End-Semester Examination

M.Stat. II Year
Academic Year: 2024 – 25
Survival Analysis
Duration: 3 hours
Total Marks: 100

Instructions:

- This is a closed-book pen-paper based descriptive answer-type examination.
- Students are allowed to use scientific calculators.
- Mobile phones are not allowed. Any electronic device that connects to internet are not allowed.
- Any notation will have its usual meaning unless otherwise specified.
- Answer any 5 out of the 6 questions.
- Please mention and describe a result or theorem or lemma clearly and completely in the event of applying it to a specific problem.
- You are allowed to use just one A4 size cheat sheet.
- Show intermediate steps (whenever applicable) to obtain full marks.

Questions:

Q.1. Let T (in years) be a non-negative random failure time with density function (parameters $\mu > 0$ and $\lambda > 0$) given by

$$f(t;\mu,\lambda) = \frac{\mu t^{\mu-1}\lambda}{(1+\lambda t^{\mu})^2}, \quad t > 0. \tag{1}$$

Based on this, answer Questions 1(a) - 1(d).

(a) [**5**] Represent T in the log-linear form explaining every component.

(b) [**2**] Let z (dimension 1) be the observed covariate. Write down the expression for accelerated failure time (AFT) regression model based on a sample of size n observed from a population having the density function as given in Equation (1).

(c) [**8**] Consider data of the form $\{(t_i, \delta_i, z_i)^\intercal : i = 1, \ldots, n\}$ follows an AFT model where δ_i denotes the type-I right censoring indicator and z_i (dimension 1) is the covariate associated with individual i. Show that the maximum likelihood estimate $\hat{\beta}$ of β can be obtained by solving

$$\sum_{i=1}^{n} z_i \left[\frac{\delta_i - \exp\{(y_i - \alpha_0 - \beta z_i)/\sigma_0\}}{1 + \exp\{(y_i - \alpha_0 - \beta z_i)/\sigma_0\}}\right] = 0$$

for given $\alpha_0 = -\frac{\log\{\lambda_0\}}{\mu_0}$ and $\sigma_0 = \frac{1}{\mu_0}$.

(d) [**5**] Let $\alpha_0 = 1.5$, $\hat{\beta} = 0.5$ and $\sigma_0 = 2$. Let $z = 1$ for an individual. Find the probability of her survival beyond 5 years.

Q.2. In a clinical trial, a new experimental drug (Drug X) is being tested for greater time to relapse (in months) of smoking after quitting. The group of patients who receive Drug X is the treatment group while there are two control groups - patients who receive an already marketed drug (Drug M) and patients who receive placebo (Drug P). The times to relapse for these three groups are given below:

- Drug X: 23^+, 17
- Drug M: 16, 22^+
- Drug P: 9, 16.

where $.^+$ indicates a right-censored observation. Please perform the following:

(a) [**12**] Perform a generalized 3-sample Gehan test for the data. Clearly mention the null and the alternative hypotheses, write down the test statistic (properly defining every component) and calculate its value. Mention the asymptotic null distribution of the test statistic.

of rank vectors has the form

$$\tilde{V} = \sum_{i=1}^{k} c_i z_{(i)} + \sum_{i=1}^{k} \sum_{j=1}^{m_i} c_i^* z_{ij},$$

where $z_{(i)}$ is the covariate associated with $y_{(i)}$ for $i = 1, \ldots, k$ and z_{ij} is the covariate associated with y_{ij} for $i = 0, 1, \ldots, k$ and $j = 1, \ldots, m_i$. Based on this, answer Questions 4(a) - 4(c).

(a) [**8**] Derive analytical expressions for c_i and c_i^* for $i = 1, \ldots, k$. State clearly every intermediate result and notation that you use. You don't have to prove the intermediate results but only mentioning the appropriate ones correctly should be fine.

(b) [**7**] Considering an uncensored data in the above set-up, i.e., considering only $y_{(i)}, i = 1, \ldots, k$, derive an analytical expression for $c_i, i = 1, \ldots, k$. State clearly every intermediate result and notation that you use. You don't have to prove the intermediate results but only mentioning the appropriate ones correctly should be fine.

