MULTIPLE LINEAR REGRESSION VIEWPOINTS
VOLUME 17, NUMBER 1, FALL 1989

Use of Linear Models vs. Multiple Discriminant Analysis with Three Groups

Samuel R. Houston
University of Northern Colorado

Clement Marcantonio
University of Southern California

ABSTRACT

It is well known that the three-group oiscriminant function cannot be expressed as a specialized case of the general linear model or multiple linear regression. However, researchers should be alert to the possibility that the set of three-group membership vectors might be adequately represented by an unidimensional bipolar variable of three points thus permitting the use of regression techniques.

A research example is presented in which data were examined both on the basis of the three-group multiple discriminant function as well as by regression procedures. Results of the comparative analysis were such that regression techniques furnished an accurate picture of the finoings. The obvious implication is to suggest that researchers consider using a three-point dependent variable and regression techniques when it makes theoretical or logical sense to conceptualize the three-group membership vectors as a single variable.

In the case of the two-group discriminant function it is well known that the discriminant weights are proportional to the weights for a multiple regression equation of a dichotomous criterion group-membership variable on a set of predictor variables. Thus, discriminant analysis for two groups is a special case of multiple linear regression in which all Group 1 members are assigned the score "1," and all Group 2 members the score "0," on a "dummy" criterion variable Y. Many early writers such as Garrett (1943) and Wherry (1947), as a result of the two-group relationship between the discriminant function and multiple linear regression, stated falsely that discriminant analysis, in general, was nothing more than a special case of multiple linear regression. It should be emphasized, at this point, that the relationship between the discriminant function and multiple linear regression holds only in the case of two groups. When there are more than two groups under investigation, the discriminant function reduces, not to multiple linear regression, but to canonical correlation analysis.

In a doctoral dissertation completed at the University of Northern Colorado, Marcantonio (1977) explored the relationships between selected demographic and personality characteristics as they relate to the variable of the Divorce Initiating Party (I, Both, H/She). The author utilized multiple linear regression techniques as he conceptualized the criterion variable of the Divorce Initiation Party to represent a three-point bipolar vector. In order to make more meaningful the comparison between the three-group discriminant function with a specialized example in which the criterion variable was scored on a three-point scale, it should be helpful to review the main findings of Marcantonio (1977) derived primarily from correlational and regression procedures.

The Ss in the original study by Marcantonio (1977) consisted of 101 formerly-married individuals who had participated in a divorce adjustment seminar presented by a trained psychologist in Colorado during the Fall Quarter, 1977. All the Ss were tested prior to the start of the seminar. The tests included the Tennessee Self Concept Scale, the revised edition of the Fisher Divorce Adjustment Scale (1976), and the demographic questions. Below are presented a list and description of the variables.

Description of the Variables

1. Tennessee Self Criticism Score
2. Tennessee Total Score
3. Tennessee Row 1 Score - Identity
4. Tennessee Row 2 Score - Self Satisfaction
5. Tennessee Row 3 Score - Behavior
6. Tennessee Column A - Physical Self
7. Tennessee Column B - Moral Ethical Self
8. Tennessee Column C - Personal Self
9. Tennessee Column D - Family Self
10. Tennessee Column E - Social Self
11. Tennessee Total Variability
12. Fisher Divorce Adjustment Scale Symptoms-of-Grief Factor

Variable 12 was the Symptoms-of-Grief Factor score obtained on the revised edition of the Fisher Divorce Adjustment Scale. The revised edition utilized items for this factor on the original Fisher Divorce Adjustment Scale (1976). For his study, Marcantonio selected out only those test items or questions which had factor loadings in excess of 0.40 and were of complexity one. This variable measures the extent to which a person mourns the death of the love relationship.

13. Fisher Divorce Adjustment Scale Disentanglement of the Love-Relationship Factor

Variable 13 was the Disentanglement of the Love-Relationship Factor score obtained on the revised edition of the Fisher Divorce Adjustment Scale. This variable measures the extent to which the person dissipates the strong emotional feelings that he/she had for the former love-object person.

14. Fisher Divorce Adjustment Scale Feelings-of-Anger Factor

Variable 14 was the Feelings-of-Anger Factor score obtained on the revised edition of the Fisher Adjustment Scale. This variable measures the anger level of the divorced party.

