Admin

- Read Chapter 13 of T&F.
- Assignment 1 due at 3pm.
- Assignment 2 due at 10am on 13 May.

Discriminant Analysis

What sort of questions are being investigated?

Factors predicting return to work following mild traumatic brain injury: A discriminant analysis

...The stepwise DFA revealed that age and three cognitive variables (verbal memory, verbal fluency, and a speed test of planning and strategy) were predictive of work status 3-15 months following a documented MTBI, correctly classifying work status 68.8% of the time.

Information processing in schizophrenia and bipolar disorder: A discriminant analysis

A study was conducted in which a computerized battery of information processing tasks (called COGLAB) was administered to three participant groups: patients with schizophrenia, patients with bipolar disorder, and normal controls. The tasks included the Mueller-Lyer illusion, reaction time, size estimation, a variant of the Wisconsin Card Sorting Test, backward masking, and Asarnow continuous performance. Discriminant analyses were used to investigate the differences among the three groups. Results indicated that COGLAB correctly classified 75.5% of the cases of schizophrenia and bipolar disorder. The Mueller-Lyer illusion and the number of perseverative errors on the card sort most powerfully discriminated the two groups.

Gender conformity, masturbation fantasy, infatuation, and sexual orientation: A discriminant analysis investigation

Investigated whether the combination of gender conformity, infatuation, and masturbation fantasies could be used to differentiate between 69 heterosexuals and 106 homosexuals. Information about childhood gender conformity/nonconformity, childhood infatuation objects, and adolescent masturbation fantasies proved to be powerful discriminators between the 2 groups, suggesting that adult sexual orientation is essentially predicted by prior childhood variables.

Attempted suicide among adolescents: A stepwise discriminant analysis

A sample of 40 adolescents (aged 13-17 yrs), some of whom had attempted suicide (n = 20) and some who had not (n = 20), were compared on a number of life history and psychological variables (affective and cognitive). Stepwise discriminant analyses produced 1 discriminant function (the lack of emotional significant other) that differentiated between the groups. This single discriminant function accounted for 77% of the variance in the data. These results are interpreted within the context of the early loss hypothesis.

Understanding the audiences of a health communication campaign: A discriminant analysis of potential organ donors based on intent to donate

A discriminant analysis of 5 relevant variables showed that individuals who had a signed organ donor card had a high level of knowledge about organ donation, were rather altruistic, and did not believe that signing a donor card was a fearful activity. Individuals who were high in intent to sign organ donor cards (but had not done so) had limited knowledge but a positive attitude toward organ donation, and they often regarded the signing as a fearful activity. Those who were low in intent to donate tended to have inaccurate knowledge, were not overly altruistic, and felt that signing would be quite frightening.

Premarital contraceptive use: A discriminant analysis approach

Investigated the ability of 7 independent variables to predict accurately, from a sample of 308 unmarried, sexually active undergraduates, which Ss use reliable or unreliable contraceptives. Using discriminant analysis, the 7 independent variables were age at which Ss started engaging in coitus, frequency of coitus, frequency of dating, length of time sex partners knew each other, number of sex partners, anticipation of coitus occurring, and number of close friends who were thought to use birth control. The results indicate that these variables were accurate in predicting which Ss were users of contraceptives, with more than 80% of males and females being correctly classified. The number of close friends thought to use contraceptives was the most influential variable for both sexes, followed by length of time the partners knew each other.

Credit Risk?

Sometimes we encounter a problem that involves a categorical variable and several continuous variables.

$$Y_1, Y_2, \ldots Y_p \quad \longleftarrow \quad X$$

p continuous variables — categorical, *k* levels

We might want to distinguish between good or bad credit risks using a number of variables.

Volatility of employment
House prices in the region
Credit score
Debt-to-income ratio

$\longleftarrow$

Good Risk
or
Bad Risk

Note that if we had a metric of credit risk (i.e., if we had enough information to form a continuous variable, then we could use multiple regression.

