TMA 4275 V2020 Home Exam

Problem ~~Exercise~~ 1

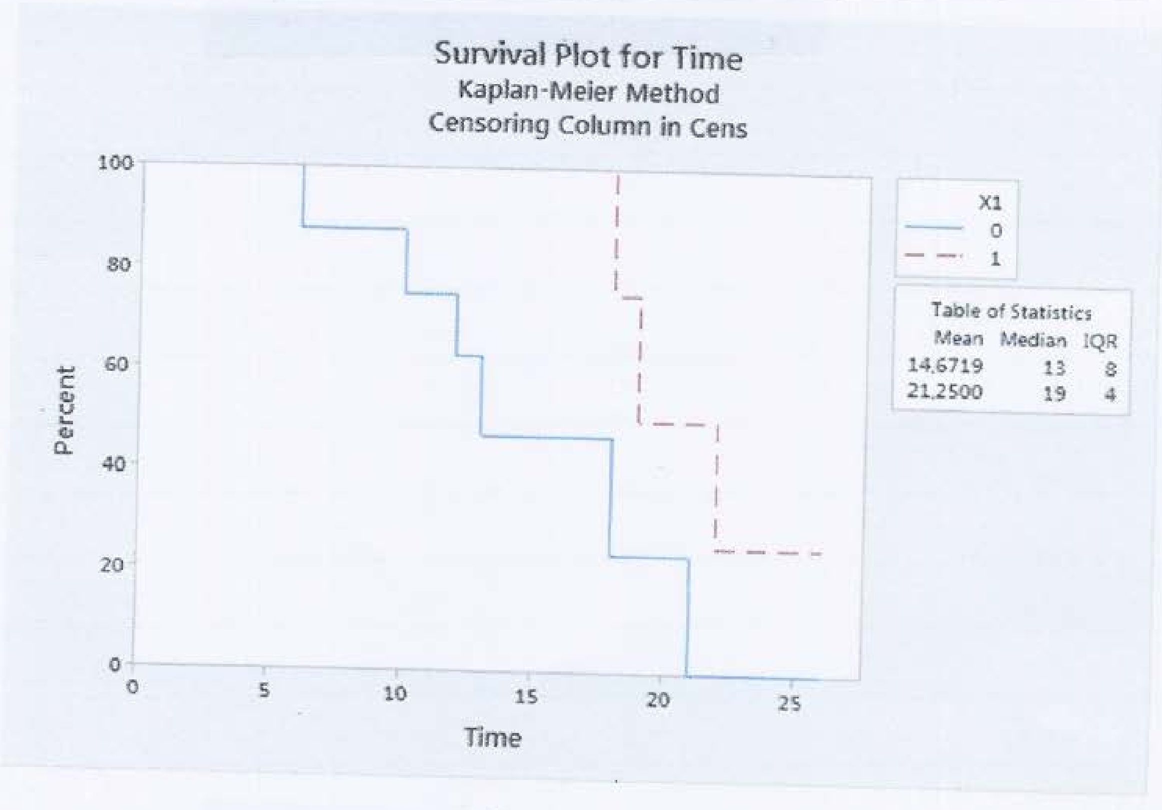
Kaplan-Meier Estimates

a) < 45, $x_1 = 0$

Time	Number at Risk	Number Failed	Survival Probability	Standard Error	95,0% Normal CI Lower	95,0% Normal CI Upper
6	8	1	0,875000	0,116927	0,645828	1,00000
10	7	1	0,750000	0,153093	0,449943	1,00000
12	6	1	0,625000	0,171163	0,289526	0,96047
13	4	1	0,468750	0,186521	0,103176	0,83432
18	2	1	0,234375	0,190167	0,000000	0,60709
21	1	1	0,000000	0,000000	0,000000	0,00000

≥ 45, $x_1 = 1$

Time	Number at Risk	Number Failed	Survival Probability	Standard Error	95,0% Normal CI Lower	95,0% Normal CI Upper
18	4	1	0,75	0,216506	0,325655	1,00000
19	3	1	0,50	0,250000	0,010009	0,98999
22	2	1	0,25	0,216506	0,000000	0,67434



Medians: $x_1 = 0$: 19, $x_2 = 1$: 19

Mean lengths are areas under the curve from 0 to the last observed time (censored or non-censored).

(See Minitab-output)

b) A positive (negative) estimated parameter corresponds to an increase (decrease) in hospital stay when x_i change from 0 to 1. ~~Opposite for the negative~~

Significant: x_1, x_5, x_6.

