# Complete Example (Mixed Factorials)

New to this section:

* With mixed designs, you will get both Levene’s and Mauchly’s tests because you have both repeated measures and between subjects!
* Main effects: interpreting each IV on its own, ignoring the effect of the other IVs.
* Interaction: interpreting the IVs together, seeing if the conditions are significantly different OR if the pattern of data across levels is different for the other IV.

Chart of ANOVA Analysis:

|  |  |  |  |
| --- | --- | --- | --- |
|  | ANOVA | | |
|  | Main Effect Between | Main Effect Repeated Measures | Interaction |
| If levels > 2  And significant | Independent t-test  Bonferroni/Tukey correction | Dependent t-test  Bonferroni/Tukey correction | SPLIT one IV column  Independent t-test OR  Dependent t-test  Bonferroni correction |
| If levels = 2 | Interpret means | Interpret means |

If the interaction is significant, often people ignore any analyses with the main effects:

* This procedure reduces Type 1 error because you are running less post hoc tests.
* You are interested in the interaction anyway, so why only interpret one variable at a time?
* Also, be sure to follow up with the correct test type – do not do dependent t on the between subjects factor.

We knew that this high BSG thing had increased ratings, which you are supposed to ignore as part of the instructions. This result is tied to people’s overestimation of how well they think they know something, which is bad for studying. So, we gave people instructions on how to ignore the BSG. Did it help?

**Datafile:** rm 2 anova.csv

**IVs:**

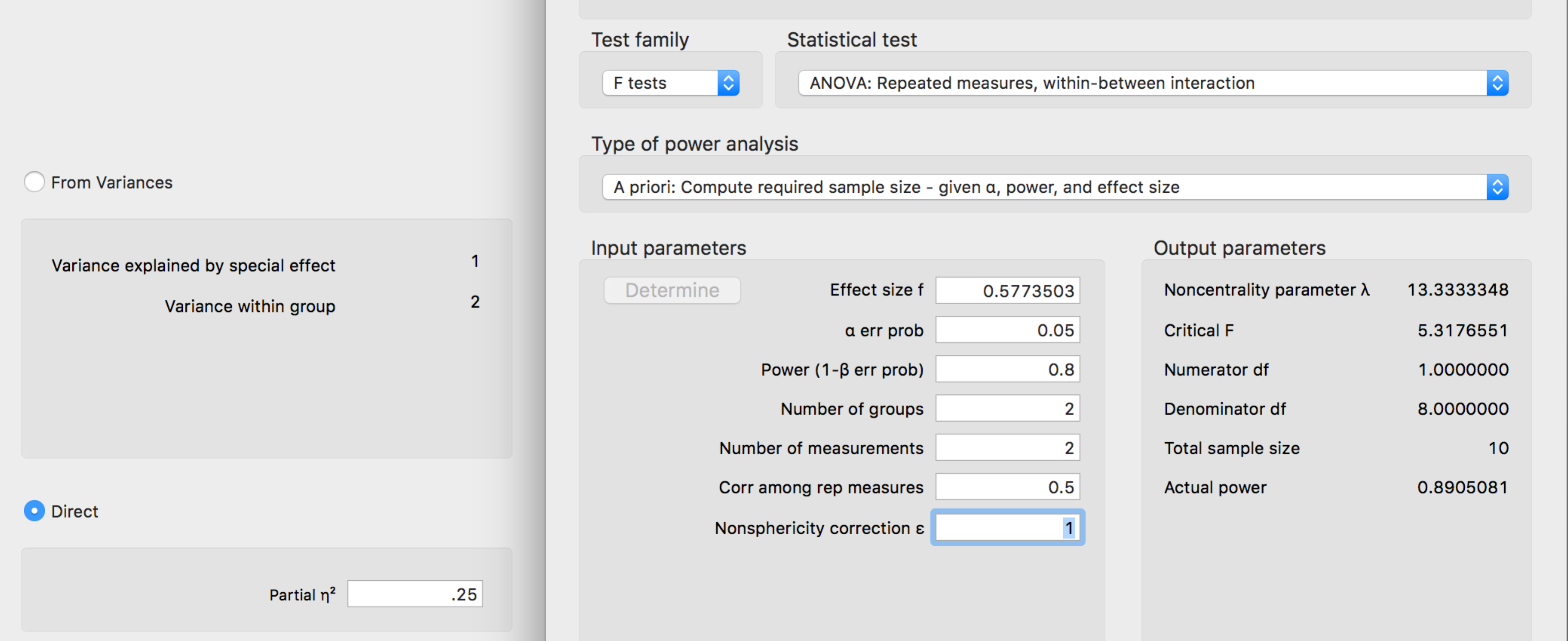
* **Group**:
  + JAM (regular judgment) versus Debias (instructions on how to do well in experiment)
  + This variable is between-subjects – they only got one set of instructions.
* **BSG:**
  + BSG low versus BSG high
  + This variable is repeated measures – they got both types of word pairs.

