# **Introduction:**

A one-way ANOVA is often described as an omnibus test because while it informs you whether there are any differences between the means of three or more groups, it does not inform you which group(s) differ from the others.

**Planned contrasts and post-hoc tests**

To know which group mean(s) differ from other group mean(s), you can run either planned contrasts (usually abbreviated to just "contrasts") or post-hoc tests. Contrasts are specific group comparisons you want to make and are decided before you analyze the data (i.e., *a priori* tests). Alternatively, post-hoc tests are for when you have no specific hypotheses about any particular group mean comparison(s). They test all possible group comparisons and are the most common approach to discovering differences between the groups.

To investigate which groups differ, you can use post-hoc tests. The options available in this guide will focus on post-hoc tests that examine all possible group combinations. The total number of possible comparisons that you will have will depend on how many groups you have in your independent variable. However, the total possible number of comparisons can be calculated using the following formula:

Number of comparisons = *k* (*k* – 1) / 2

where *k* = number of groups. So, for three groups there will be 3(3 – 1) / 2 = 3 possible comparisons and for four groups (as in this example) there will be six possible comparisons as 4(4 – 1) / 2 = 6.

When you have homogeneity of variances, a commonly used and generally good post-hoc correction is the Tukey correction on an independent *t*-test (also called the Tukey HSD).

**Effect sizes**

The one-way ANOVA, as a null hypothesis significance test, informs you whether the differences between group means are 'real', but it does not inform you of the 'size' of the difference. To try to overcome this limitation, an effect size can be calculated. There can be many different types of effect size, with different types often trying to 'capture' the importance of your results in different ways. Effect sizes have their limitations but, nonetheless, they are becoming an important part of reporting your results and, as such, you will be shown how to calculate an either η2 (eta squared) or ω2 (omega squared) for the one-way ANOVA.

**Sample size and (un)balanced designs**

Generally, your study should have six or more participants in each group in order to proceed with a one-way ANOVA, but ideally you would have more. A one-way ANOVA will run with less than six participants, but your ability to infer to a larger population will be more difficult.

If you have an equal numbers of cases (e.g., participants) in each group, you have what is called a 'balanced' design. On the other hand, if sample sizes are not the same in all groups, you have an 'unbalanced' design. Generally speaking, the more unbalanced the design, the greater the negative effect of any violation of assumptions on the validity of the test. Ideally, you want a balanced design (although this can be hard to achieve in practice).

**What problems can you solve using a one-way ANOVA?**

A one-way ANOVA is most often used for three types of study design:

**1. Determine if there are differences between three or more independent groups**

You have a study design where you are measuring the same dependent variable in three or more independent (i.e., unrelated) groups (and you want to know if there are differences between groups), a one-way ANOVA might be appropriate. For example, you could use a one-way ANOVA to compare maximal aerobic capacity in swimmers, runners and cyclists. Another example could be to investigate any differences in blood cholesterol concentration in sedentary, low, moderate and high physical activity groups.

**2. Determine if there are differences between conditions (with no pre-test measurement taken)**

If you have a study design where three or more independent groups have performed different interventions (e.g., control/interventions) and the same dependent variable is measured at the end of the study in all groups, a one-way ANOVA might be appropriate. For example, one group acted as a control, one group underwent an exercise-training program and another group underwent a dietary program. Body weight was measured at the end of each program and compared between the three separate groups to determine if there were any statistically significant mean differences between the groups.

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

If you have a study design where three or more independent groups have performed different interventions (e.g., control/interventions), the same dependent variable is measured at the beginning and end of the study in all groups, and a change score calculated (i.e., post-values minus pre-values), a one-way ANOVA might be appropriate. For example, pre- and post- blood glucose concentration measurements were taken and change scores calculated for an exercise intervention group, dietary intervention and a control group. These change scores were then compared between the three groups using a one-way ANOVA. This will determine whether the changes in blood glucose concentration between groups were equal or if there were statistically significant differences in change score (i.e., the intervention type had a differential effect on change in blood glucose concentration).

