STAT 530/J530
November 17th, 2005

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

Homework 8

The facial measurement data can be found at: face.txt. The description of the various measurements are at: facemeasure.pdf.

a) Comment on whether the assumptions for multivariate normality seem to hold for this data set or not. Indicate if you would trust the MANOVA results and why or why not.

Homework 8

b) Ignoring what you determined in part a, perform a MANOVA for this data set. Carefully state the null and alternate hypothesis, report the p-value, and give your conclusion. (Why doesn't it matter which of the four tests you use?)

Homework 8

c) Conduct Fisher's linear discriminant analysis for this problem. Give the formula for the first canonical discriminant function and give the best estimate for how accurate you think it will be. (Why is there only one discriminant function?)

Homework 8

d) One way to see how strongly the different variables are related to the discriminant function is to take the correlation between the values on the first discriminant function and each of the nine variables. Which variable seems to contribute most to being able to telling the genders apart, which seems to contribute the least?

Next

Tuesday 22nd: Homework 9 is due, final posted, optional topic, and course evals

Thursday 24th: Thanksgiving – No Class

Tuesday 29th: Optional topic continued

Thursday 1st: Homework 10 is due, ice cream field trip as penance for Homework 6 grade being late! With time for questions while we eat.

5:30pm Tuesday, December 6th – Final Exam is Due

Possible Topics

Confirmatory Factor Analysis

Hierarchical Models

Regression Trees

Loglinear Models

Item Response Theory

Canonical Correlation Analysis

Assume we have two sets of variables:
$x_1, x_2, \ldots x_{q1}$ and $y_1, y_2, \ldots. y_{q2}$

We could examine the relation ship between these sets of variables using the $(q_1+q_2)x(q_1+q_2)$ correlation matrix… but it is not clear how.

Canonical Correlation Analysis

The idea in canonical correlation analysis is to come up with new variables $u_1, u_2, \ldots$ and $v_1, v_2, \ldots$

where the u's are linear functions of the $x_1, x_2, \ldots x_{q1}$ that are orthogonal to each other and the v's are linear functions of the $y_1, y_2, \ldots. y_{q2}$ that are orthogonal to each other.

They are chosen to maximize the correlation between the u's and v's.