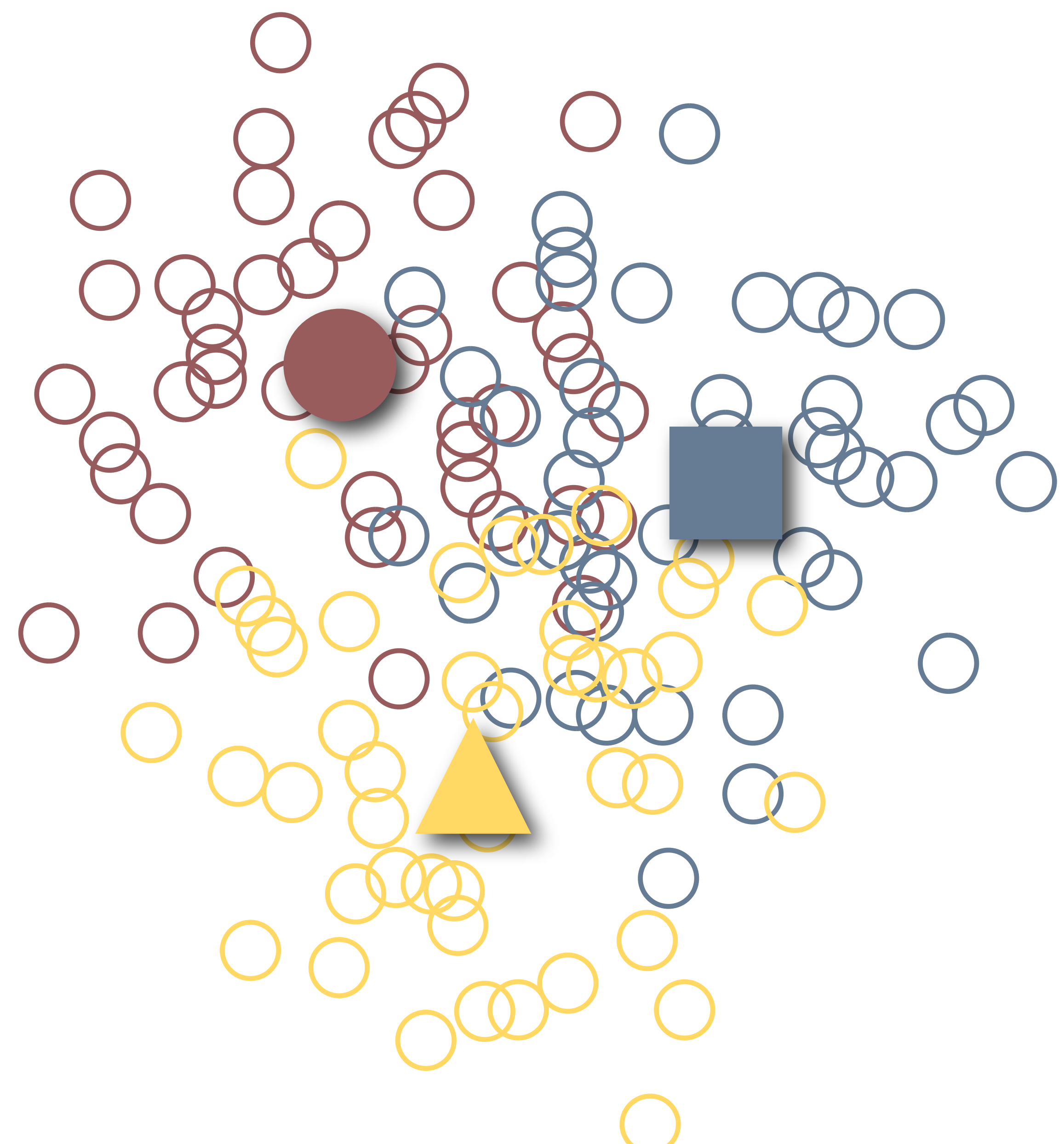
Variates and Centroids

The goal is to derive a *variate*. The discriminant variate is the linear combination of the two (or more) continuous variables that will best discriminate between the groups.

We do this by the variate's weights for each independent variable to maximise the differences between the groups (i.e., the between-group variance relative to the within-group variance).

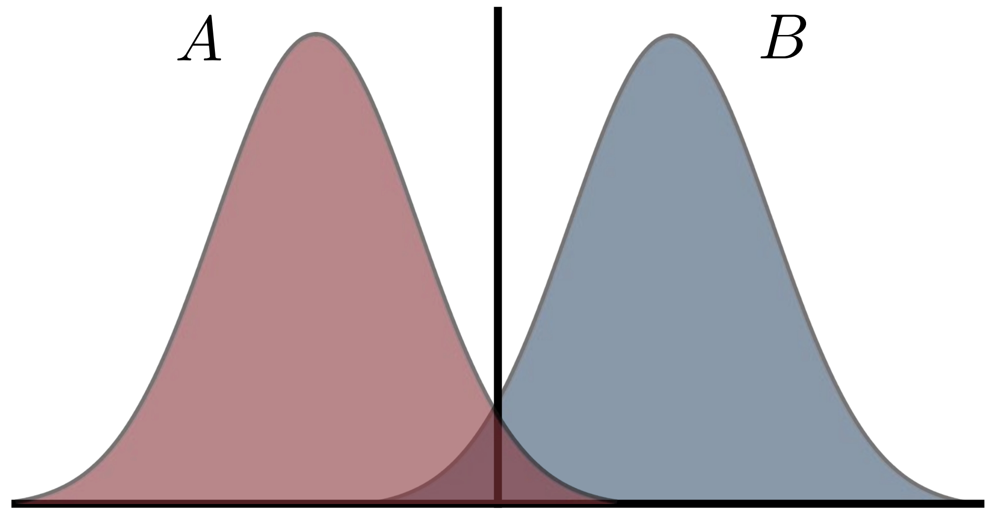
By averaging the discriminant scores for all the individuals within a particular group, we arrive at the group mean. This group mean is referred to as a *centroid*.

When the analysis involves two groups, there are two centroids. Three groups: three centroids; and so on.

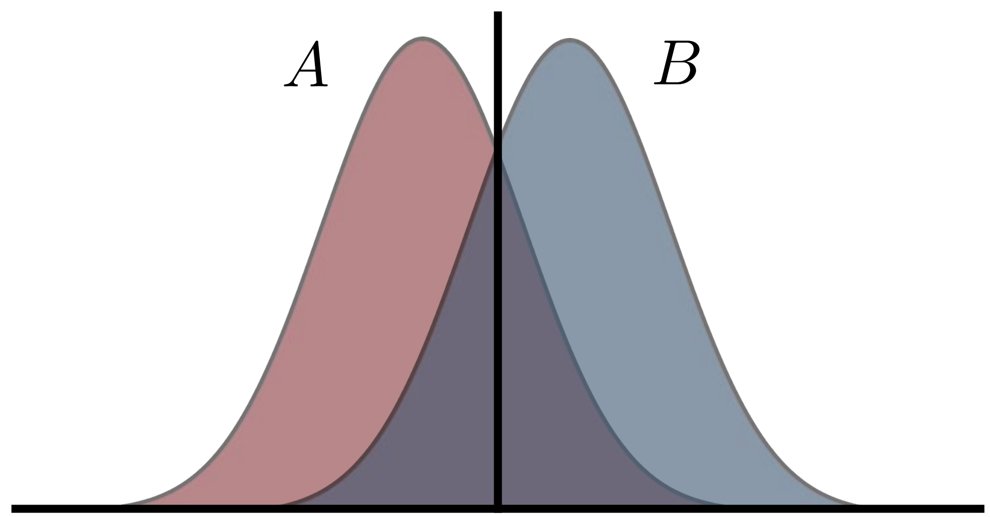


Significance of the Discriminant Function

The test for the statistical significance of the discriminant function is a generalised measure of the distance between the group centroids.

