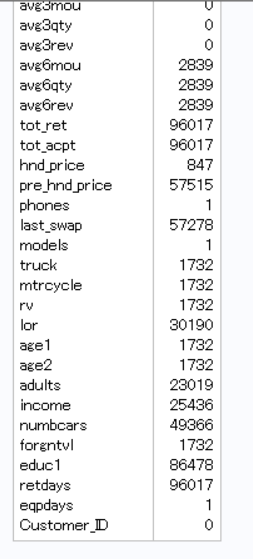
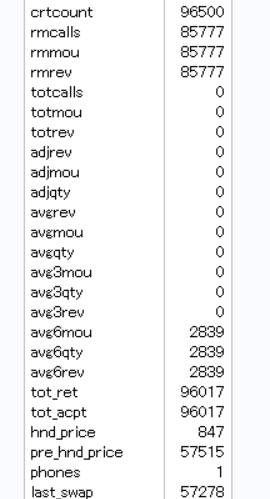
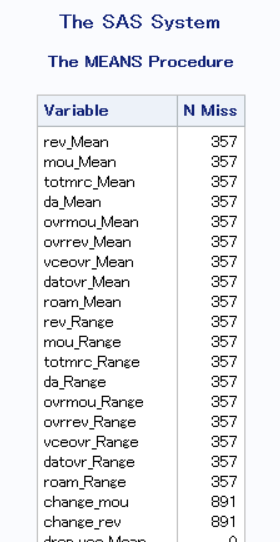
**Churn Factors Analysis for Telecom Company With SAS**

**Subject:**Use the Churn data from a telecom company to understand what factors are good predictors of churn.

**Dataset:**Churn is the dependent variable that takes the value 1 if a customer has churned (or left the company) and 0 otherwise. In the sample, 50% of the 100,000 customers have churned.

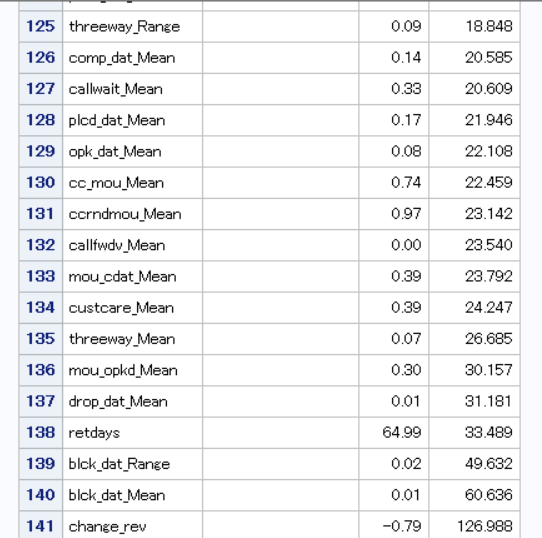
**Data Processing and Modeling:**

After removing the missing values ,we consider around 90% of the original data.



Then we use different methods to find the different importance of data.

