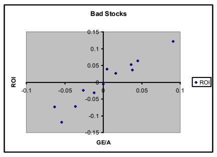
Discriminant Analysis

To find out variables that best discriminate between different clusters

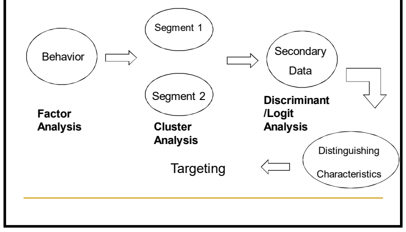
Catalog Business

- Identified two consumer segments
  - One which buys a lot
  - Other which does not buy as much
- Can we find variables that help discriminate the behavior of these two groups?
- Can we use these discriminators to classify *other consumers* into one of these two groups?

Bad Stocks



Factor/ Cluster/ Discriminant



Factor Analysis

Cluster Analysis

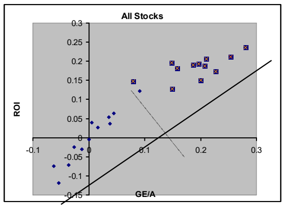
Discriminant /Logit Analysis

Targeting

Data

Stock #	GE/A	ROI		Stock #	GE/A	ROI
1	0.158	0.182		13	-0.012	-0.031
2	0.21	0.206		14	0.036	0.053
3	0.207	0.188		15	0.038	0.036
4	0.28	0.236		16	-0.063	-0.074
5	0.197	0.193		17	-0.054	-0.119
6	0.227	0.173		18	0	-0.005
7	0.148	0.196		19	0.005	0.039
8	0.254	0.212		20	0.091	0.122
9	0.079	0.147		21	-0.036	-0.072
10	0.149	0.128		22	0.045	0.064
11	0.2	0.15		23	-0.026	-0.024
12	0.187	0.191		24	0.016	0.026

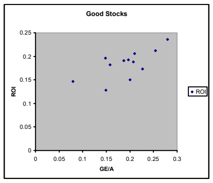
All Stocks



Web Browsing

- Cluster analysis identified two groups of consumers
  - One that visits your website frequently
  - One that doesn't
- How to find the frequent visitors for better targeting? Can the differences in behavior be related to socio-demographic variables?
- Can we use the demographic variables to classify prospects into one of these two groups?

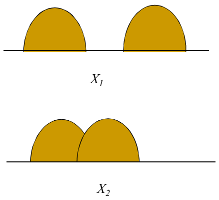
Good Stocks



Identifying the Best Discriminators

- Two groups appear to be well separated on each ratio: ROI and GE/A
- Also well separated in two dimensional space
- But this need not always be the case!

Discriminating Variables



Discriminant Analysis

- Identify a set of variables that best discriminate between the two groups
- Does so by choosing a new line that maximizes the similarity between members of the same group and minimizing the similarity between members belonging to different groups

Discriminant Function

$Z = w_1\ GEA + w_2\ ROI$

Between-Group Sum of Squares – SS_b
Within-Group Sum of Squares – SS_w

$\lambda \quad = \quad (SS_b/SS_w)$

More on the Criterion

- For Z to provide maximum separation between the groups, the following must be satisfied:
  - The means of Z for the two groups should be as far apart as possible (or high SS_b)
  - Values of Z for each group should be as homogenous as possible (or low SS_w)

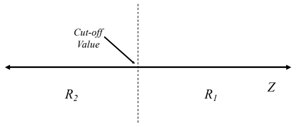
Classification

- Discriminant Function: The line that separates the members of the two groups
- Methods of Classification
  - Cut-Off Value Method
  - Decision Theory Approach
  - Classification Function Approach
  - Mahalanobis Distance Method

Cut-Off Value Method

- Uses the Discriminant Function line to score new observations (prospects) and classify them into one of two groups based on a cut-off value

Classification



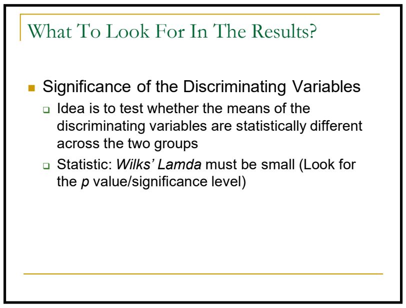
Classification Function Approach

- Classifications based on this approach are identical to those done by Decision Theory approach
- Classification functions are computed for each group:

  C_1 = -7.87 + 61.237*GEA + 21.027*ROI
  C_2 = -0.004 + 2.551*GEA – 1.404*ROI

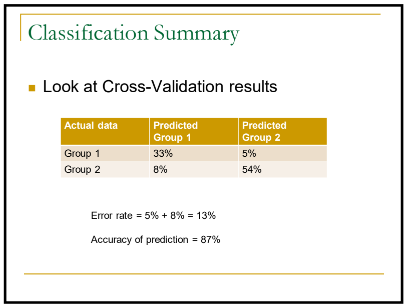
Basic Idea

- Score each new observation using these two scoring functions
- The observation gets assigned to the group with the higher score



What To Look For In The Results?

- Significance of the Discriminating Variables
  - Idea is to test whether the means of the discriminating variables are statistically different across the two groups
  - Statistic: *Wilks' Lamda* must be small (Look for the *p* value/significance level)



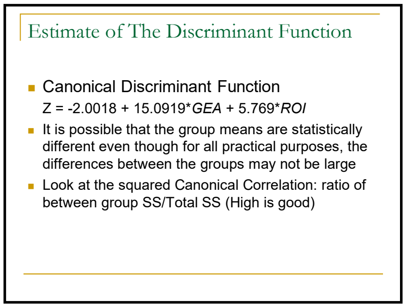
Classification Summary

- Look at Cross-Validation results

Actual data	Predicted Group 1	Predicted Group 2
Group 1	33%	5%
Group 2	8%	54%

Error rate = 5% + 8% = 13%

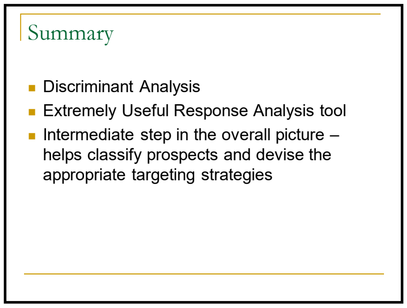
Accuracy of prediction = 87%



Estimate of The Discriminant Function

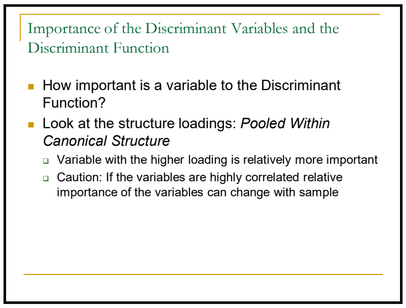
- Canonical Discriminant Function

  Z = -2.0018 + 15.0919*GEA + 5.769*ROI
- It is possible that the group means are statistically different even though for all practical purposes, the differences between the groups may not be large
- Look at the squared Canonical Correlation: ratio of between group SS/Total SS (High is good)



Summary

- Discriminant Analysis
- Extremely Useful Response Analysis tool
- Intermediate step in the overall picture – helps classify prospects and devise the appropriate targeting strategies



Importance of the Discriminant Variables and the Discriminant Function

- How important is a variable to the Discriminant Function?
- Look at the structure loadings: *Pooled Within Canonical Structure*
  - Variable with the higher loading is relatively more important
  - Caution: If the variables are highly correlated relative importance of the variables can change with sample