Statistics with jamovi

Dana Wanzer

Last Update: 2020-11-01

Contents

Welcome

This is the website for PSYC 290 and PSYC 790 at the University of Wisconsin-Stout, taught by Dana Wanzer. These resources are aimed at teaching you how to use jamovi and null hypothesis significance testing (NHST) to answer research questions.

Portions of this book may have been adapted from "Learning statistics with jamovi: A tutorial for psychology students and other beginners" by Danielle J. Navarro and David R. Foxcroft, version 0.70. Furthermore, the template and style of this book is from PsyTeachR.

Chapter 1

Introduction

This chapter will walk you through how this website/book works.

1.1 Quiz Questions

Throughout this website, there will be questions to help you test your knowledge. When you type in or select the correct answer, the dashed box will change color and become solid.

For example:

- What is 2+2?
- We attend the University of Wisconsin- Stout Madison Green Bay
- True or false: Statistics is awesome. TRUE FALSE

1.2 Errors and mistakes

I am human, therefore I err. If you find an error in the textbook or something you think might be a mistake, please let me know ASAP so I can update this for everyone else. Let me know which section you find the error or mistake in and what the error or mistake is. For example, if there was an error here you could say, "There was an error in 1.2 that the first sentence should really be 'To err is human.' "

Chapter 2

Independent t-test

2.1 What is the independent t-test?

The independent t-test is used to test the difference in our dependent variable between two different groups of observations. Our grouping variable is our independent variable. In other words, we use the independent t-test when we have a research question with a **continuous dependent variable** and a **categorical independent variable with two categories in which different participants are in each category**.

The independent t-test is also the independent samples t-test and the Student's t-test. I will use these terms interchangeably.

2.2 Data set-up

To conduct the independent t-test, we first need to ensure our data is set-up properly in our dataset. This requires having two columns: one with our continuous dependent variable and one indicating which group the participant is in. Each row is a unique participant or unit of analysis. Here's what example data may look like if we were testing for differences in a test score by students in my fall or spring semesters of this course:

Table 2.1: Example data for the independent t-test

ID	Semester	TestScore
1	Fall	86
2	Fall	80
3	Fall	75

ID	Semester	TestScore
4	Fall	79
5	Fall	82
6	Spring	84
7	Spring	90
8	Spring	72
9	Spring	75
10	Spring	81

In the example data above, what is your **independent variable**? ID Semester TestScore

In the example data above, what is your **dependent variable**? ID Semester TestScore

2.3 The math behind the independent t-test

The basic math of the independent t-test the mean difference divided by the pooled standard error.

$t = \frac{\bar{X}_1 - \bar{X}_2}{SE(\bar{X}_1 - \bar{X}_2)}$

The denominator of the equation is more difficult to calculate and depends on whether the sample size between groups is equal.

2.4 Assumptions

As a parametric test, the independent t-test has the same assumptions as other parametric tests:

1. The dependent variable is **normally distributed**
2. Variances in the two groups are roughly equal (i.e., **homogeneity of variances**)
3. The dependent variable is **interval or ratio** (i.e., continuous)
4. Scores are **independent** between groups

We cannot test the third and fourth assumptions; rather, those are based on knowing your data.

However, we can and should test for the first two assumptions. Fortunately, the independent samples t-test in jamovi has two check boxes under “Assumption Checks” that lets us test for both assumptions.

2.5 In jamovi

Let's run an example with data from lsj-data. Open data from your Data Library in "lsj-data". Select and open "Harpo". This dataset is hypothetical data of 33 students taking Dr. Harpo's statistics lectures. We have two tutors for the class, Anastasia ($n = 15$) and Bernadette ($n = 18$). Our research question is "Which tutor results in better student grades?" We don't have a hypothesis that one does better than the other.

1. To perform an independent t-test in jamovi, go to the Analyses tab, click the T-Tests button, and choose "Independent Samples T-Test".
2. Move your dependent variable `grade` to the Dependent Variables box and your independent variable `tutor` to the Grouping Variable box.
3. Under Tests, select `Student's`
4. Under Hypothesis, because we have a two-sided hypothesis select a two-sided hypothesis (Group 1 does not equal Group 2).
5. Under Additional Statistics, select `Mean difference`, `Effect size`, and `Descriptives`.
6. Under Assumption Checks, select all three options: `Homogeneity test`, `Normality test`, and `Q-Q plot`.

When you are done, your setup should look like this

2.5.1 Checking assumptions in jamovi

2.5.1.1 Testing normality

We test for normality using the Shapiro-Wilk test and the Q-Q plot. The Shapiro-Wilk test was not statistically significant (W = .98, $p = .827$); therefore, this indicates the data is normally distributed. Furthermore, the lines are fairly close to the diagonal line in the Q-Q plot. We can conclude that we satisfy the assumption of normality.

2.5.1.2 Testing homogeneity of variance

We test for homogeneity of variance using the Levene's test. The Levene's test was not statistically significant (F [1, 31] = 2.49, $p = .125$); therefore, this indicates our data satisfies the assumption of homogeneity of variance. However, I would add a caveat that we have a small sample of data ($n = 15$ for Anastasia and $n = 18$ for Bernadette) and the standard deviations are quite different from one another (SD = 9.00 vs 5.77, respectively). We should have tried to collect more data.

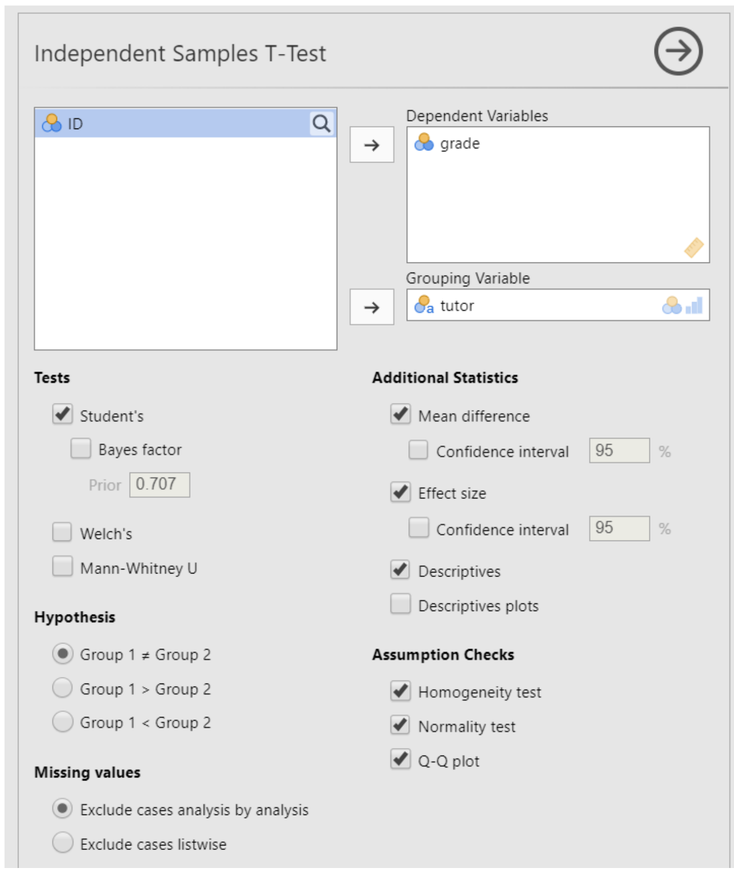


Figure 2.1: Independent t-test setup in jamovi

Assumptions

Normality Test (Shapiro-Wilk)

	W	p
grade	0.98	0.827

Note. A low p-value suggests a violation of the assumption of normality

Plots

grade

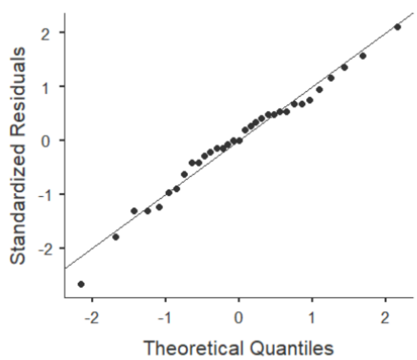


Figure 2.2: Testing normality in jamovi

Assumptions

Homogeneity of Variances Test (Levene's)

	F	df	df2	p
grade	2.49	1	31	0.125

Note. A low p-value suggests a violation of the assumption of equal variances

[3]

Figure 2.3: Testing homogeneity of variance in jamovi

2.5.2 Interpreting results

Once we are satisfied we have satisfied the assumptions for the independent t-test, we can interpret our results.

Independent Samples T-Test

Independent Samples T-Test

		Statistic	df	p	Mean difference	SE difference		Effect Size
grade	Student's t	2.12	31.00	0.043	5.48	2.59	Cohen's d	0.74

Group Descriptives

	Group	N	Mean	Median	SD	SE
grade	Anastasia	15	74.53	76.00	9.00	2.32
	Bernadette	18	69.06	69.00	5.77	1.36

Figure 2.4: Independent t-test results in jamovi

Our p-value is less than .05, so our results are statistically significant. We can write up our results in APA something like this:

> Anastasia's students ($M = 74.53$, $SD = 9.00$, $n = 15$) had significantly higher grades than Bernadette's students ($M = 69.06$, $SD = 5.77$, $n = 18$), $t\ (31) = 2.12$, $p = .043$, $d = .74$.

Sometimes, people like to put the statistics inside a parentheses. In that case, you need to change the parentheses around the degrees of freedom as brackets. Here's another example write-up of the results in APA style:

> I tested the difference in grades between Anastasia's students ($M = 74.53$, $SD = 9.00$, $n = 15$) and Bernadette's students ($M = 69.06$, $SD = 5.77$, $n = 18$). An independent samples t-test showed that the 5.48 mean difference between the tutor's student was statistically significant ($t\ [31] = 2.12$, $p = .043$, $d = .74$).

2.6 Additional information about the independent t-test

2.6.1 Positive and negative t values

Students often worry about positive or negative t-statistic values and are unsure how to interpret it. Positive or negative t-statistic values simply occur based on which group is listed first. Our t-statistic above is positive because we tested the difference between Anastasia and Bernadette: (Anastasia - Bernadette) = (74.53 - 69.06) = (5.48).

However, if we flipped it and tested the difference between Bernadette and Anastasia, our mean difference would be -5.48 and our t-statistic would be -2.12.

All that is to say, *your positive or negative t-statistic is arbitrary.* So do not fret!

However, it is important the sign of your t-statistic matches what you report. For example, notice the difference:

1. Anastasia's students had **higher** grades than Bernadette's, t (31) = **2.12**, p = .043, d = .74.
2. Bernadette's students had **lower** grades than Anastasia's, t (31) = **-2.12**, p = .043, d = .74.

One last note: this positive or negative t-statistic is only relevant for the independent and dependent t-test. You will not get negative values for the F-statistic or chi-square tests!

2.6.2 What if I violated assumptions?

The great news is that jamovi includes the Welch's t-statistic and the non-parametric version of the independent t-test (Mann-Whitney U)! The Welch's t-test has three main differences from the independent samples t-test: (a) the standard error (SE) is not a pooled estimate, (b) the degrees of freedom are calculated very different (not N - 2), and (c) it does not have an assumption of homogeneity of variance. The Mann-Whitney U is not calculated based on the mean but rather the median and compares ranks of values across the two groups: it has no assumptions about the distribution of data or homogeneity of variances.