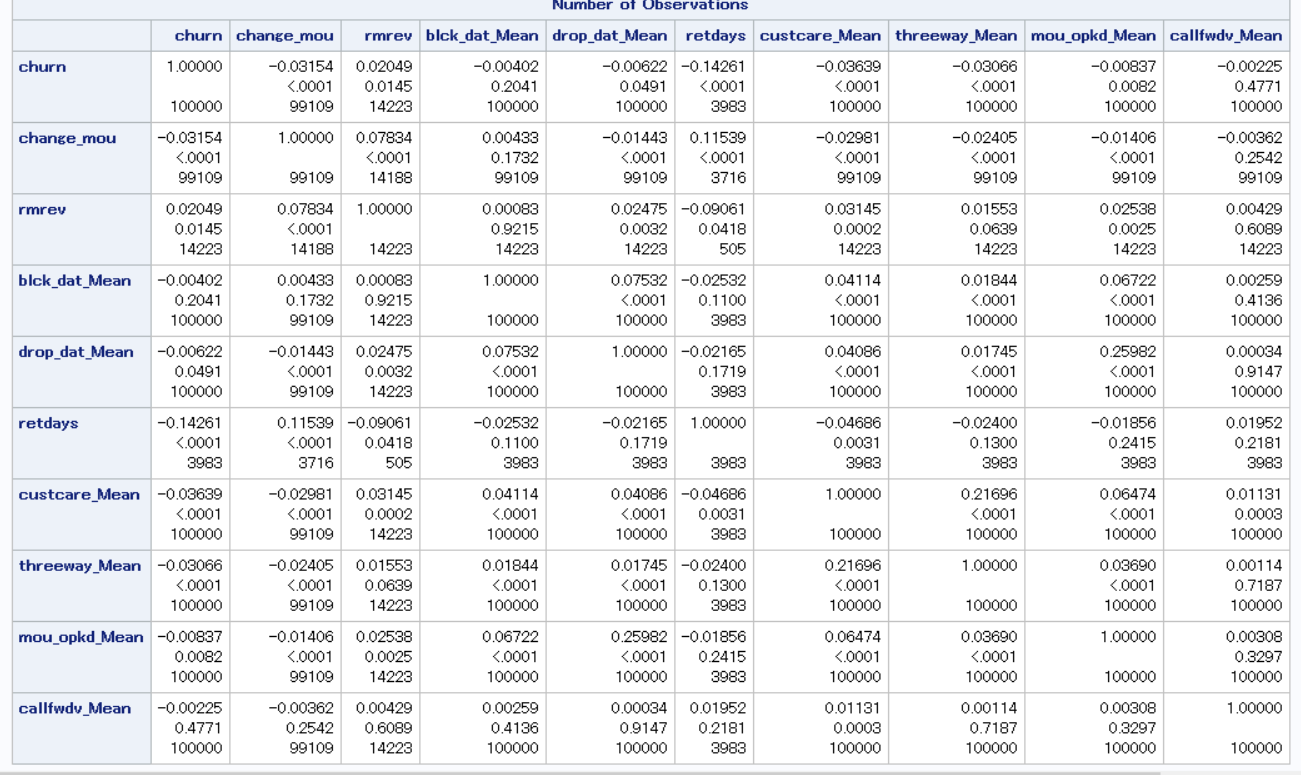
For continuous variables, we compute means of all x variables for churners and non-churners separately and compute the percentage difference for each variable.Then we find below variables：



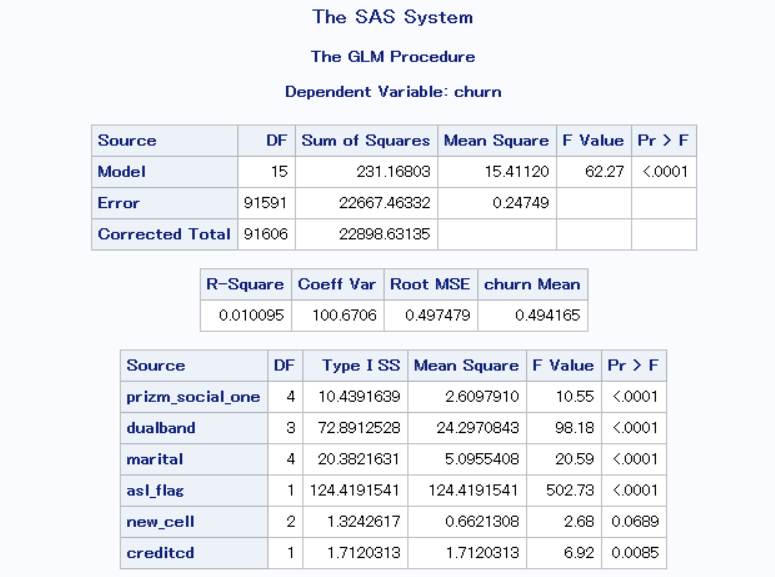
Then we sort the variables from high to low based on percentage difference in means. The we chose the top 10 variables from bottem and top 3 from top.

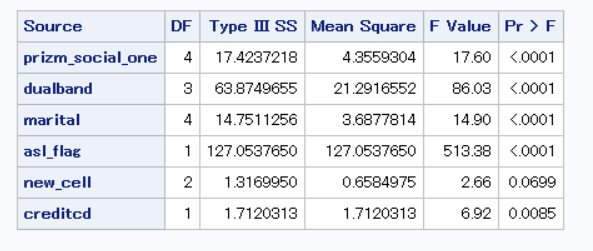
Next,we use correlation analysis to avoid high correlated problem.

