# Analysis of Variance Tutorials

These tutorials shall prepare you to be successful in performing appropriate analysis in an empirical controlled experiment.

All data files can be found in Course GDrive > Downloads > ps4hci-Wobbrock.

## Task 1: T-Test

**File used: posts.sav**

You can used R (free), or JASP (free) to complete this.

The file contains a hypothetical study of 40 college students’ Facebook posting

behavior using one of two platforms: Apple’s iOS or Google’s Android OS. The data show the number of Facebook posts subjects made during a particular week using their mobile platform.

Here, this is a one-way design with two levels and thus T-test is appropriate

**JASP**: Go to T-Tests > Independent Samples T-Test; Put Posts as Variables, and Platform as Grouping Variable. Select Tests as Welch; Check Assumption Checks > Normality as well as Equality of Variance; Check Effect Size > Cohen's d; Check Descriptive (see screenshots)

**R**: After you load the datasets, do

t.test(posts$Posts~posts$Platform, var.equal = TRUE)

For normality check:

To do a Shapiro-Wilk test, do

shapiro.test(posts$Posts)

To do a Kolmogorov-Smirnov test, do

ks.test(posts$Posts, "pnorm", mean=mean(posts$Posts), sd=sd(posts$Posts))

For homogenity of variances, do

library(car) //you may need to first install package

leveneTest(posts$Posts, posts$Platform, center=mean)

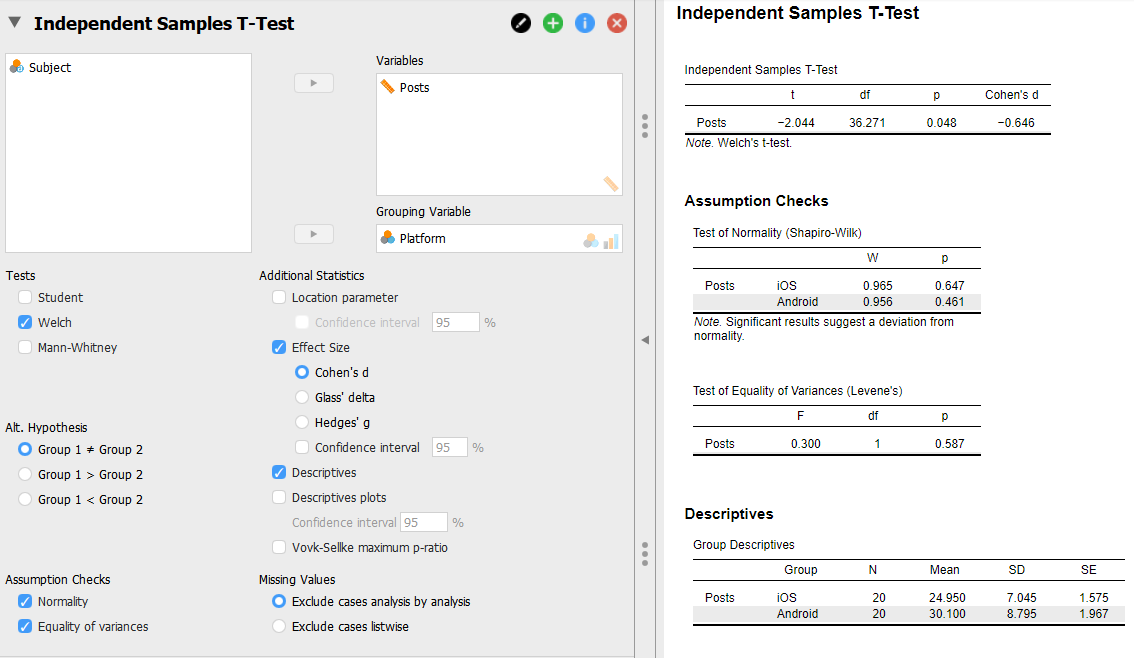
To show descriptives, do

summary(posts$Posts)

mean(posts$Posts)

sd(posts$Posts)

sd(posts$Posts) / sqrt(length(posts$Posts))



Answer the followings:

