Customer Analytics

—Mark5827 Research Project

08.11.2022

Group Numbers：

Ke Wu z5158146
Yihan Wang z5410427

Task 1

Base survival model, model summary and visualize the survival curve

```
> # ============================ Task1 =========================================
>
> mod0 <- survreg(Surv(AccountWeeks, Churn) ~ 1, data = Telco1)  # empty model without any explanatory factors.
> summary(mod0)  # summary is here

Call:
survreg(formula = Surv(AccountWeeks, Churn) ~ 1, data = Telco1)
               Value Std. Error   z       p
(Intercept)   5.4217     0.0251  216 <2e-16
Log(scale)   -1.0337     0.0357  -29 <2e-16

Scale= 0.356

Weibull distribution
Loglik(model)= -3383.9   Loglik(intercept only)= -3383.9
Number of Newton-Raphson Iterations: 10
n= 3333
```
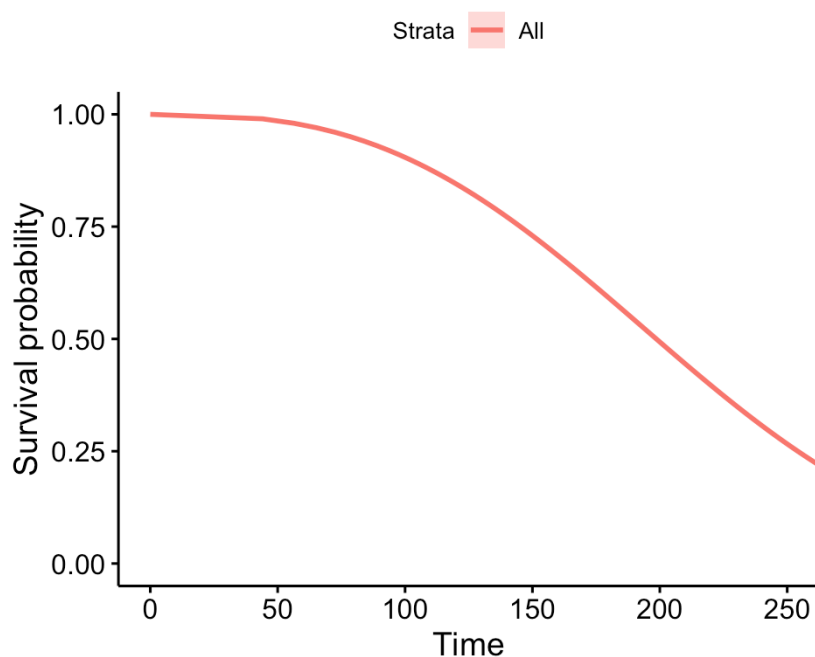
Task 2

```
> #  ============================ Task2 ========================================
>
> mod1 <- survreg(Surv(AccountWeeks, Churn) ~ DataPlan, data = Telco1)#model without one explanatory factors.
> summary(mod1)

Call:
survreg(formula = Surv(AccountWeeks, Churn) ~ DataPlan, data = Telco1)
              Value Std. Error      z        p
(Intercept)  5.3747     0.0248 216.6 < 2e-16
DataPlan     0.2273     0.0446   5.1 3.4e-07
Log(scale)  -1.0299     0.0356 -28.9 < 2e-16

Scale= 0.357

Weibull distribution
Loglik(model)= -3368.4   Loglik(intercept only)= -3383.9
        Chisq= 30.98 on 1 degrees of freedom, p= 2.6e-08
Number of Newton-Raphson Iterations: 8
n= 3333
```
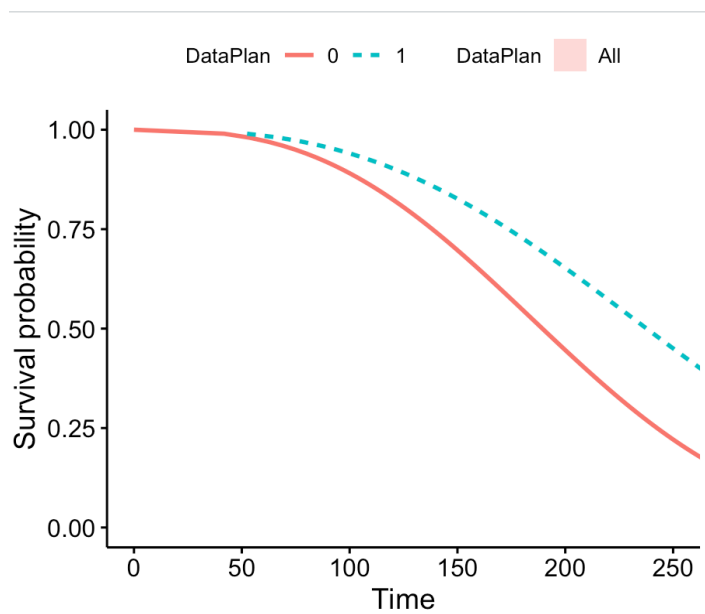
Task 3