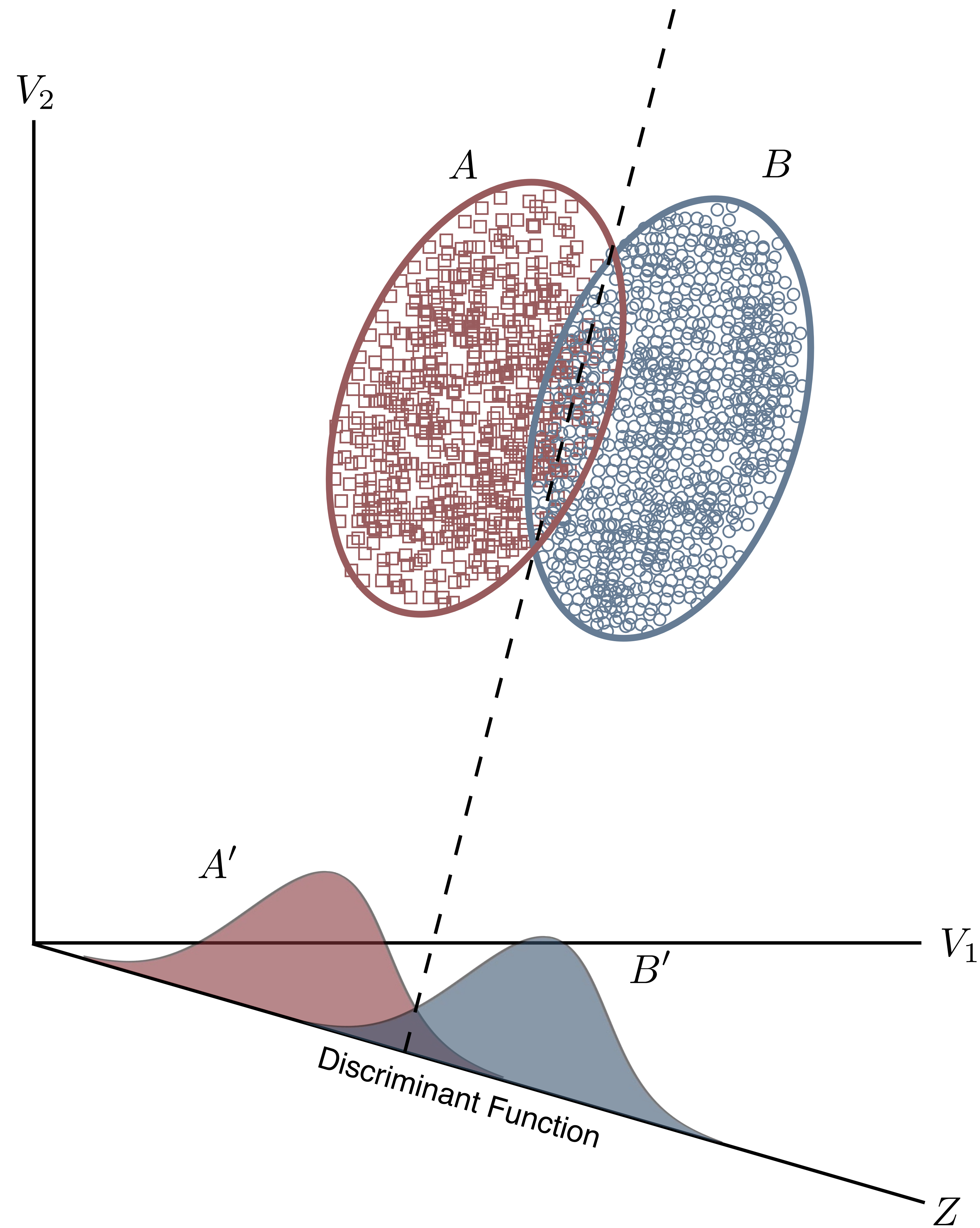


If the overlap is small, then the discriminant function separates the groups well.



If the overlap is large, then the function is a poor discriminator between groups.

Note: The shaded areas of overlap represent instances where misclassifying objects from Group A into Group B, and vice versa, can occur.


V_2
A
B
A'
V_1
B'
Discriminant Function
Z

Three Group Example

Objective

One of the emerging paradigms in marketing is the concept of a customer relationship, based on the establishment of a mutual partnership between firms over repeated transactions. The process of developing a relationship entails the formation of shared goals and values, which should coincide with improved perceptions of HBAT. Thus, the successful formation of a relationship should be seen by improved HBAT perceptions over time. In this analysis, firms are grouped on their tenure as HBAT customers. Hopefully, if HBAT has been successful in establishing relationships with its customers, then perceptions of HBAT will improve with tenure as a HBAT customer.

To test this relationship, let’s do a discriminant analysis to establish how three different customer groups (based on the length of customer relationship) differ on a set of variables.

Three Group Example

Research Design

Product Quality E-Commerce Activities Technical Support Complaint Resolution Advertising Product Line Salesforce Image Competitive Pricing Warranty & Claims New Products Order & Billing Price Flexibility Delivery Speed	$\longleftarrow$	$<$ 1 year 1 to 5 years $>$ 5 years categorical p = 3 levels

k = 13 continuous variables

We want to describe the differences between these three groups of customers on a set of 13 continuous variables.

Three Group Example

Assumptions

- True Categorical Grouping Variable
- Sample sizes
- Homoscedasticity
- Outliers
- Multicollinearity, Singularity, and Redundant Variables

Three Group Example

Syntax

```
DISCRIMINANT
  /GROUPS=x1(1 3)
  /VARIABLES=x6 x7 x8 x9 x10 x11 x12 x13 x14 x15 x16 x17 x18
  /ANALYSIS ALL (2)
  /METHOD=WILKS
  /FIN= 3.84
  /FOUT= 2.71
  /PRIORS  EQUAL
  /HISTORY
  /STATISTICS=MEAN STDDEV UNIVF BOXM CORR FPAIR TABLE
  /PLOT=COMBINED
  /CLASSIFY=NONMISSING POOLED.
```

Note: this is (2) regardless of the number of categorical levels

Three Group Example

Test for Homogeneity: Box’s M

Test Results

Box's M		364.728
F	Approx.	1.619
	df1	182
	df2	24683.7
	Sig.	.000

Tests null hypothesis of equal population covariance matrices.

From T&F: “Transform variables if there is a significant departure from homogeneity, samples are small and unequal, and inference is the major goal.”

They recommend transform if you have

1. Significant Box’s M **and**
2. Small and Unequal Sample Sizes **and**
3. The purpose is inference

That is, they recommend transformation if you have all 3.

Three Group Example

Overall statistical significance and number of functions: Wilk's Lambda

Wilk's Lambda values close to 1 indicate the group means are not different (equal to 1 indicates all means are the same).

Wilks' Lambda

Test of Function(s)	Wilks' Lambda	Chi-square	df	Sig.
1 through 2	.175	158.407	26	.000
2	.601	46.377	12	.000

Notice that both functions are significant. The number of functions that we have here depends on the number of groups being tested. If we had four groups, then there would be three functions: 1 through 2, 2 through 3, and 3.

In the first step (1 through 2 in our example in the table; 1 through k-1 in general), both (all) functions are being tested. This is the overall test. If this is not significant then our discriminant variables are not able to distinguish between our groups.

Are the groups significantly different *on any single one of the 13 continuous variables?*

Remember that the Wilk's Lambda overall test only assesses *overall* differences and does not guarantee that each group is significantly different from the others.

For the second function there are still significant differences between groups. **So two functions needed to describe the between group differences.**

Three Group Example

These measure the strength of the relationship between the discriminant scores and the groups. Values close to 1 indicate a strong correlation between the discriminant scores and the groups.

Eigenvalues

Function	Eigenvalue	% of Variance	Cumulative %	Canonical Correlation
1	2.425[a]	78.5	78.5	.841
2	.665[a]	21.5	100.0	.632

a. First 2 canonical discriminant functions were used in the analysis.

The eigenvalue is the ratio of the between-groups sum of squares to the within-groups sum of squares.

Remember:

The goal here is to maximise this discriminant ratio, this eigenvalue λ.

These eigenvalues are also used to help us compute the canonical correlations:

$$R_{Cj} = \sqrt{\frac{\lambda_j}{(1+\lambda_j)}}$$

Three Group Example

% of Variance

Eigenvalues

Function	Eigenvalue	% of Variance	Cumulative %	Canonical Correlation
1	2.425[a]	78.5	78.5	.841
2	.665[a]	21.5	100.0	.632

a. First 2 canonical discriminant functions were used in the analysis.

'% of Variance' represents the contribution of that discriminant function relative to all other functions.