Here's what statistic you should choose based on satisfying assumptions:

	Normality: satisfied	**Normality: not satisfied**
Homogeneity of Variance: satisfied	independent samples t-test	Mann-Whitney U
Homogeneity of Variance: not satisfied	Welch's t-test	Mann-Whitney U

Here is what the output for all three tests look like:

Independent Samples T-Test

Independent Samples T-Test

		Statistic	df	p	Mean difference	SE difference		Effect Size
grade	Student's t	2.12	31.00	0.043	5.48	2.59	Cohen's d	0.74
	Welch's t	2.03	23.02	0.054	5.48	2.69	Cohen's d	0.72
	Mann-Whitney U	79.50		0.046	6.00		Rank biserial correlation	0.41

Group Descriptives

	Group	N	Mean	Median	SD	SE
grade	Anastasia	15	74.53	76.00	9.00	2.32
	Bernadette	18	69.06	69.00	5.77	1.36

Figure 2.5: All independent t-test results in jamovi

2.6.2.1 Welch's t-test in jamovi

To conduct this in jamovi, under Tests select `Welch's`. You will interpret the results similarly to the independent t-test:

> Using a Welch's t-test, there was not a statistically significant difference in grades between Anastasia's students (M = 74.53, SD = 9.00, n = 15) and Bernadette's students (M = 69.06, SD = 5.77, n = 18), t (23.02) = 2.03, p = .054, d = .72.

Why is it no longer statistically significant? Which result should you trust? In reality, the difference in p-values is likely due to chance. However, the independent t-test and Welch's test have different strengths and weaknesses. If the two populations really do have equal variances, then the independent t-test is slightly more powerful (lower Type II error rate) than the Welch's test. However, if they *don't* have the same variances, then the assumptions of the independent t-test are violated and you may not be able to trust the results; you may end up with a higher Type I error rate. So it's a trade-off.

Which should you use? I tend to prefer always using Welch's t-test because if the variances are equal, then there will be practically no difference between the independent and Welch's t-test. But if the variances are not equal, then Welch's t-test will outperform the independent t-test. For that reason, defaulting to the Welch's t-test makes most sense to me.

2.6.2.2 Mann-Whitney U test

If you do not satisfy the assumption of normality (regardless of whether you satisfy the assumption of homogeneity of variance), you should either try to transform your data to be normally distributed or you will need to use a non-parametric test. In this case, if you originally wanted to perform an independent t-test, the non-parametric equivalent test is the Mann-Whitney U test.

I will not go into specifics, but the idea behind the Mann-Whitney U test is that you take all the values (regardless of group) and rank them. You then sum the ranks across groups and calculate your U statistic and p-value. You interpret the p-value like you normally would, but there are differences in how we report the results because this statistic is based on the *median* not the *mean*.

> Using the Mann-Whitney U test, there was a statistically significant difference in grades between Anastasia's students ($Mdn = 76$, $n = 15$) and Bernadette's students ($Mdn = 69$, $n = 18$), $t\,(23.02) = 2.03$, $p = .054$, $d = .72$.

2.7 Your turn!

Open the `Sample_Dataset_2014.xlsx` file that we will be using for all Your Turn exercises.

Perform independent t-tests based on the following research questions. Think critically about whether you should be using a one-tailed or two-tailed hypothesis and check your assumptions so you know which test to use!

To get the most out of these exercises, try to first find out the answer on your own and then use the drop-down menus to check your answer.

1. **Does height differ by gender (Gender: male = 0, female = 1)?**

    - Should you use a one-tailed or two-tailed hypothesis? one-tailed two-tailed
    - Which statistic should you use based on your assumptions? independent t-test Welch's t-test Mann Whitney U
    - Does height differ by gender? yes no

2. **Do athletes (Athlete: athletes = 1, non-athlete = 0) have faster sprint times than non-athletes?**

    - Should you use a one-tailed or two-tailed hypothesis? one-tailed two-tailed
    - Which statistic should you use based on your assumptions? independent t-test Welch's t-test Mann Whitney U
    - Do athletes have faster sprint times than non-athletes? yes no

3. **Do students who live on campus (LiveOnCampus: on campus = 1, off campus = 0) have higher English scores than students who live off campus?**

    - Should you use a one-tailed or two-tailed hypothesis? one-tailed two-tailed
    - Which statistic should you use based on your assumptions? independent t-test Welch's t-test Mann Whitney U
    - Does students who live on campus have higher English scores? yes no

4. **Does athletic status relate to math scores?**

    - Should you use a one-tailed or two-tailed hypothesis? one-tailed two-tailed
    - Which statistic should you use based on your assumptions? independent t-test Welch's t-test Mann Whitney U
    - Does athletic status relate to math scores? yes no

Chapter 3

Dependent t-test

3.1 What is the dependent t-test?

The dependent t-test is used to test the difference in our dependent variable between two categories in which participants are the *same* across categories. Our category variable is our independent variable. In other words, we use the independent t-test when we have a research question with a **continuous dependent variable** and a **categorical independent variable with two categories in which the same participants are in each category**.

The dependent t-test is also called a dependent samples t-test or paired samples t-test.

3.2 Data set-up

To conduct the dependent t-test, we first need to ensure our data is set-up properly in our dataset. This requires having two columns: one is our dependent variable score for the participant in one category and the other column is our dependent variable score for the participant in the other category. Each row is a unique participant or unit of analysis. Here's what example data may look like if we were testing for differences in test scores across the same participants in the fall and spring:

Table 3.1: Example data for the dependent t-test

ID	TestScore_Fall	TestScore_Spring
1	75	86
2	79	80

ID	TestScore_Fall	TestScore_Spring
3	65	75
4	81	79
5	73	82
6	72	84
7	69	90
8	60	72
9	75	75
10	74	81

In the example data above, what is your **independent variable**? ID Semester TestScore

In the example data above, what is your **dependent variable**? ID Semester Test Score

3.3 The math behind the independent t-test

The basic math of the dependent t-test is the mean difference divided by the standard error, which is estimated based on the standard deviation and sample size (N).

$t = \frac{\bar{X}_1 - \bar{X}_2}{s_d/\sqrt{N}}$

3.4 Assumptions

As a parametric test, the independent t-test has the same assumptions as other parametric tests minus the homogeneity of variance assumption because we are dealing with the same people across categories

1. The *differences in scores* in the dependent variable are **normally distributed**
2. The dependent variable is **interval or ratio** (i.e., continuous)
3. Scores are **independent** across participants

We cannot test the second and third assumptions; rather, those are based on knowing your data.

However, we can and should test for the first assumption. Fortunately, the dependent samples t-test in jamovi has two check boxes under "Assumption Checks" that lets us test normality.

3.5 In jamovi

Let's run an example with data from lsj-data. Open data from your Data Library in "lsj-data". Select and open "Chico". This dataset is hypothetical data from Dr. Chico's class in which students took two tests: one early in the semester and one later in the semester. Dr. Chico thinks that the first test is a "wake up call" for students. When they realise how hard her class really is, they'll work harder for the second test and get a better mark. Is she right? Let's test it!

1. To perform an dependent t-test in jamovi, go to the Analyses tab, click the T-Tests button, and choose "Paired Samples T-Test".
2. Move both measurements of your dependent variable (`grade_test1` and `grade_test2`) to the Paired Variables box.
3. Under Tests, select `Student's`
4. Under Hypothesis, choose the correct hypothesis: Measure 1 is not equal to Measure 2 Measure 1 > Measure 2 Measure 1 < Measure 2
5. Under Additional Statistics, select `Mean difference`, `Effect size`, and `Descriptives`.
6. Under Assumption Checks, select both options: `Normality test` and `Q-Q plot`.

When you are done, your setup should look like this

3.5.1 Checking assumptions in jamovi

3.5.1.1 Testing normality

We test for normality using the Shapiro-Wilk test and the Q-Q plot. The Shapiro-Wilk test was not statistically significant (W = .97, p = .678); therefore, this indicates the data is normally distributed. Furthermore, the lines are fairly close to the diagonal line in the Q-Q plot (although it's a bit hard to tell because our sample size is small). We can conclude that we satisfy the assumption of normality.

3.5.2 Interpreting results

Once we are satisfied we have satisfied the assumptions for the dependent t-test, we can interpret our results.

Our p-value is less than .05, so our results are statistically significant. We can write up our results in APA something like this:

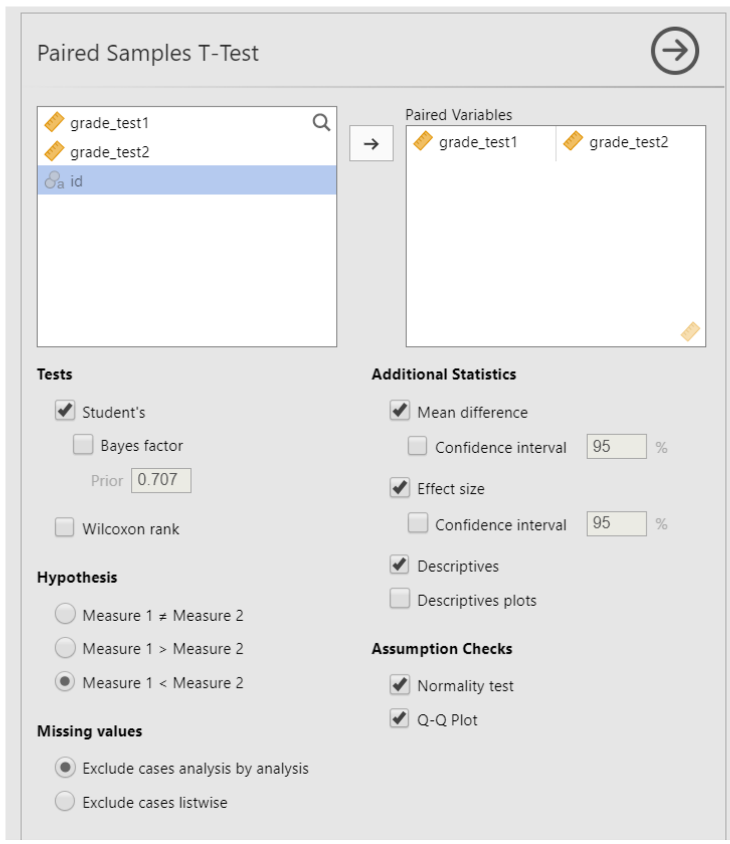


Figure 3.1: Dependent t-test setup in jamovi

Normality Test (Shapiro-Wilk)

			W	p
grade_test1	-	grade_test2	0.97	0.678

Note. A low p-value suggests a violation of the assumption of normality

Plots

grade_test1 - grade_test2

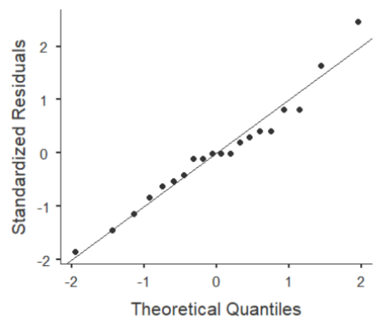


Figure 3.2: Testing normality in jamovi

Paired Samples T-Test

Paired Samples T-Test

			statistic	df	p	Mean difference	SE difference		Effect Size
grade_test1	grade_test2	Student's t	−6.48	19.00	< .001	−1.40	0.22	Cohen's d	−1.45

Note. H_a Measure 1 < Measure 2

Descriptives

	N	Mean	Median	SD	SE
grade_test1	20	56.98	57.70	6.62	1.48
grade_test2	20	58.38	59.70	6.41	1.43

Figure 3.3: Dependent t-test results in jamovi

> The 20 students in Dr. Chico's class performed worse on the first test (M = 56.98, SD = 6.62) than they did on the second test (M = 58.38, SD = 6.41), $t(19)$ = -6.48, p < .001, d = -1.45.