# **Assumptions of the Between Subjects ANOVA:**

* Assumption #1: Your dependent variable should be measured at the interval or ratiolevel (i.e., they are continuous). Examples of variables that meet this criterion 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:** Your independent variable should consist of two or more categorical, independent groups. Typically, a one-way ANOVA is used when you have three or more categorical, independent groups, but it can be used for just two groups (but an independent-samples t-test is more commonly used for two groups). Example independent variables that meet this criterion include ethnicity (e.g., 3 groups: Caucasian, African American and Hispanic), physical activity level (e.g., 4 groups: sedentary, low, moderate and high), profession (e.g., 5 groups: surgeon, doctor, nurse, dentist, therapist), and so forth.
* Assumption #3: You should have independence of observations, which means that there is no relationship between the observations in each group or between the groups themselves. For example, there must be different participants in each group with no participant being in more than one group. This is more of a study design issue than something you can test for, but it is an important assumption of the one-way ANOVA. If your study fails this assumption, you will need to use another statistical test instead of the one-way ANOVA (e.g., a repeated measures design).
* Assumption #4: There should be no significant outliers. Outliers are simply single data points within your data that do not follow the usual pattern (e.g., in a study of 100 students' IQ scores, where the mean score was 108 with only a small variation between students, one student had a score of 156, which is very unusual, and may even put her in the top 1% of IQ scores globally). The problem with outliers is that they can have a negative effect on the one-way ANOVA, reducing the validity of your results.
* Assumption #5: Your dependent variable should be approximately normally distributed for each group of the independent variable. Also, it should be noted that if the distributions are all skewed in a similar manner (e.g., all moderately negatively skewed), this is not as troublesome as when compared to the situation where you have groups that have differently-shaped distributions (e.g., the distribution of Group A is moderately 'positively' skewed, whilst the distribution of Group B is moderately 'negatively' skewed). Therefore, in this example, you need to investigate whether engagement is normally distributed. There are many different methods available to test this assumption. Other methods include skewness and kurtosis values and histograms.
* Assumption #6: There needs to be homogeneity of variances. You can test this assumption in JASP using Levene's test for homogeneity of variances. If your data fails this assumption, you will need to not only carry out a different ANOVA instead of a one-way ANOVA but also use a different post-hoc test.

These assumptions need to be tested before you can run a one-way ANOVA. Fortunately, the one-way 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 group means are equal (i.e., µ1 = µ2 = µ3 = ... = µk). There are no mean differences between GROUP1/GROUP2/GROUP3… on the dependent variable.

The alternative hypothesis is:

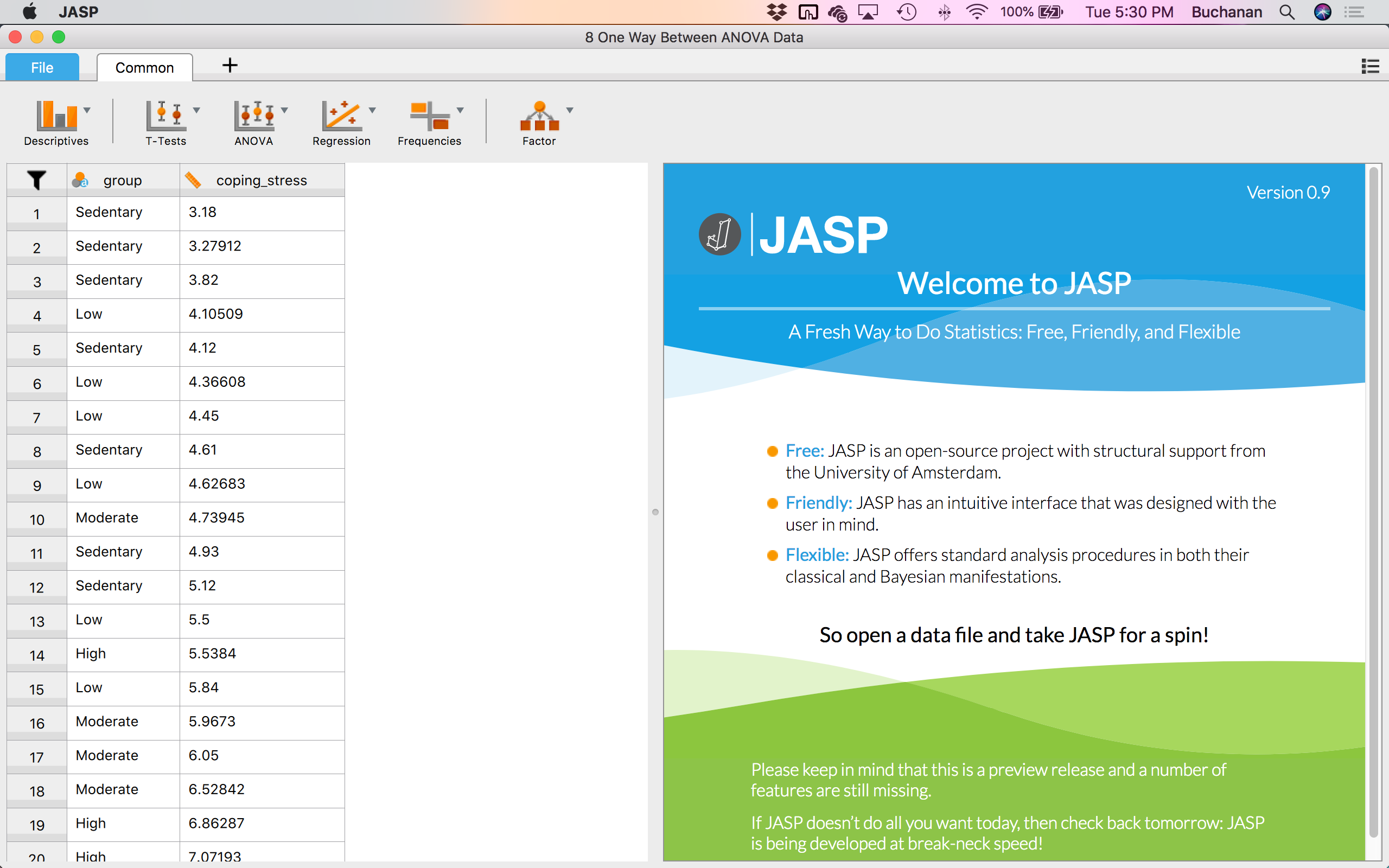
HA: at least one group mean is different (i.e., they are not all the same). There are mean differences between GROUP1/GROUP2/GROUP3… on the dependent variable.

