One-way ANOVA (between)

Relevant research questions and data requirements

Research question: An ANOVA (**An**alsyis **O**f **Va**riance) is a mean-difference test. In a one-way between-subjects ANOVA we *simultaneously compare* the means of *two or more groups* that are the result of manipulating *one independent variable* (hence "one-way" ANOVA). As such, it is an extension of the independent-samples t-test. While the independent-samples t-test can "only" compare 2 groups, ANOVA can compare two or more groups/conditions at the same time. It calculates the ratio of variability between groups and within groups. If the variability between groups exceeds the variability within groups, this may be evidence of a treatment effect.
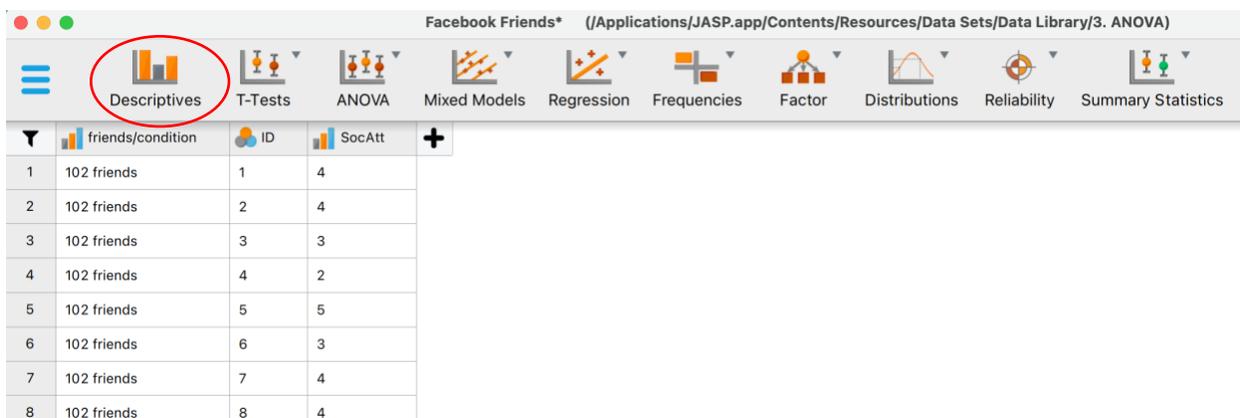
As an example, let's say we want to know if the number of friends or followers associated with a social media profile affects the social attractiveness rating of the profile. Participants in the study are shown the exact same social media profile, but the number of friends or followers listed is manipulated to say either 102, 302, 502, 702, or 902. This gives us 5 experimental conditions and participants are randomly assigned to *one* of those groups (i.e., this is a between-subjects design). After viewing the profile, participants rate the social attractiveness of the profile on a scale from 1 (not at all) to 7 (extremely).

How many groups? Two or more. Remember: ANOVA can do anything a t-test can do, but a t-test is limited to comparing 2 groups!

