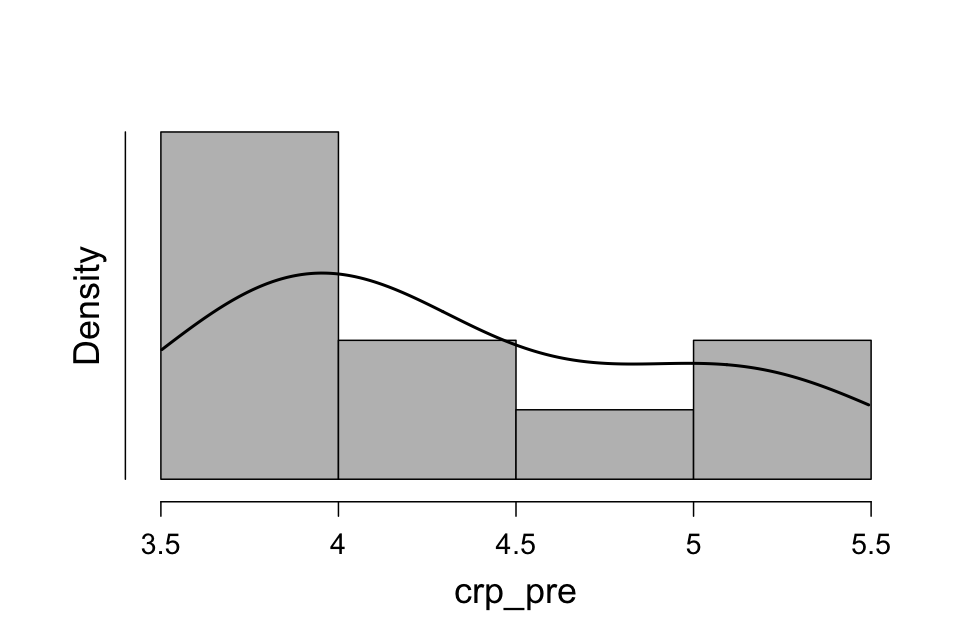
##### crp\_mid



##### crp\_post

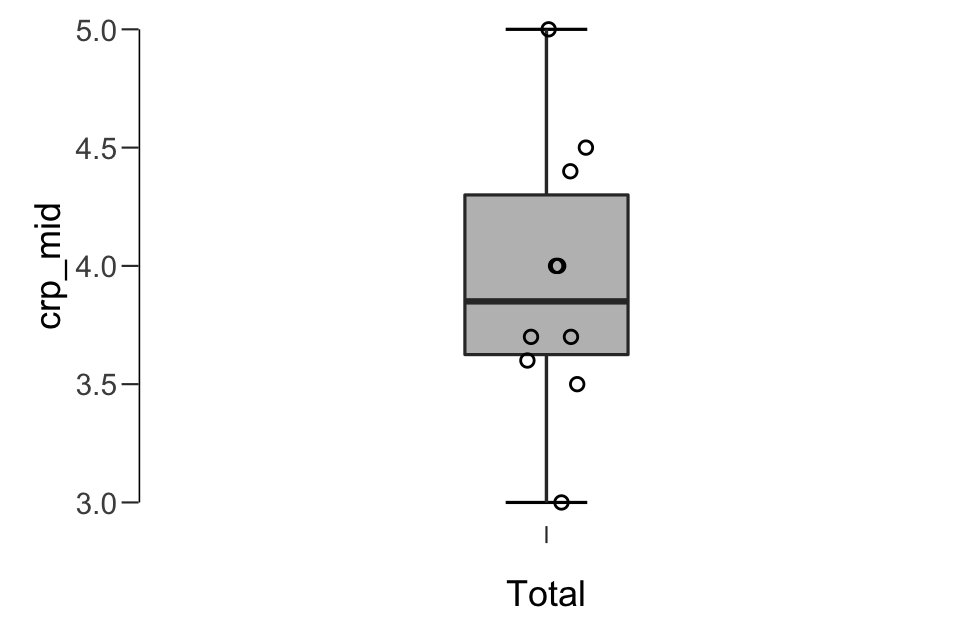


##### crp\_pre

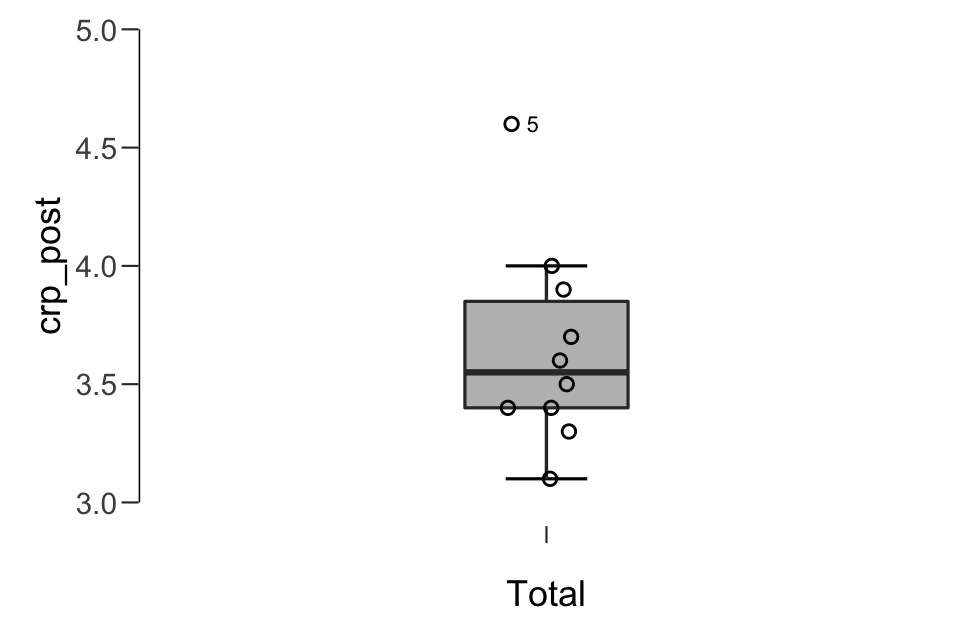


#### Boxplots

##### crp\_mid

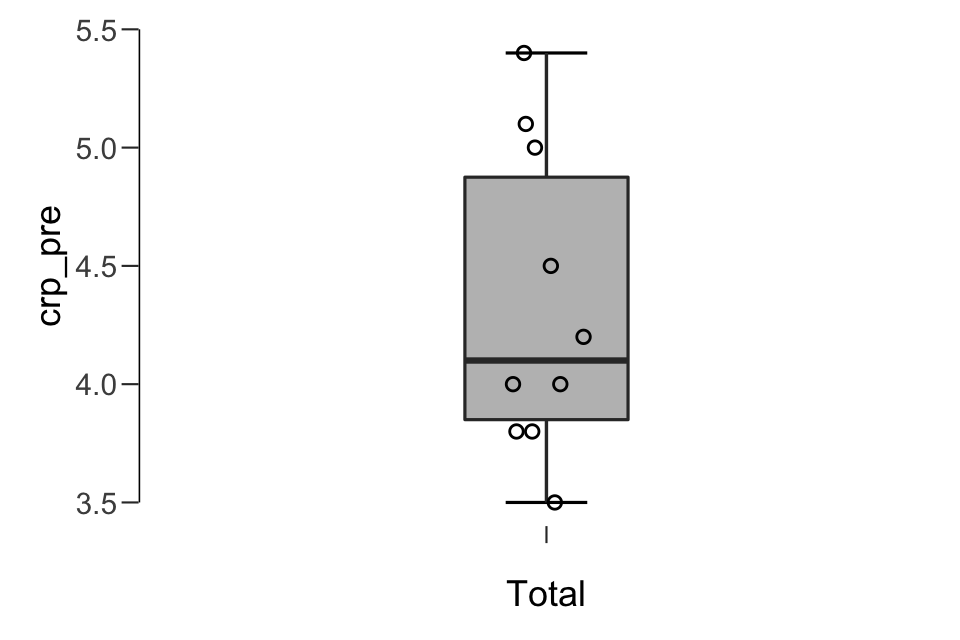


##### crp\_post



An Outlier

##### crp\_pre



##### Here we can see that the data seems to be piled on the bottom of the distribution (see below for normality tests), and we appear to have an outlier in the post test scores, as it is outside the box plot lines. The 5 indicates which ROW is the outlier for that variable.

##### 

##### We could exclude this participant for being an outlier – in this example, we will leave them in because we do not have very many participants.

**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, can simply click the Shapiro-Wilk option under Distribution on the Descriptives analysis we just conducted.

