Demonstration of ldf and MANOVA on the same data set

Let's start with an ldf looking at discriminating among students of three types of high school academic programs (general education, vocational-technical training and college preparation) based on five standardized measures of topical knowledge.

Analyze ➔ Classify ➔ Discriminant

Group Statistics

high school program		Mean	Std. Deviation
general	standardized reading test	47.8407	8.80170
	standardized writing test	49.0000	9.87808
	standardized math test	48.4519	8.33333
	standardized science test	49.3593	8.75607
	standardized civics test	50.0111	9.30216
academic prep	standardized reading test	55.5881	9.96012
	standardized writing test	54.6373	9.09049
	standardized math test	54.8150	9.42443
	standardized science test	53.3238	9.30924
	standardized civics test	55.3326	9.14935
voctech	standardized reading test	45.4061	9.00463
	standardized writing test	46.1364	10.11129
	standardized math test	45.4535	7.52200
	standardized science test	44.8267	10.31059
	standardized civics test	45.0758	8.96442
Total	standardized reading test	50.9763	10.44391
	standardized writing test	51.0112	10.21738
	standardized math test	50.7800	9.57300
	standardized science test	50.1503	10.02228
	standardized civics test	51.3573	10.05730

Tests of Equality of Group Means

	Wilks' Lambda	F	df1	df2	Sig.
standardized reading test	.811	46.349	2	397	.000
standardized writing test	.872	29.110	2	397	.000
standardized math test	.821	43.208	2	397	.000
standardized science test	.880	27.118	2	397	.000
standardized civics test	.823	42.753	2	397	.000

Test for Homogeneity – precursor to multivariate tests

Log Determinants

high school program	Rank	Log Determinant
general	5	20.056
academic prep	5	20.089
voctech	5	20.061
Pooled within-groups	5	20.196

The ranks and natural logarithms of determinants printed are those of the group covariance matrices

Test Results

Box's M		48.699
F	Approx.	1.591
	df1	30
	df2	306329.0
	Sig.	.021

Tests null hypothesis of equal population covariance matrices

When computing an ldf the matrices of correlations among the predictors for the k groups are pooled. The homogeneity test is designed to test whether they are "similar enough" to do this reasonably.

There are two *concerns* among those to "don't like" Box's M test:

1. As the sample size increases this test becomes very powerful, and so, identify trivial differences (especially with large sample sizes)
   - look at the similarity of the Log Determinants for the three groups, yet these are "highly significantly different" based on a p = .021
   - consider the df – they are very large, making the F-critical very small
2. The log determinant is a poor summary of a correlation matrix, and so is a poor basis for evaluating the equivalence of the correlation matrices of the groups

Eigenvalues

Function	Eigenvalue	% of Variance	Cumulative %	Canonical Correlation
1	.324[a]	92.8	92.8	.495
2	.025[a]	7.2	100.0	.157

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	.737	120.802	10	.000
2	.975	9.884	4	.042

The significance tests show two significant ldfs, though the canonical correlation and % of variance for the second is low.

The ß and structure weights, along with the centroids, allow us to interpret the ldfs.

Standardized Canonical Discriminant Function Coefficients

	Function	
	1	2
standardized reading test	.413	-.993
standardized writing test	.021	-.048
standardized math test	.377	-.280
standardized science test	-.060	.988
standardized civics test	.454	.582

Structure Matrix

	Function	
	1	2
standardized reading test	.844*	-.306
standardized math test	.819*	-.074
standardized civics test	.807*	.401
standardized writing test	.673*	-.022
standardized science test	.637*	.447

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

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

Functions at Group Centroids

high school program	Function	
	1	2
general	-.305	.246
academic prep	.562	-4.77E-02
voctech	-.764	-.176

Unstandardized canonical discriminant functions evaluated at group means

Classification Results[a]

