Survival Analysis HW3

Jiyang Wen, 2nd year Biostatistics PhD student

October 11, 2019

1 Problem 1

(a) We define

$$z = \begin{cases} 0 & \text{placebo group} \\ 1 & \text{6MP group} \end{cases}$$

Then the proportional hazard model is

$$\lambda(t) = \lambda_0(t)\exp\{z\beta\}$$

which is

$$\lambda_1(t) = \lambda_0(t)e^{\beta}$$

Interpretation: Conditioning on surviving to time t, the probability that patients who receive 6MP treatment die at time t is e^{β} of the probability that patients who receive placebo die at time t.
Possible constraint: The constant proportionality of the model is not automatic and needs to be confirmed by proper methods.
(b) From HW2, we can have the KM estimator for both placebo group data and 6MP group data as below.

Table 1: K-M estimator for placebo group data

Observed failure time	1	2	3	4	5	8
$\hat{S}_0(t)$	1	$\frac{19}{21}$	$\frac{17}{21}$	$\frac{16}{21}$	$\frac{14}{21}$	$\frac{12}{21}$
Observed failure time	11	12	15	17	22	23
$\hat{S}_0(t)$	$\frac{8}{21}$	$\frac{6}{21}$	$\frac{4}{21}$	$\frac{3}{21}$	$\frac{2}{21}$	$\frac{1}{21}$

Table 2: K-M estimator for 6MP group data

uncensored times	6	7	10	13	16	22	23
$\hat{S}_1(t)$	1	0.8571	0.8067	0.7529	0.6901	0.6275	0.5378

To verify the validity of proportional hazards model, we plot the $\log(S_1(t))/\log(S_0(t))$ versus time to see if it is almost a constant. The plot is shown in Figure 1.
From Figure 1, we can see that except time before 6 (since during this time 6MP group has no failures), the ratio is almost a constant around 0.2, therefore proportional hazards assumption is valid in this case.

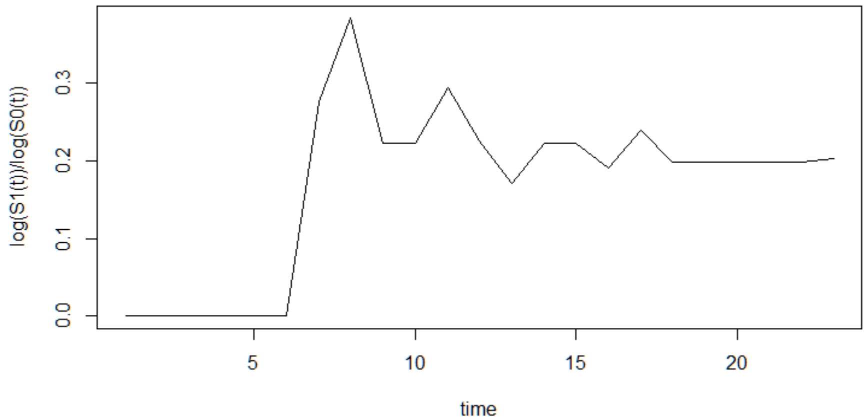

Figure 1: log survival function ratio plot

2 Problem 2

(a) $S(t) = 1 - F(t) = 1 - \int_0^t f_0(s)ds = \exp\{-\lambda t\}$. Therefore $\lambda(t) = \frac{f_0(t)}{S(t)} = \lambda$.
(b) For random variable cT_0, we can have that

$$P(cT_0 \geq t) = P(T_0 \geq \frac{t}{c}) = \exp\{-\lambda \frac{t}{c}\}$$

Therefore for cT_0, survival function is $S_1(t) = \exp\{-\frac{\lambda t}{c}\}$, density function is $f_1(t) = \frac{d}{dt}(1 - S(t)) = \frac{\lambda}{c}\exp\{-\frac{\lambda t}{c}\}$, hazard function is $\lambda_1(t) = \frac{f_1(t)}{S_1(t)} = \frac{\lambda}{c}$.
Interpretation: If one's survival time is c times the other's survival time, then his corresponding hazard is 1/c times the other's.
(c) For Model I, we can have that

$$S(t) = \exp\{-\int_0^t h(s,x)ds\} = \exp\{-\lambda e^{\beta x} t\}$$

For Model II, we can have that

$$S(t) = P(T \geq t) = P(T_0 \geq te^{-\alpha x}) = \exp\{-\lambda e^{-\alpha x} t\}$$

Compare these two results, we can see that $\beta + \alpha = 0$.
(d) Model I: If we change the covariate value of x by one unit, the corresponding hazard will be changed by the proportion e^{β}.
Model II: if we increase the covariate value of x by one unit for the ith subject, the expected value of $\log(T_i)$ is changed by α.

