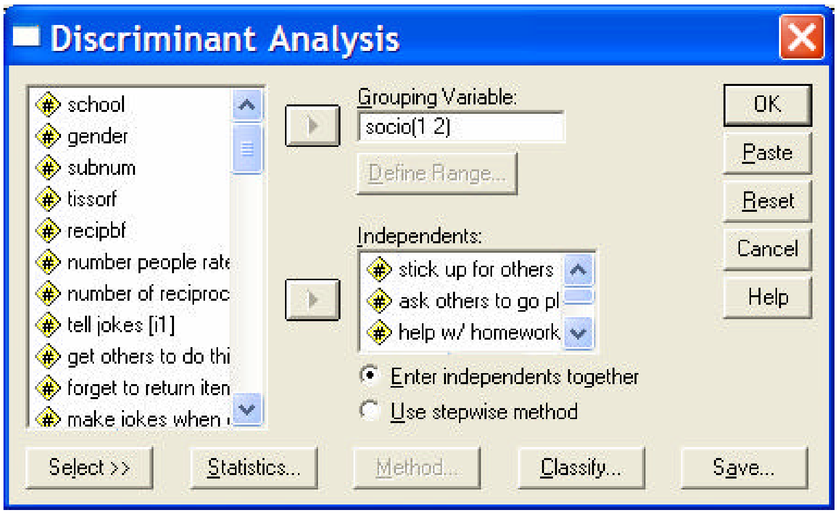
Demonstration of 2-Group Linear Discriminant Function Analysis

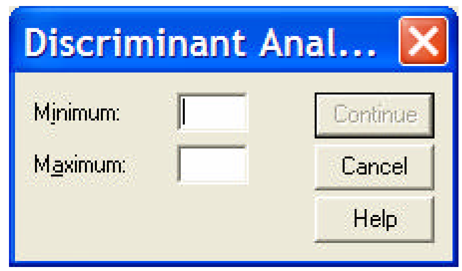
The purpose of the analysis was to identify social behaviors that would discriminate between "accepted" and "rejected' adolescents who were categorized using standard sociometric procedures. Eight behaviors were chosen based on a literature review – four "positive" (first 4) and four "negative" behaviors (last 4).

Analyze → Classify → Discriminant

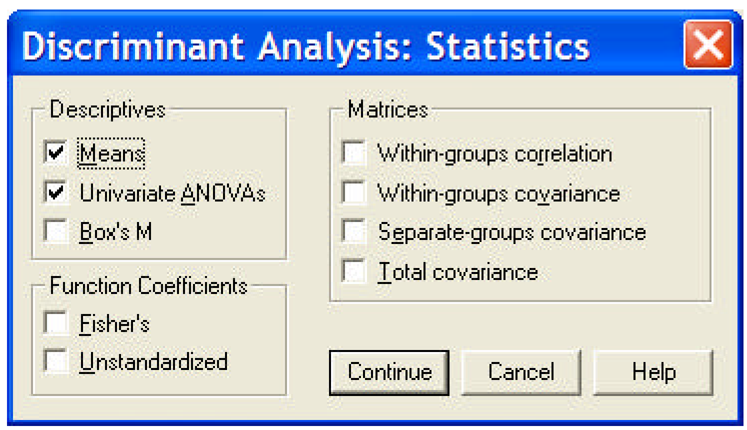


Move the criterion into the "Grouping Variable" window and click the "Define Range" button.

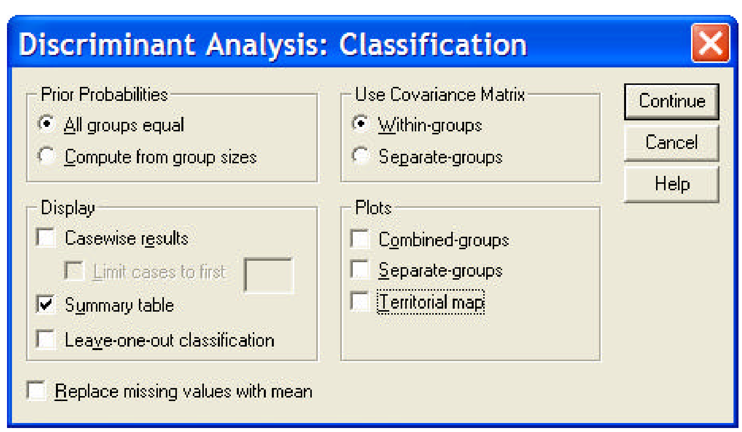
Enter the minimum and maximum code values for the groups (usually 1 & 2).



