Demonstration of 3-group Linear Discriminant Function – Concentrated Structure

All subjects underwent therapy for depression. The therapists categorized each patient as "resistant", "compliant", or "engaged" (coded as ptype 1-3, respectively). DVs (predictors) were ratings by a panel of "blind" therapists obtained at the end of 6 months of therapy. These DVs were : 1) overall psychological wellness, 2) depression change, 3) new activities, 4) new acquaintances.

Analyze → Classify → Discriminant

Group Statistics

GROUP		Mean	Std. Deviation	Valid N (listwise) Unweighted	Valid N (listwise) Weighted
resistant	change in overall psychological wellness score	17.2170	2.52070	18	18.000
	change in depression score	3.7130	1.48240	18	18.000
	number of new social activities	1.4712	.98676	18	18.000
	number of new acquaintances	.8053	.16964	18	18.000
compliant	change in overall psychological wellness score	18.4518	2.55300	16	16.000
	change in depression score	5.8487	1.91926	16	16.000
	number of new social activities	3.2208	1.07798	16	16.000
	number of new acquaintances	1.6462	.32335	16	16.000
engaged	change in overall psychological wellness score	17.5045	3.21838	17	17.000
	change in depression score	5.3690	1.80807	17	17.000
	number of new social activities	3.4122	1.24532	17	17.000
	number of new acquaintances	1.6537	.27900	17	17.000
Total	change in overall psychological wellness score	17.7002	2.77634	51	51.000
	change in depression score	4.9351	1.94091	51	51.000
	number of new social activities	2.6671	1.40630	51	51.000
	number of new acquaintances	1.3519	.48198	51	51.000

Before moving on to the multivariate analysis, consider the univariate and bivariate results.

The F-tests tell us that there are between group differences for three of the DVs, so we certainly expect that there will be a significant multivariate effect.

But, do these three groups have a concentrated or a diffuse structure when we consider these DVs?.

If we look at the pattern of mean differences among the groups for depchang, we see that the compliant and engaged groups had higher means than did the resistant patients.

We can also see that there is the same basic pattern of mean differences for newacts and newacts.

When the DVs show about the same pattern of mean differences across groups, we expect to find that there is a concentrated multivariate structure.

Tests of Equality of Group Means

	Wilks' Lambda	F	df1	df2	Sig.
change in overall psychological wellness score	.964	.897	2	48	.414
change in depression score	.769	7.194	2	48	.002
number of new social activities	.595	16.363	2	48	.000
number of new acquaintances	.284	60.397	2	48	.000

Eigenvalues

Function	Eigenvalue	% of Variance	Cumulative %	Canonical Correlation
1	3.952[a]	98.9	98.9	.893
2	.046[a]	1.1	100.0	.209

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

Wilks' Lambda

Test of Function(s)	Wilks' Lambda	Chi-square	df	Sig.
1 through 2	.193	76.477	8	.000
2	.956	2.083	3	.555

The significance tests tell us that there is a single discriminant function – a concentrated structure among these 3 groups, as we anticipated from the group's mean profiles. That ldf accounts for $.893^2 = .80$ percent of the between group variance.

With 3 groups, and so, 2 possible ldfs, the "% of Variance" statistic becomes meaningful. This tells the proportion of the between group variance that is accounted for by each ldf. In other words, it tells the relative contribution of each ldf to the model. When combined with the canonical correlation, it can be helpful when deciding whether or not the second (or third) ldf has a "meaningful" contribution, especially when there might be statistical power issues. In this case, 98.9% of the between group variation that is accounted for is accounted for by the 1st ldf, so it looks like the second ldf isn't significant, very strong (though R_c=.20 isn't tiny) or contributing much relatively speaking.

So, we know the model "works" with a single ldf and the R_c^2 tells us "how well" the model works. Now we need to "interpret" the ldf, by looking at what variables correlate with it and contribute to it. Remember, we'll only interpret the first ldf – SPSS will provide information about the 2nd, but we've decided it isn't statistically reliable function.

