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).

The null and alternative hypotheses for a repeated measures ANOVA are:

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.

H1: 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.

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 & order of testing

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:

1. There are no outliers in any of the groups.  
2. Each group's data is normally distributed.  
3. The variances of the differences between related groups are equal (the assumption of sphericity).

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. The preferred order of testing of these assumptions, and the order followed in this guide, is as follows:

1. Detect any outliers.  
2. Determine if data is approximately normally distributed in each group.  
3. Determine if there is sphericity.

Depending on the presence of outliers or the normality of the data, you will be instructed how to proceed with testing. This could include making corrections to your data or using an alternative statistical test. This will also include the route to take when, and if, any corrections you make still results in the data failing the assumptions of this test.

**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.

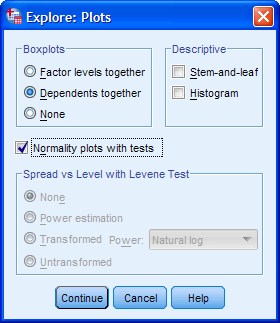
The following instructions show you how to run tests to detect outliers and check if your data is normally distributed:

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

**2.** Transfer the three variables, crp\_pre, crp\_mid and crp\_post into the Dependent List: box by clicking on them 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:

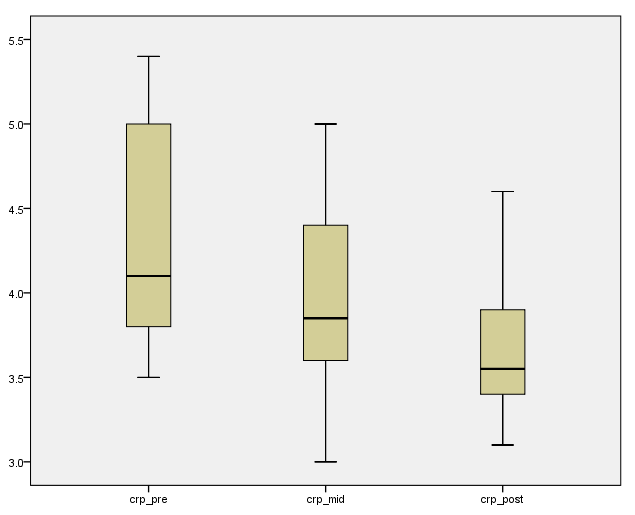


Note: The "tests" part of Normality plots with tests will generate the Kolmogorov-Smirnov and Shapiro-Wilk tests of normality. Selecting Dependents together results in one boxplot for all variables rather than a separate boxplot for each variable.

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.

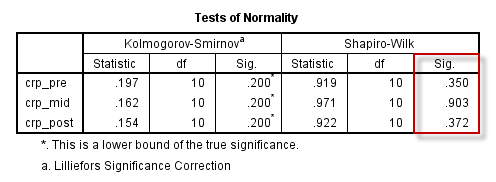
**1. Are there any outliers in our data?**



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.

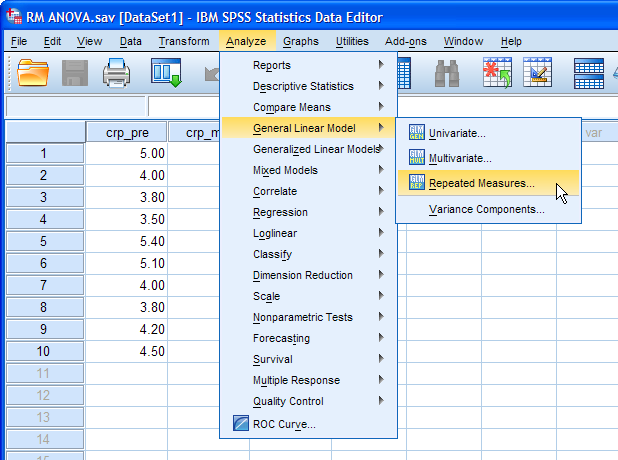
**2. Are our data normally distributed?**