1. Is this a between-subjects or within-subjects experiment? Why?
2. What is the independent variable named?
3. How many levels (i.e., values) does the independent variable have? What are they?
4. How many subjects were in this experiment?
5. How many subjects were exposed to each level of the independent variable? Is the design balanced (i.e., are the numbers equal)?
6. What are the mean and standard deviation number of posts for each level of the independent variable?
7. Assuming equal variances, what is the t statistic for this t-test? (Hint: this is also called the t Ratio.)
8. How many degrees of freedom are there (dfs)?
9. What is the two-tailed p-value resulting from this t-test? Is it significant at the α = .05 level?
10. The formulation for expressing a significant t-test result is: t(dfs) = t-statistic, p < .05. For a nonsignificant result, it is: t(dfs) = t-statistic, n.s. Write your result in the proper formulation. Read https://shengdongzhao.com/newSite/how-to-report-statistics-in-apa-format/ for more detail how to report.
11. What can we conclude about the results of this study? Which platform seems to result in more Facebook posts, if any? Write a sentence summarizing the finding from this analysis

## Task 2: One way ANOVA

**File used: posts.sav**

You can used JASP (free) or R (free) to complete this.

F-test, which can do everything a t-test can do, and more. An F-test, which is the most common analysis of variance, can handle multiple independent variables, or factors, and these factors can have more than two levels. By comparison, a t-test can only have one factor with two levels, which is not very useful for many experiment designs.

**JASP**: Select ANOVA. Posts as DV, and Platform as IV. Assumption Checks > check Homogeneity tests and Q-Q plot of residuals. Descriptive Plots > check Display error bars. Additional Options > check Descriptive Statistics and Estimates of effect size (check eta squared and partial eta squared). For normality check, Select Descriptives on the top bar, Put Posts on Variable, Under Statistics > check Shapiro Wilk test

**R**: Do:

For ANOVA:

aov <- aov(posts$Posts ~ posts$Platform)

summary(aov)

For effect size:

install.packages("lsr")

library(lsr)

etaSquared(aov)

For Q-Q plot:

qqnorm(posts$Posts)

qqline(posts$Posts)

For other R scripts, check out previous homework.

Answer the followings:

1. What is the output of the Q-Q plot test? Is the data normal? How about homogeneity of variance test?
2. Do the number of observations (N’s) and means agree with those produced by the t-test? What are they? (If they do not agree, there is an error somewhere!)
3. In the ANOVA table, what is the F-statistic? What is the p-value? Is it significant at the α = .05 level?
4. How does this p-value compare to that produced by the t-test? Did you expect it to be otherwise?
5. What is the effect size in terms of eta squared and partial eta squared? What is the interpretation?
6. The general formulation for expressing an F-test result is: F(dfnum,dfden) = F-ratio, p < .05 (or n.s.). Report the F-test result, filling in dfnum and dfden with the numerator and denominator (also called “error”) degrees-of-freedom, respectively. Also include the effect size. Read this for more detail - https://shengdongzhao.com/newSite/how-to-report-statistics-in-apa-format/

# Task 3: One-way ANOVA with 3 or more levels

**File used: postctrl.sav**

As noted, the F-test can handle more than one factor, and also, more than two levels per factor. A one-way ANOVA refers to a single factor design. Similarly, a two-way ANOVA refers to a two-factor design, i.e., two independent variables. In this part, we will still conduct a one-way ANOVA, but this time, our factor will have three levels. Thus, it cannot be analyzed with a t-test, which can only handle two levels of a single factor.

Open postsctrl.sav. This data set is the same for the iOS and Android levels, but now has added 20 new college students as a control group who did not use a mobile device for posting on Facebook but were told to use their desktop computer instead. Thus, the Platform factor now has three levels: iOS, Android, and desktop.

**JASP**: Perform everything same as previous homework on one-way ANOVA. In addition: under ANOVA > Posthoc tests, check Tukey and Bonferroni

**R**: Do:

Set the levels (unfortunately, R does not recognize the levels of Platforms):

postsctrl$Platform <- ordered(postsctrl$Platform, levels = c("1", "2", "3"))

To check the levels are set correctly:

levels(postsctrl$Platform)

Perform anova:

aov <- aov(Posts ~ Platform, data = postsctrl)

summary(aov)

Perform Posthoc:

TukeyHSD(aov) //Tukey test

pairwise.t.test(postsctrl$Posts, postsctrl$Platform, p.adj="bonferroni") //pairwise t test with Bonferroni correction

Answer the followings:

