DEDAN KIMATHI UNIVERSITY OF TECHNOLOGY
University Examinations 2017/2018

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

STA 2493: SURVIVAL ANALYSIS

DATE: 21st August 2017 TIME: 2:00-4:00 p.m.

Instructions: Answer QUESTION ONE and any other TWO QUESTIONS.

QUESTION ONE (30 marks) (COMPULSORY)

(a) An investigation into whether the rate of surrender of policies of an insurance company varies with their duration is to be conducted. The company's records show the policy issue date, the calendar year in which policy is surrendered, policy maturity date, date of exit and in case of exits due to any other reasons.
In the context of this investigation consider the following types of censoring:- Left, Right, Interval and Informative . For each type of censoring above;

(i) Describe how it occurs. **[4 marks]**

(ii) State whether or not that the particular type of censoring is present in these data. **[2 marks]**

(iii) If that particular type of censoring is present, explain how it arises. **[3 marks]**

(b) The survivor function of a random variable is $S(t) = \exp\{-[\exp(\alpha t)^{\beta} - 1]\}$. Find an approximation to the probability that a subject with this survivor function fails within the interval (1.5,2.1) given survival to point 1.5 . **[4 marks]**

(c) The following data shows the period in complete months from the initial ill health retirement to the end of observation for those members who died or withdraw with a special permission from the observation before the end of the investigation of two years.
12, 5*, 6, 15, 1*, 18*, 20, 6, 3*, 20, 10*, 23, 8*.
A censored observation is denoted by *.

(i) Compute the Kaplan-Meier estimate of the survivor function and plot the results. **[7 marks]**

(ii) If survival times have an exponential distribution, estimate the 95% confidence interval for the parameter. **[4 marks]**

(d) Write down the proportional hazards model with time-independent covariates Z, and interpret each term in your model. Define the *hazard ratio* in the context of this model. What assumptions does this model make about the effect of the covariates on the hazard function? **[6 marks]**

QUESTION TWO (20 marks) (Optional)

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

$$h(t) = \lambda\gamma t^{\gamma-1} \quad t \geq 0 \quad \lambda > 0 \quad \gamma > 0$$

(i) Explain what is meant by the *hazard function* in survival analysis. **[1 mark]**

(ii) Derive the survivor function, $S(t)$ and the density function $f(t)$ of T. **[3 marks]**

(iii) Show that the parameters are given by the intercept and slope of the theoretical relationship of the logarithm of the cumulative hazard function plotted against the logarithm of time. **[2 marks]**

(b) The table below shows the Kaplan-Meier estimate of the survival function $S(t)$ for 13 patients for the time to removal of a catheter following a kidney infection. Sometimes the catheter has to be removed for reasons other than infection, giving rise to right-censored observations

Survival analysis for Time, the number of days from insertion of catheter until removal

Time	Status of Catheter	Cumulative Survival	Standard Error	Cumulative Events	Number Remaining
8.0	removed	.9231	.0739	1	12
15.0	removed	.8462	.1001	2	11
22.0	removed	.7692	.1169	3	10
24.0	removed	.6923	.1280	4	9
30.0	removed	.6154	.1349	5	8
54.0	Censored			5	7
119.0	removed	.5275	.1414	6	6
141.0	removed	.4396	.1426	7	5
185.0	removed	.3516	.1385	8	4
292.0	removed	.2637	.1288	9	3
402.0	removed	.1758	.1119	10	2
447.0	removed	.0879	.0836	11	1
536.0	removed	.0000	.0000	12	0

Number of Cases: 13 Censored: 1 (7.69%) Events: 12

(i) Use a graphical method based on the estimated cumulative hazard function for checking whether the data may reasonably be assumed to come from a Weibull distribution.(Use value correct to 1 d.p.) **[5 marks]**

(ii) 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. **[4 marks]**

(c) Suppose there is a single sample of failure times possibly subject to censoring and that the times can be modeled by $f(x;\theta)$. Derive the loglikelihood that can used to estimate unknown parameters θ. **[5 marks]**

QUESTION THREE (20 marks) (Optional)

(a) Consider a continuous random variable T with survivor function $S(t)$. Show that the mean of T is given by $\int_0^\infty S(t)dt$. **[3 marks]**

(b) Suppose that T is a continuous, positive random variable with cumulative distribution function $F(t)$ and probability density function $f(t)$. Let $\lambda(t)$ the hazard function. Derive an expression of $f(t)$ in terms of the hazard and the cumulative hazard. **[6 marks]**