From the fitting results of the Mod1 model, it can be seen that for Mod1 and Mod0, the user retention rate is basically unchanged during the first 50 weeks. However, for Mod1, the customer retention rate drops to 50% by about 190 weeks. For Mod1, we added the dummy variable of whether or not to subscribe to a data plan as an explanatory variable (DataPlan). The model shows that the retention rate also declines gradually over time, but at a significantly slower rate. Specifically, for the first 80 weeks, the retention rate is essentially constant, but by approximately 240 weeks, the retention rate drops to 50%. The coefficient of DataPlan in the model equation is 0.2273, which is significant at the 5% level, again indicating that subscription data plans (DataPlan) can significantly improve the customer retention.

Although the comparison of these two models allows us to conclude that DataPlan can be effective in improving customer retention, there are still areas of Mod1 that require our attention, which can be apparent in terms of two aspects below:

1. The sample duration we took ranges from one year to 1.5 years. A limited sample can only lead to limited conclusions, and we do not know if subscribing to a data plan will negatively impact retention over the long term (more than 2 years). Also, with a limited sample space, our calculation of CLV will be biased.

2. The model may have the problem of missing variables, although both the prediction curve and the model equation yield a significant positive effect of data plans on customer retention. That is, what affects customer retention may be other variables that we did not include in the model. For example, the real impact on customer retention should be the communication from customer service about whether the customer subscribes to the data plan or not, and it is the customer's good service experience that increases the retention rate.

Task 4

There is a chance to improve the model performance based on the data at hand. Based on the current data, customers display significant different features on the consumption of telecommunicating services. Take the consumption of roaming service for example. With an average consumption of 10.2 minutes, the standard deviation is approximately 2.8 minutes, relatively high compared to the mean. The dispersion of the data reveals the possible existence of different segments in the current customers. The other example is the use of data plan. In the given data, around a quarter of customers use data plan while the rest do not. Customers display different consumption behavior can make different decisions on whether to churn. Among the customers who do not churn, approximately 30% of them use data plan. They consume around ten minutes of roaming service on average. While among the customers who churn, only 17% of them use data plan but they consume 11 minutes of roaming service on average. Therefore, it is worth investigating the existence of different segments in the customers. This is especially the case given the R output suggesting that the variable DataPlan is a significant explanatory variable on the response variable, which is whether a customer churn. Furthermore, there can be correlation among the explanatory factors. For instance, it is possible that customers who consume more data will consume less roaming service because of their lifestyle preference. The existence of multicollinearity among the explanatory variables is backed up by the magnitude of VIF in the first few models fitted. Therefore, the way customer segmentation is done also worth careful attention. Because of the variance in the current customers in the given data, it can be better if different models are fitted for different segments of customers and use these models for prediction respectively. The prediction performance of models might be improved in such manner.

Task 5

Based on mod1, there are two customer prototypes which are customer A (with Dataplan=0) who does not have a dataplan service, and customer B (with Dataplan=1) who does have a dataplan.

