Vivian Ellis

DSCI 449

Multiple Regression Part 2

We will use the same model from part 1,

Where we want to determine if the non-metric variables, firm size and customer type help to explain recommend. Firm size is already encoded as a binary variable, 0=small firm, fewer than 500 employees and 1=large firm, 500 or more employees. Currently customer type is encoded as following,

1=less than a year

2= between 1 and 5 years

3=longer than 5 years

Customer type will have to be recorded into two dummy variables,

If customer type is 1 then cust1=1 else cust1=0

If customer type is 2 then cust2=1 else cust2=0

If customer type is 3 the cust1=cust2=0

All variables significant at alpha=.10 except for ecommerce, cust1, and cust2. All variance inflation is less than 10.

**Block MR2-2**

|  |  |
| --- | --- |
| **Number of Observations Read** | 100 |
| **Number of Observations Used** | 100 |

| **Analysis of Variance** | | | | | |
| --- | --- | --- | --- | --- | --- |
| **Source** | **DF** | **Sum of Squares** | **Mean Square** | **F Value** | **Pr > F** |
| **Model** | 8 | 63.62373 | 7.95297 | 16.40 | <.0001 |
| **Error** | 91 | 44.13627 | 0.48501 |  |  |
| **Corrected Total** | 99 | 107.76000 |  |  |  |

|  |  |  |  |
| --- | --- | --- | --- |
| **Root MSE** | 0.69643 | **R-Square** | 0.5904 |
| **Dependent Mean** | 7.02000 | **Adj R-Sq** | 0.5544 |
| **Coeff Var** | 9.92065 |  |  |

| **Parameter Estimates** | | | | | | |
| --- | --- | --- | --- | --- | --- | --- |
| **Variable** | **DF** | **Parameter Estimate** | **Standard Error** | **t Value** | **Pr > |t|** | **Variance Inflation** |
| **Intercept** | **1** | 1.63032 | 0.95557 | 1.71 | 0.0914 | 0 |
| **Prod\_Qual** | **1** | 0.27336 | 0.06852 | 3.99 | 0.0001 | 1.86835 |
| **Ecommerce** | **1** | -0.06205 | 0.17853 | -0.35 | 0.7290 | 3.19256 |
| **Prod\_Line** | **1** | 0.14876 | 0.07727 | 1.93 | 0.0573 | 2.10809 |
| **Sales\_Image** | **1** | 0.35044 | 0.12307 | 2.85 | 0.0054 | 3.55512 |
| **Ordering** | **1** | 0.16197 | 0.09518 | 1.70 | 0.0922 | 1.59546 |
| **cust1** | **1** | -0.35850 | 0.28324 | -1.27 | 0.2089 | 3.59934 |
| **cust2** | **1** | 0.07252 | 0.22830 | 0.32 | 0.7515 | 2.44487 |
| **Firm\_size** | **1** | 0.43174 | 0.15581 | 2.77 | 0.0068 | 1.25085 |

To test significance of cust1 and cust2 we will remove them from our model and run the model again, seen in Block MR2-3

**Block MR2-3**

|  |  |
| --- | --- |
| **Number of Observations Read** | 100 |
| **Number of Observations Used** | 100 |

| **Analysis of Variance** | | | | | |
| --- | --- | --- | --- | --- | --- |
| **Source** | **DF** | **Sum of Squares** | **Mean Square** | **F Value** | **Pr > F** |
| **Model** | 6 | 61.56998 | 10.26166 | 20.66 | <.0001 |
| **Error** | 93 | 46.19002 | 0.49667 |  |  |
| **Corrected Total** | 99 | 107.76000 |  |  |  |

|  |  |  |  |
| --- | --- | --- | --- |
| **Root MSE** | 0.70475 | **R-Square** | 0.5714 |
| **Dependent Mean** | 7.02000 | **Adj R-Sq** | 0.5437 |
| **Coeff Var** | 10.03912 |  |  |