(c) A random variable, T, has the exponential distribution with hazard function

$$h(t) = \lambda \quad t \geq 0 \quad \lambda > 0$$

Show that for $\partial t > 0$, $Pr(t \leq T \leq t + \partial t | T > t)$ is independent of t. **[4 marks]**

(d) A study was made of the impact of drinking beer on men aged 60 years and over. A sample of men was followed from their 60th birthdays until they died, or left the study for other reasons. The baseline hazard of death, μ, was assumed to be constant, and a proportional hazards model was estimated with a single covariate, the average daily beer intake in standard-sized glasses consumed, x. The equation of the model is:

$$\lambda(t) = \mu \exp(\beta x)$$

where $\lambda(t)$ is the hazard of death at age $60 + t$.
The estimated value of μ, is 0.03, and the estimated value of β is 0.2.

(i) State the features of the person to whom the baseline hazard applies. **[1 mark]**

(ii) Explain how μ and β should be interpreted, in the context of this model. **[1 mark]**

(iii) Calculate the estimated hazard of death of a man aged exactly 62 years who drinks two glasses of beer a day. **[1 mark]**

(iv) A man is aged exactly 62 years and drinks two glasses of beer a day. Calculate the

(a) estimated probability that this man will still be alive in 10 years time. **[2 marks]**

(b) expectation of life at age 60 years for this man. **[2 marks]**

QUESTION FOUR (20 marks)(Optional)

(a) A careful shopkeeper takes delivery of a batch of 20 packets of cheese. Every morning at 8 a.m. precisely she checks to see if any of the cheese has gone mouldy and throws away any mouldy packets.
As she runs a high quality establishment, she has lots of customers and some of the cheese is sold. After ten days she decides the cheese will be too old to sell and throws out the remaining packets.
A curious customer observes that the shopkeeper has created an observational plan for calculating the hazard of cheese going mouldy.

(i) State, with reasons, THREE types of censoring present in this situation. **[6 marks]**

(ii) Assess, for EACH type of censoring listed in your answer to part (i), whether a change to the observational plan could be made which would remove that type of censoring. **[3 marks]**

The shopkeeper made notes at 8 a.m. each day as follows:

Day	Shopkeepers notes
1	Sold three packets already
2	Sold one more packet
3	One went mouldy
4	Two more mouldy ones, I hope my fridge is cold enough
5	Seems OK, nothing to report
6	Sold four more all to one customer!
7	Nothing to report
8	Another two mouldy ones this morning
9	Sold two more
10	Three more mouldy ones I'll throw the rest out

(iii) Calculate the Kaplan-Meier estimate of the survival function for cheese staying free from mould. **[5 marks]**

(iv) Estimate the variance of estimate at the 5^{th} day and hence lower bound value its 99% of confidence interval. **[5 marks]**

(v) Determine the probability that a packet of cheese does not become mouldy in 10 days. **[1 mark]**

QUESTION FIVE (20 marks) (Optional)

(a) A large life insurance company is conducting an investigation into the mortality of its policyholders to see if this has changed since the previous investigation ten years ago.

  (i) Describe why the crude mortality rates should be graduated. **[3 marks]**

  (ii) Below is a sample of the results:

	Current investigation		Previous investigation
Age	Exposed to risk	Observed deaths	mortality rates
55	5,842	150	0.0267
56	5,630	32	0.0278
57	4,281	126	0.0301
58	3,955	98	0.0325
59	3,879	142	0.0356
60	3,550	149	0.0387
61	4,006	162	0.0396
62	4,150	173	0.0410
63	3,520	158	0.0433
64	3,057	150	0.0458
65	3,666	200	0.0490

  (i) Conduct a chi-squared test for goodness of fit on these data. **[7 marks]**

  (ii) Discuss three situations which the test in part (i) is not able to detect and each case identify a suitable test for the detection . **[6 marks]**

(b) A researcher is investigating the contributing factors to the speed at which patients recover from a common minor surgical procedure undertaken in hospitals across the country. He has the questionnaires which each patient completed before the surgery and the length of time the patient remained in hospital after surgery and is attempting to fit a Cox proportional hazards model to the data.
He has fitted a model with what he assumes are the most common contributing factors and has calculated the parameters as shown in the table below:

Covariate	Category	Parameter
Gender	Male	0
	Female	0.065
Smoker	Non Smoker	-0.035
	Smoker	0
Drinker	Non Drinker	-0.06
	Moderate Drinker	0
	Heavy Drinker	0.085

A male moderate drinker who does not smoke has a hazard of leaving hospital after three days of 0.6. Calculate the probability that a female non drinker who smokes and who is still in hospital after three days is discharged at that point. **[4 marks]**