The George Washington University
6210,2018 spring

Report of Homework 9

Jiaying Liu

April 18, 2018

1 Part A

1.1 dataset

The data set contain 150 observation of one time response variable and one status variable explanatory variable. And we can see the group variable have 3 level of value.

Table 1: data set

Obs	Time	Status	group
1	80.97	0	1
2	268.49	0	1
3	96.08	0	1
4	8.08	0	1
5	43.06	1	1
...	...	...	...
150	207.14	0	3

1.2 Survivor Function

Before estimate the parameter of survivor. Let's introduct the survivor function. It is defined as the probability that the survival time T which is greater than or equal to t,

$$S(t) = Pr(T \geqslant t)$$

1.2.1 Product limit method

First, we use the product limit to estimate the survivor function parameters. Product limit is a nonparametric estimate method which can estimate the survivor function from censored data. The function can be show like

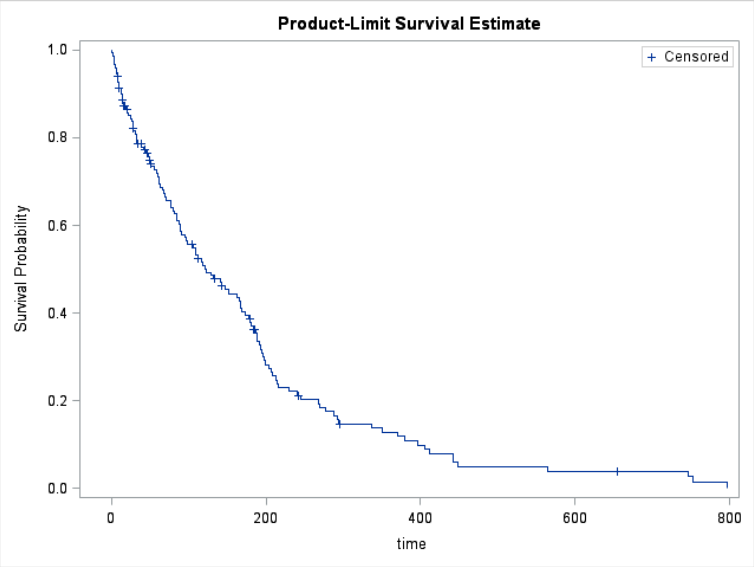
$$\hat{S}(t) = \prod_{t_j \leq t}(1 - \frac{d_j}{n_j})$$

From the table 2, it displaying the estimate result of product limit method.We can see the point estimate is 212.84 when it is in 75 percent and it is 122.26 when it is 50 percent and it is 48.41 when it is 25 percent.And the figure 1 show the fitting plot of product limit method. we can see that the curve is no that smooth. There are another method to fit the function. Let's see what is the performance of that method.

Table 2: Summary statistics for time variable time

Percent	**Point estimate**	**Transform**	**Lower CI**	**Upper CI**
75	212.840	LOGLOG	191.580	276.500
50	122.260	LOGLOG	89.520	166.990
25	48.410	LOGLOG	28.320	67.610

Figure 1: product limit survival limit



1.2.2 Life table method

In this step, we are going to use the life table method to fit the survival function.We are going to shortly introduce the life table method. This method will count the number of censored and uncensored observation and then estimate the survivor function. The formula is

$$\hat{S}(t_i) = \begin{cases} 1 & i = 0 \\ \hat{S}(t_{i-1})p_{i-1} & i > 0 \end{cases} \tag{1}$$

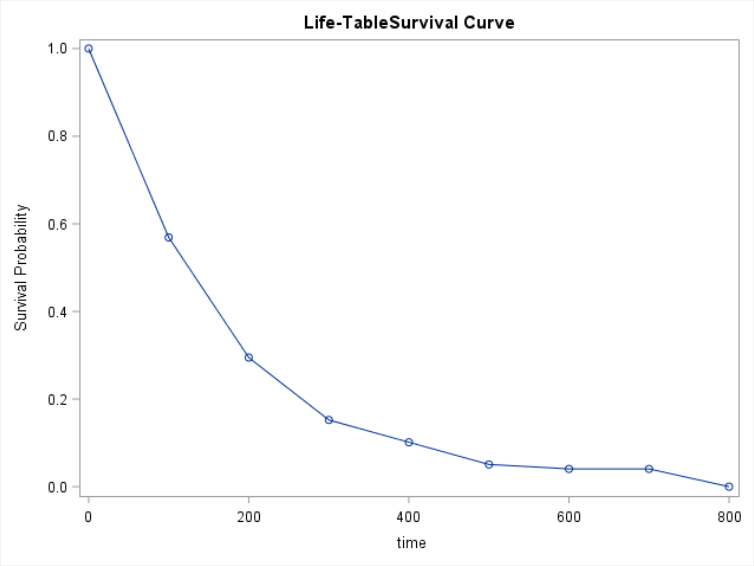
From the table 3, it show the summary statistics result of time table estimation.It calculate the number of survival failure and the number of censored and failed and get the result of survival and failure.And from the figure 2, it display the plot of survival under time table

method. We can see, compared to the product limit method, the curve of time table method goes smoothly.