Remember in the previous chapter that our t-test can be negative but we can always flip the interpretation. Here's another example of how we could write-up our results in APA style:

> Dr. Chico's hypothesis was correct in that her 20 students performed better on the second test (M = 58.38, SD = 6.41) than they did on the first test (M = 56.98, SD = 6.62), $t(19)$ = 6.48, p < .001, d = 1.45.

3.6 What if I violated assumptions?

If you violated the assumption of normality and no transformation fixed your data, then you can perform the non-parametric version of the dependent t-test called the Wilcoxon Rank test. As a reminder, non-parametric tests do not make assumptions about the distribution of data because it deals with the *median* not the *mean*.

Here is the output for both the dependent t-test and the Wilcoxon rank test:

Paired Samples T-Test

Paired Samples T-Test

			Statistic	df	p		Effect Size
grade_test1	grade_test2	Student's t	−6.48	19.00	< .001	Cohen's d	−1.45
		Wilcoxon W	2.00 ᵃ		< .001	Rank biserial correlation	−0.98

Note. H_a Measure 1 < Measure 2

ᵃ 1 pair(s) of values were tied

Descriptives

	N	Mean	Median	SD	SE
grade_test1	20	56.98	57.70	6.62	1.48
grade_test2	20	58.38	59.70	6.41	1.43

Figure 3.4: All independent t-test results in jamovi

3.6.0.1 Wilcoxon rank in jamovi

To conduct this in jamovi, under Tests select `Wilcoxon rank`. You will interpret the results similarly to the dependent t-test:

> Using Wilcoxon rank test, students' test scores were significantly higher at the second test (*Mdn* = 59.70) than at the first test (*Mdn* = 57.70), W = 2.00, $p < .001$.

The note about tied values is not necessary to discuss. It is just telling us one participant had identical values for both test1 and test2 (student15).

3.7 Your turn!

Open the `Sample_Dataset_2014.xlsx` file that we use for all Your Turn exercises.

Perform dependent t-tests based on the following research questions. Think critically about whether you should be using a one-tailed or two-tailed hypothesis and check your assumptions so you know which test to use!

To get the most out of these exercises, try to first find out the answer on your own and then use the drop-down menus to check your answer.

Note: Technically, none of our data is suitable for a dependent t-test in this dataset. We will pretend that the four test score variables (`English`, `Reading`, `Math`, and `Writing`) are really four measurements of the same underlying test. In reality, we would analyze this data using correlation.

1. **Do students perform better on the English test than they do the Writing test?**

   - Should you use a one-tailed or two-tailed hypothesis? one-tailed two-tailed
   - Which statistic should you use based on your assumptions? dependent t-test Wilcoxon rank
   - Do students perform better on the English test than they do the Writing test? yes no

2. **Does students' English scores relate to their Reading scores?**

   - Should you use a one-tailed or two-tailed hypothesis? one-tailed two-tailed
   - Which statistic should you use based on your assumptions? dependent t-test Wilcoxon rank
   - Does students' English scores relate to their Reading scores? yes no

Chapter 4

One-way ANOVA

4.1 What is the one-way ANOVA?

The one-way analysis of variance (ANOVA) is used to test the difference in our dependent variable between three or more different groups of observations. Our grouping variable is our independent variable. In other words, we use the one-way ANOVA when we have a research question with a **continuous dependent variable** and a **categorical independent variable with three or more categories in which different participants are in each category**.

The one-way ANOVA is also known as an independent factor ANOVA.

One thing to keep in mind is the one-way ANOVA is an omnibus statistic that tests against the null hypothesis that the three or more means are the same. It does not tell us where the mean differences are (e.g., that $1 > 2$); for that, we need planned contrasts or post-hoc procedures, which you'll learn about later.

4.2 Data set-up

To conduct the one-way ANOVA, we first need to ensure our data is set-up properly in our dataset. This requires having two columns: one with our continuous dependent variable and one indicating which group the participant is in. Each row is a unique participant or unit of analysis. Here's what example data may look like if we were testing for differences in a test score by students in my fall, spring, or summer semesters of my course

Table 4.1: Example data for the one-way ANOVA

ID	Semester	TestScore
1	Fall	86
2	Fall	80
3	Fall	75
4	Spring	79
5	Spring	82
6	Spring	84
7	Summer	90
8	Summer	72
9	Summer	75

In the example data above, what is your **independent variable**? ID Semester TestScore

In the example data above, what is your **dependent variable**? ID Semester TestScore

4.3 Why not multiple t-tests?

In the example above, we have three groups: fall, spring, and summer. We could just perform three separate t-tests: fall vs. spring, fall vs. summer, and spring vs. summer.

However, the reason we do not perform multiple t-tests is to reduce our Type I error rate. If I had performed three separate t-tests, set my alpha (Type I error rate) at 5%, and knew for a fact the effects do not actually exist, then each test has a probability of a Type I error rate of 5%. Because we are running three tests, our alpha rate actually becomes 1 - ($.95^3$)= 1 - .857 = 14.3%! So now our *familywise* or *experimentwise* error rate is 14.3%, not the 5% we originally set alpha at.

With three groups, that's not so bad, but let's see what happens with more tests we perform:

- **1 test**: 1 - ($.95^1$)= 1 - .95 = **5%**
- **2 tests**: 1 - ($.95^2$)= 1 - .9025 = **9.8%**
- **3 tests**: 1 - ($.95^3$)= 1 - .857 = **14.3%**
- **4 tests**: 1 - ($.95^4$)= 1 - .814 = **18.6%**
- **5 tests**: 1 - ($.95^5$)= 1 - .774 = **22.6%**
- **10 tests**: 1 - ($.95^{10}$)= 1 - .598 = **40.1%**
- **20 tests**: 1 - ($.95^{20}$)= 1 - .358 = **64.1%**

Ouch! 10 tests would have a Type I error rate of 40%! That means that if we performed 10 statistical tests (assuming the effect does not exist), then 40% of the results would be statistically significant by chance alone and would be a false positive. That's not good!

Therefore, we use the one-way ANOVA as one test to see if there is a difference overall. We can also do things to control or limit our familywise error rate, which I'll get into later.

4.4 The math behind the one-way ANOVA

The basic math of the one-way ANOVA is the between-group variance (mean squares - between groups) divided by the within-group variance (mean squares - within groups).

$F = \frac{BG\,variance}{WG\,variance} = \frac{MS_{BG}}{MS_{WG}} = \frac{\frac{SS_{BG}}{df_{BG}}}{\frac{SS_{WG}}{df_{WG}}}$

There are specific formulas for the between-group (also called the model) sum of squares (SS) and within-group (also called the residual) sum of squares. Keep in mind the following symbols:

- n/N = sample size (little n is per group, big N is the overall sample)
- k = number of groups
- X = mean
- s = standard deviation

The between-group SS is the *variation between group means*. The calculations you see below is to subtract the overall mean ($\bar{X}_{grand}$) from the group mean ($\bar{X}_k$), square that mean difference, multiply that by the sample size in that group (n_k), and then sum all the results of doing that per group.

$SS_{BG} = \sum n_k(\bar{X}_k - \bar{X}_{grand})^2$

$df_{BG} = k - 1$

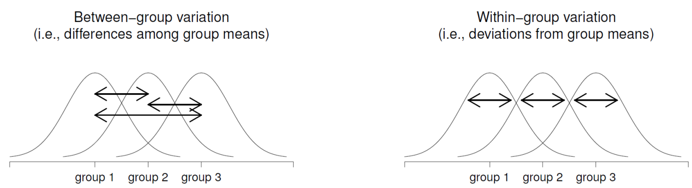
The within-group SS is the *variation of scores within groups*. The calculations you see below is to take the sample size in that group (n_k) and subtract 1 from it, then multiply it by the group's variance (s^2_k, which is standard deviation squared), and then sum all the results of doing that per group.

$SS_{WG} = \sum s^2_k(n_k - 1)$

$df_{WG} = N - k$

In other words, the F-ratio is a ratio of the group or experimental effect by the individual differences in performance.

For our F-ratio to be statistically significant, we want to *maximize* the variance between groups (numerator) and *minimize* the variance within groups (denominator). This is depicted in the image below. On the left (a) the arrows show the differences in the group means. On the right (b) the arrows highlight the variability within each group. You want to maximize (a) and minimize (b).



(a) (b)

Figure 4.1: Graphical illustration of the one-way ANOVA

4.4.1 An example with the math

In case it's useful, read below. If your eyes are glazing over, just move to the next section.

You can read the example dataset below. I've pulled the relevant data here so we can go through a hand calculation ourselves.

Drug	N	Mean	Variance
Anxifree	6	0.7167	.1537
Joyzepam	6	1.4833	.0457
Placebo	6	0.4500	.0790
Overall	**18**	**0.8833**	**.28**

Let's start with the easy stuff! Let's calculate our degrees of freedom.

$$df_{BG} = k - 1 = 3 - 1 = 2$$

$$df_{WG} = N - k = 18 - 3 = 15$$

Now let's move to the more complicated ones.

$SS_{BG} = \sum n_k(\bar{X}_k - \bar{X}_{grand})^2$
$SS_{BG} = 6(.7167 - .8833)^2 + 6(1.4833 - .8833)^2 + 6(.4500 - .8833)^2$
$SS_{BG} = 6(-.1666)^2 + 6(.6)^2 + 6(-.4333)^2$
$SS_{BG} = 6(.0278) + 6(.36) + 6(.1877)$
$SS_{BG} = .1665 + 2.16 + 1.1262$
$SS_{BG} = 3.453$

Then we can calculate our MS_{BG}.

$MS_{BG} = \frac{SS_{BG}}{df_{BG}} = \frac{3.453}{2} = 1.727$

Let's move to our SS_{WG}.

$SS_{WG} = \sum s_k^2(n_k - 1)$
$SS_{WG} = .1537(6 - 1) + .0457(6 - 1) + .0790(6 - 1)$
$SS_{WG} = .1537(5) + .0457(5) + .0790(5)$
$SS_{WG} = .7685 + .2285 + .3950$
$SS_{WG} = 1.392$

Now we can calculate our MS_{WG}.

$MS_{WG} = \frac{SS_{WG}}{df_{WG}} = \frac{1.392}{15} = .0928$

Now all that's left is to calculate F!

$F = \frac{MS_{BG}}{MS_{WG}} = \frac{1.726}{.0928} = 18.60$

Compare to the F, df_{WG}, and df_{BG} in the output below in jamovi! Notice how close we are. Also notice how many decimals I retained throughout the analyses. I was a bit off when I first did this with only two decimals throughout. Retaining four decimals throughout got me only ~one-hundredth of a decimal off from the actual results. Neat!

4.5 Assumptions

As a parametric test, the one-way ANOVA has the same assumptions as other parametric tests:

1. The dependent variable is **normally distributed**
2. Variances in the two groups are roughly equal (i.e., **homogeneity of variances**)
3. The dependent variable is **interval or ratio** (i.e., continuous)
4. Scores are **independent** between groups

We cannot test the third and fourth assumptions; rather, those are based on knowing your data. However, we can and should test for the first two assumptions. Fortunately, the one-way ANOVA in jamovi has three check boxes under "Assumption Checks" that lets us test for both assumptions.