**DV:**

* Ratings of those word pairs, ranging from 0 – 100.

**Power:**

1. Open Gpower!
   1. Test family: F-test
   2. Statistical Test: Repeated measures, within-between interaction
      1. Usually you are looking for the interaction, so you’ll use this page to estimate the number of people needed for that test.
   3. Estimate an effect size: click determine 🡪 click direct 🡪 use eta square sizes you think might be accurate, remember small, medium, and large estimates from the notes.
   4. Alpha = .05
   5. Power (1-beta .20) = .80
   6. Number of groups = number of levels of between subjects
   7. Number of measurements = number of levels of repeated measures
   8. Corr among rep measures = correlation between levels or conditions
      1. You can estimate from previous research.
      2. Look at the correlations in a pilot study, go with the lowest one you find.
      3. .50-.70 is a good estimate if you are giving them the same test a couple times.
   9. Nonsphericity correction = epsilon
      1. You will not really know this number before you start a study. More useful if you have some participants to estimate from (see below on how to get that number).
2. Let’s estimate the following:
   1. Large effect size (eta = .25)
   2. Number of groups = 2 for two sets of instructions in the between subjects.
   3. Number of measurements = 2 for two sets of BSG in the repeated measures.
   4. Correlation = estimate at .50.
   5. Epsilon = 1.
3. Says we needed to run 10 people to find a significant effect with a large effect size.



**Results**

Participants were given pairs of words and asked to rate them on how often they thought 100 people would give the second word if shown the first word. The strength of the word pairs was manipulated through the strength of the reverse rating (backward strength: BSG) to show its effect on ratings. One group of participants was given the normal judgment task, while another group of participants was given special instructions that should lower the influence of BSG on their scores. Data was screened for assumptions and outliers. One outlier was found using *z*-scores distance as a criterion and was excluded in the analysis. Levene’s test was not significant for Low BSG, *F*(1, 156) = 3.02, *p* = .08, but was significant for High BSG, *F*(1, 156) = 10.83, *p* = .001.

A significant main effect of backward strength was shown, *F*(1, 156) = 258.34, *p* < .001, *η p²* = .62. The low BSG word pairs (*M =* 64.56, *SD* = 12.19) were rated lower than the high BSG word pairs (*M* = 78.19, *SD* = 11.34). Group participation also affected ratings, *F*(1, 156) = 34.50, *p* < .001, *η p²* = .18, in which the debiasing group (*M =* 66.65, *SD* = 13.76) significantly lowered their ratings compared to the control group (*M* = 75.49, *SD =* 12.04). However, there was not an interaction between ratings and group participation, *F*(1, 156)= 1.28, *p* = .26**,** *η p²* = .008. See Figure 1 for group means.

\*\*\*if there was a significant effect, here’s how it might go\*\*\*

Each group’s ratings on low and high backward strength words were examined with a dependent t-test and Bonferroni correction to see if the influence of BSG could be reduced. The normal judgment group showed a 14 point difference between low and high word pairs, *t*(156) = -4.46, *p* < .001, *ddiff* = -0.71. The debiasing group saw a smaller difference in ratings of about 12.5 points, *t*(156) = -6.01, *p*<.001, *ddiff* = -0.96.

(I got the means from the descriptives analysis for main effects, the F values and interaction are more important).

*Figure 1*. Means and one standard deviation error bars for each condition in the experiment.

|  |  |  |  |
| --- | --- | --- | --- |
|  | JAM | Debias |  |
| BSG Low | Low, Jam | Low, Debias | Independent t |
| BSG High | High, Jam | High, Debias | Independent t |
|  | Dependent t | Dependent t |  |

Whichever variable you split, the other variable is the one you analyze:

* Split on repeated measures variable, you are analyzing the between subjects variable (independent t)
* Split on the between subjects variable, you are analyzing the repeated measures variable (dependent t)

Which way to go (across or down)?

