MS -Business Intelligence & Analytics

Fall 2015

BIA – 652 C

November 4, 2015

Mohit Ravi Ghatikar

CWID - 10405877

Multivariate Data Analytics – HW5

Ethics Statement

I pledge on my honor that I have not given or received any unauthorized assistance on this assignment /examination. I further pledge that I have not copied any material from a book, article, the Internet or any other source except where I have expressly cited the source.

Signature _Mohit Ravi Ghatikar________________ Date: 11/04/2015___

Question:

Using Churn dataset, create a categorical variable CustServ as:

CustServ_Calls < 2 then V_CSC=0;

else if CustServ_Calls < 4 then V_CSC=1;

else V_CSC=2

Perform logistic regression analysis for churn on V_csc.

Solution:

We run a logistic regression for churn on V_CSC. By using the class option, we can split the class variable V_CSC into three variables. They are:

Class Level Information

Class	**Value**	**Design Variables**	
V_CSC	**0**	0	0
	1	1	0
	2	0	1

After running the regression, we find that the overall regression is significant since the p-value is less than 0.05 for the chi-square value. For the parameter estimates, B0 and B2 are significant whereas B1 is insignificant.

Therefore, the regression equation is

p(predicted probability) = exp (-2.05 + 2.11 * V_CSC) / 1+ exp (-2.05 + 2.11 * V_CSC)

And the odds ratio is 8.318 with a c value of 62.3%

Question:

5.2 Divide the Churn dataset into two separate datasets (Churn1 and Churn2) by selecting the odd records (1,3,5…etc.) for Churn1 and the even records (2,4,6…etc.) for Churn2.

Perform logistic regression analysis for both Churn datasets on 'day minutes' and compare the results.

Solution:

For the even dataset, we run logistic regression for churn on day minutes. We find that the overall regression is significant since the p-value is less than 0.05 for the chi-square value. The individual parameters are also significant.

Therefore, the regression equation is

p(predicted probability) = exp (-3.30 + 0.008 * day_minutes) / 1+ exp(-3.30 + 0.008 * day_minutes)

And the odds ratio is 1.008 with a c value of 59.9%

Similarly, for the odd dataset, we run logistic regression for churn on day minutes. We find that the overall regression is significant since the p-value is less than 0.05 for the chi-square value. The individual parameters are also significant.

Therefore, the regression equation is

p(predicted probability) = exp (-4.61 + 0.0147 * day_minutes) / 1+ exp(-4.61 + 0.0147 * day_minutes)

And the odds ratio is 1.015 with a c value of 68.8%

Question:

5.3 Perform logistic regression analysis for Churn on :

- *International Plan indicator*
- *Voice Plan indicator*
- *V_CSC2*
- *Account Length*
- *Day Minutes*
- *Evening Minutes*
- *Night Minutes*
- *International minutes*

Solution:

International Plan indicator:

We run logistic regression for churn on International Plan indicator. We find that the overall regression is significant since the p-value is less than 0.05 for the chi-square value. The individual parameters are also significant.

Therefore, the regression equation is

p(predicted probability) = exp (-2.0411+ 1.7355* int_plan_ind) / 1+ exp(-2.0411+ 1.7355* int_plan_ind)

And the odds ratio is 5.672 with a c value of 60.9%

Voice plan Indicator:

We run logistic regression for churn on Voice plan Indicator. We find that the overall regression is significant since the p-value is less than 0.05 for the chi-square value. The individual parameters are also significant.

Therefore, the regression equation is

p(predicted probability) = exp (-2.3537 + 0.7478 * V_voiceplan) / 1+ exp(-2.3537 + 0.7478 * V_voiceplan)

And the odds ratio is 2.112 with a c value of 56.5%

V_CSC2:

We run logistic regression for churn on V_CSC2. We find that the overall regression is significant since the p-value is less than 0.05 for the chi-square value. The individual parameters are also significant.

Therefore, the regression equation is

p(predicted probability) = exp (-2.0652+ 2.1327* V_CSC2) / 1+ exp(-2.0652+ 2.1327* V_CSC2)

And the odds ratio is 8.43 with a c value of 62.0%

Account length:

We run logistic regression for churn on V_CSC2. We find that the overall regression is insignificant since the p-value is greater than 0.05 for chi-square value. Therefore, regression is not valid for account length.

Day Minutes:

We run logistic regression for churn on Day minutes. We find that the overall regression is significant since the p-value is less than 0.05 for the chi-square value. The individual parameters are also significant.

Therefore, the regression equation is

p(predicted probability) = exp (-3.9292+ 0.0113* day_minutes) / 1+ exp(-3.9292+ 0.0113* day_minutes)

And the odds ratio is 1.011 with a c value of 64.0%

Evening minutes:

We run logistic regression for churn on Evening minutes. We find that the overall regression is significant since the p-value is less than 0.05 for the chi-square value. The individual parameters are also significant.

Therefore, the regression equation is

p(predicted probability) = exp (-2.8563+ 0.00526*eve_minutes) / 1+ exp(-2.8563+ 0.00526* eve_minutes)

And the odds ratio is 1.005 with a c value of 57.3%

Night minutes:

We run logistic regression for churn on Night minutes. We find that the overall regression is significant since the p-value is less than 0.05 for the chi-square value. The individual parameters are also significant.

Therefore, the regression equation is

p(predicted probability) = exp (-2.1796+0.002*night_minutes) / 1+ exp(-2.1796+0.002*night_minutes)

And the odds ratio is 1.002 with a c value of 52.9%

International minutes:

We run logistic regression for churn on International minutes. We find that the overall regression is significant since the p-value is less than 0.05 for the chi-square value. The individual parameters are also significant.

Therefore, the regression equation is

p(predicted probability) = exp (-2.5145+ 0.0709*intl_minutes) / 1+ exp(-2.5145+ 0.0709*intl_minutes)

And the odds ratio is 1.073 with a c value of 55.0%