

Kasarani Campus Off Thika Road P. O. Box 49274, 00101 NAIROBI Westlands Campus Pamstech House Woodvale Grove Tel. 4442212 Fax: 4444175

KIRIRI WOMENS' UNIVERSITY OF SCIENCE AND TECHNOLOGY UNIVERSITY EXAMINATIONS, 2024/2025 ACADEMIC YEAR FOURTH YEAR, SECOND SEMESTER EXAMINATION FOR THE DEGREE OF BACHELOR OF SCIENCE IN MATHEMATICS

KMA 2425: SURVIVAL AND CLINICAL DATA ANALYSIS

DATE: 13TH DECEMBER, 2024 TIME: 8:30AM-10:30 AM

<u>INSTRUCTIONS TO CANDIDATES</u> <u>ANSWER QUESTION ONE (COMPULSORY)</u> AND ANY OTHER TWO QUESTIONS

QUESTION ONE: COMPULSORY (30 MARKS)

a) Define the following as used in Survival Analysis:

	i. Survival function, $S(t)$	(1 Mark)						
	ii. Hazard function, $h(t)$	(1 Mark)						
	iii. Probability density function, $f(t)$	(1 Mark)						
	iv. Write down a mathematical expression illustrating how the three functions, $S(t)$, $h(t)$ and j							
	related	(3 Marks)						
b)	Discuss the relevance of survival analysis	(3 Marks)						

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

9, 13, 13+, 18, 23, 28+, 31, 34, 45+, 48, 161+

Observations with '+' are censored. Calculate the Kaplan-Meier estimate for the survival probability S(34)

(4 Marks)

d) The National Centre for Health Services Research studied 36 for-profit nursing homes to assess the effects of different financial incentives on length of stay. "Treated" nursing homes received higher perdiems for Medicaid patients, and bonuses for improving a patient's health and sending them home. The study included 1601 patients admitted between May 1, 1981 and April 30, 1982. List six variables that can be monitored?

e) i. What do you understand by the term Censoring? (1 Mark)

ii. Distinguish right censoring from left censoring.

- f) If the survival time T is exponentially distributed with parameter, θ , i.e $T \sim \exp(\theta)$, find the corresponding expressions for:
 - i. the survival function, S(t)

(4 Marks)

(2 Marks)

(6 Marks)

QUESTION TWO: (20 MARKS)

- a) In a study, the lifetime of high- voltage power transformers is of interest. An energy company began careful archival record keeping in 1980. The dataset contains complete information on all units that were installed after 1980 (i.e., the installation dates and date of failure for those that failed). There is also information on units that were installed before January 1, 1980 and failed after January 1, 1980. There is no information on units installed and failed before January 1, 1980.
 - i. What is the appropriate time scale for this study? (1 Mark)
 - ii. For four selected units described below, what types of censoring (if any) that are represented? And why?
 (4 Marks)
 - i. A unit installed in 1990, failed in 1998
 - ii. A unit installed in 1990, still in service in 2010
 - iii. A unit installed in 1979, failed in 2001
 - iv. A unit installed in 1950, still in service in 2010
- b) Consider a small study with 8 subjects. The event times were recorded as follows:
 - 1, 3, 2+, 3+, 2, 5+, 6+, 3
 - i. Calculate the Kaplan Meier estimates for the survival function S(t) (8 Marks)
 ii. a) Make a plot of the survival function based on the estimates obtained (5 Marks)
 b) Interpret the plot (2 Marks)

QUESTION THREE: (20 MARKS)

30 newly diagnosed lung cancer patients were followed for 1 year. Time to death in weeks were recorded as below

4, 5, 6, 7+, 8, 8, 10, 11, 11, 12, 14, 14, 14, 15+, 20, 21+, 24, 25, 35, 36, 37, 40, 41+, 42, 45, 46+, 47+, 50, 50, 51

- i. Estimate the Nelson-Aalen survival function from the data (7 Marks)
- ii. Compute the point-wise 95% confidence interval (13 Marks)

QUESTION FOUR: (20 MARKS)

a) The remission times of 42 patients with acute leukemia were recorded in a clinical trial to assess the ability of 6-mercaptopurine (6-MP) to maintain remission. Each patient was randomized to receive 6-MP or a placebo. The study was terminated after one year.

6-MP (21 patients)	6	6	6	7	10	13	16	22	23	6+	9+	10+	11+	17+	19+	20+	25+	32+	32+	34+	35+
Placebo (21patients)	1	1	2	2	3	4	4	5	5	8	8	8	8	11	11	12	12	15	17	22	23

Find the log-rank test statistic from the data hence establish whether the two survival data sets are statistically similar or not. (15 Marks)

b) Use the stata output below to answer the following questions



- i. Did the treatment make a difference in the survival experience of the two groups? (2 Marks)
- ii. Mention three types of tests would you perform to confirm your answer? (3 Marks)

QUESTION FIVE: (20 MARKS)

a) Let S(t) be the survival function of T at t. Show that

$$E(T) = \int_0^\infty S(t)dt$$

b) Consider a discrete random variable *T* for the future lifetime in days of a group of insects, with the survival function given as:

t	0	1	2	3	4	5	6
S(t)	1	0.8	0.6	0.4	0.2	0.1	0

Find;i) F(t), the cumulative distribution function of T(2 Marks)ii) P(T = t), the probability mass function of T(2 Marks)iii) E(T), the mean of T(3 Marks)

- c) State whether the following statements are True or False
 - i. In Survival Analysis, we are not interested in a non-negative random variable (1 Mark)
 - ii. Cumulative distribution function gives us the probability that an individual survives until time t.
 - (1 Mark)

(5 Marks)

- iii. There are many possible distributions to describe survival time data (1 Mark)
- iv. If there is no censoring, standard regression procedures could be used (1 Mark)
- d) Describe Greenwood's formula for the standard error of the K-M estimator (4 Marks)