## **Example:**

A researcher believes that individuals that are more physically active are better able to cope with stress in the workplace. To test this theory, the researcher recruited 31 subjects and measured how many minutes of physical activity they performed per week and their ability to cope with workplace stress. The subjects were categorized into four groups based on the number of minutes of physical activity they performed: namely, "sedentary", "low", "moderate" and "high" physical activity groups. These groups (levels of physical activity) formed an independent variable called group. The ability to cope with workplace stress was assessed as the average score of a series of Likert items on a questionnaire, which allowed an overall "coping with workplace stress" score to be calculated; higher scores indicating a greater ability to cope with workplace-related stress. This dependent variable was called stress\_coping and "ability to cope with workplace-related stress" abbreviated as "CWWS" score. The researcher would like to know if CWWS score is dependent on physical activity level. In variable terms, is stress\_coping different for different levels of group?

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 One Way Between Subjects Data.



## **Check your assumptions:**

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

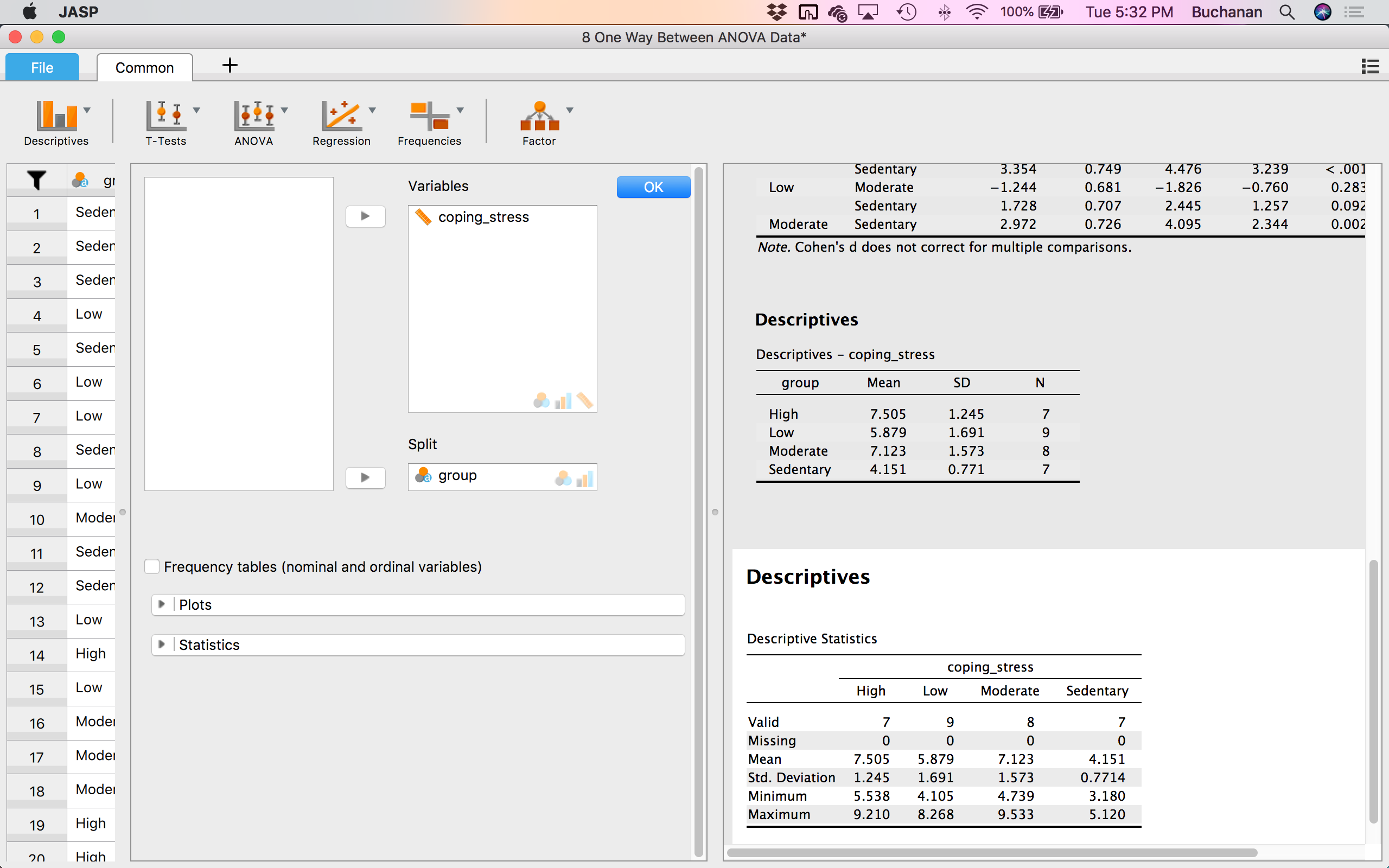
Yes, we are using ratio style data.

