Multidimensional Scaling

**Description:**

Multidimensional Scaling (MDS) is a picture type analysis, where you can see the underlying pattern to the variables from your data set. You usually scale the variables (whereas with cluster analysis you can cluster by person), similar to a factor analysis. You will end up with a picture of the data and how closely variables are related – variables that measure the same information or similar information will group closely together in space. Data can be of any type.

**Definitions/Abbreviations:**

1. Data types:
   1. Count – frequency data, a simple count of the number of times thing occur.
   2. Categorical/Binary – yes/no, off/on type of data. Usually labels like gender.
   3. Interval – continuous data, were zero is a placeholder and does not mean nothing. For example, Likert based scales.
   4. Quantitative/Ratio – things that are continuous and zero means nothing.
2. Distancing:
   1. Most common type is Euclidean distance – especially for interval and above type data. This distance measure is like looking for a hypotenuse in geometry/algebra.
   2. Chi-square is the most common measure for count data.
   3. Binary/Categorical data has many options, and they are all about the same.
3. Models:
   1. Generally, people run 1 dimension and 2 dimension models. 3 dimension models are also popular, but hard to interpret the picture of the data.
      1. You want your R2 values to be high. Usually people will justify a 2 dimension model over a 1 dimension model because of a large jump in R2.
      2. R2 is really the only fit index for MDS, unlike other forms of picture-data interpretation (EFA).

**Assumptions:**

1. Scaling – the data should be in fairly similar scales - (for example, one variable is measured in dollars and the other variable is measured in years), consider standardizing them. You can standardize automatically in SPSS.
2. Data types – make sure you select the right options in SPSS to match the type of data you have.

# Complete Example

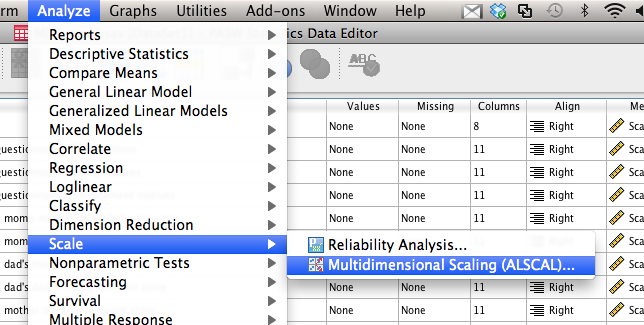
**Research Question: Is there a pattern of relationship between alcohol question answers for criminals?**

Dataset – use only the drug and alcohol related questions.

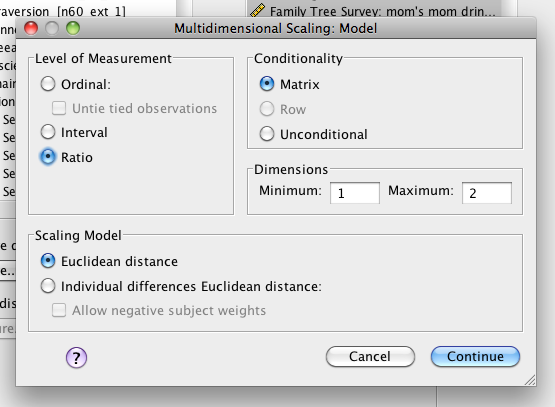
**Assumption Checks:** The drink motivation and SAD scales appear to have ratio type values. They are mildly continuous with big jumps (higher drug/alcohol use) and lots of zeros. Although they are not totally on the same scale – they are not so incredibly different we should standardize. We should use ratio/interval scale distancing for the analysis.

How to Run Analysis:

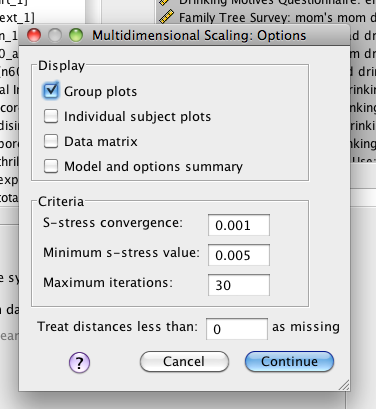
1. Analyze > Scale > Multidimensional Scaling.

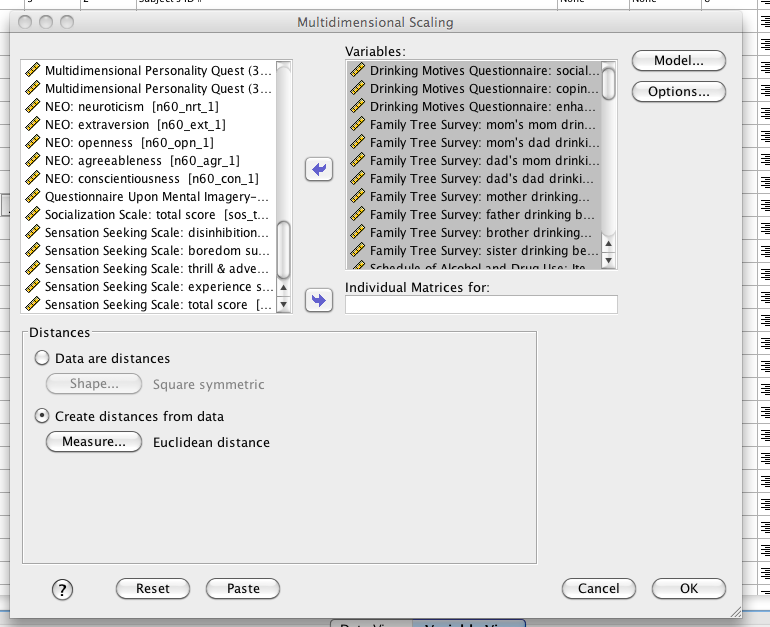


1. Move over the variables you want to use.
2. Hit model.
   1. I’m going to change the level of measurement to ratio because our zeros here mean no use. Ordinal would be for count/categorical data, interval for likert type scales.
   2. I’m going to change the dimensions to 1 and 2, so I can get both on the same output.
   3. Hit continue.

