MULTIPLE DISCRIMINANT ANALYSIS

Seth Howells
Concepts of Statistics II
Week# 5 Project – Multiple Regression
08/09/20

TABLE OF CONTENTS

OVERVIEW

Multiple discriminant analysis was performed on 4 variables on the HBAT.xls file with x_4 (region) as the nonmetric criterion, or dependent variable, and all other variables listed below as he predictors, or explanatory/independent variables. The tables provide visual understanding of the multivariate relationships.

ID	Variable	Measurement	Description	Type
x_4	Region	nonmetric	Classification	Dependent
x_{11}	Product Line	metric	Performance	Independent
x_{13}	Competitive Pricing	metric	Performance	Independent
x_{17}	Price Flexibility	metric	Performance	Independent

Interpretation of the tables will be provided in separate sections and in the event of unobtained tables due programming choice, interpretation with the absence of its respective table will be provided. List of tables used in the analysis can be found in the table of contents and the corresponding page number. In addition, textbook snippets may be added to analysis so that the SPSS output can be displayed since SAS EG/Studio will not be able to provide exact output.

Wilks' Lambda: examines the difference between multiple groups on one variable. If small, then it indicates that the dependent variable varies significantly across groups, which might mean that the different groups possess different means.

Significance Level: determines whether the independent variable is statistically significant (<0.05) or not (>0.05). If the Wilks' Lambda is small with a significant p-value, then the researcher would reject the null hypothesis.

The discriminant equation: $F = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \ldots + \beta_p X_p + \varepsilon$

F: *latent variable formed by the linear combination of metric dependent variables*
$\boldsymbol{\beta_0}$**:** *y-intercept (constant), the mean*
$\boldsymbol{\beta_0}$: *discriminant coefficients*
$\boldsymbol{X_1}$**:** *p, explanatory/independent variables*
$\boldsymbol{\varepsilon}$: *random error*
together, $b_1 x_1$ *represents the linear effect of* X_1
**please note: exponential variables, such as* x^2*, will create inflection points of curvilinear relationships*

TABLE 5-8:

Stepwise selection was performed on x_{11} (Product Line), x_{13} (Competitive Pricing), and x_{17} (Price Flexibility) as the independent metric variables. This estimation method focuses on independent variables entered one at a time based on its respective discriminating power and would be considered appropriate to use since the three independent variables are the most useful from the original 13 metric variables.

- *Significance:* Product Line proves to be statistically significant based on the 0.000 significance level

FIG 5-8

TABLE 8 Results from Step 3 of Stepwise Two-Group Discriminant Analysis

Overall Model Fit

	Value	*F* Value	Degrees of Freedom	Significance
Wilks' Lambda	.438	23.923	3, 56	.000

Variable Entered/Removed at Step 3

		F		
	Minimum D^2	Value	Significance	Between Groups
X_{11} Product Line	5.045	23.923	.000	0 and 1

Note: At each step, the variable that maximizes the Mahalanobis distance between the two closest group is entered.

Variables in the Analysis After Step 3

Variable	Tolerance	*F* to Remove	D^2	Between Groups
X_{13} Competitive Pricing	.849	7.258	4 015	0 and 1
X_{17} Price Flexibility	.999	18.416	2 822	0 and 1
X_{11} Product Line	.848	5.068	4.300	0 and 1

TABLE 5-9:

This table examines all independent variables instead of just Product Line as in Table 5-8, stepwise selection, and focuses on the *Percent of Variance.* This step displays a moderate correlation. The independent variables are statistically significant as shown (0.000). The eigenvalue of 1.282 displays the sum of squares of total variation.

