

Case Study – Classification Problem

Submitted by: Vishal Kumar

Due Date: 19th October, 2015

Date of Submission: 16th October, 2015

Supervisor's Remarks

Late Submission:

Plagiarism:

Completeness:

Quality of Content:

Results and Interpretations:

Additional Remarks:

Overall significance of the discriminant function:

Similar to multiple regression analysis, our first task is to determine whether or not there is a statistically significant relationship between the independent variables and the dependent variable. We navigate to the section of output titled "Summary of Canonical Discriminant Functions" to locate the following outputs:

Eigenvalues

Function	Eigenvalue	% of Variance	Cumulative %	Canonical Correlation
1	2.019 ^a	100.0	100.0	.818

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

Wilks' Lambda

Test of Function(s)	Wilks' Lambda	Chi-square	df	Sig.
1	.331	62.424	3	.000

The key statistic indicating whether or not there is a relationship between the independent and dependent variables is the significance test for Wilks' lambda. Wilks' lambda is the proportion of the total variance in the discriminant scores NOT explained by differences among the groups. In this example about 33% of the variance is not explained by group differences. Unlike R^2 , smaller values of Wilks' lambda are desirable.

You can also note that Wilk's Lambda (overall) is nothing but $1 - \text{square of canonical correlation}$.

Wilk's lambda is used to test the null hypothesis that the means of all of the independent variables are equal across groups of the dependent variable. If the means of the independent variables are equal for all groups, the means will not be a useful basis for predicting the group to which a case belongs, and thus there is no relationship between the independent variables and the dependent variable. If the chi-square statistic corresponding to Wilks' lambda is statistically significant we conclude that there is a relationship between the dependent groups and the independent variables. We should note that there is no correspondence between the size of Wilks' lambda and the accuracy of the classifications based on the discriminant functions.

The information from the table of Eigenvalues is often cited in analyses using discriminant analysis, but it is not as important to us as the statistical test of Wilks' lambda. The table of eigenvalues gives us information about the effectiveness of the discriminant functions. The eigenvalue is a ratio of the between-groups sum of squares to the within-groups or error sum of squares. The size of the eigenvalue is helpful for measuring the spread of the group centroids in the corresponding dimension of the multivariate discriminant space. Larger eigenvalues indicate that the discriminant function is more useful in distinguishing between the groups. The eigenvalues will always be listed in descending order since the solution in a discriminant analysis requires that the first discriminant function is the most capable in differentiating the groups; the second discriminant function is the second most useful function, etc.

The canonical correlation coefficient (.818) measures the association between the discriminant score and the set of independent variables. Like Wilks' lambda, it is an indicator of the strength of relationship between entities in the solution, but it does not have any necessary relationship to the classification accuracy which is our ultimate measure of the value of the model.