Example of MDS Analyses

Step 1 -- Preparing the Sorting Data
After the sorting is finished, the data are collected into a matrix like that below.

- There are 17 columns col1 = stim1 (pine bark), col2 = stim2 (brick), etc.
- Each row is one person's sorting result
- Stimuli that were sorted into the same group are given the same number (e.g., 1st person sorted stim 1, 4, & 13 into the same group)
- The group numbers don't have to be consistent across person's sorts

```
3 6 7 3 2 4 7 1 2 5 5 5 3 6 4 1 7
2 2 2 2 2 3 7 1 1 5 4 6 2 1 3 7 7
. . . . . . . . . . . . . . . . .
. . . . . . . . . . . . . . . . .
2 1 3 3 5 6 5 2 4 1 4 4 1 2 6 2 2
```

These data are input to a program that constructs a composite dissimilarity matrix

- In a dissimilarity matrix, larger numbers mean that the two stimuli are more dissimilar
- For each person, the program determines whether or not each pair of stimuli were sorted into the same group. A "0" means they two stimuli were sorted together (are similar), while a "1" means they were not sorted together (are different).
- The program accumulates these representations of the sorting data across persons into a dissimilarity matrix.
  - The smallest possible value -- 0 -- would mean that everybody sorted that pair of stimuli into a group
  - The largest possible value -- N -- would mean that no one sorted that pair of stimuli into a group.
- This process can also be done "by hand" -- don't!!! I'll be happy to give you the program!
- The resulting dissimilarity matrix is shown in the SPSS ALSCAL program below.

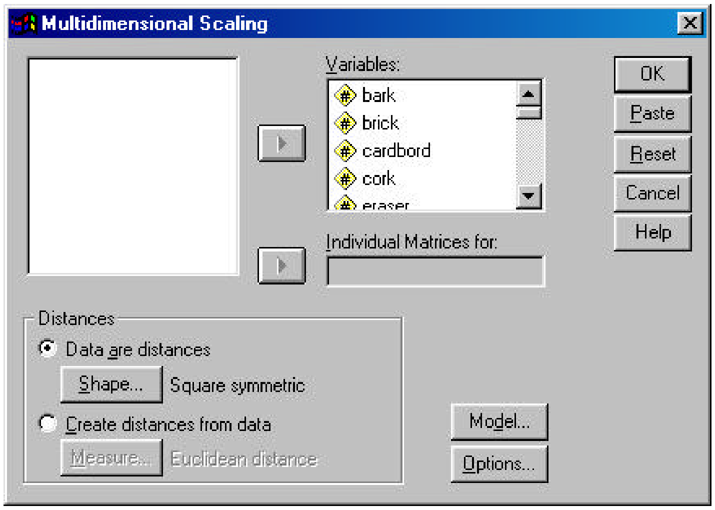
Step 2 -- Getting the Composite MDS solution

The resulting matrix is put into SPSS -- notice that this looks very different from other data sets you've worked with!! The only analysis you can do with these data is MDScaling.

```
inner surface of pine bark    0
brick                         22  0
cardboard                     23 27  0
cork                          24 27 18  0
rubber eraser                 26 27 19 15  0
felt                          27 29 28 28 28  0
leather wallet                26 29 23 25 20 24  0
rigid plastic sheet           23 28 24 26 27 28 22  0
very fine sandpaper           24 16 24 28 24 29 28 27  0
nylon scouring pad            23 18 29 28 27 26 28 29 21  0
cellulose kitchen sponge      23 28 27 20 24 26 27 29 24 22  0
woven straw                   18 25 28 27 25 29 26 27 26 16 19  0
block of styrofoam            23 24 21 10 19 28 25 25 25 25 21 26  0
unglazed ceramic tile         21 10 26 26 24 29 29 25 12 24 26 26 25  0
velvet                        28 29 28 28 29  4 24 29 29 27 27 28 29 29  0
wax paper                     24 28 24 28 24 28 21 12 29 29 29 27 26 28 27  0
glossy painted wood           22 27 23 29 28 29 20 13 27 28 27 25 29 26 26 12  0
```