Table below shows the CLV and the probability for both customer types. The result for the cash flow of customer B is 8756.73, whereas customer A is 5632.50 both with an annual interest rate of 4%. From that common CLV, we could know that customer B is more valuable. In addition, with the result from the survival model, the survival probability could make more accurate cash flow in each period(also with discount). And we could get the final lifetime value for customer B which is 3703.90 that is higher than customer A(1946.79).

In conclusion, customer B has a higher common CLV perspective and also has higher probability corrected CLV, therefore, customer B(who has a dataplan) tends to bring more profits to the company.

		Dataplan =0						Dataplan =1		
Time	Cash Flow	Discounted	Survival Probabi	Corrected CF		Time	Cash Flow	Discounted	Survival Probab	Corrected CF
1	11.26500335	11.25418202	0.99999971	11.25417876		1	17.51345737	17.49663368	0.999999846	17.49663099
2	11.26500335	11.24337109	0.999997978	11.24334835		2	17.51345737	17.47982616	0.99999893	17.47980745
3	11.26500335	11.23257054	0.999993706	11.23249984		3	17.51345737	17.46303478	0.99999667	17.46297662
4	11.26500335	11.22178037	0.999985913	11.22162228		4	17.51345737	17.44625953	0.999992547	17.4461295
5	11.26500335	11.21100056	0.999973683	11.21070552		5	17.51345737	17.42950039	0.999986076	17.4292577
6	11.26500335	11.2002311	0.999956147	11.19973994		6	17.51345737	17.41275736	0.999976799	17.41235336
7	11.26500335	11.189472	0.999932471	11.18871638		7	17.51345737	17.3960304	0.999964272	17.39540888
8	11.26500335	11.17872322	0.999901847	11.177626		8	17.51345737	17.37931952	0.99994807	17.37841701
9	11.26500335	11.16798478	0.999863492	11.16646026		9	17.51345737	17.36262469	0.999927777	17.36137071
10	11.26500335	11.15725665	0.999816642	11.15521087		10	17.51345737	17.34594589	0.999902989	17.34426315
11	11.26500335	11.14653882	0.99976055	11.14386978		11	17.51345737	17.32928312	0.99987331	17.32708767
12	11.26500335	11.13583129	0.999694483	11.1324291		12	17.51345737	17.31263636	0.999838352	17.3098378
13	11.26500335	11.12513405	0.999617725	11.12088119		13	17.51345737	17.29600558	0.999797736	17.29250722
14	11.26500335	11.11444708	0.999529569	11.1092185		14	17.51345737	17.27939078	0.999751087	17.27508972
15	11.26500335	11.10377038	0.999429321	11.09743369		15	17.51345737	17.26279194	0.999698037	17.25757922
16	11.26500335	11.09310393	0.999316298	11.08551955		16	17.51345737	17.24620905	0.999638223	17.23996977
17	11.26500335	11.08244773	0.999189826	11.07346902		17	17.51345737	17.22964209	0.999571289	17.22225555
18	11.26500335	11.07180177	0.99904924	11.06127514		18	17.51345737	17.21309104	0.99949688	17.20443079
19	11.26500335	11.06116603	0.998893885	11.04893111		19	17.51345737	17.19655589	0.999414648	17.18648985
20	11.26500335	11.05054051	0.998723113	11.03643022		20	17.51345737	17.18003662	0.999324249	17.16842719
490	11.26500335	7.034125645	4.86E-05	0.000341859		490	17.51345737	10.93580319	0.005222692	0.057114332
491	11.26500335	7.027368559	4.59E-05	0.000322556		491	17.51345737	10.9252981	0.005067884	0.055368143
492	11.26500335	7.020617965	4.33E-05	0.000303993		492	17.51345737	10.91480309	0.00491712	0.053669397
493	11.26500335	7.013873856	4.09E-05	0.000286867		493	17.51345737	10.90431817	0.004770314	0.052017022
494	11.26500335	7.007136225	3.86E-05	0.000270475		494	17.51345737	10.89384332	0.004627377	0.05040992
495	11.26500335	7.000405066	3.65E-05	0.000255515		495	17.51345737	10.88337854	0.004488225	0.048847052
496	11.26500335	6.993680373	3.44E-05	0.000240583		496	17.51345737	10.8729238	0.004352773	0.047327369
497	11.26500335	6.986962141	3.25E-05	0.000227076		497	17.51345737	10.86247911	0.004220938	0.045849851
498	11.26500335	6.980250361	3.06E-05	0.000213596		498	17.51345737	10.85204445	0.004092639	0.0444135
499	11.26500335	6.97354503	2.89E-05	0.000201535		499	17.51345737	10.84161982	0.003967796	0.043017336
500	11.26500335	6.966846139	2.73E-05	0.000190195		500	17.51345737	10.8312052	0.003846331	0.0416604
sum	5632.501675	4470.0835		1946.789978			8756.728683	6949.542257		3703.904765

