International Journal of Multidisciplinary Research and Development
Online ISSN: 2349-4182, Print ISSN: 2349-5979, Impact Factor: RJIF 5.72
www.allsubjectjournal.com
Volume 4; Issue 4; April 2017; Page No. 121-126

Survival analysis of customers in telecommunication companies by using Cox proportional hazards regression model

Ramesh Khatana, * Annasaheb Suryawanshi

Department of Statistics, Saurashtra University, Rajkot, Gujarat, India

Abstract
The objective of the study was to assess the importance of various covariates in the survival times of individuals through the hazard function. The Kaplan-Meier curve compares customer survival across different groups by using Log-Rank test but the Cox proportional hazard regression simultaneously investigate the effects of several explanatory variables on the survival of customer. Cox regression was used for data analysis on 300 customers in various telecommunication companies. This research showed that, among all factors examined, age, gender, region and sim Card Company plays important role in customer survival in the telecommunications company.

Keywords: cox proportional hazards regression, customer survival, omnibus test

1. Introduction

In survival analysis, survival time is referred as time variable. For this study, the time between fixed starting point (date of subscription to the service) and a termination event (date of cancellation of the service) is consider as survival time. The survival time not known exactly for those customers who have not cancelled their service at the end of the follow-up period and this is called as censoring. Several statistical methods are available for modeling survival data.

One of the most popular regression techniques for survival analysis is Cox proportional hazards regression. The Cox proportional hazards model (Cox, 1972) is the most commonly used multivariate approach for analyzing survival time data. It helps to assess the importance of various covariates in the survival times of individuals through the hazard function. The Cox proportional hazards regression model is popular because the result from using the Cox model will closely approximate the result for the correct parametric model. i.e. The Cox proportional model is robust model for survival analysis. In this study, we are interested in comparing groups with respect to their hazards. The Cox proportional hazards regression model is written in terms of the hazard model as,

$$h(t, X) = h_0(t) \exp(\beta_1 X_1 + \beta_2 X_2 + \ldots + \beta_p X_p)$$

This model gives an expression for the hazard at time t for an individual with a given specification of a set of explanatory variables denoted by X. That is, X represents the collection of predictor variables that is being modeled to predict an individual hazard. The term $h_0(t)$ is called the baseline hazard function and represents the hazard when all of the predictors X_1, X_2, X_P are equal to zero.

Ken Kwong-Kay Wong (2011) studied the customer retention in the context of a Canadian wireless telecommunications company and explores the predictors of churn incidence as part of customer relationship management. Their study shows that rate plan suitability plays a key role in influencing customer churn. Jones, M. *et al.* (2000) [3] studied the switching barriers and repurchase intentions in services and their result indicate that the influence of core-service satisfaction on repurchase intentions decreases under conditions of high switching barriers. Ganesh, J. *et al.* (2000) [4] examined the differences Between Switchers and stayers in service industry. Their study shows that, customers who have switched service providers because of dissatisfaction seem to differ significantly from other customer groups in their satisfaction and loyalty behaviors.

In present study, the predictor variables like age, gender, region, marital status, sim Card Company etc. are used to predict individual hazard. Result of this study gives the significant predictors to customer survival in telecommunication companies.

2. Material and methods

The data collected from 300 respondents who are using services of different telecommunication companies. This study is based on questionnaire method. The respondents are belongs to rural as well as urban area of the Rajkot. The information collected from Airtel, BSNL, Idea, JIO, Reliance, Tata DoCoMo, Uninor and Vodafone service users. The data collection took place from October to December 2016. The survival analysis technique Cox proportional hazard regression is used for this research. The SPSS software has been used for Cox proportional hazard regression analysis.

3. Result and Discussion

To increase customer survival in Telecommunications Company, the important task is to find out the most important factors which are affecting on customer survival. For this study, a random sample of customers is selected and their time spent as customers is calculated and find out the current status of each customer i.e. active

or not active with their service. For this study, Cox Regression is used to determine which factors are associated with shorter survival.