T with covariates $x_1, x_2, \dots, x_6$:

$$T \sim \text{Weibull}\left(e^{\hat{\beta}_0 + \hat{\beta}_1 x_1 + \dots + \hat{\beta}_6 x_6}, \hat{\alpha}\right) \quad \text{(estimated)}$$

Median of Weibull(α, θ):

$$t_{0.5} = \theta (\ln 2)^{1/\alpha}$$

Here: $t_{0.5} = e^{\hat{\beta}_0 + \hat{\beta}_1 x_1 + \dots + \hat{\beta}_6 x_6} \cdot (\ln 2)^{1/\hat{\alpha}}$

value of x_1

$$\frac{t_{0.5}(1) - t_{0.5}(0)}{t_{0.5}(0)} = \frac{e^{\hat\beta_0 + \hat\beta_1 + \hat\beta_2 x_2 + \cdots + \hat\beta_6 x_6} - e^{\hat\beta_0 + \hat\beta_2 x_2 + \cdots + \hat\beta_6 x_6}}{e^{\hat\beta_0 + \hat\beta_2 x_2 + \cdots + \hat\beta_6 x_6}}$$

$$= \frac{e^{\hat\beta_1} - 1}{1} = \underline{e^{\hat\beta_1} - 1}$$

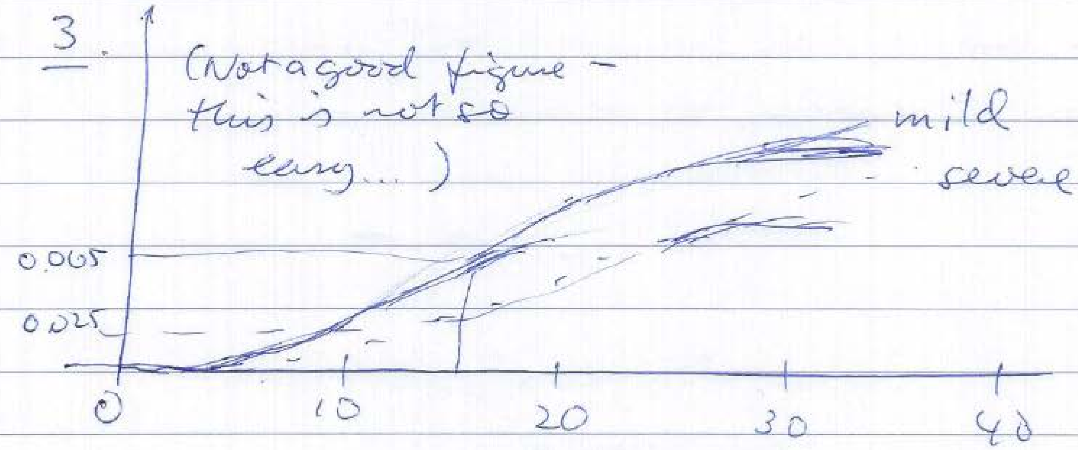
Note that $(\ln 2)^{1/\lambda}$ are cancelling in this calculation

c)

1. Median $x_6 = 0$ (mild) 19
$x_6 = 1$ (severe) 23

2. Quartiles: $x_6 = 0$: 15, 23
$x_6 = 1$: 18, 30

3.



4. These probabilities are approximately equal to the value of the hazard.
Look at derivatives

Severe: Increases by 0.5 on 20, i.e. $\hat{z}_{severe}(15) = \frac{0.5}{20} = \underline{0.025}$

Mild: " " 1.3 " ", ie $\hat{z}_{mild}(15) = \frac{1.3}{20} = \underline{\underline{0.065}}$

d) The plots ~~are~~ indicate fairly good fit, but some observations with low time seem to be questionable in the Weibull distributions.

The slopes of the probability plots correspond to the shape of the Weibull distributions (theoretically, the probability plots are close to the lines, "$y = \alpha x + \alpha \ln \theta$" so the shape is the slope of the lines).

Let α_0, α_1 be the shapes corresponding to $x_6 = 0$ and $x_6 = 1$, respectively. Then we can test

$$H_0: \alpha_0 = \alpha_1 \quad \text{vs.} \quad H_1: \alpha_0 \neq \alpha_1$$

We then calculate

$$W = 2\big(l(\hat{\alpha}_0, \hat{\alpha}_1, \cancel{\beta}\, \hat{\theta}_0, \hat{\theta}_1)$$

$$- \, l(\alpha^*, \alpha^*, \theta_0^*, \theta_1^*) \big)$$

Here, the first is the log-lik of the full ~~regression model~~, model where we have 538 data; ~~and~~ The likelihood is then the product of two likelihoods corresponding to separate models; and log-lik is hence the sum of logliks:

$$-1077.416 - 169.687 = -1247.103$$

$$\begin{array}{r} 1077.416 \\ 169.687 \\ \hline 1247.103 \end{array}$$

The second l ~~is~~ has three parameters, and is equivalent to the regression one,
-1249.846

Thus the test statistic is $W = 2(1249.846 - 1247.103)$
$= 2 \cdot 2.743 = 5.486$

and the p-value is $\underline{P(\chi_1^2 > 5.486)} = \underline{0.019169}$

so we reject at 5% level (critical value 3.84)

Problem 2

a) ~~Let $\hat{W}(t) = N(t)$. This~~

Note that $N(t)$ is not completely observed, so $\hat{W}(t) = N(t)$ is not a valid estimator.