This can also be represented cumulatively.

In English: The overall goodness of fit for the discriminant model is statistically significant and both functions are statistically significant as well. The first function accounts for 78.5 percent of the variance explained by the first two functions, with the remaining variance (21.5%) due to the second function. The total amount of variance explained by the first function is $.841^2$, or 70.7 percent. The next function explains $.632^2$, or 39.9 percent of the remaining variance (29.3%).

Three Group Example

Pooled Within-groups Correlation Matrix

		X6 – Product Quality	X7 – E-Commerce Activities	X8 – Technical Support	X9 – Complaint Resolution	X10 – Advertising	X11 – Product Line	X12 – Salesforce Image	X13 – Competitive Pricing	X14 – Warranty & Claims	X15 – New Products	X16 – Order & Billing	X17 – Price Flexibility	X18 – Delivery Speed
Correlation	X6 – Product Quality	1.000	–.083	.027	–.200	–.080	.108	–.175	–.104	–.046	–.055	–.105	–.450	–.348
	X7 – E-Commerce Activities	–.083	1.000	.009	.185	.430	.008	.799	.197	.073	–.011	.184	.247	.259
	X8 – Technical Support	.027	.009	1.000	.025	–.081	.142	.005	–.254	.795	–.082	.019	–.223	–.076
	X9 – Complaint Resolution	–.200	.185	.025	1.000	.116	.411	.166	.030	.079	.073	.643	.411	.772
	X10 – Advertising	–.080	.430	–.081	.116	1.000	–.079	.522	.153	.003	.099	.107	.301	.215
	X11 – Product Line	.108	.008	.142	.411	–.079	1.000	–.115	–.284	.203	–.010	.262	–.429	.488
	X12 – Salesforce Image	–.175	.799	.005	.166	.522	–.115	1.000	.281	.112	.051	.123	.297	.215
	X13 – Competitive Pricing	–.104	.197	–.254	.030	.153	–.284	.281	1.000	–.179	.093	–.007	.398	.111
	X14 – Warranty & Claims	–.046	.073	.795	.079	.003	.203	.112	–.179	1.000	.018	.162	–.152	.039
	X15 – New Products	–.055	–.011	–.082	.073	.099	–.010	.051	.093	.018	1.000	.084	.191	.145
	X16 – Order & Billing	–.105	.184	.019	.643	.107	.262	.123	–.007	.162	.084	1.000	.397	.626
	X17 – Price Flexibility	–.450	.247	–.223	.411	.301	–.429	.297	.398	–.152	.191	.397	1.000	.555
	X18 – Delivery Speed	–.348	.259	–.076	.772	.215	.488	.215	.111	.039	.145	.626	.555	1.000

The pooled within-group correlation matrix provides estimates of the correlations between variables with the effects of the grouping variable removed. In effect, this is as if the variables were correlated separately for each of the groups and these correlations were averaged.

If there are several strong correlations (greater than say 0.75 or less than -0.75) there may be alternative subsets of variables that would perform equally well.

What variables are important in discriminating between the groups? This isn't an easy question with no clear answers recommended in the literature.

We'll deal with five stats that can address this question:

1. Univariate F-ratio
2. F-TO-REMOVE statistics and pr^2
3. Structure Coefficients
4. Standardised discriminant function coefficients
5. Relative Weights

Three Group Example

Relative importance of variables: (1) Univariate F-ratio

Tests of Equality of Group Means

	Wilks' Lambda	F	df1	df2	Sig.
X6 - Product Quality	.570	36.652	2	97	.000
X7 - E-Commerce Activities	.982	.878	2	97	.419
X8 - Technical Support	.981	.917	2	97	.403
X9 - Complaint Resolution	.612	30.782	2	97	.000
X10 - Advertising	.959	2.050	2	97	.134
X11 - Product Line	.556	38.758	2	97	.000
X12 - Salesforce Image	.950	2.549	2	97	.083
X13 - Competitive Pricing	.739	17.172	2	97	.000
X14 - Warranty & Claims	.964	1.803	2	97	.170
X15 - New Products	.985	.730	2	97	.485
X16 - Order & Billing	.715	19.372	2	97	.000
X17 - Price Flexibility	.682	22.563	2	97	.000
X18 - Delivery Speed	.550	39.681	2	97	.000

Seven variables show statistically significant differences univariately ($p<.001$).

Again: these are simply a series of ANOVA's for each discriminant variable and don't take into account the interrelationships between the variables or the effect on the familywise error rate with multiple tests.

Three Group Example

Relative importance of variables: (2) F-TO-REMOVE statistics for each variable

Final Step

Step		Tolerance	F to Remove	Wilks' Lambda
13	X6 – Product Quality	.699	21.381	.264
	X7 – E-Commerce Activities	.326	4.578	.194
	X8 – Technical Support	.322	1.677	.182
	X9 – Complaint Resolution	.338	.043	.176
	X10 – Advertising	.672	.067	.176
	X11 – Product Line	.046	.989	.179
	X12 – Salesforce Image	.265	3.746	.191
	X13 – Competitive Pricing	.730	3.515	.190
	X14 – Warranty & Claims	.310	1.023	.180
	X15 – New Products	.908	1.690	.182
	X16 – Order & Billing	.476	.413	.177
	X17 – Price Flexibility	.042	2.294	.185
	X18 – Delivery Speed	.039	2.027	.184

From Table C.3 on page 948: the critical value of F for = .05 for testing F-TO-REMOVE is $F(2,85) = 3.15$ (actually $F(2,60) = 3.15$ and $F(2,120) = 3.07$, so be conservative and use the larger value).

The degrees of freedom are $[k-1, N - k - p + 1]$.

Four variables are significant using this critical value and contribute uniquely to the separation of the groups in addition to the other variables.

Three Group Example

Relative importance of variables: (2) F-TO-REMOVE statistics for each variable

X6 - Product Quality represents 33.47% of total variance accounted for by the grouping variable *controlling for the other variables.*

Step		Tolerance	F to Remove	Wilks' Lambda	$pr^2\%$ partial η^2
13	X6 – Product Quality	.699	21.381	.264	33.47%
	X7 – E-Commerce Activities	.326	4.578	.194	9.724%
	X8 – Technical Support	.322	1.677	.182	3.796%
	X9 – Complaint Resolution	.338	.043	.176	0.101%
	X10 – Advertising	.672	.067	.176	0.157%
	X11 – Product Line	.046	.989	.179	2.274%
	X12 – Salesforce Image	.265	3.746	.191	8.100%
	X13 – Competitive Pricing	.730	3.515	.190	7.639%
	X14 – Warranty & Claims	.310	1.023	.180	2.350%
	X15 – New Products	.908	1.690	.182	3.824%
	X16 – Order & Billing	.476	.413	.177	0.962%
	X17 – Price Flexibility	.042	2.294	.185	5.121%
	X18 – Delivery Speed	.039	2.027	.184	4.552%