(c) [**5**] For the i-th order statistic $U_{(i)}, i = 1, \ldots, k$ from a Uniform(0, 1) random sample, prove that:

$$J(1) = 1,$$

where

$$J(g(U_{(i)})) = \int \cdots \int_{u_{(1)} < \cdots < u_{(k)}} g(u_{(i)}) \left[\prod_{j=1}^{k} n_j (1 - u_{(j)})^{m_j} du_{(j)} \right] \qquad (2)$$

and $n_j = \sum_{\ell=j}^{k} (m_\ell + 1)$.

Q.5. Consider a type-I right censored data of the form:

$$\{(t_i, \delta_i)^\intercal : (2,1)^\intercal, (2,1)^\intercal, (2,1)^\intercal, (3,1)^\intercal, (3,1)^\intercal,$$

$$(3,0)^\intercal, (3,1)^\intercal, (4,1)^\intercal, (4,0)^\intercal, (5,1)^\intercal\}$$

where t_i and δ_i have their usual contextual definitions. Based on this, answer Questions 5(a) - 5(b).

(a) [8] Find the Kaplan-Meier estimates of the survival probability at distinct event times and corresponding Greenwood's standard errors.

(b) [7+5] To the above given type-I right censored data, fit a parametric survival model where the hazard function is $h(t) = \lambda t,\ t \geq 0,\ \lambda > 0$. Find out the maximum likelihood estimate of λ. Hence, perform a test of hypothesis $H_0 : \lambda = 0.5$ vs. $\lambda \neq 0.5$ at 5% level of significance, where $\chi^2_{1;0.95} = 3.8414$ and $Z_{0.975} = 1.9599$. You may test with any relevant test statistic of your choice.

Q.6. (a) [14] Consider a type-I right censored data of the form:

$$\{(t_i, \delta_i, z_i)^\top : (2, 1, 1)^\top, (2, 0, 0)^\top, (3, 1, 0)^\top, (4, 1, 1)^\top, (4, 0, 1)^\top\},$$

where t_i, δ_i and z_i have their usual contextual meanings. Calculate $r_i, i = 1, \ldots, 5$ where r_i is the i-th Cox-Snell residual based on fitting a Cox's proportional hazards model. Hence, check graphically whether the consideration of proportional hazards assumption is correct or not. The graph need not be very accurate, however, the reasoning behind your interpretation based on the graph should be correct for obtaining full marks.

(b) [6] Prove that the Weibull class of distribution is the only one which satisfies the properties of both proportional hazards and accelerated failure time models.

Finally, provide the conclusion of the test at 5% level where $\chi^2_{2:0.95} = 5.9914$.

(b) [**8**] Perform a 2-sample Mantel-Haenszel (ordinary log-rank) test between the groups receiving Drug X and Drug M. Clearly mention the null and the alternative hypotheses, write down the test statistic (properly defining every component) and calculate its value. Mention the symptotic null distribution of the test statistic. Finally, provide the conclusion of the test at 5% level of significance where $\chi^2_{1:0.95} = 3.8414$ and $z_{0.975} = 1.9599$.

Q.3. Suppose

$$\{(t_i, \delta_i, z_i)^\top : (2.0, 1, 1)^\top, (2.5, 1, -1)^\top, (3.0, 0, 1)^\top, (4.0, 0, 1)^\top, (4.5, 1, -1)^\top\}$$

be the data (with ties) obtained from a type-I right censoring scheme. Assuming a Cox's proportional hazards (PH) model

$$h(t; z) = h_0(t)e^{\beta z},$$

answer the following:

(a) [**6**] Estimate the maximum partial likelihood estimate of β.

(c) [**6**] Provide Breslow's estimate for the baseline cumulative hazard function.

(d) [**8**] Provide Miller's least square estimate for β. Performing just one full iteration is fine.

Q.4. Assume a right-censored data from accelerated failure time (AFT) model, i.e., $Y = \log\{T\} = \alpha + \beta z + W$ where

$$W \sim f(w) = \frac{e^w}{(1 + e^w)^2}, \quad -\infty < w < \infty.$$

Let $y_{(1)} < \cdots < y_{(k)}$ be k distinct ordered log-failure times and $y_{i1}, \ldots, y_{im_i}$ be m_i right-censored observations in the interval $[y_{(i)}, y_{(i+1)})$, $i = 0, 1, \ldots, k$ where $y_{(0)} = -\infty$ and $y_{(k+1)} = \infty$. Our objective is to test $H_0 : \beta = 0$ vs. $H_1 : \beta \neq 0$. A linear-rank test statistic using permutation distribution