Data requirements? Interval/ratio outcome data (here: social attractiveness ratings), ideally the data are (roughly) normally distributed and independent (scores are not related). It is also assumed that the variance in the different groups/conditions is *roughly* the same. This is called homogeneity/equality of variance. Also check for outliers.
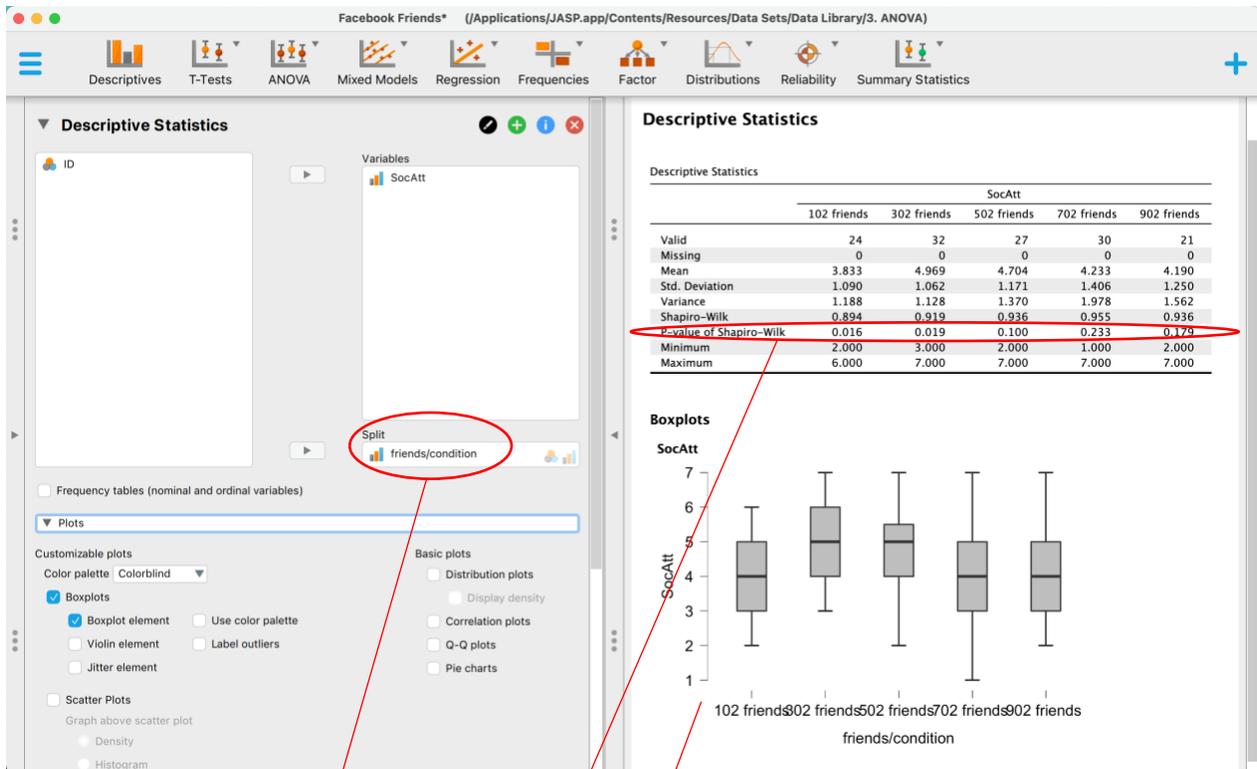
Checking our assumptions

First, let's check our assumptions and have a look at descriptives and the distribution of scores in each condition. (For more detailed information, refer to the Descriptives worksheet.)

I have asked for output by condition here (I am "splitting" the output by condition). You can ask for several plots and statistics (refer to the "Descriptives" resource sheet for more detailed information), but here I have simply asked for boxplots and M, SD, variance, and minimum and maximum.

The top table shows us the sample size of each condition, how many missing cases we have, M, SD, Variance, and the Shapiro-Wilk test for normality. Just by eye-balling the means, we can see that condition 1 ("102 friends") has the lowest social attractiveness ratings, and condition 2 ("302 friends") the highest, with the other conditions somewhere inbetween. To see if these 5 means differ statistically, we will need to run our ANOVA.

Normal distribution of data. For now, let's check on the assumption of normality. Recall that the null hypothesis of the Shapiro-Wilk test is that "normality is met" – hence, any p-value below .05 should alert us to deviations from normality. This is the case for conditions 1 (p=.016) and 2 (p=.019). Because 3 out of 5 conditions meet our assumption and ANOVA is generally known to be quite "robust" to violations of normality, it is ok to proceed with our analysis. (Alternatively, you could run the nonparametric equivalent, the Kruskal-Wallis test, which is an option at the very bottom of the ANOVA screen in JASP, as we will see below.)

Outliers. Looking at our boxplots, we see no outliers, but get a more visual representation of the 5 different groups' distribution of scores.

Homogeneity of Variance. This cannot be checked under "Descriptives." Proceed to the ANOVA (below) to check for this assumption.