4.5.1 ANOVA is robust to violations

Although we should meet the assumptions as much as possible, in general the F-statistic is *robust* to violations of normality and homogeneity of variance. This means that you can still run the one-way ANOVA if you violate the assumptions, but *only when group sizes are equal or nearly equal.* If you have vastly different variances (such as 2:1 ratio or greater) or vastly different group sizes (such as a 2:1 ratio or greater), and especially if one group is small (such as 10 or fewer cases), then your F-statistic is likely to be very wrong. For example, if your larger group has the larger variance, then your F-statistic is likely to be non-significant or smaller than it should be; however, if your larger group has smaller variance,then your F-statistic is likely to be significant or bigger than it should be!

4.6 In jamovi

Let's run an example with data from lsj-data. Open data from your Data Library in "lsj-data". Select and open "clinicaltrial". This dataset is hypothetical data of a clinical trial in which you are testing a new antidepressant drug called *Joyzepam*. In order to construct a fair test of the drug's effectiveness, the study involves three separate drugs to be administered. One is a placebo, and the other is an existing antidepressant / anti-anxiety drug called *Anxifree*. A collection of 18 participants with moderate to severe depression are recruited for your initial testing. Because the drugs are sometimes administered in conjunction with psychological therapy, your study includes 9 people undergoing cognitive behavioral therapy (CBT) and 9 who are not. Participants are randomly assigned (doubly blinded, of course) a treatment, such that there are 3 CBT people and 3 no-therapy people assigned to each of the 3 drugs. A psychologist assesses the mood of each person after a 3 month run with each drug, and the overall improvement in each person's mood is assessed on a scale ranging from -5 to +5.

1. To perform a one-way ANOVA in jamovi, go to the Analyses tab, click the ANOVA button, and choose "ANOVA". You might be asking why we aren't choosing "One-Way ANOVA" and that's because the options there are too limited.
2. Move your dependent variable `mood.gain` to the Dependent Variable box and your independent variable `drug` to the Fixed Factors box.

3. Select ω^2 for your effect size.

4. Ignore the Model drop-down menu. If you are doing more complicated ANOVAs you will need this. We will ignore it.

5. In the Assumption Checks drop-down menu, select all three options: `Homogeneity test`, `Normality test`, and `Q-Q plot`.

6. Ignore the Contrasts and Post Hoc Tests drop-down menus for now. See below for more information on them.

7. In the Estimated Marginal Means drop-down menu, move your IV `drug` to the Marginal Means box and select `Marginal means plots`, `Marginal means tables`, and `Observed scores`, in addition to the pre-selected `Equal cell weights`.

When you are done, your setup should look like this:

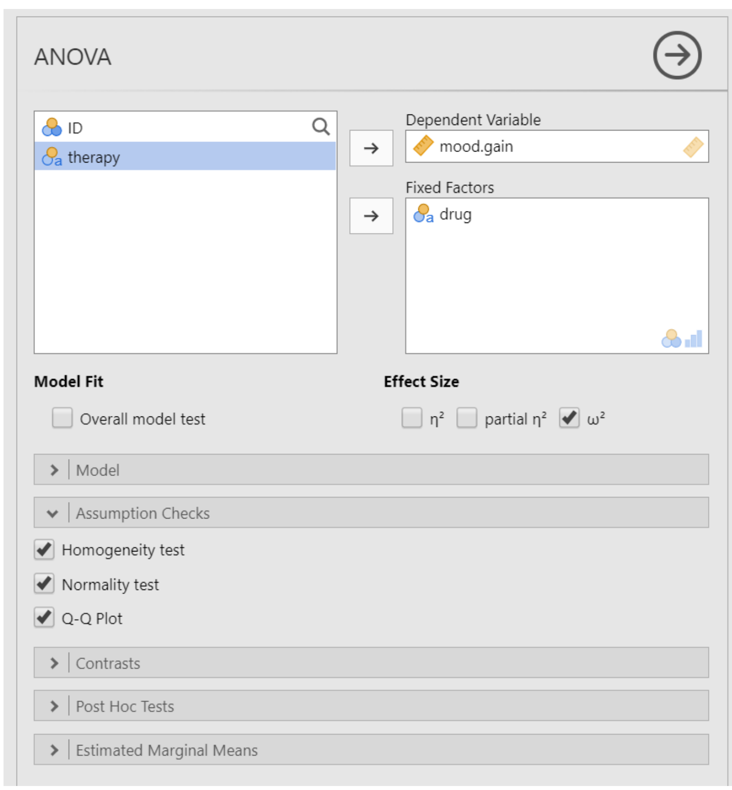


Figure 4.2: One-way ANOVA setup in jamovi

4.6.1 Checking assumptions in jamovi

We test for normality using the Shapiro-Wilk test and the Q-Q plot. The Shapiro-Wilk test was not statistically significant (W = .96, $p = .605$); therefore, this indicates the data is normally distributed. Furthermore, the lines are fairly close to the diagonal line in the Q-Q plot. We can conclude that we satisfy the assumption of normality.

We test for homogeneity of variance using the Levene's test. The Levene's test was not statistically significant (F [2, 15] = 1.45, p = .266); therefore, this indicates our data satisfies the assumption of homogeneity of variance. However, I would add a caveat that we have a small sample of data (N = 18); we should have tried to collect more data.

Assumption Checks

Normality Test (Shapiro-Wilk)

	W	p
mood.gain	0.96	0.605

Note. A low p-value suggests a violation of the assumption of normality

Homogeneity of Variances Test (Levene's)

	F	df1	df2	p
mood.gain	1.45	2	15	0.266

[3]

Figure 4.3: Testing assumptions in jamovi

4.7 Interpreting results

Once we are satisfied we have satisfied the assumptions for the one-way ANOVA, we can interpret our results.

Our p-value is less than .05, so our results are statistically significant. We can write up our results in APA something like this:

> There is a significant difference in mood gain across the three drug conditions, F (2, 15) = 18.61, $p < .001$, $\omega^2 = .66$.

Sometimes, people like to put the statistics inside a parentheses. In that case,

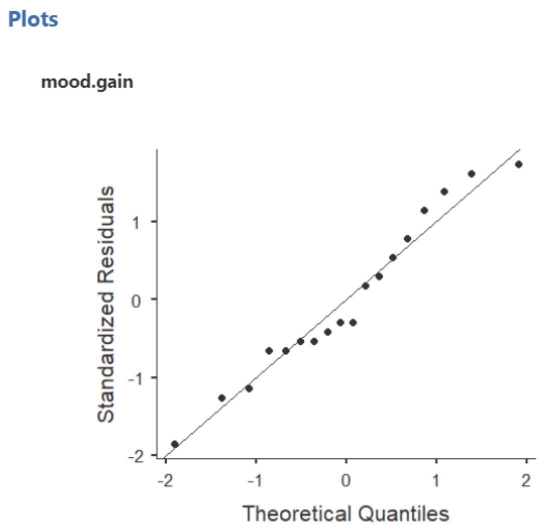


Figure 4.4: **CAPTION THIS FIGURE!!**

ANOVA

ANOVA - mood.gain

	Sum of Squares	df	Mean Square	F	p	ω^2
drug	3.45	2	1.73	18.61	< .001	0.66
Residuals	1.39	15	0.09			

[3]

Figure 4.5: One-way ANOVA results in jamovi

you need to change the parentheses around the degrees of freedom as brackets. Here's another example write-up of the results in APA style:

> There is a significant difference in mood gain across the three drug conditions (F [2, 15] = 18.61, $p < .001$, $\omega^2 = .66$).

4.7.1 A note on one-tailed vs. two-tailed tests

Unlike a t-test, we would not halve the p-value in an F-ratio because it is an omnibus test. Our planned contrasts or post-hoc tests can tell us where differences are, and we can provide directional hypotheses there if we so choose.

4.8 What if I violated assumptions?

The great news is that jamovi includes the Welch's F-statistic and the Kruskal-Wallis non-parametric test! The bad news is that you lose some functionality in jamovi when you use them. Just like with the Welch's t-statistic (for the independent t-test), it does not have the assumption of equal variances so it's appropriate to use if your data is normally distributed but does not have homogeneous variances. Similarly, the Kruskal-Wallis test is the non-parametric version of the one-way ANOVA and should be used if you do not satisfy the assumption of normality.

Here's what statistic you should choose based on satisfying assumptions:

	Normality: satisfied	**Normality: not satisfied**
Homogeneity of Variance: satisfied	one-way ANOVA	Kruskal-Wallis
Homogeneity of Variance: not satisfied	Welch's F-test	Kruskal-Wallis

4.8.0.1 Welch's F-test in jamovi

To conduct this in jamovi, you will need to use the "One-Way ANOVA" test, not the "ANOVA" test. The unfortunate thing about this test is that it strangely does not provide effect sizes.

In jamovi, under Variances select `Don't assume equal (Welch's)`. Move `mood.gain` to the Dependent Variable box and `drug` to your Grouping Variable box. You will interpret the results similarly to the one-way ANOVA:

> Using a Welch's F-test, there is a significant difference in mood gain across the three drug conditions, F (2, 9.49) = 26.32, $p < .001$.

One-Way ANOVA

One-Way ANOVA

		F	df1	df2	p
mood.gain	Welch's	26.32	2	9.49	< .001
	Fisher's	18.61	2	15	< .001

Figure 4.6: One-way ANOVA results in jamovi

4.8.0.2 Kruskal-Wallis test in jamovi

To perform the Kruskal-Wallis test in jamovi, you will need to select under the ANOVA button "One-Way ANOVA, Kruskal Wallis" towards the bottom of the list of options. Move `mood.gain` to the Dependent Variables box and `drug` to the Grouping Variable box. Select Effect size; if you need post hoc comparisons select DSCF pairwise comparisons (see section below on group differences). You will interpret the results similarly to the one-way ANOVA:

One-Way ANOVA (Non-parametric)

Kruskal-Wallis

	χ^2	df	p	ε^2
mood.gain	12.08	2	0.002	0.71

Figure 4.7: Kruskal-Wallis results in jamovi

> Using a Kruskal-Wallis test, there is a significant difference in mood gain across the three drug conditions, χ^2 (2) = 12.08, p = .002, ϵ^2 = .71.

Notice how in this case all three results converge and show there is a statistically significant difference in the results! The problem is… differences in which groups?

4.9 Finding Group Differences

Often, we're not interested in just *whether* there is a difference (which the F-statistic can tell us), but *where* the differences are between groups (which the

F-statistic cannot tell us). For that, we use either planned contrasts when you have specific hypotheses you want to test or post-hoc comparisons when you have no specific hypotheses.

Note: You do not perform contrasts or post hoc comparisons if your overall F statistic is not statistically significant. You do not interpret group differences if you fail to reject the null hypothesis that there are no group differences!

4.9.1 Planned Contrasts

If you have before-analysis hypotheses of group differences in your data, you will use planned contrasts. You can find the planned contrasts in the ANOVA (but not the one-way ANOVA) setup as a drop-down menu. Note that while I show all six contrasts that jamovi provides, you do not normally do multiple contrasts. These are just shown for illustrative purposes:

1. **Deviation**: compares the effect of each category (except the first category) to the overall experimental effect. The order of categories is alphabetical or numerical order. Notice how anxifree is considered the first category.