1. What is the output of the normality test? Is the data normal?
2. Was this a one-way, two-way, or three-way analysis of variance? What is/are the factor(s)? What are each factor’s levels?
3. How many data points are there for each level of Platform? What are the means and standard deviations for each level?
4. In this case, the overall, or omnibus, F-test is testing for whether any differences exist among the levels of the independent variable. Is the F-test significant at the α = .05 level? What is the F-ratio? What is the p-value? Construct a proper reporting formulation for this result.
5. The omnibus F-test does not tell us whether all three levels of Platform are different from one another, or whether just two levels (and which two?) are different. For this, we need post hoc comparisons, which are justified only when the omnibus F-test is significant. Examine the Post Hoc Tests output. What is the interpretation? Does Tukey and Bonferroni converge to the same interpretation? (FYI: A Bonferroni correction divides α by the number of post hoc comparisons. In this case, with three post hoc comparisons, we would use α = .05 / 3 = .0166. (Equivalently, we can multiply our p-values by 3 and then continue using α = .05, which is what SPSS’s Bonferroni output does.)
6. What is the effect size in terms of eta squared and partial eta squared? What is the interpretation?
7. Write a sentence summarizing the findings from this analysis. Read this for how to report - https://shengdongzhao.com/newSite/how-to-report-statistics-in-apa-format/

# Task 4: Non-parametric tests

**File used: posts.sav**

Mann-Whitney U test (independent) and Wilcoxon rank sums test (paired) are non-parametric t-test . The use of a t-test or an F-test requires the upholding of some assumptions. One of these assumptions is that the measures form a roughly normal distribution, or would if enough measures were taken from the population. When these assumptions are violated, a nonparametric test should be used. Such a test does not presume an underlying normal distribution. They are therefore sometimes called “distribution free.” Such tests are useful when, for example, analyzing Likert scale data, which is often non-normal.

The equivalent of a between-subjects (independent samples) t-test using nonparametric statistics is the Mann-Whitney U test. The Mann-Whitney U test is a test for an experiment containing one between-subjects factor with two-levels.

Let's perform a normality test first. Follow same step as previous homework.

**JASP**: Perform a U test on posts.sav by going to T-Tests > Independent Samples T-Test. Under Tests, select Mann-Whitney.

*\*\*\*\*UPDATE\*\*\*\*\* It seems JASP has some bug reporting Mann-Whitney which reports Wilcoxon W value instead of the U value. Please use this website instead --> https://www.socscistatistics.com/tests/mannwhitney/default2.aspx*

**R**: do wilcox.test(posts$Posts ~ posts$Platform)

//Read this for more detail -> https://www.statmethods.net/stats/nonparametric.html

1. Why does this assignment not doing a normality test nor a homogeneity of variance test?
2. What is the Mann-Whitney U value produced?
3. What is the two-tailed p-value? Is it significant at the α = .05 level? How does the result compare to the t-test and F-test results for the same data from parts 1 and 2, above?
4. The formulation for a Mann-Whitney U test result is (U = U-value, p value). Report the test, filling in the appropriate values. Note that there are no degrees of freedom in these nonparametric tests. Read this for how to report - https://statistics.laerd.com/spss-tutorials/mann-whitney-u-test-using-spss-statistics-2.php
5. What can we conclude about the results of this study? Which platform seems to result in more Facebook posts, if any?Write a sentence summarizing the finding from this nonparametric analysis.