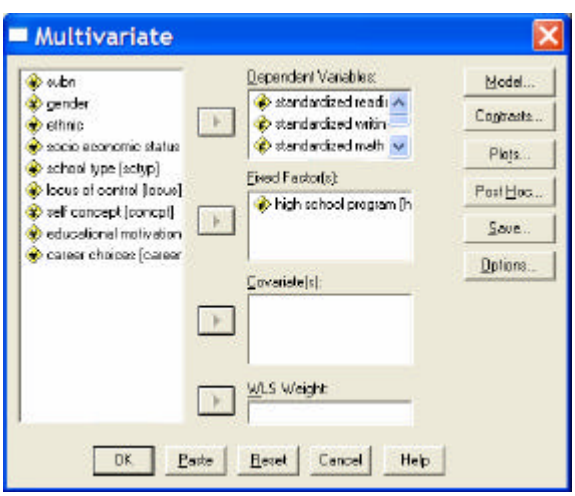
			Predicted Group Membership			
		high school program	general	academic prep	voctech	Total
Original	Count	general	35	31	42	108
		academic prep	38	124	31	193
		voctech	15	21	63	99
	%	general	32.4	28.7	38.9	100.0
		academic prep	19.7	64.2	16.1	100.0
		voctech	15.2	21.2	63.6	100.0

a. 55.5% of original grouped cases correctly classified.

The classification results converge with the centroids to suggest that while there is pretty good discrimination between the academic prep and votech groups, both of these groups are difficult to discriminate from the general education group – at least when considering only these five academic performance measures.

MANOVA

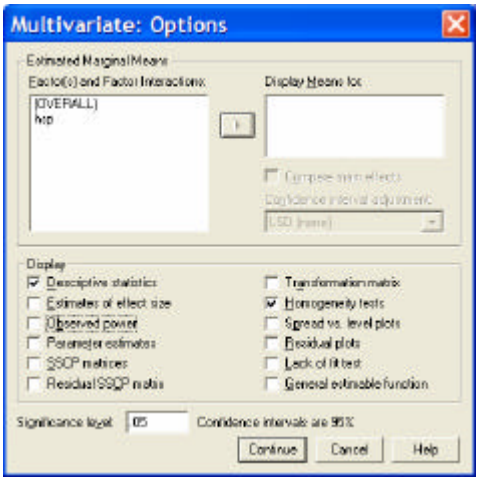
Analyze ➔ General Linear Model ➔ Multivariate



Move the multiple DVs into the Dependent Variables window.

Move the variable that defines the groups into the Fixed Factor(s) window.

You'll want to specify details of the analysis and output using the "Options" window



Descriptive statistics will get you the univariate stats of each group for each DV.

Homogeneity tests gets you the Box's M and Levine tests of homogeneity of variance and covariance.

Same values as we got from ldf – just arranged a bit differently.

Same bivariate ANOVA results, but with "all the trimmings" – Sum of Squares and Mean Squares values, etc.

Tests of Between-Subjects Effects

Source	Dependent Variable	Type III Sum of Squares	df	Mean Square	F	Sig.
Corrected Model	standardized reading test	8238.385[a]	2	4119.192	46.349	.000
	standardized writing test	5327.179[b]	2	2663.589	29.110	.000
	standardized math test	6536.458[c]	2	3268.229	43.208	.000
	standardized science test	4817.111[d]	2	2408.556	27.118	.000
	standardized civics test	7152.086[e]	2	3576.043	42.753	.000
Intercept	standardized reading test	902622.806	1	902622.806	10156.311	.000
	standardized writing test	914045.414	1	914045.414	9989.336	.000
	standardized math test	901234.899	1	901234.899	11914.887	.000
	standardized science test	886621.683	1	886621.683	9982.420	.000
	standardized civics test	921945.355	1	921945.355	11022.305	.000
HSP	standardized reading test	8238.385	2	4119.192	46.349	.000
	standardized writing test	5327.179	2	2663.589	29.110	.000
	standardized math test	6536.458	2	3268.229	43.208	.000
	standardized science test	4817.111	2	2408.556	27.118	.000
	standardized civics test	7152.086	2	3576.043	42.753	.000
Error	standardized reading test	35282.620	397	88.873		
	standardized writing test	36326.340	397	91.502		
	standardized math test	30028.842	397	75.639		
	standardized science test	35260.870	397	88.818		
	standardized civics test	33206.513	397	83.644		
Total	standardized reading test	1082952.230	400			
	standardized writing test	1082512.570	400			
	standardized math test	1068008.660	400			
	standardized science test	1046101.024	400			
	standardized civics test	1095385.450	400			
Corrected Total	standardized reading test	43521.004	399			
	standardized writing test	41653.519	399			
	standardized math test	36565.300	399			
	standardized science test	40077.982	399			
	standardized civics test	40358.599	399			

