LAIKIPIA UNIVERSITY

UNIVERSITY EXAMINATIONS

FIRST SEMESTER 2025/2026 ACADEMIC YEAR

FOURTH YEAR EXAMINATION FOR THE DEGREE OF BACHELOR OF SCIENCE (STATISTICS)

STAT 417: SURVIVAL MODELS AND ANALYSIS

STREAM: R ***TIME: 2 HRS***

DAY: FRIDAY [11.30 – 13.30 P.M] ***DATE: 06/02/2026***

THIS QUESTION PAPER CONSISTS OF SIX (6) PAGES

PLEASE DO NOT OPEN UNTIL THE INVIGILATOR SAYS SO.

INSTRUCTIONS: ANSWER QUESTION ONE AND ANY OTHER TWO QUESTIONS

QUESTION ONE (30 MARKS)

a) The discrete random variable T has times T=0, 1, 2, 3 … with corresponding survival functions S_j and probability mass function f_j .

i. Show that the expected time of failure E(T) can be given by $\sum_{j>0} S_j$. **(4 Marks)**

ii. During one July it was very cold, and one day 25 students reported to Laikipia University hospital with flu virus. Two students had only a mild dose, and were discharged immediately with medicine, but the other had to be admitted. The record number of days each student stayed in the hospital is given by the random variable T, as shown below

No. of days in hospital t	0	1	2	3	4	5
No. of students who were in hospital for t days	2	1	3	5	6	8

Determine the PMF f(t), the survival function S(t), the hazard function h(t) and the expectation E(T). **(6 Marks)**

b) For a continuous random variable T, the hazard function h(t) is given by:

$$h(t) = \begin{cases} \alpha\beta(\alpha t)^{\beta-1}, & for\ t > 0 \\ 0, & elsewhere \end{cases}$$

Determine an expression for the integrated hazard function H(t), the survival function S(t) and the PDF f(t) **(4 Marks)**

c) Suppose that in a study of the efficacy of a new drug, 20 mice with tumours are given the drug. The experimenter decides to terminate the study after 15 mice have died. The survival times are, in weeks;

5	6	6	7	10	15	18	18	18	21
23	27	31	34	55	55+	55+	55+	55+	55+

Assume that the times to death of these mice follow the lognormal distribution, estimate the mean and standard deviation of the survival time. **(6 Marks)**

d) The time taken by patients to wake up after being anaesthetised during minor surgeries can be modelled by random variable X (hours) with the probability density function f(x) being given by:

$$f(x) = \begin{cases} \sec^2(x), & \text{for } 0 \le x \le \frac{\pi}{4} \\ 0, & elsewhere \end{cases}$$

i. Determine an expression from survival function S(x) **(1 Mark)**

ii. Hence use your result to show that a patient takes $\pi/6$hours to wake up will be given by $\frac{3-\sqrt{3}}{3}$. **(2 Marks)**

iii. Also determine correct to two decimal places the probability that the wake-up time will be between 12 and 24 minutes. **(3 Marks)**

e) Given the following proportional hazards regression model for mortality of a sample of life assurance policy holders.

$$h_i(t) = h_0(t)exp\{0.01(x_i - 25) + y_i - 0.06z_i\}$$

where

$h_i(t)$ denotes the hazard function for life i at duration t,

$h_0(t)$ denotes the baseline hazard function at duration t,

x_i denotes the age of entry of life i

y_i = 1 if life i is a smoker, otherwise zero

z_i = 1if life i is female, zero if male

i. Determine the class of the lives to which the baseline hazard function applies **(2 Marks)**

ii. What does the model tell you about the relative risk of a male smoker aged 25 at entry compared to a male non-smoker aged 40 at entry? (Answer to 2 d.p.). **(3 Marks)**

iii. What does the model tell you about the relative risk of a female non-smoker aged 30 at entry compared to a male smoker aged 35 at entry? (Answer to 2 d.p.) **(2 Marks)**

QUESTION TWO (20 MARKS)

a) For a continuous variable X with parameter vector Φ, probability density function f(x, Φ), survival function S(x, Φ), and hazard function(x, Φ). Show that the logl-likelihood function l, based on a sample $(x_1, x_2, \ldots, x_n)$ of which $(x_1, x_2, \ldots, x_d)$ is uncensored and $(x_{d+1}, x_{d+2}, \ldots, x_n)$ is censored, is given by **(6 Marks)**

$$l = \sum_{i=1}^{d} \ln[h(x_i, \Phi] + \sum_{i=1}^{n} \ln[S(x_i, \Phi]$$

b) A doctor treating patients with bone cancer desires to find out whether a new drug, drug B, is more effective in increasing the time to remission of patients, than drug A, which he has been using. He took two random samples of patients of group 1 was treated with drug A, and group 2 with drug B. We know that the time to remission X days can be modelled by the PDF.

