MULTIPLE DISCRIMINANT ANALYSIS

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

TABLE OF CONTENTS

MULTIPLE DISCRIMINANT ANALYSIS

OVERVIEW

Multiple discriminant analysis, or MDA, is very similar to multiple regression analysis with a slight difference in the variate's measurement level. While multiple regression analysis requires two or more metric variables that predict one metric dependent variable, multiple discriminant analysis requires two or more metric variables that predict **one nonmetric dependent variable**. This slight difference allows the researcher to perform a multivariate dimensionality reduction technique with categorical data, to which multiple regression cannot include categorical data. Just as linear regression differs from multiple regression, a difference in the number of predictor variables, discriminant analysis differs from multiple discriminant analysis in the same way.

Thus, it is appropriate to use discriminant analysis when the dependent variable is nonmetric (categorical) and only one metric independent (predictor), where multiple discriminant analysis is appropriate with two or more metric independent variables.

EQUATION

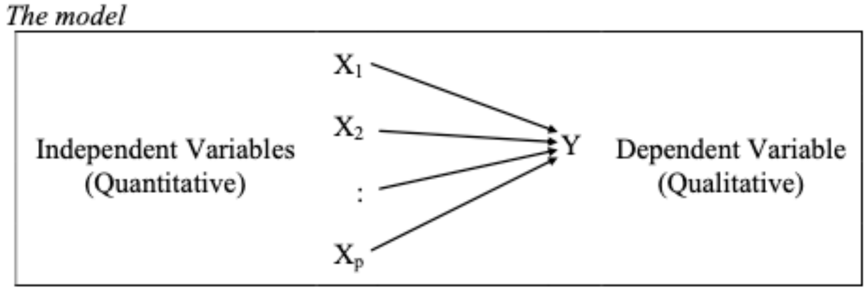
Multiple discriminant analysis derives from the slope formula (*y = mx + b*) in the sense that it is a linear combination comprised of two or more metric independent variables. The *slope line* offers the researcher a way to discriminate, or recognize a distinction, between the groups of categorical data. Basic examples of different groups within the discriminant analysis, *F*, are below. As a result of the linear combination, a series of discriminant scores is obtained for each group.

- Gender – Male vs Female
- Credit Risk – Good credit vs Poor credit
- Affiliation – Member vs Nonmember
- Customer Retention: Reoccurring vs Intermittent

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*
β_0: *y-intercept (constant), the mean*
β_0: *discriminant coefficients*
X_1: *p, explanatory/independent variables*
ε : *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*

FIG 5-1: Discriminant model


The model
Independent Variables (Quantitative)
X_1
X_2
:
X_p
Y
Dependent Variable (Qualitative)

SAMPLE SIZE

In any analysis, sample size should be questioned in efforts to provide the most meaningful output for the variables under in investigation. In multiple discriminant analysis specifically, the objective for the researcher is to determine how many categories within the dependent variable since the effects of the metric independent variables (predictors) result in the dependent variable's output. In this type of multivariate technique, a researcher can implement any number of categories, but with a few precautions:

1. Dependent variable to be mutually exclusive and exhaustive. The groups should be distinct with a unique set of independent variables to maximize the separation of groups. For example, two profiles that are highly similar will not allow the multiple discriminant function to accommodate the separation in groups that is required for analysis.
2. Smaller categories are better for interpretation than larger number of categories because the latter offers complexities in regard to unique groupings. Smaller categories, on the other hand, can still find unique groupings, but mitigates the model from being too complex to the point where analysis is difficult to understand.

As for the sample size, not the number of groups within the dependent variable, there are several general rules of thumb:

- Small sample sizes lead to high sampling error
- Oversampling leads to issues with the ratio of sampling size and the number of independent variables.
- 5 observations is the minimum while 20 observations may be the best fit.

ASSESSING OVERALL FIT

Overall fit can be assessed in different ways, but the discriminant Z-scores (numerical measurement the describes the relationship of the group observations to the mean in efforts to analyze the deviation, provide enough information for inference.

ESTIMATION METHOD

There are two estimation methods for multiple discriminant analysis: simultaneous (direct) method and stepwise method. The researcher's objective is to find the balance between control over the estimation process and maintaining a parsimonious discriminant function. That is, a method with the least assumptions and variables with great explanatory power while possessing control of the estimation process.

Simultaneous (direct): computes all independent variables concurrently.

- When to use: researcher is not interested in results based on the most discriminating variables, but rather to include all variables.
- Significance level analysis: Wilks' lambda, Hotelling's trace, Pillai's criterion

Stepwise: independent variables entered one at a time based on its respective discriminating power.

- When to use: researcher is interested in only useful discriminating variables among a large set of independent variables.
- Significance level analysis: $Mahalanobis\ D^2$, *Rao's V*
    - $Mahalanobis\ D^2$: as the number of predictors increase, this graph becomes more critical since it does not result in any reduction of dimensionality, see FIG 5-2.

CLASSIFICATION MATRIX

A **classification matrix** is comprised of each observation being classified into the dependent variable's group based on the discriminant function. This is done by calculating the **cutting score** (or critical Z value) for each discriminant function, and then placed into the appropriate group. The optimal cutting score formula is listed below:

$$Z_{CS} = \frac{N_A Z_B + N_B Z_A}{N_A + N_B}$$

, *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

The classification matrix is a cross tabulation of observed and predicted value.