Task 6
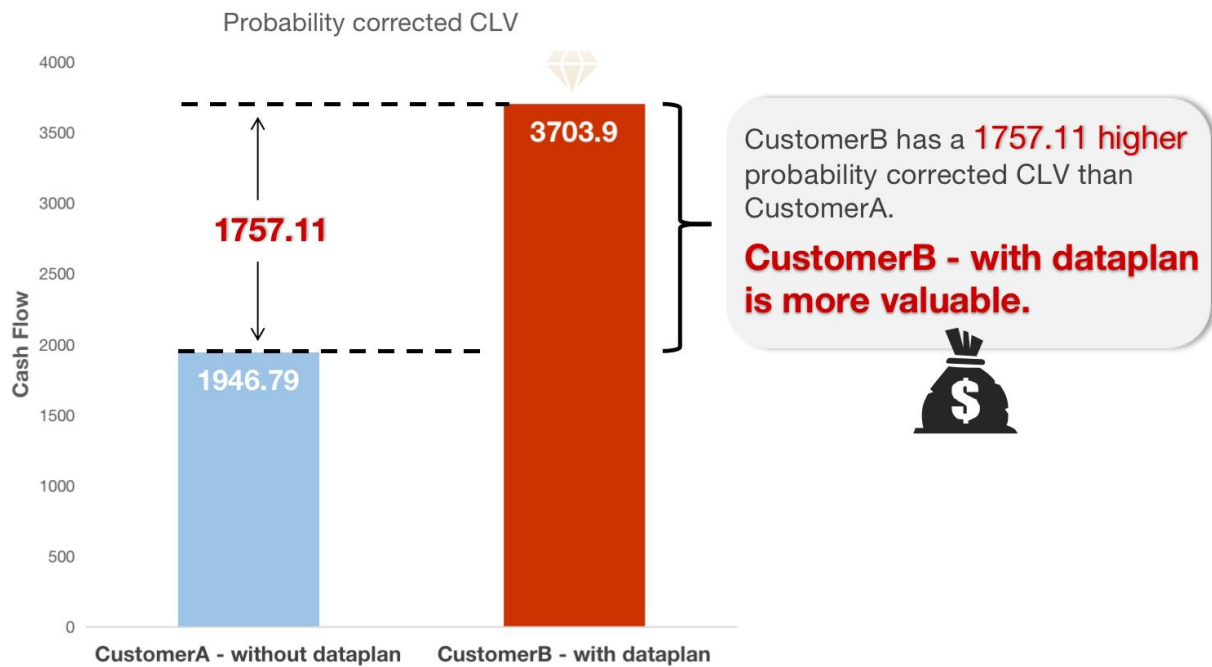
The essential difference between the two customer groups had been calculated and the generated insights indicated that CustomerB with a data plan has a higher probability corrected CLV so they are with higher value.

It is suggested to retain and develop the relationship with CustomerB who have a data plan since they are considered as a more valuable group and they might generate more profit to the company. For instance, offering a special reward or other marketing investment toward this highly profitable group(who has a data plan) to enhance their loyalty and build a stronger relationship. The difference between CustomerB and CustomerA is significant and the company could also consider decreasing the subscription price or offering a discount for attracting CustomerA to engage with a data plan.