Click the Statistics button and request univariate stats and F-tests



Click the Classify button and request a summary table of the reclassification results



SPSS Output

Group Statistics

sociometric status		Mean	Std. Deviation	Valid N (listwise) Unweighted	Weighted
popular	stick up for others	4.56	1.105	32	32.000
	ask others to go places	4.31	1.230	32	32.000
	help w/ homework when asked	5.06	1.045	32	32.000
	listen to others problems	5.41	1.160	32	32.000
	push others	1.63	.833	32	32.000
	tell others what to do	3.22	1.431	32	32.000
	ignore conversations	2.28	1.023	32	32.000
	hit others	2.31	1.554	32	32.000
rejected	stick up for others	4.00	1.291	19	19.000
	ask others to go places	4.00	1.333	19	19.000
	help w/ homework when asked	4.05	1.545	19	19.000
	listen to others problems	3.79	1.843	19	19.000
	push others	2.47	1.679	19	19.000
	tell others what to do	2.58	1.216	19	19.000
	ignore conversations	3.32	1.529	19	19.000
	hit others	3.11	1.449	19	19.000
Total	stick up for others	4.35	1.197	51	51.000
	ask others to go places	4.20	1.265	51	51.000
	help w/ homework when asked	4.69	1.334	51	51.000
	listen to others problems	4.80	1.637	51	51.000
	push others	1.94	1.271	51	51.000
	tell others what to do	2.98	1.378	51	51.000
	ignore conversations	2.67	1.322	51	51.000
	hit others	2.61	1.550	51	51.000

Tests of Equality of Group Means

	Wilks' Lambda	F	df1	df2	Sig.
stick up for others	.947	2.723	1	49	.105
ask others to go places	.985	.723	1	49	.399
help w/ homework when asked	.863	7.755	1	49	.008
listen to others problems	.768	14.843	1	49	.000
push others	.894	5.825	1	49	.020
tell others what to do	.949	2.654	1	49	.110
ignore conversations	.854	8.384	1	49	.006
hit others	.938	3.259	1	49	.077

Notice that the popular group had higher means on all of the “positive” behaviors (though only 2 were significantly higher).

However, the popular group had lower means on only 3 of the 4 “negative” behaviors (with only 2 significant differences).

Always take a moment to examine and interpret these bivariate analyses, for two reasons:

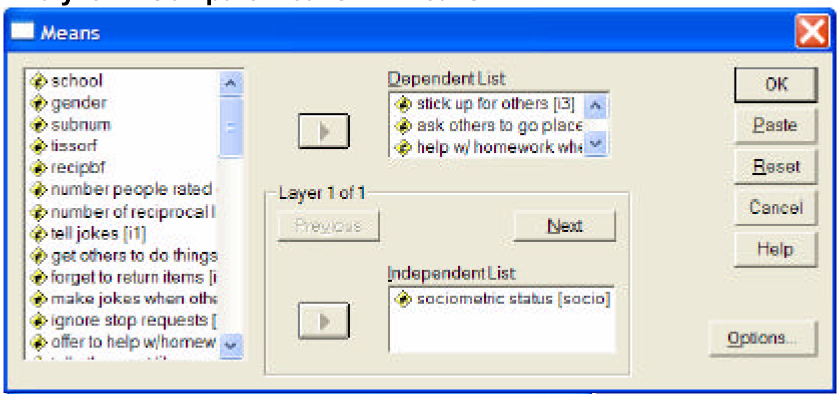
1. You need to know these results to properly interpret the multivariate results – identifying the occurrence of suppressors and other “surprises”

2. At some point you will need to determine whether to present the multivariate results or just the bivariate analyses (depending upon intent, audience, “value” of the multivariate results” etc.)

Remember it is a bad idea to exclude variables from a multivariate analysis because they don’t have bivariate differences – this ignores suppressor effects, multivariate power issues as well as the basic idea that more complicated models are more likely to capture the complexities of behavior!

Often these group means are presented as graph of group profiles, rather than as a table. SPSS will make such a graph, with a bit of persuasion…

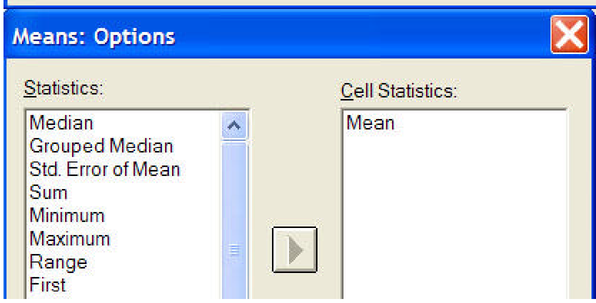
Analyze → Compare Means → Means



Move the predictors variables in to the "Dependent List" window

Move the grouping variable into the "Independent List" window

Click the Options button



You want only "Mean" in the "Cell Statistics:" window.