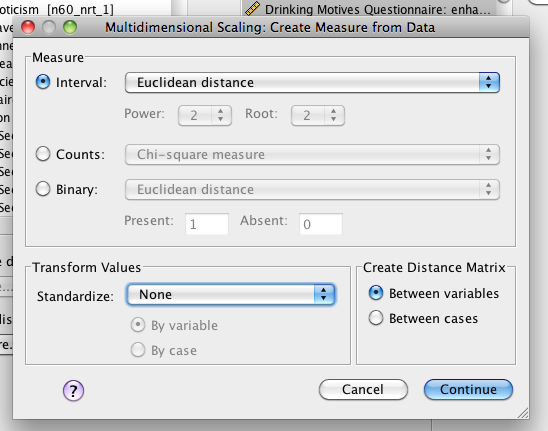


1. Hit options.
   1. Ask for group plots.





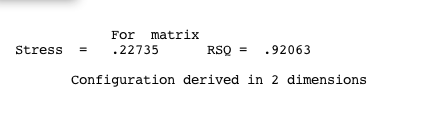
1. In the bottom left, you will most likely want to change this to “create distances” (see above). Hit measure.
   1. In the top, you can change to binary, interval, or count data. You will use interval for ratio or interval. The most common is Euclidean.
   2. In the bottom, you can change to standardize variables that have very large scale distances.

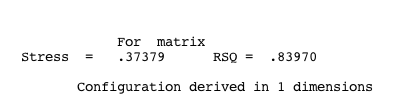


**Reading the Output:**

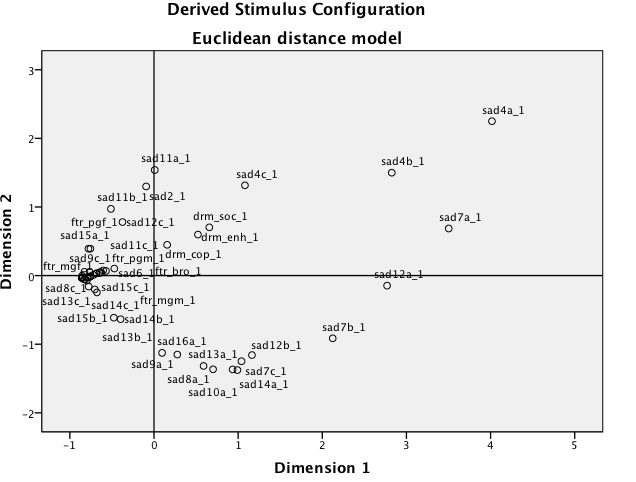
Note: This guide covers the output you need.

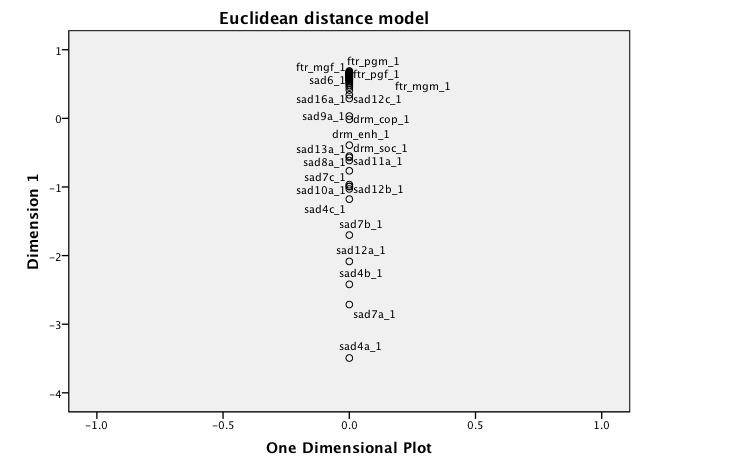
1. Look for R2 for both your models in the ASCAL output.





1. Here we are getting a 9% increase in R2 for including a second dimension in the data. If that is explainable, you could say it was the best model. Or describe both.
2. Next, people use the pictures created in the output.





1. How do you interpret these?
   1. Things that are close together in space are grouped together/have the same variance/probably measure the same construct.
   2. Generally with 2 dimensional graphs, people will describe:
      1. Left versus right
      2. Top versus bottom
      3. OR each quadrant individually.

**Example Write Up:**

Results

Multidimensional Scaling was used to analyze the Drinking Motives Questionnaire, the Family Tree Survey of family alcohol and drug use and the Schedule of Alcohol and Drug (SAD) Use scale for individual use scores. Since the scores were ratio scales, Euclidean distance was analyzed between the variables. One dimension and two dimension models were produced with the two dimensional model (R2= .92) having better fit than the one dimensional model (R2 = .84). Figure 1 depicts the two dimensional model. Generally, family drug use tended to group with hard drug use frequency on the SAD scale (crack, cocaine, barbiturate, heroin, etc.). Drinking motives were in the same quadrant as drink frequency questions on the SAD. Many of the lifetime and last 12 months (questions a and b for each drug use) questions on the SAD also grouped together in a quadrant. This model seemed to indicate that questions about overall drug use (lifetime, past years), hard drug use, and drinking motivations separated into distinct groupings in the data.

*Figure 1*. (this would be the 2D scale picture pasted above).