ANOVA

**Uses:**

One or more independent variables with different levels/conditions (groups or measurements) analyzing differences in means on a continuous dependent variable.

**Description:**

ANOVA stands for the analysis of variance. You can use t-tests to analyze two groups or two time measurements from the same people. However, when you use more than two groups or times, you increase the chance of a Type 1 error (rejecting when you shouldn’t) for each t-test you have to run. Therefore, you use an ANOVA to cover the chance of that error. However, ANOVAs have the disadvantage of involving extra steps or *post hoc* tests.

**Definitions/Abbreviations:**

IV – independent variable. This variable *has* to be a categorical/nominal variable. You can put people into groups based on any category (gender, handedness) or your experimental manipulation (instructions versus no instructions).

DV – dependent variable. The dependent variable *needs* to be a continuous variable or another type of analysis might work better (see log regression). Your dependent variable should be the measurement you took in your study or what information you are expecting to see changed over groups.

Levels versus conditions: Levels are the different groups within an IV. Conditions are the combinations of different levels. Example:

* IV: Gender
  + Levels: Male, Female
  + Conditions would be the same because you only have one variable.
* IV: Gender, Handedness
  + Levels: Male, Female, Left, Right
  + Conditions: Male Left, Male Right, Female Left, Female Right

Between subjects / independent: ANOVA for independent measures where participants are in only one level/condition.

Repeated measures / dependent / within subjects: ANOVA for dependent measures where participants are in all levels/conditions.

Mixed designs: ANOVA for when you have two variables, one between subjects and one repeated measures.

One-*way* versus two-*way*: The word “way” just means number of independent variables. So a one-way ANOVA has one IV, while a two-way has two IVs.

# X # (i.e. 2X4, 2X3) ANOVA: designates the number of levels for the multiple IVs in a two+-way design. If you say a 2X4 between subjects ANOVA, you mean:

* I have two variables.
* These variables are both between subjects.
* One variable has two levels.
* One variable has four levels.
* Together there are eight conditions.

Whereas a 2X2X3 mixed ANOVA would be:

* I have three variables.
* Some are repeated, some are between subjects variables.
* One variable has two levels.
* One variable has two levels.
* One variable has three levels.
* There are twelve conditions.

**Hypothesis testing:**

ANOVA: First you run an ANOVA to determine the following:  
Null hypothesis – level/condition means are not different.

Research hypothesis – level/conditions means are different.

Notice, this analysis does not tell you *where* group differences occur, just that groups are different. If your overall test is significant, you then run a post hoc test (different types are discussed below) to determine how groups are different.

Post Hoc Tests: depending on the type of test, you will have to run several tests to determine how levels/conditions are different, usually by comparing level/condition means *pairwise* (two at a time).

Null hypothesis – level 1 mean is equal to level 2 mean.

Research hypothesis – level 1 mean is not equal to level 2 mean.

Types of post hocs:

* *A priori* – or planned comparisons: means that you knew before you started the experiment what group means you wanted to compare.
* *A posteriori* – all comparisons: after the fact you decided to do these comparisons or you decided to do all pairwise combinations.
* We are mostly going to discuss *posteriori* tests because they are the most common.

Post hoc TEST: The actual statistical test you use to determine group differences.

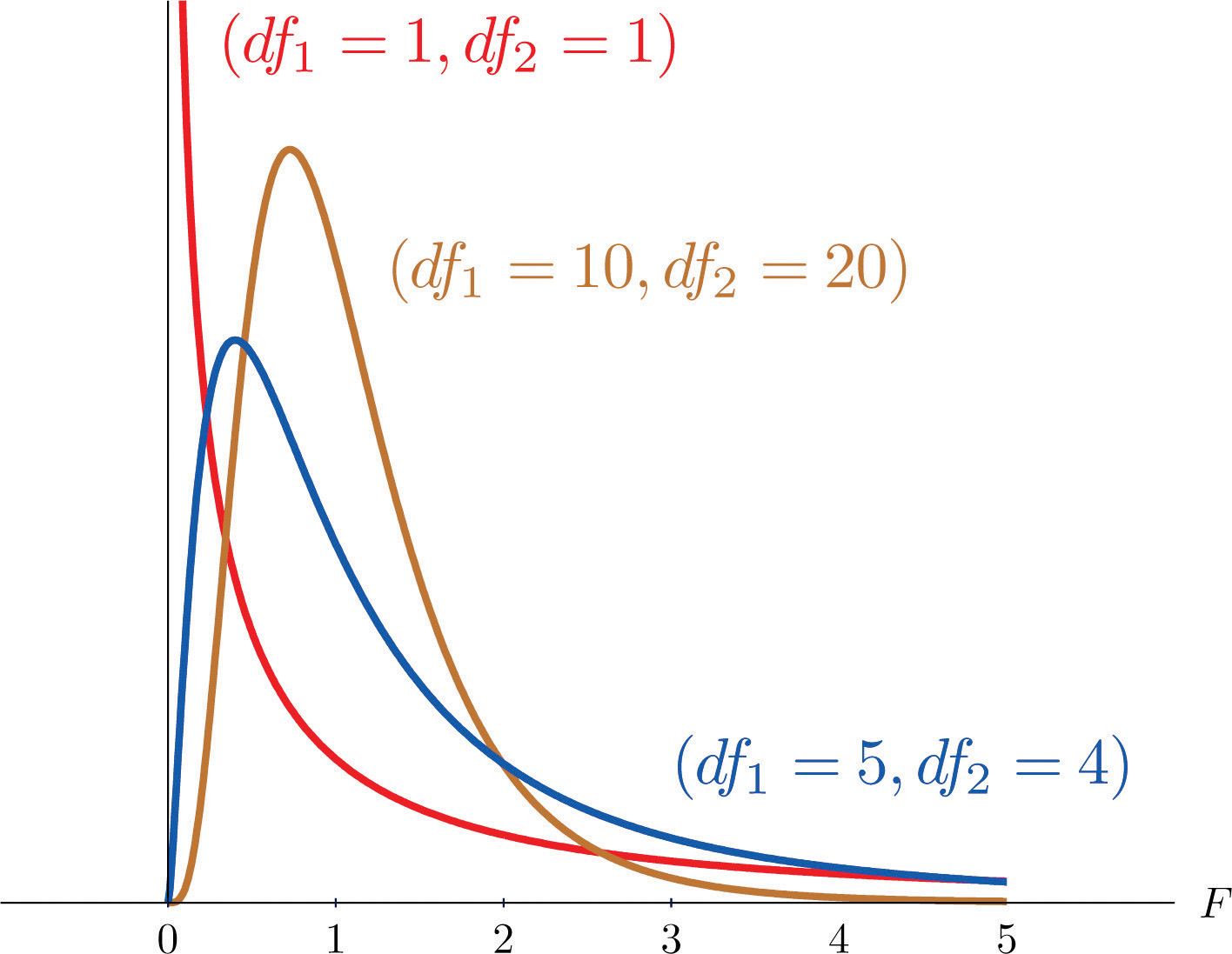
* Must match the type of analysis you are using.
* Between subjects = independent t-test.
* Repeated measures = dependent t-test.

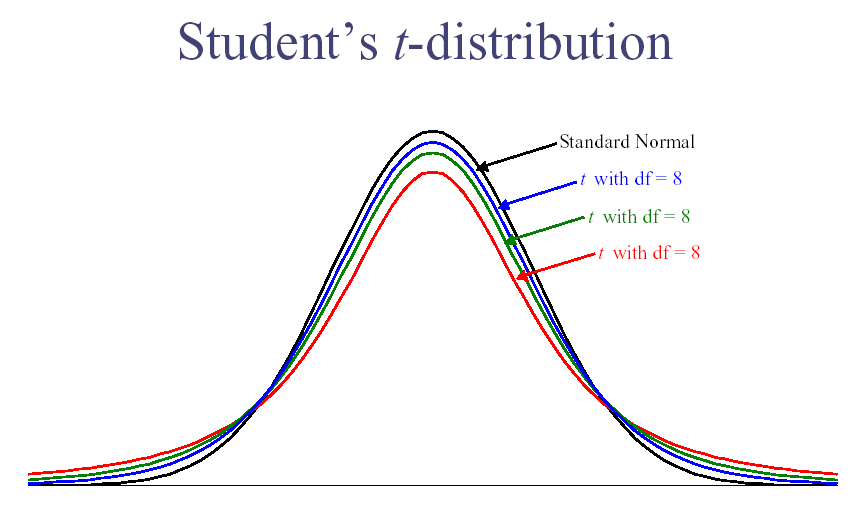
Post hoc CORRECTION:

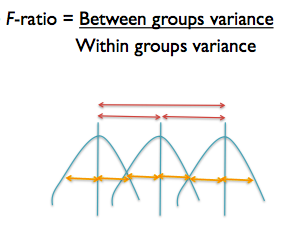
* You can still run into the problem of increasing Type 1 error rate if you run all pairwise combinations (or just have a lot of tests). We will discuss the most common solutions for Type 1 error control, given the type of test you are running.
* What do the corrections do? Control type 1 error rate by making it harder for you to find a significant difference (i.e. move the cut off score further out). However, that’s not obvious on the output, because it automatically recalculates the *p*-values based on new critical/cut off scores.
* Between subjects: Tukey.
* Repeated measures: Bonferroni.
* There are many types of corrections – these are just the most common.

***F*-ratios:**

In this chapter, we introduce a new statistical distribution and letter, *F*. We discussed previously the *z* and *t* distributions, which are related to the *F* distribution. The *F* distribution is the *t* distribution squared, so there are no negative cut off scores or *F*-values. With the *t* distribution, we had one degree of freedom (*df*) value to use to look up cut off scores or report for our results sections. With *F,* we have two *df* value to report – one for the numerator (between subjects/IV/model) and one for the denominator (within subjects/error).







*Note*. The blue lines are distributions of each group and their mean.

Between groups variance / model variance / main effects / IV variance – This variance is the noted in red on the picture. You are measuring an average distance between level/condition means and the grand mean. You want this score to be LARGE because it is the effect that your experiment had on the means. Small numbers indicate that the means are not very different (boo!), while large values indicate that at least one of the means is far apart from the rest.

Within groups variance / error variance – This variance is the error term in your experiment, noted in yellow on the picture. This score is basically the average of group variances. You want this score to be SMALL, which means that you measured your groups with precision and their scores do not vary widely. Large numbers indicate a lot of error or that levels/conditions have a wide range of scores. Small numbers indicate that you have scores that tightly cluster around their means. Sometimes, this value is called individual differences.

There are *df* values that match both of these values (hence where the numerator/denominator thing comes from.

Between groups *df* (numerator) = Number of levels/conditions – 1 (number of means for that variable)

Within groups *df* (denominator) = (Number of participants – 1) + (Number of participants – 1) + (Number of participants – 1) …

These values get more complex as we add additional independent variables, but those formulas are the basic gist. We are still using the idea of N-1 but changing N to denote the number of means or people used to calculate the between or within group variances.

**Effect size:**

* η2 is effect size (eta squared) – the proportion of variance accounted for by the between groups variance (you want this to be big).
  + Numbers range from 0 – 1.
  + Small = .01
  + Medium = .09
  + Large = .25
  + Basically, how much variance is due to the IV and not the error variance.
* You can ask SPSS to give this to you automatically (see below).
* You can also report R2 or ω2 (omega).
  + R is the same as eta squared.
  + R is more common for regression.
  + Omega coming soon to MOTE, but the interpretation is the same (the math adjusts the estimate to be less biased and smaller to better estimate the population effect size).
* *ηp*2  is effect size (partial eta squared)-ratio of variance accounted for by an effect and that effect plus its error variance.

**Power:**

You want to check how many people you need to run (or alternatively, how many more people you need in a study).

Using G\*Power, finding ideal participant numbers is fairly easy. Set up options:

* Test family: F-tests
* Statistical Test: ANOVA:
  + Between subjects only: fixed effects omnibus, one-way
  + Repeated measures only: repeated measures, within factors
  + Factorial:
    - Between only: fixed effects, special, main effects, and interactions
    - Repeated only: repeated measures, within factors
    - Mixed factorials: repeated measures, within-between interaction
* Type of power analysis: A priori (most common)
* Effect size f: either guess at an effect size based on research, use a small effect size for good measure, or after a couple subjects run a prelim ANOVA and use the current effect size. (You can click determine to convert eta squared to f).
* Alpha = .05
* Power = .80
* The rest of the settings are discussed with their particular analysis.
* Hit calculate for the number of participants needed.

**Data screening:**

Accuracy and missing data are the same as the data screening discussion. However, you have to think about filling in data for missing data because usually you only have one number to fill in (the DV for between subjects; doesn't apply to repeated measures).

Outliers:

* Univariate – outliers only on the DV. You want to check z-scores for people who are more than 3 (3 or -3) away from the mean.
* Multivariate – outliers on combination of the DV and the time measurement (for factorial or repeated measures only). This procedure uses Mahalanobis distance to make sure that people do not have a strange combination of answers on the DVs.
  + You can do both of these or just Mahalanobis. You may not want to eliminate people who are a univariate outlier on one variable, but you really should eliminate people who are multivariate outlier.

Normality:

* Univariate – you want the DV to be normally distributed. You can check this information through frequencies and asking for a histogram. Non-normal distributions also have skew and kurtosis values over 3/-3.
* Multivariate – you also want the DV/time measurement combination to be normally distributed. You can check for multivariate normality by running a fake regression and asking for a histogram of the residuals.

Linearity: Linearity between the IV and DV is a factor, although you would expect differences. You can check for this value using a fake regression or bivariate scatterplot.

Homogeneity: the variance of the levels/conditions from your IV need to be equal across the DV. You can check this information with a residual plot from your fake regression (you do not want raining or an unequal spread of the dots around 0). You can also use Levene’s test of homogeneity – you ***do not*** want p<.001.

Specific repeated measures issues are discussed in that section.

# Complete Example (Between Subjects – 1 IV only)

A health psychologist recorded the number of close inter-personal attachments of 45-year-olds who were in excellent, fair, or poor health. People in the Excellent Health group had 4, 3, 2, and 3 close attachments; people in the Fair Health group had 3, 5, and 8 close attachments; and people in the Poor Health group had 3, 1, 0, and 2 close attachments. (bn 1 anova.sav data)

**IVs:**

Health group – excellent, fair, poor

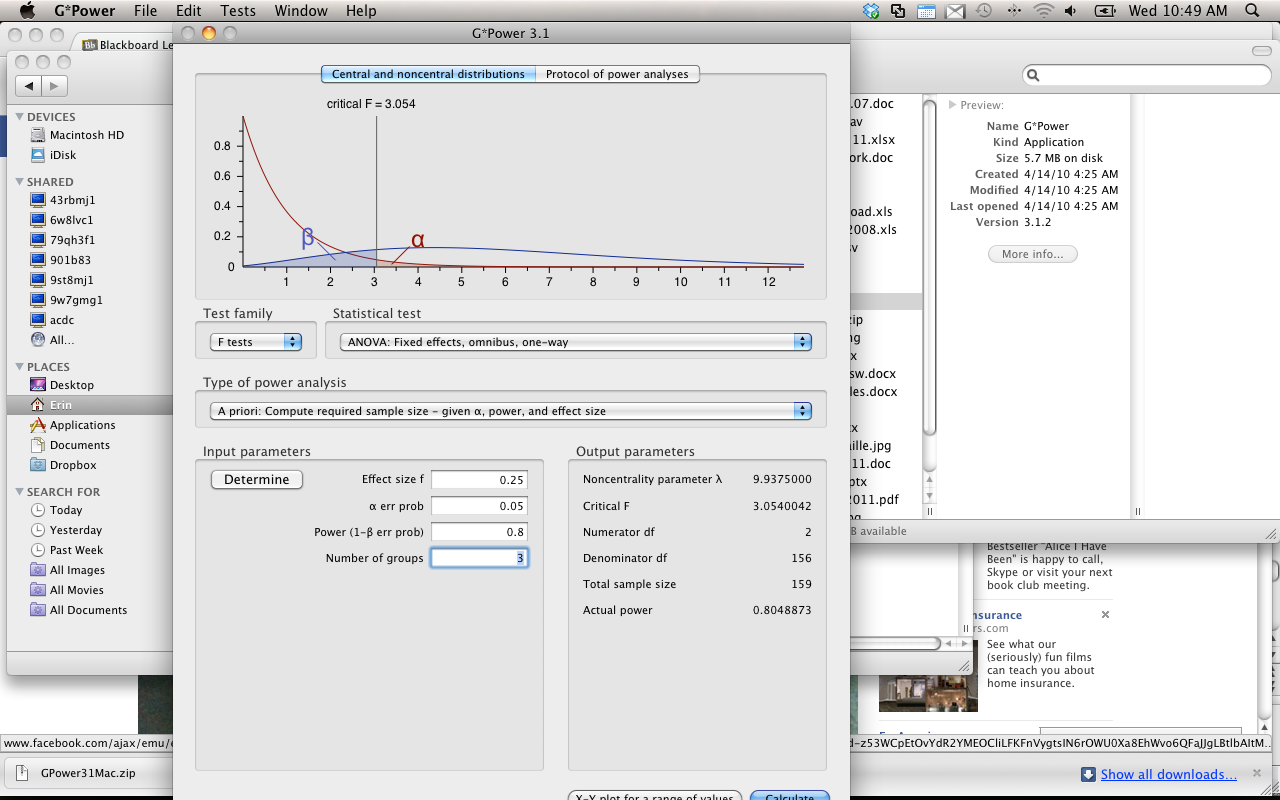
**DV:**

Friends - number of close inter-personal attachments

Remember that the data has to be set up to have one column for the IV (value labels!) and one column for the DV.

**Power:**

1. Open Gpower!
   1. Test family: F test
   2. Statistical test: ANOVA: fixed one way
   3. Pick an effect size (or calculate!)\*\*
   4. Alpha normally .05
   5. Beta normally .80
   6. Groups – number of levels.
   7. Hit ok!
2. Says we needed to run 159 people to find a significant effect with a medium effect size.

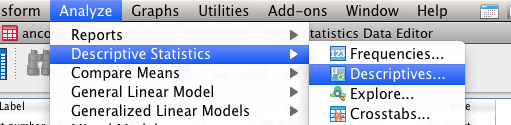
****

**Assumptions Checks:**

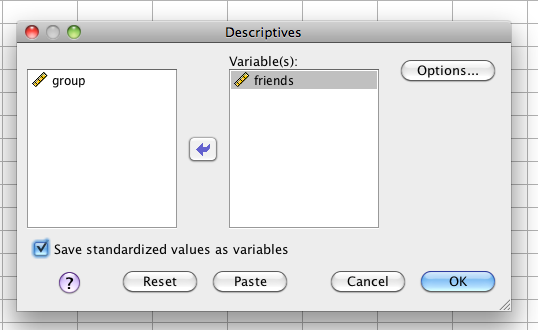
Accuracy and missing data are the same as data screening.

Univariate outliers:

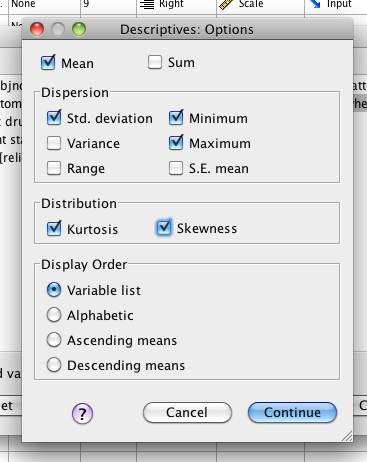
1. Analyze > Descriptive Statistics > Descriptives



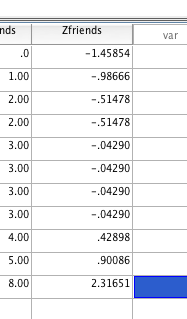
1. Move **just the DV** into the descriptives box.
2. Hit save standardized values as variables.