Define instead

$$\hat{W}(t) = \begin{cases} D_1 + D_2 + \cdots D_{\lfloor t \rfloor} = N(\lfloor t \rfloor) \\ \text{where } \lfloor\ \rfloor \text{ is the "floor" function} \end{cases}$$

The $w(t)$ can roughly be considered as the slope of the plot in Figure 4 at t.
The plot shows that the rate of deaths is first increasing until, say day 30, and is then constant until, say 45, and is then slightly decreasing.

b) $L(\theta) = P(D_1 = d_1, \ldots, D_r = d_r) = \prod_{i=1}^{r} P(D_i = d_i)$

Now $D_i \underset{\text{indep.}}{\sim} \text{Poisson}(W(i;\theta) - W(i-1;\theta))$

so (1) follows by noting that

$$\prod_{i=1}^{r} e^{-(W(i;\theta) - W(i-1;\theta))} = e^{-W(r;\theta)}$$

c) $W(t;\alpha,\beta) = \int_0^t e^{\alpha+\beta u}\,du = e^{\alpha}\int_0^t e^{\beta u}\,du$

$$= \frac{e^{\alpha}}{\beta}\left(e^{\beta t}-1\right) \quad \text{QED} \quad \circledast$$

If $\beta=0$, then $w(t;\alpha,0)=e^{\alpha}$, so

$$W(t;\alpha,0) = e^{\alpha}t \quad (\text{can let } \beta\to 0 \text{ in } \circledast)$$

Now

$$W(i;\alpha,\beta) - W(i-1;\alpha,\beta)$$

$$= \frac{e^{\alpha}}{\beta}\left(e^{\beta i}-1\right) - \frac{e^{\alpha}}{\beta}\left(e^{\beta(i-1)}-1\right)$$

$$= \frac{e^{\alpha}}{\beta}\left(e^{\beta i} - e^{\beta(i-1)}\right)$$

$$= \frac{e^{\alpha}e^{\beta i}}{\beta}\left(1-e^{-\beta}\right)$$

Now

$$l(\alpha,\beta) = \sum_{i=1}^{r} \cancel{\ln}\, d_i \ln\left(W(i;\alpha,\beta) - W(i-1;\alpha,\beta)\right)$$

$$- \sum_{i=1}^{r} \ln d_i! - W(r;\alpha,\beta))$$

$$= \sum_{i=1}^{r} d_i\left[\alpha + \beta i + \ln\frac{1-e^{-\beta}}{\beta}\right]$$

$$- \sum_{i=1}^{r} \ln d_i! - \frac{e^{\alpha}}{\beta}\left(e^{\beta r} - 1\right)$$

$$= \left(\alpha + \ln \frac{1-e^{-\beta}}{\beta}\right) \sum_{i=1}^{r} d_i + \beta \sum_{i=1}^{r} i\, d_i$$

$$- \sum_{i=1}^{r} \ln d_i! - \frac{e^{\alpha}}{\beta}\left(e^{\beta r} - 1\right)$$

If β is known; solve $\frac{\partial l}{\partial \alpha} = 0$.

$$\frac{\partial l(\alpha,\beta)}{\partial \alpha} = \sum_{i=1}^{r} d_i - \frac{e^{\alpha}}{\beta}\left(e^{\beta r} - 1\right) = 0$$

$$e^{\alpha} = \frac{\beta \sum_{i=1}^{r} d_i}{e^{\beta r} - 1}$$

So $$\hat{\alpha} = \ln \frac{\hat{\beta} \sum_{i=1}^{r} d_i}{e^{\hat{\beta} r} - 1} \quad \text{if } \beta \neq 0$$

If $\beta = 0$: let $\beta \to 0$: $$\hat{\alpha} = \ln \frac{\sum_{i=1}^{r} d_i}{r}$$

~~Tabk~~

d) $\hat{\beta} = 0.1242$

$$\Rightarrow \hat{\alpha} = \hat{\alpha}(\hat{\beta}) = \ln \frac{0.1242 \cdot 50}{e^{24 \cdot 0.1242} - 1} = -1.1026$$

Now the expected # deaths in days 25 to 31 is

$W(31) - W(24)$ which is estimated by

$$W(31; \hat{\alpha}, \hat{\beta}) - W(24; \hat{\alpha}, \hat{\beta})$$

$$= \frac{e^{\hat{\alpha}}}{\hat{\beta}} \left(e^{31\hat{\beta}} - e^{24\hat{\beta}} \right) =$$

$$= 72.98 \approx \underline{\underline{73}}$$

Actually observed: $\underline{42}$