Contrasts

Contrasts - drug

	Estimate	SE	t	p
joyzepam - anxifree, joyzepam, placebo	0.60	0.10	5.91	< .001
placebo - anxifree, joyzepam, placebo	−0.43	0.10	−4.27	< .001

Figure 4.8: **CAPTION THIS FIGURE!!**

2. **Simple**: Each category is compared to the first category. The order of categories is alphabetical or numerical order. Notice how anxifree is considered the first category.

3. **Difference**: Each category (except the first) is compared to the mean effect of all previous categories.

4. **Helmert**: Each category (except the last) is compared to the mean effect of all subsequent categories.

Contrasts

Contrasts - drug

	Estimate	SE	t	p
joyzepam - anxifree	0.77	0.18	4.36	< .001
placebo - anxifree	−0.27	0.18	−1.52	0.150

Figure 4.9: **CAPTION THIS FIGURE!!**

Contrasts

Contrasts - drug

	Estimate	SE	t	p
joyzepam - anxifree	0.77	0.18	4.36	< .001
placebo - anxifree, joyzepam	−0.65	0.15	−4.27	< .001

Figure 4.10: **CAPTION THIS FIGURE!!**

Contrasts

Contrasts - drug

	Estimate	SE	t	p
anxifree - joyzepam, placebo	−0.25	0.15	−1.64	0.121
joyzepam - placebo	1.03	0.18	5.88	< .001

Figure 4.11: **CAPTION THIS FIGURE!!**

Contrasts

Contrasts - drug

	Estimate	SE	t	p
anxifree - joyzepam	−0.77	0.18	−4.36	< .001
joyzepam - placebo	1.03	0.18	5.88	< .001

Figure 4.12: **CAPTION THIS FIGURE!!**

5. **Repeated**: Each category is compared to the last category.

6. **Polynomial**: Tests trends in the data. It will examine the *n-1*$^{\text{th}}$ degree based on the number of groups. In this case, because we have 3 groups it is testing both a linear (^{1}) and quadratic (^{2}) trend. If we had 5 groups, it would test a linear (^{1}), quadratic (^{2}), cubic (^{3}), and quartic (^{4}) trend. Note that your factor levels must be ordinal for a polynomial contrast to make sense.

Contrasts

Contrasts - drug

	Estimate	SE	t	p
linear	−0.19	0.12	−1.52	0.150
quadratic	−0.73	0.12	−5.91	< .001

Figure 4.13: **CAPTION THIS FIGURE!!**

Test yourself! Which contrast would make most sense to test given that we want to know how our drug compares to the other two drugs? deviation simple difference helmert repeated polynomial

4.9.1.1 Writing up planned contrasts

Here's some example write-ups of the above results.

> There is a significant difference in mood gain across the three drug conditions, F (2, 15) = 18.61, $p < .001$. Repeated contrasts showed that *Joyzepam* (M = 1.48, SD = .21) outperformed both *Anxifree* (M = .72, SD = .39; $p < .001$) and the placebo condition (M = .45, SD = .28; $p < .001$).
>
> (Note how this example makes no sense because our data is not ordinal) There is a significant difference in mood gain across the three drug conditions, F (2, 15) = 18.61, $p < .001$. There was not a significant linear trend across the drug conditions (p = .150).

4.9.2 Post hoc comparisons

Often, we do not have any *a priori* (or planned) predictions or hypotheses about our group differences. In this case, we use post hoc procedures. These procedures do pairwise comparisons among all of our groups, like t-tests across each of our groups. As we noted on the first page of this handout, this can be highly problematic for our Type I error rate! Therefore, we must perform corrections to control our familywise error rate.

jamovi currently supports five types of post-hoc tests; I generally only use Tukey or Bonferonni:

1. **No correction**: This doesn't correct for a Type I error at all. Don't use this! I won't even show it. It's bad. Never use it. NEVER. You are warned!
2. **Tukey**: This is the post hoc test I use most often. It controls the Type I error rate well, but isn't as conservative of a control as the Bonferonni.
3. **Scheffe**: Honestly, I've never used it. I am not sure how it's calculated.
4. **Bonferroni**: This is the most conservative test. It's good if you only have a small number of comparisons to make or if you *really* want to control your Type I error rate. If you have a lot of them to test , then you should use something else.
5. **Holm**: Honestly, I've never used it. I am not sure how it's calculated.

Games-Howell for when you have unequal variances and Tukey for when you have equal variances. They will each calculate your p-values slightly differently but in a way to control for our Type I error rate as best it can. They are interpreted very similarly, so we will proceed with the Tukey's post hoc comparisons because we satisfied the assumption of equal variances.

To request post hoc tests from the one-way ANOVA, open the collapsed menu at the bottom of the setup menu. Select `Tukey (equal variances` under Post-Hoc Test and select `Mean difference`, `Report significance`, and `Flag significant comparisons` under Statistics. Optionally, you can request the `Test results (t and df)` although this is not necessary.

Below shows the post hoc test results for our one-way ANOVA. Notice the differences in p-values across the four post hoc tests and how all other values are the same. Notice how the Bonferroni is most conservative (i.e., has the largest p-values) and the Holm's is the least conservative (i.e., has the smallest p-values). Keep in mind you do not normally ask for multiple post hoc comparisons. Just pick one! Normally, I just pick Tukey's.

4.9.2.1 Writing up post hoc results

The way we would write up each of the post hoc comparisons is very similar. Given that I usually use Tukey, here is a write-up for those results:

> There is a significant difference in mood gain across the three drug conditions, $F\ (2, 15) = 18.61$, $p < .001$. Post hoc comparisons using

Post Hoc Tests

Post Hoc Comparisons - drug

Comparison											
drug		drug	Mean Difference	SE	df	t	p_{tukey}	$p_{scheffe}$	$p_{bonferroni}$	p_{holm}	Cohen's d
anxifree	-	joyzepam	−0.77	0.18	15.00	−4.36	0.002	0.002	0.002	0.001	−2.52
	-	placebo	0.27	0.18	15.00	1.52	0.312	0.343	0.451	0.150	0.88
joyzepam	-	placebo	1.03	0.18	15.00	5.88	< .001	< .001	< .001	< .001	3.39

Note. Comparisons are based on estimated marginal means

Figure 4.14: Post hoc test results in jamovi

> Tukey's HSD revealed that our drug *Joyzepam* (M = 1.48, SD = .21) outperformed both *Anxifree* (M = .72, SD = .39; p = .002) and the placebo condition (M = .45, SD = .28; p < .001); there were no differences between *Anxifree* and the placebo condition (p

4.9.3 Group differences with violated assumptions

If you are using Welch's F-test using the One-Way ANOVA in jamovi, you should select under Post-Hoc Tests `Games-Howell (unequal variances)`. These will be interpreted similarly to the post hoc comparisons above.

If you are using the Kruskal-Wallis test, you will select the check-box for `DSCF pairwise comparisons`. This stands for the Dwass-Steel-Critchlow-Fligner test. All you need to know is that they, too, are interpreted similarly to the post hoc comparisons above.

Unfortunately, you cannot perform contrasts with either the Welch's F-test or Kruskal-Wallis test.

4.10 Relationship between ANOVA and t-test

An ANOVA with two groups is identical to the t-test. That means the F and t statistics are directly related, and you will get the same p-value. For example, imagine you run a t-test and get a t-statistic of t (16) = -1.31, p = .210. If you ran it as a one-way ANOVA, you would get an F-statistic of F (1, 16) = 1.71, p = .210.

$F = t^2$

$t = \sqrt{F}$

Just a fun little bit of trivia! So if you accidentally do an F-test with two groups, no need to go back and redo the analyses (although you should if you are sharing your code for reproducibility). You can just convert your F to a t statistic easily!

4.11 Tips on writing up results

Writing up results in APA style is both a science and an art. There's a science to what you need to report. For example, you always report the statistics exactly the same: F (df_{WG}, df_{BG}) = X.XX, p = .XXX. You also always report the group means (M) and standard deviations (SD), although you can report them in-text like I did above or in a descriptives table like you can ask from jamovi.

However, there's also an art to it. Notice how I wrote that up in a way to summarize the findings as succinctly as possible. I could have said there was a difference between *anxifree* and *joyzepam* and a difference between *joyzepam* and the placebo, but that's a lot more words and isn't written in a way to focus on what I'm hoping to see: that my drug *joyzepam* performed better than the competitor *anxifree* and a placebo condition.

This is where you need to think creatively and be very critical in checking that what you say makes sense. Read your write-ups carefully! Have someone else read it. Can they understand what you mean?

4.12 Your turn!

Open the `Sample_Dataset_2014.xlsx` file that we will be using for all Your Turn exercises.

Perform one-way ANOVAs based on the following research questions. Check your assumptions and ensure you are using the correct tests.

To get the most out of these exercises, try to first find out the answer on your own and then use the drop-down menus to check your answer.

1. **Does students differ on English scores by rank (i.e., freshmen, sophomore, junior, senior)?**

    - Do you satisfy the assumption of normality? yes no
    - Do you satisfy the assumption of homogeneity of variance? yes no
    - Which statistic should you use? one-way ANOVA Welch's F-test Kruskal-Wallis test
    - Do students differ on English scores by rank? yes no
    - Should you perform a planned contrast or post hoc comparison? yes no
    - What are the results of the post hoc comparison? N/A - Don't perform Freshmen had higher English scores than sophomores, juniors, and seniors Freshmen and sophomores had higher English scores than juniors and seniors

2. **Does smoking status (Smoking: Nonsmoker = 0, Past smoker = 1, Current smoker = 2) relate to sprint time?**

    - Do you satisfy the assumption of normality? yes no
    - Do you satisfy the assumption of homogeneity of variance? yes no
    - Which statistic should you use? one-way ANOVA Welch's F-test Kruskal-Wallis test
    - Does smoking status relate to sprint time? yes no
    - Should you perform a planned contrast or post hoc comparison? yes no
    - What are the results of the post hoc comparison? N/A - Don't perform Nonsmokers had significantly faster sprint times than current smokers Nonsmokers and past smokers had significantly faster spring times than current smokers Nonsmokers had significantly faster sprint times than both past and current smokers

Chapter 5

Repeated Measures ANOVA

5.1 What is the repeated measures ANOVA?

The repeated measures analysis of variance (ANOVA) is used to test the difference in our dependent variable between three or more groups of observations in which all participants participate in all groups or levels. Our grouping variable is our independent variable. In other words, we use the one-way ANOVA when we have a research question with a **continuous dependent variable** and a **categorical independent variable with three or more categories in which the same participants are in each category**.

The repeated measures ANOVA is also sometimes called the one-way related ANOVA.

There are two ways we could have the repeated measures ANOVA. Perhaps the same group of participants are measured in the same dependent variable at three or more time points. In this case, our independent variable is time and our dependent variable is whatever is measured at each time point.

The other way we might have the repeated measures ANOVA is if all our participants participate in all conditions of our study. In this case, our independent variable is the treatment or condition and the dependent variable is whatever is measured in each treatment or condition.