# Task 5: Repeated Measures ANOVA

**File used: postwthn.sav**

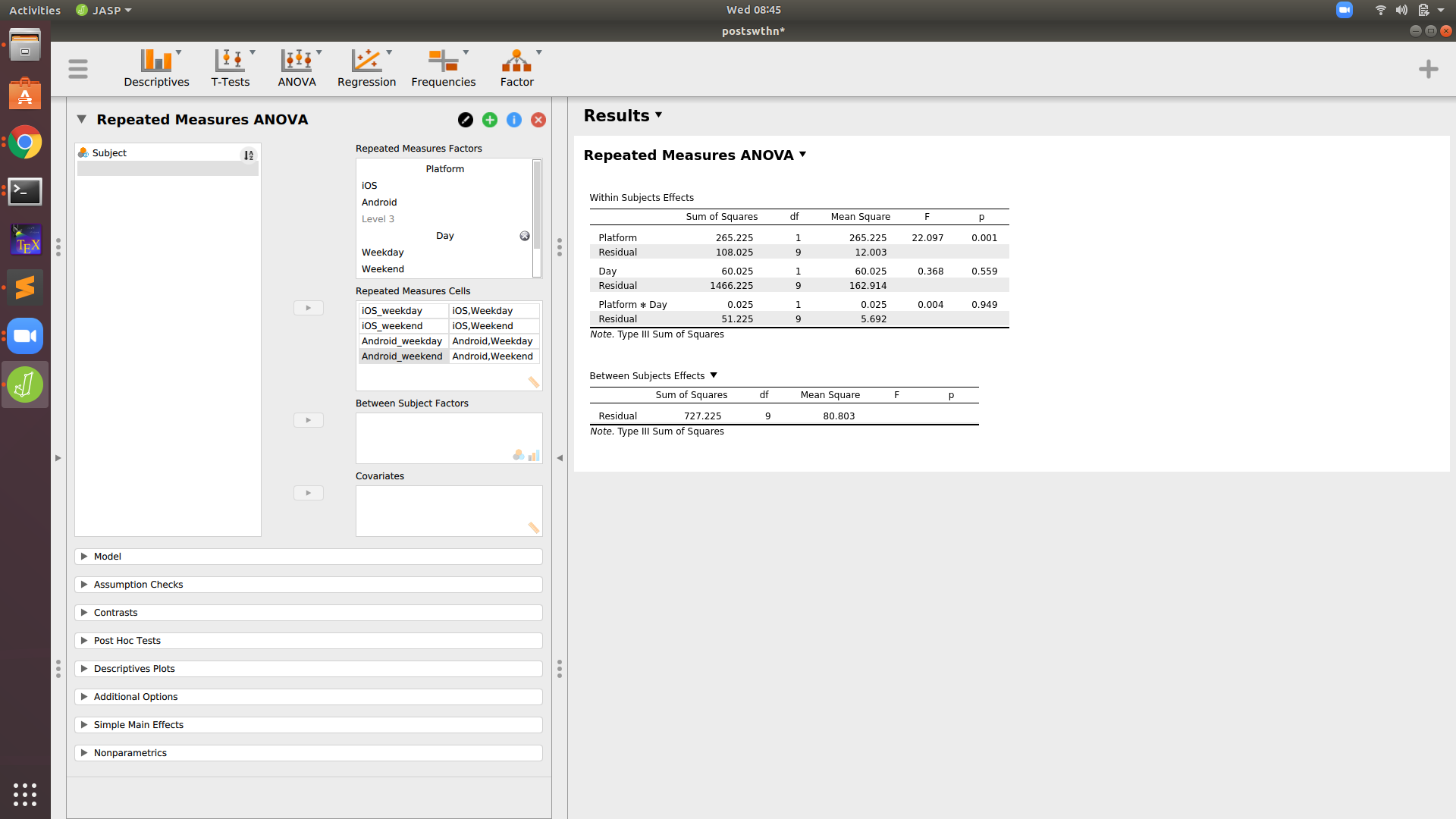
Thus far, we have only considered experiments where one subject was measured once on only one level of each factor. But often we wish to measure a subject more than once, perhaps for different levels of our factor(s), or over time, in which case time itself becomes a factor. Such designs are called “repeated measures” designs, and the factors on which we obtain repeated measures are called within-subjects factors (as opposed to between-subjects factors). For repeated measures studies, we can still use an ANOVA, but now we use a “repeated measures ANOVA,” and our data table inevitably looks different: for a wide-format table, there are now multiple measures per row (each row still corresponds to just one subject, as it has thus far).

Our current hypothetical study on Facebook posts has been modified to be a purely within-subjects study. Imagine that each college student was issued either an iOS or Android device for one week, and then the other device for the next week. Also, each college student’s posts were counted separately on weekdays and weekends. Instead of needing 40 college students as before, we now only need 10 students for the same data, which is shown in postswthn.sav. Open those files and see the wide-format data tables.

Additional Info: In general, we should expect within-subjects studies to be more statistically powerful than between-subjects studies because subjects are, in effect, compared to themselves, and any given subject is more like himself than he is like any other subject. Within-subjects designs therefore reduce variance compared to between-subjects designs, and thus result in more power for detecting differences. The downside of using within-subjects designs is that they are subject to carryover effects and require careful counterbalancing. This makes them impractical in many circumstances.

First, perform a normality test. Then:

**JASP**: Go to Repeated Measures ANOVA. Under Repeated Measures Factor, for RM Factor 1, rename to Device - for level 1, rename to iOS, for level 2, rename to Android. For RM Factor 2, double click and name it Day, for level 1, rename to Weekday, for level 2, rename to Weekend. Then under Repeated Measures Cells, transfer the variable accordingly, e.g., iOS\_weekday to iOS, Weekday (see screenshot). For homogeneity tests, since this is a repeated measures design, sphericity tests are used instead (since each of our factor has only two levels, assumption of sphericity is always met). For normality test, we need to go out of the Repeated measures, and go to Descriptives, set all variable as Variables (does not need Split variables), then check Shapiro-Wilk.