Running the test in JASP

Next, to proceed with our analysis, click on "ANOVA" in the main menu and select "ANOVA" under "Classical" (we will ignore all Bayesian analyses for now).
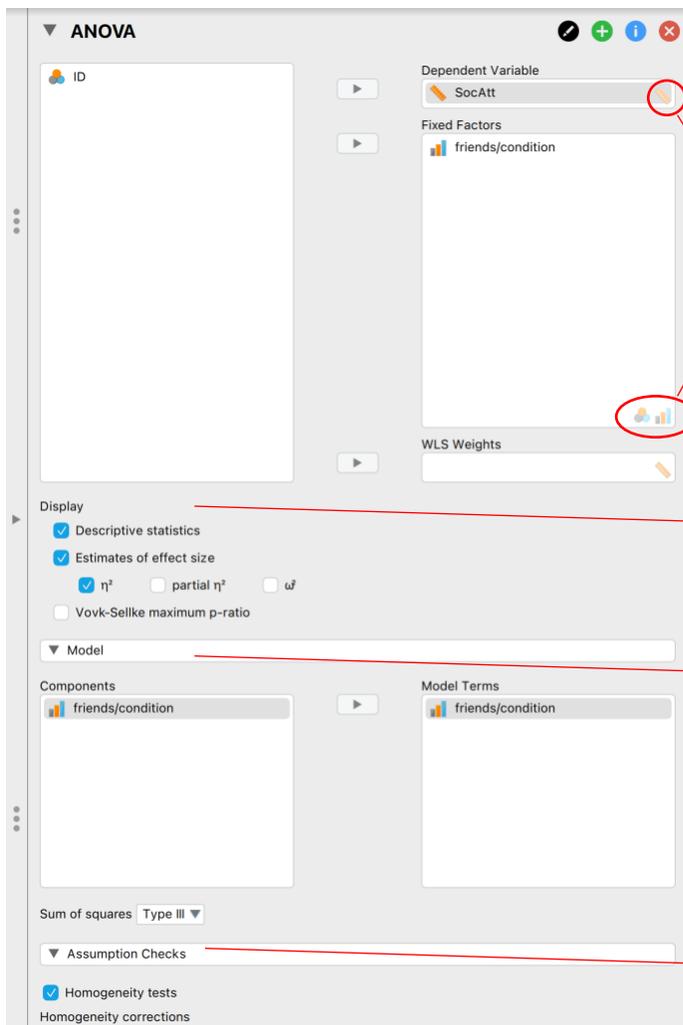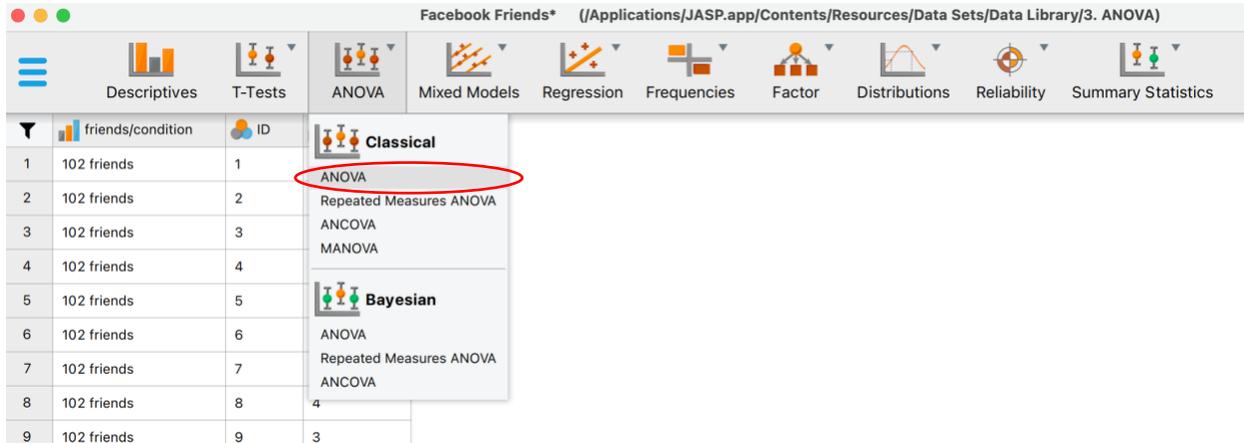
Move the outcome variable of interest into the "dependent variable" field (here "Social Attractiveness") and the independent variable into "fixed factors."

(As always, JASP reminds us what type of variable we need to enter in each field by showing us the little symbols in the bottom right)

Under **Display**, select "Descriptive Statistics" and "Estimates of effect size" and "eta squared"

Under **Model**, select your IV and move it over to "Model Terms"

