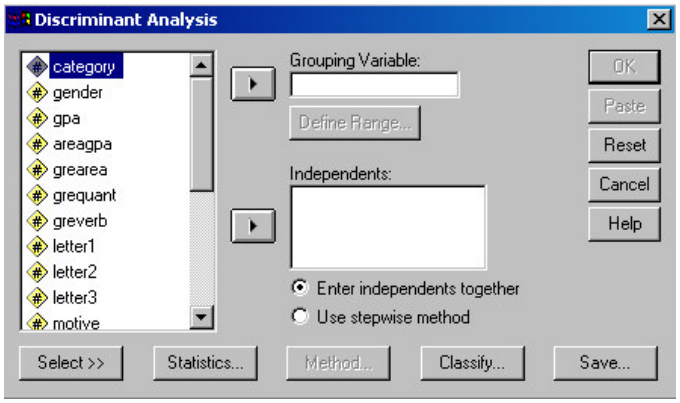
Chapter 6 Discriminant Analyses
SPSS - Discriminant Analyses

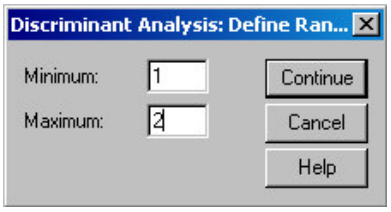
Data file used: ***graduate.sav***
In this example the topic is criteria for acceptance into a graduate program. Every year, selectors miss-guess and select students who are unsuccessful in their efforts to finish the degree. A wealth of information is collected about each applicant prior to acceptance, and department records indicate whether that student was successful in completing the course. This example uses the information prior to acceptance to predict successful completion of a graduate program. The file consists of 50 students admitted into the program between 7 and 11 years ago. The dependent variable is category (1 = finished the Ph.D., 2 = did not finish), and 17 predictor variables are utilized to predict category membership in on of these groups.

How to get there: <u>Analyze → Classify → ….</u>

<u>→ Discriminant …</u>

Discriminant Analysis is used primarily to predict membership in two or more mutually exclusive groups. This menu selection opens the following dialog box:





First enter the grouping variable (here: variable *category*). Then, define the lowest and highest coded value for the grouping variable by clicking on **Button → Define Range**. As the variable category has only two levels you enter 1 and 2 in the boxes. See the second figure standing above.
Then, select the independent variables (choose *gender, motive* and *stable*) in the 'Independents:' box.

There are several methods for discriminant analyses, but here we will only use 'Enter independents together', which is standard selected.

Button → Statistics…
Here you can indicate those statistics that are desired in discriminant analysis. Often these include:
Means: The means and standard deviations for each variable for each group (the two levels of category in this case), and for the entire sample.
Univariate ANOVAs: This compares the mean values for each group for each variable to see if there are significant univariate differences between means.
Box's M: A test for the equality of the group covariance matrices. For sufficiently large samples, a non-significant p value means there is insufficient evidence that the matrices differ. The test is sensitive to departures from multivariate normality.
Unstandardized Function Coefficients: The unstandardized coefficients of the discriminant equation based on the raw scores of discriminating variables.

Button → Classify…
Many classification options can be selected here, such as prior probabilities and plots. Also, a summary table can be requested.

Button → Save…
This option allows you to save as new variables: Predicted group membership, Discriminant Scores and Probabilities of group membership.

Output of running Discriminant Analyses

We performed a discriminant analysis selecting 'Enter independents together'. The descriptives Univariate Anova's Box's M and unstandardized function coefficients are requested. A within-groups covariance matrix is used, and a summary table and a leave-one-out classification are requested (under **Button → Classify…**).

The SPSS output is enormous, so we only indicate some of the relevant information here.

Tests of Equality of Group Means

	Wilks' Lambda	F	df1	df2	Sig.
STUDENTS EMOTIONAL STABILITY	1,000	,007	1	48	,934
1=FEMALE 2=MALE	,938	3,200	1	48	,080
STUDENTS MOTIVATION	,679	22,722	1	48	,000

In the table 'Tests of Equality of Group Means' the results of univariate ANOVA's, carried out for each independent variable, are presented. Here, only student's motivation (variable *motive*) differ (Sig. = ,000) for the two groups (PhD completed and PhD not completed).

