STAT3032 SURVIVAL MODELS

TUTORIAL SOLUTIONS WEEK SEVEN

Question One

The Nelson-Aalen estimator is

$$\exp\left(-\hat{\Lambda}(t)\right) = \exp\left(-\sum_{t_j \le t} \frac{d_j}{r_j}\right)$$

$$= \prod_{t_j \le t} \exp\left(\frac{-d_j}{r_j}\right)$$

$$\approx \prod_{t_j \le t} \left(1 - \frac{d_j}{r_j}\right)$$

$$= \prod_{t_j \le t} \left(\frac{r_j - d_j}{r_j}\right)$$ which is the Kaplan-Meier estimator.

Question Two

(a)

```
time<-c(5,9,4,8,6,15,13,1,2)
death<-c(1,1,1,1,1,0,1,0,1)
hospital<-c(0,0,1,1,1,1,1,0,0)
```

Note: It does not matter which hospital is coded as 1. The results from the fitted Cox regression will remain the same.

(b)

```
library(survival)
cox.fit<-coxph(Surv(time,death)~hospital)
 summary(cox.fit)
Call:
coxph(formula = Surv(time, death) ~ hospital)

  n= 9, number of events= 7

            coef exp(coef) se(coef)      z Pr(>|z|)
hospital -0.8821    0.4139   0.8251 -1.069    0.285
```

```
         exp(coef) exp(-coef) lower .95 upper .95
hospital    0.4139      2.416   0.08215     2.086

Concordance= 0.625  (se = 0.116 )
Rsquare= 0.116   (max possible= 0.905 )
Likelihood ratio test= 1.11  on 1 df,   p=0.2921
Wald test            = 1.14  on 1 df,   p=0.285
Score (logrank) test = 1.21  on 1 df,   p=0.2704
```

The test-statistics for testing the significance of hospital is -1.07 with a p-value of 0.285. It is clear that at the 5% level that hospital is not a significant variable. It appears that the choice of hospital does not impact on survival outcomes.

Question Three

(a)

$$PL(\beta)=\left(\frac{e^{\beta_1+\beta_3}}{e^{\beta_1}+e^{\beta_2}+2e^{\beta_3}+e^{\beta_1+\beta_2}+2e^{\beta_1+\beta_3}+e^{\beta_2+\beta_3}}\right)\left(\frac{e^{\beta_1}}{e^{\beta_1}+e^{\beta_2}+2e^{\beta_3}+e^{\beta_1+\beta_2}+e^{\beta_1+\beta_3}+e^{\beta_2+\beta_3}}\right).$$
$$\left(\frac{e^{\beta_1+\beta_2+\beta_3}}{\left(e^{\beta_2}+2e^{\beta_3}+\ e^{\beta_1+\beta_3}+e^{\beta_2+\beta_3}\right)^2}\right)\left(\frac{e^{\beta_2+\beta_3}}{2e^{\beta_3}+e^{\beta_2+\beta_3}}\right)$$

(b)

```
survtime<-c(5,6,4,1,2,8,9,5)
status<-c(1,1,0,1,1,0,1,1)
z1<-c(0,0,1,1,1,0,0,1)
z2<-c(1,1,1,0,0,0,0,0)
z3<-c(0,1,0,1,0,1,1,1)

cox.fit<-coxph(Surv(survtime,status)~z1+z2+z3)
summary(cox.fit)
Call:
coxph(formula = Surv(survtime, status) ~ z1 + z2 + z3)

  n= 8, number of events= 6

      coef exp(coef) se(coef)      z Pr(>|z|)
z1  2.4823   11.9690   1.4206  1.747   0.0806 .
z2  0.3466    1.4142   1.1579  0.299   0.7647
z3 -1.0594    0.3467   1.4935 -0.709   0.4781
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

   exp(coef) exp(-coef) lower .95 upper .95
z1   11.9690    0.08355   0.73939   193.751
z2    1.4142    0.70711   0.14618    13.682
z3    0.3467    2.88454   0.01856     6.474
```

```
Concordance= 0.833  (se = 0.174 )
Rsquare= 0.447   (max possible= 0.869 )
Likelihood ratio test= 4.74  on 3 df,   p=0.192
Wald test            = 3.35  on 3 df,   p=0.3413
Score (logrank) test = 4.59  on 3 df,   p=0.2047
```

Looking at the output from likelihood ratio test (p-value =0.192) it does not appear that any of the risk factors are significant.