**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 coping\_stress and click the arrow  to move it over to the right hand side under Variables. In this window, you will see a new  Split option. This option will allow you to separate out descriptives and outlier analysis by group! Move the group variable into the Split box.



Click on the plots options:  to see more available options.

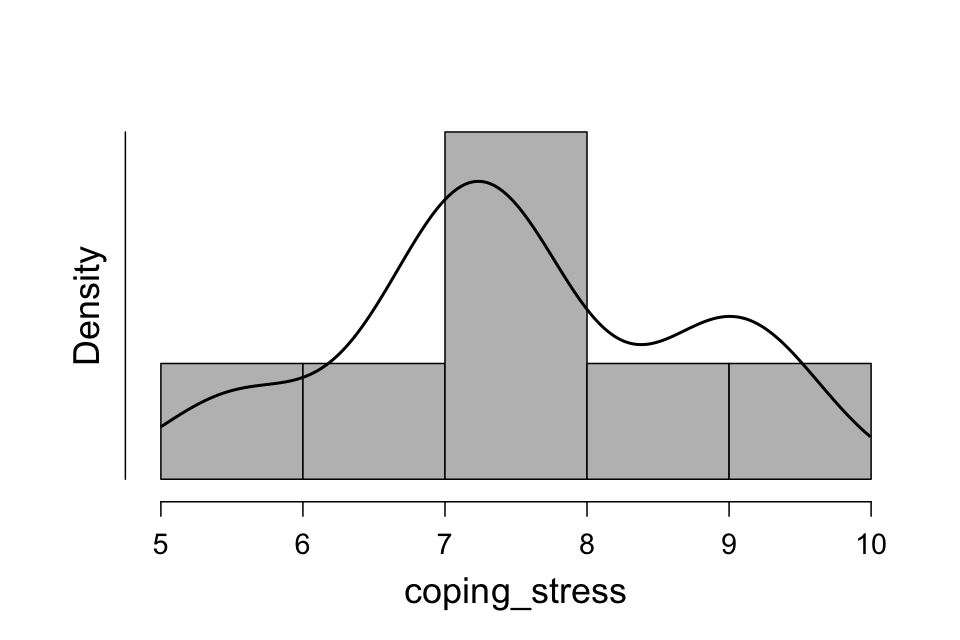
Here we can look at two different options to see if any participants scores are very different from other participants scores. First, click on Distribution plots.  You can also add display density A close up of a logo

