Example of Multiple-group ldf – with Follow-up Analyses

In this example, three sections of a research methods class were conducted using three different formats for test preparation. Group 1 was a "control group" that received the lectures, and took the exams; Group 2 received a steady stream of homework assignments, which were similar to items which appeared on the exams; Group 3 received no homework assignments, but completed "exam preps" that was similar to items which appeared on the exam. There were four "DVs" for this analysie: total points from the quizzes, Midterm Exam #1, Midterm Exam 2, and the Final (cummulative) Exam.

Group Statistics

GROUP		Mean	Std. Deviation	Valid N (listwise) Unweighted	Weighted
lecture	QUIZ	518.8628	108.77367	20	20.000
	EXAM1	43.6497	9.23430	20	20.000
	EXAM2	39.6246	10.00930	20	20.000
	FINAL	94.4252	8.10272	20	20.000
homework	QUIZ	594.3515	71.29752	20	20.000
	EXAM1	50.9138	8.81464	20	20.000
	EXAM2	49.5690	9.64761	20	20.000
	FINAL	99.4550	6.23406	20	20.000
examprep	QUIZ	472.3838	133.61485	20	20.000
	EXAM1	59.6089	7.27711	20	20.000
	EXAM2	52.9495	11.57586	20	20.000
	FINAL	130.6873	6.76939	20	20.000
Total	QUIZ	528.5327	117.32600	60	60.000
	EXAM1	51.3908	10.62168	60	60.000
	EXAM2	47.3810	11.74384	60	60.000
	FINAL	108.1892	17.60895	60	60.000

Tests of Equality of Group Means

	Wilks' Lambda	F	df1	df2	Sig.
QUIZ	.813	6.539	2	57	.003
EXAM	.616	17.741	2	57	.000
EXAM	.764	8.796	2	57	.000
FINAL	.156	154.028	2	57	.000

Eigenvalues

Function	Eigenvalue	% of Variance	Cumulative %	Canonical Correlation
1	7.110[a]	96.0	96.0	.936
2	.295[a]	4.0	100.0	.477

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	.095	130.514	8	.000
2	.772	14.348	3	.002

Structure Matrix

	Function 1	2
FINAL	.869*	-.338
QUIZ	-.116	.673*
EXAM2	.173	.572*
EXAM1	.286	.380*

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

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

Functions at Group Centroids

GROUP	Function 1	2
lecture	-2.454	-.557
homework	-1.143	.712
examprep	3.597	-.154

Unstandardized canonical discriminant functions evaluated at group means

Functions at Group Centroids

GROUP	Function 1	2
lecture	-2.454	-.557
homework	-1.143	.712
examprep	3.597	-.154

Unstandardized canonical discriminant functions evaluated at group means

Classification Results[a]

		GROUP	Predicted Group Membership lecture	homework	examprep	Total
Original	Count	lecture	17	3	0	20
		homework	6	14	0	20
		examprep	0	0	20	20
	%	lecture	85.0	15.0	.0	100.0
		homework	30.0	70.0	.0	100.0
		examprep	.0	.0	100.0	100.0

a. 85.0% of original grouped cases correctly classified.

So, we have two ldfs, that seem to do a pretty good job of discriminating between the groups. However, some presentations of these results would benefit from more formal tests of the "contributions" of the two ldfs, to the discrimination among the groups. There are three common types of such "follow-up analyses" for multiple group discriminant analyses:

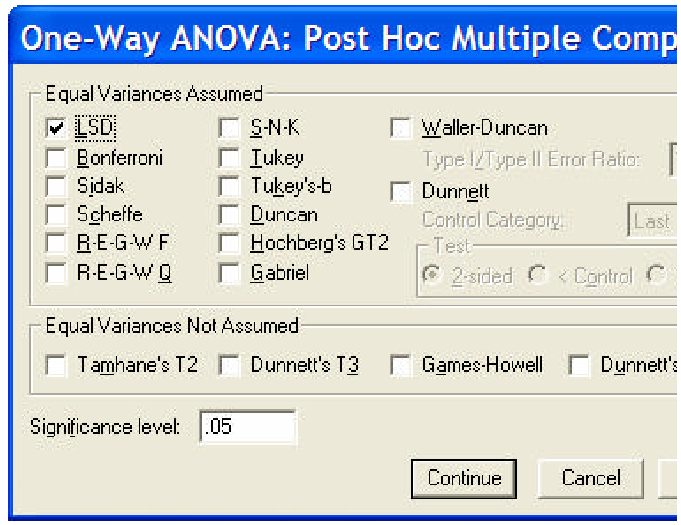
1) bivariate follow-ups – emphasis returns to how the groups differ on each of the DVs
2) multivariate pairwise ldf analyses – which groups are differentiable using which ldfs
3) multivariate pairwise group analyses – building separate ldf models for each pair of groups