15. Fisher Divorce Adjustment Scale Rebuilding Social Relationships Factor

Variable 15 was the Rebuilding Social Relationships Factor score obtained on the revised edition of the Fisher Adjustment Scale. This variable measures the extent to which a person has learned to build new friendships and to feel comfortable with friends.

16. Time Separated from One's Spouse

Variable 16 was scored on a four-point scale:

1 identifying a S separated between zero and six months;

2 identifying a person separated between six and 12 months;

3 identifying a person separated one to three years;

4 identifying a person separated more than three years.

17. Age of the Divorced Party

Variable 17 was scored on a five-point scale: 1 identifying a person who is between 20 and 29; 2 identifying a person who is between 30 and 39; 3 identifying a person who is between 40 and 49; 4 identifying a person who is between 50 and 59; 5 identifying a person who is 60 or older.

18. Sex Status of the Divorced Party

Variable 18 was binary coded: 1 identifying a female S; 2 identifying a male S.

19. Divorce Initiating Party

Variable 19 (the criterion variable) was trinary coded: 1 identifying the situation in which S who was tested also initiated the divorce; 2 identifying the situation in which both formerly-married parties initiated the divorce; 3 identifying the situation in which the S who was tested did not initiate the divorce. Since the variable of Divorce Initiating Party was essentially a bipolar concept, it was felt that a three-point numeric scale could be utilized to represent it as a one-dimensional vector.

Multiple Regression Analysis

A list of the variables and their abbreviations are presented in Table 1. In Table 2 are presented the means and standard deviations for the 19 variables studied. The intercorrelation coefficients among the 19 variables are product-moment coefficients and are presented in Table 3. Because one of the variables is binary-coded, some of the coefficients are point-biserial's. In the Marcantonio study there was an attempt to measure both the total or absolute contribution of a predictor variable to the criterion variable as well as the unique contribution of a variable or set of variables to the criterion variable. The total or absolute contribution of a predictor variable is measured by the square of the correlation coefficient between the predictor variable and the criterion variable. The unique contribution of a variable or a set of variables to the criterion variable was determined by methods described in Schmid and Reed (1966). The authors explain that the unique contribution of a predictor variable to the prediction of a criterion

TABLE 1

LIST OF VARIABLES

Number	Variable	Abbreviation
1	TSCS Self Criticism Score	SC-T
2	TSCS Total Score	TOT-T
3	TSCS Row 1 Score - Identity	R1-T
4	TSCS Row 2 Score - Self Satisfaction	R2-T
5	TSCS Row 3 Score - Behavior	R3-T
6	TSCS Column A - Physical Self	CA-T
7	TSCS Column B - Moral Ethical Self	CB-T
8	TSCS Column C - Personal Self	CC-T
9	TSCS Column D - Family Self	CD-T
10	TSCS Column E - Social Self	CE-T
11	TSCS Total Variability	TV-T
12	FDAS Symptoms-of-Grief Factor	SOG-F
13	FDAS Disentanglement of the Love-Relationship Factor	DLR-F
14	FDAS Feelings-of-Anger Factor	FOA-F
15	FDAS Rebuiling Social Relationships Factor	RSR-F
16	Time Separated from One's Spouse	TIME
17	Age of the Divorced Party	AGE
18	Sex Status of the Divorced Party	SEX
19	Divorce Initiating Party (criterion variable)	DIP

TABLE 2

MEANS AND STANDARD DEVIATIONS (N=101)

Variable	Mean	Standard Deviation
1 SC-T	35.56	5.53
2 TOT-T	264.48	12.71
3 R1-T	82.56	5.93
4 R2-T	88.25	7.08
5 R3-T	92.70	5.33
6 CA-T	55.83	4.11
7 CB-T	51.87	4.25
8 CC-T	47.86	5.11
9 CD-T	55.00	5.40
10 CE-T	53.31	3.98
11 TV-T	36.79	10.17
12 SOG-F	49.67	8.91
13 DLR-F	51.06	13.71
14 FOA-F	24.46	6.22
15 RSR-F	23.86	6.34
16 TIME	2.911	1.04
17 AGE	2.17	0.76
18 SEX	1.38	0.48
19 DIP	2.02	0.92