FIG 5-2: Mahalanobis

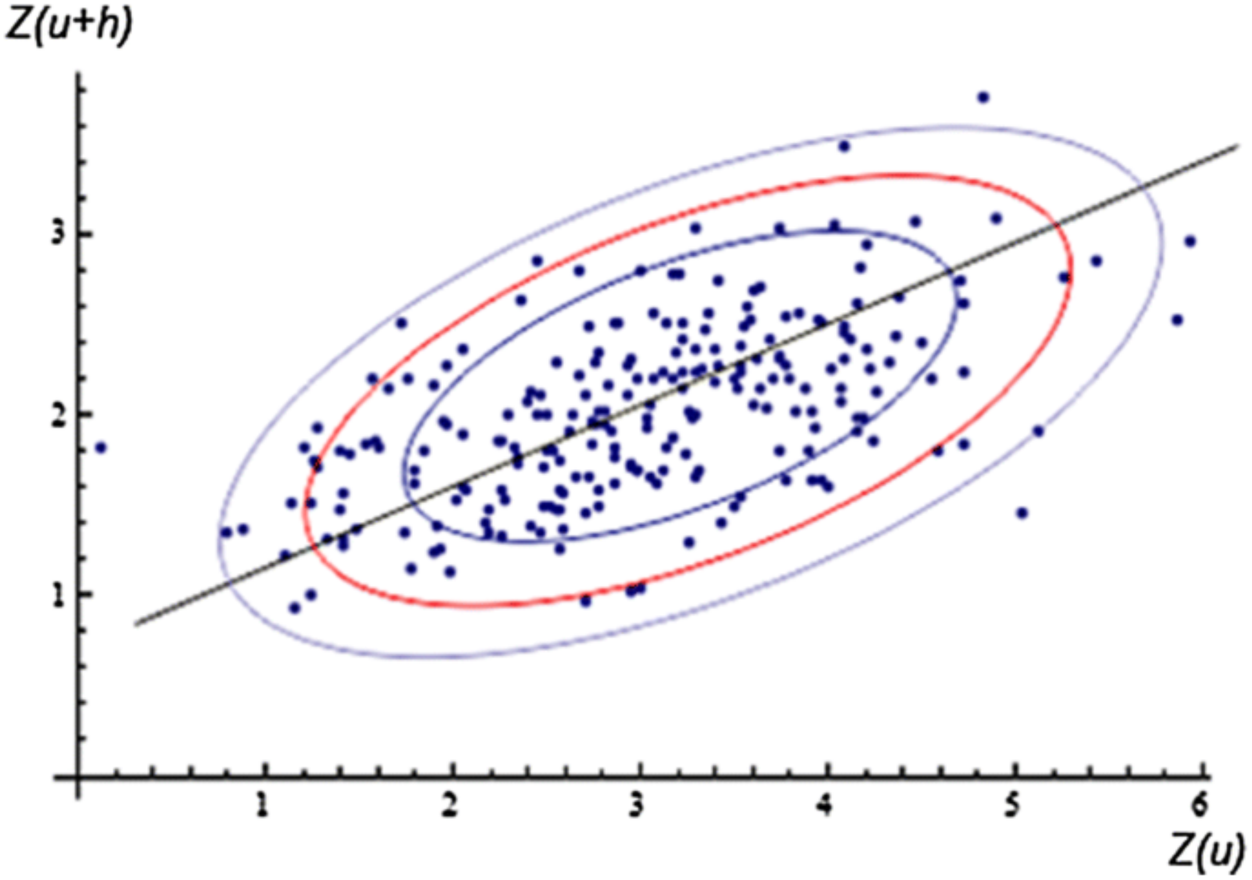


FIG 5-3: Classification Matrix

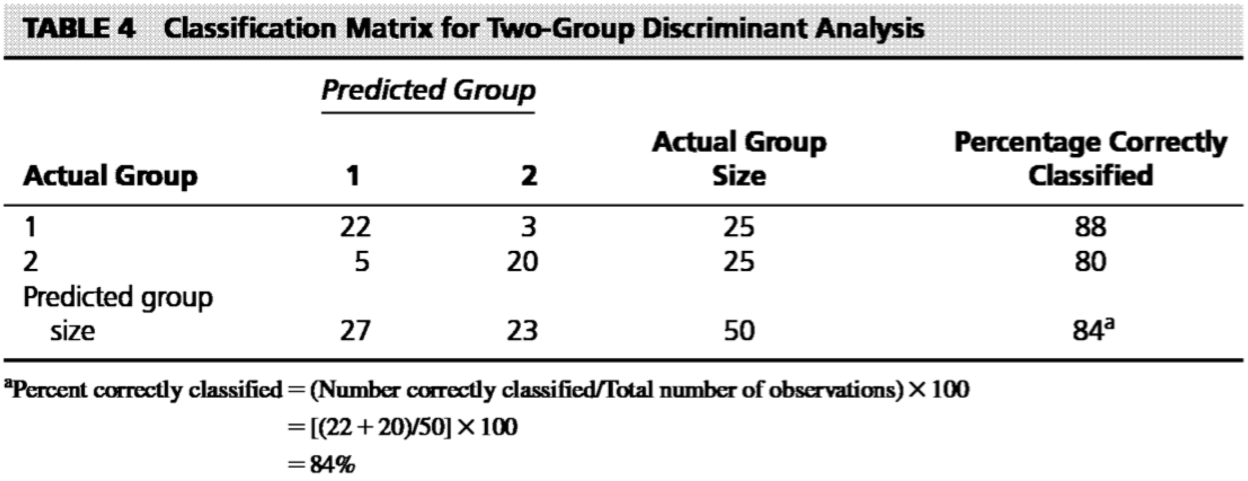


TABLE 4 Classification Matrix for Two-Group Discriminant Analysis

	Predicted Group			
Actual Group	**1**	**2**	**Actual Group Size**	**Percentage Correctly Classified**
1	22	3	25	88
2	5	20	25	80
Predicted group size	27	23	50	84[a]

[a] Percent correctly classified = (Number correctly classified/Total number of observations) × 100
= [(22 + 20)/50] × 100
= 84%