* Go with the hypothesis
* Split on the larger number of levels, which creates less post hoc tests
  + Square design, that doesn’t matter (2X2, 3X3)

If you are using JASP, and you have more than two levels of the between subjects variable:

* Need to split up the excel files into separate comparisons
* 4 levels of my between subjects variables example:
  + Split all four levels into separate files, then analyzed dependent t on the repeated measures variable (easiest)
  + Split into pairwise comparisons and analyze each repeated measures level with independent t
    - One excel file with levels 1 and 2,
    - One excel file with levels 1 and 3,
    - One excel file with levels 1 and 4 …

**Results**

**Descriptives**

| **Descriptive Statistics** | | | | | |
| --- | --- | --- | --- | --- | --- |
|  | | **bsglo** | | **bsghi** | |
| **Valid** |  | 158 |  | 158 |  |
| **Missing** |  | 0 |  | 0 |  |
| **Mean** |  | 64.34 |  | 77.90 |  |
| **Std. Deviation** |  | 12.48 |  | 11.86 |  |
| **Minimum** |  | 28.25 |  | 33.25 |  |
| **Maximum** |  | 88.67 |  | 93.67 |  |
|  | | | | | |

**Repeated Measures ANOVA**

| **Within Subjects Effects** | | | | | | | | | | | | | | | |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | | **Sum of Squares** | | **df** | | **Mean Square** | | **F** | | **p** | | **η²** | | **η² p** | |
| BSG |  | 14336.08 |  | 1 |  | 14336.08 |  | 258.335 |  | < .001 |  | 0.622 |  | 0.623 |  |
| BSG ✻ group |  | 71.08 |  | 1 |  | 71.08 |  | 1.281 |  | 0.259 |  | 0.003 |  | 0.008 |  |
| Residual |  | 8657.09 |  | 156 |  | 55.49 |  |  |  |  |  |  |  |  |  |
|  | | | | | | | | | | | | | | | |
| *Note.*  Type III Sum of Squares | | | | | | | | | | | | | | | |

| **Between Subjects Effects** | | | | | | | | | | | | | | | |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | | **Sum of Squares** | | **df** | | **Mean Square** | | **F** | | **p** | | **η²** | | **η² p** | |
| group |  | 6842 |  | 1 |  | 6842.5 |  | 34.50 |  | < .001 |  | 0.181 |  | 0.181 |  |
| Residual |  | 30936 |  | 156 |  | 198.3 |  |  |  |  |  |  |  |  |  |
|  | | | | | | | | | | | | | | | |
| *Note.*  Type III Sum of Squares | | | | | | | | | | | | | | | |

Main effects:

* BSG: *F*(1, 156) = 258.34, *p* < .001, *η p²* = .62
* Group: *F*(1, 156) = 34.50, *p* < .001, *η p²* = .18

Interaction:

* *F*(1, 156)= 1.28, *p* = .26**,** *η p²* = .008

**Assumption Checks**

| **Test of Sphericity** | | | | | | | | | |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | | **Mauchly's W** | | **p** | | **Greenhouse-Geisser ε** | | **Huynh-Feldt ε** | |
| BSG |  | 1.000 | ᵃ | NaN | ᵃ | 1.000 | ᵃ | 1.000 | ᵃ |
|  | | | | | | | | | |
| ᵃ The repeated measure has only two levels. When the repeated measure has two levels, the assumption of sphericity is always met. | | | | | | | | | |

| **Test for Equality of Variances (Levene's)** | | | | | | | | | |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | | **F** | | **df1** | | **df2** | | **p** | |
| bsglo |  | 3.018 |  | 1 |  | 156 |  | 0.084 |  |
| bsghi |  | 10.828 |  | 1 |  | 156 |  | 0.001 |  |
|  | | | | | | | | | |