A screenshot of a cell phone

Description automatically generated

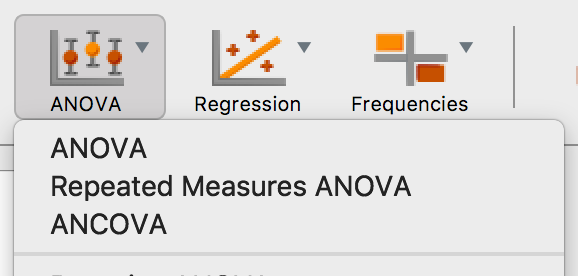
**Descriptive Statistics**

| **Descriptive Statistics** | | | | | | | |
| --- | --- | --- | --- | --- | --- | --- | --- |
|  | | **crp\_pre** | | **crp\_mid** | | **crp\_post** | |
| **Valid** |  | 10 |  | 10 |  | 10 |  |
| **Missing** |  | 0 |  | 0 |  | 0 |  |
| **Mean** |  | 4.330 |  | 3.940 |  | 3.650 |  |
| **Std. Deviation** |  | 0.6413 |  | 0.5739 |  | 0.4301 |  |
| **Shapiro-Wilk** |  | 0.9191 |  | 0.9713 |  | 0.9218 |  |
| **P-value of Shapiro-Wilk** |  | 0.350 |  | 0.903 |  | 0.372 |  |
| **Minimum** |  | 3.500 |  | 3.000 |  | 3.100 |  |
| **Maximum** |  | 5.400 |  | 5.000 |  | 4.600 |  |
|  | | | | | | | |

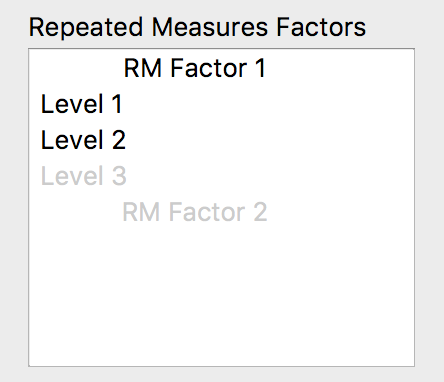
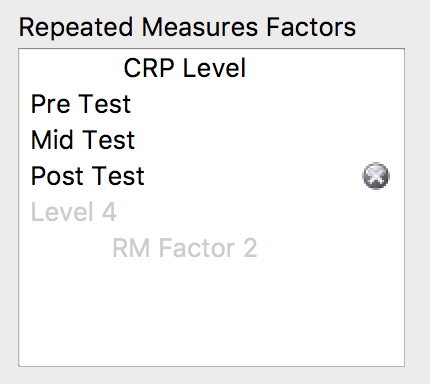
The Shapiro-Wilk test ran for each measurement separately, so we can tell if the assumptions were met for each measurement. We see that our data is normally distributed because *p* > .05 each of the measurements.

**Do you meet the assumption of sphericity?**

To get this test, we need to run the repeated measures ANOVA. Click on ANOVA  🡪 Repeated Measures ANOVA.

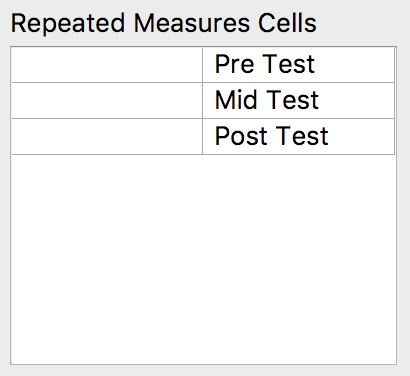
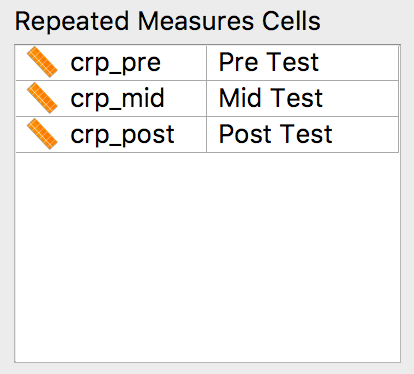


The first thing we can do in adjust the Repeated Measures Factors area to indicate what our experiment looks like. This window is where you label everything you are expecting to analyze. Click on the words to edit them.