1. Hit options > skew and kurtosis.



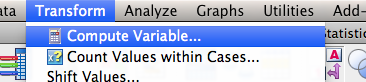
1. Look for univariate outliers.
2. Go back to the data set. You should see the z-scores for the DV.
3. Sort the variables one at a time – look for values over 3 and -3 (right click sort).



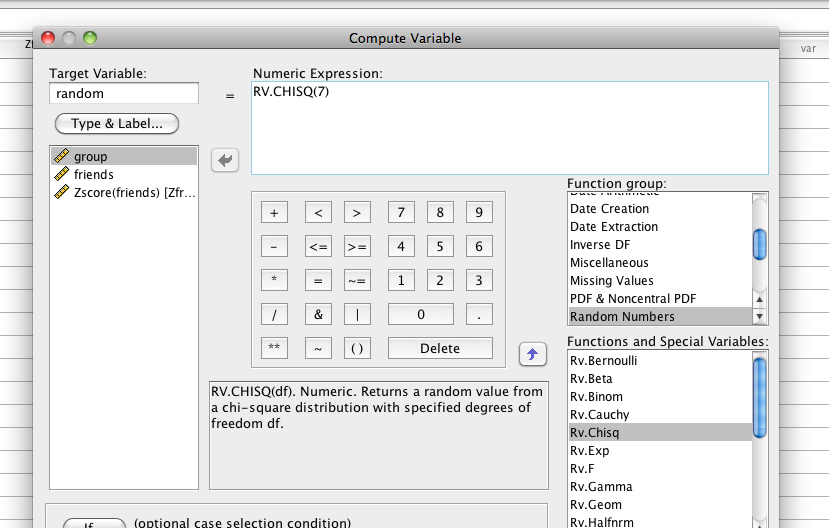
1. We do not appear to have univariate outliers.

Normality, Linearity, Homogeneity (Multicollinearity does not apply because it is one dependent variable):

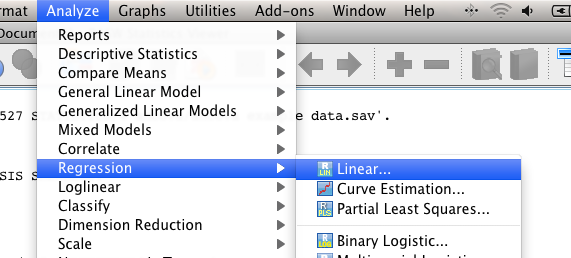
1. Make a fake variable.
   1. Transform > compute variable



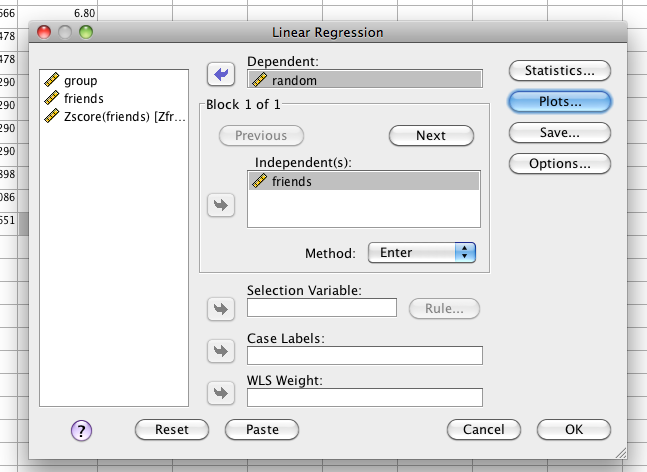
* 1. Label the random variable, then use the chi square random functions to create a random variable.



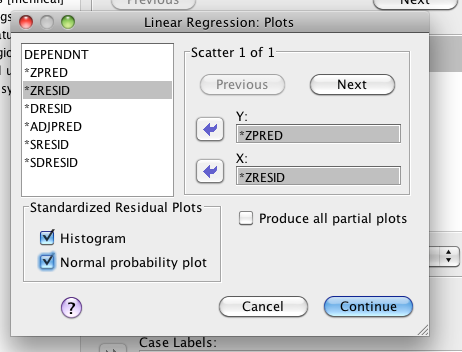
1. Run a fake regression. Analyze > Regression > Linear.



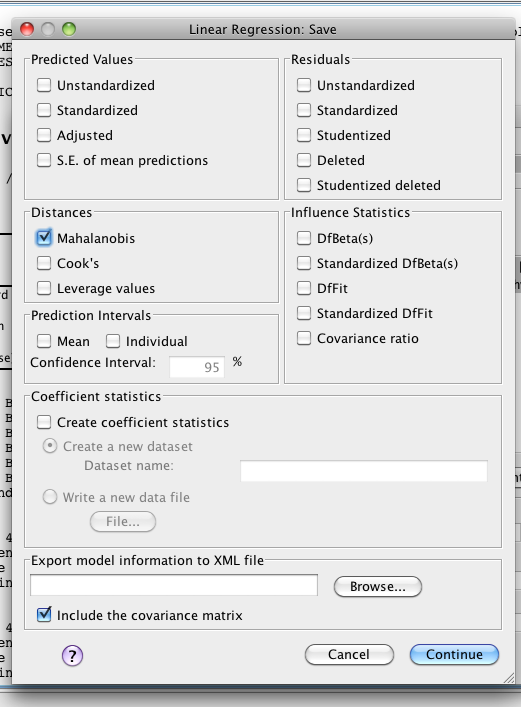
1. Move your random variable into the dependent box. Put your DV into the independent(s) box.



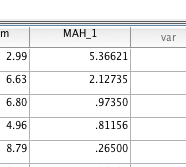
1. Hit Plots.



1. Put zpred in Y and zresid in X. This section will create the residual plots for other assumptions checks. Check histogram and normal probability plot.
2. Hit Save.



1. Check Mahalanobis – it will create another new variable for you to check.
2. Go back to the data set, and sort cases descending for Mahalanobis.
   1. You will need a cut off value to understand these scores.
   2. The cut off value is chi-square with degrees of freedom = number of variables, p<.001.
   3. Here we have one variable (DV), p<.001, chi square = 10.83 cut off score.
   4. You will look for people who have a Mahalanobis score greater than > 10.83



1. We do not have multivariate outliers.

Linearity:

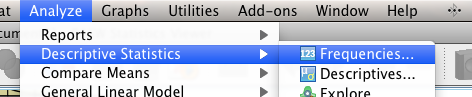
1. You can check linearity with the fake regression you’ve already run.
   1. IF YOU DELETED OUTLIERS: run the regression again, just as it describes above.
   2. IF YOU DID NOT DELETE OUTLIERS: you can use the output you already have.
2. Scroll down to the Normality Probability Plot.



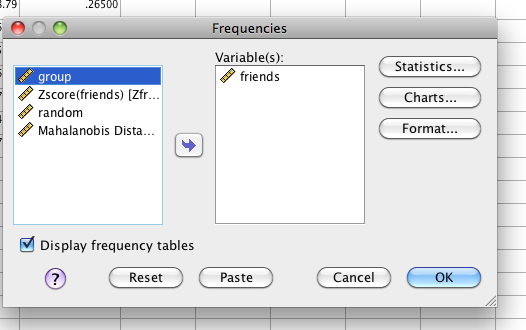
1. You want the dots to follow along the line moderately closely. This picture is starting to not be normal, but is ok.

Normality:

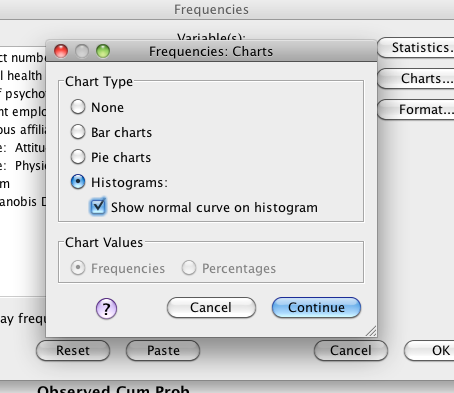
1. Univariate normality:
   1. You will need to get individual histograms of your DV.
   2. Analyze > Descriptive Statistics > Frequencies.



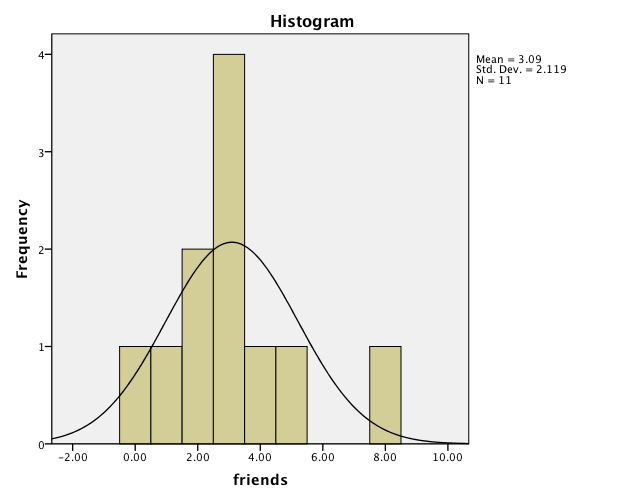
* 1. Move over the DV into the right hand box.



* 1. Hit charts. Pick Histogram and Normal Curve.



* 1. Hit continue and Ok.
  2. Now you’ll want to check the charts to make sure they look ok.

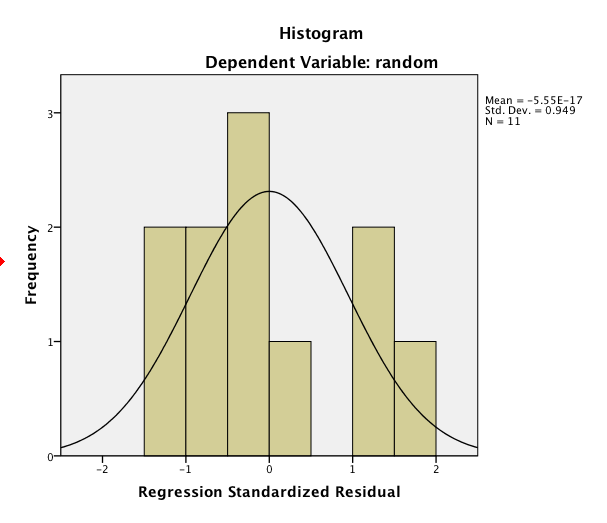


* 1. You can see this chart is fairly normal. You can go back and check skew and kurtosis if you are not sure. (See the top when you ran descriptives).



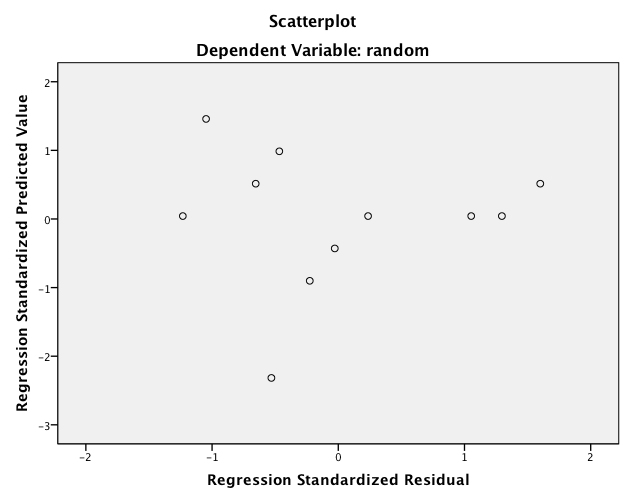
* 1. Both of these are under 3, or -3, so I’m good.

1. Multivariate Normality: use the histogram created from your fake regression to check. This test is actually a moot point because you only have ONE variable, but it doesn’t hurt to make sure the chart looks the same as the one from the univariate test.



* 1. This histogram looks fairly normal.

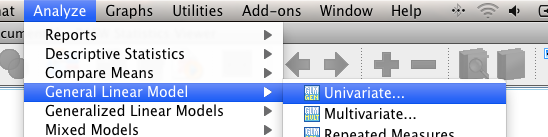
Homogeneity: you can check this assumption using the residual plot from the fake regression you ran earlier.



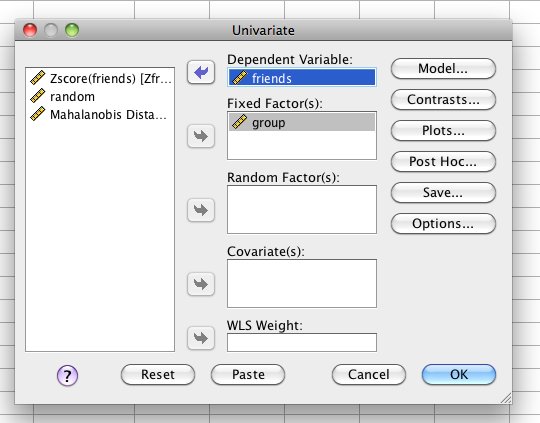
1. What to look for:
   1. Draw a line at 0.
   2. Homogeneity – is the spread above that line the same as below that line?
      1. These lines are complete lines.
      2. So it goes from 0 to 3 above and 0 to -3 below.
      3. You *do not* want a very large spread on one side and a small spread on the other side.
   3. If you encounter this problem – check Levene’s when you run the data.
      1. Levene’s is a test for homogeneity between groups, so it looks to see if the variances are equal across your IV levels.
      2. It is notoriously oversensitive, but can be a good place to start if you want to check a real number, rather than this scatterplot.
      3. With large sample sizes, it is often significant (remembering the big important rule, p<.001), and with large sample sizes it matters less. Ergo, if you have big *n* in each group, then don’t worry about it so much.

**Running the ANOVA:**

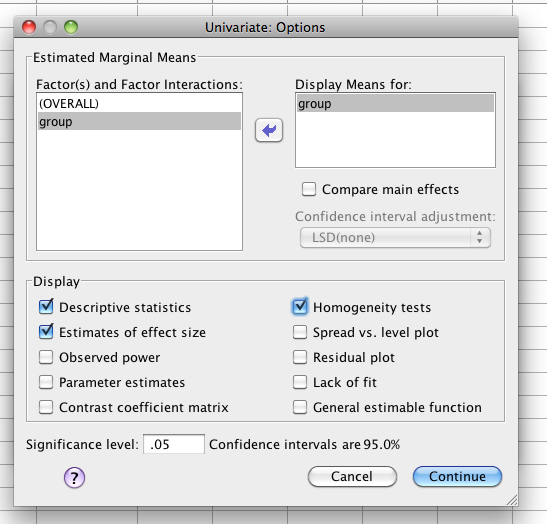
1. Analyze > General Linear Model > Univariate (because we only have between subjects variables).



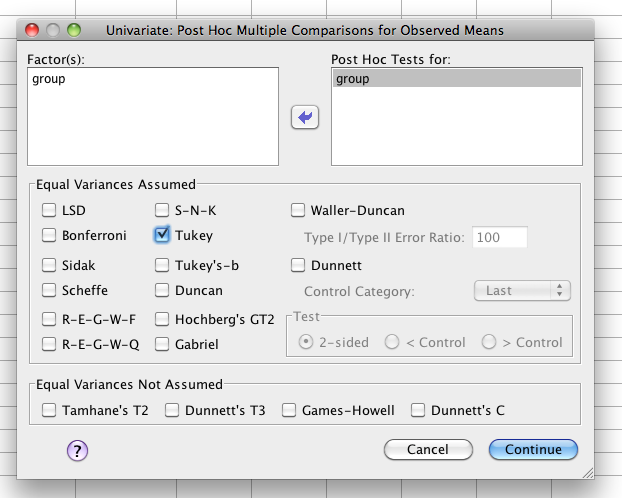
1. Your IV (group variable) goes in the fixed factor box.
2. Your DV (measured variable) goes in the dependent box
   1. Aren’t sure? Ask yourself which one you assigned (group membership) versus which one you took data on (survey, test scores, etc).
   2. AND you can’t put a continuous variable in the fixed factor box. It has to be categorical.



1. Hit options > move over the variables for descriptive statistics. Pick options descriptives, estimates of effect size, and homogeneity (this is Levene’s test).

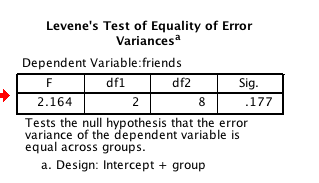


1. Hit post hoc.
2. Move the group variable over into the post hoc box.
3. Choose Tukey/Bonferroni/Sheffe (Tukey is the most common).
   1. Only pick one though!
   2. Since we only have three groups, I’m just going to do Tukey.

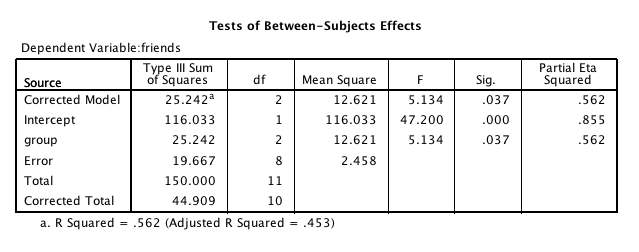
****

**Going from Output to Results Section:**

1. Check Levene’s for Homogeneity – especially if your residual plot was not evenly distributed. You do not want *p* > .001.
2. Levene’s is p=.177 so we are ok.



1. The ANOVA Box:
   1. Use the line labeled “group” or whatever you called your IV (this is why labels are good!)
   2. You are looking for a significant effect, so you want the sig column to be <.05.
   3. Write Up example: A between subjects one-way ANOVA was used to analyze differences in numbers of friends for fair, poor, and excellent health levels. Group means were found to be different, *F*(2,8)=5.13, *p*=.04, η2 = .56.



Effect sizes – to be listed with f-statistic.

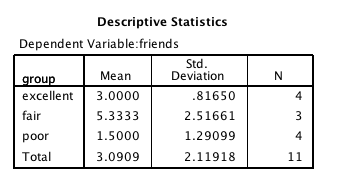
F-values – used to write after the equal signs.

Degrees of Freedom for tests.

* the first number or top box is the first number in the f-statistic.
* The error line or second box is the second number in the f-statistic.

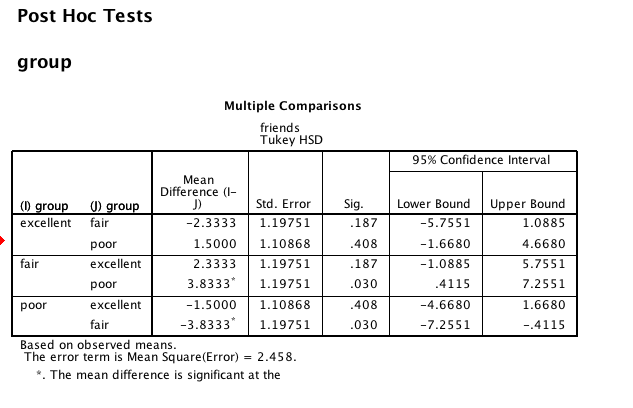
P-values, check these for p<.05

1. IF IT IS SIGNIFICANT Post Hocs.
   1. First, I find it easiest to create a table to figure out my effects. You’ll need the means to get that information.



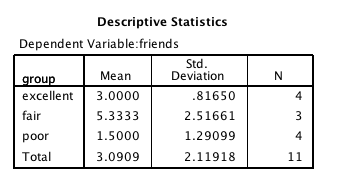
|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Mean 1** | **Mean 2** | **P-value** | **Explain?** | **Cohen’s *d*** |
| Excellent *M* = 3.00, *SD* = .82 | Fair *M* = 5.33, *SD* = 2.52 | .187 | Not significant | d = 1.35 |
| Excellent *M* = 3.00, *SD* = .82 | Poor *M* = 1.50, *SD* = 1.29 | .408 | Not significant | d = 1.39 |
| Fair *M* = 5.33, *SD* = 2.52 | Poor *M* = 1.50, *SD* = 1.29 | .030 | Significant | d = 1.91 |

Use the post hoc box to fill in the p-values.

