# STAT 530/J530
# November 1st, 2005

Instructor: Brian Habing
Department of Statistics
LeConte 203
Telephone: 803-777-3578
E-mail: habing@stat.sc.edu



## Today

- A Few Lose Ends About Multidimensional Scaling
- Overview of Analyzing Grouped Data
- Hotelling's $T^2$



## A few notes on MDS

SAS and SPSS both approach scaling from the sense of minimizing a *stress* function (like the isometric scaling does). One of the commonly used stress functions is Kruskal's first stress formula:

$$Stress1 = \sqrt{\frac{\sum_{i<j}\left(d_{ij} - \hat{d}_{ij}\right)^2}{\sum_{i<j} d_{ij}^2}}$$

## What value?

Manly (1994) suggests that you want a value at least 0.10 or smaller while increasing dimensions to get below 0.05 is supect.

Lattin, Carroll, & Green (2003) cite a value of 0.10 as fair, 0.05 as good, and 0.02 as excellent. They also note that looking at the change in Stress can be useful as well.



## A Complication

Because both SPSS and SAS approach the scaling as a minimization problem, they won't return the classical solution without a little work. In SAS you can get the classical solution by fitting a number of dimensions equal to the number of variables… and then just take the first few.



## Another Complication

For non-metric scaling the difference in algorithms can make the actual coordinates appear different from each other.

However, the results from SAS and R will agree with each other in the sense of being the same except for possible "enlargement", rotation, and/or flipping.

## Discriminating Between Groups

In cluster analysis groups of observations are formed based on the values of the variables.

In other cases we know the group membership in advance.



## Research Questions

- Are the groups significantly different based on the values of the variables?

Multivariate Analysis of Variance (MANOVA) tests the null hypothesis that the mean vectors of different groups are equal against the alternate hypothesis that they differ.



## Research Questions

- What combination of the variables does the best job of distinguishing between the groups of observations?

Fisher's Linear Discriminant Analysis searches for the linear combination of variables that "does the best job" of distinguishing between the group

## An Alternate Approach

An alternate approach is to use multiple logistic regression, with the different group memberships as the response and the variables as the predictors.



## Hotelling's $T^2$

Hotelling's $T^2$ is the special case of MANOVA where there are only two groups (populations).

If $\mu_1$ and $\mu_2$ are the (qx1) mean vectors then the procedure tests:

$$H_0: \mu_1 = \mu_2$$



## Assumptions

1. The variables in each population are multivariate normal.
2. The two populations have the same covariance matrix.
3. The observations are independent.