

HW 9 Report

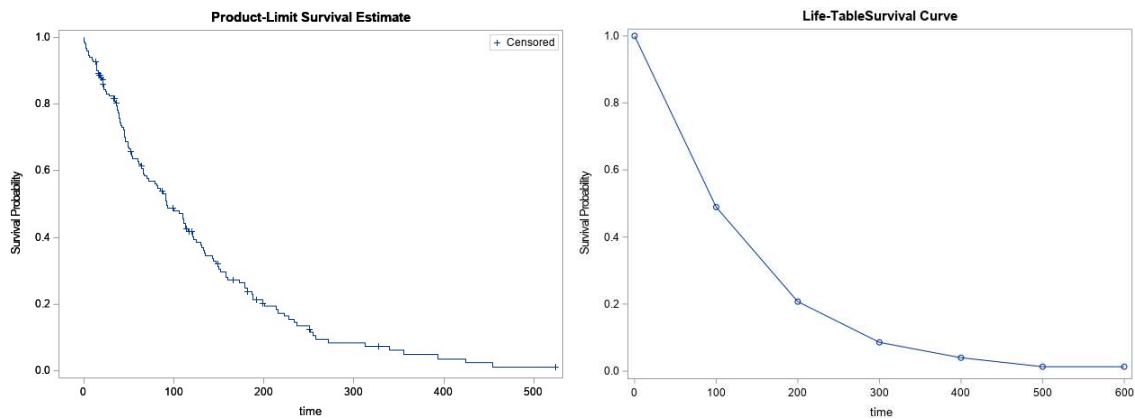
Yongchao Qiao

Part A

I Without the consideration of group information:

1. Survivor functions estimated by different methods

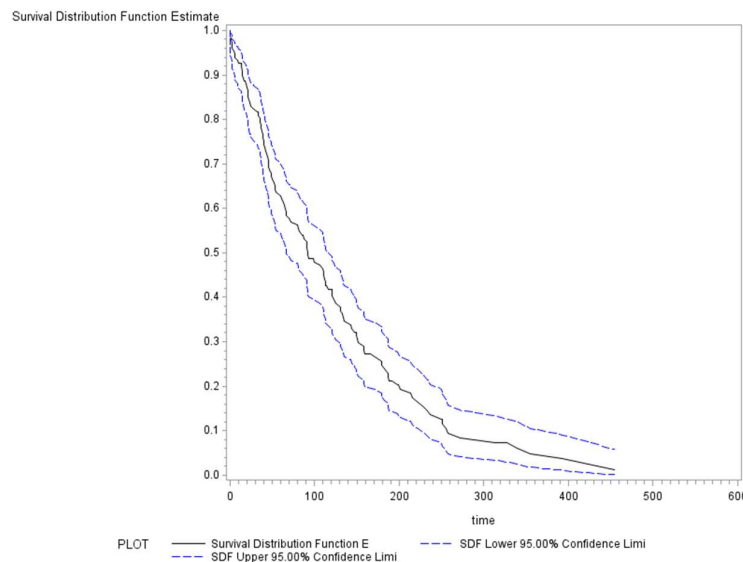
Figure 1: Survivor functions estimated by different methods



From the figure 1 we can see that without consideration of group information, the survivor functions estimated by product-limit and life-table methods are clear different, since the survivor functions generated by product-limit method can provide much more details than that generated by life-table method, although the general trends of these two plots are the same.

