A Nonparametric Statistical Analysis of

Telcom Customer Churn Analysis

Haichuan Du

Villanova University

Dr. Jesse Frey

MAT 8452

24 July 2020

Customer churn is a major problem and one of the biggest concerns for the telecommunication industry. Since the cost of retaining an existing customer is much lower than acquiring a new one, identifying the factors that increase customer churn is important.

The raw data contains a total of 7043 observations and 21 variables. I classified variables into three groups, which were the customer demographic features, service attributes and contract attributes. For the customer demographic features, I examined both parametric and nonparametric Chi-Square Tests. At the conventional significance level of 0.05, both tests concluded that, in addition to gender, there was an association between churn and user attributes. Through the histogram, we found that senior citizens and customers without partners and dependents were more likely to churn.

For contract attributes, I ran the Kolmogorov-Smirnov test to conclude that customer retention and customer churn do not come from the same distribution. I also performed a sign test, and the results showed that the 50th of the monthly charge in my dataset is significantly different from the national statistics ($65.5). The monthly charge of losing customers is relatively high, I was interested in knowing if the monthly charge differs between partner. The Wilcoxon rank sum test provided us with a W test statistic of 6954306, with a p value of <0.0001, and concluded that having a partner will increase the monthly charge.

Since the longer the contract period, the less likely to lose customers, I wanted to compare the difference in the monthly charge between the churn and contract. For retained loyal customers, I used Ansari-Bradley and the normal F test to conclude that the variability of a One-Year contract is greater than that of a monthly contract. Second, I conducted a rank sum test, and the results showed that we did not have enough evidence at the 0.05 level to conclude that the monthly charge of the one-year and two-year contracts is not equal. With the three numeric

variables I used throughout this analysis, I also wanted to see if monthly charges had any sort of association with each tenure. To accomplish this, I made a scatterplot and found that there is a fairly strong, positive, linear association between monthly charge and tenure. To verify this result, I ran Spearman's and Pearson correlation test, both tests concluded that these two variables are linearly related.

For service attributes, I decided to determine if there exists a correlation between the churn and the services customers choose. For comparison, I examined both parametric and nonparametric Chi-Square Tests and showed that there exists an association between churn and services customers choose excepting PhoneService.

Finally, I was interested in finding the scale effect of how variables can affect the churn. I used the variables tenure, MonthlyCharges, InternetService and PhoneService to address this problem. I applied the bootstrap method for this logistic regression and found that 95% bootstrap confidence interval for the coefficient of tenure, churn, and PhoneService are consistent with the results what we have been found before, such as customers with internet service, in particular, fiber service is more likely to churn.

In summary, there appeared to be several findings that could reduce customer churn. First, customized services, such as family packages for the senior citizens, and users who have no dependencies and partner. For new customers, a half-year discount is given to pass the peak period of user loss. On the other hand, the new customers are promised free network upgrades and free TV, movie and other monthly services to improve user stickability. For monthly contract users, the annual contract payment discount activity is launched to convert monthly contract users into annual contract users to achieve a higher user retention rate. For monthly charge, it is best to set up the price around 64 dollars to retain the customers.