Stats 756 — Assignment 2

December 8, 2022

---

1. In the frailty model introduced, prove that the density of lifetime T, given the frailty λ, is

$$f(t \mid \lambda) = \lambda h_0(t) e^{-\lambda H_0(t)}.$$

   (a) Then, using the fact that the joint density is

$$f(t, \lambda) = g(\lambda) f(t \mid \lambda),$$

   find an expression for the joint density when frailty density $g(\lambda)$ is that of $Gamma(\alpha, \theta)$, where α is the shape parameter and θ is the scale parameter;

   (b) Prove then that the conditional density $g(\lambda \mid t)$ is $Gamma(\alpha + 1, \theta(t))$, where $\theta(t) = \theta + H_0(t)$;

   (c) Prove that the density of frailty among the survivors at time t, i.e., $g(\lambda \mid T > t)$, is $Gamma(\alpha, \theta(t))$.

2. Consider the data on weeks to death of 51 adults with recurrent gliomas, *A-astrocytoma* and *G-glioblastoma*, given in Assignment 1 (Problem No. 9). Compare the survival functions of the two types of gliomas using the log-rank test, and draw conclusions.

3. In a study of cancer of the tongue, patients were divided into two groups and the data on weeks until death were reported by Klein and Moeschberger (1997) as follows:

Atypical tumour

1	3	3	4	10	13	13	16	16	24	26	27
28	30	30	32	41	51	65	67	70	72	73	77
91	93	96	100	104	157	167					

Right censored observations

61	74	79	80	81	87	87	88	89	93	97	101
104	108	109	120	131	150	231	240	400			

Typical tumour

1	3	4	5	5	8	12	13	18	23	26	27
30	42	56	62	69	104	104	112	129	181		

Right censored observations

8	67	76	104	176	231

(a) Fit a Weibull distribution to each group of patients separately. Estimate the parameters η and α and their standard errors;

(b) In each group, test the hypothesis that the shape parameter η is 1 (i.e., suitability of the exponential distribution), using the Wald test and also using the likelihood ratio test;

(c) Find the maximum likelihood estimate of the median survival time in each group of patients;

(d) Fit a Weibull regression model to the whole data, using one covariate which takes the value 1 for patients in the atypical tumour group and 0 for patients belonging to the typical tumour group. Then, test the hypothesis of no difference between these two groups;

(e) Test the difference between the two groups using a Weibull regression model.

4. Besides the data presented in the last problem, there are two further groups of patients who had an extra stage of therapy, and these data are as follows:

NHL	28	32	49	84	357	933*	1078*	1183*	1560*	2114*	2144*
HL	2	4	72	77	79			.	.		.

In the above, * denotes right censored observations.

Then, fit a Weibull regression model with two covariates (disease – NHL/HL, and therapy – one stage/two stages), and test the effect of therapy on survival of patients.

5. Consider the frailty proportional hazards model of the form

$$h(t \mid y) = y h_0(t) e^{\boldsymbol{\beta}^T \mathbf{x}},$$

where $h(t \mid y)$ is the conditional hazard function of lifetime T, given y, y denotes the unobserved frailty, $\mathbf{x}$ is the associated set of covariates, and $\boldsymbol{\beta}$ is the set of regression coefficients.

Then, by considering the situation in which you have observed n observations, $(t_i, \delta_i, \mathbf{x}_i),\ i = 1, \ldots, n$, from the above frailty model, derive the likelihood function in an explicit form for the observed data.

6. Consider the "*catheter infection data*" given by McGilchrist and Aisbett (1991, *Biometrics*, Vol. **47**, Pages 461-466), dealing with time from insertion of a catheter into dialysis patients until it has to be removed due to infection (some were removed for other reasons, resulting in censoring).

   Restricting the analysis just to *Age* and *Sex* as covariates, and treating the two failure times of every individual as independent, a Cox hazard model can be fitted to the data.

   Also, to account for heterogeneity within the observed failure times, a *gamma frailty model* can be fitted to the data.

   (a) Present the results of these two model fits, giving estimates of model parameters, standard errors, etc.;

   (b) Compare and comment on the suitability of gamma frailty model in its ability to account for the unobserved heterogeneity in the observed failure times.

7. A breast cancer study, reported by Hajan-Tilaki et al. (2011), presented the following test results for various cut-off points of *BMI* (kg/m^2) for a number of breast cancer patients:

	($n = 100$) Breast cancer patients		($n = 200$) Normal subjects	
Cut-off	True positive	False negative	False positive	True negative
18	100	0	200	0
20	100	0	198	2
22	99	1	177	23
24	95	5	117	83
26	85	15	80	120
28	66	34	53	147
30	47	53	27	173
32	34	66	17	183
34	21	79	14	186
36	17	83	6	194
38	7	93	4	196
40	1	99	1	199

   (a) Calculate various indices of accuracy and inaccuracy, and comment;

   (b) Comment on the relationship between sensitivity and specificity;

(c) Draw the corresponding ROC curve and provide an estimate of AUC (you may either use empirical method from the above or use a nonparametric smoothing method in R or in SAS), and present an approximate 95% confidence interval for it;

(d) Comment on the ability of *BMI* as a biomarker in terms of its predictive ability to discriminate breast cancer patients from normal subjects.

<u>**Good luck!**</u>