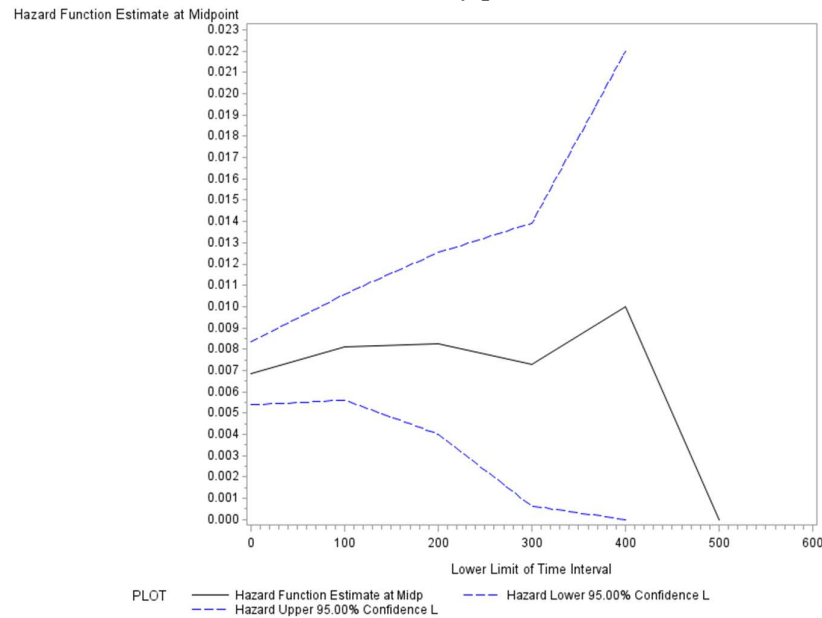
2. The survivor function estimated by the product-limit method with confidence intervals

Figure 2: The survivor function estimated by product-limit method with 95% CI



3. The hazard function estimated by the life-table method with confidence intervals

Figure 3: The survivor function estimated by product-limit method with 95% CI



From the figure 2 and figure 3, we can see that there is a clear downward trend for each survival function while there is not a clear trend for the hazard function since for the hazard function, it is going up until time=400 and then going down after time=400.

II With the consideration of group information:

Compare the survivor functions of different groups estimated by the product-limit method

Figure 4: The survivor functions of different groups estimated by the product-limit method

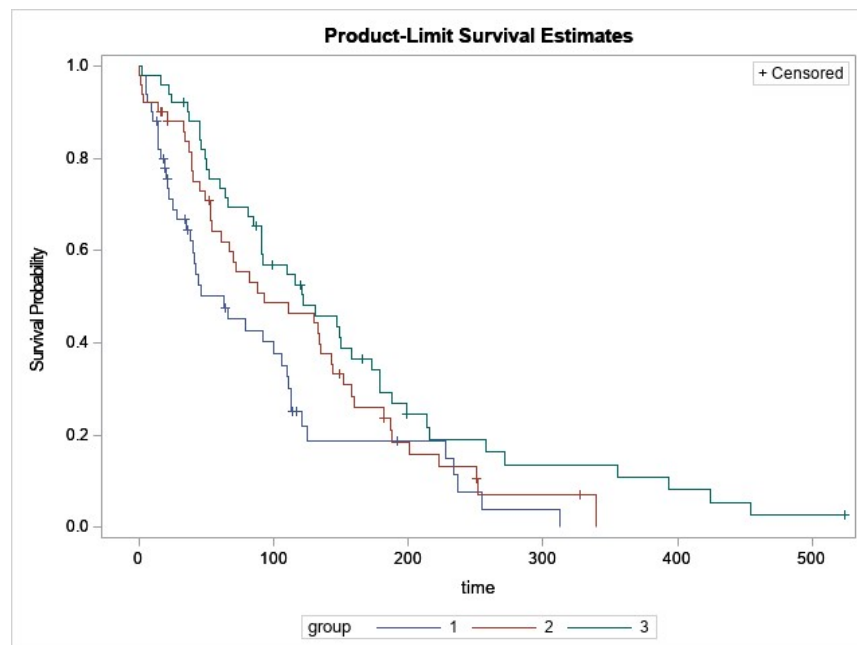


Table 1: Test of Equality over Strata

Test	Chi-Square	DF	Pr > ChiSq
Log-Rank	8.8729	2	0.0118
Wilcoxon	10.9545	2	0.0043
Tarone	10.0042	2	0.0067
Peto	10.5961	2	0.0050
Modified Peto	10.5926	2	0.0050
Fleming(1)	10.5964	2	0.0050

From table 1 we can see, for the Test of Equality over Strata, the P-Values of all six tests are less than 0.05 which means the estimated survivor functions for different group are significant different. Besides, we can also visually find the difference between different survivor functions in the figure 4.