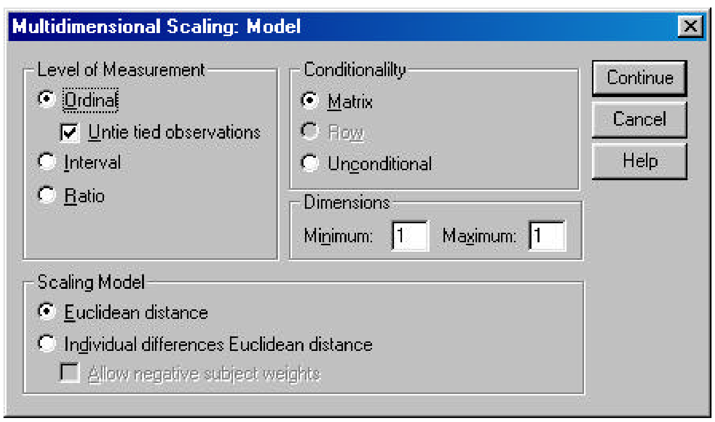
Example of a Composite MDScaling Analysis

Analyze → Scale → Multidimensional Scaling



Move the stimulus variables into the window

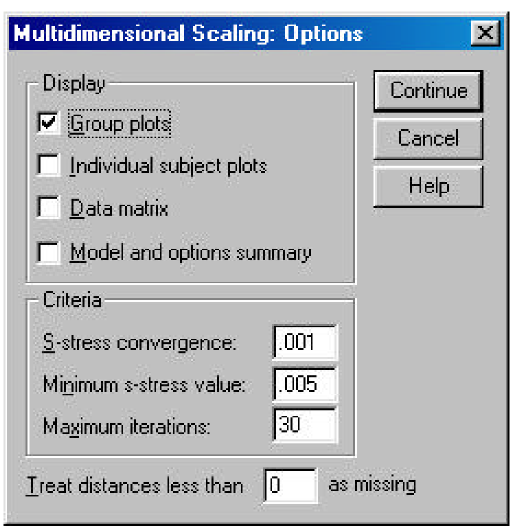
Use the "Model" and "Option" windows to select the analysis you want.



Ordinal analyses with the "Untie" option are the most common and usually produce the most replicable results

While SPSS will perform 1-6 dimensional solutions, the output is "unruly". It often works better to get each dimensional analysis separately.

Usually we will want analyses in 1-6 dimensions, so we can make the scree plot. However, with smaller stimulus sets you might not be able to get larger solutions -- sometimes 1-3 is all the program can provide (and it will warn you about the small number of stimuli involved).



Be sure to click the "group plots".

Occasionally an MDS solution won't converge -- this is where to increase the number of iterations.

SPSS MDS Output

Iteration history for the 3 dimensional solution (in squared distances)

Iteration	S-stress	Improvement
1	.31365	
2	.25158	.06206
3	.24603	.00556
4	.24530	.00072

Iterations stopped because
S-stress improvement is less than .001000

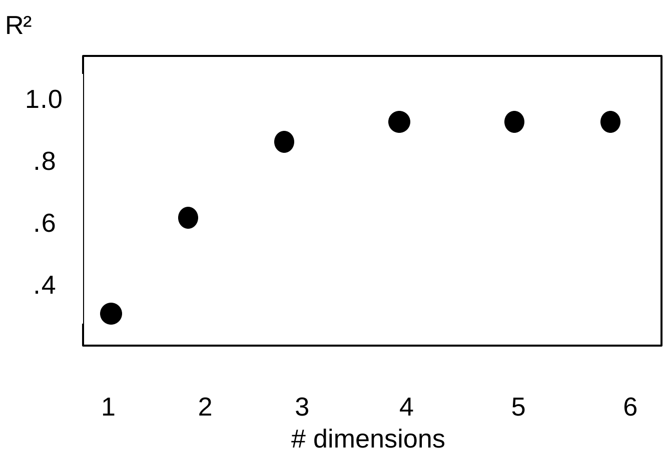
Stress and squared correlation (RSQ) in distances

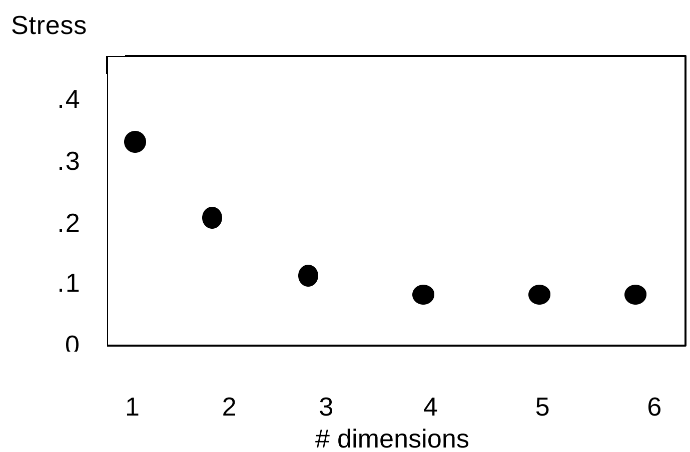
RSQ values are the proportion of variance of the scaled data (disparities)
in the partition (row, matrix, or entire data) which
is accounted for by their corresponding distances.
Stress values are Kruskal's stress formula 1.

