Strathmore
UNIVERSITY

STRATHMORE INSTITUTE OF MATHEMATICAL SCIENCES

BACHELOR OF BUSINESS SCIENCE IN ACTUARIAL SCIENCE

END OF SEMESTER EXAMINATION

BSA 3218: ACTUARIAL MODELING II

DATE: 16th September 2021 TIME: [08:00] 2 HRS

ANSWER QUESTION ONE AND ANY OTHER TWO QUESTIONS

QUESTION 1 (30 MARKS)

a. After a power outage, an intern at a certain hospital is forced to record patient's data on a notebook. You happen to get a glimpse of the note book and below is the information you see.
   - A certain patient was admitted to the hospital but she has put an asterisk on the date
   - The calendar year a certain patient left, if they left after getting well.
   - The date a certain patient died
   - The date a certain patient was transferred to another hospital with ICU facility.

   If you are interested in the patients who left the hospital after getting well, by defining the type of censoring, discuss the possible types of censoring in the above records (7 marks)

b. Derive the hazard function for the Weibull distribution. State the values of γ for which this is: decreasing, constant, increasing? (4 marks)
c. Explain why the likelihood function for estimating the parameters of the Cox proportional hazard model is referred to as "partial likelihood" (2 marks)

d. The following are Cox regression results obtained from R, adequately discuss the results. (6 marks)

```
Call:
coxph(formula = Surv(time, status) ~ sex, data = lung)
  n= 228, number of events= 165
        coef exp(coef) se(coef)      z Pr(>|z|)
sex -0.5310    0.5880   0.1672 -3.176  0.00149 **
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
    exp(coef) exp(-coef) lower .95 upper .95
sex     0.588      1.701    0.4237     0.816
Concordance= 0.579  (se = 0.022 )
Rsquare= 0.046   (max possible= 0.999 )
Likelihood ratio test= 10.63  on 1 df,   p=0.001111
Wald test            = 10.09  on 1 df,   p=0.001491
Score (logrank) test = 10.33  on 1 df,   p=0.001312
```

e. Briefly outline the three methods of graduation (3 marks)
f. Mortality of a group of lives is assumed to follow Gompertz' law. Calculate μ_x for a 30 and a 70 year old, given that $\mu_x = 0.003$ for a 50 year old and 0.01 for a 60 year old. (4 marks)
g. The following is data on the number of individuals in a certain pension scheme in the year 2019.

Time	Population
1 January 2019	34,453
21 June 2019	39,656
3 September 2019	29,469
13 December 2019	37,256

Using census approximation, calculate the central exposed to risk for the year 2019. Assume that $P_{x,t}$ is linear between census dates. (4 marks)

Question Two (20 Marks)

a. Suppose that a mortality investigation covers the period 1 January 2015 to 31 December 2017. In this investigation, the age label used is 'age last birthday'. The table below gives information about three males involved in the investigation.

Life	Date of birth	Date of joining	Date of exit	Reason for exit
A	25.04.83	01.01.15	30.10.16	Death
B	01.07.83	12.09.16	–	–
C	04.09.82	22.07.17	04.12.17	Withdrawal

Use the data to determine the range of dates for which the lives contribute to E_x^c at each age where they make a contribution, show your workings (10 marks).

b. The table below is data on 5 ages from a certain research in 2019. Assuming $\log(e^{\mu_x} - 1)$ satisfies a polynomial of order one, obtain the graduated rates q_x^0. Determine whether your graduated results are smooth and whether the actual deaths are consistent with those of the model. (10 marks).

x	l_x	d_x	μ_x
36	98291	74	0.000 72
37	98217	81	0.000 79
38	98136	88	0.000 86
39	98048	96	0.000 94
40	97952	105	0.001 02
41	97847	114	0.001 12

<u>Question Three (20 Marks)</u>

You want to use a Cox regression model to estimate the force of mortality for a group of endowment assurance policyholders. You propose using a model that takes account of duration (*ie* the time that has elapsed since the policy was issued) and the age and sex of the policyholder. You start by investigating the model:

$$\boldsymbol{\mu(x, z_1, z_2)} = \mu_0(\boldsymbol{x})\boldsymbol{e}^{\boldsymbol{\beta_1 z_1 + \beta_2 z_2}}$$

Where:

x denotes the age of the policy holder,

$$Z_1 = \begin{cases} 0 & \text{if the duration is less than 1 year} \\ 1 & \text{if the duration is at least 1 year} \end{cases}$$

$$Z_2 = \begin{cases} 0 & \text{for males} \\ 1 & \text{for females} \end{cases}$$

You have estimated the values of the parameters Z_1 and Z_2 , and have obtained the following results:

Covariate	Parameter	Standard error
Duration	0.416	0.067
Sex	–0.030	0.017

(i) State the class of policyholders to which the baseline hazard refers. (2 marks)
(ii) Explain whether the duration covariate is significant in determining mortality. (2 marks)

(iii) Compare the force of mortality for a new female policyholder to that of a male policyholder of the same age, who took out a policy 2 years ago. (6 marks)

b. You have been give data to analyze and you discover it has 30 variables. You want to select those variables with significant effects. Discuss in detail how you will go about it. (10 marks)

Question Four (20 Marks)

a. The following data is obtained from a mortality investigation. Each of the lives was under observation at all ages from age 55 until they died or were censored. The table below shows disease status, age at exit and reason for exit from the investigation.

Life	Disease Status	Age at Exit	Reason for Exit
1	Positive	56	Death
2	Negative	62	Censored
3	Negative	63	Death
4	Positive	66	Death
5	Positive	67	Censored
6	Positive	67	Censored

Using the following model for mortality: $\mu(x|Z = z) = \mu_0(x)e^{\beta z}$ and where x denotes the age, $\mu_0(x)$ the baseline hazard, and $z = 0$ for positive status and 1 otherwise, obtain the value of β that maximizes the partial likelihood (12 marks)

b. Carry out the smoothness test on the following set of graduated rates: (8 marks)

x	$\mathring{\mu}_x$
55	0.00429
56	0.00478
57	0.00535
58	0.00596
59	0.00667
60	0.00754
61	0.00867

Question Five (20 Marks)

Army training is usually very demanding and a good number of individuals do not make it. The following records were pulled out of the army database on the number of trainees who quit or were lost to follow up.

	Week	Event		Week	Event
1	17	Quit	**11**	19	Lost to follow up
2	13	Quit	**12**	15	Quit
3	15	Lost to Follow up	**13**	4	Quit
4	7	Lost to follow up	**14**	11	Quit
5	21	Quit	**15**	14	Lost to follow up
6	18	Lost to follow up	**16**	18	Quit
7	5	Quit	**17**	10	Quit
8	18	Quit	**18**	10	Quit
9	6	Lost to follow up	**19**	8	Lost to follow up
10	22	Quit	**20**	17	Quit

i. State the values n, m, k, t_j, d_j, c_j and n_j for these data, assuming that censoring occurs just after the failures were observed. (3 marks)

ii. Calculate the Kaplan-Meier estimate of the trainee survival function. (5 marks)

iii. Using Greenwood's formula, estimate $var[\tilde{S}(16)]$ (4 marks)

iv. Calculate the Nelson-Aalen estimate of the cumulative hazard function using the given data values (4 marks)

v. Use the data to estimate $var\widehat{\Lambda}(16)$ (4 marks)