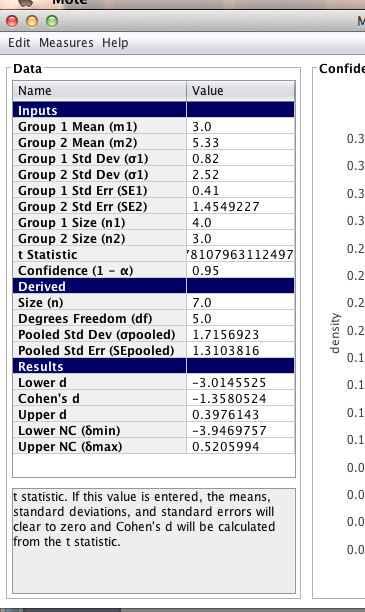


You’ll notice that you do not get any sort of t-test value when using the automatic post hoc correction options. In the background, SPSS calculates an independent t-test for you, and then uses the Tukey correction to calculate *p* values for this box. You will report the *p* values only since you do not have the t-test values.

What about effect size? You can use MOTE to calculate the *d* values for this test (use independent t page) by entering the means and standard deviations – like what I had in my chart. You also want to make sure to enter the right *n* values for each group.



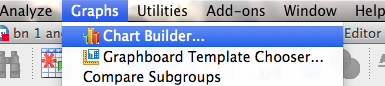
You can see that the Mean, standard deviation (SD), and group size (N) of the Excellent group and fair group was entered into MOTE. The values for the excellent were entered into the Group 1 Mean (3.0), Group 1 SD (.82), and Group 1 size (4). The values for the fair group were entered into MOTE under the Group 2 Mean (5.33), Group 2 SD (2.52), and Group 2 size (3).



**Charts:**

The best type of chart for anything analyzing group means is a bar chart.

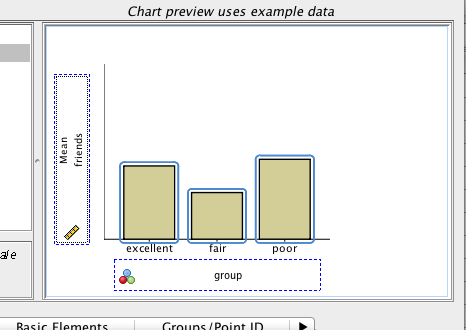
Graphs > chart builder:



Pick bar charts on the left > double click / drag & drop the simple bar into the editor field.

Your IV goes onto the X-axis (the levels/conditions variable).

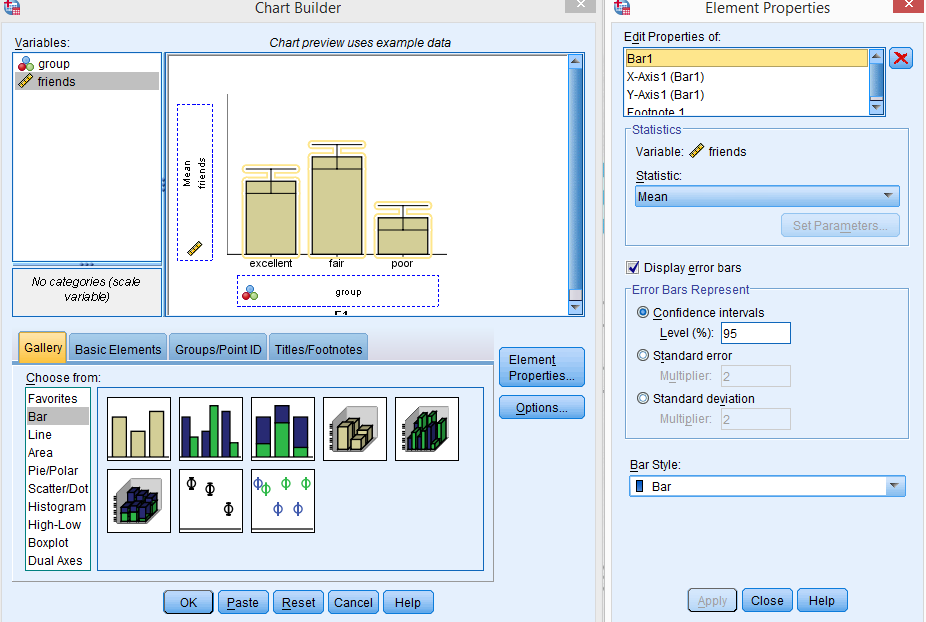
Your DV goes into the Y-axis.



Be sure to add error bars, x-axis, and y-axis labels (chart notes). See figure 1 below.

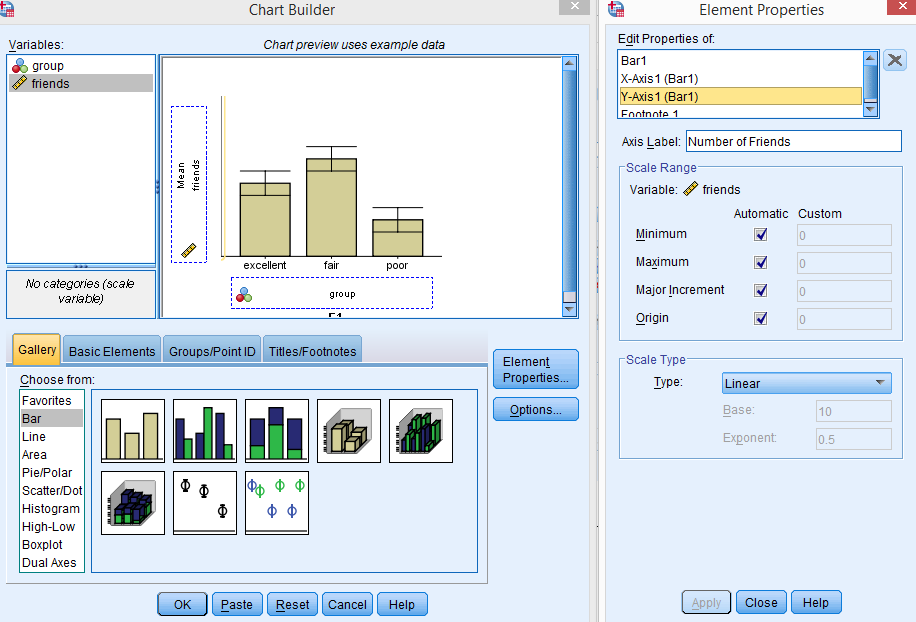
**To Add Error Bars:**

Select Bar1 (Under Edit Properties of)>Check the Display Error Bars Box> Hit Apply



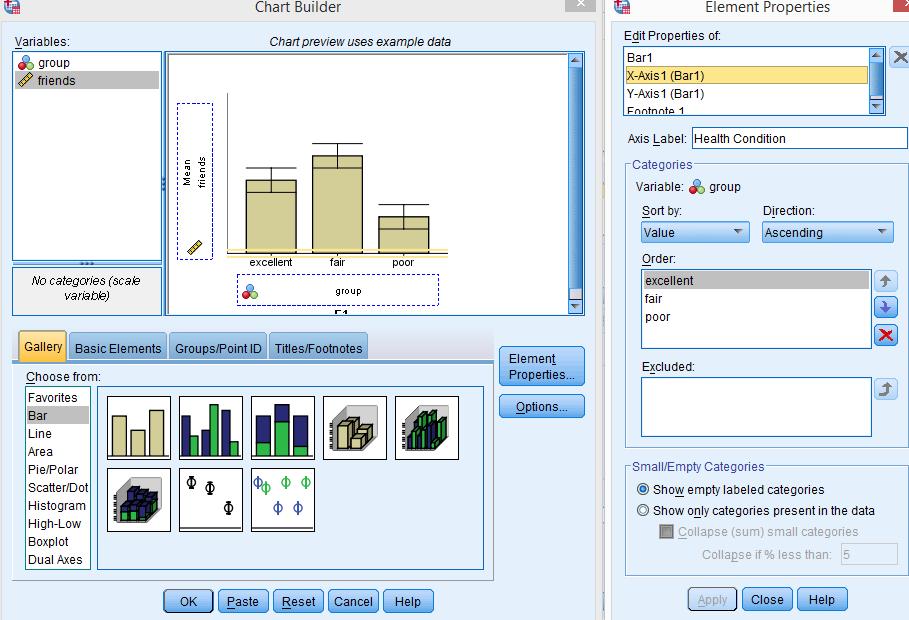
**To Add Labels to the Y-Axis:**

Drag the IV (“friends”) into the blue Y-Axis box>Select Y-Axis (under the Edit Properties of)> Change the name of the Y-Axis in the Axis Label box>Hit Apply



**To Add Labels to the X-Axis:**

Drag the DV (“group”) into the blue X-Axis box>Select X-Axis (under the Edit Properties of)> Change the name of the X-Axis in the Axis Label box>Hit Apply

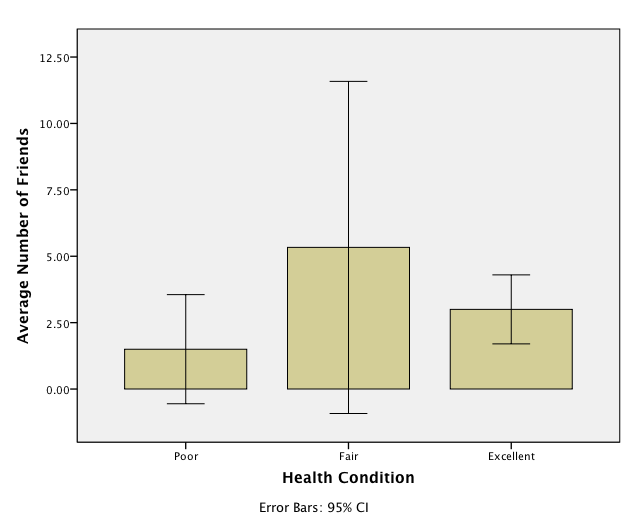


Example write up:

**Results**

**Example write-up when you do not have a figure**: Participants’ numbers of inter-personal connections were examined in relationship to their overall health. The data was screened for assumptions and outliers and found to be satisfactory (Levene’s *p* = .18). A between subjects ANOVA was used to analyze the number of inter-personal attachments with excellent, fair, and poor health groups. These group means were found to be significantly different, *F*(2,8)=5.13, *p*=.04, η2 = .56. A Tukey post correction test was used to examine differences between groups. The excellent health group (*M* = 3.00, *SD* = .82) was not different from the fair health group (*M* = 5.33, *SD* = 2.52, *p* = .19, *d* = 1.35). The excellent and poor health groups (*M* = 1.50, *SD* = 1.29) were also found to have equal numbers of inter-personal attachments (*p* = .41, *d* = 1.39). However, the fair health group was found to have significantly more attachments than the poor health groups (*p* = .03, *d =* 1.91).

**Example write-up without means when you have a figure:** Participants’ numbers of inter-personal connections were examined in relationship to their overall health. The data was screened for assumptions and outliers and found to be satisfactory (Levene’s *p* = .18). A between subjects ANOVA was used to analyze the number of inter-personal attachments with excellent, fair, and poor health groups. These group means were found to be significantly different, *F*(2,8)=5.13, *p*=.04, η2 = .56. A Tukey post correction test was used to examine differences between groups. As shown in Figure 1, the excellent health group was not different from the fair health group (*p* = .19, *d* = 1.35). The excellent and poor health groups were also found to have equal numbers of inter-personal attachments (*p* = .41, *d* = 1.39). However, the fair health group was found to have significantly more attachments than the poor health groups (*p* = .03, *d =* 1.91).



*Figure 1.*

# Complete Example (Repeated Measures – 1 IV only)

A note on repeated measures designs:

* People sometimes find it difficult to separate IV and DV in these designs because they are related. Here’s a simple trick:
  + The columns in your SPSS document are the IV … like the titles of them. Here it’s the manipulation of the different days (cardio). It might be time or pre-test/post-test, etc.
  + The DV is the *number* in the columns in SPSS, what the participant did/gave you in the study. Here’s it’s heart rate, but might be test scores in a different example.
* These designs are more powerful (need less people!) because you are measuring the same people a couple times. That means that you reduce the error variance (within subjects) because you can control for the fact that everyone is slightly different and take that out of the error variance. However, the draw back is that there is an interaction with the study sometimes (i.e. if you take the same test over and over, of course you’ll get better … or worse because you are bored.). So, you have to balance these effects (carry over effects, fatigue) with the ability to control for participant wackiness.

Participants were tested over several days to measure variations in their pulse. Day 1 tested each participant for their “at rest” heart rate. Day 2 tested each participant after slow jogging. Day 3 tested each participant after 30 minutes of intense cardio exercise. Where there differences in pulse for each participant across all three days? (rm 1 anova.sav)

**IV:**

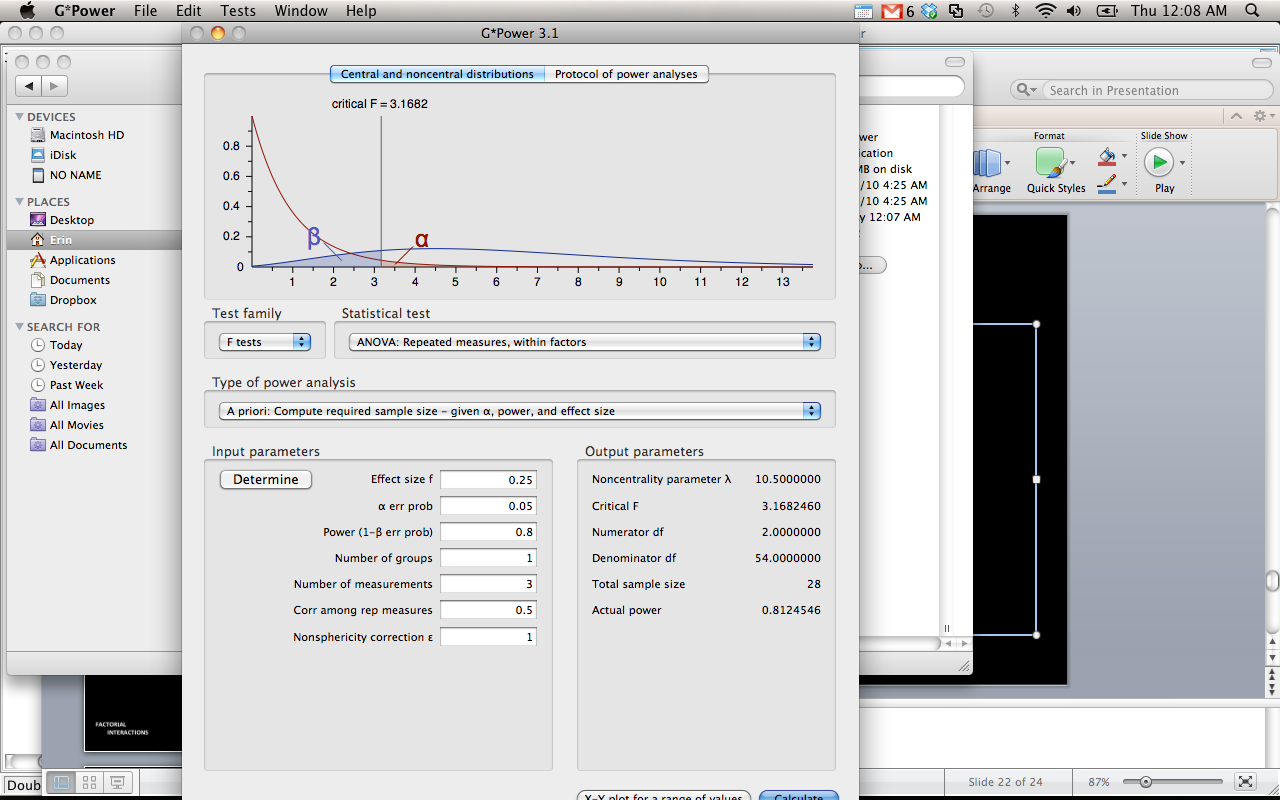
Activity level: at rest, slow jog, cardio (each will have their own column because participants did all of them).

**DV:**

Pulse/heart rate

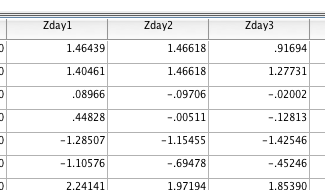
**Power:**

* F-test, ANOVA repeated measures, within factors
* Alpha = .05
* Beta = .80
* Number of groups = number of IVs
* Number of measurements = number of levels
* Corr among rep measures = correlation between levels (you can estimate from previous research or the first couple participants … go with the lowest correlation you find … .5-.7 is a good estimate if you are giving them the same test a couple times).
* Nonsphericity correction = epsilon … 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).

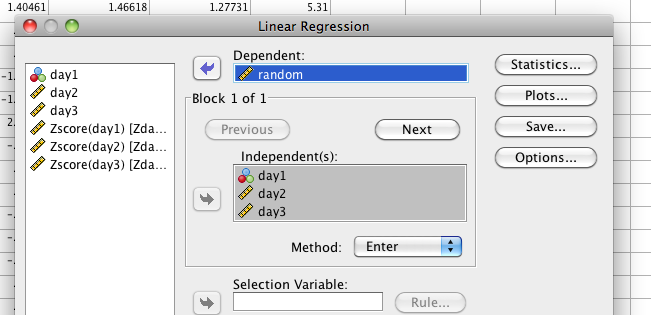
****

**Assumptions: (a lot of the pictures will be left out, you can see them above)**

1. Accuracy and missing data, just as in data screening.
2. Outliers
   1. Univariate – check each time measurement to make sure there’s nothing crazy going on.
   2. Analyze > descriptive statistics > descriptive
      1. Move over ALL time measurements
      2. Be sure to hit save standardized values.
      3. Hit options.
      4. Skewness and Kurtosis (for normality later)
   3. Checking my z-scores, I don’t see any univariate outliers.



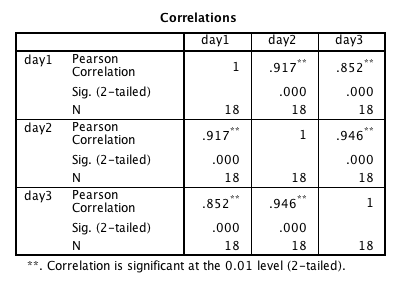
* 1. Multivariate – you need to check these here because you have multiple time measurements. In a between subjects with one variable, it’s not that big a deal, but anything with multiple measurements needs to have multivariate outliers check.
     1. First make a fake variable.
        1. Transform > compute
        2. Type in a variable name (random)
        3. Use a random variable creator for the numeric expression box. (Rv.chisq(7)).
     2. Run a fake regression.
        1. Analyze > regression > linear
        2. Put your random variable in the dependent box, and ALL of your levels in the IV box.



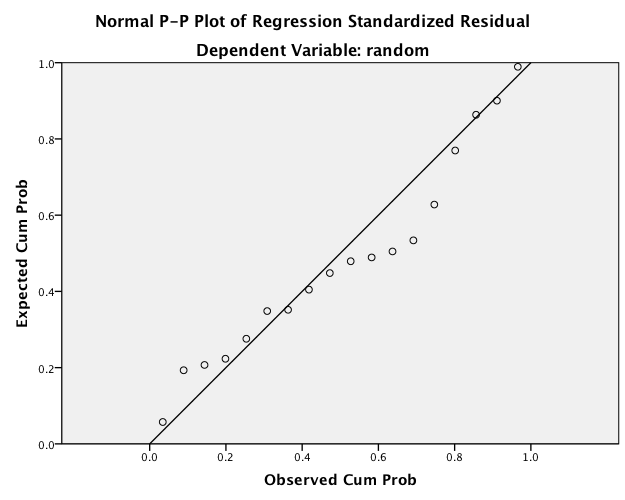
* + - 1. Under plots > zpred in Y, zresid in X, click both histogram and normal probability plot.
      2. Under save > Mahalanobis
    1. Check out the mahalanobis scores
       1. We have three levels (activity levels), so the cut off score is *χ2*(3) p<.001 = 16.27
       2. We do not have any multivariate outliers.