In Case Processing Summary (table-1) the status variable shows whether the event has occurred for a given case (sim card not active). If the event has not occurred (sim card active), so it is called censored cases. Here, censored cases are not used in the computation of the regression coefficients, but they are used to compute the baseline hazard. The case processing summary shows that, 251 cases are censored. i.e. These customers are still in service. The case processing summary also shows that, 40 customers are left the service i.e. Event occurred. The total number of customers was 300.

Table 1: Case Processing Summary

		N	Percent
Cases available in analysis	Event[a]	40	13.3%
	Censored	251	83.7%
	Total	291	97.0%
Cases dropped	Cases with missing values	0	0.0%
	Cases with negative time	0	0.0%
	Censored cases before the earliest event in a stratum	9	3.0%
	Total	9	3.0%
Total		300	100.0%

a. Dependent Variable: Customer survival time (Months)

The table-2 shows that the categorical variable coding. These are useful to the interpretation of regression coefficients for categorical covariates. For this analysis the reference category is the "last" category. In the variable gender, female and male customers have variable values coded as 1 and 2 in the data file, but for regression analysis purposes it is coded as 1 and 0 respectively.

Table 2: Categorical Variable Codings

		Frequency	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Gender	1=Female	59	1						
	2=Male	241	0						
Marital status	1=married	83	1						
	2=Unmarried	217	0						
Region	1=Rural area	124	1						
	2=Urban area	176	0						
Sim card company	1=Airtel	61	1	0	0	0	0	0	0
	2=BSNL	17	0	1	0	0	0	0	0
	3=Idea	90	0	0	1	0	0	0	0
	4=JIO	16	0	0	0	1	0	0	0
	5=Reliance	14	0	0	0	0	1	0	0
	6=Tata-docomo	19	0	0	0	0	0	1	0
	7=Uninor	15	0	0	0	0	0	0	1
	8=Vodafone	68	0	0	0	0	0	0	0
Type of service plan	1=Postpaid	17	1						
	2=Prepaid	283	0						

Block 1: Forward Stepwise Method (Likelihood Ratio)

The omnibus tests (table-3) are measures of how well the model performs. The chi-square change from previous step is the difference between the –2 log-likelihood of the model at the previous step and the current step. In omnibus tests if the step was to add a variable, the inclusion makes sense if the significance of the change is less than 0.05. If the step was to remove a variable, the exclusion makes sense if the significance of the change is greater than 0.10.

In Omnibus Tests of Model Coefficients table-3, in the first, second, third and fourth steps the variable age, gender, sim Card Company and region respectively are added to the model. The change from previous step and change from previous block both report the effect of adding Sim Card Company to the model selected in Block-1. Since the significance value of the change is less than 0.05 at step-4 it indicates that Sim Card Company contributes to the model.

Table 3: Omnibus Tests of Model Coefficients

Step	-2 Log Likelihood	Overall (score)			Change From Previous Step			Change From Previous Block		
		Chi-square	df	Sig.	Chi-square	df	Sig.	Chi-square	df	Sig.
1[a]	387.820	11.710	1	0.001	17.683	1	0.000	17.683	1	0.000
2[b]	380.946	18.389	2	0.000	6.874	1	0.009	24.557	2	0.000
3[c]	367.870	33.621	9	0.000	13.076	7	0.070	37.633	9	0.000
4[d]	362.919	38.496	10	0.000	4.951	1	0.026	42.584	10	0.000

a. Variable(s) Entered at Step Number 1: Age
b. Variable(s) Entered at Step Number 2: Gender
c. Variable(s) Entered at Step Number 3: Sim card company
d. Variable(s) Entered at Step Number 4: Region

In table-4, the final model (step-4) includes age, gender, region, and Sim Card Company. To understand the effects of individual predictors, see the Exp (B), which interprets the predicted change in the hazard for a unit increase in the predictor. There is one assumption for Cox's regression which is the proportional hazards assumption that the hazard ratio between two groups remains constant over time. The Exp(B) value for Age is 0.880341, so an increase in Age of 1 year will be associated with a 0.880341 fold increase in the hazard. The Exp (B) value for variable gender is 3.193557, that the hazard for an female customer is 3.193557 times that of a male customer. The Exp (B) value for variable region is 2.170379, that the hazard for an Rural area customer is 2.170379 times that of a urban area customer. The regression coefficients for the first seven levels of *Sim Card Company* are relative to the reference category, which corresponds to *Vodafone* customers.

