STAT 530/J530
November 8th, 2005

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

MANOVA

Multivariate Analysis of Variance (MANOVA) is designed to test:

$$H_0: \mu_1 = \mu_2 = \cdots = \mu_k$$

where the μ_i are the (qx1) mean vectors.

The assumptions are the same as before: independence, multivariate normality, and equal variances.

Four Statistics

If H is the between-group sum of squares and E is the within-group sum of squares (E), then

Wilk's Lamda: $\Lambda = \frac{|\mathrm{E}|}{|\mathrm{H}+\mathrm{E}|}$

Roy's Greatest Root: Max Eigen value of $E^{-1}H$

Lawley-Hotelling Trace: Trace of $E^{-1}H$

Pillai Trace: Trace of $H(H+E)^{-1}$

Crab Example

```
MANOVA Test Criteria and F Approximations for the
      Hypothesis of No Overall CrabType Effect
        H = Type III SSCP Matrix for CrabType
               E = Error SSCP Matrix
            S=3    M=0    N=25.5

Statistic     Value    F Value   Num DF    Den DF  Pr > F

Wilks' Lambda   0.026  34.19      12    140.52    <.0001
Pillai's Trace  1.607  15.87      12       165    <.0001
Hotelling-Law  13.308  57.87      12    88.531    <.0001
Roy's Greatest 11.371  156.36      4        55    <.0001
```


Discriminant Analysis

Discriminant Analysis tries to find the linear combinations of variables that do the best job at classifying observations into one of several groups.

There are many ways to pursue discriminant analysis such as quadratic discriminant analysis, neural networks, regression trees and support vector machines. (e.g. Venables and Ripley, MASS2002)

Fisher's Linear Discriminant Analysis

The idea of Fisher's Linear Discriminant Analysis is to find the linear combination of variables ($a^T x$) that "best distinguishes between groups" by maximizing the between group sum of squares (H) divided by the within group sum of squares (E)

How?

Using theory similar to that of PCA, the coefficient vector a that maximizes $a^T Ha / a^T Ea$ is the first eigen vector of $E^{-1}H$.

$a^T x$ is Fisher's linear discriminant function or the first canonical discriminant function

Each observation X is allocated to the group i such that $a^T x$ is closest to $a^T \overline{x}_i$

STAT 530/J530 B.Habing Univ. of S.C. 7

Notes

- For two groups this is the same as classifying each observation with the group whose mean it is closest to using Mahalanobis distance (they differ, but are often similar for more than 2)
- For two groups this is the same as using maximum likelihood and assuming multivariate normality and equal covariances

STAT 530/J530 B.Habing Univ. of S.C. 8

Notes

- The corresponding eigen Value is Roy's greatest root (and can also be shown related to Wilk's lamda)
- There are other canonical discriminant functions based on the other eigen vectors (although they do not form an orthogonal basis)
- Test's can be used to see how many discriminant functions are needed

STAT 530/J530 B.Habing Univ. of S.C. 9

Example

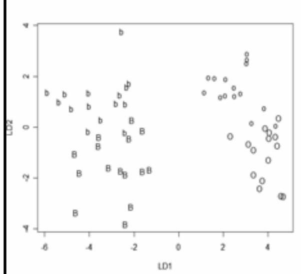
```
> library(MASS)
> attach(crabdata)
> lda(cbind(sFL,sRW,sCL,sCW),CrabType)

Coefficients of linear discriminants:
            LD1         LD2        LD3
sFL  0.55668852 -0.06333663  1.0947794
sRW  0.01345867  2.49170184 -0.1427386
sCL  2.56898619 -1.00234308  1.6110090
sCW -6.23384916 -0.87881080 -2.0107102

Proportion of trace:
   LD1    LD2    LD3
0.8544 0.1431 0.0025
```


Example

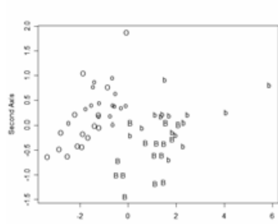
Canonical Discriminant Analysis
Test of H0: The canonical correlations in the current row and all that follow are zero

	Eigenvalue	Proportion	Approximate F Value	Pr > F
1	11.3714	0.8544	34.19	<.0001
2	1.9039	0.1431	13.18	<.0001
3	0.0336	0.0025	0.92	0.4031

Example



LDA



MDS

Example

```
> predict(crablda,dim=3)
$class
[1] B b B B B B B B B B B B B B B b b b b B b
   b b b b b b b b b O O O O O O O O o O O
[41] O O O O O o o O o o o o o o o O o o o o
Levels: b B o O

$posterior
       b    B    o    O
[1,] 0.16 0.84 0.00 0.00
[2,] 0.67 0.33 0.00 0.00
[3,] 0.25 0.75 0.00 0.00
[4,] 0.01 0.99 0.00 0.00
```


About Prediction

The predicted probabilities are calculated assuming normality and follow from Bayes rule.

It can be shown that the linear discriminant analysis is equivalent to logistic regression for 2 dimensions.

Allowing for unequal covariances results in quadratic discriminant analysis.

Example

```
> table(Original=CrabType,
  Predicted=predict(crablda,dim=3)$class)

        Predicted
Original  b  B  o  O
       b 14  1  0  0
       B  1 14  0  0
       o  0  0 13  2
       O  0  0  1 14
```

Need for Cross-Validation

A difficulty with the previous slide is that we are testing how good we are at predicting the very data we used to come up with the rule.

This will almost always overestimate the accuracy.

The jack-knife (or leave-one-out cross-validation) can give a better idea of the true accuracy.

Jack-Knife

The jack-knife is when a statistical procedure is repeated *n* times, once for each observation, where that observation is removed.

In measuring the accuracy of LDA the classification of each observation is determined by performing the LDA without that observation.

Example

```
> crabcv<-lda(cbind(sFL,sRW,sCL,sCW),CrabType,
  CV=T)
> table(Original=CrabType,
  Predicted=crabcv$class)

        Predicted
Original  b  B  o  O
       b 14  1  0  0
       B  1 14  0  0
       o  0  0 12  3
       O  0  0  1 14
```