FIG 5-9

TABLE 9 Summary Statistics for Two-Group Discriminant Analysis

Overall Model Fit: Canonical Discriminant Functions

		Percent of Variance						
Function	**Eigenvalue**	**Function %**	**Cumulative %**	**Canonical Correlation**	**Wilks' Lambda**	**Chi-Square**	***df***	**Significance**
1	1.282	100	100	.749	.438	46.606	3	.000

Discriminant Function and Classification Function Coefficients

	Discriminant Functions		*Classification Funct ons*	
Independent Variables	**Unstandardized**	**Standardized**	**Group 0: USA/North America**	**Group 1: Outside North America**
X_{11} Product Line	–.363	–.417	7.725	6.909
X_{13} Competitive Pricing	.398	.490	6.456	7.349
X_{17} Price Flexibility	.749	.664	4.2 1	5.912
Constant	–3.752		–52 800	–60.623

TABLE 5-10:

Table 5-10 focuses on the predicted groups. The cutting score, calculation to separate the groups based on the number of observations in the groups and their respective centroids of the groups, was determined to be -0.2997 based on the following equation and calculation:

$$Z_{CS} = \frac{N_A Z_B + N_B Z_A}{N_A + N_B}$$, *where*:

Z_{CS}: optimal cutting score between groups A and B
N_A: number of observations in group A
N_B: number of observations in group B
Z_A: centroid for group A
Z_B: centroid for group B

And as a result:

$$Z_{CS} = \frac{(26 * 0.973) + (34 * -1.273)}{26 + 34} = -0.2997$$

The right-hand most column (Total) shows the distribution of counts within the matrix and each region column showing the proportion of that group.

FIG 5-10

TABLE 10 Classification R s lts for Two-Group Discriminant Analysis

Classification Results[a, b,]

Sample	Actual Group	Predicted Group Membership: USA/ North America	Predicted Group Membership: Outside North America	Total
Estimati n Sample	USA/North America	25	1	26
		96.2%	3.8%	
	Outside North America	7	27	34
		20.6%	79.4%	
Cross-validated[d]	USA/North America	24	2	26
		92.3	7.7	
	Outside North America	8	26	34
		23.5	76.5	
Holdout Sample	USA/North America	9	4	13
		69.2	30.8	
	Outside North America	2	25	27
		7.4	92.6	

TABLE 5-11:

Table 5-11 analyzes the predict group with the actual group. Table 5-11 displays the probabilities of predicting the groups with a "*" to indicate a misclassified observation. It is important to note that the probabilities listed in Table 5-11 do not provide as much insight as the textbook's version which replaces the probability with a discriminant z-score. The z-score provides an understanding by how far away from the mean the predict value was. That is, Table 5-11 listed in the *TABLES* (page 8) does not provide information in regard to cost of misclassified observations.

The classification accuracy level exceeded 85% (14 misclassified observations out of 100 observations) which is greater than the calculated proportional chance value of 52.4%. This indicates that the model is statistically significant and will yield a higher percentage of correct predictions when chosen at random.

TABLE 5-13:

Table 5-13 analyzes the discriminant weights in unstandardized and standardized forms. Unstandardized weights (including the constant) calculate the discriminant score, which affects the scale of the independent variable. Recall the Discriminant Equations listed in *OVERVIEW* section to computationally explain how an increase in the coefficients can significantly increase the resulting dependent value. That is, an unstandardized form represents the amount of change in a dependent variable due to a change in 1 unit of the independent variable (for example, X_1).

The two highest ranks are x_{13} (Competitive Pricing) and x_{17} (Price Flexibility). Both have the highest loading value and consequently the highest F-value. However, it is important to refer to the

