What is Discriminant Function Analysis and how is it used in SPSS Data Analysis?

Authored by

stats writer

June 29, 2024

Discriminant Function Analysis (DFA) is a statistical method used to determine the relationship between a set of predictor variables and a categorical outcome variable. It is commonly used in data analysis to identify the variables that best differentiate between two or more groups. In SPSS, DFA is used to analyze data by creating a discriminant function that maximally separates the groups based on their characteristics. This function can then be used to predict the group membership of new cases based on their values for the predictor variables. DFA is particularly useful for understanding the factors that contribute to group differences and can provide valuable insights for decision-making in various fields such as marketing, healthcare, and social sciences.

Discriminant Function Analysis | SPSS Data Analysis Examples

Version info: Code for this page was tested in IBM SPSS 20.

Linear discriminant function analysis (i.e., discriminant analysis) performs a multivariate test of differences between groups. In addition, discriminant analysis is used to determine the minimum number of dimensions needed to describe these differences. A distinction is sometimes made between descriptive discriminant analysis and predictive discriminant analysis. We will be illustrating predictive discriminant analysis on this page.

Please note: The purpose of this page is to show how to use various data analysis commands. It does not cover all aspects of the research process which researchers are expected to do. In particular, it does not cover data cleaning and checking, verification of assumptions, model diagnostics or potential follow-up analyses.

Examples of discriminant function analysis

Example 1.
A large international air carrier has collected data on employees in three different job classifications: 1) customer service personnel, 2) mechanics and 3) dispatchers. The director of Human Resources wants to know if these three job classifications appeal to different personality types. Each employee is administered a battery of psychological test which include measures of interest in outdoor activity, sociability and conservativeness.

Example 2.

There is Fisher's (1936) classic example of discriminant analysis involving three varieties of iris and four predictor variables (petal width, petal length, sepal width, and sepal length). Fisher not only wanted to determine if the varieties differed significantly on the four continuous variables, but he was also interested in predicting variety classification for unknown individual plants.

Description of the data

Let's pursue Example 1 from above.

We have included the data file, which can be obtained by clicking on discrim.sav. The dataset has 244 observations on four variables. The psychological variables are outdoor interests, social and conservative. The categorical variable is job type with three levels; 1) customer service, 2) mechanic, and 3) dispatcher.

Let's look at the data. It is always a good idea to start with descriptive statistics.

```
get file='d:datadiscrim.sav' .

descriptives variables=outdoor social conservative.
```

Descriptive Statistics

	N	Minimum	Maximum	Mean	Std. Deviation
OUTDOOR	244	0	28	15.64	4.840
SOCIAL	244	7	35	20.68	5.479
CONSERVATIVE	244	0	20	10.59	3.727
Valid N (listwise)	244				

```
means tables=outdoor social conservative by job.
```

Report

JOB		OUTDOOR	SOCIAL	CONSERVATIVE
1	Mean	12.52	24.22	9.02
	N	85	85	85
	Std. Deviation	4.649	4.335	3.143
2	Mean	18.54	21.14	10.14
	N	93	93	93
	Std. Deviation	3.565	4.551	3.242
3	Mean	15.58	15.45	13.24
	N	66	66	66
	Std. Deviation	4.110	3.767	3.692
Total	Mean	15.64	20.68	10.59
	N	244	244	244
	Std. Deviation	4.840	5.479	3.727

correlations

variables=outdoor social conservative.

Correlations

		OUTDOOR	SOCIAL	CONSERVATIVE
OUTDOOR	Pearson Correlation	1	-.071	.079
	Sig. (2-tailed)		.267	.217
	N	244	244	244
SOCIAL	Pearson Correlation	-.071	1	-.236
	Sig. (2-tailed)	.267		.000
	N	244	244	244
CONSERVATIVE	Pearson Correlation	.079	-.236	1
	Sig. (2-tailed)	.217	.000	
	N	244	244	244

frequencie

Statistics

JOB

N	Valid	244
	Missing	0

JOB

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	1	85	34.8	34.8	34.8
	2	93	38.1	38.1	73.0
	3	66	27.0	27.0	100.0
	Total	244	100.0	100.0	

s variables=job.

Analysis methods you might consider

Below is a list of some analysis methods you may have encountered. Some of the methods listed are quite reasonable, while others
have either fallen out of favor or have limitations.

Discriminant function analysis

We will run the discriminant analysis using the discriminant procedure in SPSS.