a. R Squared = .189 (Adjusted R Squared = .185)

b. R Squared = .128 (Adjusted R Squared = .123)

c. R Squared = .179 (Adjusted R Squared = .175)

d. R Squared = .120 (Adjusted R Squared = .116)

e. R Squared = .177 (Adjusted R Squared = .173)

Box's Test of Equality of Covariance Matrices[a]

Box's M	48.699
F	1.591
df1	30
df2	306329.0
Sig.	.021

Tests the null hypothesis that the observed covariance matrices of the dependent variables are equal across groups.

a. Design: Intercept+HSP

Levene's Test of Equality of Error Variances[a]

	F	df1	df2	Sig.
standardized reading test	1.211	2	397	.299
standardized writing test	2.076	2	397	.127
standardized math test	5.859	2	397	.003
standardized science test	1.266	2	397	.283
standardized civics test	.223	2	397	.800

Tests the null hypothesis that the error variance of the dependent variable is equal across groups.

a. Design: Intercept+HSP

And the multivariate tests…

Multivariate Tests[c]

Effect		Value	F	Hypothesis df	Error df	Sig.
Intercept	Pillai's Trace	.978	3415.512[a]	5.000	393.000	.000
	Wilks' Lambda	.022	3415.512[a]	5.000	393.000	.000
	Hotelling's Trace	43.454	3415.512[a]	5.000	393.000	.000
	Roy's Largest Root	43.454	3415.512[a]	5.000	393.000	.000
HSP	Pillai's Trace	.270	12.274	10.000	788.000	.000
	Wilks' Lambda	.737	12.987[a]	10.000	786.000	.000
	Hotelling's Trace	.350	13.702	10.000	784.000	.000
	Roy's Largest Root	.324	25.547[b]	5.000	394.000	.000

a. Exact statistic

b. The statistic is an upper bound on F that yields a lower bound on the significance level.

c. Design: Intercept+HSP

These tests are identical when there are only 2 groups. They usually provide equivalent conclusions with multiple groups

Wilk's, Hotelling's & Pillai's all pool information from all available dimensions, whereas Roy's tests only the first dimension – and so is weaker than the others if there is a diffuse structure

Pillai's is more "robust" in situations of heterogeneity, unequal samples and/or unequal samples.

Choosing between ldf and MANOVA

One "tradition" is based on the types of groups:

- Use ldf for naturally occurring groups and MANOVA for experimentally created groups

Another tradition is to use MANOVA as the "omnibus multivariate significance test" and then, if there is a significant multivariate effect use the ldf as "follow-up analyses"

Keep in mind that the two models are statistically and mathematically identical !! All the apparent differences are the results of decisions about what to include in the output.

- Both models extract ? as the basic summary statistic which is then converted to X^2 for a significance test in ldf and to F (in differing versions) in MANOVA
- Older versions of SPSS provided canonical correlations, ß and structure weights, and even the option to save "canonical variate" scores (i.e., all the output from ldf except for the reclassification table)
- The multiple multivariate tests available from MANOVA can be helpful, if they differ importantly and the homogeneity tests reject H0:, then it is probably time more closely examine how the groups differ besides mean differences.