FIG 5-13

TABLE 13 Summary of Interpretive Measures for Two-Group Discriminant Analysis

Independent Variables	*Discriminant Coefficients*		*Discriminant Loadings*		*Wilks' Lambda*	*Univariate F Ratio*		
	Unstandardized	**Standardized**	**Loading**	**Rank**	**Value**	***F* Value**	**Sig.**	**Rank**
X_6 Product Quality	NI	NI	−.418	5	.801	14.387	.000	4
X_7 E-Commerce Activities	NI	NI	.429	4	.966	2.054	.157	6
X_8 Technical Support	NI	NI	−.136	11	.973	1.598	.211	7
X_9 Complaint Resolution	NI	NI	−.181	8	.986	.849	.361	8
X_{10} Advertising	NI	NI	.238	7	.987	.775	.382	9
X_{11} Product Line	−.363	−.417	−.586	3	.695	25.500	.000	3
X_{12} Salesforce Image	NI	NI	.164	9	.856	9.733	003	5
X_{13} Competitive Pricing	−.398	.490	.656	1	.645	31. 92	.000	1
X_{14} Warranty & Claims	NI	NI	−.329	6	.992	. 53	.503	11
X_{15} New Products	NI	NI	.041	13	.990	.600	.442	10
X_{16} Order & Billing	NI	NI	−.149	10	.999	.087	.769	13
X_{17} Price Flexibility	.749	.664	.653	2	.647	31.699	.000	2
X_{18} Delivery Speed	NI	NI	−.060	12	.997	.152	.698	12

NI = Not included in estimated discriminant function.

TABLE 5-8

Step	Entered	Removed	Label	Partial R-Square	F Value	Pr > F	Wilks' Lambda	Pr < Lambda	Average Squared Canonical Correlation	Pr > ASCC
1	x17		x17	0.3802	60.12	<.0001	0.61976947	<.0001	0.38023053	<.0001
2	x11		x11	0.1872	22.34	<.0001	0.50376233	<.0001	0.49623767	<.0001
3	x13		x13	0.0585	5.96	0.0164	0.47430427	<.0001	0.52569573	<.0001

Total-Sample						
Variable	Label	N	Sum	Mean	Variance	Standard Deviation
x11	x11	100	580.50000	5.80500	1.72997	1.3153
x13	x13	100	697.40000	6.97400	2.38720	1.5451
x17	x17	100	461.00000	4.61000	1.45444	1.2060

x4 = 0						
Variable	Label	N	Sum	Mean	Variance	Standard Deviation
x11	x11	39	261.40000	6.70256	0.74710	0.8643
x13	x13	39	231.00000	5.92308	1.74814	1.3222
x17	x17	39	143.70000	3.68462	0.37239	0.6102

x4 = 1						
Variable	Label	N	Sum	Mean	Variance	Standard Deviation
x11	x11	61	319.10000	5.23115	1.52285	1.2340
x13	x13	61	466.40000	7.64590	1.65486	1.2864
x17	x17	61	317.30000	5.20164	1.25150	1.1187

TABLE 5-9

Canonical Discriminant Analysis

	Canonical Correlation	Adjusted Canonical Correlation	Approximate Standard Error	Squared Canonical Correlation	Eigenvalues of Inv(E)*H = CanRsq/(1-CanRsq)				Test of H0: The canonical correlations in the current row and all that follow are zero				
					Eigenvalue	Difference	Proportion	Cumulative	Likelihood Ratio	Approximate F Value	Num DF	Den DF	Pr > F
1	0.725049	0.720178	0.047669	0.525696	1.1084		1.0000	1.0000	0.47430427	35.47	3	96	<.0001

TABLE 5-10

Number of Observations and Average Posterior Probabilities Classified into x4		
From x4	0	1
0	36 0.8832	3 0.6314
1	11 0.8267	50 0.9261
Total	47 0.8700	53 0.9094
Priors	0.5	0.5

TABLE 5-11

Posterior Probability of Membership in x4					
Obs	From x4	Classified into x4		0	1
1	1	1		0.1240	0.8760
2	0	0		0.9470	0.0530
3	1	0	*	0.9648	0.0352
4	1	1		0.1339	0.8661
5	0	0		0.7989	0.2011
6	1	1		0.0506	0.9494
7	1	1		0.0131	0.9869
8	1	1		0.0070	0.9930
9	1	1		0.0196	0.9804
10	1	1		0.0414	0.9586
11	0	0		0.8632	0.1368
12	1	1		0.0317	0.9683
13	0	0		0.5791	0.4209
14	0	0		0.9914	0.0086
15	1	1		0.0512	0.9488
16	0	0		0.9733	0.0267
17	1	1		0.0675	0.9325
18	1	1		0.0768	0.9232
19	1	1		0.0211	0.9789
20	1	1		0.0189	0.9811
21	1	1		0.0407	0.9593
22	1	0	*	0.9963	0.0037
23	0	0		0.9895	0.0105
24	1	0	*	0.8262	0.1738