Table 3: Summary statistics for time table estimate method

lower	upper	number failed	number censored	survival	failure
0	100	61	17	1.0000	0
100	200	33	7	0.5689	0.4311
200	300	15	2	0.2948	0.7052
300	400	5	0	0.1522	0.8478
400	500	5	0	0.1014	0.8986
500	600	1	0	0.0507	0.9493
600	700	0	1	0.0406	0.9594
700	800	3	0	0.0406	0.9594
800	.	0	0	0	1.0000

Figure 2: time table survival function



1.2.3 conclusion

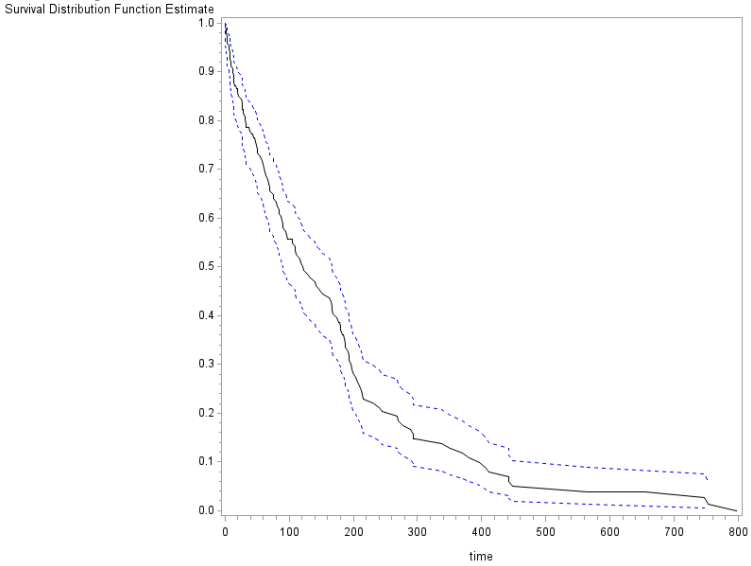
From above analysis, we can answer whether there a clear different between the survivor functions estimated by the product limit and life table method. The answer is not. They have the similar trend unless one is more smoothly.

1.3 Confidence interval plot

1.3.1 Confidence interval for survivor function by product limit method

From the figure 3, it show the confidence interval plot of survivor function by product limit method.

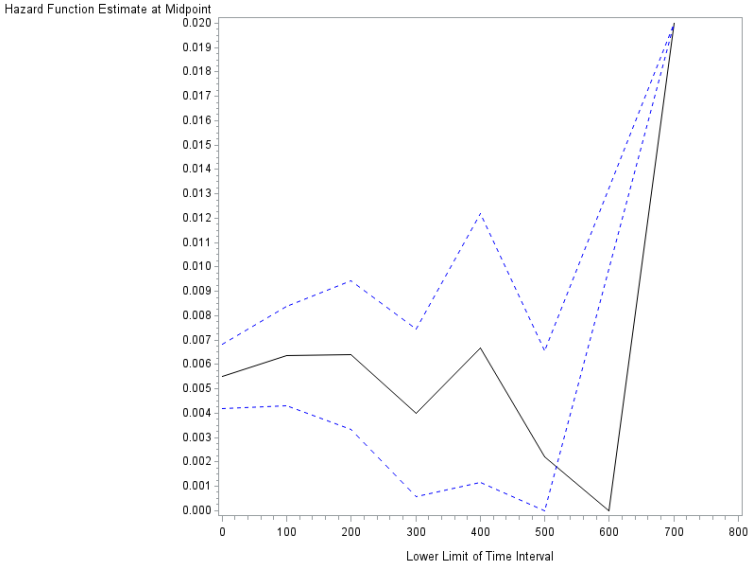
Figure 3: The confidence interval of survival function



1.3.2 Confidence interval for Hazard function

From the figure 4, it show the confidence interval plot of Hazard function.

Figure 4: The confidence interval of survival function



1.3.3 conclusion

From the above figure, we can see the trend of survival function and Hazard are clear.

1.4 survivor function under different group

Now let's consider the survival function under different group and verify the relation between each group.From the following table, it show the result of survival under different group.And we visualize the result thought the plot figure to directly the result of survival result.

