P8110: Applied Regression II
Homework #3
[10 points]

NOTE: Please do not hand in computer output seperately from responses to the questions below. Instead, cut and paste relevant SAS output. Attach SAS codes at the end of your homework.

The "HW3Data" gives the time until staphylococcus infection (in days) for 154 patients with a burn wound. The purpose of this study was to compare a routine bathing care treatment with a body-cleansing method. The time until staphylococcus infection was recorded, along with an indicator variable – whether or not an infection had occured. The severity of the burn was measured by percentage of total surface area of body burned.

ID	=	Patient ID
treatment	=	0-routine bathing & 1-body cleansing
percentBurned	=	Percentage of total surface area burned
timeToInfection	=	Time to straphylocous aureaus infection or on study time
Infection	=	1-infection occured & 0-infection did not occur

1. **Use SAS to compute the Kaplan-Meier estimator and 95% CI of the survival function for the routine bathing group (only keep the K-M estimates for unique event time points). Generate a graph of the survival functions in the two treatment groups. Interpret the graph. (hint: What do you observe about the difference of the survival functions between the two groups?) [2 points]**

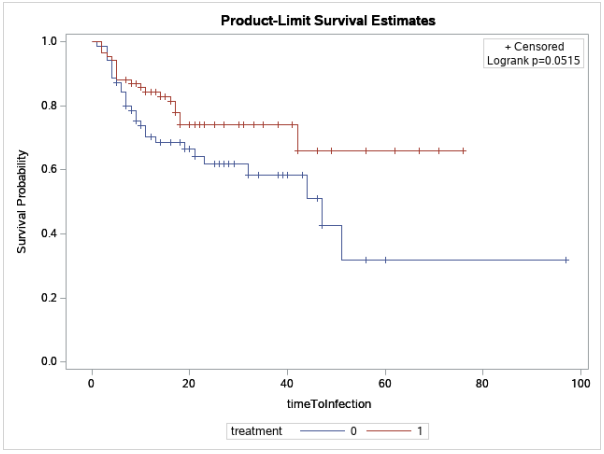

Figure 1: The survival functions for two treatment groups

SAS output:

Obs	treatment	timeToInfection	SURVIVAL	SDF_STDERR	SDF_LCL	SDF_UCL
1	0	0	1.00000	0.00000	1.00000	1.00000
2	0	1	0.98571	0.01418	0.90290	0.99798
3	0	3	0.94286	0.02774	0.85488	0.97816
4	0	4	0.88571	0.03803	0.78445	0.94114
5	0	5	0.87143	0.04001	0.76749	0.93093
6	0	6	0.84238	0.04363	0.73345	0.90946
7	0	7	0.79881	0.04808	0.68400	0.87559
8	0	8	0.78374	0.04948	0.66704	0.86359
9	0	9	0.75239	0.05223	0.63194	0.83832
10	0	10	0.73638	0.05352	0.61422	0.82520
11	0	11	0.70291	0.05607	0.57735	0.79752
12	0	13	0.68534	0.05736	0.55808	0.78288
13	0	19	0.66457	0.05926	0.53434	0.76612
14	0	21	0.64242	0.06129	0.50905	0.74825
15	0	23	0.61862	0.06347	0.48192	0.72908
16	0	32	0.58425	0.06862	0.43834	0.70456
17	0	44	0.51122	0.09095	0.32361	0.67098
18	0	47	0.42602	0.10860	0.21600	0.62182
19	0	51	0.31951	0.12305	0.10959	0.55501

Figure 2: The Kaplan-Meier estimator for the routine bathing group

Before around 8 days, the two survival curves are virtually indistinguishable, with little visual evidence of treatment effect. After around 8 days, there exists an obvious gap between the two survival curves.