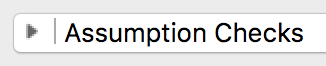
 

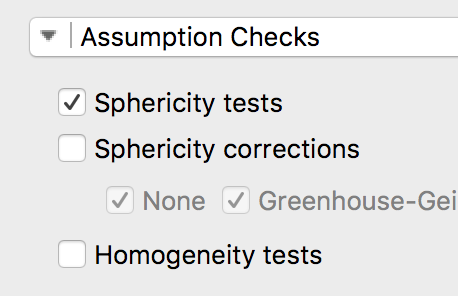
Edit and add levels

When you edit the Repeated Measures Factors window, the Repeated Measures Cells window changes to match.

Match the variables to the levels

Now, we can use the arrow to move the variables over into each of the appropriate boxes. You want to make sure they match so you know that the output is correct when you run the test. To view the sphericity assumption, click Assumption Checks  and Sphericity test to get Mauchly’s test.



### Assumption Checks

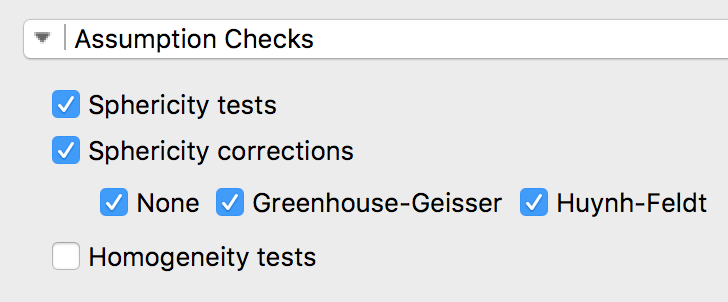
| **Test of Sphericity** | | | | | | | | | |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | | **Mauchly's W** | | **p** | | **Greenhouse-Geisser ε** | | **Huynh-Feldt ε** | |
| CRP Level |  | 0.457 |  | 0.043 |  | 0.648 |  | 0.711 |  |
|  | | | | | | | | | |

The assumption of sphericity is tested with Mauchly's Test for Sphericity. Sphericity is the condition where the variances of the differences between all combinations of related measurements (levels) are equal. Violation of sphericity is when the variances of the differences between all combinations of related groups are not equal. Sphericity can be likened to homogeneity of variances in a between-subjects ANOVA and can be tested for with Mauchly's Test of Sphericity. This test has been heavily criticized as it often fails to detect departures from sphericity in small samples and over-detects them in large samples, but it is very commonly used and very easy to interpret.

Mauchly's Test of Sphericity tests the null hypothesis that the variances of the differences are equal. Thus, if Mauchly's Test of Sphericity is statistically significant (p < .05), you can reject the null hypothesis and accept the alternative hypothesis that the variances of the differences are not equal (i.e., sphericity has been violated). The results of this test show that sphericity has been violated (p = .043).

Reporting Sphericity Results: Mauchly's Test of Sphericity indicated that the assumption of sphericity had been violated, *p* = .043.

If you have violated the assumption of sphericity, you will need to apply a correction to the repeated measures ANOVA so that the result is still valid. To get those results, click on Sphericity Corrections underneath Assumption Checks.

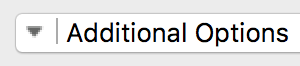


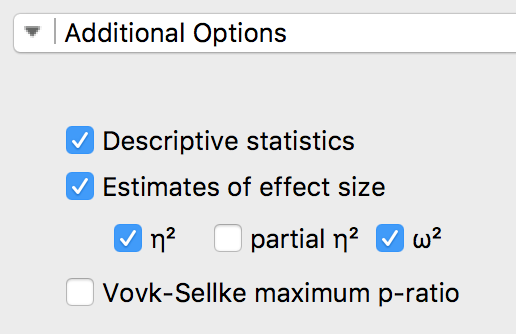
You can see that the table references the terms **Epsilon**, **Greenhouse-Geisser**, and **Huynh-Feldt**. The degree to which sphericity is present, or not, is represented by the statistic called **epsilon (ε)**. An epsilon of 1 (i.e., ε = 1) indicates that the condition of sphericity is exactly met. The further epsilon decreases below 1 (i.e., ε < 1), the greater the violation of sphericity. Therefore, you can think of epsilon as a statistic that describes the degree to which sphericity has been violated. The lowest value that epsilon (ε) can take is called the lower-bound estimate, while both the Greenhouse-Geisser and the Huynd-Feldt procedures attempt to estimate epsilon (ε) albeit in different ways (it is an estimate as we are dealing with samples, not populations). The estimates of sphericity (ε) tend to always be different depending on which procedure is used. After estimating epsilon (ε), all these procedures then use their sphericity estimate (ε) to correct the degrees of freedom for the F-distribution. In this way, these corrections attempt to overcome the fact that sphericity has been violated. Generally, the recommendation is to use the Greenhouse-Geisser correction, especially if estimated epsilon (ε) is less than 0.75. However, some statisticians recommend using the Huynh-Feldt correction if estimated epsilon (ε) is greater than 0.75. In practice, both corrections produce very similar corrections, so if estimated epsilon (ε) is greater than 0.75, you can easily justify using either.