Here are Examples of each of the approaches:

Bivariate Follow-ups:

Analyze → Compare Means → One-way ANOVA

Click on "Post Hoc"



ANOVA

QUIZ

	Sum of Squares	df	Mean Square	F	Sig.
Between Groups	151566.5	2	75783.262	6.539	.003
Within Groups	660591.5	57	11589.325		
Total	812158.1	59			

Multiple Comparisons

Dependent Variable: QUIZ
LSD

(I) GROUP	(J) GROUP	Mean Difference (I-J)	Std. Error	Sig.	95% Confidence Interval Lower Bound	95% Confidence Interval Upper Bound
lecture	homework	-75.4887*	34.04310	.031	-143.6589	-7.3186
	examprep	46.4790	34.04310	.178	-21.6911	114.6491
homework	lecture	75.4887*	34.04310	.031	7.3186	143.6589
	examprep	121.9677*	34.04310	.001	53.7976	190.1379
examprep	lecture	-46.4790	34.04310	.178	-114.6491	21.6911
	homework	-121.9677*	34.04310	.001	-190.1379	-53.7976

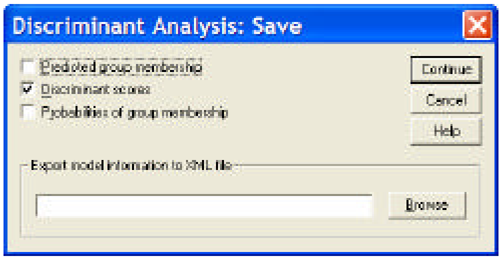
*. The mean difference is significant at the .05 level.

Using LSD has the advantage that you get the exact probabilities for each pairwise comparison. Then you can apply whatever amount of inflation control you think appropriate.

This approach emphasizes "specificity" but doesn't take advantage of any multivariate analyses.

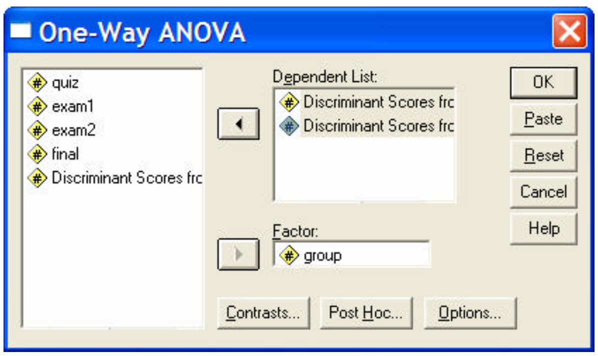
Pairwise ldf Follow-ups:

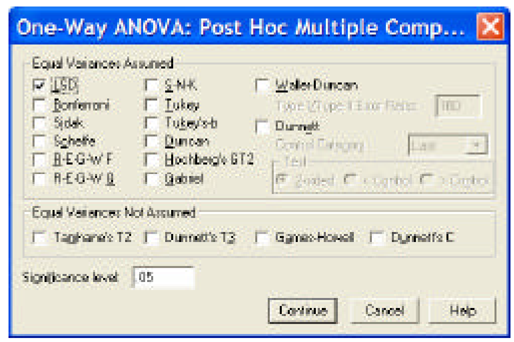
When getting the ldf analysis click the “Save” button and check “Discriminant scores”



This approach is an obvious extension of the descriptive procedures we were using earlier. It emphasizes the ldfs that were identified and interpreted, and gives statistical information about which groups can be discriminated based on each ldf. Remember, larger F ~ less overlap ~ better classification.

