Stats 756 - Assignment 2

Tommy Flynn (400121496)

December 23, 2022

Question 1:

We are given the frailty model for lifetime T, thus, $h(t|\lambda) = \lambda h_0(t)$. The cumulative hazard is

$$H(t|\lambda) = \int_0^t h(s|\lambda)\, ds = \int_0^t \lambda h_0(s)\, ds = \lambda \int_0^t h_0(s)\, ds = \lambda H_0(t)$$

This means the survival function is

$$S(t|\lambda) = e^{-H(t|\lambda)} = e^{-\lambda H_0(t)}$$

Therefore, the density is

$$f(t|\lambda) = h(t|\lambda)S(t|\lambda) = \lambda h_0(t) e^{-\lambda H_0(t)}$$

a) Let the joint density be $f(t, \lambda) = g(\lambda) f(t|\lambda)$ with $g(\lambda) \sim \text{Gamma}(\alpha, \theta)$. Then we have

$f(t, \lambda)$

$= g(\lambda) f(t|\lambda)$

$= \dfrac{\lambda^{\alpha-1} e^{-\lambda/\theta}}{\Gamma(\alpha)\theta^{\alpha}} \cdot \lambda h_0(t) e^{-\lambda H_0(t)}$

$= \dfrac{\lambda^{\alpha} e^{-\lambda(1/\theta + H_0(t))}}{\Gamma(\alpha)\theta^{\alpha}} h_0(t)$

b) For the conditional density $g(\lambda|t)$ we have

$g(\lambda|t)$

$$= f(t,\lambda)\frac{1}{f(t)}$$

$$= \frac{\lambda^\alpha e^{-\lambda(1/\theta+H_0(t))}}{\Gamma(\alpha)\theta^\alpha}h_0(t)\cdot\frac{\Gamma(\alpha)\theta^\alpha}{h_0(t)\int_0^\infty \lambda^\alpha e^{-\lambda(1/\theta+H_0(t))}\,d\lambda}$$

$$= \lambda^\alpha e^{-\lambda(1/\theta+H_0(t))}\frac{1}{\Gamma(\alpha+1)(1/\theta+H_0(t))^{\alpha+1}} \sim \text{Gamma}\left(\alpha+1, \frac{1}{\theta}+H_0(t)\right)$$

c) The frailty among the survivors at time t is as follows

$g(\lambda|T>t)$

$$\int_t^\infty \frac{\lambda^\alpha e^{-\lambda(1/\theta+H_0(u))h_0(u)}}{\Gamma(\alpha)\theta^\alpha}\,du\cdot\frac{\Gamma(\alpha)\theta^\alpha}{\int_0^\infty\int_t^\infty \lambda^\alpha e^{-\lambda(1/\theta+H_0(u))h_0(u)}\,du d\lambda}$$

$$\lambda^\alpha e^{-\lambda/\theta}\left[\frac{-e^{-\lambda H_0(u)}}{\lambda}\right]\Bigg|_{u=t}^{\infty}\cdot\frac{1}{\int_0^\infty \lambda^\alpha e^{-\lambda/\theta}\left[\frac{-e^{-\lambda H_0(u)}}{\lambda}\right]\Big|_{u=t}^{\infty} d\lambda}$$

$$= \lambda^{\alpha-1}e^{-\lambda(1/\theta+H_0(t))}\frac{1}{\int_0^\infty \lambda^{\alpha-1}e^{-\lambda(1/\theta+H_0(t))}\,d\lambda}$$

$$= \lambda^{\alpha-1}e^{-\lambda(1/\theta+H_0(t))}\frac{1}{\Gamma(\alpha)(1/\theta+H_0(t))^{\alpha}} \sim \text{Gamma}\left(\alpha, \frac{1}{\theta}+H_0(t)\right)$$

Question 2:

We will perform the log-rank test on the glioma data with the null hypothesis that there is no difference between survival curves of astrocytoma and glioblastoma. The test statistic is $Q = 7.5 \sim \chi^2(1)$. This implies a p-value of 0.006. Therefore, we have sufficient evidence to reject the null hypothesis meaning there is a significant difference between the survival curves of astrocytoma and glioblastoma. This is a reasonable conclusion as seen by the survival curves below.