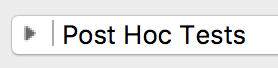
In this example, epsilon (ε) = 0.648, and therefore, the Greenhouse-Geisser correction will be used.

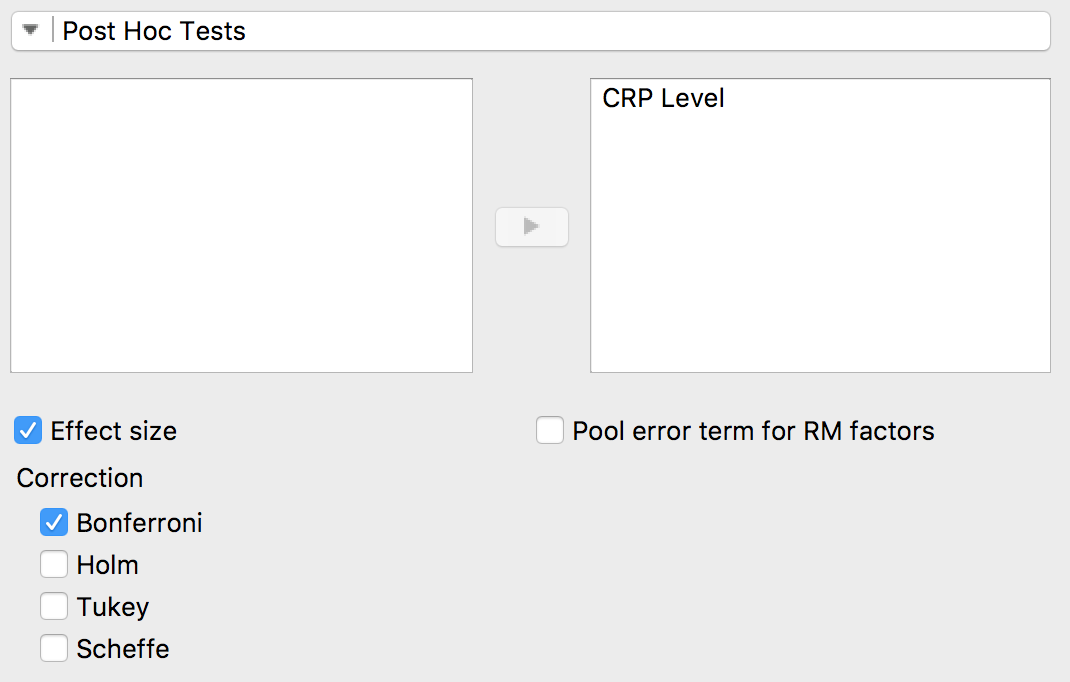
## **The ANOVA and effect size:**

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

You will want to add effect size η2 (eta squared) or ω2 (omega squared) for the overall test. Click on Additional Options  🡪 click on Descriptive Statistics, and Estimates of Effect Size.



To get the post hoc tests, click on Post Hoc Tests  🡪 Move the CRP Level to the right side. Click on effect size (*d*) to get the effect size for the pairwise tests and pick the correction suggested by your instructor (Holm is now automatic, so be sure to check which one you want!).