The P-value for the second, fifth and seventh categories, corresponding to BSNL, Reliance and *Uninor* customers, are less than 0.05, which means they are statistically different from the vodafone customers. However, the P-value for first, third, fourth and sixth categories, corresponding to Airtel, Idea, JIO and Tata-docomo customers are greater than 0.05, which means they are statistically not different from the vodafone customers. So any observed difference between these customer categories could be due to chance.

The regression coefficient for the first category, corresponding to Airtel customers, suggests that the hazard for Airtel customers is 2.362795 times that of Vodafone customers. The regression coefficient for the second category, corresponding to BSNL customers, suggests that the hazard for BSNL customers is 4.517930 times that of Vodafone customers. The regression coefficient for the third category, corresponding to Idea customers, suggests that the hazard for Idea customers is 1.662418 times that of Vodafone customers.

The regression coefficient for the fourth category, corresponding to JIO customers, suggests that the hazard for JIO customers is 0.000033 times that of Vodafone customers. The regression coefficient for the fifth category, corresponding to Reliance customers, suggests that the hazard for Reliance customers is 8.490337 times that of Vodafone customers. The regression coefficient for the sixth category, corresponding to Tata-docomo customers, suggests that the hazard for Tata-docomo customers is 3.437694 times that of Vodafone customers. The regression coefficient for the seventh category, corresponding to Uninor customers, suggests that the hazard for Uninor customers is 10.969276 times that of Vodafone customers.

Table 4: Variables in the Equation

		B	SE	Wald	df	Sig.	Exp(B)
Step 1	Age	-.107	.035	9.402	1	.002	0.898915
Step 2	Age	-.118	.038	9.549	1	.002	0.888925
	Gender	.963	.346	7.762	1	.005	2.620515
Step 3	Age	-.121	.040	9.244	1	.002	0.885878
	Gender	1.054	.360	8.591	1	.003	2.869318
	Sim card company			13.098	7	.070	
	Sim card company (1)	.958	.536	3.190	1	.074	2.605244
	Sim card company (2)	1.365	.674	4.102	1	.043	3.917361
	Sim card company (3)	.545	.552	.974	1	.324	1.724561
	Sim card company (4)	-10.158	362.605	.001	1	.978	0.000039
	Sim card company (5)	1.736	.744	5.449	1	.020	5.671955
	Sim card company (6)	1.002	.741	1.830	1	.176	2.724714
	Sim card company (7)	2.250	.761	8.745	1	.003	9.483655
Step 4	Age	-.127	.039	10.471	1	.001	0.880341
	Gender	1.161	.365	10.143	1	.001	3.193557
	Region	.775	.352	4.858	1	.028	2.170379
	Sim card company			16.018	7	.025	
	Sim card company (1)	.860	.539	2.548	1	.110	2.362795

	Sim card company (2)	1.508	.678	4.947	1	.026	4.517930
	Sim card company (3)	.508	.552	.849	1	.357	1.662418
	Sim card company (4)	-10.333	355.874	.001	1	.977	0.000033
	Sim card company (5)	2.139	.772	7.674	1	.006	8.490337
	Sim card company (6)	1.235	.750	2.709	1	.100	3.437694
	Sim card company (7)	2.395	.766	9.773	1	.002	10.969276

Variables left out of the model all have score statistics with significance values greater than 0.05. This is shown in table-5.

Table 5: Variables not in the Equation

		Score	df	Sig.
Step 1	Gender	8.366	1	.004
	Marital status	.538	1	.463
	Region	1.449	1	.229
	Sim card company	14.007	7	.051
	Sim card company (1)	1.101	1	.294
	Sim card company (2)	2.078	1	.149
	Sim card company (3)	1.167	1	.280
	Sim card company (4)	.507	1	.477
	Sim card company (5)	2.483	1	.115
	Sim card company (6)	.057	1	.811
	Sim card company (7)	4.908	1	.027
	Type of service plan	.184	1	.668
Step 2	Marital status	.788	1	.375
	Region	2.414	1	.120
	Sim card company	16.117	7	.024
	Sim card company (1)	.661	1	.416
	Sim card company (2)	1.594	1	.207
	Sim card company (3)	.802	1	.370
	Sim card company (4)	.455	1	.500
	Sim card company (5)	2.986	1	.084
	Sim card company (6)	.158	1	.691
	Sim card company (7)	7.258	1	.007
	Type of service plan	.360	1	.548
Step 3	Marital status	.169	1	.681
	Region	5.047	1	.025
	Type of service plan	.512	1	.474
Step 4	Marital status	.394	1	.530
	Type of service plan	.275	1	.600