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

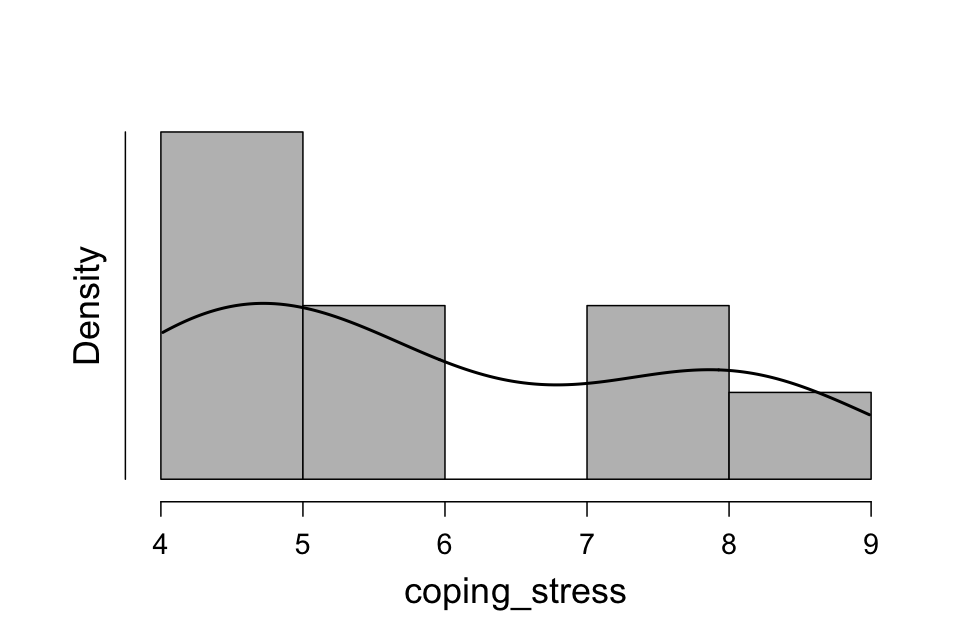
#### Distribution plots

##### coping\_stress

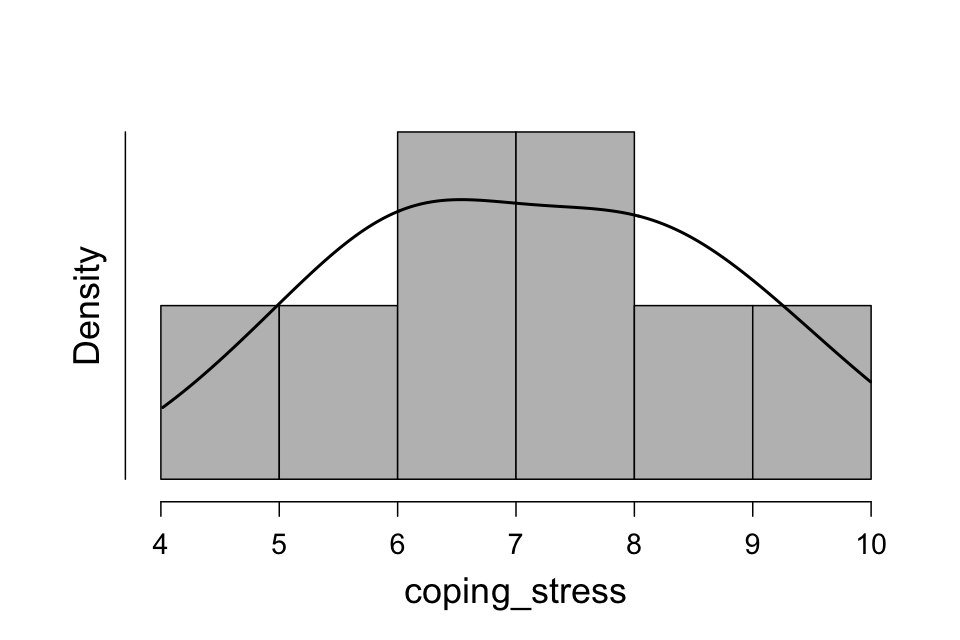
###### High



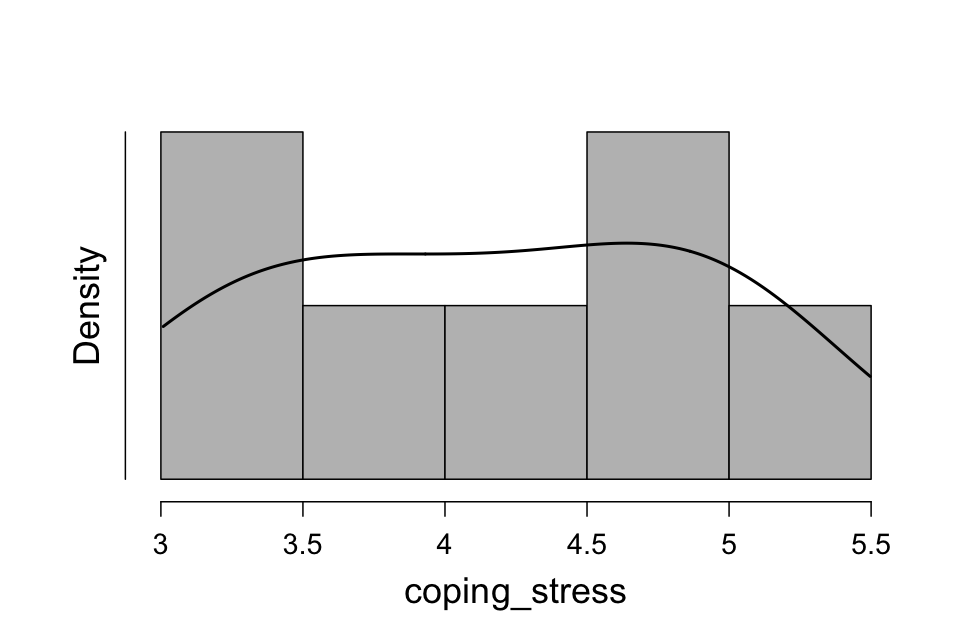
###### Low



###### Moderate



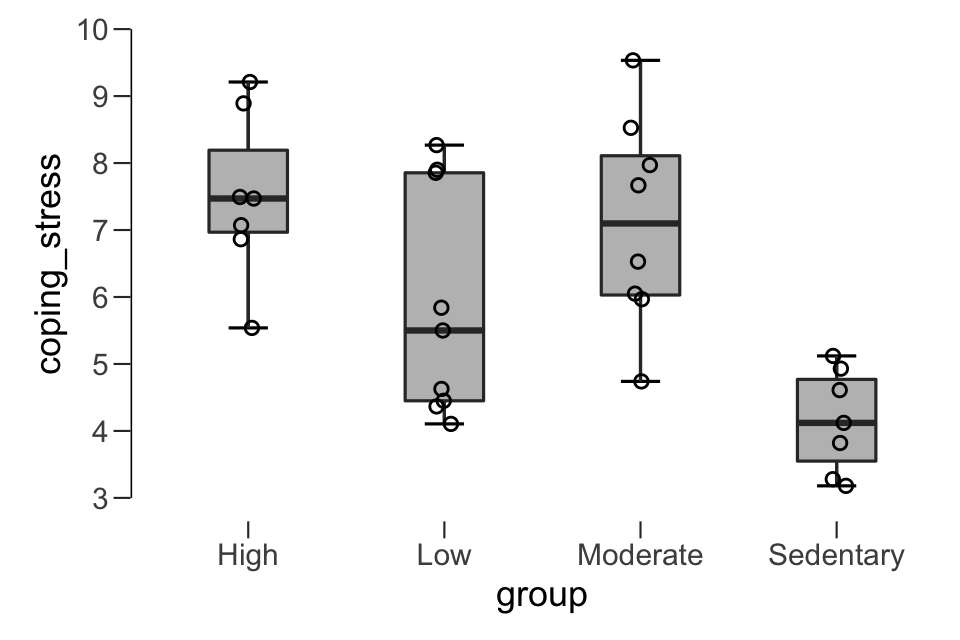
###### Sedentary



Four histograms will appear – one for each of the groups (levels) of the independent variable. 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.

##### coping\_stress



##### 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, 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**

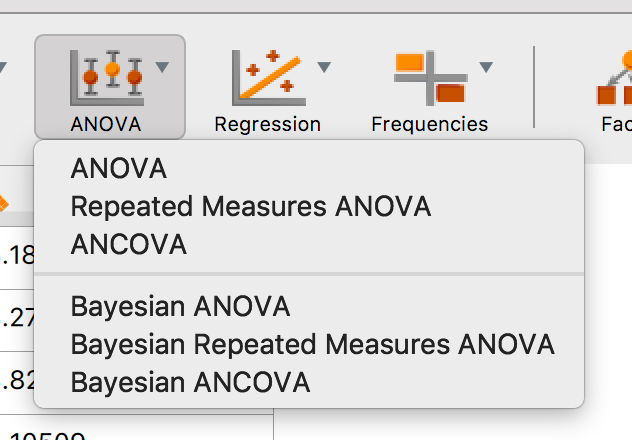
| **Descriptive Statistics** | | | | | | | | | |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | | **coping\_stress** | | | | | | | |
|  | | **High** | | **Low** | | **Moderate** | | **Sedentary** | |
| **Valid** |  | 7 |  | 9 |  | 8 |  | 7 |  |
| **Missing** |  | 0 |  | 0 |  | 0 |  | 0 |  |
| **Mean** |  | 7.505 |  | 5.879 |  | 7.123 |  | 4.151 |  |
| **Std. Deviation** |  | 1.245 |  | 1.691 |  | 1.573 |  | 0.7714 |  |
| **Shapiro-Wilk** |  | 0.9436 |  | 0.8408 |  | 0.9759 |  | 0.9284 |  |
| **P-value of Shapiro-Wilk** |  | 0.671 |  | 0.059 |  | 0.940 |  | 0.538 |  |
| **Minimum** |  | 5.538 |  | 4.105 |  | 4.739 |  | 3.180 |  |
| **Maximum** |  | 9.210 |  | 8.268 |  | 9.533 |  | 5.120 |  |
|  | | | | | | | | | |

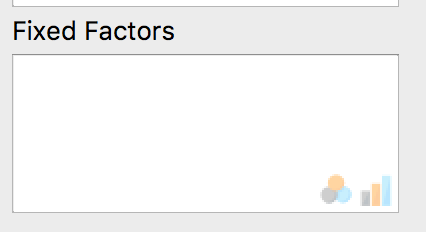
The Shapiro-Wilk test ran for each group separately, so we can tell if the assumptions were met for each sample. We see that our data is normally distributed because *p* > .05 for the dependent variable. You might report that like:

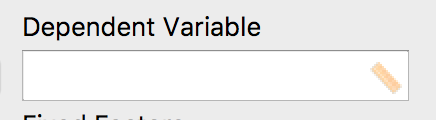
The assumption of normality was met, assessed by a Shapiro-Wilk test, *p*s =.671, .059, .940, and .538 for each group.

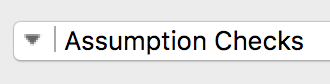
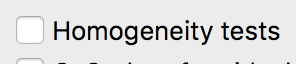
**Do you have homogeneity of variances?**

To get this analysis, we will run the ANOVA to get Levene’s test. Click ANOVA  🡪 ANOVA.



First, move the independent variable (the one with the groups) into the Fixed Factors box .

Next, move the dependent variable into the Dependent Variable box. 

Click on Assumption Checks:  🡪 Homogeneity Tests .

### Assumption Checks

| **Test for Equality of Variances (Levene's)** | | | | | | | |
| --- | --- | --- | --- | --- | --- | --- | --- |
| **F** | | **df1** | | **df2** | | **p** | |
| 2.129 |  | 3 |  | 27 |  | 0.120 |  |
|  | | | | | | | |

At this stage of the analysis you will have analyzed your data for outliers and normality, but not yet interpreted whether the third major assumption – homogeneity of variances – is violated or not. If the variances are unequal, this can affect the Type I error rate. In this example, the (population) variance for CWWS scores, stress\_coping, for all levels of group should be equal. If this is not the case, corrections can be applied to the calculations of the one-way ANOVA so that any violation of homogeneity of variances can be compensated for and the test remains valid. The assumption of homogeneity of variances is tested for using Levene's Test of Equality of Variances.

