DEDAN KIMATHI UNIVERSITY OF TECHNOLOGY

University Examinations 2015/2016

FOURTH YEAR SPECIAL/SUPPLEMNTARY EXAMINATION FOR THE DEGREE OF BACHELOR OF SCIENCE IN **ACTUARIAL SCIENCE**

STA 2493: SURVIVAL ANALYSIS

DATE: 14$^{\text{th}}$ March 2016 TIME: 8:30 AM - 10:30 AM

Instructions: Answer QUESTION ONE and any other TWO QUESTIONS.

QUESTION ONE (30 Marks) (COMPULSORY)

(a) Define the following

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

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

(iii) Right censoring. **[1 mark]**

(iv) Left censoring. **[1 mark]**

(b) Consider a discrete random variable T taking the values $0, 1, 2, \ldots,$. Obtain $E(T)$ in terms of the survival function and hence find the mean of the random variable whose survival function given by

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

[3 marks]

(c) Consider a continuous random variable T whose p.d.f is

$$f(t) = \alpha\beta(\beta t)^{\alpha-1}\exp[(-\beta t)^{\alpha}] \qquad t \geq 0$$

Obtain the hazard function of T^{α} **[5 marks]**

(d) A study of the mortality of a certain species of insect reveals that for the first 30 days of life, the insects are subject to a constant force of mortality of 0.05. After 30 days, the force of mortality increases according to the formula

$$\mu_{30+x} = 0.05\exp(0.01x)$$

where x is the number of days after day 30. Calculate

(i) The probability that a newly born insect will survive for at least 10 days. **[2 marks]**

(ii) The probability that an insect aged 10 days will survive for at least a further 30 days. **[3 marks]**

(iii) The age in days by which 90 per cent of insects are expected to have died. **[5 marks]**

(e) Losses from arising from a group of policies are assumed to follow exponential distribution with parameter θ. The following data is available.

- The exact amounts $x_1, x_2, \ldots, x_n$ paid by the insurers in respect to the n losses
- a further m losses in respect of which the insurer paid an amount M. The actual loss amount exceeds M but it is not known by how much.

Obtain the m.l.e. of θ. **[5 marks]**

(f) (i) Write down the equation of the Cox proportional hazards model in which the hazard function depends on duration t and a vector of covariates z. You should define all the other terms that you use. **[2 marks]**

(ii) Why is the Cox model sometimes described as semi-parametric . **[1 mark]**

QUESTION TWO (20 marks) (Optional)

(a) A lecturer at a university gives a course on Survival Models consisting of 8 lectures. 50 students initially register for the course and all attend the first lecture, but as the course proceeds the numbers attending lectures gradually fall. Some students switch to another course. Others intend to sit the Survival Models examination but simply stop attending lectures because they are so boring. In this university, students who decide not to attend a lecture are not permitted to attend any subsequent lectures. The table below gives the number of students switching courses
The university's Teaching Quality Assurance Service has devised an Index of Lecture Boringness. This index is defined as the Kaplan-Meier estimate of the proportion of students remaining registered for the course who attend the final lecture. In calculating the Index, students who switch courses are to be treated as censored after the last lecture they attend.

(i) Calculate the Index of Lecture Boringness for the Survival Models course. **[6 marks]**

(ii) What type of censoring is in this example?. Explain. **[3 marks]**

Lecture Number	Number of students switching courses	Number of students ceasing to attend lectures but remaining registered for Survival Models
1	5	1
2	3	0
3	2	3
4	0	1
5	0	2
6	0	1
7	0	0

(iii) Estimate the variance of Lecture Boringness Index and hence the 95% confidence interval **[5 marks]**

(b) The hazard of a life aged x is $\lambda_x = BC^x$, show that its survival function

$$S_x(t) = \left[exp\left(\frac{-B}{\ln C}\right)\right]^{C^x(C^t-1)}.$$

If $x = 50$, $S_x(1) = 0.995$ and $S_x(2) = 0.989$ find B and C **[6 marks]**

QUESTION THREE (20 marks) (Optional)

(a) An investigation was carried out into the effects of lifestyle factors on the mortality of people aged between 50 and 65 years. The investigation took the form of a prospective study following a sample of several hundred individuals from their 50th birthdays until their 65th birthdays and collecting data on the following covariates for each person
X_1 Sex (a categorical variable with 0 = female, 1 = male)
X_2 Cigarette smoking (a categorical variable with 0 = non-smoker, 1 = smoker)
X_3 Alcohol consumption (a categorical variable with 0 = consumes fewer than 21 units of alcohol per week, 1 = consumes 21 or more units of alcohol per week)
In addition, data were collected on the age at death for persons who died during the period of investigation. In order to analyse the data, it was decided to use a Gompertz hazard, $\lambda_x = BC^x$, where x is the duration since the start of the observation.