Then use oneway to get pairwise comparisons using these ldf scores as the DVs.





ANOVA

		Sum of Squares	df	Mean Square	F	Sig.
Discriminant Scores from Function 1 for Analysis 1	Between Groups	405.263	2	202.632	202.632	.000
	Within Groups	57.000	57	1.000		
	Total	462.263	59			
Discriminant Scores from Function 2 for Analysis 1	Between Groups	16.815	2	8.408	8.408	.001
	Within Groups	57.000	57	1.000		
	Total	73.815	59			

Multiple Comparisons

LSD

Dependent Variable	(I) GROUP	(J) GROUP	Mean Difference (I-J)	Std. Error	Sig.	95% Confidence Interval Lower Bound	95% Confidence Interval Upper Bound
Discriminant Scores from Function 1 for Analysis 1	lecture	homework	-1.3107532*	.31622777	.000	-1.9439883	-.6775180
		examprep	-6.0503915*	.31622777	.000	-6.6836267	-5.4171563
	homework	lecture	1.3107532*	.31622777	.000	.6775180	1.9439883
		examprep	-4.7396383*	.31622777	.000	-5.3728735	-4.1064031
	examprep	lecture	6.0503915*	.31622777	.000	5.4171563	6.6836267
		homework	4.7396383*	.31622777	.000	4.1064031	5.3728735
Discriminant Scores from Function 2 for Analysis 1	lecture	homework	-1.2689528*	.31622777	.000	-1.9021880	-.6357176
		examprep	-.4032513	.31622777	.207	-1.0364865	.2299839
	homework	lecture	1.2689528*	.31622777	.000	.6357176	1.9021880
		examprep	.8657015*	.31622777	.008	.2324663	1.4989367
	examprep	lecture	.4032513	.31622777	.207	-.2299839	1.0364865
		homework	-.8657015*	.31622777	.008	-1.4989367	-.2324663

*. The mean difference is significant at the .05 level.

Pairwise group comparisons

This involves getting separate 2-group analyses for each pair of groups. It can be especially helpful when the overall ldf doesn't discriminate between one or more group pairs.

You have to recode the group variable to get comparison of nonadjacent groups, like this.

Recode group (1=4) (3=5). Then ask for a discriminant analysis of groups 4 & 5.

Eigenvalues

Function	Eigenvalue	% of Variance	Cumulative %	Canonical Correlation
1	.987[a]	100.0	100.0	.705

a. First 1 canonical discriminant functions were used in the analysis.

Wilks' Lambda

Test of Function(s)	Wilks' Lambda	Chi-square	df	Sig.
1	.503	24.711	4	.000

Structure Matrix

	Function
	1
EXAM2	.522
QUIZ	.424
EXAM1	.416
FINAL	.359

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

Standardized Canonical Discriminant Function Coefficients

	Function
	1
QUIZ	.838
EXAM1	.713
EXAM2	.270
FINAL	.577

Functions at Group Centroids

	Function
GROUP	1
4.00	-.968
5.00	.968

Unstandardized canonical discriminant functions evaluated at group means

Classification Results[a]

			Predicted Group Membership		
		GROUP	4.00	5.00	Total
Original	Count	4.00	16	4	20
		5.00	5	15	20
		Ungrouped cases	0	20	20
	%	4.00	80.0	20.0	100.0
		5.00	25.0	75.0	100.0
		Ungrouped cases	.0	100.0	100.0

a. 77.5% of original grouped cases correctly classified.

How to choose the best follow-up ??

If the multivariate results "add nothing" to the bivariate analyses and you won't be doing classifications, then the bivariate follow-ups might be easier for your audience to understand. This approach is also "more specific" in that it uses the original DVs that you selected, rather than constructing new variates out of them.

If you have carefully identified and interpreted the ldfs, then the pairwise ldf analysis "completes" the story by telling giving a statistical description of which groups are different on which ldfs.

If you are interested in how pairs of groups differ from each other (especially if the full model ldf doesn't separate particular groups) the multivariate group comparison approach may give useful information.