Discriminant Analysis

Clustering and Classification

Lecture 3

2/14/06

Today’s Class

- Introduction to Discriminant Analysis
  - From Sage book with same title by Klecka (1980).
  - From Johnson & Wichern, Chapter 11.
- Assumptions of DA.
- How DA works.
  - How to arrive at discriminant functions.
  - How many discriminant functions to use.
  - How to interpret the results.
- How to classify objects using DA.

General Introduction

General Introduction

- DA is a statistical technique that allows the user to investigate the differences between multiple sets of objects across several variables simultaneously.

- DA works off of matrices used in Multivariate Analysis of Variance (MANOVA).

When to Use Discriminant Analysis

- Data should be from distinct groups.
  - Group membership must already be known prior to initial analysis.
- DA is used to interpret group differences.
- DA is used to classify new objects.

Assumptions

- Data must not have linear dependencies.
  - Must be able to invert matrices.
- Population covariance must be equal for each group.
- Each group must be drawn from a population where the variables are multivariate normal (MVN).

Notation

- g = number of groups
- p = number of discriminating variables
- n_i = number of cases in group i
- $n_.$ = number of cases over all the groups

More Assumptions

1. two or more groups $g >= 2$
2. at least two cases per group $n_i >= 2$
3. any number of discriminating variables, provided that they are less than the total number of cases minus two: $0 < p < (n_. - 2)$
4. discriminating variables are measured at the interval level
5. no discriminating variable may be a linear combination of the other discriminating variables

More Assumptions

5. no discriminating variable may be a linear combination of the other discriminating variables
6. the covariance matrices for each group must be (approximately) equal, unless special formulas are used
7. each group has been drawn from a population with a MVN distribution on the discriminating variables.

Example from Klecka

- To demonstrate DA, Klecka (1980) uses an example of data taken from senatorial factions (citing Bardes, 1975 and 1976).
- Bardes wanted to know how US Senate voting factions changed over time
  - How stable they were from year to year
  - How much they were influenced by other issues.

Groups of Senators

- Known Groups of Senators:
  1. Generally favoring foreign aid (9)
  2. Generally opposing foreign aid (2)
  3. Opposed to foreign involvements (5)
  4. Anti-Communists (3)

Variables

- Six variables (from roll call votes):
  1. CUTAID – cut aid funds
  2. RESTRICT – add restrictions to the aid program
  3. CUTASIAN – cut funds for Asian nations
  4. MIXED – Mixed issues: liberal aid v. no aid to communists
  5. ANTIYUGO – Anti-aid to Yugoslavia
  6. ANTINUET – Anti-aid to neutral countries

Univariate Statistics

TABLE 1
Means for "Known" Senators

Variable	Group 1	2	3	4	Total
CUTAID	1.422	3.000	2.200	2.100	1.900
RESTRICT	1.944	1.000	2.000	2.333	1.921
CUTASIAN	1.000	3.000	2.000	1.333	1.526
MIXED	2.667	2.000	1.800	1.667	2.211
ANTIYUGO	1.556	2.500	2.600	3.000	2.158
ANTINEUT	1.259	1.667	2.133	2.444	1.719

How Discriminant Analysis Works

Canonical Discriminant Analysis

- The canonical discriminant function looks like this:

$$f_{km} = u_0 + u_1 X_{1km} + u_2 X_{2km} + \ldots + u_p X_{pkm}$$

- Here:
  - f_{km} = the value (score) on the canonical discriminant function for case m in the group k
  - X_{ikm} = the value on discriminating variable X_i for case m in group k
  - u_i = coefficients which produce the desired characteristics of the function.

Number of Functions

- Because Canonical DA makes use of methods similar to Canonical Correlations, a set of discriminant functions are derived.
  – The first function is built to maximize group differences.
  – The next functions are built to be orthogonal to the first, and still maximize group differences.
- The number of functions derived is equal to max(g-1,p)
  – In the example, this would be max(4-1,6)=6.

Deriving the Canonical Discriminant Functions

- To get at the canonical discriminant functions, we must first construct a set of sums of squares and crossproducts (SSCP) matrices.
  - A total covariance matrix
  - A within group covariance matrix
  - A between group covariance matrix

- Once we have the between and within matrices, we take the eigenvalues and eigenvectors of each.

Total SSCP Matrix

- Each element of the total SSCP matrix:

$$t_{ij} = \sum_{k=1}^{g} \sum_{m=1}^{n_k} (X_{ikm} - X_{i..})(X_{jkm} - X_{j..})$$

- g = number of groups
- n_k = number of cases in group k
- $n_.$ = total number of cases over all groups
- X_{ikm} = the value of variable i for case m in group k
- $X_{ik.}$ = mean value of variable i for cases in group k
- $X_{i..}$ = mean value of variable i for all cases

Within SSCP Matrix

- Each element of the within SSCP matrix:

$$w_{ij} = \sum_{k=1}^{g} \sum_{m=1}^{n_k} (X_{ikm} - X_{ik\cdot})(X_{jkm} - X_{jk\cdot})$$

- g = number of groups
- n_k = number of cases in group k
- $n_{.}$ = total number of cases over all groups
- X_{ikm} = the value of variable i for case m in group k
- $X_{ik.}$ = mean value of variable i for cases in group k
- $X_{i..}$ = mean value of variable i for all cases

Between SSCP Matrix

- Once we have **W** and **T**, we can compute **B** by the following formula:

$$\mathbf{B} = \mathbf{T} - \mathbf{W}$$

- When there are no differences between the group centroids (the mean vectors of each group), **W** = **T**.