Table-6 shows the Covariate Means and Pattern. It is Values displays the average value of each predictor variable, plus a pattern for each level of Sim Card Company. The eight patterns correspond to each of the Sim card companies, each with otherwise "average" predictor values.

Table 6: Covariate Means and Pattern Values

	Mean	Pattern							
		1	2	3	4	5	6	7	8
Age	25.216	25.216	25.216	25.216	25.216	25.216	25.216	25.216	25.216
Gender	.192	.192	.192	.192	.192	.192	.192	.192	.192
Marital status	.285	.285	.285	.285	.285	.285	.285	.285	.285
Region	.416	.416	.416	.416	.416	.416	.416	.416	.416
Sim card company (1)	.206	1.000	.000	.000	.000	.000	.000	.000	.000
Sim card company (2)	.055	.000	1.000	.000	.000	.000	.000	.000	.000
Sim card company (3)	.302	.000	.000	1.000	.000	.000	.000	.000	.000
Sim card company (4)	.048	.000	.000	.000	1.000	.000	.000	.000	.000
Sim card company (5)	.045	.000	.000	.000	.000	1.000	.000	.000	.000
Sim card company (6)	.065	.000	.000	.000	.000	.000	1.000	.000	.000
Sim card company (7)	.048	.000	.000	.000	.000	.000	.000	1.000	.000
Type of service plan	.058	.058	.058	.058	.058	.058	.058	.058	.058

Fig-1 shows the plot of the survival curves for each covariate pattern gives a visual representation of the effect of sim Card Company. Vodafone and JIO customers have higher survival curves. However, Uninor and Reliance customers have lower survival curves

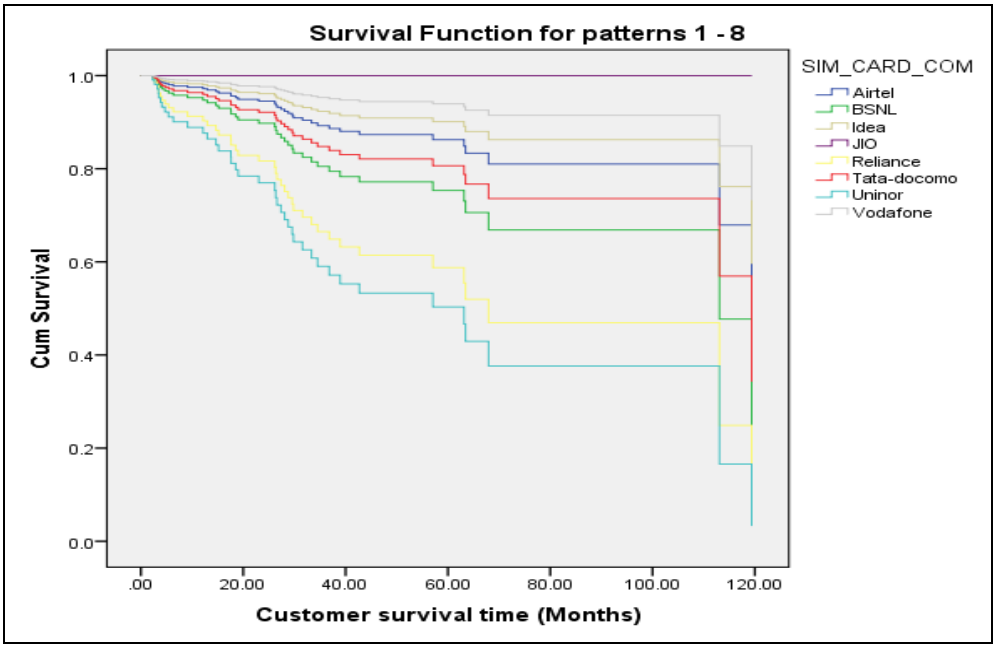

Fig 1

Fig - 2 shows the plot of the hazard curves for each covariate pattern gives a visual representation of the effect of sim Card Company. Vodafone and JIO customers have lower hazard curves. However, Uninor and Reliance customers have higher hazard curves.