5.2 Data set-up

To conduct the repeated measures ANOVA, we first need to ensure our data is set-up properly in our dataset. This requires having two columns: one with our

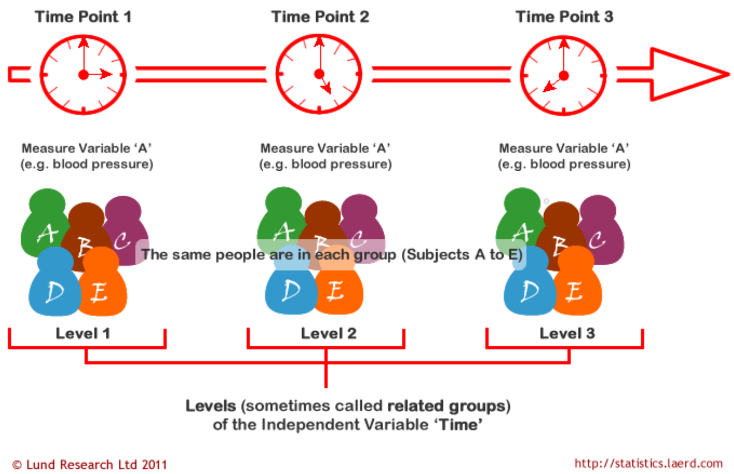


Figure 5.1: Repeated measures ANOVA by Time

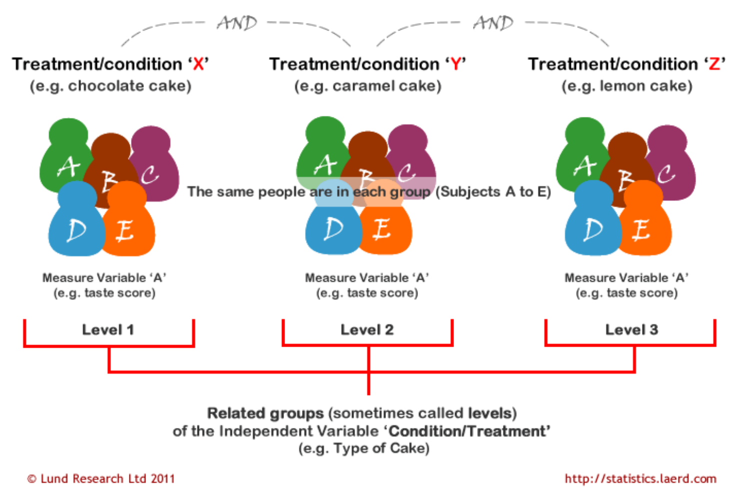


Figure 5.2: Repeated measures ANOVA by Conditions

continuous dependent variable and one indicating which group the participant is in. Each row is a unique participant or unit of analysis. Here's what example data may look like if we were testing for differences in a test score by students in my fall, spring, or summer semesters of my course in which three participants are in each of my three courses.

Table 5.1: Example data for the repeated measures ANOVA

ID	Fall	Spring	Summer
1	86	79	90
2	80	82	72
3	75	84	75

In the example data above, what is your **independent variable**? ID Semester TestScore

In the example data above, what is your **dependent variable**? ID Semester TestScore

5.3 The math behind the repeated measures ANOVA

The basic math of the repeated measures ANOVA is very similar to the one-way ANOVA except that the SS_{WG} is partitioned into two parts. Since there are the same subjects in each group, we can remove the variability due to the individual differences between subjects from the within-groups variability.

$F = \frac{BG\, variance}{WG\, variance} = \frac{MS_{BG}}{MS_{WG}} = \frac{\frac{SS_{BG}}{df_{BG}}}{\frac{SS_{WG}}{df_{WG}}}$

So whereas our denominator in the one-way ANOVA has SS_{WG}, our denominator for the repeated measures ANOVA has $SS_{WG} - SS_{subjects}$. You can see this in more detail in the figure below by Laerd Statistics. The one-way ANOVA (also known as the independent samples ANOVA) splits the total variability in scores by the between-group (conditions) variability and the within-groups (error) variability. The repeated measures anova takes that within-groups variability and splits out the subject variability from the error variability.

I won't go into any more detail on the math of the repeated measures ANOVA, but you can read more on Laerd Statistics guide for the repeated measures ANOVA.

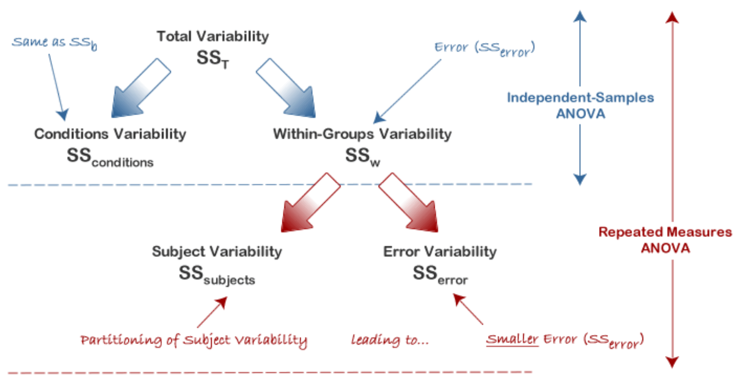


Figure 5.3: Repeated measures ANOVA by Time

5.4 Assumptions

As a parametric test, the repeated measures ANOVA has the same assumptions as other parametric tests:

1. The dependent variable is **normally distributed**
2. Variances in the two groups are roughly equal (i.e., **homogeneity of variances**); in repeated measures ANOVA this is called the assumption of **sphericity**
3. The dependent variable is **interval or ratio** (i.e., continuous)
4. ~~Scores are **independent** between groups~~ (this assumption is not relevant because all participants participate in all conditions)

We cannot test the third and fourth assumptions; rather, those are based on knowing your data. However, we can and should test for the first two assumptions. Fortunately, the one-way ANOVA in jamovi has three check boxes under "Assumption Checks" that lets us test for both assumptions.

5.4.1 Sphericity Assumption

The sphericity assumption is essentially the repeated measures ANOVA equivalent of homogeneity of variances. Sphericity means there is equality of variances of the *differences* between treatment levels. For example, if there are three groups, then the difference in all three pairs of differences (1-2, 1-3, 2-3) need to have approximately equal variances. You only need to care about sphericity when there are at least three conditions, which is why we did not talk about this with the dependent t-test.

Fortunately, like the other assumption checks, testing for sphericity is as simple as a checkbox in jamovi.

5.5 In jamovi

Let's run an example with data from lsj-data. Open data from your Data Library in "lsj-data". Select and open "broca".

This dataset is hypothetical data in which six patients suffering from Broca's Aphasia (a language deficit commonly experienced following a stroke) complete three word recognition tasks. On the

first (speech production) task, patients were required to repeat single words read out aloud by the researcher. On the second (conceptual) task, designed to test word comprehension, patients were required to match a series of pictures with their correct name. On the third (syntax) task, designed to test knowledge of correct word order, patients were asked to reorder syntactically incorrect sentences. Each patient completed all three tasks. The order in which patients attempted the tasks was counterbalanced between participants. Each task consisted of a series of 10 attempts. The number of attempts successfully completed by each patient are provided in the dataset.

1. To perform a one-way ANOVA in jamovi, go to the Analyses tab, click the ANOVA button, and choose "Repeated Measures ANOVA".
2. Under "Repeated Measures Factors" name your independent variable. In this case you can name it "Task". Rename the three levels of Task: Speech, Conceptual, and Syntax.
3. Under "Repeated Measures Cells" move the given variable into the correct level. For example, you'll move Speech to the Speech cell.
4. Select Generalised η^2 as your measure of effect size.
5. In the Assumption Checks drop-down menu, select `Sphericity tests`. You'll note that if you violate the assumption of sphericity, there are two corrections provided. These will be discussed later.
6. In the Post Hoc Tests drop-down menu, select `Tukey`. Remember that we only interpret these if the overall F is statistically significant.
7. In the Estimated Marginal Means drop-down menu, move Task to the Marginal Means box, select `Marginal means tables`, and select `Observed scores.`

When you are done, your setup should look like this:

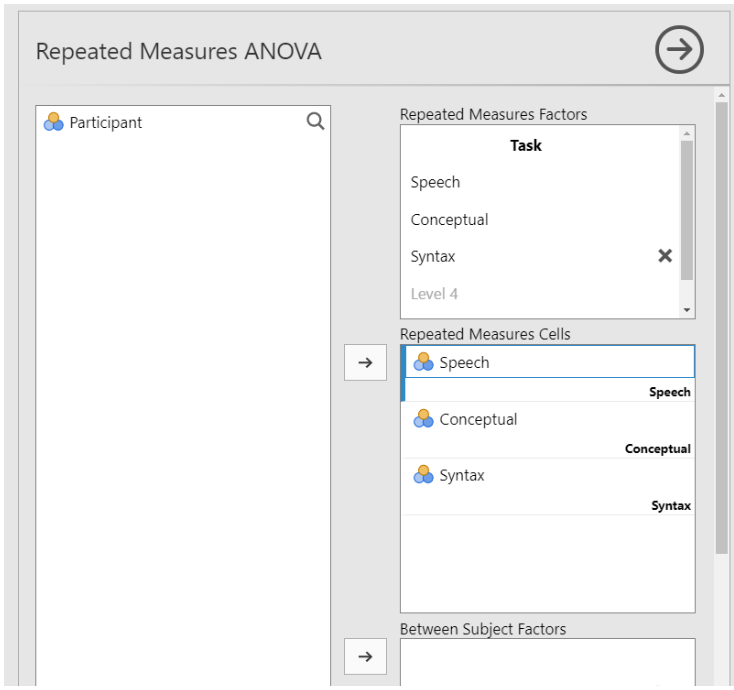


Figure 5.4: Repeated Measures ANOVA setup in jamovi

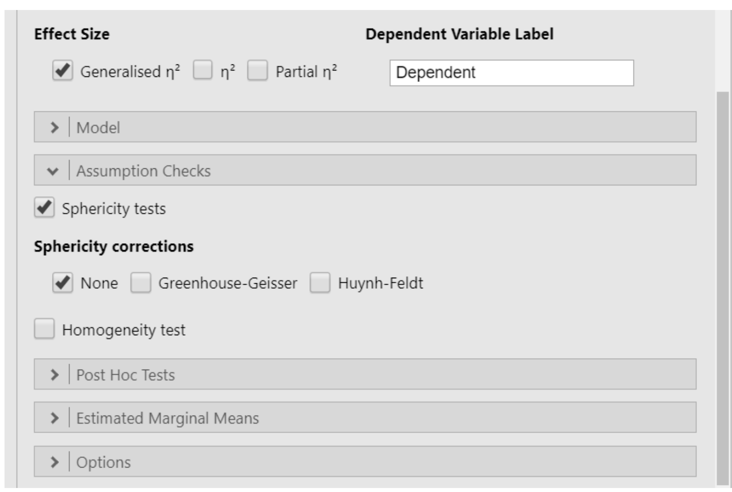


Figure 5.5: Repeated Measures ANOVA setup in jamovi

5.5.1 Checking assumptions in jamovi

You'll notice there are no options to check for normality in the repeated measures ANOVA in jamovi. There's an interesting conversation on the topic in the jamovi forums. Suffice to say, it's complicated and maybe someday they will implement it. For now, we just won't worry about it for the repeated measures ANOVA.