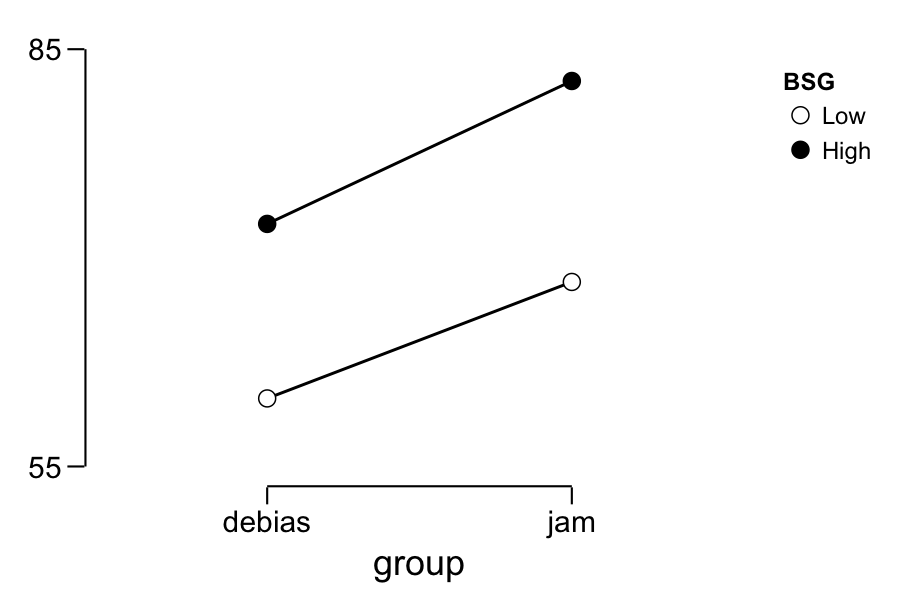
**Levene’s test:**

* **Low, *F*(1, 156) = 3.02, *p* = .08, not significant**
* **High, *F*(1, 156) = 10.83, *p* = .001, significant**

**Descriptives**

| **Descriptives** | | | | | | | | | |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **BSG** | | **group** | | **Mean** | | **SD** | | **N** | |
| Low |  | debias |  | 59.89 |  | 12.779 |  | 74 |  |
|  |  | jam |  | 68.27 |  | 10.841 |  | 84 |  |
| High |  | debias |  | 72.44 |  | 12.932 |  | 74 |  |
|  |  | jam |  | 82.72 |  | 8.292 |  | 84 |  |
|  | | | | | | | | | |

**Descriptives Plot**



**T-Test**

| **Independent Samples T-Test** | | | | | | | | | |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | | **t** | | **df** | | **p** | | **Cohen's d** | |
| bsglo |  | -4.456 |  | 156.0 |  | < .001 |  | -0.710 |  |
| bsghi |  | -6.014 |  | 156.0 |  | < .001 | ᵃ | -0.959 |  |
|  | | | | | | | | | |
| *Note.*  Student's T-Test. | | | | | | | | | |
| ᵃ Levene's test is significant (p < .05), suggesting a violation of the equal variance assumption | | | | | | | | | |

**Descriptives**

| **Group Descriptives** | | | | | | | | | | | |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | | **Group** | | **N** | | **Mean** | | **SD** | | **SE** | |
| bsglo |  | debias |  | 74 |  | 59.89 |  | 12.779 |  | 1.486 |  |
|  |  | jam |  | 84 |  | 68.27 |  | 10.841 |  | 1.183 |  |
| bsghi |  | debias |  | 74 |  | 72.44 |  | 12.932 |  | 1.503 |  |
|  |  | jam |  | 84 |  | 82.72 |  | 8.292 |  | 0.905 |  |
|  | | | | | | | | | | | |

T-test output has no correction on it.

Bonferroni correction:

* Alpha (remember this is type one error, we set this value, p < .05) / number of comparisons that we ran
* .05 / 2 = .025 new alpha, compare that value to p found in the experiment

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| IV 1 | Group 1 | Group 2 | Test | Effect Size | Interpretation |
| BSG Low | Debias  M = 59.89  SD = 12.78  N = 74 | JAM  M = 68.27  SD = 10.84  N = 84 | *t*(156) = -4.46, *p* < .001 | *d* = -0.71 | Significant (because p found is less than alpha) |
| BSG High | Debias  M = 72.44  SD = 12.93  N = 74 | JAM  M = 82.72  SD = 8.29  N = 84 | *t*(156) = -6.01, *p* < .001 | *d* = -0.96 | Significant |

While there are new options for effect size in JASP, those work for one-way tests or main effects only. You could still calculate Cohen’s d as described here.