For matrix
Stress = .13077 RSQ = .89894 ← fit indices for the solution

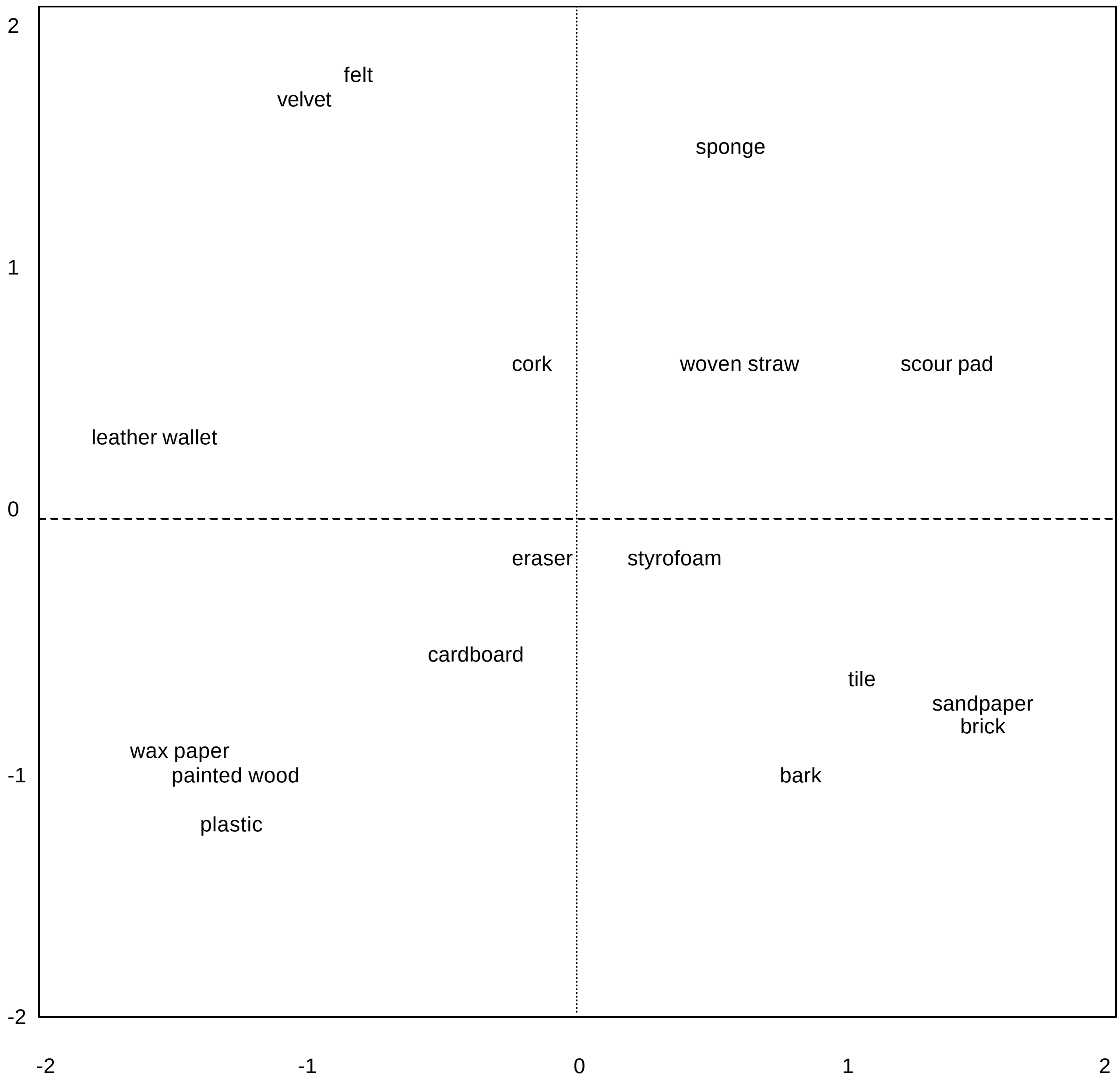
Stimulus Coordinates — Configuration derived in 3 dimensions