So what we need to worry about is testing our assumption of sphericity. You should have checked the box `Sphericity tests` under the Assumption Checks drop-down menu. That produces the following output:

Assumptions

Tests of Sphericity

	Mauchly's W	p	Greenhouse-Geisser ε	Huynh-Feldt ε
Task	0.85	0.720	0.87	1.00

Figure 5.6: Testing sphericity in jamovi

Mauchly's test of sphericity tests the null hypothesis that the variances of the differences between the conditions are equal. Therefore, just like with our previous assumption checks, if Mauchly's test is non-significant (i.e., $p > .05$, as is the case in this analysis) then it is reasonable to conclude that the variances of the differences are not significantly different. This means we satisfy the assumption of sphericity and can conclude that the variances of the differences are roughly equal.

If Mauchly's test had been statistically significant ($p < .05$), then we would conclude that the assumption had *not* been met. In this case, we should apply a correction to the F-value obtained in the repeated measures ANOVA:

- If the Greenhouse-Geisser value in the "Tests of Sphericity" table is $> .75$ then you should use the Huynh-Feldt correction.
- If the Greenhouse-Geisser value is $< .75$, then you should use the Greenhouse-Geisser correction.

You can select these corrections in the Assumption Checks drop-down menu.

5.6 Interpreting results

Once we are satisfied we have satisfied the assumptions for the repeated measures, we can interpret our results.

Repeated Measures ANOVA

Within Subjects Effects

	Sum of Squares	df	Mean Square	F	p	η^2_G
Task	24.78	2	12.39	6.93	0.013	0.41
Residual	17.89	10	1.79			

Note. Type 3 Sums of Squares

[3]

Between Subjects Effects

	Sum of Squares	df	Mean Square	F	p	η^2_G
Residual	17.11	5	3.42			

Note. Type 3 Sums of Squares

Figure 5.7: One-way ANOVA results in jamovi

You'll notice that jamovi provides you both a Within Subjects Effects table and Between Subjects Effects table. However, we only have a within-subjects effect (Task). Why did it give us a between-subjects table? With the repeated-measures ANOVA (which only has within-subjects IVs), this is just our SS_{BG}. However, because we don't have one, it's not calculating anything. We can ignore it. It is only useful if we are conducting a mixed factorial ANOVA with both between-subjects and within-subjects effects.

Therefore, the Within Subjects Effects table is of most use to us. We can see that the overall effect of Task is statistically significant (p = .013). Therefore we can look at our Post Hoc Tests results.

The Tukey post hoc differences show that there was a significant difference between speech and syntax (p = .011), but not between conceptual and both speech and syntax. Last, we can look at the Estimated Marginal Means - Task table to see the group means for reporting purposes. This shows us that participants recognized significantly more words in the speech task than in the syntax task.

We can write this up in APA style similar to the one-way ANOVA.

Post Hoc Tests

Post Hoc Comparisons - Task

Comparison							
Task		Task	Mean Difference	SE	df	t	p_{tukey}
Speech	-	Conceptual	1.00	0.77	10.00	1.29	0.429
	-	Syntax	2.83	0.77	10.00	3.67	0.011
Conceptual	-	Syntax	1.83	0.77	10.00	2.37	0.091

Figure 5.8: One-way ANOVA results in jamovi

> A repeated measures ANOVA was performed examining how three tasks affected word recognition in patients suffering from Broca's Aphasia. Task significantly affected word recall, F (2, 10) = 6.93, p = .013, η^2_G = .41. Tukey's post hoc difference tests indicated that participants recognized significantly more words in the speech task (M = 7.17, SE = .62) than participants in the syntax task (M = 4.33, SE = .62; p = .011). There were no differences between the conceptual task (M = 6.17, SE = .62) and both the speech and syntax tasks.

5.7 What if I violated assumptions?

We have already discussed what to do if you violate the assumption of sphericity above; you select one of the two sphericity corrections.

If you violate the assumption of normality or if the dependent variable is ordinal, then you can use the Friedman test. You can select this using the Repeated Measures ANOVA - Friedman option under the ANOVA analysis.

Friedman's test can only examine one within-subjects variable, so you will move all three conditions (Speech, Conceptual, and Syntax) to the Measures box. Select `Pairwise comparisons (Durbin-Conover` for post hoc comparisons and `Descriptives` for the Means and Medians. Optionally, you can select to plot either the Means or Medians. The Setup is shown below.

Once you've set-up the analysis, you can interpret the results. Overall, we continue to see a statistically significant result and that there is only a significant difference between speech and syntax.

We can write up the results similarly as before:

> Friedman's test was performed examining how three tasks affected word recognition in patients suffering from Broca's Aphasia. Task

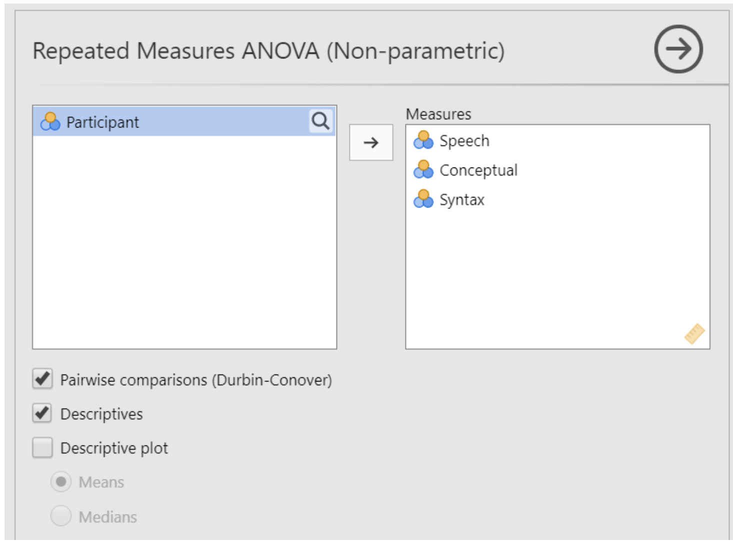


Figure 5.9: Repeated Measures ANOVA setup in jamovi

> significantly affected word recall, χ^2 (2) = 6.64, p = .036. Pairwise comparisons using Durbin-Conover indicated that participants recognized significantly more words in the speech task (M = 7.17, Mdn = 7.50) than participants in the syntax task (M = 4.33, Mdn = 6.50; p = .006). There were no differences between the conceptual task (M = 6.17, Mdn = 6.50) and both the speech and syntax tasks.

5.8 Tips on writing up results

Writing up results in APA style is both a science and an art. There's a science to what you need to report. For example, you always report the statistics exactly the same: F (df_{WG}, df_{BG}) = X.XX, p = .XXX. You also always report the group means (M) and standard deviations (SD), although you can report them in-text like I did above or in a descriptives table like you can ask from jamovi.

However, there's also an art to it. Notice how I wrote that up in a way to summarize the findings as succinctly as possible. I could have said there was a difference between *anxifree* and *joyzepam* and a difference between *joyzepam* and the placebo, but that's a lot more words and isn't written in a way to focus on what I'm hoping to see: that my drug *joyzepam* performed better than the competitor *anxifree* and a placebo condition.

This is where you need to think creatively and be very critical in checking that what you say makes sense. Read your write-ups carefully! Have someone else read it. Can they understand what you mean?

Repeated Measures ANOVA (Non-parametric)

Friedman

χ^2	df	p
6.64	2	0.036

Pairwise Comparisons (Durbin-Conover)

			Statistic	p
Speech	-	Conceptual	1.44	0.180
Speech	-	Syntax	3.50	0.006
Conceptual	-	Syntax	2.06	0.067

[5]

Descriptives

	Mean	Median
Speech	7.17	7.50
Conceptual	6.17	6.50
Syntax	4.33	4.50

Figure 5.10: Repeated Measures ANOVA setup in jamovi

5.9 Your turn!

Open the `Sample_Dataset_2014.xlsx` file that we will be using for all Your Turn exercises.

Perform a repeated measures ANOVA based on the following research questions. Check your assumptions and ensure you are using the correct tests.

To get the most out of these exercises, try to first find out the answer on your own and then use the drop-down menus to check your answer.

1. **Does students differ on their test scores (English, Reading, Math, Writing)?**

    - Based on your understanding of the nature of the test scores, which statistic should you use? repeated measures ANOVA Friedman's test
    - Should you apply a sphericity correction? If so, which one? N/A - using Friedman's test no, assumption satisfed yes, Greenhouse-Geisser yes, Huynh-Feldt
    - Do students differ on their test scores? yes no
    - Should you perform a planned contrast or post hoc comparison? yes no
    - What are the results of the post hoc comparison? N/A - Don't perform All test scores were significantly different from one another All test scores were significantly different from one another except for English and Reading

Chapter 6

Factorial ANOVA

6.1 What is factorial ANOVA?

Factorial ANOVA allows us to examine two or more independent variables (IVs) simultaneously. There are several types of factorial designs:

- **Independent factorial design**: several between-group (independent) IVs
- **Repeated measures factorial design**: several within-group (repeated-measures) IVs
- **Mixed factorial design**: some between-group and some within-group IVs

Furthermore, you may read about ANOVAs referred to as "one-way", "two-way", "three-way" or greater. This simply refers to how many independent variables there are. Factorial ANOVAs are sometimes also referenced by how many groups per IV there are; for example, a 2 x 3 ANOVA is a factorial ANOVA in which the first IV has two conditions and the second IV has three conditions. You would also specify which IVs are between-group and which are within-group. For example, you might write that your design is a 2 (between-subjects: gender) x3 (within-subjects: task) mixed factorial.

We won't be going into too much detail on the different factorial ANOVA designs. Instead, I will walk through illustrative cases so that if you want to apply them in the future you can mimic the procedures below.

6.2 Independent Factorial ANOVA

The independent factorial ANOVA has multiple between-group IVs. Let's run an example with data from lsj-data. Open data from your Data Library in "lsj-data". Select and open "rtfm". This data has two IVs: attend (whether or not the student turned up to lectures) and reading (whether or not the student actually read the textbook). 1 = they did and 0 = they did not.

Because we do not have a within-group IV, we will select the ANOVA analysis. Move grade to your Dependent Variable box and both attend and reading to your Fixed Factors. Select all the same options as you did for the one-way ANOVA (i.e., ω^2, assumption checks, Tukey's post hoc tests for the two variables attend and reading, estimated marginal means).

Let's go through the output (check that your output matches!) and then discuss how to write up the results in APA format. First, we need to check assumptions. The Levene's test and Shapiro-Wilk's test are shown below. We can see that we meet the assumption of normality but fail to meet the assumption of homogeneity of variance. Unfortunately, we cannot perform a Welch's F-test with more than one independent factorial so we will note this failed assumption and move on.

Assumption Checks

Homogeneity of Variances Test (Levene's)

F	df1	df2	p
6.80e+32	3	4	< .001

[3]

Normality Test (Shapiro-Wilk)

Statistic	p
0.96	0.851

Figure 6.1: Assumption check results in jamovi

Let's look at the main results next. We got three lines of results in addition to the typical residuals (error). The first two lines are our main effects of `attend` and `reading` on grades. The p-values for both are statistically significant indicating attend affects grades and reading affects grades. However, it also added an interaction term of `attend * reading`, which is not statistically significant.

This means we do not have an interaction between attend and reading on grades. Interactions will be discussed in more detail in the next section.

