MATH3085/6143 Survival Models – Worksheet 3 Solutions

1. The fitted Cox proportional model is

$$\hat{h}_i(t) = h_0(t)\exp(0.005A_i + 0.05S_i),$$

where $\hat{h}_i(t)$ is the estimated hazard of individual i at duration t, $h_0(t)$ is the baseline hazard, the covariate A_i is the age of individual i in years, while covariate S_i is a unit-specific smoking indicator taking value 1 for cigarette smokers and 0 otherwise.

Suppose x is the age of a non-smoker with the same estimated hazard of death as a smoker aged 55 years, then we have

$$h_0(t)\exp(0.005 \times 55 + 0.05 \times 1) = h_0(t)\exp(0.005x + 0.05 \times 0),$$

whence

$$\begin{aligned} &\Rightarrow \quad \exp(0.325) = \exp(0.005x) \\ &\Rightarrow \quad 0.325 = 0.005x \\ &\Rightarrow \quad x = 65. \end{aligned}$$

2. i) The Cox proportional hazards regression model for this study assumes independent time variables, with hazards given by

$$h_{T_i}(t) = h_0(t)\exp(\beta_A A_i + \beta_F F_i + \beta_E E_i), \qquad i = 1, \ldots, n$$

where $h_i(t)$ is the hazard of unit i at duration t while $h_0(t)$ is the baseline hazard.

ii) A female aged exactly 16 years when she first claimed benefit who had not passed the school mathematics examination.

iii) "The hazard of resuming work for males aged 17 years who had passed the mathematics examination was 1.5 times the hazard for males aged 16 years who had not passed the mathematics examination" implies that

$$\frac{h_0(t)\exp(\beta_A \times 1 + \beta_F + \beta_E)}{h_0(t)\exp(\beta_A \times 0 + \beta_F)} = \exp(\beta_A + \beta_E) = 1.5. \tag{1}$$

"Females who had passed the examination were twice as likely to take up a new job as were males of the same age who had failed" implies that

$$\frac{h_0(t)\exp(\beta_E)}{h_0(t)\exp(\beta_F)} = \exp(\beta_E - \beta_F) = 2 \tag{2}$$

since the age terms cancel each other out.
"Females aged 20 years who had passed the examination were twice as likely to resume work as were males aged 16 years who had also passed the examination" implies that

$$\frac{h_0(t)\exp(\beta_A \times 4)}{h_0(t)\exp(\beta_F)} = \exp(4\beta_A - \beta_F) = 2. \tag{3}$$

(1)÷(2) gives

$$\exp(\beta_A + \beta_F) = \frac{1.5}{2} = \frac{3}{4}. \tag{4}$$

(3)×(4) yields

$$\begin{aligned} \exp(5\beta_A) &= 2 \times \frac{3}{4} = 1.5 \\ \Rightarrow \quad \beta_A &= \frac{1}{5}\log 1.5 = 0.0811. \end{aligned}$$

From (1), we then obtain

$$\beta_E = \log 1.5 - \beta_A = \log 1.5 - 0.0811 = 0.3244.$$

Finally, from (2), we obtain

$$\beta_F = \beta_E - \log 2 = 0.3244 - \log 2 = -0.3687.$$

3. i) A female chicken kept in the old enclosure.

ii) The Cox proportional hazards model for hazard function here is

$$h_{T_i}(t) = h_0(t)\exp(\beta_D D_i + \beta_G G_i + \beta_E E_i + \beta_S S_i), \qquad i = 1, \ldots, n$$

where D_i, G_i, E_i, S_i are dummy variables such that

$$D_i = \begin{cases} 1 & \text{for ducks} \\ 0 & \text{otherwise} \end{cases} \qquad G_i = \begin{cases} 1 & \text{for geese} \\ 0 & \text{otherwise} \end{cases}$$

$$E_i = \begin{cases} 1 & \text{for birds in new enclosure} \\ 0 & \text{for birds in old enclosure} \end{cases} \qquad S_i = \begin{cases} 1 & \text{for male} \\ 0 & \text{for female} \end{cases}$$

and β_D, β_G, β_E, β_M are the corresponding regression parameters associated. Using the usual formula for constructing 95% confidence interval (CI), $[\hat{\beta}_i - 1.96 s.e.(\hat{\beta}_i), \hat{\beta}_i + 1.96 s.e.(\hat{\beta}_i)]$, we have that:

CI for β_D: $[-0.210 - 1.96 \times \sqrt{0.002}, -0.210 + 1.96 \times \sqrt{0.002}] = [-0.2977, -0.1223]$.
CI for β_G: $[0.075 - 1.96 \times \sqrt{0.004}, 0.075 + 1.96 \times \sqrt{0.004}] = [0.0490, 0.1990]$.
CI for β_E: $[0.125 - 1.96 \times \sqrt{0.0015}, 0.125 + 1.96 \times \sqrt{0.0015}] = [0.0491, 0.2009]$.
CI for β_M: $[0.2 - 1.96 \times \sqrt{0.0026}, 0.2 + 1.96 \times \sqrt{0.0026}] = [0.1000, 0.2999]$.

iii) The parameter of the new enclosure is $+0.125$, so the ratio of the hazards of two otherwise identical birds is $\exp(0.125) = 1.133$. Hence the hazard appears to have got worse for birds in the new enclosure! In fact, the 95% confidence interval was constructed to be $[0.0491, 0.2009]$ in part (ii), which is entirely positive (OR does not include 0), so the deterioration is statistically significant at 5% level, contradicting the farmer's belief that new enclosure will result in an increase in his birds life expectancy.

iv) Note that under the Cox proportional hazards model, the survival functions of 2 individuals A and B with covariates $\boldsymbol{x}_A$ and $\boldsymbol{x}_B$ respectively can be written down as

$$\begin{aligned} S_B(t) &= \exp\left[-\int_0^t h_B(s)\mathrm{d}s\right] \\ &= \exp\left[-\int_0^t h_0(s)\exp(\boldsymbol{x}_B^\top\boldsymbol{\beta})\mathrm{d}s\right] \\ &= \exp\left[-\exp(\boldsymbol{x}_B^\top\boldsymbol{\beta})\int_0^t h_0(s)\mathrm{d}s\right] \\ &= S_0(t)^{\exp(\boldsymbol{x}_B^\top\boldsymbol{\beta})} \end{aligned}$$

Similarly

$$S_A(t) = S_0(t)^{\exp(\boldsymbol{x}_A^\top\boldsymbol{\beta})} \quad \Rightarrow \quad S_0(t) = S_A(t)^{\frac{1}{\exp(\boldsymbol{x}_A^\top\boldsymbol{\beta})}} \quad \Rightarrow \quad S_B(t) = S_A(t)^{\frac{\exp(\boldsymbol{x}_B^\top\boldsymbol{\beta})}{\exp(\boldsymbol{x}_A^\top\boldsymbol{\beta})}}$$

Thus, given that $F_A(6) = 0.1$, the required probability is simply

$$\begin{aligned} F_B(6) &= 1 - S_B(6) \\ &= 1 - S_A(6)^{\frac{\exp(-0.210\times1+0.075\times0+0.125\times1+0.2\times0)}{\exp(-0.210\times0+0.075\times1+0.125\times0+0.2\times1)}} \\ &= 1 - (1 - F_A(6))^{\frac{\exp(-0.085)}{\exp(0.275)}} \\ &= 1 - (1 - 0.1)^{\exp(-0.36)} \\ &= 0.0709. \end{aligned}$$

4. i) Recall from lecture notes that the partial likelihood is computed as $L(\boldsymbol{\beta}) = \prod_{i:d_i=1} \frac{\exp(\boldsymbol{x}_i^\top \boldsymbol{\beta})}{\sum_{j \in R_i} \exp(\boldsymbol{x}_j^\top \boldsymbol{\beta})}$, where R_i is the risk set at time t_i, and $\exp(\boldsymbol{x}_i^\top \boldsymbol{\beta}) = 1$ for units i in group A and $\exp \beta$ for units i in group B. Also note that only at times of failures that the data will contribute to elements of the partial likelihood, censored data are only taken into account indirectly through the risk sets. In this problem, the failure times are at 2, 4, 6, 8, 9, 10, 14, the partial likelihood is hence given by

$$
\begin{aligned}
L(\beta) &= \frac{1}{9+9\exp\beta} \times \frac{\exp\beta}{8+8\exp\beta} \times \frac{1}{7+7\exp\beta} \\
&\quad \times \frac{1}{3+3\exp\beta} \times \frac{\exp\beta}{2+3\exp\beta} \times \frac{1}{2+2\exp\beta} \times \frac{1}{1+\exp\beta},
\end{aligned}
$$

which simplifies to

$$
L(\beta) = \frac{\exp(2\beta)}{3024(1+\exp\beta)^6(2+3\exp\beta)}
$$

whence

$$
l(\beta) = \log L(\beta) = 2\beta - \log(3024) - 6\log(1+\exp\beta) - \log(2+3\exp\beta).
$$

ii) Differentiate the partial log-likelihood w.r.t. β to obtain

$$
\begin{aligned}
\frac{\mathrm{d}l(\beta)}{\mathrm{d}\beta} &= 2 - \frac{6\exp\beta}{1+\exp\beta} - \frac{3\exp\beta}{2+3\exp\beta} \\
&= 2 - 6\left(1 - \frac{1}{1+\exp\beta}\right) - \left(1 - \frac{2}{2+3\exp\beta}\right) \\
&= -5 + \frac{6}{1+\exp\beta} + \frac{2}{2+3\exp\beta}.
\end{aligned}
$$