**Repeated Measures ANOVA**

| **Within Subjects Effects** | | | | | | | | | | | | | | | | | |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | | **Sphericity Correction** | | **Sum of Squares** | | **df** | | **Mean Square** | | **F** | | **p** | | **η²** | | **ω²** | |
| Time Point |  | None |  | 2.329 | ᵃ | 2.000 | ᵃ | 1.164 | ᵃ | 26.94 | ᵃ | < .001 | ᵃ | 0.218 |  | 0.195 |  |
|  |  | Greenhouse-Geisser |  | 2.329 | ᵃ | 1.296 | ᵃ | 1.797 | ᵃ | 26.94 | ᵃ | < .001 | ᵃ | 0.218 |  | 0.195 |  |
|  |  | Huynh-Feldt |  | 2.329 | ᵃ | 1.423 | ᵃ | 1.637 | ᵃ | 26.94 | ᵃ | < .001 | ᵃ | 0.218 |  | 0.195 |  |
| Residual |  | None |  | 0.778 |  | 18.000 |  | 0.043 |  |  |  |  |  |  |  |  |  |
|  |  | Greenhouse-Geisser |  | 0.778 |  | 11.663 |  | 0.067 |  |  |  |  |  |  |  |  |  |
|  |  | Huynh-Feldt |  | 0.778 |  | 12.803 |  | 0.061 |  |  |  |  |  |  |  |  |  |
|  | | | | | | | | | | | | | | | | | |
| Note.  Type III Sum of Squares | | | | | | | | | | | | | | | | | |
| ᵃ Mauchly's test of sphericity indicates that the assumption of sphericity is violated (p < .05). | | | | | | | | | | | | | | | | | |

The actual results of the repeated measures ANOVA are presented in the **Tests of Within-Subjects Effects** table. If your data did not violate the assumption of sphericity, you need to consult the "None" row under correction (because no need to correct if you meet the assumption). However, if sphericity was violated, you need to consult one of the other rows. In this example, the assumption of sphericity was violated, and epsilon (ε) = 0.648, so you need to consult the "Greenhouse-Geisser" rows (highlighted above).

The *p* column (p-value) indicates whether or not the repeated measures ANOVA is statistically significant (i.e., whether at least one mean is statistically significantly different from another mean or not). If p < .05, you can reject the null hypothesis and accept the alternative hypothesis that the group means are not equal. If p > .05, you must fail to reject the null hypothesis and conclude that the group means are equal. That is, not all group means are equal; somewhere, at least one group mean is different from another group mean. This is as far as you can go with a repeated measures ANOVA. In order to discover where the group mean differences lie, you will need to interpret the post-hoc tests that you ran as part of the repeated measures ANOVA procedure. If you found that your repeated measures ANOVA is not statistically significant, this is telling you that all group means are equal. In this case, you would not follow up the repeated measures ANOVA result with any post-hoc analysis, but just report the result of the repeated measures ANOVA.

Reporting ANOVA Results: CRP concentration was statistically significantly different at the different time points during the exercise intervention, *F*(1.30, 11.66) = 26.94, *p* < .001, η2 = .22.

|  |  |
| --- | --- |
| **Part** | **Meaning** |
| *F* | Indicates that we are comparing to a *F*-distribution (*F*-test). |
| 1.296 in (1.296, 11.663) | Indicates the corrected degrees of freedom for CRP level |
| 11.663 in (1.296, 11.663) | Indicates the corrected degrees of freedom for Error / Residual |
| 26.94 | Indicates the obtained value of the *F*-statistic (obtained *F*-value) |
| *p* < .001 | Indicates the probability of obtaining the observed *F*-value if the null hypothesis is correct. |
| η2 = .22 | A measure of effect size. |

| **Between Subjects Effects** | | | | | | | | | | | | | | | |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | | **Sum of Squares** | | **df** | | **Mean Square** | | **F** | | **p** | | **η²** | | **ω²** | |
| Residual |  | 7.552 |  | 9 |  | 0.839 |  |  |  |  |  |  |  |  |  |
|  | | | | | | | | | | | | | | | |
| Note.  Type III Sum of Squares | | | | | | | | | | | | | | | |

You will ignore this box because you do not use it for the one way repeated measures ANOVA.