TABLE 3

INTERCORRELATION MATRIX*

Variable		1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18
SC-T	1																		
TOT-T	2	-08																	
R1-T	3	-05	58																
R2-T	4	-03	67	37															
R3-T	5	-05	54	10	20														
CA-T	6	04	14	38	23	12													
CB-T	7	-02	54	47	54	38	17												
CC-T	8	-16	39	35	44	15	-04	23											
CD-T	9	-07	65	45	49	43	-08	41	17										
CE-T	10	04	33	23	28	16	10	07	-00	10									
TV-T	11	01	-06	-16	-18	19	09	05	-28	-01	22								
SOG-F	12	-03	-18	-15	-18	-09	-02	-15	-17	-18	-02	-10							
DLR-F	13	25	-00	06	-02	-06	00	-04	-11	03	08	-16	49						
FOA-F	14	03	12	10	14	05	-00	-00	-02	18	-01	-14	39	42					
RSR-F	15	-16	13	14	20	07	17	12	08	13	-02	-22	17	19	47				
TIME	16	08	07	10	-09	-00	-06	08	-15	18	03	10	13	28	21	-06			
AGE	17	-12	11	19	-02	10	14	-02	17	01	02	-02	-17	-12	-00	01	21		
SEX	18	-06	-12	-18	04	-09	-14	06	-02	-20	-12	-07	-01	-26	07	06	-16	-11	
DIP	19	-20	04	-02	17	-00	-03	12	09	10	-06	04	-18	-46	-10	13	-08	-18	16

variable may be interpreted in a couple of ways. If a predictor variable is making a unique contribution, then knowledge of that variable furnishes information about the criterion. Secondly, if a variableis making a unique contribution, then two Ss, who are different on the variable but who are alike on the other predictor variables, will differ on the criterion. Thus, according to Schmid and Reed, the magnitude of the unique contribution of a set of variables to prediction may be measured by the difference between two squares of multiple correlation coefficients (RSs), one obtained for a linear regression model in which all predictors are used, called the full model (FM), and the other obtained for a linear regression equation in which the proper subset of variables under consideration have been deleted; this model is called the restricted model, (RM). The RS for the RM can never be larger than the RS for the FM. The difference between the two RSs may be tested for statistical significance with the variance-ratio or F test. The formula for this test is as follows:

$$F = \frac{(RS_{FM} - RS_{RM}) / (DF_{FM} - DF_{RM})}{(1 - RS_{FM}) / (N - DF_{FM})}$$

in which N = the size of the sample,

RS_{FM} = the square of the multiple correlation coefficient for the full model,

RS_{RM} = the square of the multiple correlation coefficient for the restricted model,

DF_{FM} = the degrees of freedom associated with the full model, that is, the number of parameters to be estimated in the full model, and

DF_{RM} = the degrees of freedom or number of parameters to be estimated in the restricted model.

For a determination of which variables made a significant total or unique contribution, see Tables 3 and 4. Marcantonio's major findings include:

1) There was found to be a negative significant correlation between the Tennessee Self Criticism (TSC) scores and the Divorce Initiating Party (DIP) scores. This indicates that Ss with a healthy openness and a higher capacity for self-criticism tend to be the initiating party in the divorce procedure.

2) There was found to be a positive significant correlation between the Row 2 TSC scores and the DIP scores. This indicates that Ss with a lower self-satisfaction score tend to be the initiating party in the divorce procedure.

3) There was found to be a negative significant correlation between the Fisher Divorce Adjustment Scale (FDAS) Symptoms-of-Grief Factor scores and the DIP scores. This indicates that Ss with high grief scores tend to be the initiating party in the divorce procedure.

4) There was found to be a negative significant correlation between the FDAS Disentanglement of the Love-relationship Factor scores and the DIP scores. This indicates that SS with high disentanglement scores tend to be the initiating party in the divorce procedure.

5) There was found to be a negative significant correlation between the Age of the Divorced Party scores and the DIP scores. This indicates that Ss who were older tend to be the initiating party in the divorce procedure.

6) The variables of the Age of the Divorced was found to be making a significant unique contribution to the explanation of the DIP scores.

7) The variable of the FDAS Disentanglement of the Love-relationship Factor was found to be making a significant unique contribution to the explanation of the DIP scores.