III Conclusion

From the figure 1 we can see that without consideration of group information, the survivor functions estimated by product-limit and life-table methods are clear different, since the survivor functions generated by product-limit method can provide much more details than that generated by life-table method, although the general trends of these two plots are the same.

From the figure 2 and figure 3, we can see that there is a clear downward trend for each survival function while there is not a clear trend for the hazard function since for the hazard function, it is going up until time=400 and then going down after time=400.

From table 1 we can see, for the Test of Equality over Strata, the P-Values of all six tests are less than 0.05 which means the estimated survivor functions for different group are significant different. Besides, we can also visually find the difference between different survivor functions in the figure 4.

Part B

I Run the Cox's regression analysis with both X and Z included

Table 2: Testing Global Null Hypothesis: BETA=0

Test	Chi-Square	DF	Pr > ChiSq
Likelihood Ratio	0.3077	2	0.8574

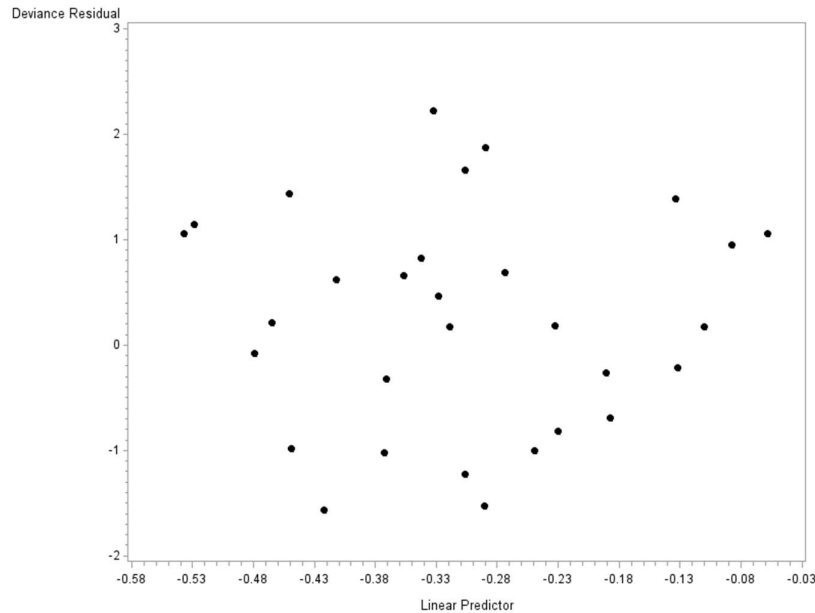
Table 3: Analysis of Maximum Likelihood Estimates

Parameter	DF	Estimate	Standard		Pr > ChiSq	Hazard Ratio	95% Hazard Ration CI
			Error	Chi-Square			
X	1	-0.0317	0.0967	0.1072	0.7434	0.969	0.802 1.171
Z	1	-0.0028	0.0066	0.1775	0.6735	0.997	0.984 1.010

From table 2 and table 3 we can see, for the Likelihood Ratio Test, the P-Value of the model is greater than 0.05 which means the model is not significant. Besides, the P-values of variable X and variable Z are both greater than 0.05, which means variable X and variable Z are not significant to explain the change of survival time.

II Plot of the deviance residuals vs. linear predictor scores

Figure 5: Plot of the deviance residuals vs. linear predictor scores



From the figure 5 we can see that there is no special pattern for the deviance residuals vs. linear predictor scores.

III Conclusion

Based on previous analysis, we can see that there are no explanatory variables significantly associated with the survival time and the model is not appropriate since the P-Value of the Likelihood Ratio Test is greater than 0.05 which means the model is not significant.