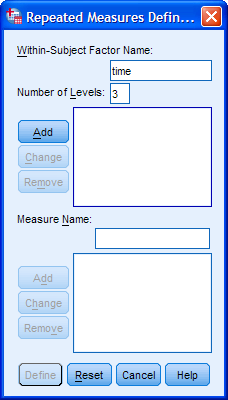
If the assumption of normality has been violated, the "**Sig.**" value will be less than .05 (i.e., the test is significant at the p < .05 level). If the assumption of normality has not been violated, the "**Sig.**" value will be greater than .05 (i.e. p > .05). This is because the Shapiro-Wilk test is testing the null hypothesis that your data's distribution is equal to a normal distribution. Rejecting the null hypothesis means that your data's distribution is not equal to a normal distribution. In the table above, you can see that all of the "**Sig.**" values are greater than .05 (they are .350, .903 and .372). Therefore, crp\_pre, crp\_mid and crp\_post are all normally distributed.

One-Way Repeated Measures ANOVA

1. Click **Analyze > General Linear Model > Repeated measures...** on the top menu, as shown below:

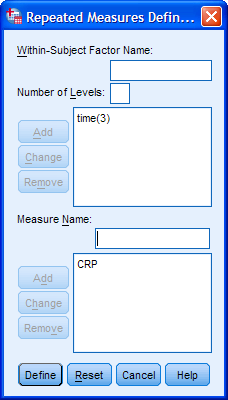


2. In the Within-Subject Factor Name: box, replace "factor1" with a more meaningful name for your independent variable. In this example, you will replace it with time, as this is the independent variable. Enter into the Number of Levels: box the number of time points (i.e., the number of levels of the independent variable). In this case, enter "3", representing crp\_pre, crp\_mid and crp\_post, as shown below:



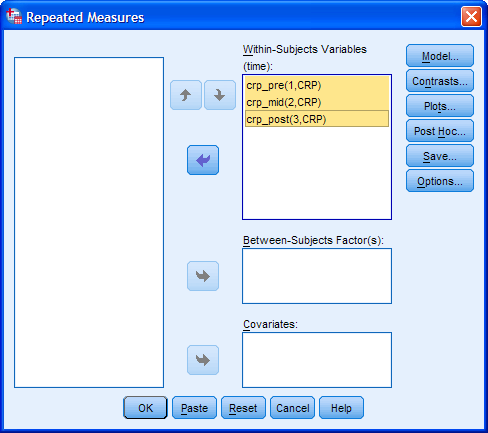
Click the https://statistics.laerd.com/premium/rma/img/add-button.pngbutton.

Put an appropriate name into the Measure Name: box. Basically, this is the name of the dependent variable, which is CRP in this example. Therefore, enter CRP and click the https://statistics.laerd.com/premium/rma/img/add-small-d-button.pngbutton, and you will end up with the screen below:



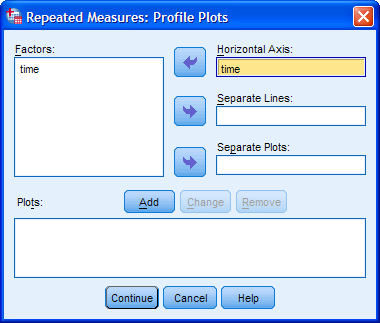
3. Click the https://statistics.laerd.com/premium/rma/img/define-button.pngbutton and you will be presented with the **Repeated Measures** dialogue box.

4. Transfer crp\_pre, crp\_mid and crp\_post into the Within-Subjects Variables (time): box by highlighting all the variables (clicking on them while holding down the shift-key) in the left-hand box and clicking the top https://statistics.laerd.com/premium/rma/img/right-arrow-button.pngbutton. You will end up with the following screen:



5. Click the https://statistics.laerd.com/premium/rma/img/plots-button.pngbutton. You will be presented with the **Repeated Measures: Profile Plots** dialogue box.