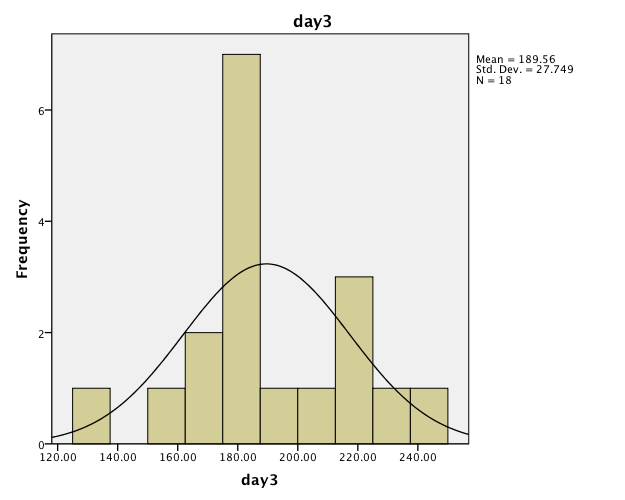
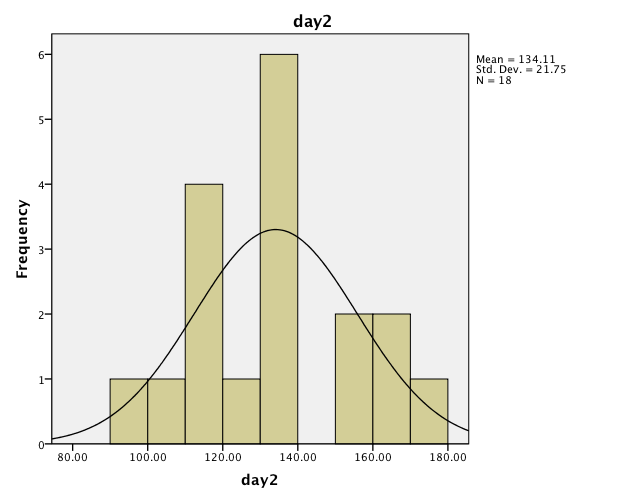
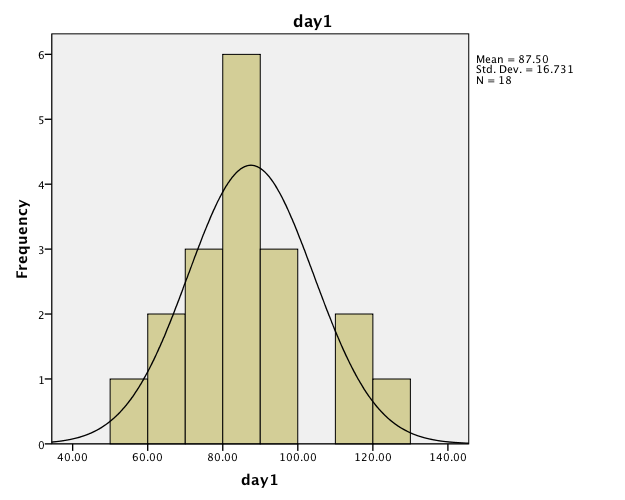
1. Multicollinearity – for repeated measures, you expect there to be a correlation between the levels … that’s sort of the point. So you will check a correlation table to make sure there isn’t a *perfect* correlation or SPSS will freak out.
   1. Analyze > correlate > bivariate.
   2. Check for .999 or 1.00 (don’t forget this is different than the other multicollinearity checks, as you can see in this chart, we have some over >.90 but that’s ok).



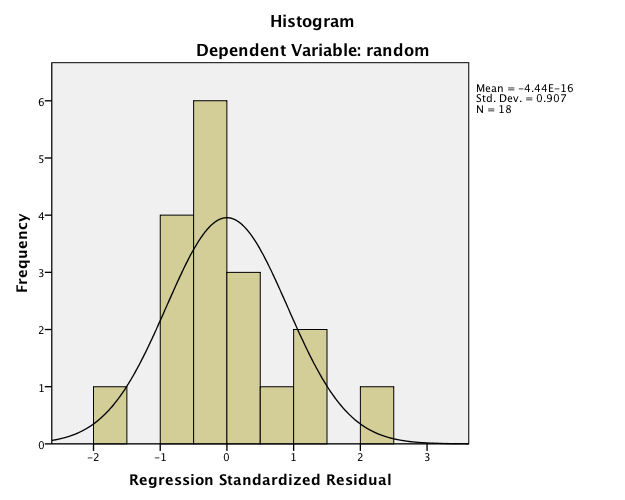
1. Linearity
   1. Check out the Normal PP Plot and make sure the dots are close to the line. You really want the relationship between time measurements to be linear. If this looks “funky” then look at each combination of time measurements together separately (using a bivariate scatter plot).



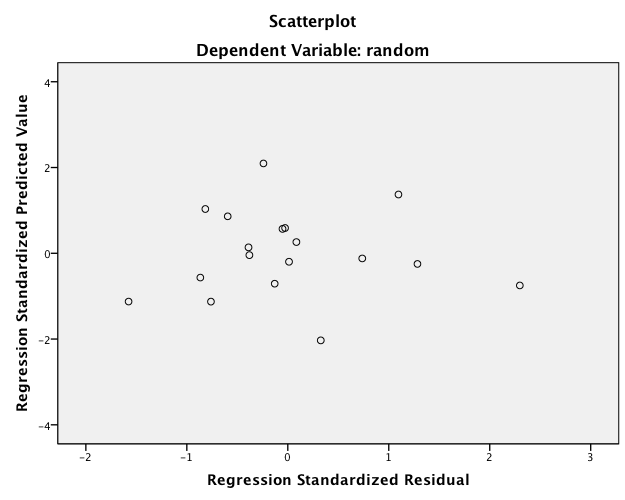
1. Normality
   1. Univariate – each variable should be normal individually.
      1. Analyze > descriptive statistics > frequencies
      2. Move over all levels of the IV
      3. Charts > histograms > show normal curves



* 1. Multivariate normality – the combination of all time measurements should be normal
     1. Use the histogram from your fake regression.
     2. Notice that this chart mirrors the other three, and is more normal than that cardio level.



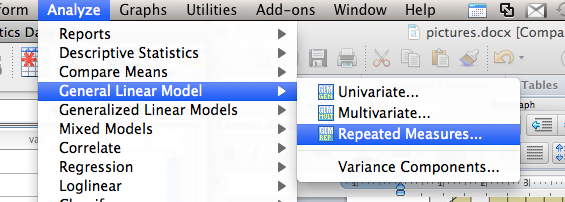
1. Homogeneity
   1. Some people argue that you do not expect things to have the same variance across time because you are obviously trying to change people across time (like losing weight for example). However, you don’t want it to be crazy different.
      1. This analysis/idea also applies to sphericity, which is special assumption for repeated measures (see below).
   2. Check out the residual plot from your fake regression.



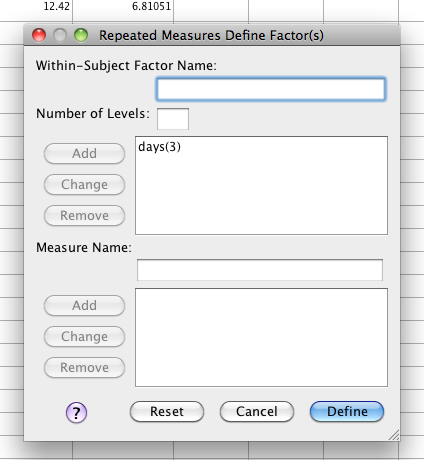
1. Sphericity – this assumption will be checked when you run the actual test. The assumption is considered *compound symmetry*:
   1. The correlations between all the levels are equal.
   2. The variance of the difference scores between each level combination is the same.
   3. It is almost impossible to meet this assumption:
      1. Generally, you are examining if there are differences in levels.
      2. They are often taking the same thing over and over.
      3. So, the variances often get much smaller or larger across the levels.
      4. It’s such a problem, people often ignore sphericity.
   4. More discussion below.

**Running the analysis:**

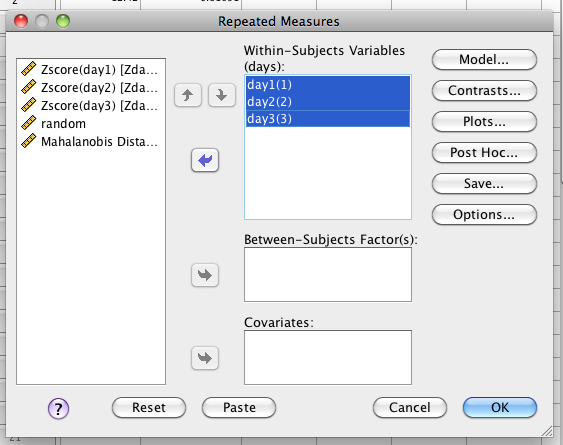
1. Analyze > general linear model > repeated measures (any analysis with a repeated measures variable must use repeated measures GLM).



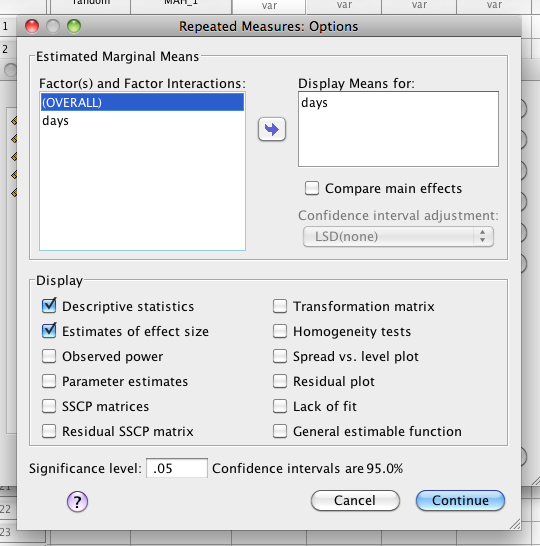
1. When the define box comes up:
   1. Make up a name for the variable (so you can find it later).
   2. List how many times that variable was measured.
   3. Hit add.
   4. Hit define.



1. Move over the level measurements (in the right order! If it matters) into the ?\_\_(1) spots.

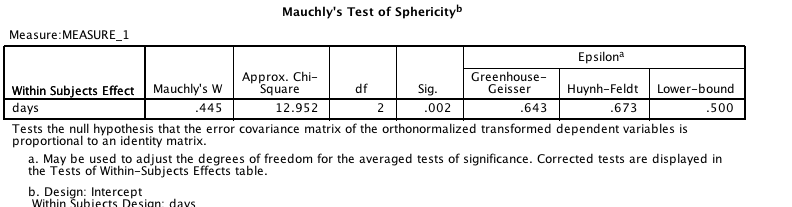


1. Hit options
   1. Move over the mean to the right.
   2. Hit descriptive statistics, effect size, (we normally click homogeneity tests, and I always click it so I don’t forget when I’m doing a mixed design).
   3. Hit continue and ok.

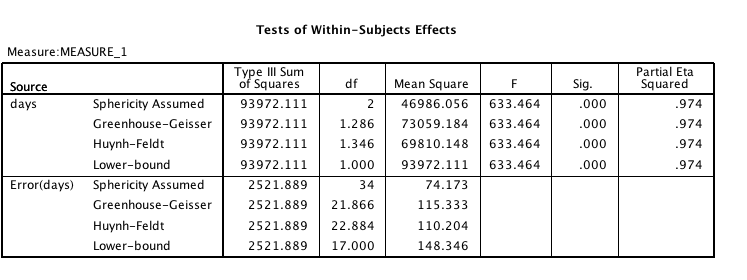


**Translating the output:**

1. Ignore the multivariate box, contrasts and between subject boxes.
2. Check out the Sphericity box – you want this test to be p > .001.
   1. If it is p<.001, then you’ll want to use a correction (in theory …)
   2. Which one?
      1. Greenhouse-Geisser is the most common but you want to look at epsilon. If both the GG and Huynh-Feldt epsilons are < .75, then use GG.
      2. If >.75, then use Huynh-Feldt.
      3. What does that mean?! You will look at the repeated measures ANOVA box and use that particular line.
   3. In this case, I scraped by! But if I hadn’t I would have used the GG.



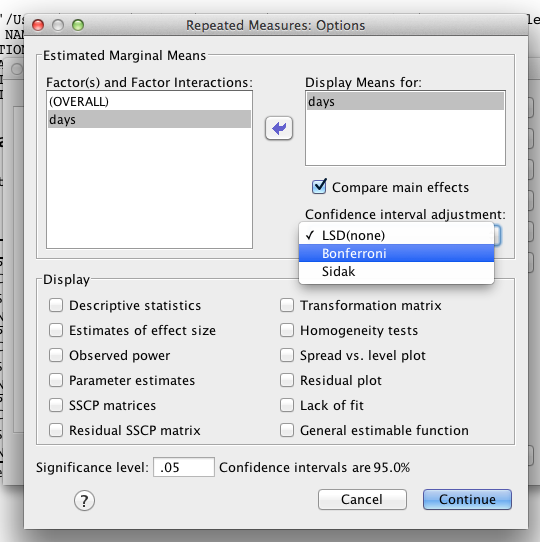
1. The ANOVA box – Tests of within-subjects effects:
   1. You will use only the correction line you choose (most commonly sphericity assumed).
   2. You’ll get the df, F value and Sig - you want Sig to be p<.05.
   3. *F*(2, 34) = 633.46, *p*<.001, *ηp*2 = .97
   4. Please Note: If you had not meet spherecity than you would have looked at the Greenhouse-Geisser correction line across the Within-Subjects Effects box which give you the results, *F*(1.29, 21.87)=633.46, *p*<.001, *ηp*2 = .97



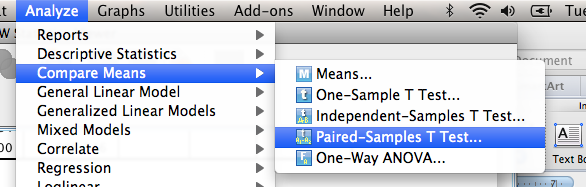
df p

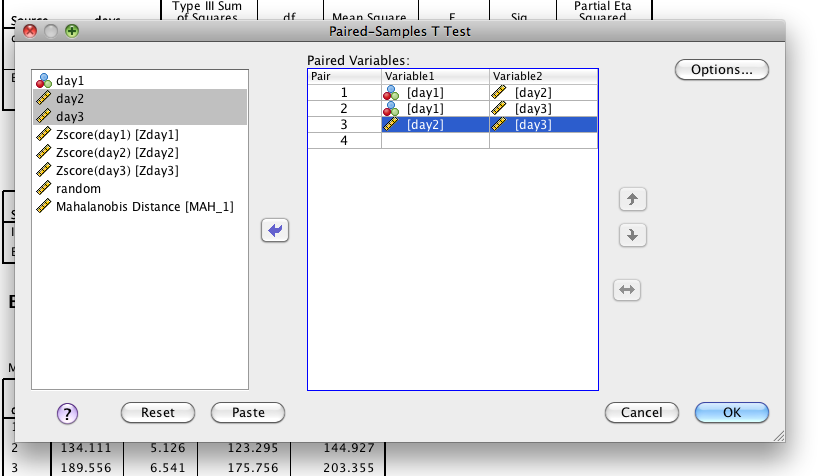
F eta

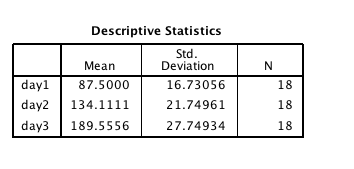
1. **If the ANOVA is significant Post Hocs:**
   1. The matching test for repeated measures is a dependent t-test.
   2. You will want to run the combinations of dependent t’s you hypothesized about (most people have a specific question OR will run all of them).
      1. If you run ALL of them and have a lot, then you will want to correct.
      2. The easiest correction is a Bonferroni – divide your p-value (.05) by the number of tests and look for that new p-value in the sig column. So instead of looking for Sig < .05, you might be looking for Sig < .017.
   3. There is also a Bonferroni option in the Options menu when you are running the repeated measures. (THIS IS EASIER).
      1. See notes above, click options.
      2. Move the variable over to get the means.
      3. Then click compare main effects.
      4. Then choose:
         1. LSD = dependent t with no correction.
         2. Bonferroni = dependent t with p value correction.

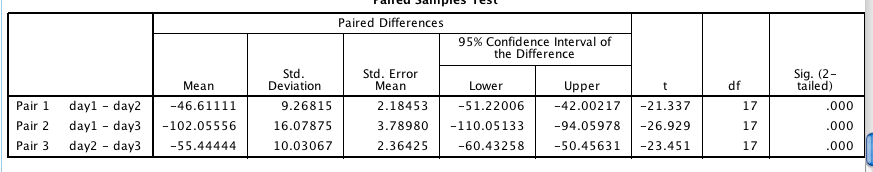
****

* 1. Continuing with dependent t option:
  2. Analyze > compare means > paired samples
     1. Create all combinations of the time measurements in the paired variables box and hit ok.
  3. Again, a chart really helps:

****

****

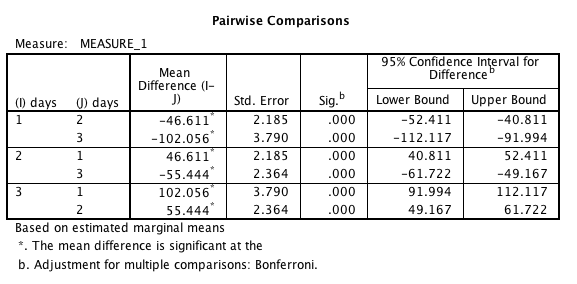
****

****

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Group 1** |  | **Group 2** | **P value** | **ddiff** |
| Day 1 (M=87.5) | < | Day 2 (M=134.11) | .001 | 5.03 |
| Day 1 (M=87.5) | < | Day 3 (M=189.56) | .001 | 6.35 |
| Day 2 (M=134.11) | < | Day 3 (M=189.56) | .001 | 5.53 |

So here, they are all significant if I correct or not (remember new *Bonferroni p* = .017).

Here’s the Bonferroni output for comparison.



Either way is appropriate.

**Effect size:** MOTE – use the dependent t options, remembering that there are two different ones (davg versus ddiff).

**What to enter into MOTE for Ddiff**

|  |  |
| --- | --- |
| **MOTE Column Name** | **Where to find the necessary values** |
| Mean Difference | Pairwise Comparisons box (46.61) |
| Standard Error Difference | Pairwise Comparisons box (2.185) |
| Size | Descriptive statistics box (18) |
| df | Paired samples test box (17) |

**What to enter into MOTE for Davg**

|  |  |
| --- | --- |
| **MOTE Column Name** | **Where to find the necessary values** |
| Mean Difference | Pairwise Comparisons box (46.61) |
| Time 1 SD | Descriptive statistics box (16.73) |
| Time 2 SD | Descriptive statistics box (21.75) |
| size | Descriptive statistics box (18) |
| df | Paired samples test box (17) |

|  |  |
| --- | --- |
| Differences | Averages |

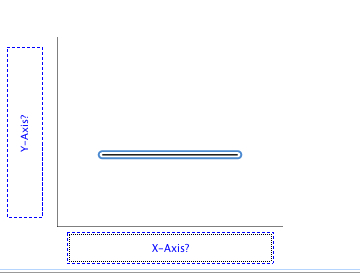
Make sure to indicate if you use ddifference or daverages because the numbers will be different. \*Please note that the t value will not match on daverages because of the math used.

**Charts:**

Bar charts would be appropriate for levels that are not time related (so they are separate things). Line graphs are appropriate for levels that are time based, so it’s sort of continuous on the bottom.

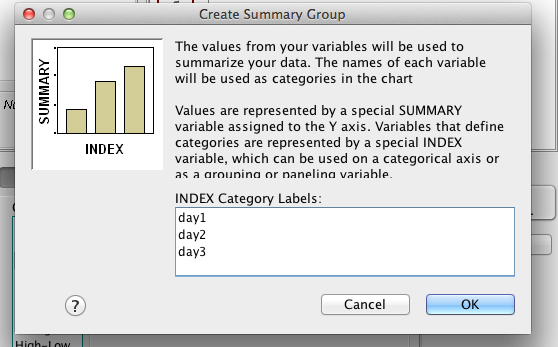
Graphs > chart builder.