Part C

I Use Cox's regression to analyze the dataset with a time-dependent covariate

Table 4: Testing Global Null Hypothesis: BETA=0

Test	Chi-Square	DF	Pr > ChiSq
Likelihood Ratio	21.1335	2	<0.0001

Table 5: Analysis of Maximum Likelihood Estimates

Parameter	DF	Estimate	Standard Error	Chi-Square	Pr > ChiSq	Hazard Ratio	95% Ration CI	Hazard
X	1	0.01973	0.06065	0.1058	0.7450	1.020	0.906 1.149	
Z	1	0.12296	0.02982	17.0084	<0.0001	1.131	1.067 1.199	

From table 4 and table 5 we can see, for the Likelihood Ratio Test, the P-Value of the model is less than 0.05 which means the model is significant. Besides, the P-value of variable X is greater than 0.05, which means variable X is not significant to explain the change of survival time. However, the P-value of variable Z is less than 0.05, which means variable Z is significant to explain the change of survival time. The Hazard ratio of variable X is 1.020 with

confidence interval as (0.906, 1.149) and the Hazard ratio of variable Z is 1.131 with confidence interval as (1.067, 1.199).

II Use Cox's regression to perform a stratified analysis (based on X)

Table 6: Testing Global Null Hypothesis: BETA=0

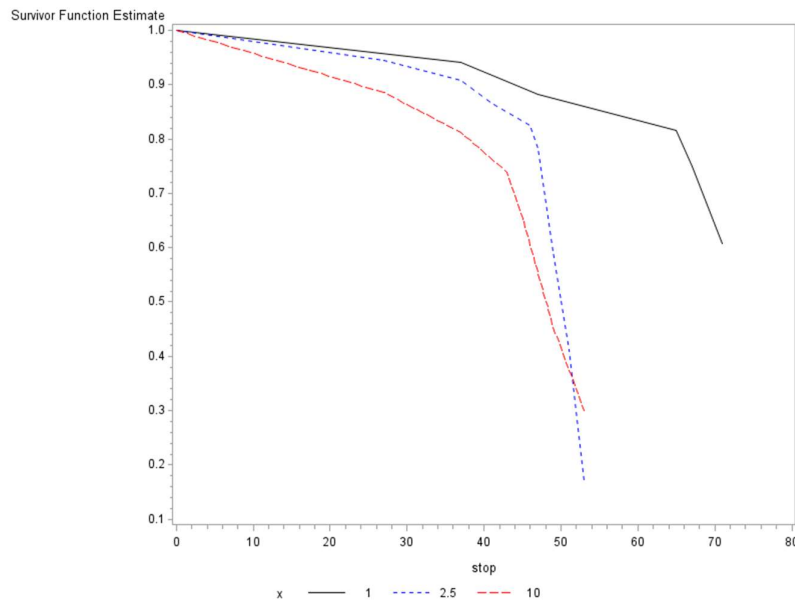
Test	Chi-Square	DF	Pr > ChiSq
Likelihood Ratio	10.0398	1	0.0015

Table 7: Analysis of Maximum Likelihood Estimates

Parameter	DF	Estimate	Standard Error	Chi-Square	Pr > ChiSq	Hazard Ratio	95% Hazard Ration CI
Z	1	0.11222	0.03653	9.4395	0.0021	1.119	1.041 1.202

From table 4 and table 5 we can see, for the Likelihood Ratio Test, the P-Value of the model is less than 0.05 which means the model is significant. Besides, the P-value of variable Z is less than 0.05, which means variable Z is significant to explain the change of survival time. The Hazard ratio of variable Z is 1.119 with confidence interval as (1.041, 1.202).

Figure 6: Three estimated baseline survivor functions in the stratified analysis



From the figure 6, we can see these three estimated baseline survivor functions are visually different.

III. Conclusion

Based on previous analysis, we can see that the explanatory variable Z is significantly associated with the survival time and in the stratified analysis, these three estimated baseline survivor functions are visually different.