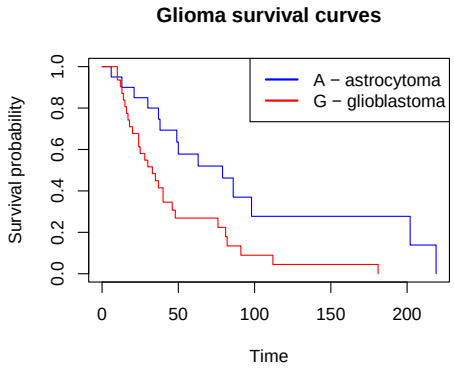

Figure 1: Kaplan-Meier curves by class.

Question 3:

a) Below are the results of the Weibull(η, α) fits.

Table 1: Weibull(η, α) fits.

Parameter	**Estimate**	**Std Error**
η_{Atypical}(shape)	0.8320822	0.1279216
α_{Atypical}(scale)	142.6771166	31.6729573
η_{Typical}(shape)	0.7746011	0.1357687
α_{Typical} (scale)	73.4608222	20.2847609

b) To decide if the exponential is adequate, we perform a likelihood ratio test and Wald test for each class. That is,

$$H_0 : \eta = 1 \quad \text{v.s.} \quad H_1 : \eta \neq 1$$

with the following test statistics:

$$\Lambda = -2(\ell_{\text{Exponential}} - \ell_{\text{Weibull}}) \sim \chi^2(1)$$

$$W = \frac{\hat{\eta} - 1}{\text{SE}(\hat{\eta})} \sim Z$$

Atypical:

$$\Lambda_{\text{Atypical}} = -2(-183.2484 + 182.4678) = 1.56114 \implies \text{p-value} = 0.2114984$$

$$W_{\text{Atypical}} = \frac{0.8320822 - 1}{0.1279216} = -1.312662 \implies \text{p-value} = 0.1892969$$

Both tests fail to reject the null-hypothesis at the 0.05 significance level thus exponential is adequate.

Typical:

$$\Lambda_{\text{Typical}} = -2(-116.3991 + 117.5897) = 2.381136 \implies \text{p-value} = 0.1228082$$

$$W_{\text{Typical}} = \frac{0.7746011 - 1}{0.1357687} = -1.660169 \implies \text{p-value} = 0.09688041$$

Both tests fail to reject the null-hypothesis at the 0.05 significance level thus exponential is adequate.

c) Since both classes are adequately modelled by exponential we may use the median $\log(2)/\lambda$ which is estimated by $\overline{x}\log(2)$. This means the atypical and typical median survival time estimates are

$$\text{Median}_{\text{Atypical}} = (80.96154)\log(2) = 56.11826 \text{ weeks}$$

$$\text{Median}_{\text{Typical}} = (60.57143)\log(2) = 41.98491 \text{ weeks}$$

d) The Weibull regression model with one covariate for tumour group is shown below.

Table 2: Weibull regression fit.

Parameter	**Estimate**	**Std Error**
β_0	4.3025923	0.2651153
β_1	0.6689624	0.3509176

To test if there is no difference between tumour group we perform the Wald test.

$$H_0 : \beta_1 = 0 \text{ v.s. } H_1 : \beta_1 \neq 0$$

$$W = \frac{0.6689624}{0.3509176} = 1.906323 \sim Z \implies \text{p-value} = 0.05660831$$

Therefore, at the 0.05 significance level there is no difference in survival time between tumour groups.

e) To test the difference we use a Weibull regression model with the intercept and use the likelihood ratio test.

$$H_0 : \beta_1 = 0 \text{ v.s. } H_1 : \beta_1 \neq 0$$

$$\Lambda = -2(-300.7041 + 298.9146) = 3.579 \sim \chi^2(1) \implies \text{p-value} = 0.05851437$$

At the 0.05 significance level there is no difference in survival time between tumour groups.

Question 4:

The results of the Weibull fit with two covariates for disease and therapy are below.

Table 3: Weibull regression fit.

Parameter	**Estimate**	**Std Error**
β_0	6.9439300	0.8067342
β_{Disease}	-0.6518663	0.9249562
β_{Therapy}	-0.0350508	0.9501682

Now we test whether therapy was effective.

$$H_0 : \beta_{\text{Therapy}} = 0 \text{ v.s } H_1 : \beta_{\text{Therapy}} \neq 0$$
$$W = \frac{-0.0350508}{0.9501682} \sim Z \implies \text{p-value} = 0.9705735$$

At the 0.05 significance level, the additional therapy did not have a significant effect on the survival time. The likelihood ratio test with the model containing disease and intercept results in the same conclusion $\Lambda = 0.001361753 \sim \chi^2(1) \implies$ p-value $= 0.9705632$.