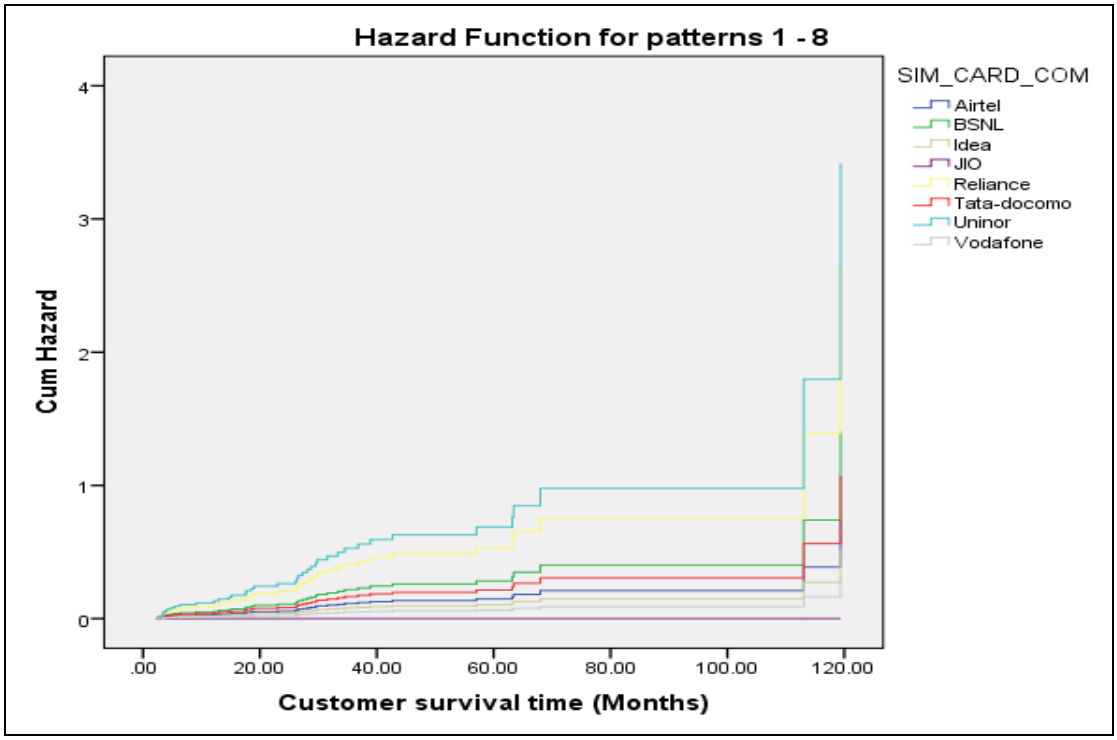

Fig 2

4. Conclusions

The findings of this study show that, as customer age increase above average age (25.0833 years) will be associated with decrease in the hazard (i.e. increase in survival). The Vodafone and JIO companies have higher survival (lower hazard). The male customers have longer survival (shorter hazard) as compared to female customers. The customers living in urban area has better

survival as compare to rural area customers. These results are helpful to Telecommunication Companies to predict customer survival.

5. References

1. Cox DR. Regression Models and Life-Tables (with discussion). Journal of the Royal Statistical Society – Series B. 1972; 34:187-220.
2. Ken Kwong-Kay Wong. Using Cox regression to model customer time to churn in the wireless telecommunications industry, Journal of Targeting, Measurement and Analysis for Marketing. 2011; 19(1):37-43.
3. Jones M, Mothersbaugh D, Betty S. Switching barriers and repurchase intentions in services. Journal of Retailing. 2000; 76(2):257-272.
4. Ganesh J, Arnold M, Reynolds K. Understanding the customer base of service providers: An examination of the differences between switchers and stayers. Journal of Marketing. 2000; 64(3):65-87.