2. Test whether the survival curves for the two treatment groups are the same using the log-rank test at $\alpha = 0.05$. Give the null and alternative hypothesis, test statistic, degrees of freedom, p-value, and conclusion. Does the Wilcoxon test lead to the same conclusion? Briefly explain why the test statistic of the Wilcoxon test is smaller than that of the log-rank test in this application. [5 points]

Let $S_1(t)$ and $S_2(t)$ denote the corresponding survival functions of the routine bathing and body cleansing treatment groups, respectively.

$$H_0 : S_1(t) = S_2(t) \text{ for } \textbf{all t} \quad \text{v.s.} \quad H_\alpha : S_1(t) \neq S_2(t) \text{ for } \textbf{some t}$$

Log-rank test: *The test statistics is*

$$Q_{log\text{-}rank} = 3.7924 \textit{ with 1 degree of freedom}$$

Therefore

$$P\text{-}value = Pr(\mathcal{X}_1^2 \geq 3.7924) = 0.0515$$

Given that p-value $= 0.0515 > 0.05$*, we fail to reject* H_0 *under the level of 0.05 though, we may conclude that there is a marginally significant difference between the survival functions of the two treatment groups.*

Wilcoxon test: *The test statistics is*

$$Q_{Wilcoxon} = 2.8639 \text{ with 1 degree of freedom}$$

Therefore

$$P\text{-value} = Pr(\mathcal{X}_1^2 \geq 2.8639) = 0.0906$$

Since p-value $=0.0906 > 0.05$*, we lead to the same conclusion as the log-rank test.*

SAS output:

Test of Equality over Strata			
Test	**Chi-Square**	**DF**	**Pr > Chi-Square**
Log-Rank	3.7924	1	0.0515
Wilcoxon	2.8639	1	0.0906
-2Log(LR)	3.4994	1	0.0614

Because log-rank test is more sensitive to the difference between groups in later time points and Wilcoxon test is more sensitive to the difference between groups in early time points. In our case, the difference in later time points is much greater than the early points. Therefore the log-rank test has a smaller p-value than Wilcoxon test. (Or log-rank test has a larger test statistic than the Wilcoxon test.)

3. **We can categorize the percentage of total surface area burned into four groups using the sample quartiles** ($< 12\%, [12\%, 20\%), [20\%, 30\%),$ **and** $\geq 30\%$)**. Test whether the survival functions are the same among the four groups using the generalized log-rank test at** $\alpha = 0.05$**. Give the null and alternative hypothesis, test statistic, degrees of freedom, p-value, and conclusion. [3 points]**

$$H_0 \ : \ S_{<12\%}(t) = S_{[12\%,20\%)}(t) = S_{[20\%,30\%)}(t) = S_{\geq 30\%}(t) \quad \text{for all } t$$
$$H_\alpha \ : \ \text{At least one of } S(t)\text{'s is different from the others for some } t$$

The test statistics is

$$Q_{log\text{-}rank} = 5.4722 \text{ with 3 degree of freedom}$$

Therefore

$$P\text{-}value = Pr(\mathcal{X}_3^2 \geq 5.4722) = 0.1403$$

Since p-value $=0.1403 > 0.05$*, we fail to reject* H_0 *at* $\alpha = 0.05$*, and conclude that there is no statistically significant difference among the survival functions of the four groups.*

SAS output:

Test of Equality over Strata			
Test	Chi-Square	DF	Pr > Chi-Square
Log-Rank	5.4722	3	0.1403
Wilcoxon	4.3055	3	0.2303
-2Log(LR)	3.8573	3	0.2773

SAS code:

```
data data1;
infile 'HW3Data.csv' delimiter = ',' MISSOVER DSD;
input ID treatment percentBurned timeToInfection Infection;
run;

ods graphics on;
proc lifetest data=data1 method=KM alpha=0.05 conftype=loglog plots=survival(test) outsurv=A
stderr;
time timeToInfection*Infection(0);
strata treatment;
run;



data A ;
     set A ;
     if _censor_ = 1 then delete ;
run;

proc print data = A ;
where treatment = 0;
var treatment timeToInfection SURVIVAL SDF_STDERR SDF_LCL SDF_UCL;
run;

ods graphics off;
```

```
ods graphics on;
proc lifetest data=data1 method=KM alpha=0.05 conftype=loglog plots=survival(test) outsurv=B
stderr;
time timeToInfection*Infection(0);
strata percentBurned (12 20 30);
run;
ods graphics off;
```