Structure Matrix

	Function	
	1	2
number of new acquaintances	.697*	-.406
change in depression score	.470	.629*
number of new social activities	.412	-.517*
change in overall psychological wellness score	.172	.415*

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

andardized Canonical Discriminant Function Coefficien

	Function	
	1	2
change in overall psychological wellness score	.045	.707
change in depression score	.164	.629
number of new social activities	.158	-.592
number of new acquaintances	.476	.029

The structure matrix shows that new acquaintances, change in depression scores, and new activities are correlated with the ldf, while change in wellness scores isn't. The standardized coefficients tell us that number of new acquaintances has the only large unique contribution.

Functions at Group Centroids

	Function	
GROUP	1	2
resistant	-2.609	1.332E-02
compliant	1.540	.259
engaged	1.313	-.257

Unstandardized canonical discriminant functions evaluated at group means

The centroids give the group mean on the ldf.

Remember that the weightings of the predictors to compose the ldf are chosen so that the F-test of these centroids is as large as possible.

This is done by selecting weightings that simultaneously maximize the between group centroids differences and minimize the within group variability on the ldf scores. This combination gives the least "overlap" among the group's ldf scores, maximizing classification accuracy.

We can combine the information from the structure weights with the group centroids, to develop a graphical depiction of this ldf.

resistant compliant engaged

-1 0 1

fewer new activities	more new activities
fewer new acquaintances	more new acquaintances
less depression change	more depression change

From this graphic we would anticipate that we can use this ldf to discriminate the resistant group from the other two with considerable accuracy, but might not do so well discriminating between the compliant and engaged groups. We can also look at the reclassification table to evaluate the classification accuracy of this ldf, as well as any pattern of pairwise group discriminability.

Classification Results[a]

			Predicted Group Membership			
		GROUP	resistant	compliant	engaged	Total
Original	Count	resistant	18	0	0	18
		compliant	0	10	6	16
		engaged	1	5	11	17
	%	resistant	100.0	.0	.0	100.0
		compliant	.0	62.5	37.5	100.0
		engaged	5.9	29.4	64.7	100.0

a. 76.5% of original grouped cases correctly classified.

Notice that the resistant group is not often confused with the other two groups, whereas there is considerable confusion between the compliant and engaged groups.

Write-up:

The write-up for a k-group ldf proceeds much like that for a 2-group, except…

- must explicitly tell whether each of the possible ldf "works" (and report the significance test and R^2 for each)
- must be sure to tell which groups are and are not differentiated using the ldf(s)

Discriminant analysis was used to determine if therapy patients identified as resistant, compliant and engaged differed in terms of the number of new acquaintances, number of new social activities, decrease in depressive symptomology, and general psychological wellness. Table 1 presents a summary of the univariate and bivariate analyses. Three of the measures produced significant differences between the groups; only general wellness did not.

Multivariate analysis revealed that the first discriminant function reliably differentiated among the patient groups, $\lambda = .193$, $X^2(8) = 76.477$, $p < .001$, R^2-canonical = .80, but that the second function did not provide reliable further differentiation ($\lambda = .956$, $X^2(3) = 2.083$, $p = .555$, R^2-canonical = .044). Table 2 shows the structure weights for the first discriminant function, revealing that, consistent with the bivariate results, the number of new acquaintances, number of new social activities, and decrease in depressive symptomology contributed to the discrimination among the groups. Inspection of the standardized canonical coefficients also shown in Table 2 reveals that, because of the collinearity between the new acquaintances and new activities measures, only the depression change measure has a strong unique contribution to the function.

Figure 1 gives a graphical depiction of the multivariate results and Table 3 shows the results with the ldf was used to re-classify patients into their treatment groups. Both of these show that the discriminant function did very well at differentiating the resistant patients, who had fewer new acivities, fewer new acquaintances, and less depression change, from the other two groups. However, there was little success at differentiating between the compliant and engaged patients.

Table 1. Means (standard deviations) for each treatment group and related F-tests.

Table 2. Standardized Canonical Coefficients and Structure weights for the discriminant function.

Table 3. Results from re-classifying patients into groups.

Figure 1. Graphical Representation of ldf