MERU UNIVERSITY OF SCIENCE AND TECHNOLOGY

P.O. Box 972-60200 – Meru-Kenya.
Tel: +254(0) 799 529 958, +254(0) 799 529 959, +254 (0)712 524 293
Website: www.must.ac.ke Email: info@mucst.ac.ke

UNIVERSITY EXAMINATIONS 2023/2024

FIRST YEAR FIRST SEMESTER AND FIRST YEAR SECOND SEMESTER
EXAMINATION FOR DEGREE OF MASTERS OF SCIENCE IN APPLIED STATISTICS

SMS 5178: SURVIVAL AND CLINICAL DATA ANALYSIS

DATE: APRIL 2023 **TIME: 2 HOURS**

INSTRUCTIONS: Answer Question ONE and any other THREE questions.

QUESTION ONE (30MARKS)

a) Briefly explain the following
   i. Survival Analysis
   ii. Hazard function
   iii. Censoring.

b) Distinguish between right and left censoring giving examples in each case. (4marks)

c) Show that the probability that an individual lives longer that $t_1 + t_2$ years given has attained t_1 years is equal to the unconditional probability that he survives least t_2 years in and only if the survival distribution is of exponential form.

d) Show that if the hazard function is given by $\lambda(t) = \alpha\beta(\alpha t)^{\beta-1}\exp\{(\alpha t)^{\beta}\}$ Then the survival function will be $s(t) = \exp\{-[\exp(\alpha t)^{\beta} - 1]\}$

e) The random variable T represents the application of medication, for pimples to disappear from a sensitive teenager's face. The probability mass function $f(t)$ is given below.

t	0	1	2	3	4	5
F(t)	0.449	0.359	0.144	0.038	0.008	0.001

Determine;

i. $S(t)$

ii. $h(t)$

iii. $E(t)$

f) For a random variable X, the probability density function $f(x)$ is given by $f(x) = sinx$ for $0 \leq X \leq \frac{n}{2}$ and $f(x) = 0$ elsewhere. Determine the expressions for the survival function and the hazard function and show that the integrated hazard function is Insec (X) (6 marks)

g) Using a figure describe the survival function (2marks)

QUESTION TWO (20 MARKS)

The table below gives the number r_j enrolled at the start of the year in a school, the number d_j who dropped out of at the end of the year and c_j who absconded during the year for each year. Time in school (in years)

Time in school (in years)	r_j	d_j	c_j
$0 \leq x < 1$	150	6	5
$1 \leq x < 2$		10	3
$2 \leq x < 3$		2	7
$3 \leq x < 4$		7	2

Using this information as a representative sample , determine the Kaplan Meier estimator of the survivor function $S_{(x)}$ over each interval and also the 95% confidence interval for $S_{(x)}$ of x just after the interval $1 \leq x < 2$ using Greenwood formula, and also for the x just after the interval $2 \leq x < 3$ using the formula

$$\text{var}[s(x)] = \frac{\{[\hat{s}(x)]^2[1-\hat{s}(x)]\}}{r(x)}$$

(12 marks)

b) Let T be a continuous random variable with density given by $f(t) = \begin{cases} \lambda e^{-2t}, t > 0 \\ 0 \; elsewhere \end{cases}$

i) Find the expression for the P[th] quantile (4marks)

ii) Hence or otherwise determine the median life time in terms of λ (4marks)

OUESTION THREE (20 MARKS)

a. The times (in months) taken by customers from two different banks to repay their loans are shown below (censored times shown with + for those customers who defaulted).

Bank A: 47 42 25 49+ 95+ 75

Bank B: 32 89 87 55+ 63+ 91

i. Show how R-software can be used to calculate the survival function for customers in Bank A (Show the steps and codes) (5 marks)

ii. Use the log-rank test at 5 % level to determine whether there is a significant. difference between the customers in the two banks in the times taken to repay their loans. (9 marks)

b. Given the survival function S(x) below

$$S(x) = \begin{cases} 1 - \frac{x}{100}, & 0 < x \leq 25 \\ \frac{1}{x^{1.5}}, & x > 25 \end{cases}$$

k is constant

i. Determine the value of *k* (3 marks)

ii. Determine the life expectancy (3 marks)

QUESTION FOUR (20 MARKS)

a. The table below gives some details of fitting a proportional hazards regression model to times to recurrence of a certain disease. The data were obtained during a randomized clinical trial of a new treatment. The factors investigated were treatment (coded by for placebo xy=l for treatment) stage of disease (coded by x_2=0) for stage 1, x_2=l for stage II x_2=2 for stage III) and the interaction between treatment and stage of disease coded by x_3.

	Variable	Coefficient	Standard error
Treatment	X_1	-0.18	0.10
Stage	X_2	+0.32	0.21
interaction	x_3	-0.66	0.1 1

i. Specify the form of the proportional hazards model used for this analysis interms of baseline Lazard function $h_0(t)$ and the covariates. (5 marks)

ii. Describe in detail the effects of those factors on the time to recurrence of the disease. (6 marks)

b. Consider a study carried out to evaluate the effect of a drug for high blood pressure. Elders volunteers are randomly allocated to one of the three treatments; a placebo, the drug that is currently most common and a new drug that has been developed. The study last for ten years and the event of interest is occurrence of any kind of heart disease including death resulting from heart disease. Briefly describe three situations from this study that might lead to an observation that is right censored. [6 marks]

c. List three reasons for modeling survival data [3 marks]

OUESTION FIVE (20 MARKS)

A clinical trial to evaluate the efficiency of maintained chemotherapy for acute myelogenous leukemia (AMLO was conducted. After reaching a state of remission through treatment by chemotherapy, the patient who entered the study were randomized into two groups. The first group received maintainance chemotherapy. The second group or control group did not. The objective of the trial was to see if maintenance chemotherapy prolonged the time until relapse i.e for a preliminary during the course of the trial the date was as follows.

maintained group

9,13,13+,18,23,28+,31,34,45+ 48,16+

Non —maintained group

5, 5, 8, 8, 12, 16,+, 23, 27, 30, 33, 43, 45

a. Using the Wilcoxon test, compare the survival function for the two group. (14 marks)

b. Show how the test in a above can be done using R-software (6 marks)