And we found some variables are high correlated,such as roam\_Mean and rmrev(we chosed rmrev finally ),mou\_cdat\_Mean and mou\_opkd\_Mean( we chosed mou\_opkd\_Mean finaly based on the higher correlation with churn)(the final correlations as follows:)

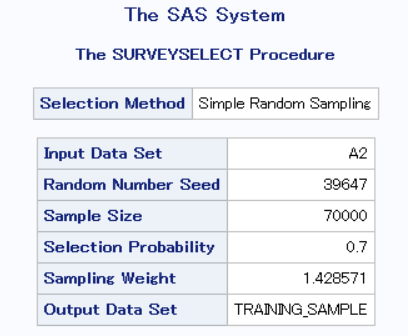


For category variables,we tried to use glm (or anova or chi-test) to find the relationship between those category variables and churn,then we found that some variables ,such as asl\_flag can have significant effect on the churn.

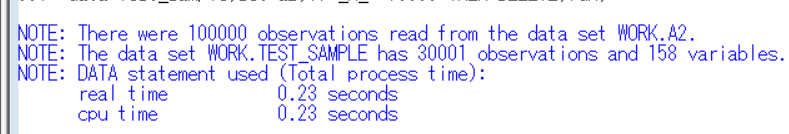




Then we created a random sample of 70,000 customers using PROC SURVEYSELECT in SAS. Call this the estimation sample (training sample)

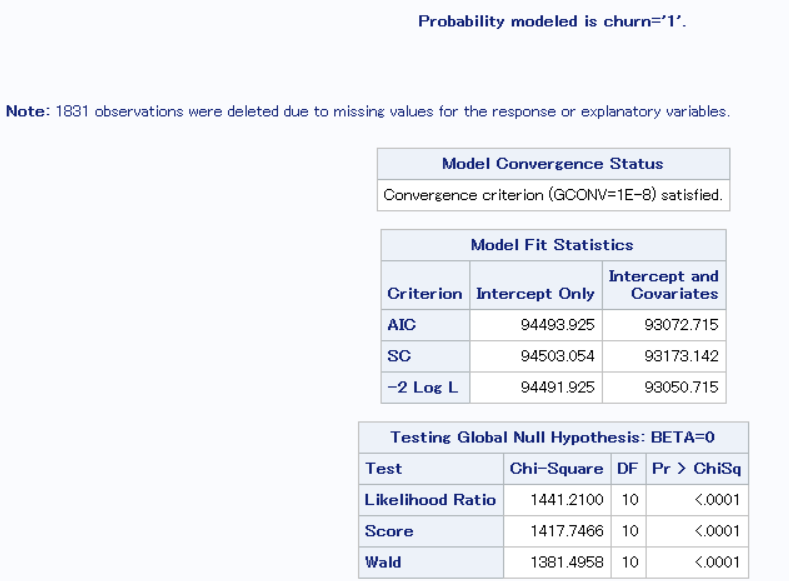
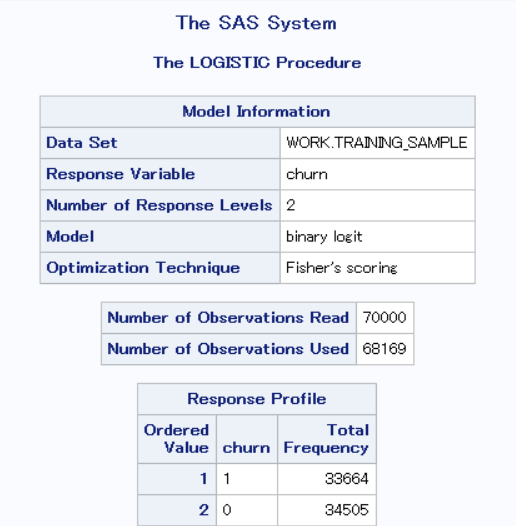