Stimulus Number	Stimulus Name	Dimension 1	Dimension 2	Dimension 3		
1	S101	.4310	-1.0150	.7713	1	inner surface of pine bark
2	S102	1.4625	-.8565	.5499	2	brick
3	S103	-.5619	-.5124	-1.4518	3	cardboard
4	S104	.0939	.4911	-1.6814	4	cork
5	S105	-.0484	.2154	-1.6038	5	rubber eraser
6	S106	-.7855	1.7860	.7767	6	felt
7	S107	-1.6664	.2428	-.1925	7	leather wallet
8	S108	-1.3502	-1.1892	.2552	8	rigid plastic sheet
9	S109	1.4719	-.7621	.0604	9	very fine sandpaper
10	S110	1.3014	.5978	1.0184	10	nylon scouring pad
11	S111	.8247	1.4068	-.3880	11	cellulose kitchen sponge
12	S112	.8557	.6318	1.2145	12	woven straw
13	S113	.3874	.2774	-1.5097	13	block of styrofoam
14	S114	1.2491	-1.0667	.1517	14	unglazed ceramic tile
15	S115	-.8749	1.6053	.9492	15	velvet
16	S116	-1.5241	-.8947	.3250	16	wax paper
17	S117	-1.2660	-.9578	.7549	17	glossy painted wood





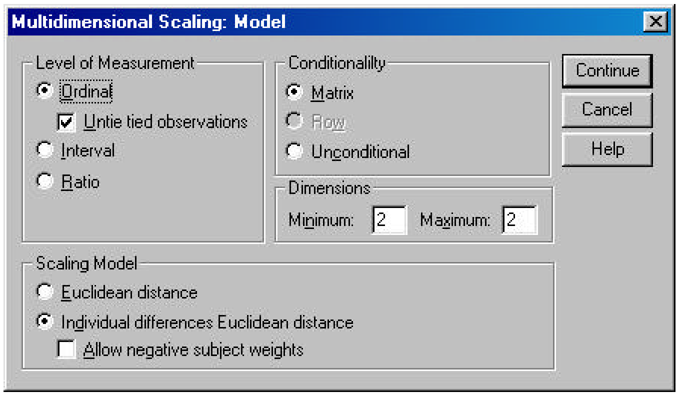
Both of the scree plots suggest a 3-dimensional solution -- there is substantial improvement in fit with the 2nd and 3rd dimensions are added, but no further improvement when the 4th through 6th are added.



Dimension 1 - horizontal
Dimension 2 - vertical

Example of Individual Differences Scaling

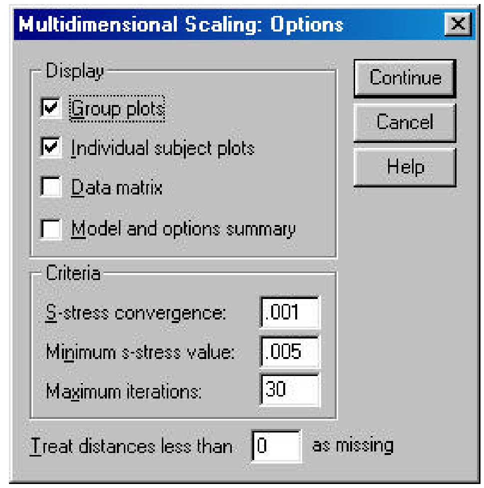
We start with a separate dissimilarity matrix for each participant. For these data, each was a 17x17 matrix. Because free-sorting procedure was used, each matrix includes just "0"s (for pairs grouped together) and "1"s (for pairs not grouped together).



Ordinal/untied solutions are also very good with INDSCAL.

Be sure to select "Individual differences.." under Scaling Model

INDSCAL solutions must have at least 2 dimensions



Request the Individual subject plots (gives you the "subject space"), as well as the group plots (gives you the "stimulus space").