6. Transfer time from the Factors: box into the Horizontal Axis: box by clicking the top https://statistics.laerd.com/premium/rma/img/right-arrow-button.pngbutton. You will end up with the following screen:

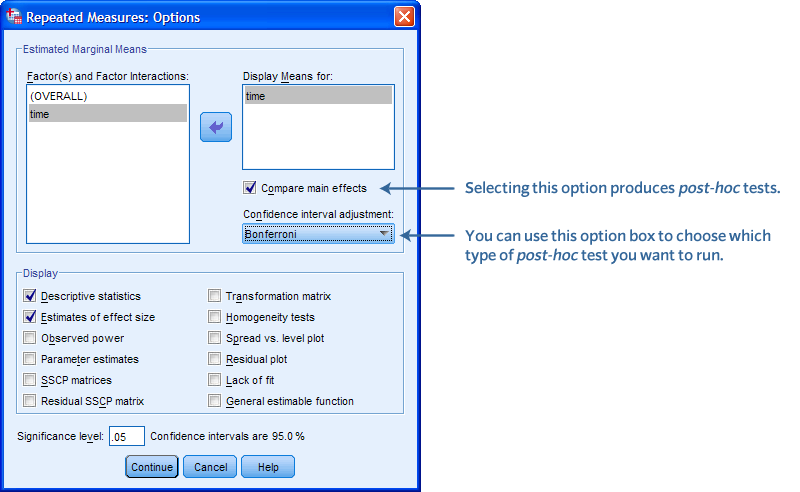


7. Click the https://statistics.laerd.com/premium/rma/img/add-button.pngbutton and you will transfer the plot into the Plots: box.

8. Click the https://statistics.laerd.com/premium/rma/img/continue-button.pngbutton and you will returned to the **Repeated Measures** dialogue box.

9. Click the https://statistics.laerd.com/premium/rma/img/options-button.pngbutton. You will be presented with the **Repeated Measures: Options** dialogue box.

10. Transfer time from the Factor(s) and Factor Interactions: box to the Display Means for: box using the https://statistics.laerd.com/premium/rma/img/right-arrow-button.pngbutton. This will activate the Compare main effects checkbox (i.e., it will no longer be grayed out). Tick this checkbox and select "Bonferroni" from the drop-down menu under Confidence interval adjustment:. In the -Display- area, tick the Descriptive statistics and Estimates of effect size checkboxes. After you have done all this, you will be presented with the following screen:



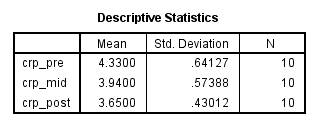
Note: The Bonferroni correction is generally considered one of the most suitable adjustments for making multiple post-hoc comparisons for a repeated measures ANOVA.

11. Click the https://statistics.laerd.com/premium/rma/img/continue-button.pngbutton and you will returned to the **Repeated Measures** dialogue box.

12. Click the https://statistics.laerd.com/premium/rma/img/ok-button.pngbutton. This will generate the output.

## Descriptive statistics and estimates

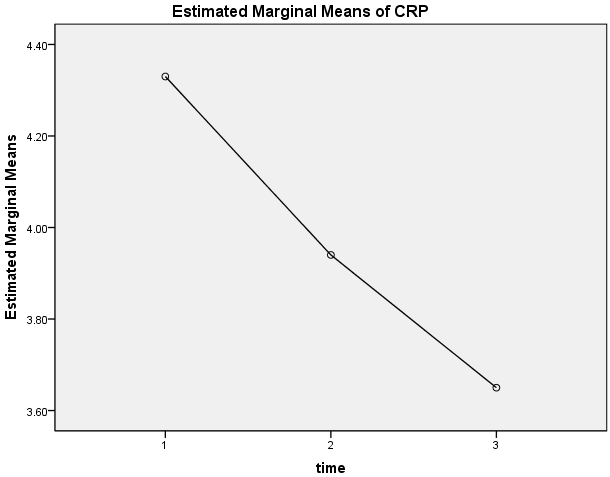
The **Descriptive Statistics** table (shown below) provides the necessary information to describe your dependent variable based on the levels of your independent variable:



You should use the statistics in the **Descriptive Statistics** table to see if there are any obvious trends or points of interest in your data. In this example, you can see that CRP concentration decreased at each successive time point. That is, the effect of the exercise-training programme was to keep decreasing CRP as the intervention continued.