Posterior Probability of Membership in x4					
Obs	From x4	Classified into x4		0	1
25	1	1		0.0714	0.9286
26	1	1		0.0053	0.9947
27	0	0		0.9737	0.0263
28	1	1		0.0301	0.9699
29	0	0		0.9737	0.0263
30	1	1		0.0239	0.9761
31	0	0		0.9789	0.0211
32	1	0	*	0.6573	0.3427
33	1	1		0.0911	0.9089
34	1	1		0.0041	0.9959
35	1	1		0.2200	0.7800
36	0	0		0.9541	0.0459
37	0	0		0.7989	0.2011
38	1	0	*	0.7043	0.2957
39	1	1		0.0029	0.9971
40	1	1		0.0317	0.9683
41	1	1		0.0101	0.9899
42	0	1	*	0.3590	0.6410
43	0	0		0.5025	0.4975
44	1	1		0.0636	0.9364
45	0	0		0.9477	0.0523
46	1	1		0.0768	0.9232
47	0	0		0.8632	0.1368
48	1	1		0.0056	0.9944
49	1	0	*	0.9214	0.0786
50	0	0		0.9197	0.0803
51	1	1		0.4460	0.5540
52	0	0		0.9470	0.0530
53	1	0	*	0.7870	0.2130
54	0	0		0.9789	0.0211
55	1	1		0.0041	0.9959
56	0	0		0.8626	0.1374
57	1	1		0.0310	0.9690
58	0	0		0.9406	0.0594
59	0	0		0.8632	0.1368
60	1	1		0.4979	0.5021
61	0	0		0.9733	0.0267

Posterior Probability of Membership in x4					
Obs	From x4	Classified into x4		0	1
62	1	1		0.1256	0.8744
63	0	0		0.6115	0.3885
64	1	0	*	0.7473	0.2527
65	1	1		0.4460	0.5540
66	1	1		0.0056	0.9944
67	1	1		0.0675	0.9325
68	1	1		0.0940	0.9060
69	1	1		0.0265	0.9735
70	1	1		0.0053	0.9947
71	1	1		0.0189	0.9811
72	0	0		0.9821	0.0179
73	1	1		0.0782	0.9218
74	1	0	*	0.9214	0.0786
75	1	1		0.1240	0.8760
76	0	0		0.7755	0.2245
77	1	1		0.0268	0.9732
78	0	1	*	0.3590	0.6410
79	0	0		0.9609	0.0391
80	1	0	*	0.6070	0.3930
81	0	0		0.7755	0.2245
82	0	0		0.8632	0.1368
83	0	0		0.8823	0.1177
84	1	1		0.0602	0.9398
85	0	0		0.9208	0.0792
86	1	1		0.0265	0.9735
87	1	1		0.0477	0.9523
88	0	0		0.9208	0.0792
89	0	1	*	0.3877	0.6123
90	1	1		0.0386	0.9614
91	0	0		0.8899	0.1101
92	1	1		0.0247	0.9753
93	0	0		0.9914	0.0086
94	1	0	*	0.9609	0.0391
95	0	0		0.8626	0.1374
96	0	0		0.7847	0.2153
97	1	1		0.0330	0.9670
98	0	0		0.9541	0.0459

Posterior Probability of Membership in x4					
Obs	From x4	Classified into x4		0	1
99	1	1		0.1265	0.8735
100	1	1		0.0087	0.9913

* Misclassified observation