Box's Test of Equality of Covariance Matrices

Log Determinants

1=COMPLETED PHD, 2=DID NOT	Rank	Log Determinant
FINISH	3	-,952
NOT FINISH	3	-1,023
Pooled within-groups	3	-,890

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

Test Results

Box's M		4,679
F	Approx.	,727
	df1	6
	df2	16693,132
	Sig.	,628

Tests null hypothesis of equal population covariance matrices.

The significance value of 0,628 indicates that the data *do not differ significantly* from multivariate normal. This means one can proceed with the analysis.

Summary of Canonical Discriminant Functions

Eigenvalues

Function	Eigenvalue	% of Variance	Cumulative %	Canonical Correlation
1	,514[a]	100,0	100,0	,583

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

An eigenvalue indicates the proportion of variance explained. (Between-groups sums of squares divided by within-groups sums of squares). A large eigenvalue is associated with a strong function.

The canonical relation is a correlation between the discriminant scores and the levels of the dependent variable. A high correlation indicates a function that discriminates well. The present correlation of 0.583 is not extremely high (1.00 is perfect).

Wilks' Lambda

Test of Function(s)	Wilks' Lambda	Chi-square	df	Sig.
1	,661	19,286	3	,000

Wilks' Lambda is the ratio of within-groups sums of squares to the total sums of squares. This is the proportion of the total variance in the discriminant scores *not* explained by differences among groups. A lambda of 1.00 occurs when observed group means are equal (all the variance is explained by factors other than difference between those means), while a small lambda occurs when within-groups variability is small compared to the total variability. A small lambda indicates that group means appear to differ. The associated significance value indicate whether the difference is significant. Here, the Lambda of 0,661 has a significant value (Sig. = 0,000); thus, the group means appear to differ.

Canonical Discriminant Function Coefficients

	Function
	1
STUDENTS EMOTIONAL STABILITY	-,001
1=FEMALE 2=MALE	-,595
STUDENTS MOTIVATION	1,169
(Constant)	-8,327

Unstandardized coefficients

The 'Canonical Discriminant Function Coefficients' indicate the unstandardized scores concerning the independent variables. It is the list of coefficients of the unstandardized discriminant equation. Each subject's discriminant score would be computed by entering his or her variable values (raw data) for each of the variables in the equation.

Functions at Group Centroids

1=COMPLETED PHD, 2=DID NOT COMPLETE PHD	Function
	1
FINISH	,702
NOT FINISH	-,702

Unstandardized canonical discriminant functions evaluated at group means

'Functions at Group Centroide' indicates the average discriminant score for subjects in the two groups. More specifically, the discriminant score for each group when the variable means (rather than individual values for each subject) are entered into the discriminant equation. Note that the two scores are equal in absolute value but have opposite signs.

Classification Results[b,c]

		1=COMPLETED PHD, 2=DID NOT COMPLETE PHD	Predicted Group Membership		Total
			FINISH	NOT FINISH	
Original	Count	FINISH	20	5	25
		NOT FINISH	6	19	25
	%	FINISH	80,0	20,0	100,0
		NOT FINISH	24,0	76,0	100,0
Cross-validated[a]	Count	FINISH	20	5	25
		NOT FINISH	7	18	25
	%	FINISH	80,0	20,0	100,0
		NOT FINISH	28,0	72,0	100,0

a. Cross validation is done only for those cases in the analysis. In cross validation, each case is classified by the functions derived from all cases other than that case.

b. 78,0% of original grouped cases correctly classified.

c. 76,0% of cross-validated grouped cases correctly classified.

'Classification Results' is a simple summary of number and percent of subjects classified correctly and incorrectly. The 'leave-one-out classification' is a cross-validation method, of which the results are also presented.