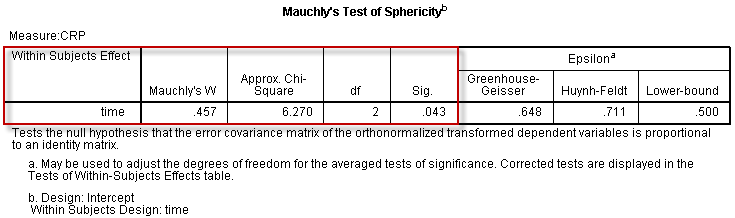
## Profile Plot

The profile plot that is generated should also be used to see if there are any obvious trends or points of interest. From the profile plot below, you can clearly see that CRP concentration decreased at each successive time point, with the decrease from time point "1" to "2" seemingly slightly greater than from time point "2" to "3":



## Assumption #3: Mauchly's Test of Sphericity

The assumption of sphericity is tested with Mauchly's Test for Sphericity found in the aptly named **Mauchly's Test of Sphericity** table. Sphericity is the condition where the variances of the differences between all combinations of related groups (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. This is probably due to its automatic print out in SPSS for repeated measures ANOVA and the lack of an otherwise readily available test.

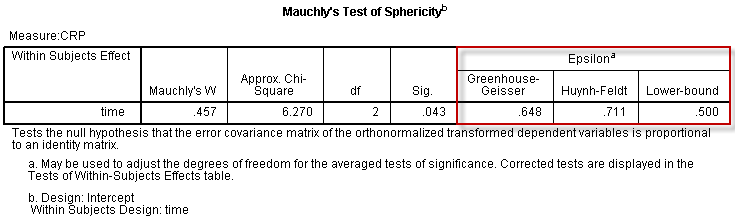


Mauchly's Test of Sphericity tests the null hypothesis that the variances of the differences are equal. Thus, if Mauchly's Test of Sphericity is statistically significant (p < .05), you can reject the null hypothesis and accept the alternative hypothesis that the variances of the differences are not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

Reporting Sphericity Results: Mauchly's Test of Sphericity indicated that the assumption of sphericity had been violated, χ2(2) = 6.270, p = .043.

## Violation of the assumption of sphericity

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. These corrections are located in the right-hand columns of the **Mauchly's Test of Sphericity** table, as shown below:



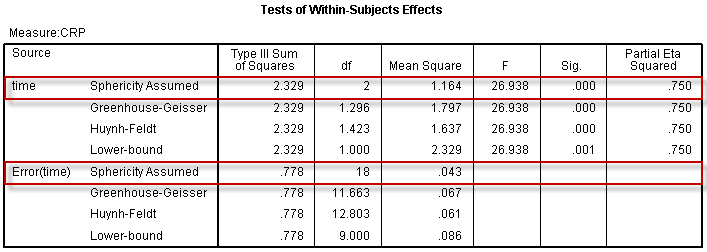
You can see that the table references the terms **Epsilon**, **Greenhouse-Geisser**, **Huynh-Feldt**, and **lower-bound**. The degree to which sphericity is present, or not, is represented by the statistic called **epsilon (ε)**. An epsilon of 1 (i.e., ε = 1) indicates that the condition of sphericity 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. Firstly, **never use the lower-bound estimate**. 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 Huynd-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.