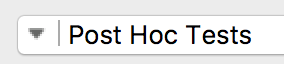
If Levene's test is statistically significant (i.e., p < .05), you do not have equal variances and have violated the assumption of homogeneity of variances (i.e., you have heterogeneous variances). On the other hand, if Levene's test is not statistically significant (i.e., p > .05), you have equal variances and you have not violated the assumption of homogeneity of variances.

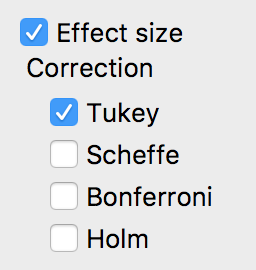
We could report that we have homogeneity of variances by saying:

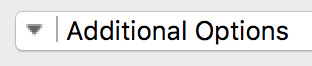
The assumption of homogeneity was assessed by using Levene’s test and was not significant, *p* = .120, therefore, this assumption was met.

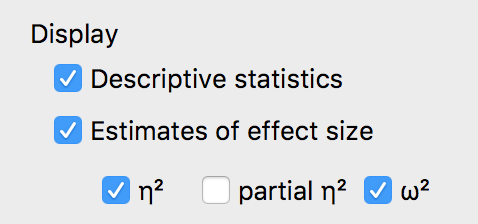
## **The ANOVA and effect size:**

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

Click on Post Hoc Tests:  🡪 Move group over to the right 🡪 Click on Effect size and a post hoc test. Tukey is automatically selected, but you can turn that one off and select Bonferroni (for example) if that is what your instructor requests:



Click on Additional Options:  🡪 Under Marginal means, move group to the right. Under display, click on descriptive statistics and estimates of effect size:



Depending on your instructor, you may want to select η2 (eta squared) or ω2 (omega squared).

## **Interpretation and Reporting:**

### ANOVA

| **ANOVA - coping\_stress** | | | | | | | | | | | | | | | |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **Cases** | | **Sum of Squares** | | **df** | | **Mean Square** | | **F** | | **p** | | **η²** | | **ω²** | |
| group |  | 49.03 |  | 3 |  | 16.344 |  | 8.316 |  | < .001 |  | 0.480 |  | 0.415 |  |
| Residual |  | 53.07 |  | 27 |  | 1.965 |  |  |  |  |  |  |  |  |  |
|  | | | | | | | | | | | | | | | |
| Note.  Type III Sum of Squares | | | | | | | | | | | | | | | |

If the ANOVA is statistically significant (i.e., p < .05), it can be concluded that not all group means are equal in the population (i.e., at least one group mean is different to another group mean). Alternatively, if p > .05, you do not have any statistically significant differences between the group means. Using our *p* < .001 from this example, it can be concluded that there is a statistically significant difference in coping\_stress scores for the different levels of group.

You can report that information using the following style:

* *F*(df1, df2) = *F* value, *p* = *p* value, effect size = effect size value.
* df1is found in the top row next to the name of your independent variable, which is group here. So df1 = 3.
* df2 is found in the bottom row next to Residual, so df2 = 27.
* F values are listed under F, and p values are listed under p.
* Eta or omega squared (not both usually you just pick one) are a representation of effect size for ANOVA. Because you cannot simply say that group 1 is different than group 2 for the overall ANOVA test, you have to estimate how much the difference is across all the groups. Therefore, eta and omega squared are used to determine the amount of variance accounted for (out of 100% variance) that the IV explains in the DV. These values are listed as proportions, so they never go over the value of one. Therefore, we could say that 48% or 42% of the variance in coping stress group differences is accounted for by the group variable. Eta squared is the more commonly reported effect size, while omega squared is often used to help represent the true population effect size.

With the meaning of each part as follows:

|  |  |
| --- | --- |
| **Part** | **Meaning** |
| *F* | Indicates that you are comparing to an *F*-distribution (*F*-test). |
| 3 in (3,27) | Indicates the Between Groups degrees of freedom ("df1") |
| 27 in (3,27) | Indicates the Within Groups [Error] degrees of freedom ("df2") |
| 8.316 | 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. |
| Effect size | The amount of variance in the DV accounted for by the groups or IV. |

Therefore, we would report this as: F(3,27) = 8.32, p < .001, η² = .48 OR F(3,27) = 8.32, p < .001, ω² = .42.

### Assumption Checks

| **Test for Equality of Variances (Levene's)** | | | | | | | |
| --- | --- | --- | --- | --- | --- | --- | --- |
| **F** | | **df1** | | **df2** | | **p** | |
| 2.129 |  | 3 |  | 27 |  | 0.120 |  |
|  | | | | | | | |

Levene’s test is described above in the assumptions section.