Question 5:

Given n observations $(t_i, \delta_i x_i), i = 1, \ldots, n$ with $j = 1, \ldots, m_i$ for each group we have hazard and survival given frailty of the form:

$$h_{ij}(t_{ij}|y_i) = y_i h_0(t_{ij}) e^{\beta' x_{ij}}$$

$$S_{ij}(t_{ij}|y_i) = S_0(t_{ij})^{y_i e^{\beta' x_{ij}}}$$

Thus, the conditional likelihood contribution given the frailty of each group is

$$\prod_{j=1}^{m_i} y_i h_0(t_{ij}) e^{\beta' x_{ij}} S_0(t_{ij})^{y_i e^{\beta' x_{ij}}}, \quad i = 1, \ldots, n$$

Unconditioning each contribution using the respective frailty distribution $g(y_i)$ gives the unconditional likelihood contributions.

$$\int_0^\infty \prod_{j=1}^{m_i} y_i h_0(t_{ij}) e^{\beta' x_{ij}} S_0(t_{ij})^{y_i e^{\beta' x_{ij}}} g(y_i)\, dy_i, \quad i = 1, \ldots, n$$

Thus, the unconditional likelihood is the product of the unconditional likelihood contributions

$$\prod_{i=1}^{n} \int_0^\infty \prod_{j=1}^{m_i} y_i h_0(t_{ij}) e^{\beta' x_{ij}} S_0(t_{ij})^{y_i e^{\beta' x_{ij}}} g(y_i)\, dy_i$$

Question 6:

a) Below are the results of the cox hazard and gamma frailty fits.

Table 4: Cox hazard fit.

Parameter	**Estimate**	**Std Error**
β_{Age}	0.002032	0.009246
β_{Sex}	-0.829314	0.298955

Table 5: Gamma frailty fit.

Parameter	**Estimate**	**Std Error**
β_{Age}	0.00525	0.01168
β_{Sex}	-1.58749	0.44948

Note that the frailty variance is 0.412.

b) The gamma frailty model accounts for some heterogeneity in the observed failure times as seen by the sex variable which is nullified in the cox model. Although, we must note that the frailty variance is quite small and close to zero indicating possible homogeneity. Therefore, it is not obvious that the gamma frailty is suitable here.

Question 7:

a) Below we have the accuracy, sensitivity, and specificity at each cut-off.

Table 6: Gamma frailty fit.

Cut-off	**Accuracy**	**Sensitivity**	**Specificity**
18	0.33	1.00	0.00
20	0.34	1.00	0.01
22	0.41	0.99	0.12
24	0.59	0.95	0.42
26	0.68	0.85	0.60
28	0.71	0.66	0.74
30	0.73	0.47	0.87
32	0.72	0.34	0.92
34	0.69	0.21	0.93
36	0.70	0.17	0.97
38	0.68	0.07	0.98
40	0.67	0.01	0.995

b) We can see that as sensitivity increases specificity decreases. Thus, sensitivity and specificity are inversely related.

c) Below is the ROC curve. It has an estimated AUC of 0.7811497 with 95% confidence interval $(0.7398955, 0.8224039)$.

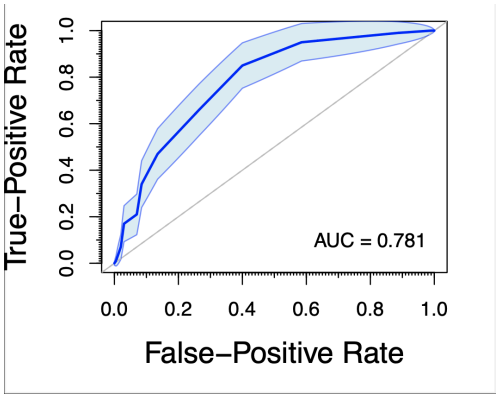

Figure 2: ROC curve.

d) Overall, BMI is a somewhat unimpressive as it invokes massive tradeoffs between sensitivity and specificity. A 90% level in one causes a sub-50% level in the other. The optimal point on the ROC curve indicated $(0.60, 0.85)$ which is quite imprecise in cancer discrimination.