| **Parameter Estimates** | | | | | | |
| --- | --- | --- | --- | --- | --- | --- |
| **Variable** | **DF** | **Parameter Estimate** | **Standard Error** | **t Value** | **Pr > |t|** | **Variance Inflation** |
| **Intercept** | **1** | 0.83871 | 0.63636 | 1.32 | 0.1908 | 0 |
| **Prod\_Qual** | **1** | 0.29118 | 0.05860 | 4.97 | <.0001 | 1.33460 |
| **Ecommerce** | **1** | -0.14869 | 0.17319 | -0.86 | 0.3928 | 2.93408 |
| **Prod\_Line** | **1** | 0.18017 | 0.06951 | 2.59 | 0.0111 | 1.66599 |
| **Sales\_Image** | **1** | 0.41829 | 0.11840 | 3.53 | 0.0006 | 3.21288 |
| **Ordering** | **1** | 0.24194 | 0.08793 | 2.75 | 0.0071 | 1.32947 |
| **Firm\_size** | **1** | 0.44985 | 0.15741 | 2.86 | 0.0053 | 1.24674 |

The test will be defined as,

H\_0 :

H\_a : at least one

With the sum of squares error in the reduced model, the sum of square error in the full model, the degrees of freedom for the full model, the degrees of freedom for the reduced model and the mean square error of the full model. Thus,

The p-value is calculated using a numerator df=2 and denominator df=2. A calculator gives p=.126259. The result is not significant thus we accept the null hypothesis because cust1 and cust2 do not help explain recommend. I would suggest removing ecommerce from the model since it proves to not be significant in the full and reduced model. Block MR2-4 gives the final model,

Where all variance inflation is less than 10 and all variables are significant. I would not suggest removing any more variables. This model appears to be “better” than the model produces in part 1, in part 1 the adjusted and the final model has a higher adjusted. In the final model, a larger firm size, 500 or more employees, are likely to recommend HBAT to others firms by.48962 units than smaller firms, holding all other variables constant.

**Block MR2-4**

|  |  |
| --- | --- |
| **Number of Observations Read** | 100 |
| **Number of Observations Used** | 100 |

| **Analysis of Variance** | | | | | |
| --- | --- | --- | --- | --- | --- |
| **Source** | **DF** | **Sum of Squares** | **Mean Square** | **F Value** | **Pr > F** |
| **Model** | 5 | 61.20388 | 12.24078 | 24.71 | <.0001 |
| **Error** | 94 | 46.55612 | 0.49528 |  |  |
| **Corrected Total** | 99 | 107.76000 |  |  |  |

|  |  |  |  |
| --- | --- | --- | --- |
| **Root MSE** | 0.70376 | **R-Square** | 0.5680 |
| **Dependent Mean** | 7.02000 | **Adj R-Sq** | 0.5450 |
| **Coeff Var** | 10.02507 |  |  |

| **Parameter Estimates** | | | | | | |
| --- | --- | --- | --- | --- | --- | --- |
| **Variable** | **DF** | **Parameter Estimate** | **Standard Error** | **t Value** | **Pr > |t|** | **Variance Inflation** |
| **Intercept** | **1** | 0.67597 | 0.60663 | 1.11 | 0.2680 | 0 |
| **Prod\_Qual** | **1** | 0.29213 | 0.05851 | 4.99 | <.0001 | 1.33411 |
| **Prod\_Line** | **1** | 0.18339 | 0.06931 | 2.65 | 0.0095 | 1.66114 |
| **Sales\_Image** | **1** | 0.33696 | 0.07092 | 4.75 | <.0001 | 1.15603 |
| **Ordering** | **1** | 0.23889 | 0.08773 | 2.72 | 0.0077 | 1.32729 |
| **Firm\_size** | **1** | 0.48962 | 0.15023 | 3.26 | 0.0016 | 1.13878 |