Set $\frac{\mathrm{d}l(\beta)}{\mathrm{d}\beta} = 0$ to solve for the maximum partial likelihood estimate, $\hat{\beta}$,

$$
\begin{aligned}
& -5 + \frac{6}{1+\exp(\hat{\beta})} + \frac{2}{2+3\exp(\hat{\beta})} = 0 \\
\Rightarrow\ & -5(1+\exp(\hat{\beta}))(2+3\exp(\hat{\beta})) + 6(2+3\exp(\hat{\beta})) + 2(1+\exp(\hat{\beta})) = 0 \\
\Rightarrow\ & 15\exp(2\hat{\beta}) + 5\exp(\hat{\beta}) - 4 = 0 \\
\Rightarrow\ & \exp(\hat{\beta}) = \frac{-5 \pm \sqrt{5^2 - 4(15)(-4)}}{2 \times 15} \text{ (quadratic formula)} \\
\Rightarrow\ & \exp(\hat{\beta}) = -0.7093 \text{ (invalid) or } 0.376 \\
\Rightarrow\ & \hat{\beta} = \log 0.376 = -0.9783.
\end{aligned}
$$

Differentiate the partial log-likelihood again to obtain the observed information matrix,

$$
I(\beta) = -\frac{\mathrm{d}^2 l(\beta)}{\mathrm{d}\beta^2} = \frac{6\exp\beta}{(1+\exp\beta)^2} + \frac{6\exp\beta}{(2+3\exp\beta)^2}.
$$

Therefore,

$$
\begin{aligned}
s.e.(\hat{\beta}) &= [I(\hat{\beta})^{-1}]^{\frac{1}{2}} \\
&= \left[\frac{6\exp(-0.9783)}{(1+\exp(-0.9783))^2} + \frac{6\exp(-0.9783)}{(2+3\exp(-0.9783))^2}\right]^{-\frac{1}{2}} \\
&= 0.8386.
\end{aligned}
$$

iii) Here we use the Wald test. Under the null hypothesis H_0,

$$\frac{\hat{\beta}}{s.e.(\hat{\beta})} \sim N(0,1).$$

H_0 is rejected when

$$\left|\frac{\hat{\beta}}{s.e.(\hat{\beta})}\right| > 1.96$$

for a 5% significance level test. Thus, since

$$\left|\frac{-0.9783}{0.8386}\right| = 1.17 < 1.96$$

for this problem, we **do not** reject H_0 at 5% significance level.
Equivalently, the 95% confidence interval is constructed to be

$$-0.9783 \pm 1.96 \times s.e.(\hat{\beta}) = [-2.6220, 0.6654].$$

So since the 95% CI contains 0, we **do not** reject H_0.

5. (a) Partial likelihood is

$$L(\beta) = \frac{1}{3+3e^{\beta}} \times \frac{e^{\beta}}{2+3e^{\beta}} \times \frac{1}{2+2e^{\beta}} = \frac{e^{\beta}}{6\left(1+e^{\beta}\right)^2\left(2+3e^{\beta}\right)}.$$

(b) As all patients, without relapse, were censored at 70 days, so analysis could be carried out, this censoring is non-informative.

6. (a) $t_1, \ldots, t_n$ are observations of independent random variables $T_1, \ldots, T_n$ with

$$T_i \sim \text{Weibull}\left[\alpha, \exp\left(-\beta_0 - \beta_1 x_i\right)\right].$$

(b) In the Weibull model, we specify a distribution for T_i. In the Cox model, we do not.

(c) 95% CI for β is

$$\begin{array}{rcl} \hat{\beta} & \pm & z_{0.975}\text{s.e.}(\hat{\beta}) \\ 0.14 & \pm & 1.96 \times 0.09 \\ (-0.0364 & , & 0.316) \end{array}$$

Test statistic is

$$\left|\frac{\hat{\beta}}{\text{s.e.}(\hat{\beta})}\right| = \left|\frac{0.14}{0.09}\right| = 1.56 < z_{0.975} = 1.96.$$

Hence do not reject $\text{H}_0 : \beta = 0$. Conclude no evidence of a difference between old and new sealants.