**MARKETING PREDICTIVE ANALYTICS WITH SAS**

**Group 4**

|  |  |  |
| --- | --- | --- |
| **S.no** | **Name** | **NetID** |
| **1** | Xuan Luo | Xxl180010 |
| **2** | Dinesh Varma Indukuri | Dvi170030 |
| **3** | Zhijia Yang | Zxy180017 |
| **4** | Garima Bajaj | Gxb180003 |
| **5** | Yuting Kuang | Yxk162830 |
| **6** | Sehrish Rizvi | Sxr095220 |

**Feature Selection**

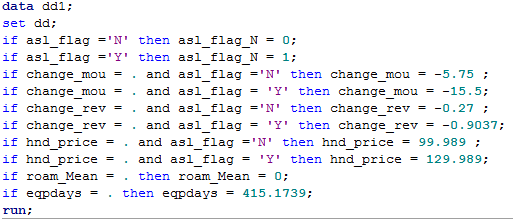
Telecome Customer Churn analysis is done on a dataset with 100000 observations with 173 features. Since it’s hard to perform analysis on all 173 features so we have to choose that features that cause maxmum impact on customer’s churn. So we proceeded as below to select the features for analysis.

1. Group the observations by churn class( 0/1).
2. Compute means of continuous variables for both the classes
3. Take percentage change of both the classes
4. Sort the % change means and choose the top 10-15 variables for analysis.
5. Be sure that these 10-15 variables are not correlated.
6. Now we are done with continuous variables, it’s time for categorical variables.
7. We choose that categorical variables that has zero missing values and used stepwise method to find variables that effects % concordance.
8. Used some more continuous variables with critical thinking and common sense that are not included in above 10-15 variables selected using % change in means technique.
9. Finally selected variables are tested for correlation and uncorrelated variables are used for analysis.

***Features Selected :*** avgqty hnd\_price mou\_opkv\_ Mean months eqpdays change\_mou change\_rev roam\_Mean threeway\_Mean asl\_flag

**Data Preparation**

The above selected features have missing values which need to be imputed. Below is how we imputed. Used asl\_flag categorical variable as class and found the mean, median for the features with missing values. Then imputed the missing values with the median of respective feature under each class.



Since we imputed the missing values with median values we used data to 100% with out any loss.

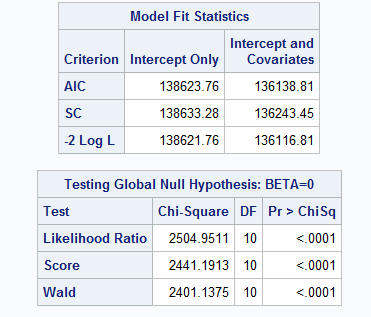
**Model Fitness**:

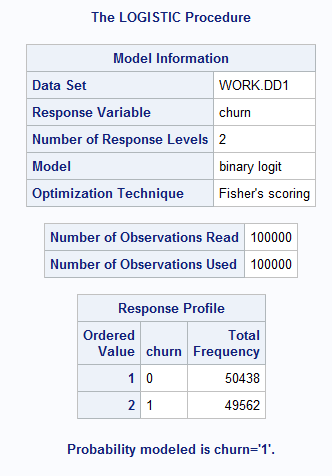
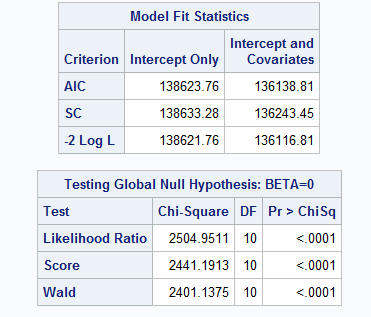
Both AIC & BIC(SC) are lower than intercept only model suggesting our model is a good fit.

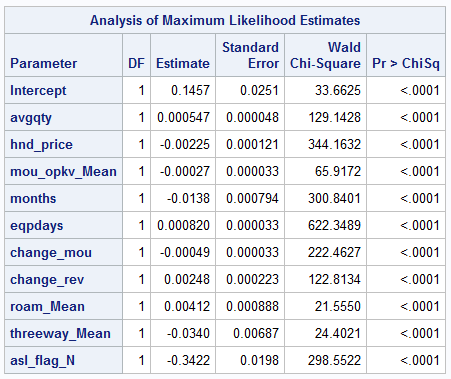
|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **Analysis of Maximum Likelihood Estimates** | | | | | |  |
| **Parameter** | **DF** | **Estimate** | **Standard** | **Wald** | **Pr > ChiSq** | **t-value** |
| **Error** | **Chi-Square** |
| **Intercept** | 1 | 0.1457 | 0.0251 | 33.6625 | <.0001 | 5.804781 |
| **Avgqty** | 1 | 0.000547 | 0.000048 | 129.143 | <.0001 | 11.39583 |
| **hnd\_price** | 1 | -0.00225 | 0.000121 | 344.163 | <.0001 | -18.595 |
| **mou\_opkv\_Mean** | 1 | -0.00027 | 0.000033 | 65.9172 | <.0001 | -8.18182 |
| **Months** | 1 | -0.0138 | 0.000794 | 300.84 | <.0001 | -17.3804 |
| **Eqpdays** | 1 | 0.00082 | 0.000033 | 622.349 | <.0001 | 24.84848 |
| **change\_mou** | 1 | -0.00049 | 0.000033 | 222.463 | <.0001 | -14.8485 |
| **change\_rev** | 1 | 0.00248 | 0.000223 | 122.813 | <.0001 | 11.12108 |
| **roam\_Mean** | 1 | 0.00412 | 0.000888 | 21.555 | <.0001 | 4.63964 |
| **threeway\_Mean** | 1 | -0.034 | 0.00687 | 24.4021 | <.0001 | -4.94905 |
| **asl\_flag\_N** | 1 | -0.3422 | 0.0198 | 298.552 | <.0001 | -17.2828 |

Likelihood ratio is 2504.9511 with p-value < 0.0001 indicating that our model is better than empty model.