Table 4: Group 1

Percent	**Point estimate**	**Transform**	**Lower CI**	**Upper CI**
75	166.990	LOGLOG	111.530	214.340
50	80.970	LOGLOG	42.010	116.030
25	27.530	LOGLOG	13.600	58.300

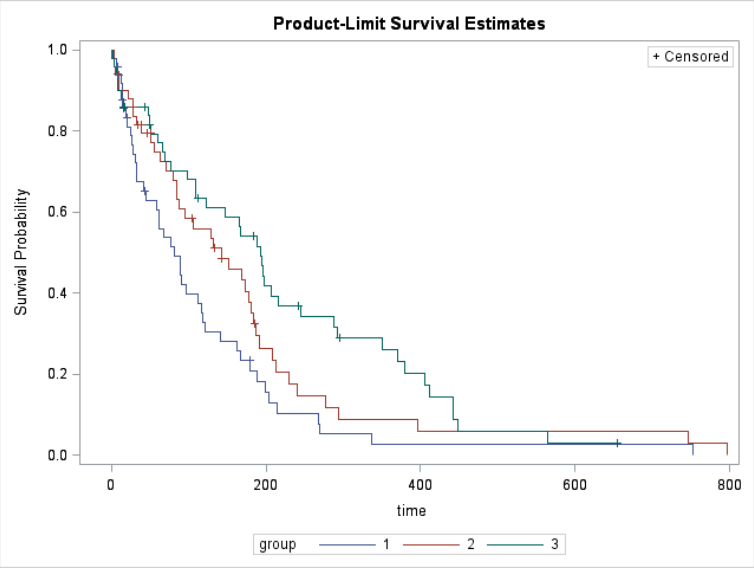
Table 5: Group 2

Percent	**Point estimate**	**Transform**	**Lower CI**	**Upper CI**
75	208.920	LOGLOG	176.930	276.500
50	141.910	LOGLOG	83.700	180.710
25	54.740	LOGLOG	21.730	87.830

Table 6: Group 3

Percent	**Point estimate**	**Transform**	**Lower CI**	**Upper CI**
75	369.960	LOGLOG	207.140	441.740
50	193.530	LOGLOG	109.230	244.490
25	66.680	LOGLOG	14.740	122.260

Figure 5: The confidence interval of survival function



To evaluate the different between estimated survivor function for different group, we can see the table 7 which contain a lot of statistical test. and we can see the all the p value of test is smaller than 0.05 so that the survivor function of different groups are significantly different.

Table 7: Test of Equality over strata

Test	Chi Square	DF	P value
Log-Rank	10.4587	2	0.0054
Wilcoxon	8.7038	2	0.0129
Tarone	10.6033	2	0.0050
Peto	9.5432	2	0.0085
Modified Peto	9.5032	2	0.0086
Fleming(1)	9.5519	2	0.0084

2 B Part

2.1 data set

The data set contains 100 observation and there are one time variable, one status variable and x and z variable.

Table 8: data set

Obs	Time	Status	x	z
1	44.7	1	26.5	8.52
2	31.56	0	65.7	6.73
3	6.31	0	61	5.23

4	1.95	0	67	1.06
5	78.21	1	3.9	8.11
6	75.63	0	98.2	3.36
7	34.22	0	64.4	4.88
8	51.39	1	18.2	5.14
9	0.05	1	1.8	5.19
10	39.31	0	92	5.41
...	...	...	...	

2.2 Cox's regression analysis

In the following step, we are going to run the cox's regression analysis with both x and z. The table 7 show the result of maximum likelihood estimates of variable x and y, and we an see both the p value of x and z are greater than 0.05, so we have the reason to believe there are not significant variable associated with the survival time.

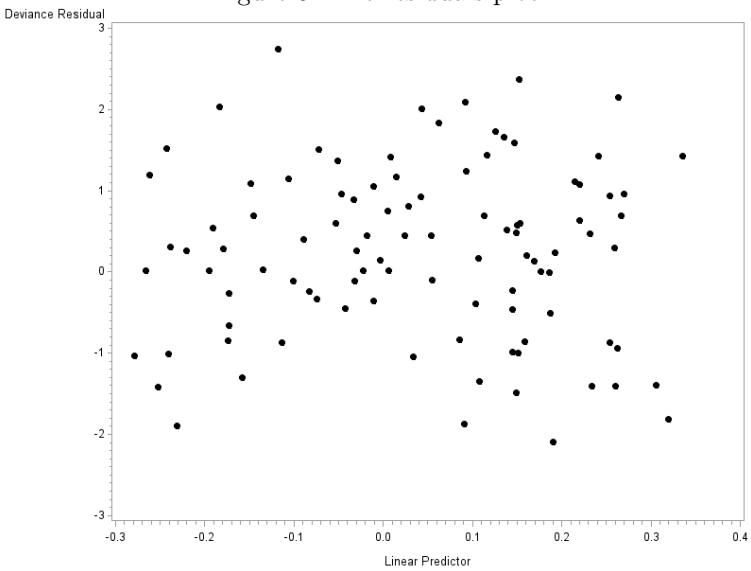
Table 9: Analysis of maximum likelihood estimates