We can use the F-TO-REMOVE values to calculate an estimate of the effect size for the difference between groups for a variable controlling for the other variables. It's equivalent to pr^2, the squared partial-correlation coefficient. For the ith variable controlling for the other variables:

$$pr_i^2 = \frac{\frac{(k-1)F_{tr_i}}{(N-k-p+1)}}{\left(\frac{(k-1)F_{tr_i}}{(N-k-p+1)} + 1\right)} = \frac{\frac{(3-1)(21.381)}{(100-3-13+1)}}{\left(\frac{(3-1)(21.381)}{(100-3-13+1)} + 1\right)} = .335$$

Three Group Example

Relative importance of variables: (3) Structure Coefficients (s)

Structure Matrix

	Function	
	1	2
X11 - Product Line	.566*	-.187
X6 - Product Quality	.496*	-.490
X14 - Warranty & Claims	.118*	-.074
X8 - Technical Support	.088*	.016
X17 - Price Flexibility	-.059	.829*
X18 - Delivery Speed	.491	.592*
X9 - Complaint Resolution	.446	.479*
X16 - Order & Billing	.339	.427*
X13 - Competitive Pricing	-.320	.398*
X12 - Salesforce Image	.046	.267*
X10 - Advertising	.066	.219*
X7 - E-Commerce Activities	-.045	.141*
X15 - New Products	.041	-.128*

Pooled within-groups correlations between discriminating variables and standardized canonical discriminant functions
Variables ordered by absolute size of correlation within function.

*. Largest absolute correlation between each variable and any discriminant function

These simply represent the correlations between the 13 variables and each of the two discriminant functions. They are calculated within each group and then pooled together.

These are handy when you're trying to assign a meaningful label to each function. The first function (for the most part) has to do with: Product line and quality. The second function (for the most part) has to do with: Price flexibility and delivery speed.

These may nicely describe a “macro” label... or not.

Three Group Example

Relative importance of variables: (4) Standardised Discriminant Function Coefficients (d)

Standardized Canonical Discriminant Function Coefficients

	Function	
	1	2
X6 - Product Quality	.816	-.133
X7 - E-Commerce Activities	-.575	-.401
X8 - Technical Support	.000	.544
X9 - Complaint Resolution	-.008	-.086
X10 - Advertising	-.054	-.024
X11 - Product Line	-.140	1.093
X12 - Salesforce Image	.591	.384
X13 - Competitive Pricing	-.363	.167
X14 - Warranty & Claims	.046	-.431
X15 - New Products	-.029	-.322
X16 - Order & Billing	-.123	.155
X17 - Price Flexibility	-.169	1.739
X18 - Delivery Speed	1.097	-.871

These serve the same purpose as beta weights in multiple regression: they indicate the relative importance of the independent variables in predicting the dependent variables.

These indicate the partial contribution of each variable to the discriminant functions, controlling for other independents entered in the equation. The structure coefficients (s) indicate the simple correlations between the variables and the discriminant functions. Use the structure coefficients (s) to assign meaningful labels to the discriminant functions, and the standardised discriminant function coefficients (d) to assess each independent variable's unique contribution to the discriminant function.

Three Group Example

Relative importance of variables: (5) Relative Weights (d x s)

Structure Matrix

	Function 1	Function 2
X11 - Product Line	.566*	-.187
X6 - Product Quality	.496*	-.490
X14 - Warranty & Claims	.118*	-.074
X8 - Technical Support	.088*	.016
X17 - Price Flexibility	-.059	.829*
X18 - Delivery Speed	.491	.592*
X9 - Complaint Resolution	.446	.479*
X16 - Order & Billing	.339	.427*
X13 - Competitive Pricing	-.320	.398*
X12 - Salesforce Image	.046	.267*
X10 - Advertising	.066	.219*
X7 - E-Commerce Activities	-.045	.141*
X15 - New Products	.041	-.128*

Pooled within-groups correlations between discriminating variables and standardized canonical discriminant functions
Variables ordered by absolute size of correlation within function.
*. Largest absolute correlation between each variable and any discriminant function

Standardized Canonical Discriminant Function Coefficients

	Function 1	Function 2
X6 - Product Quality	.816	-.133
X7 - E-Commerce Activities	-.575	-.401
X8 - Technical Support	.000	.544
X9 - Complaint Resolution	-.008	-.086
X10 - Advertising	-.054	-.024
X11 - Product Line	-.140	1.093
X12 - Salesforce Image	.591	.384
X13 - Competitive Pricing	-.363	.167
X14 - Warranty & Claims	.046	-.431
X15 - New Products	-.029	-.322
X16 - Order & Billing	-.123	.155
X17 - Price Flexibility	-.169	1.739
X18 - Delivery Speed	1.097	-.871