(c)

(i) multiplies the base hazard by 11.966.

(ii) raises the base survival probability to the power of 11.966.

Question Three

The output from a Cox proportional-hazards regression analysis of a Recidivism dataset (Rossi, Berk and Lenihan 1980) is provided below. The purpose of the analysis was to investigate whether certain covariates were related to survival time (in this context survival time being the time until first-arrest upon release from prison). The covariates included in the fitted model are:

- *fin*: a categorical variable taking the value 1 if financial aid was received and 0 otherwise.
- *age*: age in years.
- *race*: a categorical variable taking the value 1 for blacks and 0 otherwise.
- *mar*: a categorical variable taking the value 1 if married and 0 otherwise.
- *prio*: the number of prior convictions

```
> summary(fit)
Call:
coxph(formula = Surv(week, arrest) ~ fin + age +
I(age^2) + race +
    mar + prio, data = Rossi)

  n= 432, number of events= 114

             coef     exp(coef)  se(coef)   z      Pr(>|z|)
fin       -0.373246  0.688496  0.191009 -1.954  0.05069
age       -0.276401  0.758509  0.136125 -2.031  0.04231
age^2      0.003907  1.003915  0.002409  1.622  0.10485
race       0.344845  1.411770  0.308424  1.118  0.26353
mar       -0.417402  0.658756  0.378661 -1.102  0.27033
prio       0.099941  1.105106  0.027367  3.652  0.00026
---

Concordance= 0.642  (se = 0.027 )
Rsquare= 0.078   (max possible= 0.956 )
Likelihood ratio test= 34.99  on 6 df,   p=4.336e-06
```

```
Wald test            = 35.54  on 6 df,   p=3.389e-06
Score (logrank) test = 37.09  on 6 df,   p=1.693e-06
```

Note in the questions that follow, β_1 refers to the parameter corresponding to the variable *fin*, β_2 to the parameter corresponding to the variable *age*, and so on.

Using the R-output above answer the following questions:

a)

$$
\begin{aligned}
&= \frac{\exp(-0.373 - 23\times 0.276 + 23^2 \times 0.0039 + 0.345 + 2\times 0.1)}{\exp(-26\times 0.276 + 26^2 \times 0.0039 + 0.345 + -0.417)} \\
&= 1.64
\end{aligned}
$$

b) Provide a 95% confidence interval for the multiplicative increase in the hazard ratio for an increase in the number of prior convictions of two, everything else held constant.

Want a 95% CI for $\exp(2\beta_6)$.

95% CI for β_6 is $0.1 \pm 2\times 0.027 \Rightarrow$ 95% CI for $\exp(2\beta_6)$ is $\exp(2\times(0.1\pm 2\times 0.027)) = [1.1, 1.36]$.

c) Provide a standard error for the estimate of $\exp(\beta_1)$ obtained from the fitted model.

$\text{var}(\hat{\beta}_1) = 0.19^2 \Rightarrow \text{var}(\exp(\hat{\beta}_1)) \approx 0.19^2 \times \exp(2\hat{\beta}_1) = 0.017.$

d) Is marital status related to time until first-arrest? You must provide statistically sound reasons for your answer.

No. The test of the hypothesis that $\beta_5 = 0$ versus $\beta_5 \neq 0$ has a test-statistics of negative 1.1, which is less than 2 in absolute value.

e) Does the positive coefficient estimate obtained for the covariate representing the number of prior convictions seem reasonable? You must provide justification for your answer.

Yes. A positive coefficient estimate suggests that has the number of prior convictions increases that the risk of being re-arrested also increases.