Under **Assumption Checks**, check "Homogeneity of Variance" – this tests the null hypothesis that the variance in all groups is roughly equal. This is the last assumption we need to check.
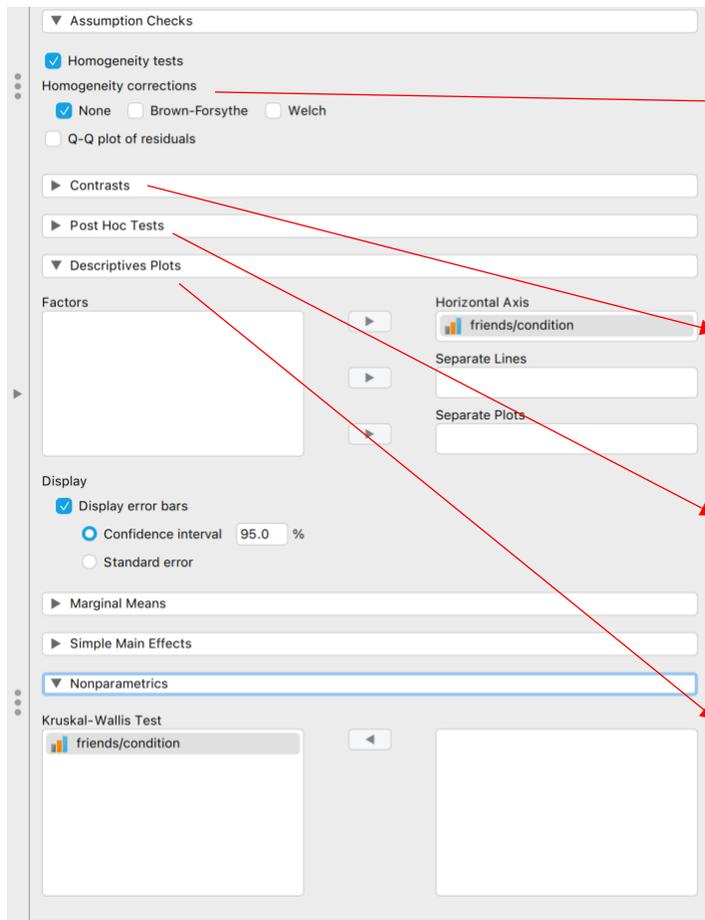
Assumption Checks, continued: IF your homogeneity assumption is violated, you can go under "Homogeneity Corrections" and select "Brown-Forsythe" or "Welch", both of which correct the df to make up for the violation and to keep the type I error low. The Welch adjustment is commonly used.

Contrasts: These are specific ways to compare the different groups. We don't teach this at the undergraduate level, so you can skip this.

Post-Hoc Tests: IF your overall ANOVA is significant, you will want to run post-hoc tests to see *which groups* differ (the overall test only asks *if* any groups differ). We will return to this.

Descriptive Plots: Select this if you would like a visual representation of your means and the 95% Confidence Intervals

Marginal Means: This becomes relevant in a factorial ANOVA (i.e., when you have two or more IVs). Because we are running a one-way ANOVA here, we can skip this. Our marginal means are identical to the means of the five groups that we get by selecting "Descriptive Statistics" above.

Simple Main Effects: Simple main effects help us make sense of interactions in factorial ANOVAs (i.e., ANOVAs with 2 or more IVs). This, too is irrelevant for a one-way ANOVA.

Nonparametrics: This is where you can select the "Kruskal-Wallis Test" if you have an ordinal DV, or if your data vastly violate the assumption of normality (again, though, ANOVA, especially with larger samples, is believed to be robust against this violation). Still, you may run this nonparametric test and report it alongside the ANOVA to let your reader know if the parametric and nonparametric analyses are different or identical in their conclusion.

Assumption Checks ▼

Test for Equality of Variances (Levene's) ▼

F	df1	df2	p
0.921	4.000	129.000	0.454

Even though Assumption checks are reported at the bottom of the output, it is important to look at this first. Here we see that $p = .454$, so our null hypothesis that variances are equal across all groups cannot be rejected, so our assumption is met. IF $p < .05$, select the Brown-Forsythe or Welch test under Homogeneity corrections as indicated on the previous page to correct for the violation of the assumption.

ANOVA ▼

ANOVA – SocAtt

Cases	Sum of Squares	df	Mean Square	F	p	η²
friends/condition	22.217	4	5.554	3.841	0.006	0.106
Residuals	186.536	129	1.446			

Note. Type III Sum of Squares

This is our hypothesis test. Shown in order are the SS, df, MS (aka variance, which is SS/df), the value of the F-ratio, the p-value, and our effect size eta squared.
Note: For SS, df, and MS, the top row always shows the values between groups, the bottom row shows values within groups!