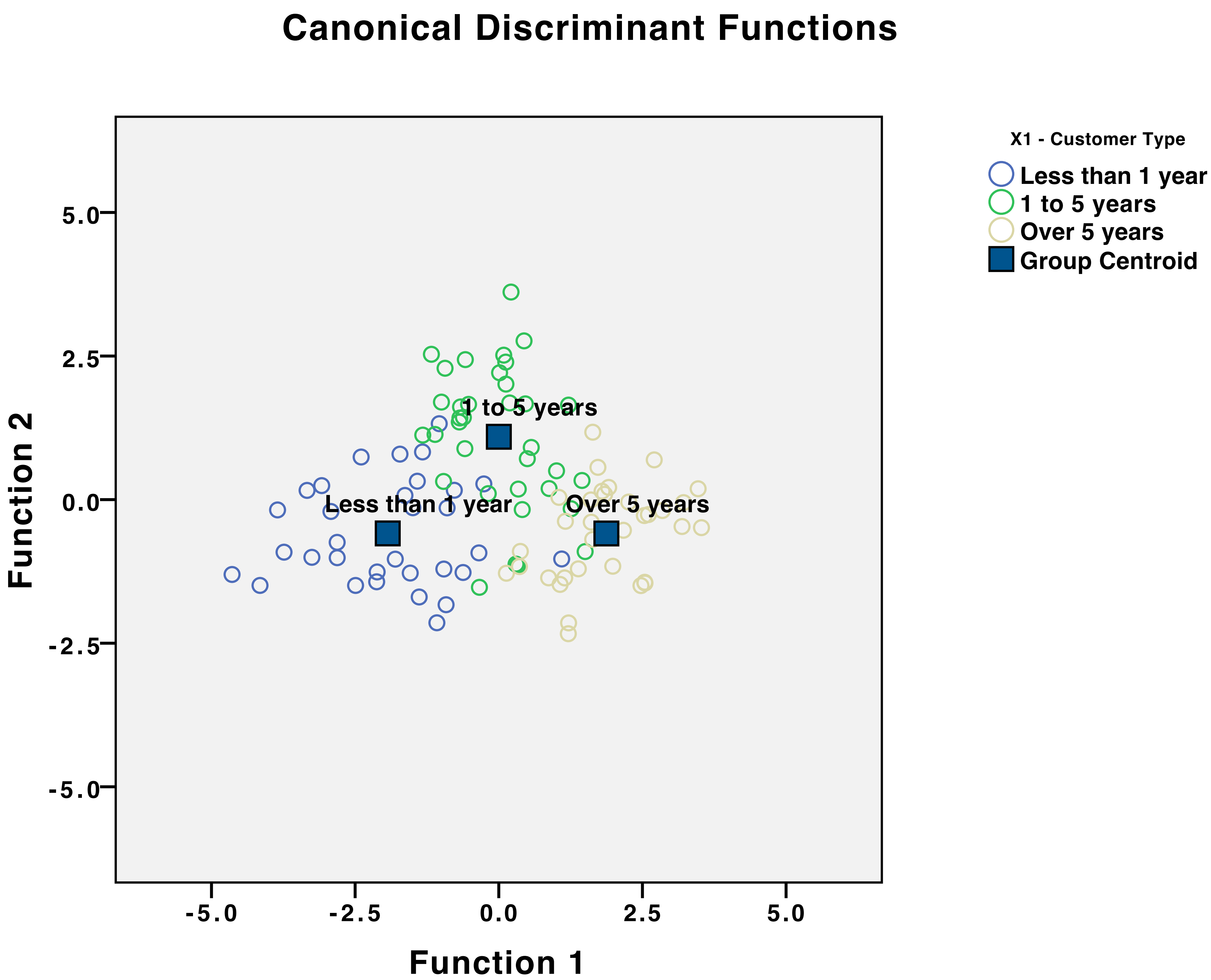
	Function 1	Function 2
X6 - Product Quality	40.48%	6.52%
X7 - E-Commerce Activities	2.59%	-5.65%
X8 - Technical Support	0.00%	0.85%
X9 - Complaint Resolution	-0.35%	-4.13%
X10 - Advertising	-0.36%	-0.53%
X11 - Product Line	-7.89%	-20.46%
X12 - Salesforce Image	2.74%	10.25%
X13 - Competitive Pricing	11.64%	6.63%
X14 - Warranty & Claims	0.54%	3.20%
X15 - New Products	-0.12%	4.14%
X16 - Order & Billing	-4.16%	6.59%
X17 - Price Flexibility	1.00%	144.17%
X18 - Delivery Speed	53.87%	-51.57%
Total	100%	100%

Just multiply the values in the structure matrix by the standardised discriminant function coefficients.

Three Group Example

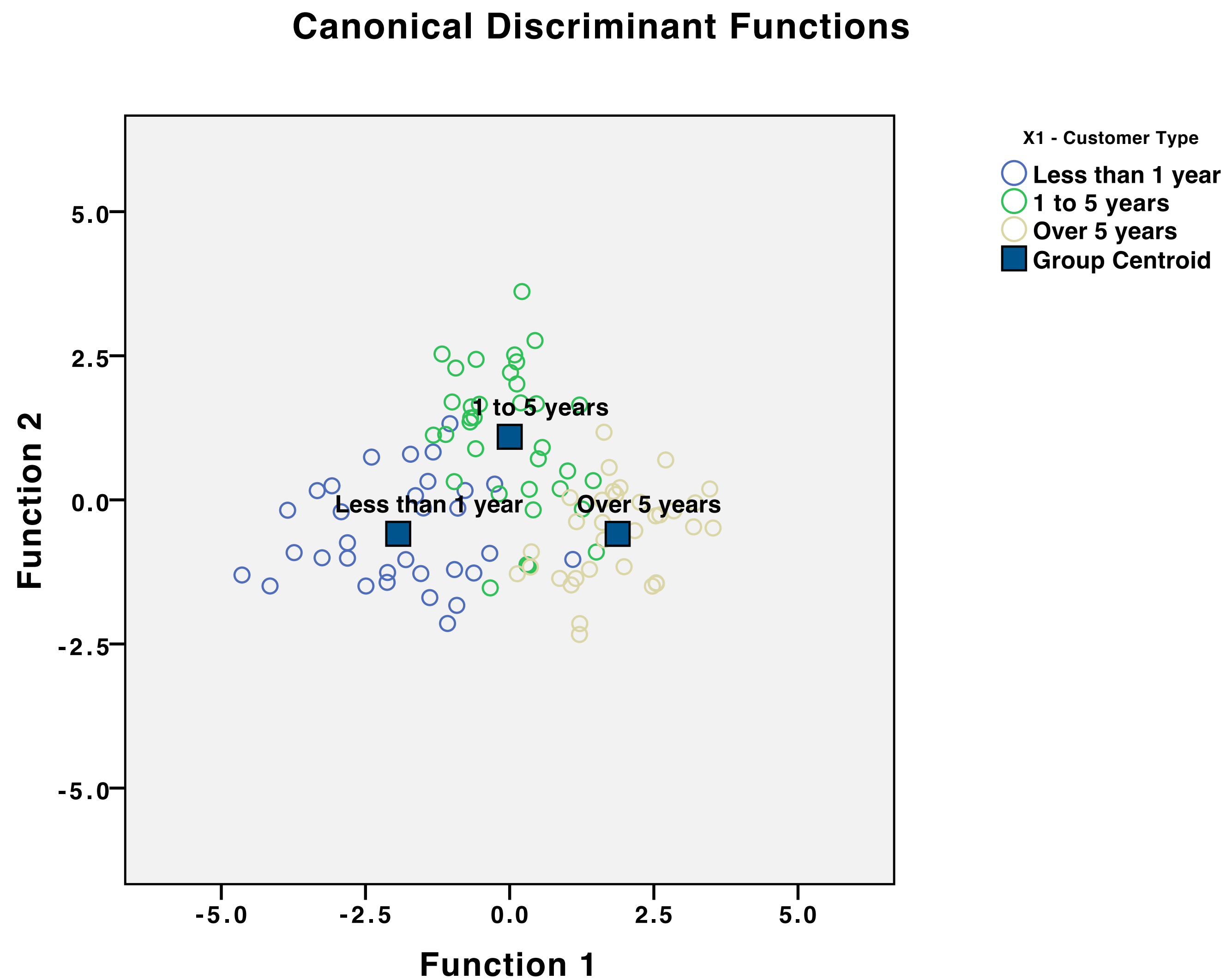
Group separation: Centroid Plots in reduced discriminant space

How are the groups separated? This is answered by plotting the group centroids (looking at the combined-groups plot or plotting them yourselves from the table) and by labelling the discriminant functions with the names of the important variables. This shows the use of discriminant analysis as a data reduction method.

