

Multidimensional Scaling

Description of MDS

- *Multidimensional scaling* (MDS) is a mathematical dimension reduction technique that maps the distances between observations in a high dimensional space into a lower (for example, two) dimensional space.

How does MDS Work?

- MDS attempts to locate the n observations in a reduced dimensional space so that the differences between pairs of points in this reduced space match, as closely as possible, the true-ordered differences between the observations.

Recommended Distance Measure

Standardized Euclidean Distance

$$d_{2s}(\mathbf{x}_r, \mathbf{x}_s) = \left[(\mathbf{z}_r - \mathbf{z}_s)' (\mathbf{z}_r - \mathbf{z}_s) \right]^{\frac{1}{2}}$$

Matching Distances in MDS

Let $D_{r_1 s_1}$ = distance between the 2 closest points

$D_{r_2 s_2}$ = distance between the next 2 closest points

⋮

$D_{r_{N(N-1)/2} s_{N(N-1)/2}}$ = distance between the 2 farthest points

Matching Distances in MDS

Consider plotting the points in a 2-dimensional space. Let the distance between the r^{th} and s^{th} points in the 2-D space be d_{rs} .

Let $d_{r_1 s_1}$ = distance between the 2 closest points

$d_{r_2 s_2}$ = distance between the next 2 closest points

⋮

$d_{r_{N(N-1)/2} s_{N(N-1)/2}}$ = distance between the 2 farthest points

Matching Distances in MDS

Multidimensional Scaling tries to locate the N points in a 2-D space so the the distances d_{rs} match as closely as possible the true ordered distances:

$$D_{r_1 s_1} < D_{r_2 s_2} < \dots < D_{r_{N(N-1)/2} s_{N(N-1)/2}}$$

How does MDS Work?

- Specifically, MDS attempts to locate the observed data in a reduced dimensional space so that the following equation is minimized:

$$E = \frac{\sum_{r=1}^N \sum_{s=1}^{r-1} (D_{rs} - d_{rs})^2}{\sum_{r=1}^N \sum_{s=1}^{r-1} D_{rs}}$$

The MDS Procedure

General form of the MDS procedure:

```
PROC MDS <options>;  
  VAR variables;  
  INVAR variables;  
  ID variable;  
  MATRIX variable;  
  BY variables;  
RUN;
```

PROC MDS Requires Distance Data

- The input data must be expressed as a square matrix of dissimilarities between the variables.

```
data city;  
  title 'Analysis of Flying Mileages Between Ten U.S. Cities';  
  input (atlanta chicago denver houston losangeles  
        miami newyork sanfran seattle washdc) (5.)  
        @56 city $15.;  
  datalines;  
    0  
    587 0  
    1212 920 0  
    701 940 879 0  
    1936 1745 831 1374 0  
    604 1188 1726 968 2339 0  
    748 713 1631 1420 2451 1092 0  
    2139 1858 949 1645 347 2594 2571 0  
    2182 1737 1021 1891 959 2734 2408 678 0  
    543 597 1494 1220 2300 923 205 2442 2329 0  
  ;
```

Multidimensional Scaling Plots