There is a lot of output so we will comment at various places
along the way.

discriminant

```
/groups=job(1 3)
/variables=outdoor social conservative
/analysis all
/priors equal
/statistics=boxm table
/plot=combined map
/classify=pooled.
```

Analysis Case Processing Summary

Unweighted Cases		N	Percent
Valid		244	100.0
Excluded	Missing or out-of-range group codes	0	.0
	At least one missing discriminating variable	0	.0
	Both missing or out-of-range group codes and at least one missing discriminating variable	0	.0
	Total	0	.0
Total		244	100.0

Group Statistics

JOB		Valid N (listwise)	
		Unweighted	Weighted
1	OUTDOOR	85	85.000
	SOCIAL	85	85.000
	CONSERVATIVE	85	85.000
2	OUTDOOR	93	93.000
	SOCIAL	93	93.000
	CONSERVATIVE	93	93.000
3	OUTDOOR	66	66.000
	SOCIAL	66	66.000
	CONSERVATIVE	66	66.000
Total	OUTDOOR	244	244.000
	SOCIAL	244	244.000
	CONSERVATIVE	244	244.000

Analysis 1

Box's Test of Equality of Covariance Matrices

Log Determinants

JOB	Rank	Log Determinant
1	3	8.162
2	3	7.865
3	3	8.078
Pooled within-groups	3	8.134

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

Test Results

Box's M		26.123
F	Approx.	2.137
	df1	12
	df2	233001.647
	Sig.	.012

Tests null hypothesis of equal population covariance matrices.

Summary of Canonical Discriminant Functions

Eigenvalues

Function	Eigenvalue	% of Variance	Cumulative %	Canonical Correlation
1	1.081[a]	77.1	77.1	.721
2	.321[a]	22.9	100.0	.493

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	.364	242.552	6	.000
2	.757	66.723	2	.000

Standardized Canonical Discriminant Function Coefficients

	Function	
	1	2
OUTDOOR	.379	.926
SOCIAL	-.831	.213
CONSERVATIVE	.517	-.291

Structure Matrix

	Function	
	1	2
SOCIAL	-.765*	.266
CONSERVATIVE	.468*	-.259
OUTDOOR	.323	.937*

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

Functions at Group Centroids

JOB	Function	
	1	2
1	-1.219	-.389
2	.107	.715
3	1.420	-.506

Unstandardized canonical discriminant functions evaluated at group means

Classification Processing Summary

Processed		244
Excluded	Missing or out-of-range group codes	0
	At least one missing discriminating variable	0
Used in Output		244

Prior Probabilities for Groups

JOB	Prior	Cases Used in Analysis	
		Unweighted	Weighted
1	.333	85	85.000
2	.333	93	93.000
3	.333	66	66.000
Total	1.000	244	244.000