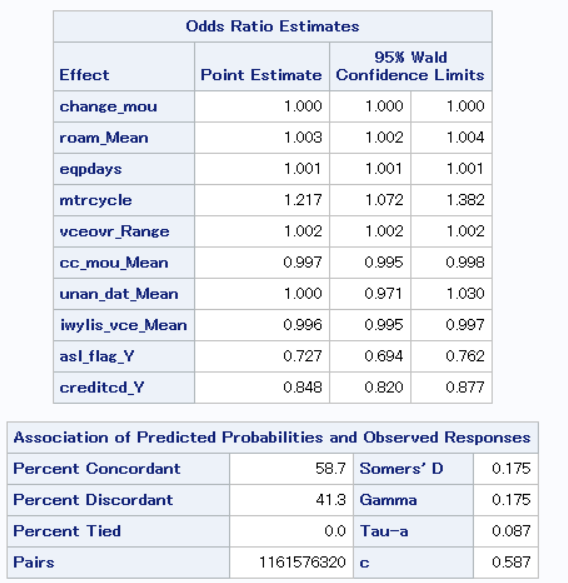
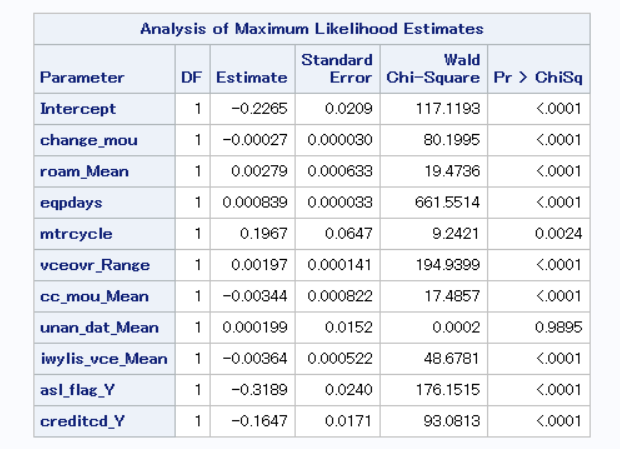
Created a holdout sample with the rest of the 30000 customers (call it the test sample).



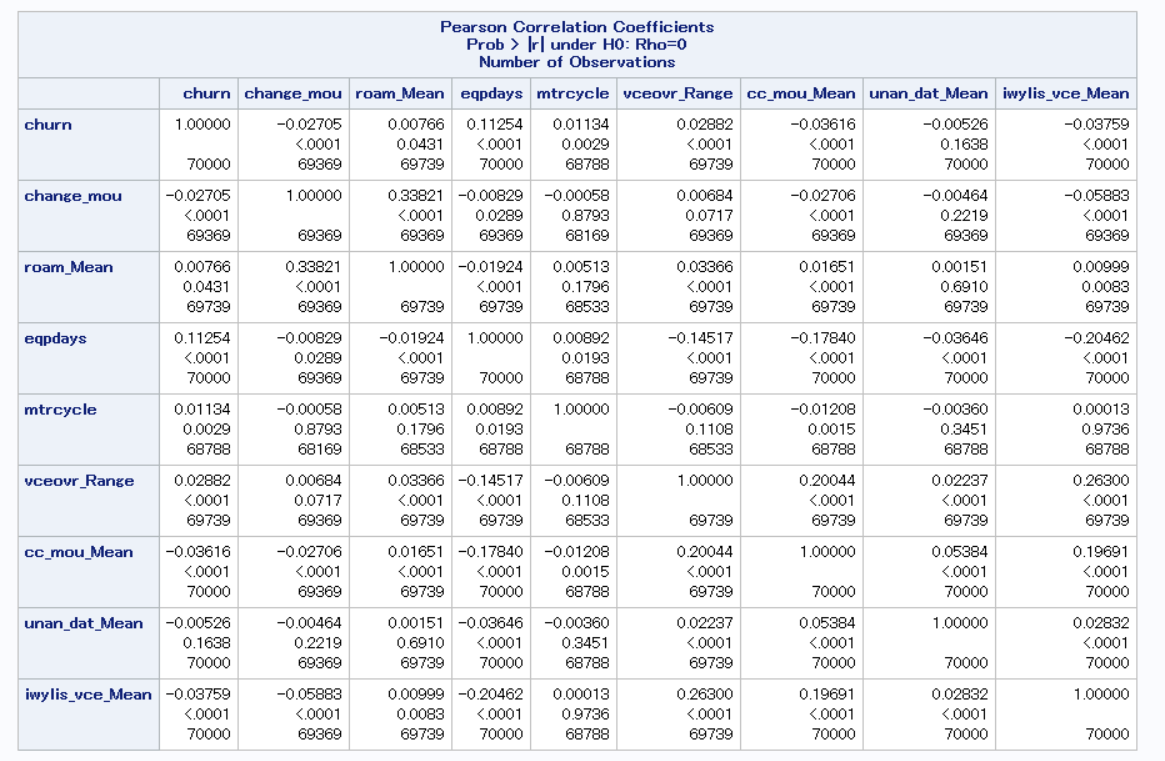
Finally ,we used a logistic regression model to build the best model and used correlation analysis or VIF to determine the correlation between the variables.