ANOVA

ANOVA - grade

	Sum of Squares	df	Mean Square	F	p	ω^2
attend	648.00	1	648.00	18.25	0.013	0.26
reading	1568.00	1	1568.00	44.17	0.003	0.64
attend ✻ reading	8.00	1	8.00	0.23	0.660	−0.01
Residuals	142.00	4	35.50			

Figure 6.2: Independent factorial ANOVA results in jamovi

Although we could simply look at the means to know whether attending or reading had higher grades than not attending or not reading because there are only two conditions, we can also look at the post hoc tests and definitely need to look at them if we have three or more conditions per IV. These are shown below. Because the mean differences are negative, we can determine that the second group had higher means than the second group. We can confirm that with the estimated marginal means (not shown here).

Post Hoc Tests

Post Hoc Comparisons - reading

Comparison								
reading		reading	Mean Difference	SE	df	t	p_{tukey}	Cohen's d
0	-	1	−28.00	4.21	4.00	−6.65	0.003	−4.70

Note. Comparisons are based on estimated marginal means

Post Hoc Comparisons - attend

Comparison								
attend		attend	Mean Difference	SE	df	t	p_{tukey}	Cohen's d
0	-	1	−18.00	4.21	4.00	−4.27	0.013	−3.02

Note. Comparisons are based on estimated marginal means

Figure 6.3: Post hoc results in jamovi

Last, we can write-up our results!

> We were interested in knowing how attendance and reading affected student grades. An independent factorial ANOVA showed that both

> attendance (F [1, 4] = 18.25, p = .013, ω^2 = .26) and reading (F [1, 4] = 44.17, p = .003, ω^2 = .64) affected student grades; there was no significant interaction between attendance and reading (F [1, 4] = 8.00, p = .660, ω^2 = -.01). Post hoc comparisons using Tukey's HSD show that students who attended lectures (M = 75.50, SE = 2.98) had higher grades than students who did not (M = 57.50, SE = 2.98; p = .003, d = 4.70); furthermore, students who read (M = 80.50, SE = 2.98) had higher grades than students who did not (M = 52.50, SE = 2.98; p = .013; d = 3.02).

6.2.1 Interactions

Interactions occur when the effect of one IV on the DV depends on the lvel of the other IV. If you did not want to test for interaction effects, you could remove them from the Model Terms in the Model drop-down menu.

However, by default they will include them. If you have 2-3 IVs, it may be reasonable to look at these interactions. However, 3-variable interactions (e.g., IV1 * IV2 * IV3) are pushing it and 4-variable interactions are highly implausible. Be critical in which interaction terms you include!

jamovi can provide a plot of your interaction, which can be helpful to help interpret your results. Below is the plot for our interaction of attendance on reading.

The parallel lines that are sloping upward tell me there is a significant main effect for both IVs but no interaction. How do I know that? With two variables, there are only so many interaction shapes possible. This website does a fantastic time showing you all 8 combinations of the three effects (2 main effects and 1 interaction effect). Spend some time looking through it and familiarizing yourself with the plots!

6.3 Repeated Measures Factorial ANOVA

This is also sometimes called the two-way (or three-way or n-way, depending on the n of IVs you have) repeated measures ANOVA. Let's go through an example repeated measures factorial ANOVA. The dataset is courtesy of Real Statistics Using Excel.

> A company has created a new training program for their customer service staff. To test the effectiveness of the program they took a sample of 10 employees and assessed their performance in three areas: Product (knowledge of the company's products and services), Client (their ability to relate to the customer with politeness and

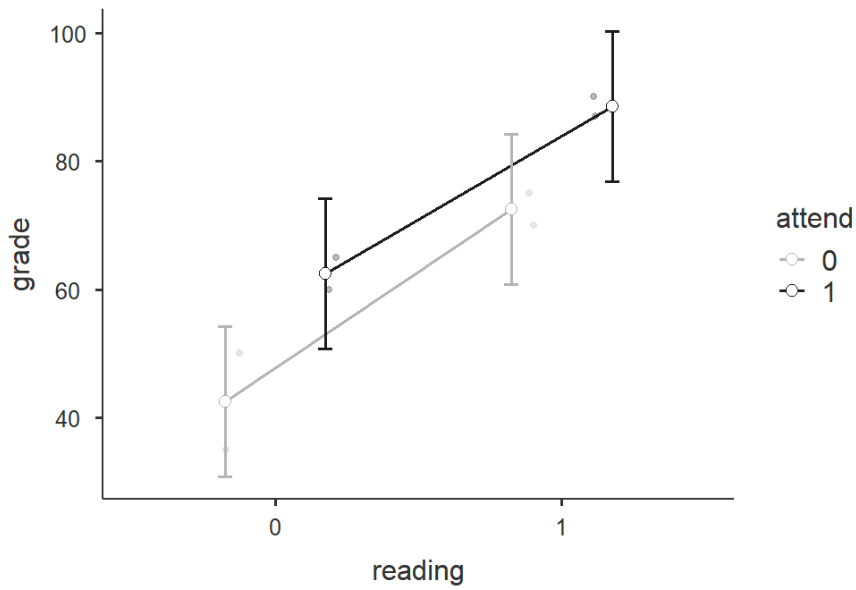


Figure 6.4: Interaction plot in jamovi

> empathy) and Action (their ability to take action to help the customer). They then had the same 10 employees take the training course and rated their performance after the program in the same three areas. -Real Statistics Using Excel

Table 6.1: Dataset for the repeated measures factorial ANOVA

ID	Pre-Product	Pre-Client	Pre-Action	Post-Product	Post-Client	Post-Action
1	13	12	17	18	30	34
2	12	19	18	6	18	30
3	17	19	24	21	31	32
4	12	25	25	18	39	40
5	19	27	19	18	28	27
6	6	12	6	6	18	23
7	17	18	30	24	36	38
8	18	29	36	22	36	40
9	23	30	24	18	38	32
10	18	12	24	24	25	34

You can copy-paste the table into Microsoft Excel, save as an .xlsx or .csv file, and import into jamovi to follow-along.

In jamovi, select Repeated Measures ANOVA under the ANOVA analysis option. Here are the general steps:

1. In Repeated Measures Factors, you'll need to name both factors. Rename `RM Factor 1` to Time and `RM Factor 2` to Area. Under Time, specify two levels: Pre and Post. Under Area, specify three levels: Product, Client, and Action.
2. In Repeated Measures Cells, you'll now have six cells with the combina-

Now let's go over selected output. First, we need to check our assumption of sphericity. All the Mauchly's W's are not statistically significant so we satisfy the assumption of sphericity and do not need to apply any sphericity corrections.

Assumptions

Tests of Sphericity

	Mauchly's W	p	Greenhouse-Geisser ε	Huynh-Feldt ε
Area	0.97	0.878	0.97	1.00
Time	1.00	NaN [a]	1.00	1.00
Area ✻ Time	0.62	0.152	0.73	0.83

[a] The repeated measures has only two levels. The assumption of sphericity is always met when the repeated measures has only two levels

Figure 6.5: Assumption testing in jamovi

Next let's look at the within subjects effects table. Remember, we do not need to worry about the between subjects effects table because we do not have one; it will be used in the mixed factorial design below. Overall, we see a significant main effect of area, a significant main effect of time, and a significant interaction effect of both area and time. Neat!

Repeated Measures ANOVA

Within Subjects Effects

	Sum of Squares	df	Mean Square	F	p	η^2_G
Area	1365.23	2	682.62	26.96	< .001	0.36
Residual	455.77	18	25.32			
Time	828.82	1	828.82	33.85	< .001	0.25
Residual	220.35	9	24.48			
Area ✻ Time	224.43	2	112.22	12.63	< .001	0.08
Residual	159.90	18	8.88			

Note. Type 3 Sums of Squares

Figure 6.6: Repeated measures ANOVA in jamovi

Next, we can look at post hoc comparisons because the main effects were all statistically significant. For area, we can see that client and action had significantly higher means than product, but there was no difference between client and action. Furthermore, post-intervention performance was significantly higher than pre-intervention.

Post Hoc Tests

Post Hoc Comparisons - Area

Comparison							
Area		Area	Mean Difference	SE	df	t	p_{tukey}
Product	-	Client	−8.60	1.59	18.00	−5.40	< .001
	-	Action	−11.15	1.59	18.00	−7.01	< .001
Client	-	Action	−2.55	1.59	18.00	−1.60	0.270

Post Hoc Comparisons - Time

Comparison							
Time		Time	Mean Difference	SE	df	t	p_{tukey}
Pre	-	Post	−7.43	1.28	9.00	−5.82	< .001

Figure 6.7: Post hoc tests in jamovi

Last, let's look at the interaction to get a sense of what the interaction looks like. It appears that although there are no differences between pre- and post-intervention for product, there are significant differences from pre- to post-intervention for both client and action. To be more specific on where the statistically significant differences are, you can also ask for post hoc tests for the interaction term. This is where including a plot can be very helpful for your audience!

Now we have everything we need (in addition to the estimated marginal means) and can write-up our results.

> We tested a 2 (time: pre- and post-intervention) x 3 (area: product, client, action) repeated measures factorial design to examine how both time and area affected performance. We satisfied the assumption of sphericity for all effects. There was a significant main effect of time (F [1, 9] = 33.85, $p < .001$, $\eta^2_G = .25$) such that performance at post-intervention ($M = 26.80$, $SE = 1.84$) was higher than at pre-intervention ($M = 19.37$, $SE = 1.84$). There was also a significant main effect of area (F [2, 18] = 26.96, $p < .001$, $\eta^2_G = .36$) such that both client ($M = 25.10$, $SE = 1.95$) and action ($M = 27.65$, $SE = 19.5$) performance was higher than product performance ($M = 16.50$, $SE = 1.95$), but there was no difference between client and action performance. Lastly, there was a significant interaction between time and area such that there were no differences in product

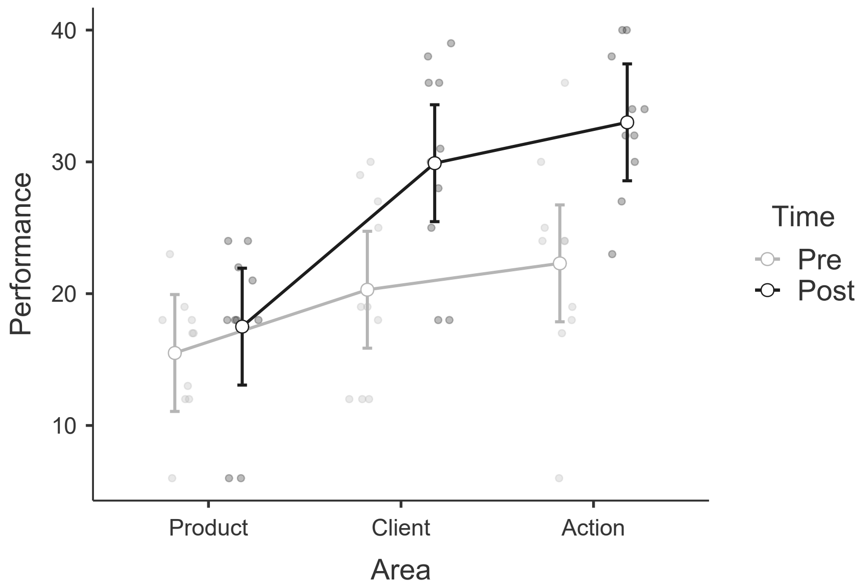


Figure 6.8: Interaction plot in jamovi

performance from pre- to post-intervention but there was for client and action performance (see Figure 1).

6.4 Mixed Factorial ANOVA

Under construction.

Appendix A

References