parameter	**DF**	**parameter estimate**	**SE**	**CHI square**	**p value**	**Hazard ratio**
x	1	-0.00339	0.00437	0.6030	0.4375	0.997
z	1	0.04109	0.03787	1.1776	0.2778	1.042

2.3 Model Diagnostics

To verify the assumption of the model, we need to plot the residual plot and check the dispersion of the plot. From the figure 6, it show the result of residual plot, and find that the plot distributed randomly between y=0. So we can conclude that the model is appropriate.

Figure 6: The residuals plot



3 Part c

3.1 data set

The data set with 45 subject and a time dependent variable .start and stop is a time interval variable, the status variable contain 0 and 1 which is for censored variable. X variable is a fixed variable and the z is a time varying variable. There are 428 observation.

Table 10: data set

Obs	ID	start	stop	x	z	Status
1	1	0	27	1	0	0
2	1	27	34	1	5	0
3	1	34	37	1	6	0
4	1	37	41	1	8	0
5	1	41	43	1	10	0
6	1	43	45	1	10	0
7	1	45	46	1	10	0
8	1	46	47	1	10	1
9	2	0	27	1	0	0
10	2	27	34	1	0	0
...	...	...	...			

3.2 Cox's regression with time dependent covariate

In this step, we are going to use the cox's regression to analyze the data. When both x and z included, we can see the performance of the model from the table 11.Consider the p

value of variable x and z, we can say that since the p value of z is smaller than 0.05, the variable z is significant. otherwise, because the p value of x variable is greater than 0.05, it is no a significant variable.We can see the confidence interval of Hazard ration of x variable is (0.855,1.181) and the CI of z variable is (1.056,1.180).

Table 11: Analysis of maximum likelihood estimates

par	DF	par estimate	SE	CHI square	p value	Hazard ratio	Lower CI	Upper CI
x	1	0.06013	0.05429	1.2266	0.2681	1.062	0.955	1.181
z	1	0.11014	0.02837	15.0723	0.0001	1.116	1.056	1.180

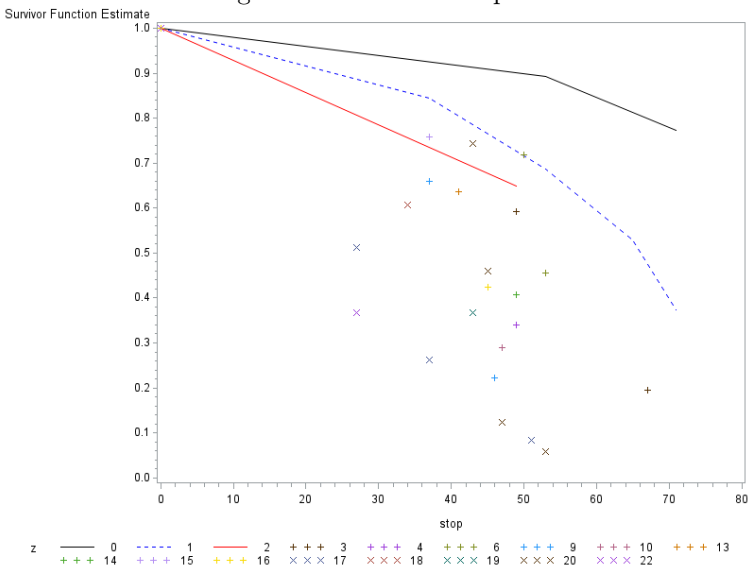
3.3 Cox's regression with stratified analysis(based on x)

With the variable z is the time dependent covariate, from the table 11 we can see the result.The variable of x become significant when the variable z is the time dependent covariate. And we can also conclude the confidence interval of Hazard ration is (0.901,1.282)

Table 12: Analysis of maximum likelihood estimates

par	DF	par estimate	SE	CHI square	p value	Hazard ratio	Lower CI	Upper CI
x	1	0.07183	0.08997	0.6376	0.4246	1.074	0.901	1.282

Figure 7: The residuals plot



From the figure 7, we can conclude that the three function are visually different.