Appendix:

/* Part A */

```
proc import datafile = 'E:/GW/Textbook/Data Analysis/HW9/HW9a.csv' /*read the file into  
sas*/
```

```
dbms = csv /*specify the format of the file*/
```

```
out=work.HW9a; /*specify the saved dataset in sas*/
```

```
getnames=yes; /*get the name of the variables from the original file*/
```

```
run; /*run this procedure*/
```

/* Sort the dataset by time */

```
proc sort data=HW9a;
```

```
by time;
```

```
run;
```

/* Estimate the survivor function by product-limit method */

```
proc lifetest data=HW9a plots=(s)outsurv=temp;
```

```
time time*status(1);
```

```
run;
```

/* Estimate the survivor function by life-table method */

```
proc lifetest data=HW9a method=life plots=(s) outsurv=templ;
```

```
time time*status(1);
```

```
run;
```

/* Plot the Survivor function and 95% CI generated by product-limit method */

```
goptions reset=all;
```

```
symbol1 i=join l = 1 v=none c=black;
```

```
symbol2 i=join l = 3 v=none c=blue;
```

```
symbol3 i=join l = 3 v=none c=blue;
```

```
proc gplot data=temp;
```

```
plot (survival sdf_lcl sdf_ucl)*time / overlay legend;
```

```
run;
```

/* Hazard function by life-table method */

```
proc lifetest data=HW9a method=life plots=(h) outsurv=templh;
```

```
time time*status(1);
```

```
run;
```

/* Plot the Hazard function and 95% CI generated by life-table method */

```
goptions reset=all;
```

```
symbol1 i=join v=none l = 1 c=black;
```

```
symbol2 i=join v=none l = 3 c=blue;
```

```
symbol3 i=join v=none l = 3 c=blue;
```

```
proc gplot data=templh;
```

```
plot (hazard haz_lcl haz_ucl)*time / overlay legend;  
run;
```

```
/* Strata */
```

```
goptions reset=all;  
symbol1 i=join v=none c=black;  
symbol2 i=join v=none c=blue;  
symbol3 i=join v=none c=red;
```

```
proc lifetest data=HW9a plots=(s);  
    time time*status(1);  
    strata group / test=(all);  
run;
```

```
/* Part B */
```

```
proc import datafile = 'E:/GW/Textbook/Data Analysis/HW9/HW9b.csv' /*read the file into  
sas*/  
dbms = csv /*specify the format of the file*/  
out=work.HW9b; /*specify the saved dataset in sas*/  
getnames=yes; /*get the name of the variables from the original file*/  
run; /*run this procedure*/
```

```
/* Proportional hazard model */
```

```
proc phreg data=HW9b;  
    model time*status(1)=X Z / rl;  
    output out=Outp xbeta=xb resdev=dev;  
run;
```

```
/* Plot of deviance residuals vs. linear predictor scores */
```

```
goptions reset=all;  
symbol i=none v=dot c=black;  
proc gplot data=Outp;  
    plot dev*xb;  
run;
```

```
/* Part C */
```

```
proc import datafile = 'E:/GW/Textbook/Data Analysis/HW9/HW9c.csv' /*read the file into  
sas*/  
dbms = csv /*specify the format of the file*/  
out=work.HW9c; /*specify the saved dataset in sas*/  
getnames=yes; /*get the name of the variables from the original file*/  
run; /*run this procedure*/
```

```

/* Cox's regression */
proc phreg data=HW9c;
    model (start,stop)*status(0)=X Z / rl;
run;

/* Straified analysis (Based on X) */
proc phreg data=HW9c;
    model (start,stop)*status(0) = Z / rl;
    strata X;
    baseline out=phout survival=sfest;
run;

symbol1 v=none i=join c=black r=1;
symbol2 v=none i=join l=2 c=blue r=1;
symbol3 v=none i=join l=3 c=RED r=1;
proc gplot data=phout;
    plot sfest*stop=X;
run;

```