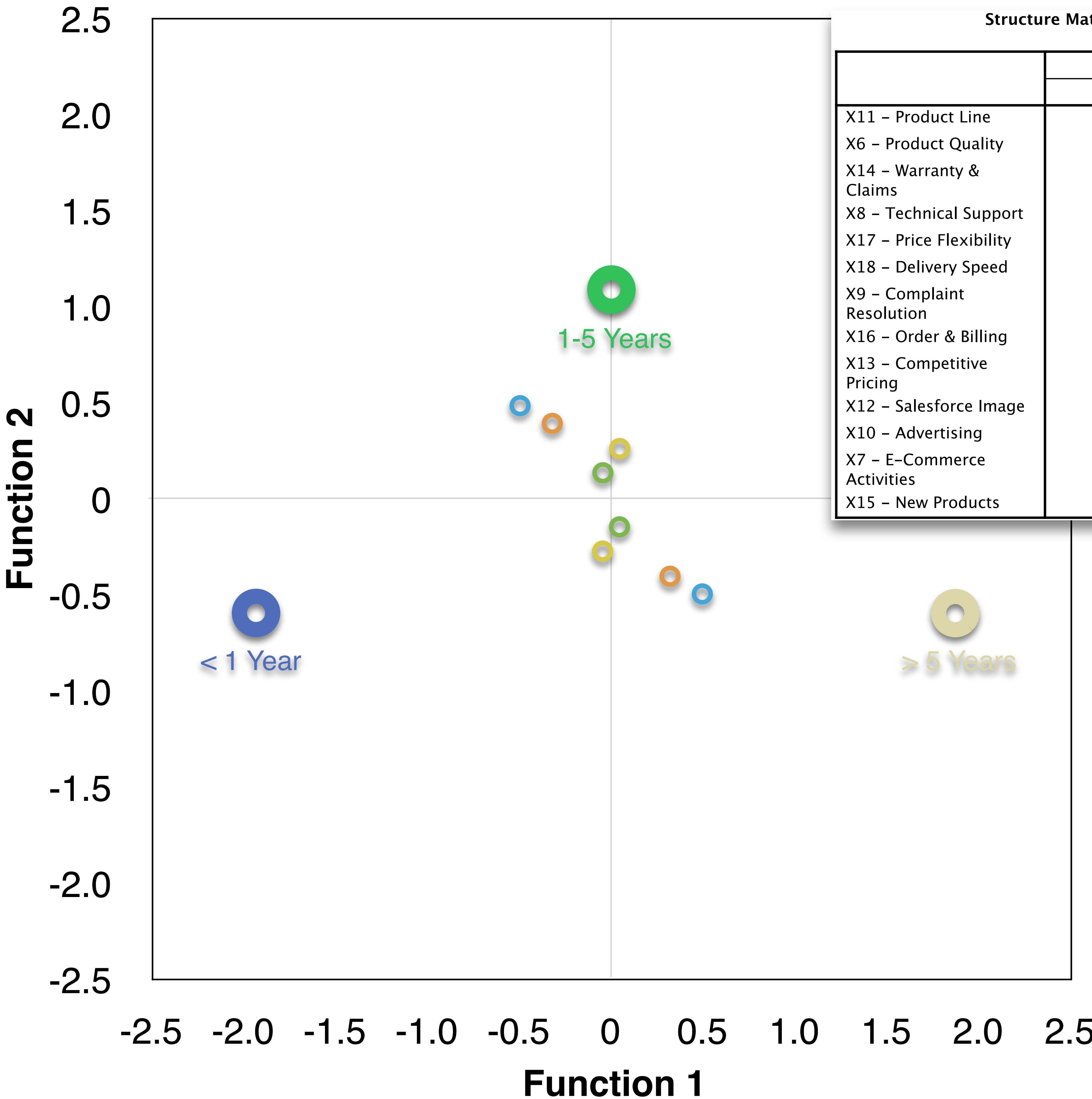


Three Group Example

Group separation: Centroid Plots in reduced discriminant space

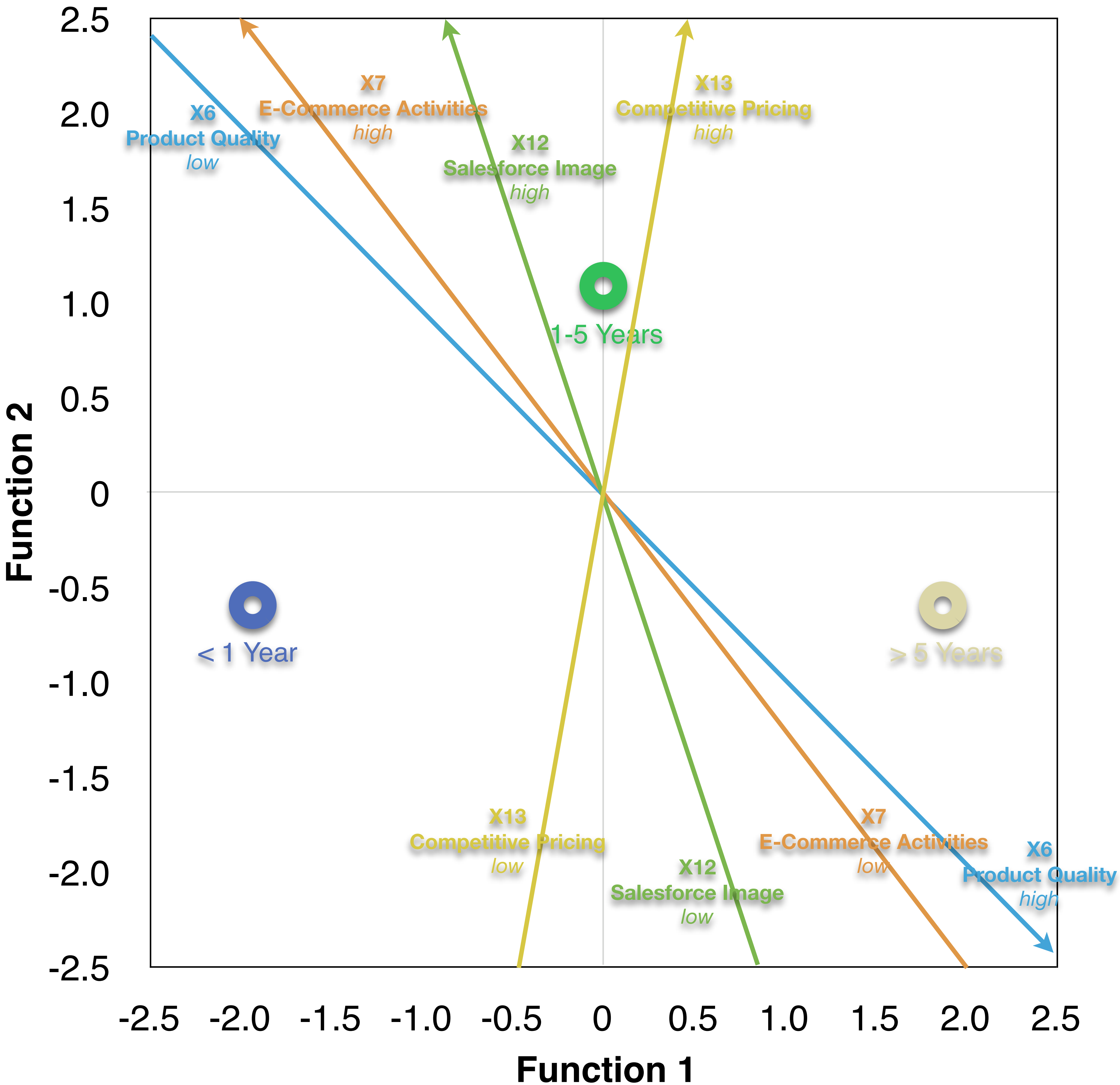


Examining the group centroids and the distribution of cases in each group, we see that Function 1 primarily differentiates between < 1 year vs > 5 years, whereas Function 2 distinguishes between > 5 years vs Groups < 1 and 1-5 years.



Structure Matrix

	Function 1	Function 2
X11 – Product Line	.566*	–.187
X6 – Product Quality	.496*	–.490
X14 – Warranty & Claims	.118*	–.074
X8 – Technical Support	.088*	.016
X17 – Price Flexibility	–.059	.829*
X18 – Delivery Speed	.491	.592*
X9 – Complaint Resolution	.446	.479*
X16 – Order & Billing	.339	.427*
X13 – Competitive Pricing	–.320	.398*
X12 – Salesforce Image	.046	.267*
X10 – Advertising	.066	.219*
X7 – E–Commerce Activities	–.045	.141*
X15 – New Products	.041	–.128*



2.5
2.0
1.5
1.0
0.5
0
-0.5
-1.0
-1.5
-2.0
-2.5
Function 2
-2.5 -2.0 -1.5 -1.0 -0.5 0 0.5 1.0 1.5 2.0 2.5
Function 1
X6
Product Quality
low
X7
E-Commerce Activities
high
X12
Salesforce Image
high
X13
Competitive Pricing
high
1-5 Years
< 1 Year
> 5 Years
X13
Competitive Pricing
low
X12
Salesforce Image
low
X7
E-Commerce Activities
low
X6
Product Quality
high

Three Group Example

Group separation: Discriminant variable mean differences at the group level

Group Statistics

X1 - Customer Type		Mean	Std. Deviation
Less than 1 year	X6 - Product Quality	7.097	1.0219
	X7 - E-Commerce Activities	3.675	.6998
	X8 - Technical Support	5.091	1.6747
	X9 - Complaint Resolution	4.350	.9333
	X10 - Advertising	3.725	1.0122
	X11 - Product Line	4.831	1.0532
	X12 - Salesforce Image	4.863	.9517
	X13 - Competitive Pricing	7.491	1.2830
	X14 - Warranty & Claims	5.894	.9442
	X15 - New Products	5.144	1.6582
	X16 - Order & Billing	3.559	.8728
	X17 - Price Flexibility	4.234	1.0012
	X18 - Delivery Speed	3.172	.6150
1 to 5 years	X6 - Product Quality	7.240	1.3720
	X7 - E-Commerce Activities	3.780	.6521
	X8 - Technical Support	5.391	1.5056
	X9 - Complaint Resolution	5.943	.8876
	X10 - Advertising	4.277	1.1083
	X11 - Product Line	5.603	.9922
	X12 - Salesforce Image	5.431	.9539
	X13 - Competitive Pricing	7.557	1.3834
	X14 - Warranty & Claims	5.977	.8178
	X15 - New Products	4.940	1.3574
	X16 - Order & Billing	4.649	.7493
	X17 - Price Flexibility	5.523	1.1840