Including other stats will complicate things later.

Initial output

Report

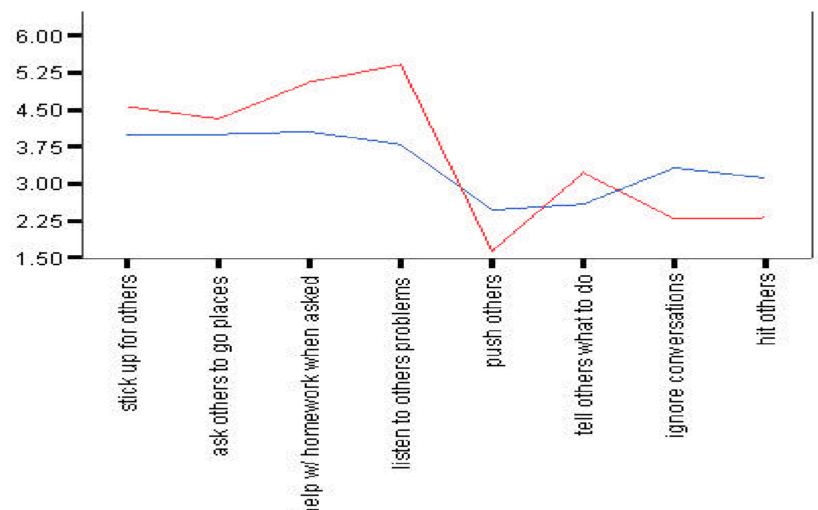
Mean

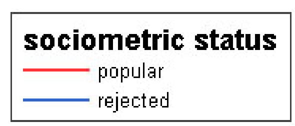
sociometric	stick up or other	k others go places	help w/ homework hen aske	listen to others problems	ush other	ell others hat to d	ignore nversation	it others
popular	4.56	4.31	5.06	5.41	1.63	3.22	2.28	2.31
rejected	4.00	4.00	4.05	3.79	2.47	2.58	3.32	3.11
Total	4.35	4.20	4.69	4.80	1.94	2.98	2.61	2.61

Double-click the table in the SPSS output, putting you into edit mode.

Right-click the table.
Create Graph → Line

For the version below I edited out the Total row before making the graph.





Careful arrangement of the variables on the X axis can make it each for the reader to see the overall pattern of the group differences.

On to the multivariate analyses!

SPSS Output

Eigenvalues

Function	Eigenvalue	% of Variance	Cumulative %	Canonical Correlation
1	.859[a]	100.0	100.0	.680

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

This model accounts for $.68^2$ = 46% of the between group variance

This is one of the statistics used to answer the question, "How well does the model work?"

Wilks' Lambda

Test of Function(s)	Wilks' Lambda	Chi-square	df	Sig.
1	.538	27.902	8	.000

This shows the significance test used to answer the question, "Does the model work?" This model does "work" better than would be expected by chance for this sample size.

Standardized Canonical Discriminant Function Coefficien

	Function
	1
stick up for others	-.271
ask others to go places	-.011
help w/ homework when asked	.524
listen to others problems	.654
push others	-.330
tell others what to do	.247
ignore conversations	-.774
hit others	.264

Structure Matrix

	Function
	1
listen to others problems	.594
ignore conversations	-.446
help w/ homework when asked	.429
push others	-.372
hit others	-.278
stick up for others	.254
tell others what to do	.251
ask others to go places	.131

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

There are two sets of values that are routinely used to interpret ldfs. The ***standardized ldf coefficients*** are just that, the ß weights we use with participant's standardized predictor scores to compute their ldf score. As ß weights, they reflect the unique contribution of each predictor to the ldf.

SPSS does not provide significance tests of these B weights. Effect size-based rules of thumb are usually used, with values in the +/- .3-.4 range indicating a "contributing predictor"

A structure weight is the correlation between scores on that predictor and scores on the ldf. The use of structure weights in ldf analyses comes from factor analysis, and the tradition of "naming a variate in terms of the variables with which it correlates."

Structure weights are correlations and significance tests using df= N-2 are possible, but the usual rule of thumb is to interpret predictors that have structure weights in the +/- .3-.4 range (though some folks/texts will suggest higher or lower values – wouldn't you just figure).

"Interpretation in multiple regression involves: identifying the contributing predictors and 2) noting any suppressor effects. In ldf, however, "interpretation" may involve providing a "name" for the ldf.

