Vivian Ellis Logistic Regression Multivariate Data Analysis

The purpose of logistic regression is to determine whether differences in the perceptions of HBAT exists between customer region. The customer region is determined by whether a customer resides in North America and outside of North America. I will use stepwise logistic regression to control for multicollinearity. The three variables that have a variance inflation factor > 10 are Prod\_line, Price\_flex, and Del\_speed with a VIF of 37.97842, 33.33234, and 44.00376 respectively. The preliminary analysis will provide an insight as to which metric independent variables might help in predicting the dependent variable, region. Using the F test,

To determine if the variances are equal. F(60,38)= 2.85, with a p-value=0.0008 and the null hypothesis is accepted. The table below gives a summary of the T tests on each of the 13 metric independent variables.

|  |  |  |  |
| --- | --- | --- | --- |
| Variable | T Value | Df | p-value |
| ***Product Quality*** | ***5.95*** | ***98*** | ***<.0001*** |
| Ecommerce | -1.89 | 84.496 | 0.0618 |
| Tech Support | 1.80 | 87.403 | 0.0757 |
| Complaint | .10 | 87.881 | 0.9227 |
| ***Advertising*** | ***-2.01*** | ***88.389*** | ***0.0476*** |
| ***Product Line*** | ***6.49*** | ***98*** | ***<.0001*** |
| ***Sales Image*** | ***-4.48*** | ***89.158*** | ***<.0001*** |
| ***Pricing*** | ***-6.42*** | ***79.479*** | ***<.0001*** |
| Warranty | 1.53 | 87.386 | 0.1305 |
| New Product | -1.10 | 83.93 | 0.2726 |
| Ordering | -.38 | 98 | 0.7027 |
| ***Price Flex*** | ***-7.75*** | ***98*** | ***<.0001*** |
| Delivery Speed | -.19 | 95.131 | 0.8474 |

There are six variables, product quality, advertising, product line, sales image, pricing, and price flex that have significantly different means from region at alpha=.05. For each of these six variables the mean for region group 1 is higher than the mean for region group 0. Respondents who reside outside of North America do not want to partner with HBAT have lower average perceptions of HBAT on those six variables when compared to respondents who reside in North America. These six variables are potential predictor or discriminating variables to predict or group the response variable region.

Using the overall model test let,

Ho: None of the independent variables help to explain region

Ha: at least one independent variable helps

Then using the likelihood ratio test, and p-value<.0001 the null hypothesis is rejected and at least one independent variable helps to explain region. The Hosmer and Lemeshow Goodness-of-Fit test is used to assess how well the logistic model fits the data. Let,

Ho: The model fits the data

Ha: The model does not fit the data

and p-value=0.7319 we fail to reject the null hypothesis. Thus, the model fits the data. The following table gives the logistic regression model and the individual parameter test using the Wald Chi-square test. All variables are highly significant.

| **Analysis of Maximum Likelihood Estimates** | | | | | |
| --- | --- | --- | --- | --- | --- |
| **Parameter** | **DF** | **Estimate** | **Standard Error** | **Wald Chi-Square** | **Pr > ChiSq** |
| **Intercept** | 1 | 3.4071 | 3.6412 | 0.8755 | 0.3494 |
| **Ecommerce** | 1 | 4.0558 | 1.2920 | 9.8547 | 0.0017 |
| **Adv** | 1 | 1.2507 | 0.5647 | 4.9063 | 0.0268 |
| **Prod\_Line** | 1 | 2.5162 | 0.7859 | 10.2496 | 0.0014 |
| **Sales\_Image** | 1 | -5.5335 | 1.7723 | 9.7479 | 0.0018 |
| **Price\_Flex** | 1 | -2.5699 | 0.8453 | 9.2435 | 0.0024 |

The following table shows the Hosmer and Lemeshow Test is used to test the overall fit of the model. Let,

Ho: The model fits the data

Ha: The model does not fit the data

χ^2 (8)=6.7285 and p-value=0.5662 we fail to reject the null hypothesis. Thus, the model fits the data.

| **Hosmer and Lemeshow Goodness-of-Fit Test** | | |
| --- | --- | --- |
| **Chi-Square** | **DF** | **Pr > ChiSq** |
| 6.7285 | 8 | 0.5662 |