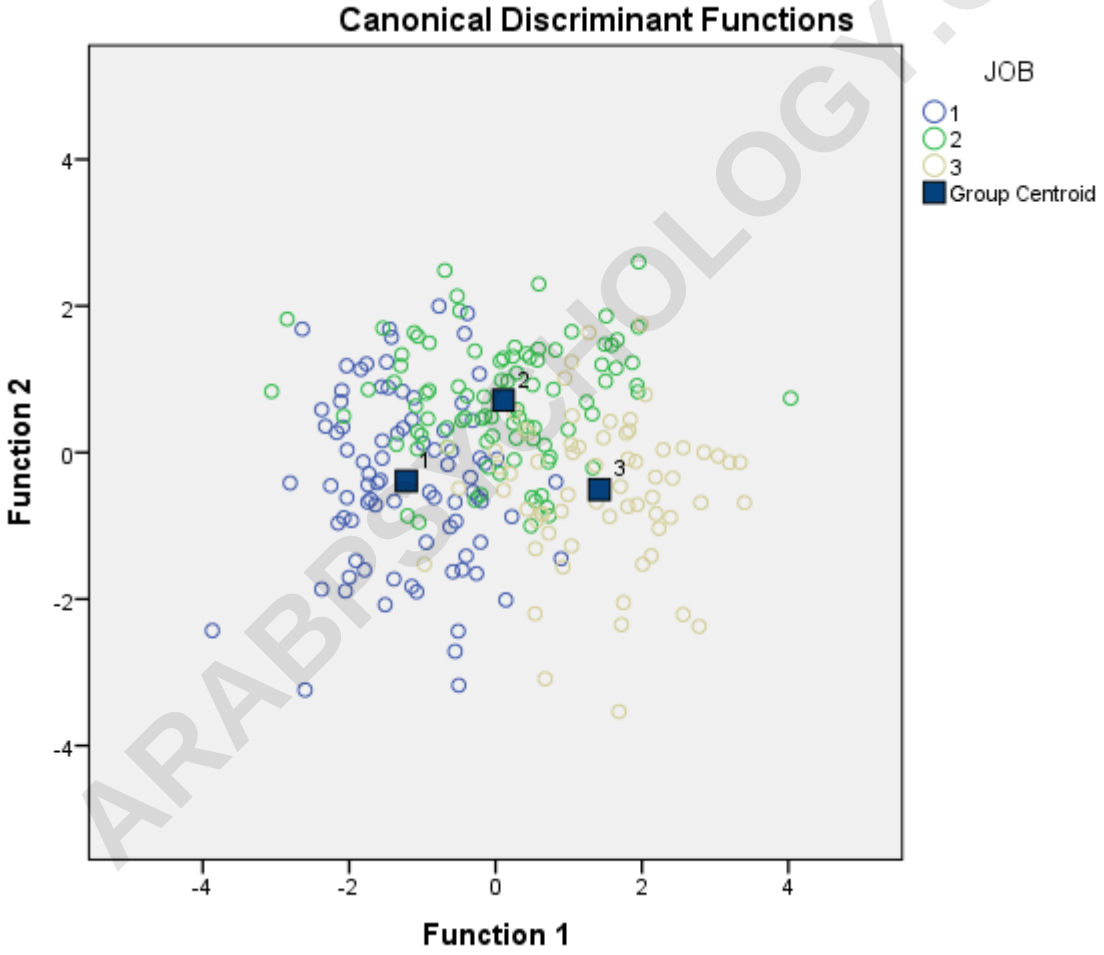
Classification Results[a]

		JOB	Predicted Group Membership			Total
			1	2	3	
Original	Count	1	70	11	4	85
		2	16	62	15	93
		3	3	12	51	66
	%	1	82.4	12.9	4.7	100.0
		2	17.2	66.7	16.1	100.0
		3	4.5	18.2	77.3	100.0

a. 75.0% of original grouped cases correctly classified.

Next, we will plot a graph of individuals on the

discriminant dimensions. Due to the large number of subjects we will shorten the labels for the job groups to make the graph more legible. As long as we don't save the dataset these new labels will not be made permanent.



The discriminant functions are:

discriminant_score_1 = 0.517*conservative + 0.379*outdoor - 0.831*social.

discriminant_score_2 = 0.926*outdoor + 0.213*social - 0.291*conservative.

As you can see, the customer service employees tend to be at the more social (negative) end of dimension 1; the dispatchers tend to be at the opposite end, with the mechanics in the middle. On dimension 2 the results are not as clear; however, the mechanics tend to be higher on the outdoor dimension and customer service employees and dispatchers lower.

SPSS also produces an ASCII territorial map plot which shows the relative location of the boundaries of the different categories. The territorial map is shown below.

Territorial Map
Canonical Discriminant
Function 2
-6.0 -4.0 -2.0 .0 2.0 4.0 6.0
.....

6.0 . 122 .

. 112 2.

. 12 223.

. 122 233 .

. 112 223 .

. 122 233 .

4.0 . 112 . . . . 223 .

. 12 233 .

. 122 223 .

. 112 2233 .

. 12 233 .

. 122 223 .

2.0 . . 112 . . 233 . .

. 122 223 .

. 112 233 .

. 12 223 .

. 122 * 233 .

. 112 223 .

.0 . . . 122. 233 . . .

. * 112 223 .

. 1233 * .

. 13 .

. 13 .

. 13 .

-2.0 . . . 13 . . .

. 13 .

. 13 .

. 13 .

. 13 .

. 13 .

-4.0 . . . 13 . . .

. 13 .

. 13 .

. 13 .

. 13 .

. 13 .

. 13 .

-6.0 . 13 .

.....

-6.0 -4.0 -2.0 .0 2.0 4.0 6.0

Canonical Discriminant Function 1

Symbols used in territorial map

Symbol Group Label

------ ----- --------------------

1 1 customer service

2 2 mechanic

3 3 dispatch

* Indicates a group centroid

Things to consider

See also

References