***Feature significance and t-values***

 McFadden’s R-sq: 1.807%



**Analysis of Maximum Likelihood**

All the variables considered for analysis are significant at 95% confidence level.

Below is interpretation of each feature in model using Log odds having other factors fixed

*Avgqty :* For every unit increase in average number of calls then log odds of churn increases by 0.000547 compared to no churn.

*hnd\_price :* For every unit increase in handset price then log odds of churn decreases by 0.00225 compared to no churn.

*mou\_opkv\_Mean:* For every unit increase in mou\_opkv\_Mean then log odds of churn decreases by 0.00027 compared to no churn.

*Months*: For every unit increase in months of usage then log aodds of churn decreases by 0.0138 compared to no churn.

*Eqpdays:* For every unit increase in eqpdays then log odds of churn increases by 0.00082 compared to no churn.

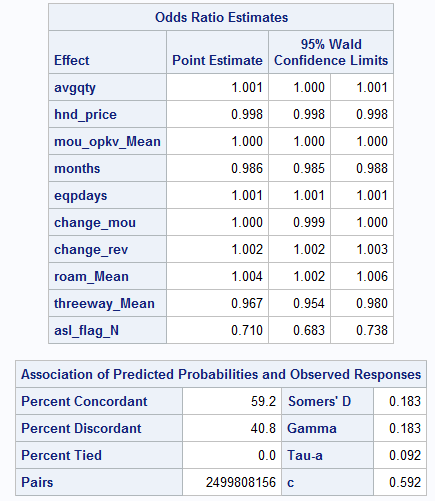
*change\_mou:* For every unit increase in change\_mou then log odds of churn decreases by 0.00049 compared to no churn.

*change\_rev:* For every unit increase in change\_rev then log odds of churn increases by 0.0024 compared to no churn

*roam\_Mean:* For every unit increase in roam\_Mean then log odds of churn increase by 0.00412 compared to no churn

*threeway\_Mean :* For every unit increase in threeway\_Mean then log odds of churn decreases by 0.0340 compared to no churn.

*asl\_type:* For a person with Account Spending Limits = ‘Y’ the the log odds of churn is decreases by 0.3422 compated to a person with Account Spending Limits = ‘N’



**Odd Ratio Estimators interpretation**.

Below is interpretation of each feature in model using odds ratio having other factors fixed

*Avgqty :* For every unit increase in average number of calls then odds of churn increases by 1.001 compared to no churn.

*hnd\_price :* For every unit increase in handset price then odds of churn decreases by 0.998 compared to no churn.

*mou\_opkv\_Mean:* For every unit increase in mou\_opkv\_Mean then odds of churn is same as no churn.

*Months*: For every unit increase in months of usage then aodds of churn increases by 0.986 compared to no churn.

*Eqpdays:* For every unit increase in eqpdays then odds of churn increases by 1.001 compared to no churn.

*change\_mou:* For every unit increase in change\_mou then odds of churn is same as no churn.

*change\_rev:* For every unit increase in change\_rev then odds of churn increases by 1.002 compared to no churn

*roam\_Mean:* For every unit increase in roam\_Mean then odds of churn increases by 1.004 compared to no churn

*threeway\_Mean :* For every unit increase in threeway\_Mean then odds of churn increases by 0.967 compared to no churn.

*asl\_type:* For a person with Account Spending Limits = ‘Y’ the the odds of churn is increases by 0.71 compated to a person with Account Spending Limits = ‘N’

**Top 3 Variables**: Found based on Wald Chi-square value

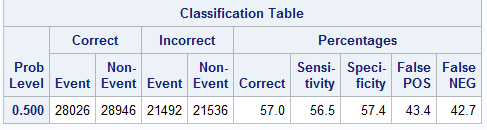
eqpdays

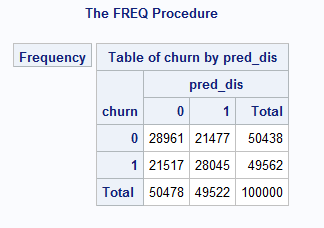
hnd\_price

months

**Other Features that can improve model**

1. Deals – Deals that offer lower charges for calls per minute, high speed internet at lower price can reduce the churn rate.
2. Having data about network availability or frequency of call drop can help model fit better.
3. For targeting international students like us – night call rates effects churn rate.





Percentage Concordance = 59.2%

Hit Ratio = % of events correctly classified

= (True positive + False Neagtive)/Total = (28045+28961)/100000 = 57%

Sensitivity = 56.5%