### Post Hoc Tests

| **Post Hoc Comparisons - group** | | | | | | | | | | | | | |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | |  | | **Mean Difference** | | **SE** | | **t** | | **Cohen's d** | | **p tukey** | |
| High |  | Low |  | 1.626 |  | 0.707 |  | 2.302 |  | 1.073 |  | 0.122 |  |
|  |  | Moderate |  | 0.383 |  | 0.726 |  | 0.527 |  | 0.267 |  | 0.952 |  |
|  |  | Sedentary |  | 3.354 |  | 0.749 |  | 4.476 |  | 3.239 |  | < .001 |  |
| Low |  | Moderate |  | -1.244 |  | 0.681 |  | -1.826 |  | -0.760 |  | 0.283 |  |
|  |  | Sedentary |  | 1.728 |  | 0.707 |  | 2.445 |  | 1.257 |  | 0.093 |  |
| Moderate |  | Sedentary |  | 2.972 |  | 0.726 |  | 4.095 |  | 2.344 |  | 0.002 |  |
|  | | | | | | | | | | | | | |
| *Note.*  Cohen's d does not correct for multiple comparisons. | | | | | | | | | | | | | |

The Tukey correction on an independent *t*-test is a good test if you wish to compare all possible combinations of group differences when the assumption of homogeneity of variances is not violated. As well as showing whether any differences between groups are statistically significant, this post-hoc test also provides confidence intervals for the differences between the group means. This box only shows you the unique possible combinations of each group two at a time. Remember that you could switch the order of them (i.e. High to Low is the same test as Low to High, so it is not repeated).

The information in each column in the above table has the following meaning:

|  |  |
| --- | --- |
| **Column Name** | **Column Meaning** |
| Mean Difference (I - J) | Mean difference between group I and group J (I minus J) |
| Std. Error | Standard error of the difference between group I and J |
| t | The independent *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 an independent *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 Tukey 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 had six two-way comparisons. If you were to run an independent *t*-test for each of these comparisons, these values would be different because those would not be corrected. |

Here, you would interpret each row one at a time to determine where the group differences are found. It appears that High versus Sedentary and Moderate versus Sedentary are significantly different, but High versus Low and Low versus Sedentary also have large effect sizes. This table highlights the need for both statistical (*p* value based) and practical (effect size based) interpretations of importance. You can now work through each comparison in turn and report the result. Alternatively, you can report the results in a table or graph.

### Marginal Means

| **Marginal Means - group** | | | | | | | | | |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **group** | | **Marginal Mean** | | **SE** | | **Lower CI** | | **Upper CI** | |
| High |  | 7.505 |  | 0.530 |  | 6.418 |  | 8.593 |  |
| Low |  | 5.879 |  | 0.467 |  | 4.920 |  | 6.838 |  |
| Moderate |  | 7.123 |  | 0.496 |  | 6.106 |  | 8.140 |  |
| Sedentary |  | 4.151 |  | 0.530 |  | 3.064 |  | 5.239 |  |
|  | | | | | | | | | |

### Descriptives

| **Descriptives - coping\_stress** | | | | | | | |
| --- | --- | --- | --- | --- | --- | --- | --- |
| **group** | | **Mean** | | **SD** | | **N** | |
| High |  | 7.505 |  | 1.245 |  | 7 |  |
| Low |  | 5.879 |  | 1.691 |  | 9 |  |
| Moderate |  | 7.123 |  | 1.573 |  | 8 |  |
| Sedentary |  | 4.151 |  | 0.771 |  | 7 |  |
|  | | | | | | | |

To really understand those post hoc test differences, we need the means. Using both marginal means and descriptives, we now can see the M, SD, SE, and N values for each group, as well as the confidence interval for each group separately. These numbers can be used to make a graph of the means (see Excel graphs how-to guide) and to help interpret the post hoc test. For example, the High group has more coping stress than the Sedentary group because their mean (7.51) is larger than the Sedentary mean (4.15).

## **Reporting All Together:**

A one-way ANOVA was conducted to determine if the ability to cope with workplace-related stress (CWWS score) was different for groups with different physical activity levels. Participants were classified into four groups: sedentary (*n* = 7), low (*n* = 9), moderate (*n* = 8) and high levels of physical activity (*n* = 7). There were no outliers, as assessed by boxplot. The assumption of normality was met, assessed by a Shapiro-Wilk test, *p*s =.671, .059, .940, and .538 for each group.; and there was homogeneity of variances, as assessed by Levene's test of homogeneity of variances (*p* = .120). Data is presented as mean ± standard deviation. CWWS scores were significantly different between different physical activity groups, *F*(3, 27) = 8.32, *p* < .001, ω2 = .42. CWWS scores increased from the sedentary (*M* = 4.15, *SD* = 0.77) to the low (*M* = 5.88, *SD* = 1.69), moderate (*M* = 7.12, *SD* = 1.57) and high (*M* = 7.51, *SD* = 1.24) physical activity groups, in that order. An independent *t*-test with Tukey post hoc correction revealed that the mean increase from sedentary to moderate was statistically significant (*p* = .002, *d* = 2.34), as well as the increase from sedentary to high (*p* < .001, *d* = 3.24), but no other group differences were significant.