Israel Nolazco

israelnolazco@lewis.edu

Overview of Multiple Regression Analysis

August 2, 2020

Table of Contents

# Background.

# Executive Summary.

# Analysis

# Estimating the Regression Model and Assessing Overall Model Fit

## Evaluating the Variate for the Assumptions of Regression Analysis

## Summary

1. Background.

The purpose of this project is to perform various techniques of Multiple Regression Analysis in order to further understand the relationship between several independent variables compare to single dependent variables. During the regressing analysis, there will be also a discovery of factors influencing variables the most; which then can be further manipulate for a desired outcome.

The variables on the HBAT(2).xls dataset are presented in the table below:

* •  ‘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 variables listed along with several visualizations were created could be used to visually identify variables affecting an independent variable with multiple dependent variables. In a broader scope this could potentially allow for a better manipulation of data for forecasting purposes. 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 Regression Analysis that were obtained from the HBAT(2).xls dataset. Within the aforementioned dataset we are primarily looking at vendor information regarding a product(s). Based on the metrics gather on this dataset, a bigger picture will be painted in order to understand consumer behavior and therefore, predict any future behavior. Additionally, we could move to create educated suggestions to improve Product Quality, Competitive Pricing and features customers focus on whilst shopping. Therefore, driving force of this analysis falls under the p-value for the effect of the classification variable on the response. This measures how large the t statistic is for this analysis. The smaller the value, the more likely the model is fits in identifying the dependent variables.

1. Analysis
   1. Estimating the Regression Model and Assessing Overall Model Fit

A screenshot of a cell phone

Description automatically generated

Figure 1. Stepwise estimation.

As seen in figure 1 the variables from our data set were compile and compare to five independent variables (‘x6 Product Quality’ , ‘x7 E-commerce’ ,  ‘x9 Complaint Resolution’,’x10 Advertising’,  ‘x11 Product Line’ ,  ‘x12 Salesforce Image’) to a single dependent variable ‘x19 Customer Satisfaction.’ In this scenario our R-Square is that of .79 or roughly 80%. This value, therefore, shows the regression model we have calculated explains almost 80% of the variance of customer satisfaction. Additionally, looking at our p-value we noticed that is less than 0.5% therefore, the null hypothesis can be rejected, and we can rest assure the model shows statistical significance between the dependent and independent variables.

A picture containing large

Description automatically generated

Figure 2. Analysis of Standardized Residuals

* 1. Evaluating the Variate for the Assumptions of Regression Analysis

There are several assumptions to analyze, Linearity, Homoscedasticity and Normality. Linearity is asses through an Analysis of residuals and partial regression plots. As seen on Figure 2 there are no linear patters to the residuals thus ensuring the overall equation is linear. However, there must be an analysis of partial regression for each independent variable in the equation, which can be seen in Figure 3. However, moving on to Homoscedasticity, again figure 2 shows no pattern of increasing or decreasing residuals and this indicates Homoscedasticity in the multivariate. Finally, Normality; this assumption will check normality of the error term of the variate with a visual examination of the normal probability plots of residuals which can be seen in Figure 4. One key detail to take away from the image is the values following a linear path; therefore representing a normal distribution and the variate is found to meet the assumption of normality.

A picture containing dog, large, kitchen, light

Description automatically generated

Figure 3. Standardized Partial Regression Plots

A close up of a map

Description automatically generated

Figure 4. Normal Probability Plot: Standardized Residuals.

1. Summary

Multiple Regression Analysis is a wonderful tool to understand relationship between independent and dependent variables. These relationships can then be used a forecasting tool to better amplify the desired results. In the analysis above, there was a strong statistical relationship of Customer Satisfaction (x19) and being affected by Product Quality, Sales Force Image, E-commerce, Product Line and Compliant Resolution. These five independent variables can then be modified and alter to improve Customer Satisfaction. Lastly, SAS as a statistical tool presents and manipulates data with ease. SAS is a great tool to use for those not familiar with code based tools such as R and Python.