	X18 - Delivery Speed	4.243	.5731
Over 5 years	X6 - Product Quality	9.106	.6509
	X7 - E-Commerce Activities	3.555	.7517
	X8 - Technical Support	5.603	1.4095
	X9 - Complaint Resolution	5.970	1.0406
	X10 - Advertising	4.003	1.2133
	X11 - Product Line	6.964	.9243
	X12 - Salesforce Image	5.048	1.2392
	X13 - Competitive Pricing	5.855	1.3514
	X14 - Warranty & Claims	6.258	.6558
	X15 - New Products	5.379	1.4741
	X16 - Order & Billing	4.582	.7568
	X17 - Price Flexibility	4.006	.7814
	X18 - Delivery Speed	4.200	.4500
Total	X6 - Product Quality	7.810	1.3963
	X7 - E-Commerce Activities	3.672	.7005
	X8 - Technical Support	5.365	1.5305
	X9 - Complaint Resolution	5.442	1.2084
	X10 - Advertising	4.010	1.1269
	X11 - Product Line	5.805	1.3153
	X12 - Salesforce Image	5.123	1.0723
	X13 - Competitive Pricing	6.974	1.5451
	X14 - Warranty & Claims	6.043	.8197
	X15 - New Products	5.150	1.4930
	X16 - Order & Billing	4.278	.9288
	X17 - Price Flexibility	4.610	1.2060
	X18 - Delivery Speed	3.886	.7344

X1 Customer Type	X6 Product Quality	X7 E-Commerce Activities	X12 Salesforce Image	X13 Competitive Pricing
Less than 1 year	7.097	3.675	4.863	7.491
1 to 5 years	7.240	3.780	5.431	7.557
Over 5 years	9.106	3.555	5.048	5.855
Total	7.810	3.672	5.123	6.974

Three Group Example

Group separation: Matrix of pairwise F values and Group means

Step	X1 - Customer Type		Less than 1 year	1 to 5 years	Over 5 years
13	Less than 1 year	F		7.405	15.856
		Sig.		.000	.000
	1 to 5 years	F	7.405		7.257
		Sig.	.000		.000
	Over 5 years	F	15.856	7.257	
		Sig.	.000	.000	

The matrix of pairwise F values between the groups tests which groups are different from one another over all the variables. This can be useful when describing the differences between the groups in the group-centroid plot.

Three Group Example

Classification: Prediction of group membership

How well do the discriminant functions predict group membership?

Classification Results[a]

		X1 - Customer Type	Predicted Group Membership: Less than 1 year	Predicted Group Membership: 1 to 5 years	Predicted Group Membership: Over 5 years	Total
Original	Count	Less than 1 year	27	4	1	32
		1 to 5 years	1	29	5	35
		Over 5 years	0	1	32	33
	%	Less than 1 year	84.4	12.5	3.1	100.0
		1 to 5 years	2.9	82.9	14.3	100.0
		Over 5 years	.0	3.0	97.0	100.0

a. 88.0% of original grouped cases correctly classified.