**R**:

First, let R understand your table structure:

Platform <- factor(c("iOS", "iOS", "Android", "Android"))

Day <- factor(c("weekday", "weekend", "weekday", "weekend"))

factor <- data.frame(Platform, Day)

factor

The first row will be iOS weekday, second row will be iOS weekend, third row will be Android weekday and so on. The first row must match with the first column of the table, for example, the first column is iOS\_weekday, which match with the factor first row.

Second, create a multivariate model with only the intercept as predictor

library(car)

model = lm(cbind(postswthn$iOS\_weekday, postswthn$iOS\_weekend, postswthn$Android\_weekday, postswthn$Android\_weekend)~1)

Third, run the anova (idata is the structure, while idesign is the DV)

aov <- Anova(model, idata=factor, idesign=~Platform\*Day, type=3)

summary(aov, multivariate = FALSE)

Fourth, run the Mauchly test to confirm the anova did not violate the sphericity assumptions. We have to do for Platform, Day, and Platform:Day separately:

mauchly.test(model, M = ~Day, X = ~1, idata=factor)

mauchly.test(model, M = ~Platform, X = ~1, idata=factor)

mauchly.test(model, M = ~Platform:Day, X = ~Platform+Day, idata=factor)

1. What is the output of the normality test? Is the data normal? How about the test of sphericity? Why all sphericity test has p-value of 1?
2. Was this a one-way, two-way, or three-way analysis of variance? What is/are the factor(s)? What are each factor’s levels? Express the design using N1 × N2 × … × Nn notation.
3. For each identified factor, was it between-subjects or within-subjects? How do you know?
4. Write the statistical result for the Device factor. (Hint: the table you care about is called Within-Subjects Effects. )
5. Write the statistical result for the Day factor.
6. Write the statistical result for the Platform\*Day interaction.
7. What is the effect size in terms of eta squared and partial eta squared? What is the interpretation?
8. Write the result in APA format.
9. What can we conclude about the results from this study?

# Task 6: Two-way between-subject ANOVA

**File used: postsbtwn.sav**

It is often the case that we wish to examine the effects of more than one factor, and we also care about the interaction among factors. Because multiple factors are involved, this is called a factorial design, expressed as N1 × N2 × … × Nn for an arbitrary number n of factors, and where each Ni is an integer indicating the number of levels of that factor. In practice, it is difficult to interpret experiments with more than three factors, especially if those factors each have more than two levels. For this part, we will examine an augmented version of our current study that adds another factor. Open postsbtwn.sav. You will see another column labeled Day with values “weekday” and “weekend.” These values correspond to the days of the week the subject was allowed to post to Facebook.

**JASP**: Very similar to previous homework on ANOVA. You have to simply add one more factor "Day" under Fixed Factors.

**R**: Do:

Anova(lm(Posts ~ Platform \* Day, data = postsbtwn, type="III"))

FYI: For the car library, it explicitly requires us to build a linear model in order to use type 3 model. Type 3 is used here as it does not depend on the order of the factors (Type 1 does!). By default, aov is of type 1 while lm can specify the type. In the previous homework, we can use aov since we only have one factor.

Answer the followings:

1. What is the output of the normality test? Is the data normal? How about the homogeneity tests?
2. Was this a one-way, two-way, or three-way analysis of variance? What is/are the factor(s)? What are each factor’s levels? Express the design using N1 × N2 × … × Nn notation.
3. For each identified factor, was it between-subjects or within-subjects? How do you know?
4. What were the means and standard deviations for the number of posts on weekdays? weekends?
5. Write the F-test result for the Platform factor. Is it significant at the α = .05 level? What is its p-value? Does this differ from the finding prior to the inclusion of the Day factor (part 2, above)? If so, how?
6. Write the F-test result for the Day factor. Is it significant at the α = .05 level? What is its p-value?
7. Write the F-test result for the Platform\*Day interaction. Is it significant at the α = .05 level? What is its p-value?
8. Within each factor, why don’t we need to perform any post hoc pairwise comparison tests?
9. What is the effect size? What is the interpretation?
10. Notice in JASP, there is a non-parametric section under ANOVA. Try put both factors under Kruskal-Wallis Test. What is the p-value? Does it differ from the parametric test? (For R people, try to find how to perform Kruskal-Wallis)
11. Interpret these results and craft three sentences describing the results of this experiment, one for each factor and one for the interaction. What can we say about the findings from this study? (Hint: p-values between .05 and .10 are often called “trends” or “marginal results,” and are often reported, although they cannot be considered strong evidence. Be wary of ever calling such results “marginally significant.” A result is either significant or it is not; there is no “marginal significance.”)