SPSS forms a composite matrix (by adding). The composite matrix is the basis for the initial solution in the prescribed dimensions, all further iteration is done to maximize fit of the mds solution to the individual dissimilarity matrices.

```
Iteration history for the 2 dimensional solution (in squared distances)

                  Young's S-stress formula 1 is used.
                Iteration     S-stress      Improvement
                    0          .22635
                    1          .23423
                    2          .20612         .02811
                    3          .20070         .00542
                    4          .19872         .00198
                    5          .19773         .00099       Iterations stopped because
                                                         S-stress improvement <  .001

            Stress and squared correlation (RSQ) in distances

RSQ values are the proportion of variance of the scaled data (disparities)in the data
which is accounted for by their corresponding distances.

             Stress values are Kruskal's stress formula 1.

        Matrix      Stress      RSQ    Matrix      Stress      RSQ
          1          .137      .845       2          .231     .431
          3          .188      .810       4          .133     .951
          5          .111      .943       6          .195     .860
          7          .161      .846       8          .122     .912
          9          .196      .855      10          .143     .959

        Averaged (rms) over  matrices
    Stress  =   .15408      RSQ =  .81122
```

SPSS returns goodness of fit indices for each participant -- things to notice…

- **Relative fit among the participants** -- notice that #2 has a much poorer fit to the composite than the others
  - This suggests they were "using different attributes" than the rest of the group when they were doing their sorting (or whatever data collection procedure was used)
  - This participant's data would probably be excluded from the composite as an "outlier" (and like any other such exclusion, we would want to know more about them and their data - they might be a member of an interesting "subpopulation" or they might just have hurried through the task
- **Fit compared to the composite solution**
  - If individual's data are fit substantially more poorly than is the composite matrix from "regular" mdscaling, there is the chance that the composite matrix is an "average that represents no one" (consider a strongly bimodal distribution -- there is a mean, but it represents almost no one)
  - Because of the "sparse" data in individual matrices (only values of 0 & 1), there will always be a drop in fit -- but is should change more than 1-2 values in the first decimal
  - It can also be useful to compare the composite and indscal solutions -- major differences also suggest that the composite is a "misrepresentative aggregate"

SPSS also shows the scaling solution -- called the "stimulus space" -- both the coordinates and the corresponding plot or "map" (not shown here to save paper -- it was virtually identical to that from the composite scaling)

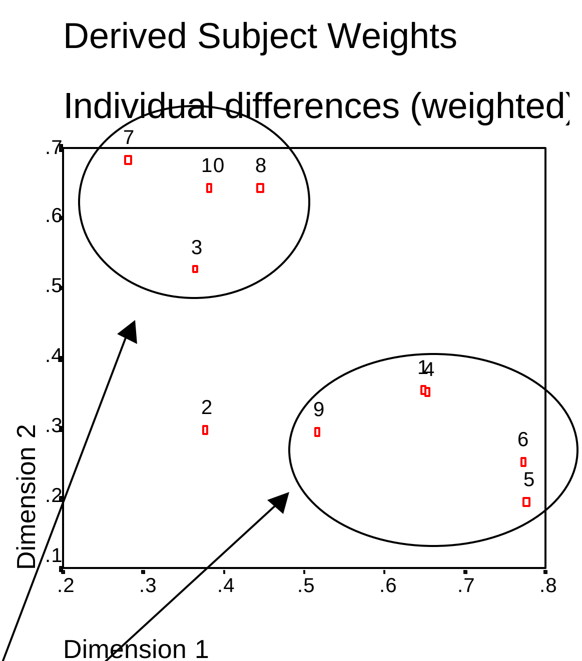
SPSS then gives the "subject weights" -- called the "subject space" -- and the corresponding graph

```
Subject weights measure the importance of each dimension to each subject.

A subject with weights proportional to the average weights has a weirdness of
zero, the minimum value.
A subject with one large weight and many low weights has a weirdness near one.
A subject with exactly one positive weight has a weirdness of one,
the maximum value for nonnegative weights.
```

		Dimension	
Subject Number	Weird-ness	1	2
1*	.2467	.6474	.3542
2	.0183	.3772	.2984
3^	.3464	.3647	.5267
4*	.2555	.6533	.3521
5*	.6181	.7775	.1958
6*	.5132	.7722	.2528
7^	.5894	.2805	.6836
8^	.3466	.4451	.6430
9*	.2196	.5172	.2961
10^	.4288	.3806	.6437

Overall importance of each dimension: .3011 .2101



Participants who seem to "favor" dimension 2 (hard vs. soft)

Participants who seem to "favor" dimension 1 (smooth vs. rough)

Notice subject #2 has small values for both dimensions -- isn't "using" either dimension

Notice that in this solution there is no one who seems to be using the two "attributes" equally -- each favors one or the other.