(i) Show that the substitution $B = exp(\beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3)$, in the Gompertz model (where $\beta_0, \ldots, \beta_3$ are parameters to be estimated), leads to a proportional hazards model for this particular analysis. **[3 marks]**

(ii) Using the Gompertz hazard, the parameter estimates in the proportional hazards model were as follows

(a) Write the baseline hazard and describe the characteristics of the person to whom it applies to in the model. **[2 marks]**

(b) Calculate the estimated hazard for a female cigarette smoker aged 55 years who does not consume alcohol. **[2 marks]**

Covariate	Parameter estimate	Parameter
Sex	β_1	+0.40
Cigarette smoking	β_2	+0.75
Alcohol consumption	β_3	-0.20
	β_0	-5.00
	C	+1.10

(c) Show that, according to this model, a cigarette smoker at any age has a risk of death roughly equal to that of a non-smoker aged eight years older. **[6 marks]**

(b) A group of six lives was observed over a period of time as part of a mortality investigation. Each life was observed at all ages from 55 until they died or were censored. The table shows the sex, age at exit and reason for exit. The model $\lambda(x; z) = \lambda_0 \exp(\beta z)$ has been suggested. where x is the age, λ_0 is the baseline hazard and $z = 0$ for male, $z = 1$ for females.

Life	**Sex**	**Age at exit**	**Reason for exit**
1	M	56	death
2	M	56	censored
3	F	62	censored
4	F	62	censored
5	F	63	death
6	M	66	death
7	M	67	censored
8	M	67	censored

Obtain the maximum likelihood estimate of β. **[7 marks]**

QUESTION FOUR (20 Marks) (Optional)

A life assurance company carried out an investigation of the mortality of male life assurance policyholders. The investigation followed a group of 100 policyholders from their 60th birthday until their 65th birthday, or until they died or cancelled their policy (whichever event occurred first).
The ages at which policyholders died or cancelled their policies were as follows:

(i) Identify and explain which types of censoring are present in the investigation. **[4 marks]**

(ii) Calculate the Nelson-Aalen estimate of the integrated hazard for these policyholders **[6 marks]**

(iii) Estimate the variance of Nelson-Aalen estimate at the age of 63 and hence its 95% confidence interval **[4 marks]**

Died	Cancelled Policy
Age in years and months	Age in years and months
60y 5m	60y 2m
61y 1m	60y 3m
62y 6m	60y 8m
63y 0m	61y 0m
63y 0m	61y 0m
63y 8m	61y 0m
64y 3m	61y 5m
	62y 2m
	62y 9m
	63y 9m
	64y 5m

(iv) Sketch the estimated integrated hazard function. **[3 marks]**

(v) Estimate the probability that a policyholder will survive to age 65. **[3 marks]**

QUESTION FIVE (20 Marks)(Optional)

(a) Describe three shortcomings of the χ^2 test for comparing crude estimates of mortality with a standard table and why they may occur **[3 marks]**

(b) The following table gives an extract of data from a mortality investigation conducted in the rural highlands of a developed country. The raw data have been graduated by reference to a standard mortality table of assured lives.

Age x	Expected death	Observed death	z_x
60	36.15	35	-0.191
61	28.92	24	-0.915
62	31.34	27	-0.775
63	38.01	35	-0.488
64	26.88	32	0.988
65	37.59	36	-0.259
66	33.85	34	0.026
67	26.66	32	1.034
68	22.37	26	0.767
69	18.69	33	3.310
70	18.24	22	0.880

Carry out each of the following tests at 5% on the data above

(i) Individual standardized deviations test **[4 marks]**

(ii) Cumulative deviations test **[5 marks]**

(iii) Serial correlations test **[8 marks]**