$$f(x) = \begin{cases} \beta e^{-\beta x}, & \text{for } x > 0 \\ 0, & elsewhere \end{cases}$$

i. Determine the CDF, the survival function and hazard function for the random variable X. **(5 Marks)**

ii. Show that if d of the times in the sample of 10 are uncensored, and the rest are censored, then the log likelihood function l can be expressed in the form

$$l = d\ ln\beta - \beta \sum_{i=1}^{10} x_i,$$

and hence determine an expression for the maximum likelihood estimate $\hat{\beta}$ of β. **(4 Marks)**

iii. The times to remission of two groups are recorded by doctor as follows

Group 1 (time in days) 19, 40, 11, 30*, 6, 14, 6, 15, 25, 6*

Group 2 (time in days) 10, 179, 53, 9, 141, 37, 2, 384*, 70, 44

*indicates a censored time.

Determine the maximum likelihood estimate of β for each sample of ten patients. Give each estimate correct to three decimal places. **(5 Marks)**

QUESTION THREE (20 MARKS)

A clinical trial to evaluate the efficacy of maintenance chemotherapy for acute myelogenous leukemia (AML) was conducted. The following table shows times of remission (i.e. freedom from symptoms in a precisely defined sense) of AML patients received chemotherapy

9, 13, 13*, 18, 23, 28*, 31, 34, 45*, 48, 161*.

Observations with star (*) are right censored.

a) Calculate the Kaplan-Meier estimate for the survival probability *S* (48). **(6 Marks)**

b) Find a 95% log-transformed confidence interval for *S* (48). **(8 Marks)**

c) Calculate the 95% CI for S (48) using the formula

$$Var[S(t)] = \left\{\left[\hat{S}(t)\right]^2\left[1 - \hat{S}(t)\right]\right\}/r(t)$$ **(6 Marks)**

QUESTION FOUR (20 MARKS)

a) Thirty melanoma patients (stages 2 to 4) were studied to compare the immunotherapies BCG (*Bacillus* Calmette-Guerin) – Treatment 1 and *Corynebacterium parvum* – Treatment 2 for their abilities to prolong remission duration. The data is shown below:

Patient	Treatment received	Remission duration	Patient	Treatment received	Remission duration
1	1	33.7+	16	2	16.0+
2	1	3.9	17	2	6.9
3	1	10.5	18	2	11.0+
4	1	5.4	19	2	24.8+
5	1	19.5	20	2	23.0+
6	1	23.8+	21	2	8.3
7	1	7.9	22	2	10.8+
8	1	16.9+	23	2	12.2+
9	1	16.6+	24	2	12.5+
10	1	33.7+	25	2	24.4
11	1	17.1+	26	2	7.7
12	2	8.0	27	2	14.8+
13	2	26.9+	28	2	8.2+
14	2	21.4+	29	2	8.2+
15	2	18.1+	30	2	7.8+

Compare the survival distributions of the two treatments using both the Log rank and generalized Wilcoxon tests **(10 Marks)**

b) Consider the following data of patients with cancer of the ovary

i. Perform a complete life-table analysis of the data **(6 Marks)**

ii. Plot the three survival functions **(4 Marks)**

Year after diagnosis	Number entering interval	Number lost to follow-up	Number dying
0-1	718	0	132
1-2	568	8	80
2-3	468	8	57
3-4	398	7	52
4-5	331	7	58
5-6	264	12	45
6-7	198	11	32
7-8	151	12	20
8-9	116	14	11
9-10	89	7	14
10-11	67	2	13
11-12	52	3	11
12-13	36	2	8
13-14	23	3	6
14-15	13	4	5
15+	2	1	1

QUESTION FIVE (20 MARKS)

a) The data below are remission times in weeks for a group of 30 patients with a disease who received a similar treatment: 1, 1, 2, 4, 4, 6, 6, 6, 7, 8, 9, 9, 10, 12, 13, 14, 18, 19, 24, 26, 29, 31^+, 42, 45^+, 50^+, 57, 60, 71^+, 85^+, 91

   i) Obtain and plot the K-M estimate of the survivor function for the remission time. **(10 Marks)**

   ii) Obtain the 95% confidence Interval for the median remission time. **(5 Marks)**

   iii) Determine the 95% confidence interval for the probability that remission lasts over 26 weeks. **(5 Marks)**