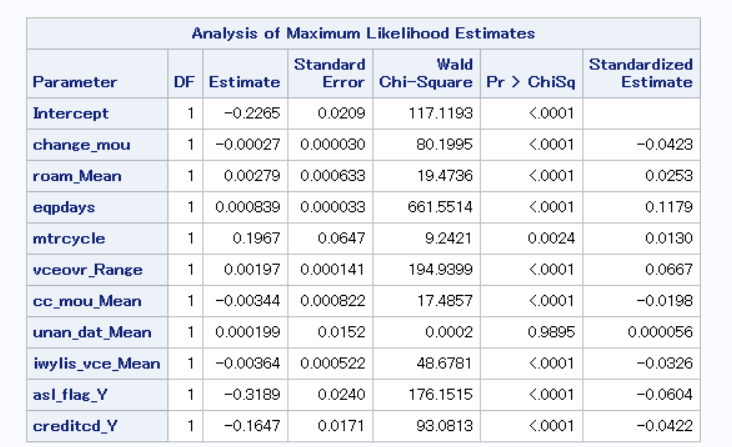
Final model:



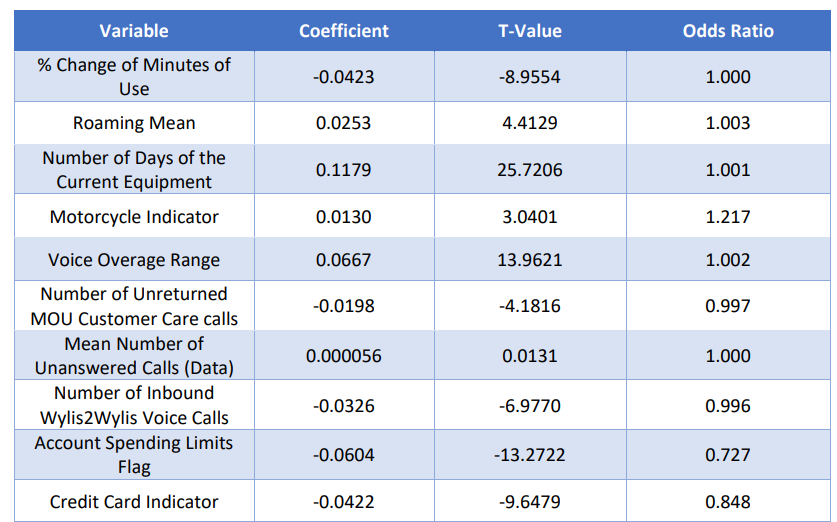


correlation analysis：





**Conclusions:**



The betas are coefficients which represent the effect of a variable on the log-odds of churn. In

other words, when a variable increases by one unit, the log-odds of churn increases or

decreases based on its beta. Meanwhile, the odds ratio represents a multiple of times the odds

of an event will happen for every unit change in size. For example, when the number of days of

the current equipment changes by one, then the odds of churn increases by a multiple of 1.001.

These are outlined in the table above.The AIC was 94493, where a lower AIC indicates a model with a better fit. The prediction accuracy (percent concordance) was 58.7%, indicating the percentage of pairs where churn events had a higher predicted probability of churn than non-churn.

Out of 68,169 training cases, our model correctly predicted 39,130 of them, resulting in

a 57% hit ratio.5. Using the model parameters predict the churn for the holdout sample (test data) as

well and compute the hit ratio.

Out of 29,222 test cases, our model correctly predicted 16,889 of them, resulting in a

58% hit ratio

Top three factors that affect churn:

Based on the odds ratio, the top three factors that affect churn in our model are:

• Motorcycle Indicator: 1.217

• Account Spending Limits Flag: 0.727

• Credit Card Indicator: 0.848

Interpreting the odds ratio gives us the following effect sizes on churn:

• Motorcycle Indicator: 21.70% increase in odds of churn when the account has a

motorcycle indicator (when mtrcycle = 1)

• Account Spending Limits Flag: 27.30% decrease in odds of churn when there is a

flag on account spending limit (when asl\_flag\_Y = 1)

• Credit Card Indicator: 15.20% decrease in odds churn when there is a credit card

indicator (when creditcd\_Y = 1)

Some other variables that could help to improve the fit of the model of the model are

variables relating to how customers are interacting with the competitors – for example,

the number of calls a customer had with a competitor, number of search results for a

competitor by a customer, or number of clicks on competitor marketing advertisements.

Other variables that may help to improve the fit of the model are customer satisfaction

data. This may include ratings from surveys after customer service calls, rankings of

services they prioritize or prefer most, and number of time a customer has engaged