Pick line at the bottom left, and double click/ drag and drop the first line graph into the builder area.



Here’s the odd part … put all the levels into the Y axis at once. (remember you’ll get the red cross plus sign).

You’ll get a pop up window that shows you that it’s going to create a “fake variable” for the x-axis.

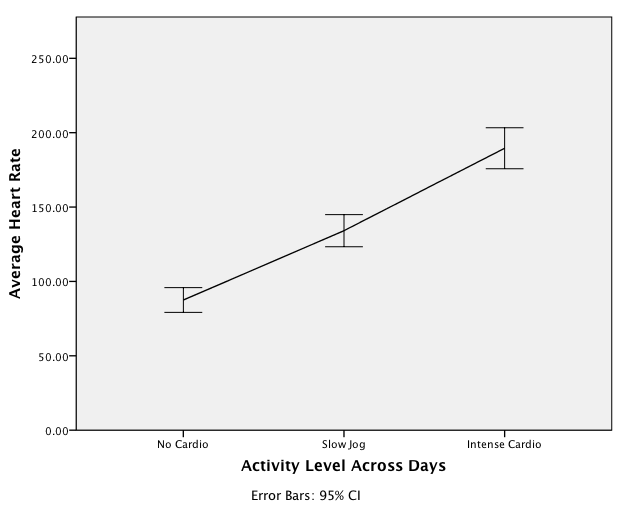


Change the x and y axis labels, and add error bars in the properties window (see graph below).

**Write up example:**

**Results**

Participants were tested on three successive days for pulse rate. On Day 1, they were told to sit at rest. Day 2 measured pulse rates after a slow jog, and Day 3 measured pulse rates after a cardio workout. Using a repeated measures ANOVA, different days were found have different heart rates, *F*(2,34)=633.46, *p*<.001, *ηp*2= .97. Post hoc comparisons were analyzed using dependent samples t-tests. Day 1 was found to have a significantly lower mean than Day 2 (*Mdiff* = -46.11, *SDdiff* = 9.27, *t*(17) = -21.34, *p*<.001, *ddiff* = 5.03), as well as Day 3 (*Mdiff* = -102.06, *SDdiff* = 16.08, *t*(17) = -26.93, *p*<.001, *ddiff* = 6.35). Day 2 showed lower pulse rates than Day 3 (*Mdiff* = -55.44, *SDdiff* = 10.03, *t*(17) = -23.45, *p*<.001, *ddiff* = 5.53). Therefore, pulse rates were lowest at rest, followed by slow jogs, and highest with more intensive workouts. Figure 1 displays the average pulse rates for each day.



*Figure 1.*

# Complete Example (Between Subjects – Factorial)

We want to know if we are spending different amounts of money for different years and types of transactions on our credit cards. (bn 2 anova.sav)

**IV:**

Year: 2007 versus 2008

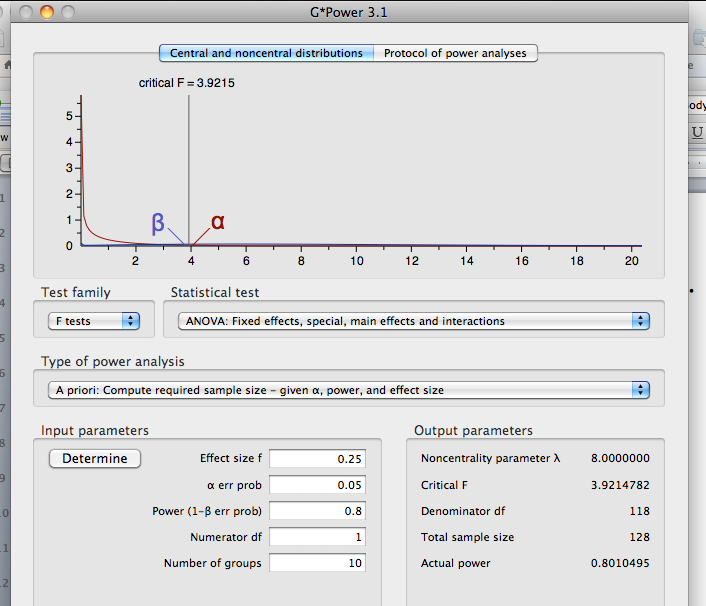
Transaction type: grocery, retail, entertainment, travel, other

**DV:**

Amount of money spent

**Power:**

* F-test
* ANOVA, fixed effects, main effects and interactions
* Alpha = .05
* Beta = .80
* Numerator df = number of IVs-1
* Number of groups = the total number of levels or conditions (depending on which analysis you want to check, usually people want the interaction = conditions)



So, we need 128 people to find the interaction.

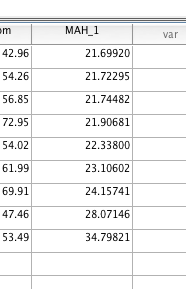
**Assumptions:**

1. Accuracy and missing data as described in data screening.
2. Outliers
   1. Univariate
      1. Analyze > descriptive statistics > descriptive
      2. Move over JUST the DV (no IVs)
      3. Save standardized values
      4. Options > Skew, kurtosis

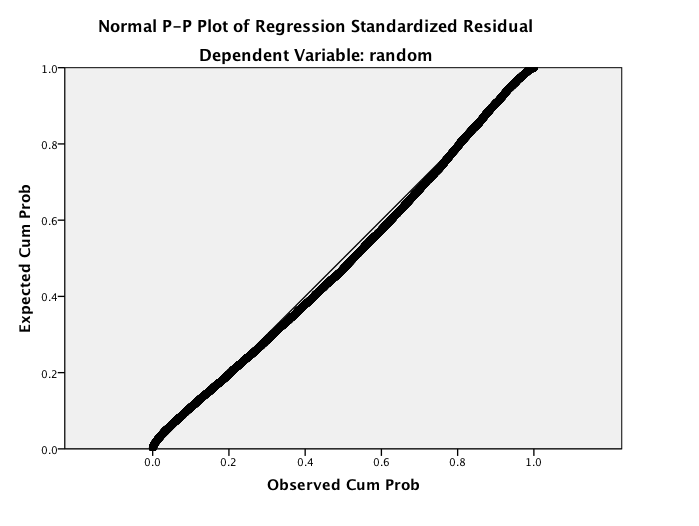


So we have a lot of outliers in this dataset.

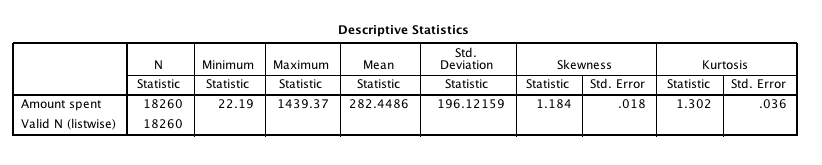
* 1. Multivariate – still only one DV, but you need to run the fake regression for the rest of the assumptions anyway.
     1. First make a fake variable.
        1. Transform > compute
        2. Type in a variable name (random)
        3. Use a random variable creator for the numeric expression box (Rv.Chisq(7)).
     2. Run a fake regression.
        1. Analyze > regression > linear
        2. Put your random variable in the dv box, and your DV in the IV box.
        3. Under plots > zpred in Y, zresid in X, click both histogram and normal probability plot.
        4. Under save > Mahalanobis
     3. Check out the mahalanobis scores
        1. One variable = cut off *p*<.001, 10.83
  2. We have some crazy outliers. However, that is money that you spent, and you know it’s not a typo. In that case, I’d say to leave them in because of the specific circumstances of this dataset.

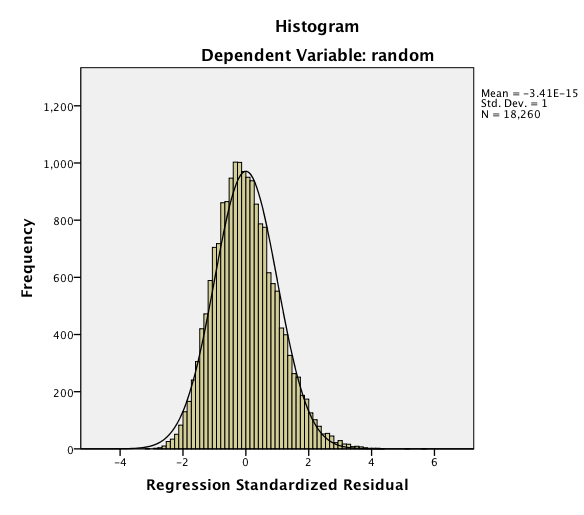


1. Linearity – looks ok!

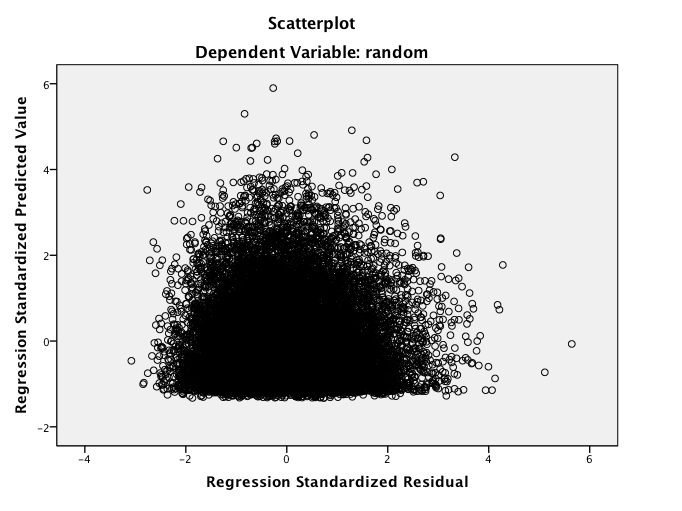


1. Normality – looks ok!
   1. Skew and Kurtosis values are ok.
   2. You can check the single frequency chart (see above in first example), but they are the same as the fake regression ones.



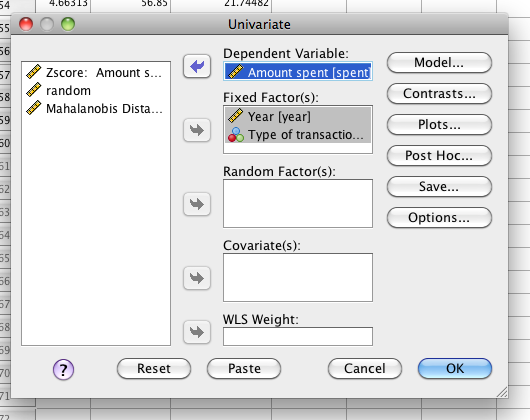


1. Homogeneity
   1. This graph is a little nutty. You can see the couple outliers, and a very odd shape above zero.
   2. I would check Levene’s to see if you get something bad before worrying (because we are way over the central limit theorem number of DFs).

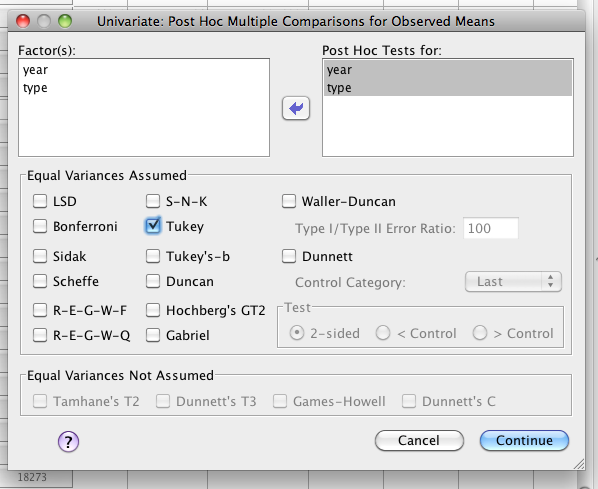


**Running the ANOVA:**

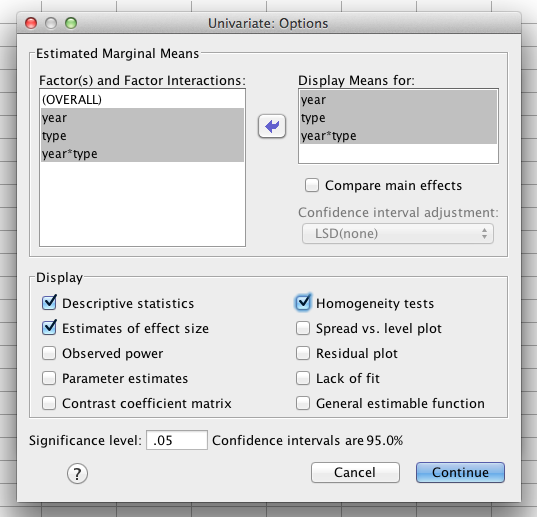
1. Analyze > general linear model > univariate – you will use this version because it’s all between subjects.
2. Move the DV into the DV box and both IVs into the fixed factor box.



1. Under post hoc > move both over to the right and hit Tukey.

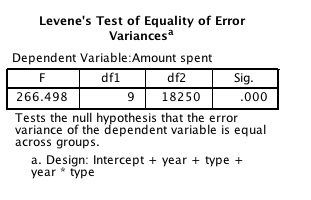


1. Options
   1. Move all means over to the right.
   2. Hit buttons for descriptive statistics, effect size, homogeneity tests

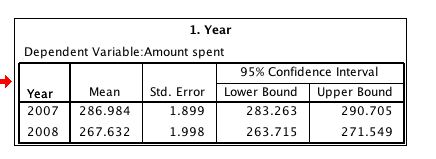


**Translating the Output:**

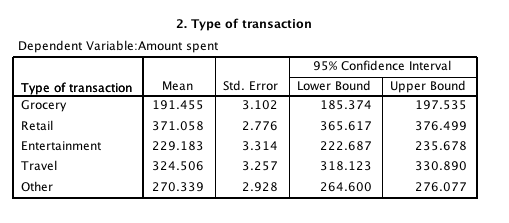
1. Levene’s:
   1. Significant (*p*<.001) – yikes! But we have a very large *N* values, so you should be ok.

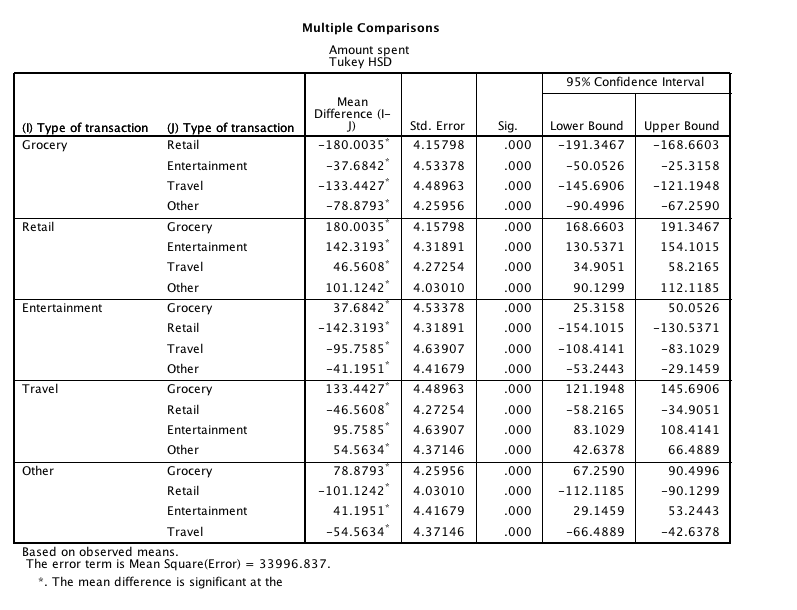


1. ANOVA Box
   1. Main Effect 1: you look at the first IV by itself.
      1. Year appears to have a significant effect on how much you spent.
      2. *F*(1,18250) = 49.29, *p*<.001, *ηp*2 =.003
      3. Year only has two levels – **so no post hoc.**
      4. So, let’s look at the means.
      5. You spent more per transaction in 2007 than 2008.



* 1. Main effect 2: you look at the second IV by itself.
     1. Transaction type also has a significant difference.
     2. *F*(4,18250) = 580.40, *p*<.001, *ηp*2 =.11
     3. However, there are 5 types, so you’ll have to look at the post hoc.
     4. This post hoc is going to create a LOT of tests…
        1. So you will hold onto the Tukey box until you see the interaction.
        2. If you have an interaction, **most people don’t even talk about the individual main effects.**

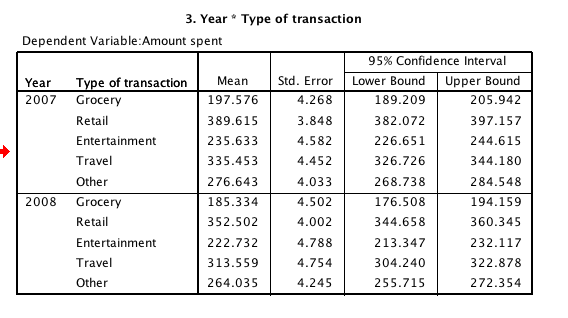




If our interaction were not significant, we would look at these to figure out what was different. They are all less than *p* < .05, and remember Tukey is a correction, so those *p*-values take into account the fact that it was a bunch of pairwise comparisons.

* 1. The interaction was also significant: *F*(4,18250) = 3.47, *p* = .008, *ηp*2 =.001
     1. Interactions occur when you have different patterns of data across the two variables. For example, see these charts of the data in each direction:
     2. What you are seeing is that the dots do not line up. When they do not line up, you get an interaction.