- The extent they differ will define the distinctions among the observed variables.

Obtaining Discriminant Functions

- Once we have **B** and **W**, we then find the solutions (v_i) to the following equations:

$$\begin{gathered}\sum b_{1i}v_i = \lambda \sum w_{1i}v_i \\ \sum b_{2i}v_i = \lambda \sum w_{2i}v_i \\ \vdots \\ \sum b_{pi}v_i = \lambda \sum w_{pi}v_i\end{gathered}$$

- There is also a constraint that the sum of the squared v_i equal one (as typical in PCA).

Step 2: Converting to Functions

- Once the λ and v_i parameters are found, one then converts these into the weights for the discriminant functions:

$$u_i = v_i \sqrt{n. - g}$$

$$u_0 = -\sum_{i=1}^{p} u_i X_{i..}$$

Interpreting the Discriminant Functions

Example Results

TABLE 4
Unstandardized Discriminant Coefficients

Variable	Unstandardized Coefficient		
	Function 1	**Function 2**	**Function 3**
Constant (u_0)	5.4243	3.5685	–4.3773
CUTAID	.8078	–.5225	1.6209
RESTRICT	.7940	–1.1177	–.3339
CUTASIAN	–4.6004	–1.1228	–1.1431
MIXED	–.6957	–1.3160	1.1418
ANTIYUGO	–1.1114	1.1132	.3781
ANTINEUT	1.4387	1.0422	.2000

Example Function Scores for an Observation

Computation of Discriminant Scores for Senator Aiken

Variable	FUNCTION 1 Coeff.	× Value	= Contribution	FUNCTION 2 Coeff.	× Value	= Contribution	FUNCTION 3 Coeff.	× Value	= Contribution
Constant			5.4243			3.5685			−4.3773
CUTAID	.8078	1.0	.8078	−.5225	1.0	−.5225	1.6209	1.0	1.6209
RESTRICT	.7940	3.0	2.3820	−1.1177	3.0	−3.3531	−.3339	3.0	−1.0017
CUTASIAN	−4.6004	1.0	−4.6004	−1.1228	1.0	−1.1228	−1.1431	1.0	−1.1431
MIXED	−.6957	3.0	−2.0871	−1.3160	3.0	−3.9480	1.1418	3.0	3.4254
ANTIYUGO	−1.1114	1.0	−1.1114	1.1132	1.0	1.1132	.3781	1.0	.3781
ANTINEUT	1.4387	1.0	1.4387	1.4387	1.0	1.0422	.2000	1.0	.2000
discriminant score			2.2539			−3.2225			−.8977

Example Interpretation

- In the example, we saw that Senator Aiken had discriminant scores of 2.25, -3.22, and -0.90.
  - These scores are in standard deviation units…of the discriminant space
- Positive values shows an object being high on a dimension.
- Negative values shows an object being low on a dimension.
- We will come to learn how to interpret the dimensions.

Group Centroids

- What we are really after is the group means for each of the discriminant functions.

- The means in this case are:
  1. 1.74, -0.94, 0.02
  2. -6.93, -0.60, 0.28
  3. -1.48, 0.69, -0.30
  4. 1.86, 2.06, 0.25

- These will be used to classify our observations.

Standardized Coefficients

- To interpret each dimension, we look at the standardized coefficients.
- Standardized coefficients are created by:

$$c_i = u_i \sqrt{\frac{w_{ii}}{n. - g}}$$

TABLE 6
Standardized Discriminant Coefficients

	Standardized Coefficient		
Variable	**Function 1**	**Function 2**	**Function 3**
CUTAID	.6094	−.3942	1.2227
RESTRICT	.7068	−.9950	−.2973
CUTASIAN	−2.1859	−.5335	−.5432
MIXED	−.4760	−.9004	.7812
ANTIYUGO	−.8077	.8090	.2748
ANTINEUT	1.0168	.7365	.1414

How Many Significant Functions?

- To see how many functions are needed to describe group differences, we need to look at the eigenvalues, λ, for each dimension.
- We will have a test statistic based on the eigenvalue.
- The statistic provides the result of a hypothesis test testing that the dimension (and all subsequent dimensions) are not signficant.

Example Test Statistics

TABLE 10
Residual Discrimination and Test of Significance

Functions Derived, k	Wilks's Lambda	Chi-Square	Degrees of Freedom	Significance Level
0	.0345	43.760	18	.001
1	.3680	12.996	10	.224
2	.9492	.678	4	.954

Classifying Objects

Classifying Objects

- Several methods exist for classifying objects.
- Each is based on the distance of an object from each group’s centroid.
  - The object is then classified into the group with the smallest distance
- Many classification methods use the raw data.
- The canonical discriminant functions can be used as well.

Validation of Classification

- We will show more about classification in the next class.

- Basically, once we classify objects, we need to see how good we are at putting our objects into groups.

- There are multiple ways to test whether or not we do a good job.
  - Most easy is to just classify all of our objects and see how good we recover our original groups.

Classification Matrix Example

TABLE 14

Classification Matrix

Original Group	Predicted Group 1	2	3	4
1	8	0	0	1
2	0	2	0	0
3	0	0	5	0
4	0	0	0	3
Unknown	33	10	27	4

Wrapping Up

- Discriminant Analysis is a long-standing method for deriving the dimensions along which groups differ.
- We will see that it is often the first method used when approaching a classification problem
- We must have a training data set in place to be able to use this method.
  - All of our other methods will not require this.

Next Time

- How to do discriminant analysis in R
- Presentation of Anderson (2005) article.