Probably the most complete interpretation of an ldf involves starting with the structure weights to tell which predictors are correlated with and therefore tell the attributes related to the ldf. Then augment this description by using the ß weights to tell which predictors have unique contributions

Interpretation of the ldf requires knowing which group is on which "end of the ldf.

A "group centroids" is the mean value of the ldf scores for that group.

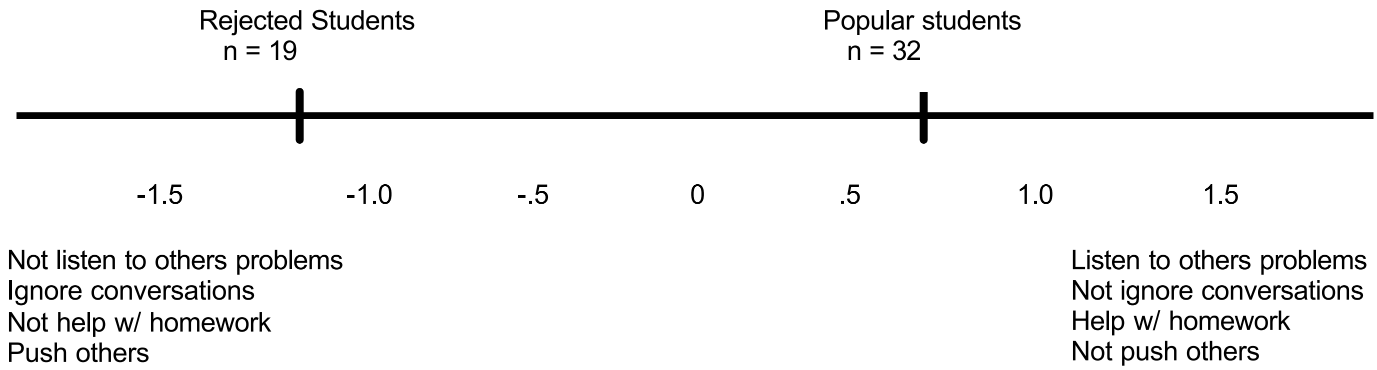
Larger centroids differences reflect better group discriminability. The centroids will be equidistant from the means when there are 2 groups or equal sample size.

Functions at Group Centroids

	Function
sociometric status	1
popular	.700
rejected	-1.179

Unstandardized canonical discriminant functions evaluated at group means

Probably the best way to portray ld results is to use a graphic display that shows the position of each group on the ldf, and then labels the ldf using the information from the structure matrix. Here's an example…



Interpretation is pretty simple in this case, because the variables with structure weights above the cutoff are also those with ß weights above the cutoff. This isn't always the case, because or colinearity among the predictors.

The another thing to examine is the reclassification table. This show how accurately the model can assign participants to their correct groups.

The % correct is the average %correct for each group.

This % correct is probably an overestimate of the classification accuracy of the model, because it is being "tested" using the same sample that was used to construct the model. Cross-validation with subsequent analyses is a good idea, especially if you are planning to use the model for actual classification (as opposed to using it to describe group differences for theoretical reasons).

Classification Results[a]

			Predicted Group Membership		
		sociometric status	popular	rejected	Total
Original	Count	popular	27	5	32
		rejected	4	15	19
		Ungrouped cases	64	66	130
	%	popular	84.4	15.6	100.0
		rejected	21.1	78.9	100.0
		Ungrouped cases	49.2	50.8	100.0

a. 82.4% of original grouped cases correctly classified.

Second Example – With Write-up → Look for "multivariate power"

Participants were given either "success" or "failure" feedback (randomly chosen) on the first 25 of 50 trials, and after completing the next 25 trials, were then asked to give ratings of their ability on those latter 25 trials and the difficulty of the task.

SPSS Code:

```
data list free / subn  grp  ability  diffic.

value labels  grp 1'success feedback'  2 'failure feedback'.

variable labels  ability  'self-report of ability for last 25 trials'
                /diffic   'rating of task difficulty'.
```

SPSS Output:

	Group means		Group Standard Deviations	
GRP	ABILITY	DIFFIC	ABILITY	DIFFIC
Success 1	4.375	5.250	1.84681	1.035
Failure 2	2.750	6.125	1.38873	.640

Wilks' Lambda (U-statistic) and univariate F-ratio
with 1 and 14 degrees of freedom

Variable	Wilks' Lambda	F	Significance	
ABILITY	.77966	3.957	.0666	These results reveal that neither
DIFFIC	.77209	4.133	.0615	variable has a difference between grps.