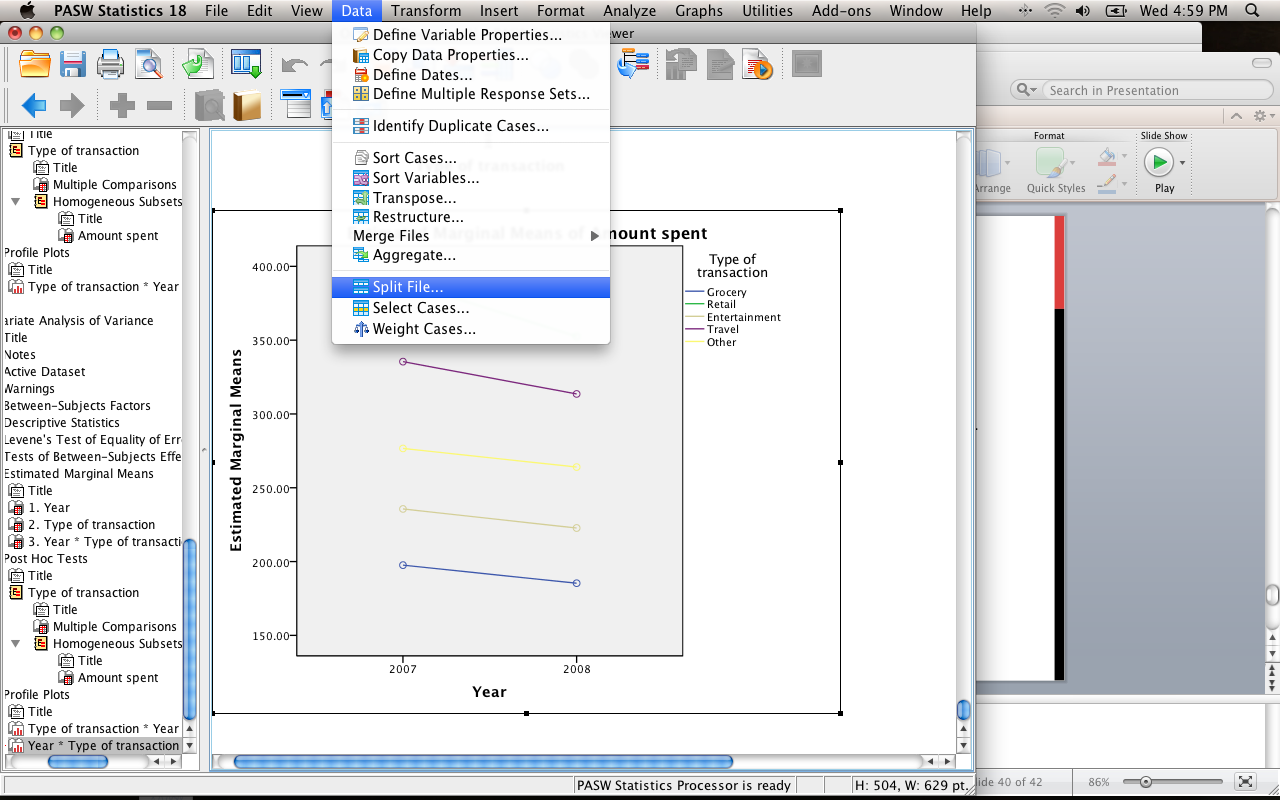
1. **Post Hoc Interactions**
   1. First, always go back and make a chart:



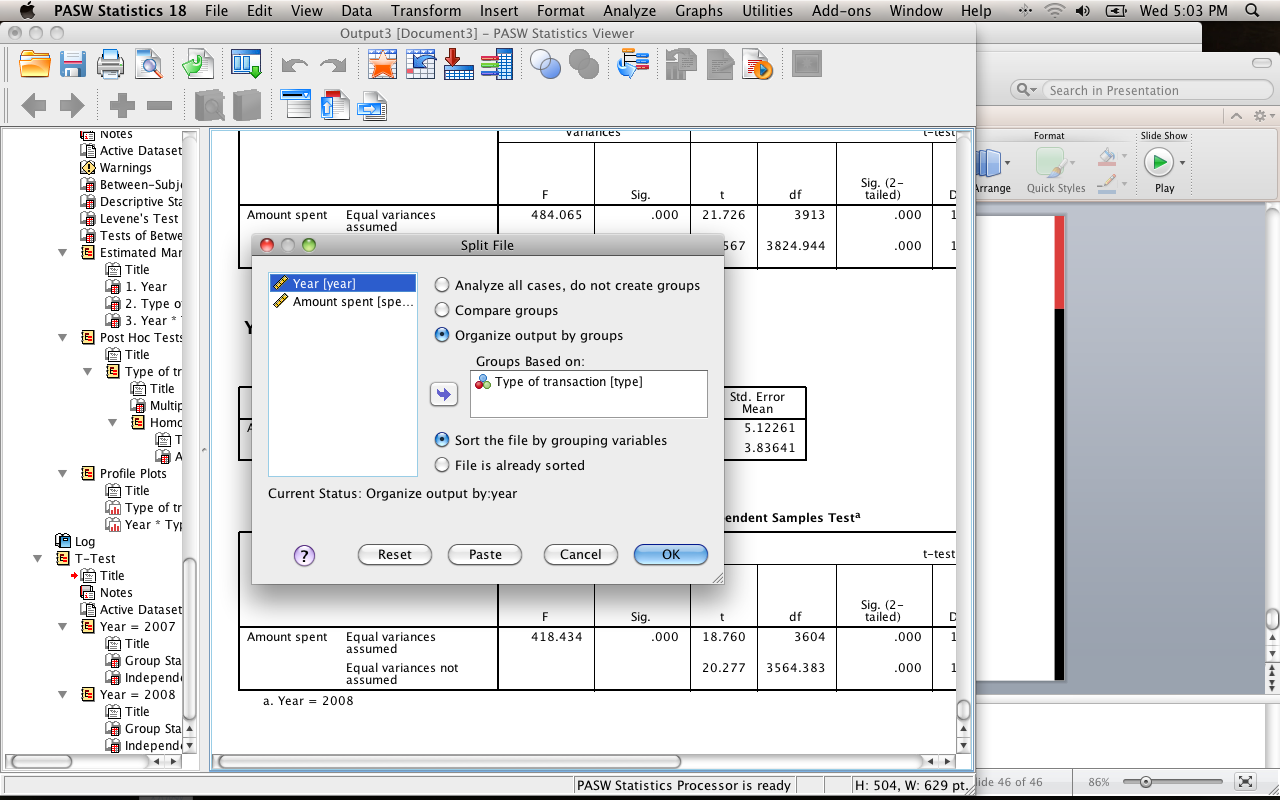
|  |  |  |
| --- | --- | --- |
|  | **2007** | **2008** |
| **Grocery** | 197.58 | 185.33 |
| **Retail** | 389.62 | 352.50 |
| **Entertainment** | 235.63 | 222.73 |
| **Travel** | 335.45 | 313.56 |
| **Other** | 276.64 | 264.04 |

* 1. Now that I have that laid out, you can analyze this two different ways:
     1. Going across
        1. Compare grocery 2007 to grocery 2008
        2. Compare retail 2007 to retail 2008
        3. Etc.
     2. Going down
        1. Compare grocery 2007 to retail 2007 to entertainment 2007…
        2. Compare grocery 2008 to retail 2008 to entertainment 2008…
  2. Which way is right? Either. Depends on your question.
  3. I want to know if I was able to reduce my retail (shoes!) and (entertainment) budgets from 2007 to 2008.

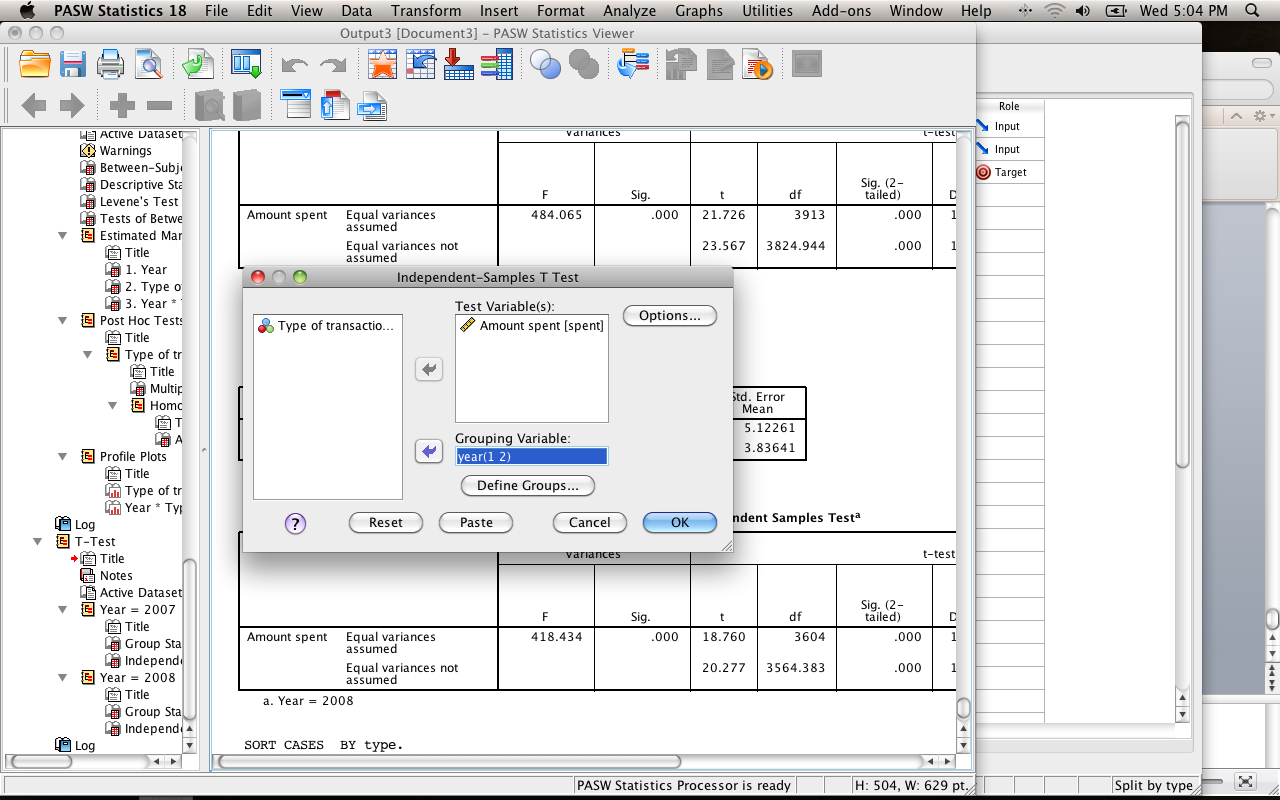
1. IMPT! Post hocs for only between subjects are always independent t-test!
   1. Problem: SPSS needs you to take one extra stepola to make this happen.
   2. Data > Split file



* 1. Pick the variable you want to “split” on.
     1. If you go across – you are splitting up the lines or the entertainment variable.
     2. If you go down – you are splitting up the columns or the year variable.
     3. Do it wrong (you’ll get an error message when you try to run it)? Data > split file and change it!

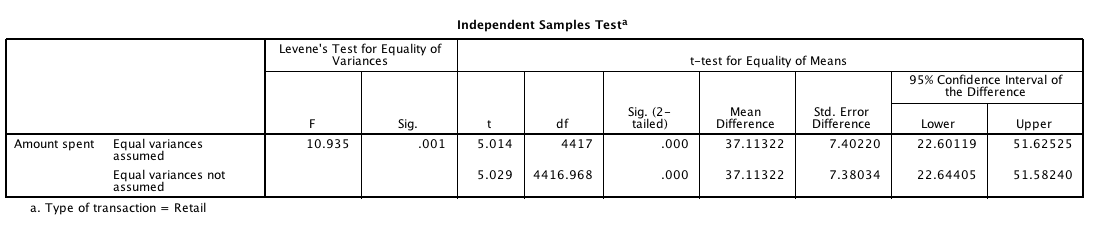


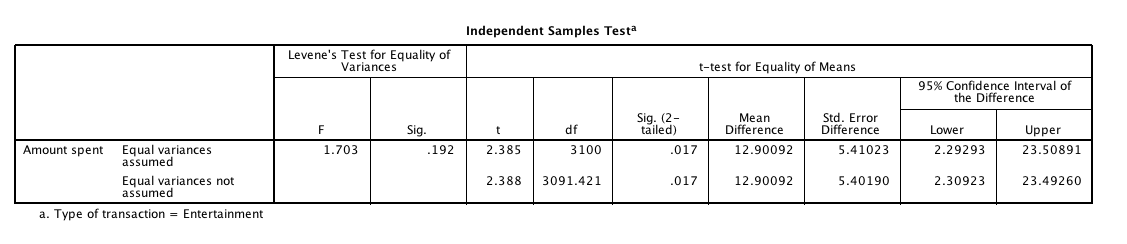
* 1. Now run an independent t-test on the thing you DIDN’T split on. If you do it wrong you’ll get a SPSS error.
     1. Analyze > compare means > independent samples t-test
     2. You’ll fill in the group numbers depending on which way you went. If you have more than two groups, you will have to run this several times with different group numbers (lame).



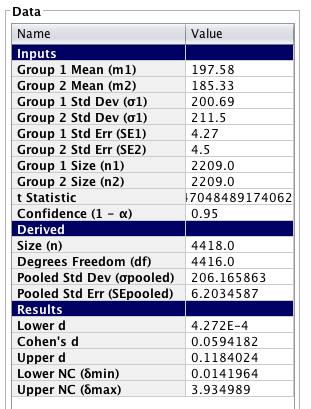
* + 1. Now you’ll see that you get output for all the entertainment categories, testing 2007 versus 2008. I just want to see retail and entertainment.

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  | **2007** | **2008** | **P-value** | **Explain** | **Cohen’s *d*** |
| **Retail** | 389.62 | 352.50 | .001 | Significant | d = .15 |
| **Entertainment** | 235.63 | 222.73 | .02 | Significant | d = .09 |





You can use these values to add effect sizes to your post hoc tests. Since it’s independent t, you’ll use the independent t options in MOTE.



**Charts:**

Graphs > chart builder

Pick bar graph on the left side.

Since we have two variables, we are going to use a clustered bar graph.

Put one IV in the x-axis, one IV in the set color option. Usually you want to pair together the options that you used in your post hoc (i.e. here I would put my categories together because that’s what I compared.). Put the DV in the y-axis.

Be sure to use error bars, x and y axis labels. See figure below.

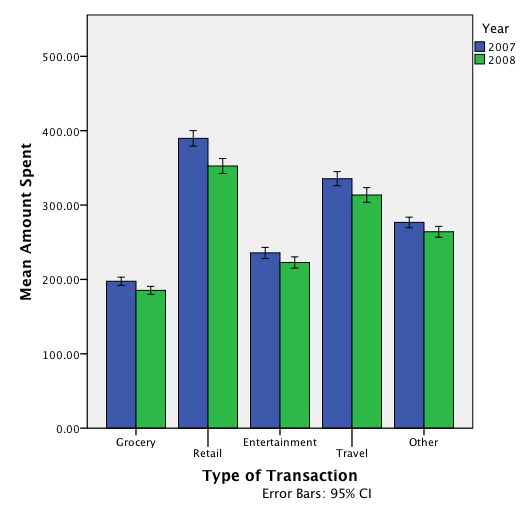
Write Up Example:

**Results**

Monthly budget transactions were examined to determine if type of transaction (grocery, retail, entertainment, travel and other) changed over a two year period (2007, 2008). Data were screened for assumptions (linearity, homogeneity, normality, outliers), and two problems were found. While Levene’s test indicated a potential problem with homogeneity (*p* <.001) the large sample size may have influenced this factor and a residual plot showed homogeneous groups. Several outliers were found with high standardized scores, which were left in analyses after determining the data was correct.

A 2X5 factorial ANOVA was analyzed on year and transaction type. Both main effects of year (*F*(1,18250) = 49.29, *p*<.001, *ηp*2 =.003) and transaction type (*F*(4, 18250) = 580.40, *p*<.001, *ηp*2=.11) were significant. However, since the interaction between year and transaction type was also significant, *F*(4,18250) = 3.47, *p* = .008, *ηp*2 =.001, only post hoc analyses are described below. Figure 1 shows the interaction. Independent t-tests were performed to examine if average budgets had decreased across time. Retail (*t*(4417) = 5.01, *p* < .001, *d* = .15)and entertainment (*t*(3100) = 2.39, *p* = .02, *d* = .09) scores were both found to decrease from 2007 to 2008. Retail money decreased approximately $35 in 2008, while entertainment money decreased about $12 in 2008.

(\*\*\* you would go on and talk about all the different comparisons\*\*\*)



*Figure 1.*

# Complete Example (Repeated Measures With Interactions)

In this experiment people were given word pairs to rate based on their “relatedness”. How many people out of a 100 would put LOST-FOUND together? (rm 2 anova.sav)

**IVs:**

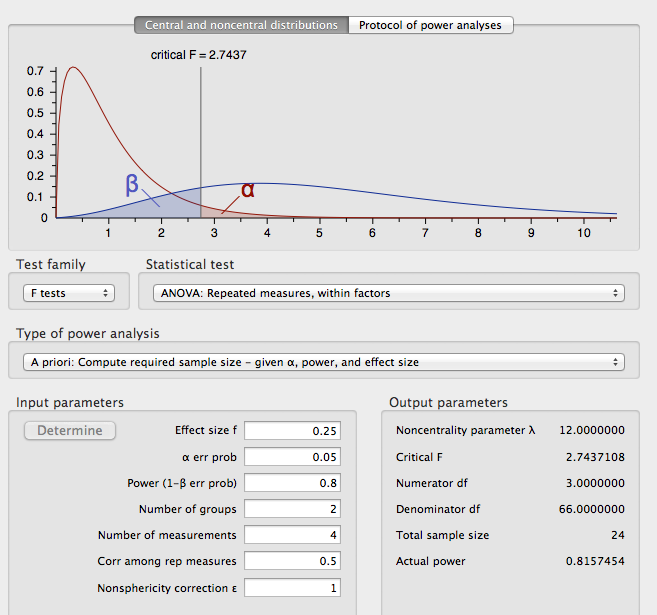
* FSG – how strong the relationship was between LOST-FOUND
  + FSG – two levels (low versus high)
* BSG – how strong the relationship was between FOUND-LOST
  + BSG – two levels (low versus high)
* Creates 4 conditions (lo lo, lo hi, hi lo, hi hi).
* Everybody got all four types of word pairs.

**DV:**

Rating of each word pair type.

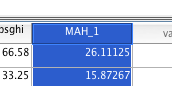
**Power:**

* F-test, ANOVA repeated measures, within factors
* Alpha = .05
* Beta = .80
* Number of groups = number of IVs
* Number of measurements = number of conditions
* Corr among rep measures = correlation between levels (you can estimate from previous research or the first couple participants … go with the lowest correlation you find … .5-.7 is a good estimate if you are giving them the same test a couple times).
* Nonsphericity correction = epsilon … 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).



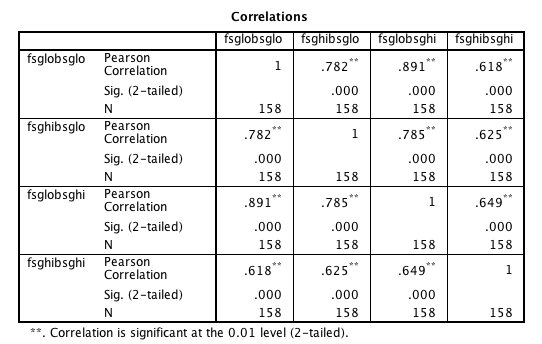
**Assumptions:**

1. Accuracy and missing are just like described in data screening.
2. Outliers
   1. Univariate
      1. Analyze > descriptive statistics > descriptive
         1. Move over ALL conditions (since they technically are all measurements of the DV).
         2. Be sure to hit save standardized values.
         3. Hit options.
         4. Skewness and Kurtosis (for normality later).
         5. Checking my z-scores, I don’t see any univariate outliers.
   2. Multivariate
      1. First make a fake variable.
         1. Transform > compute
         2. Type in a variable name (random).
         3. Use a random variable creator for the numeric expression box (Rv.Chisq(7)).
      2. Run a fake regression.
         1. Analyze > regression > linear
         2. Put your random variable in the dv box, and your condition measurements in the IV box.
         3. Under plots > zpred in Y, zresid in X, click both histogram and normal probability plot.
         4. Under save > Mahalanobis
      3. Check out the Mahalanobis scores
         1. 4 combinations = cut off *p*<.001, 18.47

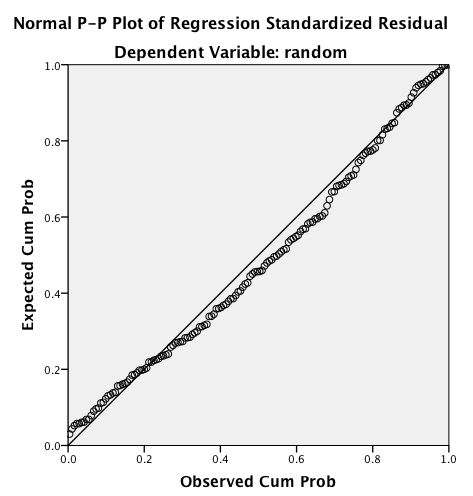


* + 1. At least one multivariate outlier. You could easily eliminate them or leave them in.

1. Multicollinearity – remember for repeated measures, you basically don’t want perfect correlations (*r* = .999 – 1).
   1. Analyze > correlate > bivariate.
   2. Move over all the conditions
   3. Check out the correlation table.

