DEDAN KIMATHI UNIVERSITY OF TECHNOLOGY
University Supplementary & Special Examinations 2021

FOURTH YEAR SEMESTER EXAMINATION FOR THE DEGREE OF BACHELOR OF SCIENCE IN **ACTUARIAL SCIENCE**

SAS 4101: SURVIVAL ANALYSIS

DATE: th February 2021 TIME:
Instructions: Answer QUESTION ONE and any other TWO QUESTIONS.

QUESTION ONE (30 marks) (COMPULSORY)

(a) Define the following as used in survival analysis

(i) Survivor function. **[1 mark]**

(ii) Hazard function. **[1 mark]**

(b) An investigation was undertaken into the length of post-operative stay in hospital after a particular type of surgery. All patients undergoing this surgery between 1^{st} and 31^{st} January 2013 were observed until either they left the hospital, died, or underwent a second operation. The event of interest was leaving the hospital. Patients who died or underwent a second operation during the period of investigation were treated as censored at the date of death or second operation respectively. The investigation ended on 28 February 2013, and patients who were still in the hospital at that time were treated as censored.
State, with reasons, whether the following types of censoring are present in this investigation:

- Type I
- Type II
- Random

Comment on whether censoring in this investigation is likely to be informative.

[4 marks]

(c) For a particular investigation the hazard of mortality is assumed to take the form:

$$\lambda(t) = a + bt$$

where a and b are constants and t represents time.
For each life i in the investigation ($i = 1, \ldots, n$) information was collected on the length of time the life was observed t_i and whether the life exited due to death ($d_i = 1$ if the life died, 0 otherwise).

(i) Show that the likelihood of the data is given by:

$$L = \prod_{i=1}^{n} (a + bt_i)^{d_i} \exp(-at - \frac{1}{2} bt_i^2)$$

[3 marks]

(ii) Derive two simultaneous equations from which the maximum likelihood estimates of the parameters a and b can be calculated. **[4 marks]**

(d) An Institute conducts tuition classes as part of their preparation for the professional exams. The management of the institute is concerned with the withdrawal rates of the students and hence it is testing a new tuition method to improve the persistency rates. Data have been collected and a Cox proportional hazards model has been fitted for the hazard of students leaving the course. Symmetric 95% confidence intervals (based upon standard errors) for the regression parameters are as shown below.

		Parameter Confidence Interval
Course	Engineering	0
	Medical	[0.08, 0.25]
Tuition method	Traditional	0
	New	[-0.05, 0.05]
Sex	Boys	[0.02, 0.12]
	Girls	0

(i) Write down the expression for the Cox proportional hazards model used, defining all terms that you use. **[4 marks]**

(ii) Describe the class of students to which the baseline hazard applies. **[1 mark]**

(iii) Discuss the suggestion that the new tuition method has improved the chances of students continuing with the tuition classes. **[2 marks]**

(e) The following are remission times(in weeks) of Leukemia patients.

6 9 13 13* 18 23 28 31 34 45* 48 56

where a right-censored observation is denoted by *. Assuming that the observation have exponential distribution with parameter μ, obtain the 95% confidence interval for the estimate of μ. **[6 marks]**

(f) If a right-censored observation is denoted by *, derive the Nelson-Aalen estimate of the cumulative hazard for the following data.

1* 3 4* 5 5 6* 7* 7 7

[4 marks]

QUESTION TWO (20 marks) (Optional)

(a) A random variable, T, has the Weibull distribution with probability density function

$$f(t) = \begin{cases} \lambda\gamma t^{\gamma-1}\exp(-\lambda t^{\gamma}), & t > 0 \quad \lambda > 0 \quad \lambda > 0 \\ 0, & \text{otherwise} \end{cases}$$

(i) Derive the survivor function, $S(t)$, of T. **[3 marks]**

(ii) Show that $\log\{-\log(S(t))\}$ is a linear function of $\log(t)$. State the intercept and slope of the line if $\log\{-\log(S(t))\}$ on the vertical axis is plotted against log(t) on the horizontal axis. **[4 marks]**

(b) 20 refrigerator motors of a particular type were each tested on an accelerated life test, and their times till first failure (hours) were recorded. The results are listed below, where * denotes that the motor was still functioning properly when the test was brought to an end.

