MANOVA

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

\[ H_0: \mu_1 = \mu_2 = \cdots = \mu_k \]

where the \( \mu_i \) are the \( (q \times 1) \) 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 Lambda: \( \Lambda = \frac{|E|}{|H + E|} \)

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

\[ H = \text{Type III SSCP Matrix for CrabType} \]
\[ E = \text{Error SSCP Matrix} \]

\[ S=3 \quad N=0 \quad N=25.5 \]

<table>
<thead>
<tr>
<th>Statistic</th>
<th>Value</th>
<th>F Value</th>
<th>Num DF</th>
<th>Den DF</th>
<th>Pr &gt; F</th>
</tr>
</thead>
<tbody>
<tr>
<td>Wilks' Lambda</td>
<td>0.026</td>
<td>34.19</td>
<td>12</td>
<td>140.52</td>
<td>&lt;.0001</td>
</tr>
<tr>
<td>Pillai's Trace</td>
<td>1.607</td>
<td>15.87</td>
<td>12</td>
<td>165</td>
<td>&lt;.0001</td>
</tr>
<tr>
<td>Hotelling-Law</td>
<td>13.308</td>
<td>57.87</td>
<td>12</td>
<td>88.531</td>
<td>&lt;.0001</td>
</tr>
<tr>
<td>Roy's Greatest</td>
<td>11.371</td>
<td>156.36</td>
<td>4</td>
<td>55</td>
<td>&lt;.0001</td>
</tr>
</tbody>
</table>

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 \((a^Tx)\) 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 $a^T H a / a^T E a$ 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 $i$.

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.

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.

• Test's can be used to see how many discriminant functions are needed.
Example

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

Coefficients of linear discriminants:

<table>
<thead>
<tr>
<th></th>
<th>LD1</th>
<th>LD2</th>
<th>LD3</th>
</tr>
</thead>
<tbody>
<tr>
<td>sFL</td>
<td>0.55668852</td>
<td>-0.06333663</td>
<td>1.0947794</td>
</tr>
<tr>
<td>sRW</td>
<td>0.01345867</td>
<td>2.49170184</td>
<td>-0.1427386</td>
</tr>
<tr>
<td>sCL</td>
<td>2.56898619</td>
<td>-1.00234308</td>
<td>1.6110090</td>
</tr>
<tr>
<td>sCW</td>
<td>-6.23384916</td>
<td>-0.87881080</td>
<td>-2.0107102</td>
</tr>
</tbody>
</table>

Proportion of trace:

<table>
<thead>
<tr>
<th></th>
<th>LD1</th>
<th>LD2</th>
<th>LD3</th>
</tr>
</thead>
<tbody>
<tr>
<td>Value</td>
<td>0.8544</td>
<td>0.1431</td>
<td>0.0025</td>
</tr>
</tbody>
</table>

Example

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

Approximate

<table>
<thead>
<tr>
<th>Eigenvalue</th>
<th>Proportion</th>
<th>F Value</th>
<th>Pr &gt; F</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0.8544</td>
<td>34.19</td>
<td>&lt;.0001</td>
</tr>
<tr>
<td>2</td>
<td>0.1431</td>
<td>13.18</td>
<td>&lt;.0001</td>
</tr>
<tr>
<td>3</td>
<td>0.0025</td>
<td>0.92</td>
<td>0.4031</td>
</tr>
</tbody>
</table>

Example

LDA
MDS
Example

```r
> predict(crablda, dim=3)
$\text{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 b b b b b b b b b b
[41] o o 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

$\text{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

```r
> 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

```r
> 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
```