**Assumption Checks**

| **Test of Sphericity** | | | | | | | | | |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | | **Mauchly's W** | | **p** | | **Greenhouse-Geisser ε** | | **Huynh-Feldt ε** | |
| CRP Level |  | 0.457 |  | 0.043 |  | 0.648 |  | 0.711 |  |
|  | | | | | | | | | |

**Post Hoc Tests**

| **Post Hoc Comparisons - CRP Level** | | | | | | | | | | | | | |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | |  | | **Mean Difference** | | **SE** | | **t** | | **Cohen's d** | | **p bonf** | |
| Pre Test |  | Mid Test |  | 0.390 |  | 0.050 |  | 7.732 |  | 2.445 |  | < .001 |  |
|  |  | Post Test |  | 0.680 |  | 0.115 |  | 5.899 |  | 1.865 |  | < .001 |  |
| Mid Test |  | Post Test |  | 0.290 |  | 0.100 |  | 2.886 |  | 0.913 |  | 0.054 |  |
|  | | | | | | | | | | | | | |
| *Note.*  Cohen's d does not correct for multiple comparisons. | | | | | | | | | | | | | |

Each of the pairwise combinations is shown above. Remember that pre compared to mid is the same thing as mid compared to pre, so each pairwise combination is only shown once. You would want to check out the *p* values to see if they are less than your alpha (i.e., *p* < .05). For each one, you also get the effect size for just that combination.

|  |  |
| --- | --- |
| **Column Name** | **Column Meaning** |
| Mean Difference (I - J) | Mean difference between measurement I and measurement J (I minus J) |
| Std. Error | Standard error of the difference between measurements I and J |
| t | The dependent *t*-test value for group I compared to J |
| d | Cohen’s *d* for the comparison of group I to J, which is the same as a dependent *t*-test Cohen’s *d* value. |
| p | Significance level (*p*-value) of the difference between group I and J. Notice that the *p* value says bonf next to it, indicating that these *p* values are corrected or adjusted for the number of comparisons that you could possibly run. Therefore, this *p* value accounts for the fact that we ran three comparisons. If you were to run a dependent *t*-test for each of these comparisons, these values would be different because those would not be corrected. |

### Marginal Means

| **Marginal Means - Time Point** | | | | | | | | | |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | | | | | | **95% CI** | | | |
| **Time Point** | | **Marginal Mean** | | **SE** | | **Lower** | | **Upper** | |
| Pre |  | 4.330 |  | 0.176 |  | 3.943 |  | 4.717 |  |
| Mid |  | 3.940 |  | 0.176 |  | 3.553 |  | 4.327 |  |
| Post |  | 3.650 |  | 0.176 |  | 3.263 |  | 4.037 |  |
|  | | | | | | | | | |

**Descriptives**

| **Descriptives** | | | | | | | |
| --- | --- | --- | --- | --- | --- | --- | --- |
| **CRP Level** | | **Mean** | | **SD** | | **N** | |
| Pre Test |  | 4.330 |  | 0.641 |  | 10 |  |
| Mid Test |  | 3.940 |  | 0.574 |  | 10 |  |
| Post Test |  | 3.650 |  | 0.430 |  | 10 |  |
|  | | | | | | | |

**Reporting Results:** There was a decrease in CRP concentration pre-intervention (*M* = 4.33, *SD* = 0.64 mg/mL) to 3 months into the exercise intervention (*M* = 3.94, *SD* = 0.57 mg/mL), a statistically significant mean decrease of 0.39 mg/mL, *p* < .001, *d* = 2.45.

## **Reporting All Together:**

A repeated measures ANOVA was conducted to determine whether there were statistically significant differences in CRP concentration over the course of a 6-month exercise intervention. There was one outlier in the post test measurement, and the data was normally distributed for each group, as assessed by boxplot and Shapiro-Wilk test (ps = .350, .903, .372), respectively. The assumption of sphericity was violated, as assessed by Mauchly's Test of Sphericity, p = .043. Therefore, a Greenhouse-Geisser correction was applied (ε = 0.648). The exercise intervention elicited statistically significant changes in CRP concentration over time, F(1.30, 11.66) = 26.94, p < .001, η2 = .22, with CRP concentration decreasing from pre-intervention (M = 4.33, SD = 0.64 mg/mL) to 3 months (M = 3.94, SD = 0.57 mg/mL) to 6 months (post-intervention) (M = 3.65, SD = 0.43 mg/mL). Post-hoc analysis with a Bonferroni adjustment revealed that CRP concentration was statistically significantly decreased from pre-intervention to 3-months (M = 0.39 mg/mL, *SE* = 0.05, p < .001, *d* = 2.45), and from pre-intervention to post-intervention (M = 0.68 mg/mL, *SE* = 0.12, p = .001, *d* = 1.87) but not from 3 months to post-intervention (M = 0.29 mg/mL, *SE* = 0.10, *p* = .054, *d* = 0.91).