We see that the value of our F-ratio (recall: $F = MS_{between}/ MS_{within}$) is $F(4,129) = 3.84$, and the p-value is .006, which is well below .05. Thus, we reject the null hypothesis that $\mu1 = \mu2 = \mu3 = \mu4 = \mu5$. The variability between groups much exceeds the variability within groups, which suggests a treatment effect. In other words: The means of our five groups significantly differ – however, we only know *that* there is a difference among the five means, not yet *where* it is, i.e. which groups, exactly, differ from one another. For that, we will need to run post-hoc tests.

Eta squared is 0.106 and tells us that 10.6% of the variability in social attractiveness ratings can be explained by the number of friends/followers. *Note:* This is considered a medium-to-large effect. General guidelines say that for eta squared, .01 is a small effect, .06 a medium effect, and .14 or above a large effect.

ANOVA

ANOVA – SocAtt

Homogeneity Correction	Cases	Sum of Squares	df	Mean Square	F	p	η²
None	friends/condition	22.217	4.000	5.554	3.841	0.006	0.106
	Residuals	186.536	129.000	1.446			
Welch	friends/condition	22.217	4.000	5.554	4.407	0.003	0.106
	Residuals	186.536	61.825	3.017			

Note. Type III Sum of Squares

FYI: This is what the output would look like if we ran the Welch correction should the assumption of homogenous variances be violated. You would report the third and fourth row (notice the denominator df changed).

Descriptives ▼

Descriptives – SocAtt

| friends/condition	Mean	SD	N
102	3.833	1.090	24
302	4.969	1.062	32
502	4.704	1.171	27
702	4.233	1.406	30
902	4.190	1.250	21

Descriptives plots ▼
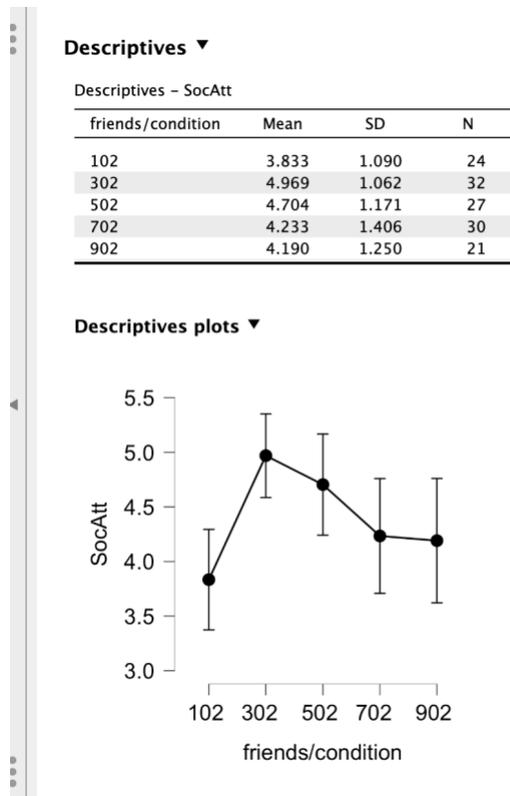
Here are our **descriptive statistics** and descriptive plots. Condition labels will be reflected here, so always label your conditions in a way that makes the output easy to read (here: "102", "302", etc. rather than "1", "2", etc.).

Our plot shows the means and their 95% confidence intervals. Right away, we can see that social attractiveness ratings are lowest in the "102 friends" group and confidence intervals do not overlap between that and the "302 friends groups", so those two groups likely differ from each other because the ranges of true population means don't overlap. We will test this with post-hoc tests below.

Also notice that the y-axis is scaled from 3-5.5, not 1-7, so differences between the groups appear slightly more pronounced in this image than they would if we scaled the y-axis from 1-7 (which JASP does not let us do).

