Israel Nolazco

israelnolazco@lewis.edu

Overview of Multiple Discriminant Analysis

August 9, 2020

Table of Contents

# Background.

# Executive Summary.

# Analysis

* 1. Estimation of the Discriminant Model and Assessing Overall Fit
  2. Classification Results for Two-Group Discriminant Analysis

## Summary

1. Background.

The purpose of this project is to perform various techniques of Multiple Discriminant Analysis in order to further understand the relationship between several independent variables compare to single nonmetric dependent variables. During the discriminant analysis we will use 13 independent variables that percept variables from the database (x6 to x18) to discriminate between firms in each geographic area evaluated in the categorical variable in x4

The variables on the  HBAT \_with SPLITS\_ revised.xls and HBAT.xls dataset are presented in the table below:

* •  ‘x4 Region’
* •  ‘x6 Product Quality’
* •  ‘x7 E-commerce’
* •  ‘x8 Technical Support’
* •  ‘x9 Complaint Resolution’,’x10 Advertising’
* •  ‘x11 Product Line’
* •  ‘x12 Salesforce Image’
* •  ‘x13 Competitive Pricing’
* •  ‘x14 Warranty & Claimns’
* •  ‘x15 Packaging’
* •  ‘x16 Order & Billing’
* •  ‘x17 Price Fexibility’
* •  ‘x18 Delivery Speed’

The analysis will then use a discriminant analysis that will create weights for each variable and access those weight to determine where customer’s Region along with determining what variables provide the greatest factor in determining the category. The next section provides a high-level synopsis of details that are presented in the remainder of this document. Additionally, the information was analyze using the programming language SAS.

1. Executive Summary

This document presents an analysis of Multiple Discriminant Analysis that were obtained from the HBAT \_with SPLITS\_ revised.xls and HBAT.xls datasets. Within the aforementioned dataset we going to join both datasets to create a testing sample. It is important to ensure randomness in the selection of the holdout sample so that nay ordering of the observations does not affect the processes of estimation and validation. Additionally, there will be an observation of the level of significance the discrimination analysis creates. This measures how significant the dependent variable Region is truly a categorical variable and if actually plays a role in the overall analysis.

1. Analysis
   1. Estimation of the Discriminant Model and Assessing Overall Fit

**A screenshot of a social media post

Description automatically generated**

**A screenshot of a cell phone

Description automatically generated**

**A screenshot of a cell phone

Description automatically generated**

Figure 1. Two-group Discriminant Analysis.

It is important to point out that at this point SAS does not provide a stepwise discriminant analysis option. Therefore, the variables being analyze are only columns X11 Product Line, x13 Competitive Pricing and x17 Price Flexibility. Figure 1 reviews the results of the Two-group Discriminant Analysis proves the statistical significance as evidence by the low Wilks’ lambda value 0.4. Thus, the review of those three columns prove sufficient for this discriminant function. This is further reinforcing by looking at the Canonical Correlation which accounts for the variance in the dependent variable X4 and that of the three variables. 77% result thus give a strong sense of the model.

A screenshot of a cell phone

Description automatically generated

Figure 2. Canonical Discriminant Analysis

The group centroids are also reported, and they represent the mean of the individual discriminant functions score for each group. Group centroids provide a summary measure of the relative position of each group on the discriminant functions(s) which can be found on Figure 2.

A close up of a piece of paper

Description automatically generated

Figure 3. Classification weight for each observation

* 1. Classification Results for Two-Group Discriminant Analysis

As seen in Figure 3 each observation carries a weight for either category (Inside USA/Outside USA). Depending on the weight the column carries a decision will determine where the observation falls under. However, in order to ascertain accuracy and confidence in results a cross-validation is implemented and errors are accounted for.

A screenshot of a cell phone

Description automatically generated

A screenshot of a cell phone

Description automatically generated

Figure 4. Group Prediction for individual cases/classification results

The classification matrices for the observation in both the analysis and holdout samples were calculated and the results shown in figure 3 and 4 contain discriminant score for each observation as well the actual predicted groups. Additionally, the cross validated sample achieved a prediction of accuracy of 93%. However, there were classification errors. Make notice of observation 3. That observation was classified mistakenly.

1. Summary

Multiple Discriminant Analysis seem fairly accurate in classifying and developing models in which categories are needed. However, making assumptions in the data could lead to skewness and misrepresentation. Unlike other Multivariate Analysis, discriminant analysis is binary at its core and misclassification will happen. However, the researcher must be weary to modify the model or further interpret those misclassifications since they could potentially create further complexities.