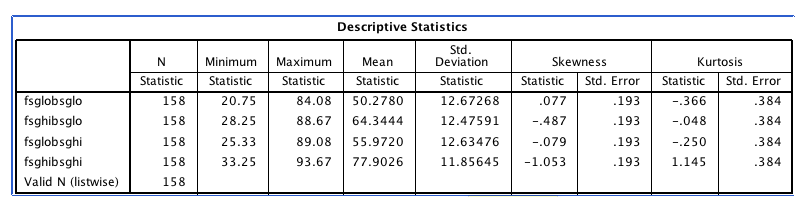


1. Linearity – looks good!

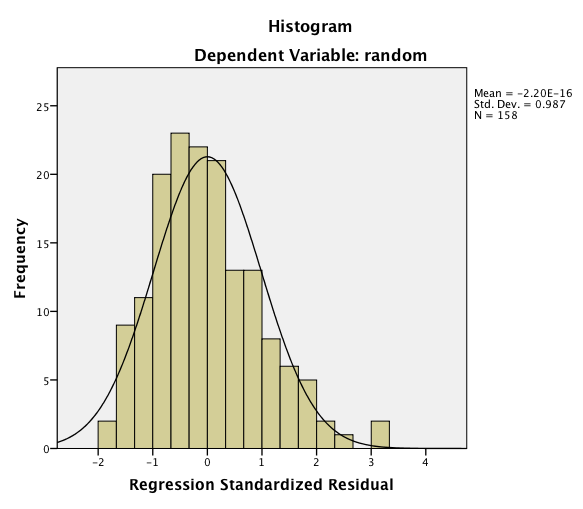


1. Normality - you can also check the skew and kurtosis values, as well as the individual condition normality charts.
   1. Univariate example:

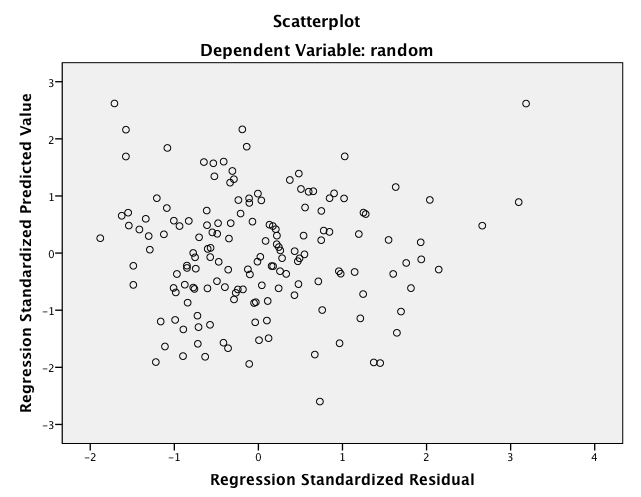
|  |  |
| --- | --- |
|  |  |



* 1. Multivariate example

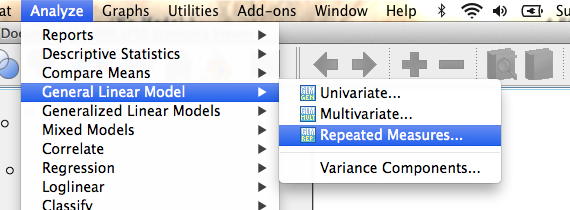


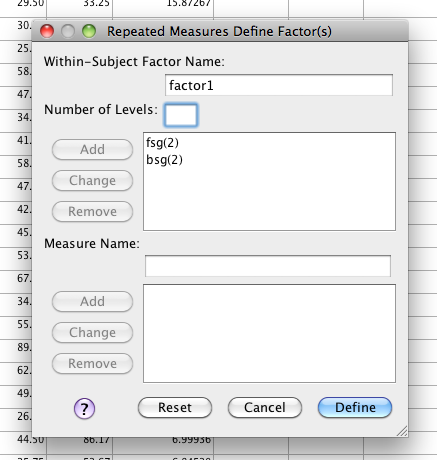
1. Homogeneity – might have some issues on the right side.



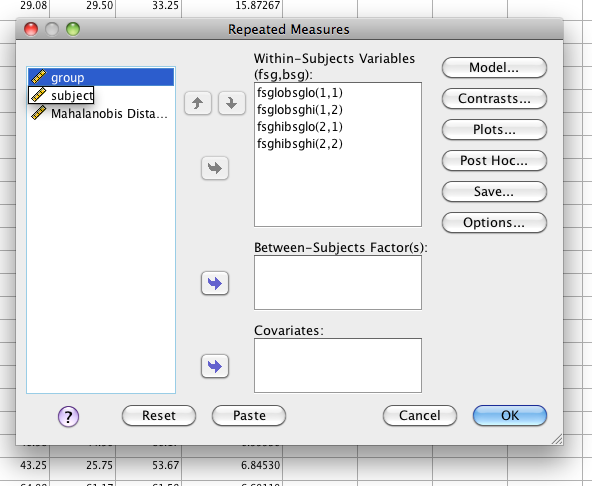
**Running the analysis:**

1. Analyze > general linear model > repeated measures
   1. You’ll need to create two variables (don’t create one with four levels … why? You would not be able to check out the interaction if you did that – it would be collapsing into one giant variable).
   2. Each variable has two levels (lo versus hi)

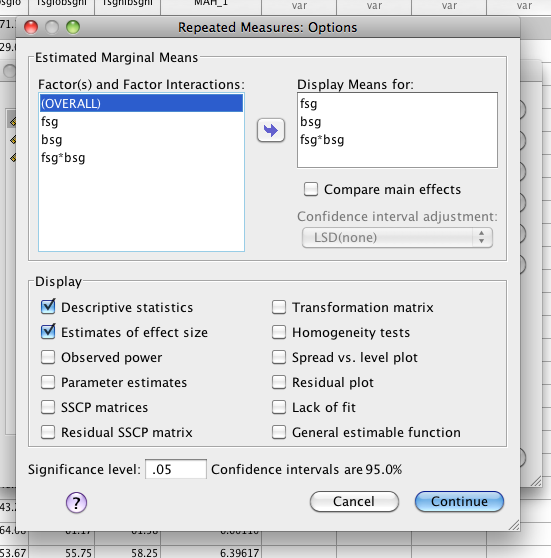




1. You will now have more options to move things over under the within-subjects variables box. As you move them over make sure they match the numbers in commas.
   1. I.E. 1 = low, 2 = high

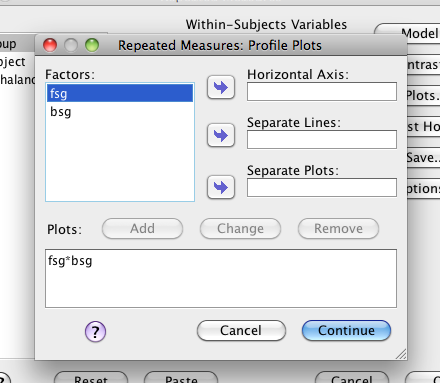


1. Options > descriptives, effect size, move over means (you do not have to click homogeneity because you do not get Levene’s with repeated measures only, but remember you can click it if you just want to always remember to check it, and you’ll just a warning in the output).



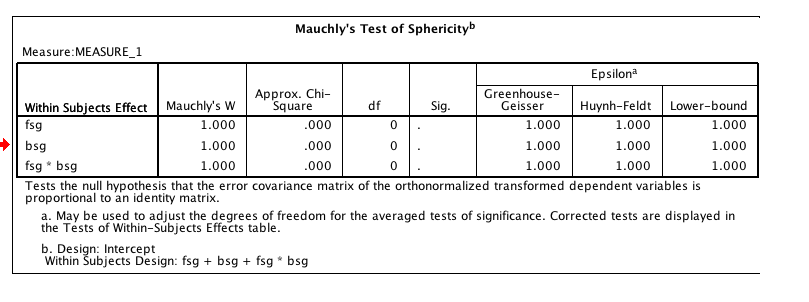
Don’t forget that here is where you can get repeated measures post hocs for the main effects only (i.e. each IV separately). You would click compare main effects, and then pick which correction you want to use (dependent t with Bonferroni or Sidak-Bonferroni are your only options). I did not do that here because I have two levels of each of the IVs, which does not require a post hoc (remember you just look at them).

1. Plots > if you want to see the data lined up, you can create plots. You move one to horizontal axis and one to separate lines and hit add.

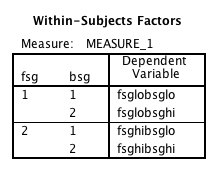


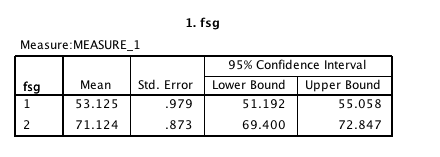
**Translating the output:**

1. On these types of designs, you will ignore:
   1. Multivariate tests
   2. Within subjects contrast
   3. Between subjects effects
2. Sphericity – totally useless in this case because there were only two level measurements for each IV.

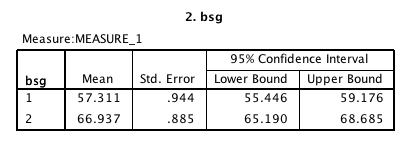


1. ANOVA box.
   1. You will get two main effects and an interaction – just like the between subjects. However, there are different error terms for each main effect and interaction. So you can pretend that each of the two rows is a small ANOVA. You would pick which correction line to use if you wanted to use a correction (they are all the same here because we only had two levels).
   2. The first main effect FSG is significant – *F*(1,157) = 965.99, *p*<.001, *ηp*2 = .86.
      1. Since there are only two levels you can use the means to see what that shows. Basically, people rate the hi FSG larger than the lo FSG (which is what we wanted them to do in the experiment).
   3. Don’t remember what the numbers mean? Go back to the very first box in the output.

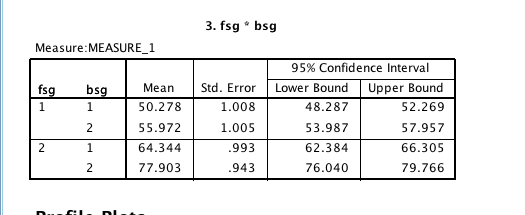


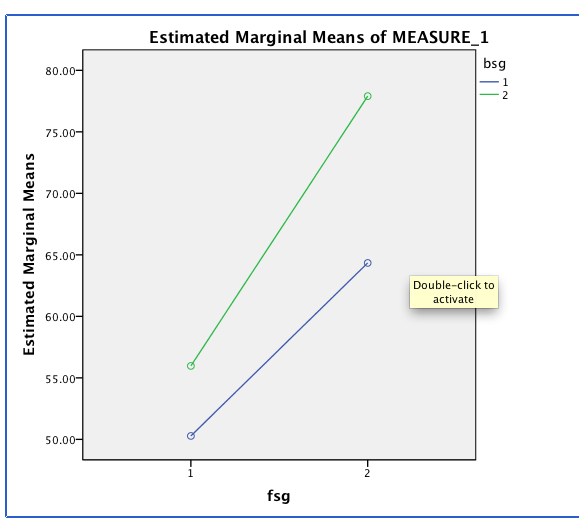


* 1. The second main effect was also significant BSG – *F*(1,157) = 378.40, *p*<.001, *ηp*2 = .71.
     1. Again, people rate the hi BSG larger than the lo BSG (not exactly surprising, but they weren’t supposed to use that information in the study).



* 1. The interaction was also significant, *F*(1,157) = 70.99, *p*<.001, *ηp*2= .31.
     1. Again, since the interaction is significant we are going to ignore follow up tests on the main effects (well we didn’t really have to do any, but if you would have …).



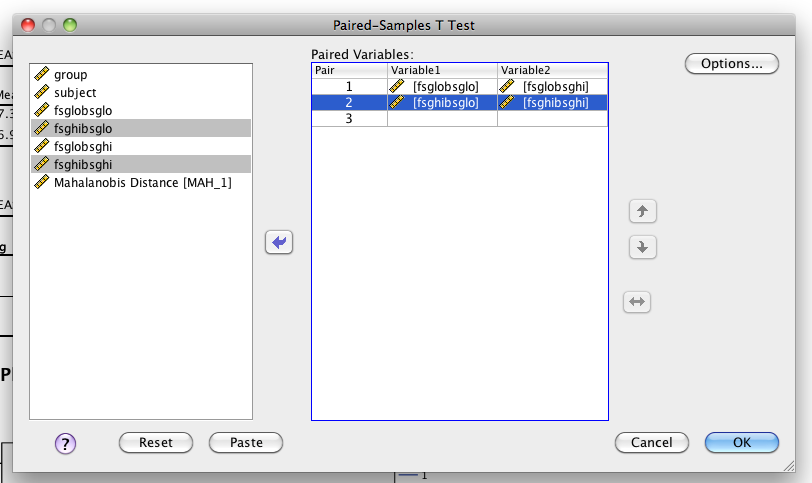


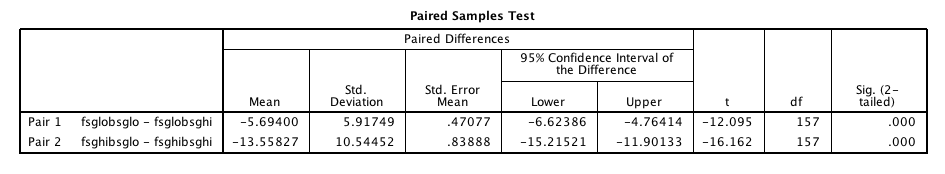
1. Post hocs! - Remember that follow up tests for only repeated measures is dependent t-tests.
   1. Make that chart!

|  |  |  |
| --- | --- | --- |
|  | **FSG LO** | **FSG HI** |
| **BSG LO** | M=50.28 | M=64.34 |
| **BSG HI** | M=55.97 | M=77.90 |

* 1. Now you can go down or across – just like with the between subjects tests.
  2. Do you want to compare BSG LO to BSG HI or FSG LO to FSG HI?
  3. I compared going down because I wanted to know what happened with the FSG ratings when BSG was amped up.

1. To run:
   1. Analyze > compare means > paired samples t-test
   2. Make sure you pair the right ones together.

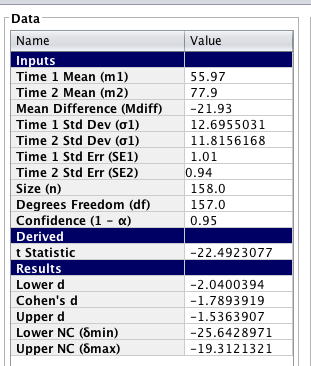




|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  | **FSG LO** | **FSG HI** | **p value** | **Explain** | **Cohen’s d** |
| **BSG LO** | M=50.28 | M=64.34 | <.001 | Significant | davg = 1.10 |
| **BSG HI** | M=55.97 | M=77.90 | <.001 | Significant | davg = 1.79 |

**Effect size:**

You would use either the dependent t averages or the dependent t sd differences to calculate these values in MOTE.



**Charts:**

Two-way repeated measures ANOVAs are one of the problems with chart builder in SPSS. You will have to just fake it, so to speak (or use Excel).

Graphs > chart builder.

Pick bar on the left hand side (we still have distinct levels, so you want bar graphs).

Pick simple bar (this is the fake it part).

Put all the conditions into the y axis – remember you’ll get the red plus sign. Put nothing in the x axis, it will auto create a variable for you.

Be sure to create error bars, x and y axis labels.

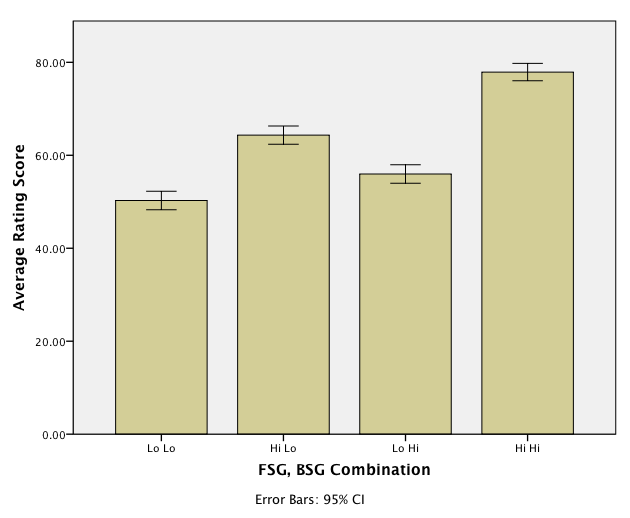
Write up example:

**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 actual rating (forward strength: FSG) and the strength of the reverse rating (backward strength: BSG). Data was screened for assumptions and outliers. One multivariate outlier was found using Mahalanobis distance as a criterion, but was included in the analysis with no discernible effect on the outcome.

Figure one shows the mean ratings for each combination of low and high forward and backward strength. FSG condition had a significant effect on ratings, *F*(1,157)=965.99, *p*<.001, *ηp*2=.86, which indicated that participants could correctly rate low related pairs (*M*=53.12, *SE*=.98) lower than high related pairs (*M*=71.12, *SE*=.87). However, BSG condition had a significant effect on ratings, *F*(1,157)=378.41, *p*<.001, *ηp*2 =.71. Participants again were able to rate low related pairs (*M*=57.31, *SE*=.94) lower than high related pairs (*M*=66.94, *SE*=.89). Finally, the interaction between FSG and BSG was significant, *F*(1,157) = 70.99, *p*<.001, *ηp*2 = .31. Dependent t-tests were used for the post hoc analysis of the interaction between FSG and BSG. BSG had an approximately 5 point effect on ratings for low FSG pairs, *t*(157)=12.09, *p*<.001, *davg =* 1.10. However, for high related FSG pairs, BSG increase from low to high also increased ratings about 13 points, *t*(157)=16.16, *p*<.001, *davg =* 1.79. See Figure 1 for the interaction.

(\*\*note I could have also used mean differences here instead of each mean and standard error separately, both are options – just be consistent\*\*)



*Figure 1.*

# Complete Example (Mixed Factorials)

We knew that this high BSG thing had increased ratings (that is not what you are supposed to do in the experiment, tied to overestimating how well you know something … bad for studying). So, we gave people instructions on how to ignore the BSG. Did it help? (rm 2 anova.sav)

**IVs:**

Between subjects: Group – JAM (regular judgment) versus Debias (instructions on how to do well in experiment).

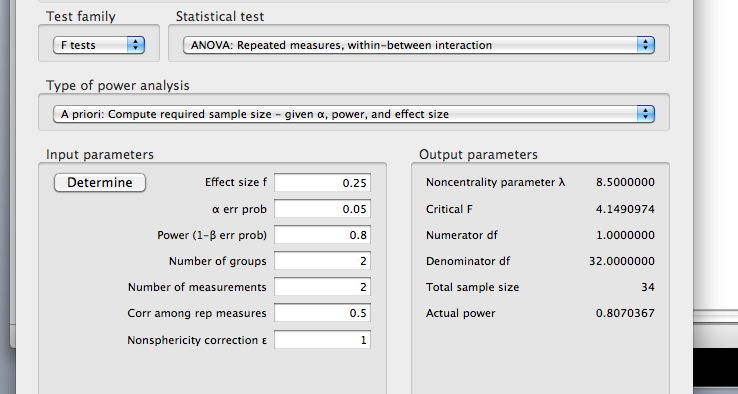
Repeated Measures: BSG lo versus BSG hi (when FSG is hi only, ignored the FSG lo ones)

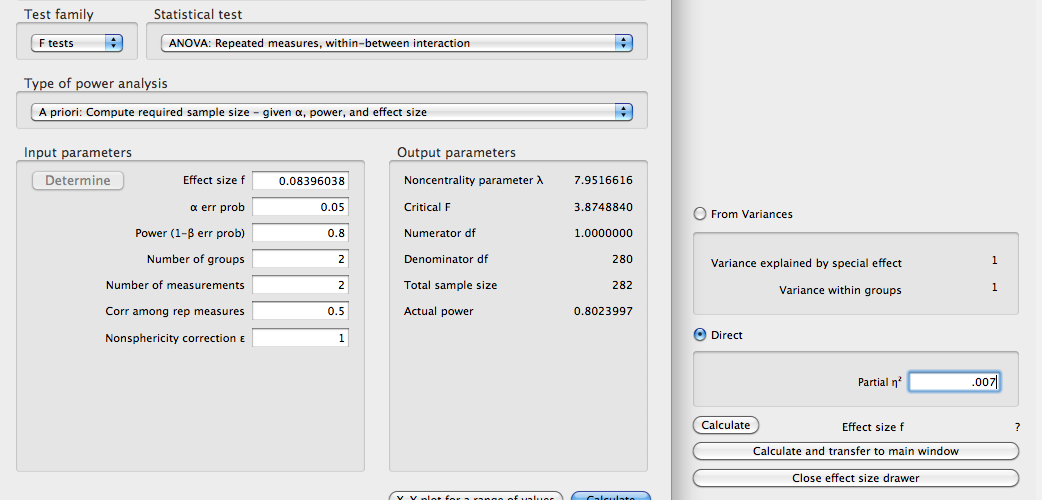
**DV:**

Ratings of those word pairs.