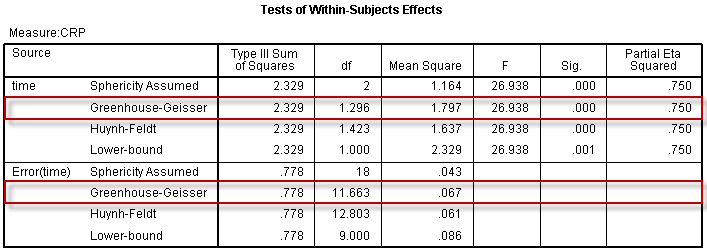
## Repeated measures ANOVA

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 "Sphericity Assumed" rows (and ignore the rest), as shown below:



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 as shown below:



The "**Sig.**" 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. The p-value in this example would appear to be .000 (obtained from the "**Sig.**" column). However, if you ever see SPSS print out a p-value of .000, do not interpret this as a significance level that is actually zero; it actually means p < .0005. Therefore, as p < .0005 in this example, you can conclude that the repeated measures ANOVA is statistically significant. 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(1.296, 11.663) = 26.938, p < .001, partial η2 = 0.750.

The breakdown of the last part (i.e., *F*(1.296, 11.663) = 26.938, *p* < .0005, partial η2 = 0.750) is as follows:

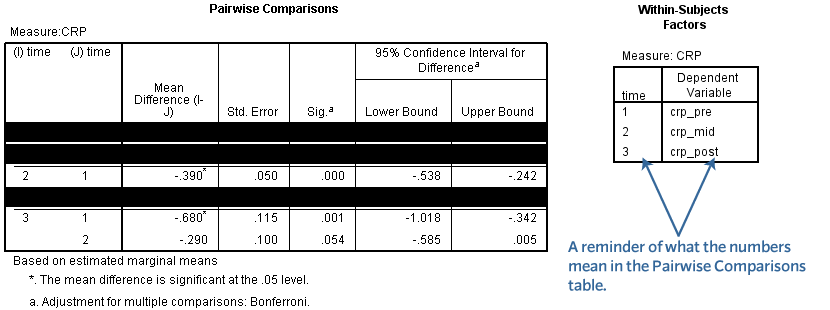
|  |  |
| --- | --- |
| **Part** | **Meaning** |
| *F* | Indicates that we are comparing to a *F*-distribution (*F*-test). |
| 1.296 in (1.296, 11.663) | Indicates the degrees of freedom for time |
| 11.663 in (1.296, 11.663) | Indicates the degrees of freedom for Error(time) |
| 26.938 | Indicates the obtained value of the *F*-statistic (obtained *F*-value) |
| *p* < .0005 | Indicates the probability of obtaining the observed *F*-value if the null hypothesis is correct. |
| partial η2 = 0.750 | A measure of effect size. |

## Post-hoc analysis

The Bonferroni post-hoc test is a good test if you wish to compare all possible combinations of group differences. This post-hoc test provides both confidence intervals for the differences between group means and whether the differences are statistically significant.

With three groups there will be three possible combinations of group differences, but in the **Pairwise Comparisons** table above, there are six combinations. This is because the data is repeated twice for each group combination (e.g., Group 1 vs. Group 2 and

then the reverse, Group 2 vs. Group 1). You can, therefore, remove the duplicates in the table to leave only the three unique combinations, as shown below:



As the "**Sig.**" value (p-value) is less than .05 (it is p < .0005), the difference between these two groups is statistically significant (i.e., their group means are not equal). In the above example, 95% confidence intervals have been used to provide a measure of spread of the differences, but it is perfectly acceptable to use the standard error ("**Std. Error**" column) if that is what you prefer (or have been taught). You can now work through each comparison in turn and report the result. Alternatively, you can report the results in a table or graph.

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, 95% CI [0.24, 0.54], p < .001.

## Putting it all together

In conclusion, reporting all the findings, and adding in the information about the assumptions run, you can write the complete results as:

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 were no outliers and the data was normally distributed for each group, as assessed by boxplot and Shapiro-Wilk test (p > .05), respectively. The assumption of sphericity was violated, as assessed by Mauchly's Test of Sphericity, χ2(2) = 6.270, 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(1.298, 11.663) = 26.938, p < .001, partial η2 = 0.750, 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, 95% CI [0.24 to 0.54], p < .001), and from pre-intervention to post-intervention (M = 0.68 mg/mL, 95% CI [0.34 to 1.02], p = .001) but not from 3 months to post-intervention (M = 0.29 mg/mL, 95% CI [-0.01 to 0.59], p > .05).