Canonical Discriminant Functions

Fcn	Eigen	Pct of Variance	Cum Pct	Canonical Corr		After Fcn	Wilks' Lambda	Chisquare	DF	Sig
					:	0	.6278	6.052	2	.0485
1	.5928	100.00	100.00	.6101	:					

What to look for in this part of the printout:
The X^2 is the sphericity test – H0: there are no between group differences from which to build a model
The "Cannonical $Corr^2$" is much like the R^2 you are used to -- between group variance accounted for by the model

So, from this, we would decide that the "proper" combination of these two variables will produce a new variable (ldf) which can differentiate between the groups.

Classification Results -

Actual Group		No. of Cases	Predicted Group Membership 1	2
Group	1	8	5 62.5%	3 37.5%
Group	2	8	2 25.0%	6 75.0%

Percent of "grouped" cases correctly classified: 68.75% (vs. chance of 50%)

Standardized Canonical Disc Coefficients

	FUNC 1
ABILITY	-.70881
DIFFIC	.72363

Unstandardized Canonical Disc Function Coefficients

	FUNC 1
ABILITY	-.4338145
DIFFIC	.8405891
(constant)	-3.235386

Structure Matrix:

Pooled-within-groups Function correlations between discriminating variables and canonical discriminant functions

	FUNC 1
ABILITY	-.69044
DIFFIC	.70563

Remember: When interpreting the canonical coefficients (weights), remember that they are influenced by the collinearity among the variables in the model.
When interpreting the structure weights, remember that they do not tell the unique contribution of Variables – but rather, tell how strongly each variable is correlated with the ldf.

Combining the information from the structure weights and canonical coefficients, we would conclude that both variables are strontly related to the ldf (large structure weights) and that each has a unique contribution to the model (large standardized canonical coefficients). You should also notice that the signs of the variables are opposite each other – more on that below.

Canonical Discriminant Functions evaluated at Group Means (Group Centroids)

Group	FUNC 1
1	-.72023
2	.72023

These are the group means on the ldf (new variable) The overall mean ldf is always zero. The two group centroids will be symmetrical around zero when the two groups have equal N.

Graphic Display of Results & Interpretation

I like to use a graphical display to Help summarize and interpret the in information from the centroids and the structure matrix.

```
          Grp 1                  Grp 2

         (-.72)                  (.72)
----------|------^------------------|-----------------^-----|------  <<-  the ldf
       -1.00                  0                1.00


     higher ability            lower ability
     lower difficulty          higher difficulty
```

Write-up -- the write-up of an LDF analysis is very much like that of a multiple regression

- Describe the variables involved and tell the statistical model employed
- Present the univariate and bivariate summary data (probably in a Table1)
- Report whether or not the model "works" and the related statistical test (present λ and the X^2 with its df & p-value)
- Describe "how well" the model works, using both the R^2-canonical and the % correct re-classification
- Tell which variables contribute to the model (based on the structure weights & a cutoff of +/- .30
- (probably put the weights in a Table 2)
- Present a graphical depiction of the results

Discriminant analyses were used to determine if groups given "false success" and "false failure" feedback during the first 25 trials of a conceptual matching tasks gave differential reports of ability and task difficulty following a second block of 25 trials. Table 1 presents a summary of the univariate and bivariate analyses. Neither of the variables produced significant differences between the groups.

Multivariate analysis revealed a significant difference between the two feedback groups (λ = .628, $X^2(2)$ = 6.05, p =.049), with an R^2-canonical = .372, and 69% correct re-classification (11 out of 16; chance was 50%). Table 2 shows the standardized canonical coefficients and the structure weights, revealing that both of the variables contributed to the multivariate effect.

Figure 1 gives a graphical depiction of the multivariate results. As can be seen, students who had been given “failure” feedback following the first block of 25 trials gave higher difficulty ratings and lower ability ratings than did those students who had been given “success” feedback.

Table 1. Means (standard deviations) and ANOVA results

Variable	"Success" Feedback	"Failure" Feedback	F(p)
Ability rating	4.375 (1.847)	2.750 (1.389)	3.957 (.067)
Difficulty rating	5.250 (1.035)	6.125 (.640)	4.133 (.062)

Table 2. Standardized Canonical Coefficients and Structure weights from the discriminant model

Variable	Standardized Coefficients	Structure Weights
ABILITY	-.709	-.690
DIFFIC	.724	.706

`

Figure 1. Graphical depiction of the discriminant function results.

```
              Success Group                             Failure Group
                 (-.72)                                     (.72)
----------|------^-----------------|----------------^-----|------
        -1.00                      0                        1.00
      Lower Difficulty Ratings                      Higher Difficulty Ratings
      Higher Ability Ratings                        Lower Ability Ratings
```