3 Problem 3

(a) 1, 2, 4 models all imply constant relative hazard between the treatment and control groups because the exponential part do not involve time.
(b) 5, 6, 7, 8 model can allow possible cross overs in hazard functions.
For model 1, 2, 3, 4, the hazard ratio between two groups is a constant at any time t, therefore there will not be cross overs.
For model 5, the hazard can cross over before and after time 5. (For example, $\beta = 1, \gamma = -2$)
For model 6, if $\beta_0 < 0$, then the hazard function will cross over at time $t = -\beta_0/\beta_1$.
For model 7, if $f(t) = \beta_0 + \beta_1 t + \beta_2 t^2 = 0$ has one root $t > 0$, then the hazard functions can have cross overs.
For model 8, the argument is the same as that in model 7.
(c) 3, 8 models enforce the hazard functions of two groups start with the same value at $t = 0$, because only these two models have the same value at $t = 0$ for $x = 0, 1$.
(d) 3, 5, 6, 7, 8 models satisfy these conditions.
In order to allow for the possibility that the hazard functions start with the same value at $t = 0$ and gradually become different, we must have:
For model 1, β must equal to 0, but then the two groups' hazard functions are the same, therefore it does not satisfy the conditions.
For model 2, since $\beta < 0$, there is no possibility that the hazard functions start with the same value at $t = 0$.
For model 3, it automatically satisfies the hazard functions start with the same value at $t = 0$, and with t growing, the two functions should be different!
For model 4, since the hazard functions start with the same value at $t = 0$, it must be $\beta = \gamma$, then the hazard functions for the two groups are the same. Therefore it does not satisfy the conditions.
For model 5, since the hazard functions start with the same value at $t = 0$, it must be $\beta = 0$, then if $\gamma \neq 0$, when $t \geq 5$, the hazards functions can be different.
For model 6, since the hazard functions start with the same value at $t = 0$, it must be $\beta_0 = 0$, then as t growing the hazard functions can be different.
For model 7 and model 8, the argument is the same as in model 6, they all satisfy the conditions.

4 Problem 4

There are six uncersored observations in this dataset. First we calculate the corresponding six 2×2 tables.

Table 3: $y_{(1)} = 2$

	D	$\bar{D}$	
A	1	6	7
B	0	6	6
	1	12	13

Table 4: $y_{(2)} = 3$

	D	$\bar{D}$	
A	1	5	6
B	0	6	6
	1	11	12

Table 5: $y_{(3)} = 6$

	D	$\bar{D}$	
A	1	3	4
B	0	6	6
	1	9	10

Table 6: $y_{(4)} = 12$

	D	$\bar{D}$	
A	0	3	3
B	1	4	5
	1	7	8

Table 7: $y_{(5)} = 20$

	D	$\bar{D}$	
A	1	0	1
B	0	3	3
	1	3	4

Table 8: $y_{(6)} = 25$

	D	$\bar{D}$	
A	0	0	0
B	1	2	3
	1	2	3

From these tables, we can calculate that:
For $y_{(1)} = 2$, $E_{(i)} = 1$, $E_0[E_{(i)}] = \frac{7}{13}$, $Var(E_{(i)}) = \frac{1\times12\times7\times6}{13^2\times12} = 0.248$.
For $y_{(2)} = 3$, $E_{(i)} = 1$, $E_0[E_{(i)}] = \frac{1}{2}$, $Var(E_{(i)}) = \frac{1\times11\times6\times6}{12^2\times11} = 0.25$.
For $y_{(3)} = 6$, $E_{(i)} = 1$, $E_0[E_{(i)}] = \frac{2}{5}$, $Var(E_{(i)}) = \frac{1\times9\times4\times6}{10^2\times9} = 0.24$.
For $y_{(4)} = 12$, $E_{(i)} = 0$, $E_0[E_{(i)}] = \frac{3}{8}$, $Var(E_{(i)}) = \frac{1\times7\times3\times5}{8^2\times7} = 0.234$.
For $y_{(5)} = 20$, $E_{(i)} = 1$, $E_0[E_{(i)}] = \frac{1}{4}$, $Var(E_{(i)}) = \frac{1\times3\times1\times3}{4^2\times3} = 0.1875$.
For $y_{(6)} = 25$, $E_{(i)} = 0$, $E_0[E_{(i)}] = 0$, $Var(E_{(i)}) = 0$.

(a) log rank test statistic is

$$z = \frac{\sum_{i=1}^{6}(E_{(i)} - E_0[E_{(i)}])}{\sqrt{\sum_{i=1}^{6} Var(E_{(i)})}} = \frac{1.9365}{1.077} = 1.798$$

The p-value is $2 * (1 - pnorm(1.798, 0, 1)) = 0.07 > 0.05$, therefore we accept the null hypothesis, there is no statistical difference between these two groups.
(b) Gehan test statistic is

$$z = \frac{\sum_{i=1}^{6} N_{(i)}(E_{(i)} - E_0[E_{(i)}])}{\sqrt{\sum_{i=1}^{6} N_{(i)}^2 Var(E_{(i)})}} = 1.6433$$

The p-value is $2 * (1 - pnorm(1.6433, 0, 1)) = 0.10 > 0.05$, therefore we accept the null hypothesis, there is no statistical difference between these two groups.
(c) Taron-Ware test statistic is

$$z = \frac{\sum_{i=1}^{6} \sqrt{N_{(i)}}(E_{(i)} - E_0[E_{(i)}])}{\sqrt{\sum_{i=1}^{6} N_{(i)} Var(E_{(i)})}} = 1.7090$$

The p-value is $2 * (1 - pnorm(1.7090, 0, 1)) = 0.0875 > 0.05$, therefore we accept the null hypothesis, there is no statistical difference between these two groups.
(d) The sample size is too small, therefore the three tests' result are not very convincing. Normal approximation requires larger sample size.

5 Problem 5

The answer is (b).
If two hazard functions $\lambda_1(t), \lambda_2(t)$ do not cross over at any time $t > 0$, then it must be $\lambda_1(t) > \lambda_2(t)$ or $\lambda_2(t) > \lambda_1(t)$ for all $t > 0$. Without loss of generality, we may assume $\lambda_1(t) > \lambda_2(t)$. Then we can have that

$$S_1(t) = e^{-\int_0^t \lambda_1(s)ds} < e^{-\int_0^t \lambda_2(s)ds} = S_2(t)$$

Therefore two survival function will not cross as well.