8) The variable of the FDAS Rebuilding Social Relationship Factor was found to be making a significant unique contribution to the explanation of the DIP scores.

9) The set of 11 TSC Scales was not found to be making a significant unique contribution to the explanation of the DIP scores.

TABLE 4

SUMMARY TABLE OF LINEAR MODELS TESTED

Model	RS Values	Y-Variable	X-Variables	Degrees of Freedom	F	P
1	0.3707	19	1-18	Numerator--18	2.68	0.001*
0	0.0000	19	None	Denominator--82		
Model	RS Values	Y-Variable	X-Variables	Degrees of Freedom	F	P
1	0.3707	19	1-18	Numerator--1	0.05	0.83
2	0.3703	19	1-17	Denominator--82		
Model	RS Values	Y-Variable	X-Variables	Degrees of Freedom	F	P
1	0.3707	19	1-18	Numerator--1	7.22	0.01*
3	0.3152	19	1-16,18	Denmominator--82		
Model	RS Values	Y-Variable	X-Variables	Degrees of Freedom	F	P
1	0.3707	19	1-18	Numerator--1	2.50	0.11
4	0.3514	19	1-15,17-18	Denominator--82		
Model	RS Values	Y-Variable	X-Variables	Degrees of Freedom	F	P
1	0.3707	19	1-18	Numerator--1	4.43	0.04*
5	0.3367	19	1-14,16-18	Denominator--82		
Model	RS Values	Y-Variable	X-Variables	Degrees of Freedom	F	P
1	0.3707	19	1-18	Numerator--1	0.04	0.84
6	0.3704	19	1-13,15-18	Denominator--82		
Model	RS Values	Y-Variable	X-Variables	Degrees of Freedom	F	P
1	0.3707	19	1-18	Numerator--1	21.43	0.001*
7	0.2062	19	1-12,14-18	Denominator--82		
Model	RS Values	Y-Variable	X-Variables	Degrees of Freedom	F	P
1	0.3707	19	1-18	Numerator--1	0.03	0.86
8	0.3705	19	1-11,13-18	Denominator--82		
Model	RS Values	Y-Variable	X-Variables	Degrees of Freedom	F	P
1	0.3707	19	1-18	Numerator--11	0.35	0.97
9	0.3410	19	12-18	Denominator--82		

*Significant beyond 5% level.

Multiple Discriminant Analysis

The problem of studying the direction of group differences is essentially a problem of finding a linear combination of the original set of predictor variables that shows large differences in group means or centroids. Discriminant Analysis is such a method for determining the linear combinations. A very readable and mathematical treatment of discriminant analysis may be found in Tatsuoka (1970). In addition, a mathematical proof that the discriminant analysis and canonical correlational approaches yield identical results was given also by Tatsuoka (1953).

In Table 5 are presented the actual classification results for N = 101, derived by the multiple discriminant function for three groups when the 18 variables are used. From Table 5 it can be seen that 70 (69.31%) of the cases

TABLE 5

DISCRIMINANT CLASSIFICATION RESULTS ON ALL VARIABLES WITH ACTUAL PRIOR PROBABILITIES

Actual Group	Number of Cases	Predicted Group Membership 1	2	3
Group 1	42	34 (83.3%)	1 (2.4%)	6 14.3%)
Group 2	15	6 (40.0%)	6 (40.0%)	3 (20.0%)
Group 3	44	11 (25.0%)	4 (9.1%)	29 (65.9%)

were correctly classified. The numbers along the diagonals represent correct classifications, while the off-diagonal numbers represent misclassifications. The discriminant function was especially accurate for Group 1 (83.3%) and Group 3 (65.9%).