2 4* 5 5 5 6* 7 7 7* 8 8 9* 11 11 12 12* 12* 12* 16 18*

(i) Use the Kaplan-Meier method with these data to estimate the survivor function, $S(t)$, for the time to first failure of a motor of this type. **[5 marks]**

(ii) Referring to the result from part (a)(ii) and (b)(i), use a suitable graphical method to investigate whether or not these data come from a Weibull distribution. **[5 marks]**

(iii) Draw a straight line through the points on your graph by eye and use it to estimate the parameters, λ and γ, of a Weibull distribution fitted to these data. **[3 marks]**

QUESTION THREE (20 marks) (Optional)

154 subjects with burns were monitored in a study of a new treatment to prevent burn wounds becoming infected. 70 of the subjects were given standard care (Treat = 0) while the other 84 had additional care thought to make infection less likely (Treat = 1). The time (in days) until the wound became infected was recorded; for 106 of the subjects, no infection was discovered during the period of follow up and for them total time in the study was treated as a censored survival time. Further information recorded about each subject included their Sex (Male = 0, Female = 1), Race (Non White = 0, White = 1) and Severity (the initial severity of their burns, in percent of body surface area).

(a) It was decided to fit a Cox proportional hazards model to the data, with Treat, Sex, Race and Severity as explanatory variables. Write down the form of this model, interpreting clearly each of the terms in it. **[5 marks]**

(b) The model was fitted and the results shown below were obtained.

	Coefficient	Standard Error
Treat	0.606	0.296
Sex	0.631	0.390
Race	2.12	1.01
Severity	0.00404	0.00703

(i) What can be concluded about the effects of the four explanatory variables? **[8 marks]**

(ii) Obtain and interpret a 95% confidence interval for the hazard ratio of a female subject given additional care compared to a female subject given standard care, assuming that the two subjects are white and have the same initial severity of burns. **[5 marks]**

(iii) Information was also recorded about the cause of the burns, which was characterised as either chemical (9 cases), scald (18), electric (11) or flame (116). Describe briefly how you would extend the model in order to make full use of this new information. **[2 marks]**

QUESTION FOUR (20 marks) (Optional)

(a) State and explain any two features which are desirable when a graduation is performed. **[4 marks]**

(b) The actuary to a large pension scheme has attempted to graduate the schemes recent mortality experience with reference to a table used for similar sized schemes in a different industry. He has calculated the standardized deviations between the crude and the graduated rates, z_x, at each age and has sent you a printout of the

figures over a small range of ages. Unfortunately the dot matrix printer on which he printed the results was very old and the dots which would form the minus sign in front of numbers no longer function, so you cannot tell which of the standardized deviations is positive and which negative. Below are the data which you have.

Age	60	61	62	63	64	65	66	67	68	69	70
z_x	2.40	0.08	0.80	0.76	1.04	0.77	1.30	1.76	0.28	0.68	0.93

(i) Carry out an overall goodness-of-fit test on the data. Comment on your result. **[5 marks]**

(ii) List four defects of a graduation which the test you have carried out would fail to detect. For each of the defects, suggest a test which could be used to detect it. **[8 marks]**

(iii) Carry out one of the tests suggested in part(ii). **[3 marks]**

QUESTION FIVE (20 marks)(Optional)

A study was made of a group of people seeking jobs. 700 people who were just starting to look for work were followed for a period of eight months in a series of interviews after exactly one month, two months, etc. If the job seeker found a job during a month, the job was assumed to have started at the end of the month. Unfortunately, the study was unable to maintain contact with all the job seekers.
The data from the study are shown in the table below:

Months since start of study	Found employment	Contact lost
1	100	50
2	70	0
3	50	20
4	40	20
5	20	30
6	20	60
7	12	38
8	6	0

(i) Describe two types of censoring present in the investigation and an example of a person to whom each type applies. **[4 marks]**

(ii) Calculate the Nelson-Aalen estimate of the integrated hazard for the job seekers and hence give estimate of the survival function. **[7 marks]**

(iii) Estimate the variance of Nelson-Aalen estimate at the fourth month and hence its 95% confidence interval **[4 marks]**

(iv) Sketch a graph of the estimated survival function. **[3 marks]**

(v) Estimate the probability that a person will be employed in the fifth month. **[2 marks]**