**Power:**

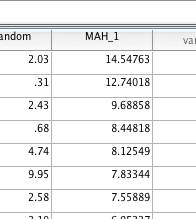
* F-test
* ANOVA: Repeated measures, within-between interaction (usually you are looking for the interaction, so you’ll use this page to estimate the number of people needed for that test).
* A priori
* Effect size f – you can determine from previous study (see below, since we had an idea from running the last analysis), or estimate (hover over the box to see options).
* Alpha = .05
* Beta = .80
* Number of groups = number of levels of between subjects
* Number of measurements = number of levels of repeated measures
* Corr among rep measures – estimate a correlation between the repeated level measurements (you can estimate from previous research or the first couple participants … go with the lowest correlation you find … .5-.7 is a good estimate if you are giving them the same test a couple times).
* Nonsphericity correction = epsilon … 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).



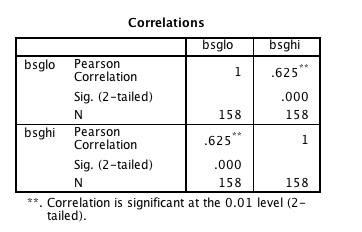


**Assumptions:**

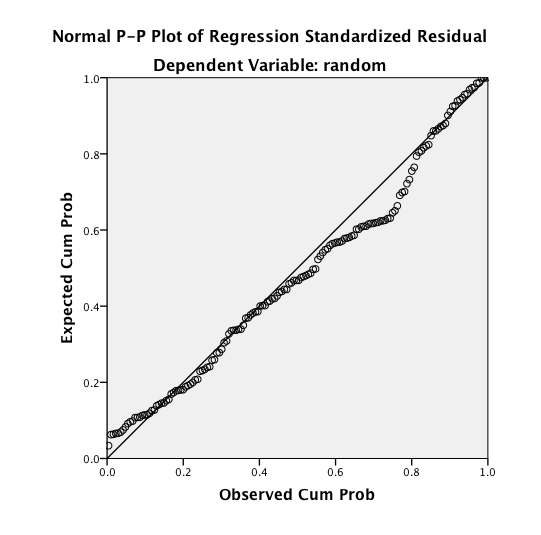
1. Accuracy and missing data are same as data screening.
2. Outliers
   1. Univariate
      1. Analyze > descriptive statistics > descriptive
         1. Move over ALL repeated levels
         2. Be sure to hit save standardized values.
         3. Hit options.
         4. Skewness and Kurtosis (for normality later).
         5. Checking my z-scores, I don’t see any univariate outliers.
   2. Multivariate
      1. First make a fake variable.
         1. Transform > compute
         2. Type in a variable name (random)
         3. Use a random variable creator for the numeric expression box (rv.chisq(7)).
      2. Run a fake regression.
         1. Analyze > regression > linear
         2. Put your random variable in the dv box, and your repeated levels in the IV box.
         3. Under plots > zpred in Y, zresid in X, click both histogram and normal probability plot.
         4. Under save > Mahalanobis
      3. Check out the mahalanobis scores
         1. 2 repeated measures = cut off *p*<.001, 13.82
         2. One multivariate outlier
            1. As you can see below, it does seem to be causing us some problems with normality and homogeneity.



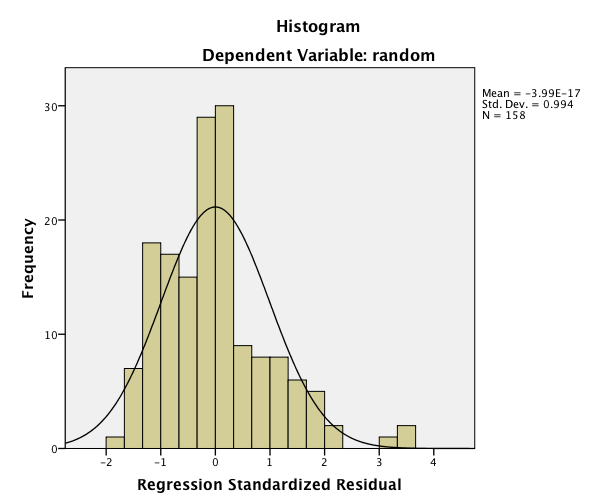
1. Multicollinearity – remember for repeated measures, you basically don’t want perfect correlations (*r* = .999 – 1).
   1. Analyze > correlate > bivariate.
   2. Move over all the repeated levels
   3. Check out the correlation table.



1. Linearity – looks ok!



1. Normality – in this analysis you really want to check multivariate, but again you can check the individual ones using frequencies or skew/kurtosis values.

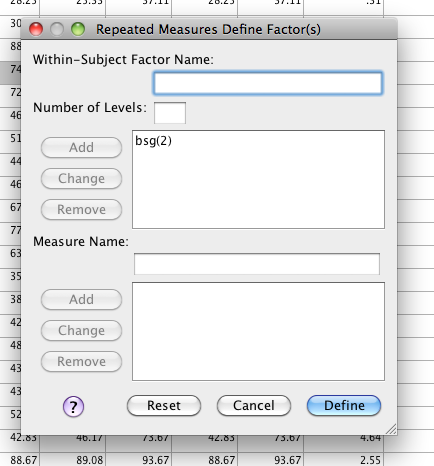


1. Homogeneity – this chart looks fine too.

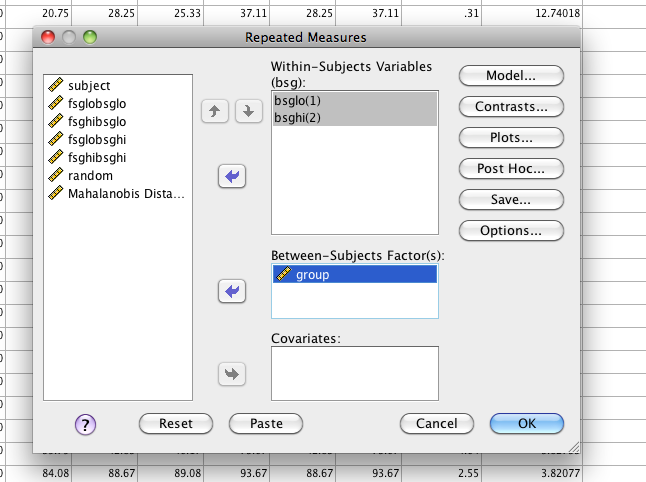


**Running the test:**

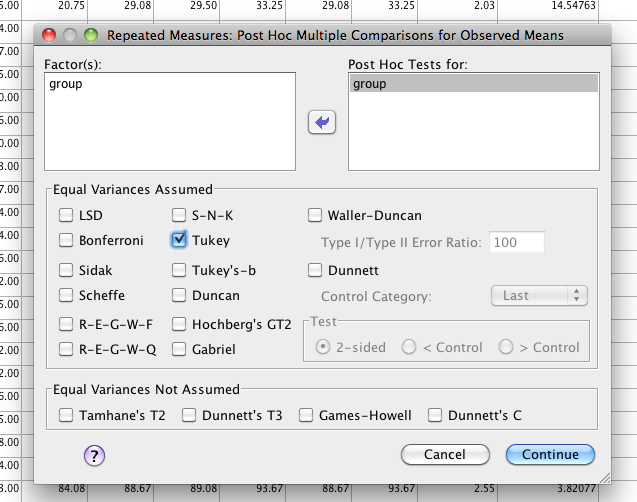
1. Analyze > general linear model > repeated measures
   1. Anytime you have a RM variable, you always run it as RM, even if the other variables are between subjects
   2. Put in your RM variables and levels



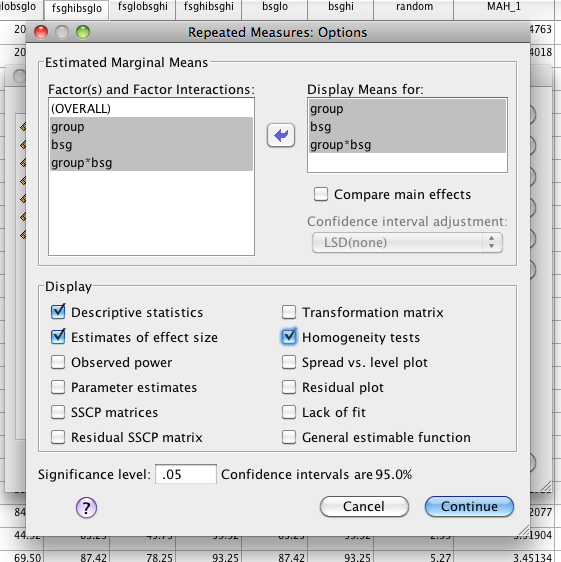
* 1. Move over the RM variables



* 1. Under post hoc > move over the group and ask for Tukey (really we won’t need this because we only have two levels, but it won’t hurt to get used to asking for that output).



* 1. Options > move over the means, descriptives, effect size, and homogeneity (since we have a between subjects variable)

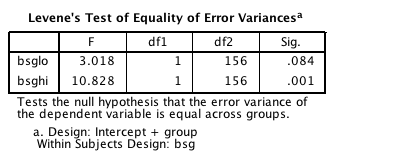


**Translating the output:**

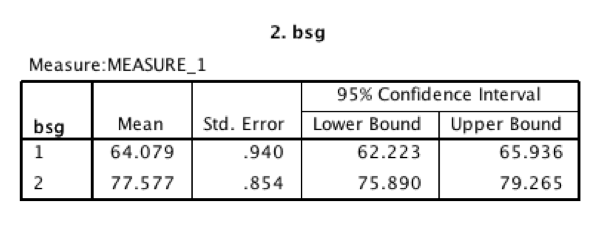
1. Remember to ignore: multivariate output, contrasts. You will need the between subjects box this time!
2. Sphericity – same problem as before, not enough repeated measures levels.



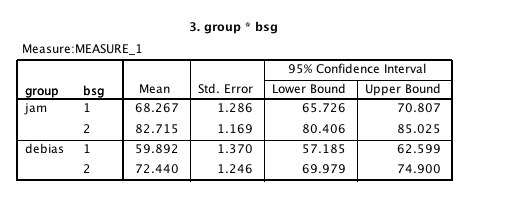
1. Levene’s – both are good, remember you do not want it to be less than .001 (splitting hairs, I know, but also removing the outlier would have helped.



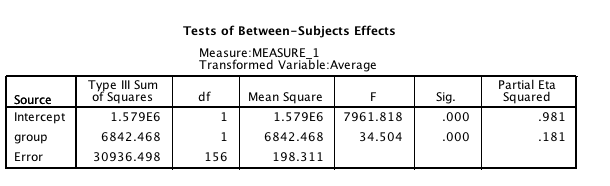
1. The repeated measures ANOVA box: gives you anything with RM levels/conditions involved, so you get both the RM main effect AND the interaction for the two variables. 
   1. There’s a repeated measures main effect of BSG: *F*(1, 156) = 258.34, *p*<.001, *ηp*2 = .62.
      1. The low BSG word pairs (*M=*64.08, *SE*=.94) were rated lower than the high BSG word pairs (*M*=77.58, *SE*=.85).

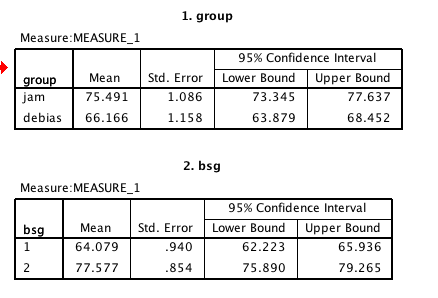


* 1. Does not appear to be an interaction, *F*(1, 156) = 1.28, *p* = .26, *ηp*2= .01.



1. The between subjects ANOVA box: The second box shows you the between subjects effects only.
   1. There was also a significant main effect of group, *F*(1,156)=34.50, *p*<.001, *ηp*2=.18.
   2. The debiased group (*M=*66.17, *SE*=1.09) significantly lowered their ratings compared to the control group (*M*=75.49, *SE=*1.16).





1. Pretending to do a Post Hoc: Two Options (with a chart!)

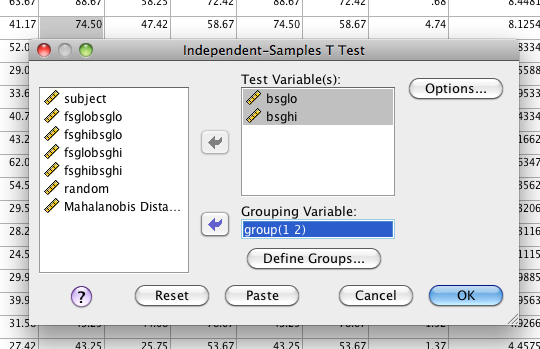
|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  | **BSG LO** | **BSG HI** | **p value** | **Explain** | **Cohen’s *d*** |
| **Jam** | 68.27 | 82.72 | <.001 | Significant | ddiff = 1.41 |
| **Debias** | 59.89 | 72.44 | <.001 | Significant | ddiff = 1.16 |

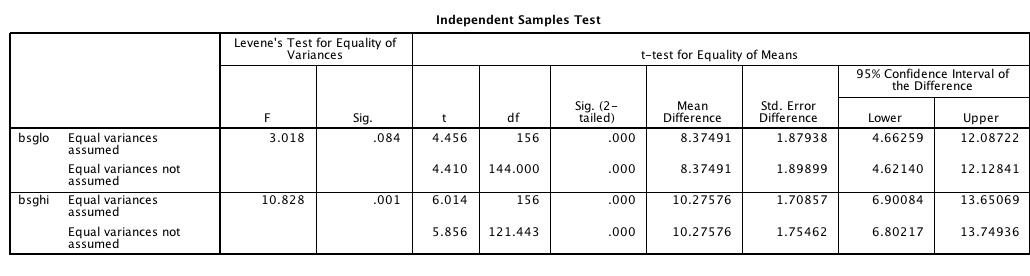
Choose a Post Hoc test based on the following:

1. Hypothesis
2. Smaller number of levels to be run (you can compare down or across)

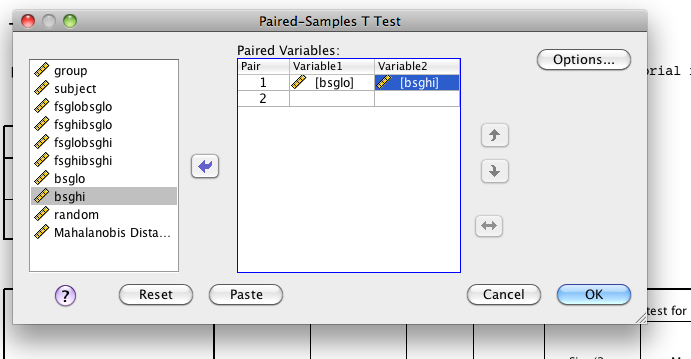
h

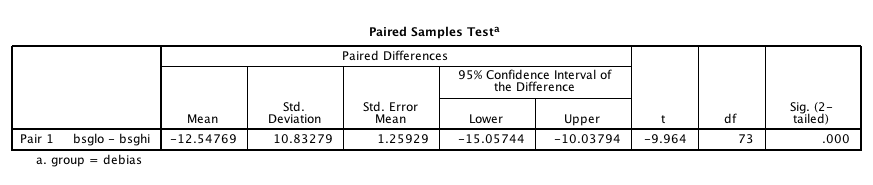
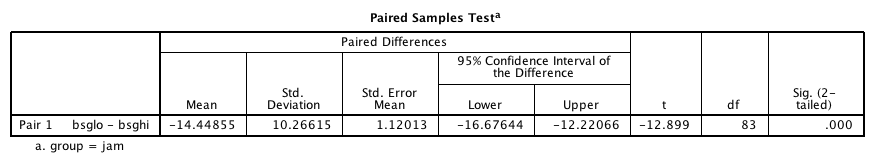
* 1. Comparing down:
     1. I want to compare the control and experimental group for both low and high BSG.
     2. So I’m comparing the BN levels using each RM level separately
     3. This way would be independent t-tests because each comparison is different groups of people.
     4. I.E. each person only got one set of instructions.
     5. Analyze > compare means > independent t-test
     6. Move over the repeated measure levels
     7. Enter and define your groups in the grouping variable
     8. Both of these tests are significant!





* 1. Comparing across:
     1. Instead let’s say I want to compare BSGlo versus hi for each instruction level (BN) separately.
     2. So I’m comparing the RM levels using each BN level separately.
     3. Easiest way to run this is dependent t-test.
     4. However, if you just run that test, it’ll give you both control and experiment groups together as one.
     5. Split file first (data > split file on the between subject variable)
     6. It’ll run each group separately.
     7. Analyze > compare means > paired samples t-test
     8. Move over the two repeated measures levels to compare them.
     9. Both tests are significant!





**Effect size:**

As shown in earlier examples, you can use the outputs from either the independent or dependent test entered in MOTE to determine the *d* values for each test. Be sure to match the post hoc test to the right *d* test.

**Graphs:**

Graphs > chart builder.

We still have distinct levels/conditions, so we are still going to use a bar chart.

Click bar chart on the bottom left.

This graph would be a clustered bar graph.

Put both repeated measures levels into the y axis (remember red cross will appear, and then you’ll get a pop up window asking about the combined variable).

Put the between subjects variable into set color.

Include error bars, appropriate x and y axis labels.

See figure below.

Write Up Example:

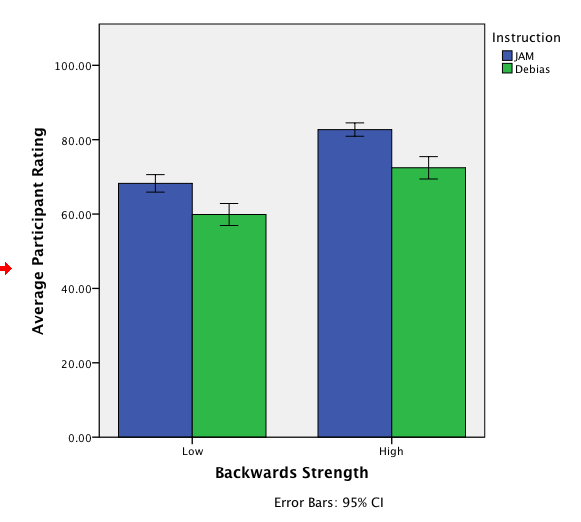
**Results**

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 multivariate outlier was found using Mahalanobis distance as a criterion, but was included in the analysis.

A significant main effect of backward strength was shown, *F*(1, 156) = 258.34, *p*<.001, *ηp*2= .62. The low BSG word pairs (*M=*64.08, *SE*=.94) were rated lower than the high BSG word pairs (*M*=77.58, *SE*=.85). Group participation also affected ratings, *F*(1,156)=34.50, *p*<.001, *ηp*2=.18, in which the debiasing group (*M=*66.17, *SE*=1.09) significantly lowered their ratings compared to the control group (*M*=75.49, *SE=*1.16). However, there was not an interaction between ratings and group participation, *F*(1, 156) = 1.28, *p* = .26, *ηp*2 = .01. 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 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*(83) = 12.90, *p*<.001, *ddiff* = 1.41. The debiasing group saw a smaller difference in ratings of about 12.5 points, *t*(73) = 9.96, *p*<.001, *ddiff* = 1.16.



*Figure 1.*