In a forward-selection procedure it was found that Variables 13, 15 and 17 were sufficient variables contributing to group separation in reduced space

with alpha = 0.05. It is interesting to observe in Table 4 that the unique contributions of Variables 13, 15 and 17 were also significant beyond the 0.05 level when multiple linear regression procedures were employed. The classification results based on the use of Variables 13, 15 and 17 in reduced space with the prior probabilities the actual probabilities are presented in Table 6. From an analysis of Table 6 it can be seen that 65 (64.4%) of the cases are now correctly classified. Again, the most accurate predictions are

TABLE 6

DISCRIMINANT CLASSIFICATION RESULTS ON THREE VARIABLES WITH ACTUAL PRIOR PROBABILITIES

Actual Group	Number Of Cases	Predicted Group Membership 1	2	3
Group 1	42	34 (81.0%)	1 (2.4%)	7 (16.7%)
Group 2	15	8 (53.3%)	0 (0.0%)	7 (46.7%)
Group 3	44	11 (25.0%)	2 (4.5%)	31 (70.5%)

associated with Group 1 (81.0%) and with Group 3 (70.5%). In Table 7 are presented the classification results based on the use of Variables 13, 15, and 17 in reduced space with the prior probabilities for each group set at one-third. From the results of Table 7 it can be seen that 62 (61.4%) are now correctly classified. It is interesting to observe that by setting each of the prior probabilities to one-third for each of the three groups, the accuracy associated with predicting membership to Group 2 has increased from 0.0% (Table 6) to 53.3% (Table 7) even though the overall accuracy has slipped from 64.4% (Table 6) to 61.4% (Table 7).

TABLE 7

DISCRIMINANT CLASSIFICATION RESULTS ON THREE VARIABLES WITH EQUAL PRIOR PROBABILITIES

Actual Group	Number Of Cases	Predicted Group Membership 1	2	3
Group 1	42	29 (69.0%)	12 (28.6%)	1 (2.4%)
Group 2	15	3 (20.0%)	8 (53.3%)	4 (26.7%)
Group 3	44	6 (13.6%)	13 (29.5%)	25 (56.8%)

In an attempt to produce a classification via multiple linear regression models alone, a series of binary-coded criterion variables were generated in which "1" designated membership in a particular group and "0" represented membership in one of the other two groups. Using this procedure repeatedly, the researchers produced a classification table which is presented in Table 8. The three variables used as predictors for Table 8 include Variables 13, 15 and 17. From Table 4 it was determined that each of them was making a significant unique contribution beyond the 0.05 level. From Table 8 it can be seen that 69 (68.3%) of the individuals were correctly classified by means of the series of binary-coded multiple linear regression models. The square of the multiple correlation coefficient for the series of binary-coded regression models ranged from 0.266 to 0.366.

TABLE 8

MULTIPLE LINEAR REGRESSION CLASSIFICATION RESULTS ON THREE VARIABLES

Actual Group	Number Of Cases	Predicted Group Membership 1	2	3
Group 1	42	35 (83.3%)	0 (0.0%)	7 (16.7%)
Group 2	15	7 (46.7%)	5 (33.3%)	3 (20.0%)
Group 3	44	14 (31.8%)	1 (2.3%)	29 (65.9%)

Summary Comments

While it is well known that the three-group discriminant function is not a specialized case of multiple linear regression, researchers should consider the possibility that the three groups might form three points on a bipolar continuum. If the set of three-group membership vectors can be captured by a one-dimensional vector, then multiple regression techniques certainly would be appropriate in the analysis of the data. Results from this study furnish an example in which the ability to classify correctly increased from 68.3% to just 69.31% by using the discriminant function instead of multiple linear regression. The slight increase incorrect classification hardly justifies the use of the discriminant function in this case.

REFERENCES

Fisher, B. Identifying and Meeting Needs of Formerly Married People Through a Divorce. Published Doctor of Education Dissertation, University of Northern Colorado, 1976.

Garrett, H. E. The discriminant function and its use in psychology. Psychometrika, 1943, 8, 65-79.

Marcantonio, C. A Profile Analysis of Selected Demographic and Personality Characteristics of Divorced Parties. Published Doctor of Education Dissertation, University of Northern Colorado, 1977.

Schmid, J. and Reed, S. R. Factors in Retention of Residence Hall Freshmen. Journal of Experimental Education, 1966, 35(1), 28-35.

Tatsuoka, M. M. The relationship between canonical correlation and discriminant analysis. Cambridge, Mass.: Educational Research Corporation, 1953.

Tatsuoka, M. M. Discriminant analysis: The study of group differences. Champaign, Ill.: Institute for Personality and Ability Testing, 1970.

Wherry, R. J. Multiple bi-serial and multiple bi-serial correlation. Psychometrika, 1947, 12, 189-195.