Running and understanding post-hoc tests

Because our overall ANOVA showed that there was a significant difference somewhere between the 5 means, we need to run **post-hoc tests** to see *what groups* differ from each other, exactly. To run those, we go back to our analysis options on the left hand side of the screen and scroll to post-hoc tests
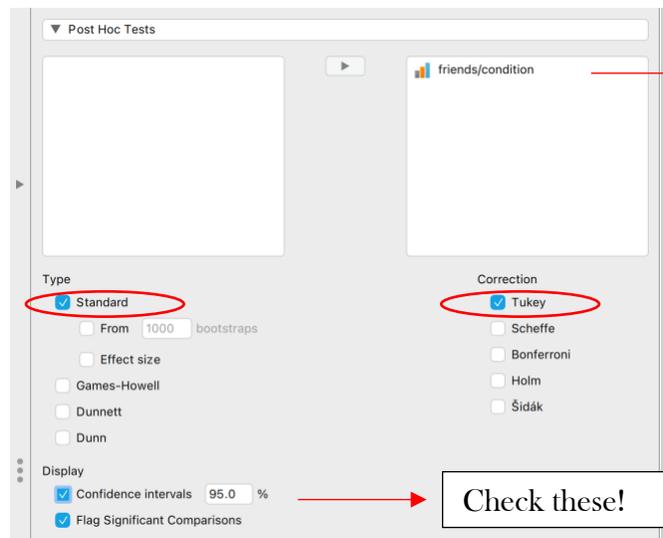
First, move your **IV** into the window on the right.

Under post-hoc tests, you have several options. For simplicity, here are some general guidelines:

Type: Standard runs pairwise t-tests and adjusts the type 1 error rate by one of the corrections on the right. **Tukey** is widely believed to be the least conservative of the ones listed. It reduces Type 1 error rate without reducing power, i.e. the probability of finding an effect if it exists.

If you have unequal variances (see above), choose **Games-Howell** as your post-hoc test.
If you ran the nonparametric Kruskal-Wallis test, choose **Dunn** as your post-hoc test.

Post Hoc Tests ▾

Standard ▾

Post Hoc Comparisons – friends/condition

			95% CI for Mean Difference				
		Mean Difference	Lower	Upper	SE	t	p_{tukey}
102	302,	−1.135	−2.034	−0.237	0.325	−3.497	0.006**
	502	−0.870	−1.804	0.063	0.337	−2.580	0.080
	702,	−0.400	−1.311	0.511	0.329	−1.215	0.743
	902	−0.357	−1.351	0.637	0.359	−0.994	0.858
302,	502	0.265	−0.604	1.134	0.314	0.843	0.916
	702,	0.735	−0.110	1.581	0.306	2.407	0.120
	902	0.778	−0.156	1.713	0.338	2.305	0.150
502	702,	0.470	−0.412	1.353	0.319	1.475	0.581
	902	0.513	−0.455	1.481	0.350	1.467	0.586
702,	902	0.043	−0.904	0.989	0.342	0.125	1.000

Note. P–value and confidence intervals adjusted for comparing a family of 5 estimates (confidence intervals corrected using the tukey method).
* p < .05, ** p < .01

Our post-hoc test output shows us the comparison of each pair of groups. The first row compares group 102 to group 302, the second row compares group 102 to group 502, etc., and the last row compares group 702 to group 902.

Because we asked significant comparisons to be flagged, we see quickly that only groups "102" and "302" differ from each other. This supports our hunch from looking at the means plot and its 95% CIs earlier. To remind ourselves which group had higher social attractiveness ratings, we can refer back to the Descriptives.

Writing up results in APA style

A one-way ANOVA indicated that social attractiveness ratings differed by number of friends/followers, $F(4,129) = 3.84$, $p = .006$, $\eta^2 = 0.11$. Post-hoc comparisons using Tukey's correction shows that profiles with 102 followers were rated as significantly less socially attractive than profiles with 302 friends or followers, $p = .006$. No other differences were significant. For all descriptive statistics, see Table 1.

Table 1

Descriptive Statistics for the five experimental conditions

	N	Mean	SD
102 friends	24	3.83	1.09
302 friends	32	4.97	1.06
502 friends	27	4.70	1.17
702 friends	30	4.23	1.41
902 friends	21	4.19	1.25

*Note that there are spaces before and after equal signs; M, SD, F, p are italicized; everything is rounded to two digits *except* for p-values, which should be reported exactly as given in the output. Only report leading zeros for values that can exceed 1 (hence, p-values should *not* be reported with leading 0s). For F-ratios, you *always* list the df$_{between}$ first, then the df$_{within}$.

If there are **three groups** (three conditions), you can probably describe all comparisons (and each group's M and SD) in the text.

If there are **more than three groups** (3+ conditions), you might want to move to minimal in-text description and then display your findings in a table (again, to give the reader each group's M and SD).

APA-style tables *never* have vertical lines, only horizontal lines that frame the table as shown above.