Strathmore
UNIVERSITY

INSTITUTE OF MATHEMATICAL SCIENCES
MSC (STATISTICAL SCIENCE)
END OF SEMESTER EXAMINATION
STA 8403: SURVIVAL ANALYSIS

DATE 15th August **TIME: 2.5 Hours**

Instruction:
1. This examination consists of **FOUR** questions.
2. Answer **Question ONE (COMPULSORY)** and any other **TWO** questions.

Question 1 (20 marks)

(a) A large number of individuals were sampled randomly from a population of adults, were enrolled in a study and were followed for 30 years to assess the age at which a disease symptom first appeared.
   i. What are the time origin and time scale for this study? **(2 marks)**
   ii. What is the failure event for this study? **(1 mark)**

(b) For 3 selected individuals described below, discuss in detail the types of censoring and/or truncation that is represented by the following individuals. **(6marks)**
   i. The first individual, enrolled in the study at age 45, entered the study with the symptom already present.
   ii. The next two healthy (without the symptom present) individuals enrolled in the study at ages 30 and 42 and never showed the symptom.
   iii. The next two healthy individuals, enrolled in the study at ages 35 and 40, exhibited the symptom at the second and fifth exam after enrollment (6 years and 15 years after enrollment), respectively. The symptom (which could only be identified by a clinical exam and clinical testing) may have appeared between exams.

(c) One wants to test the null hypothesis that the lifetime distribution for the brake pads is the same for the two year models Year=0 and Year=1, when the other factors are ignored.
   i. Formulate a null hypothesis and alternative hypothesis for this in term of hazard function **(2marks)**
   ii. Formulate the hypothesis in terms of cox model **(2marks)**
   iii. This test was performed, and the p-value was 0.206. Using this information can we conclude that the lifetime of the brake pads is independent of the car's year model (Year). Explain your answer **(2marks)**

(d) Describe how a graph of either estimate of the cumulative hazard might be used to select a suitable parametric model for the data from a set of possible candidates. **(3marks)**

(e) What is a frailty model? **(2marks)**

Question 2 (20 marks)

(a) What are typical causes leading to a censored observation **(3marks)**

(b) Differentiate between Type I and Type II censoring **(4marks)**

(c) The survival function of the three-parameter Weibull distribution is given by $S(t) = \exp[-(t-\gamma)^{\alpha}]$.
Find the hazard rate and the density function of the three- parameter Weibull distribution **(4marks)**.

(d) Derive an expression for the mean residual life for an exponential distribution. **(3marks)**

(e) You are given the following survival function for a newborn:

$$S_0(t) = \frac{(121 - t)^{1/2}}{k}, \quad t \in [0, \omega]$$

i. Show that K must be 11 for $S_0(t)$ to be a valid survival function. **(3marks)**

ii. Calculate the mean survival time for this survival model. **(3marks)**

Question 3 (20 marks)

(a) Consider the following sequence of survival times for a set of individuals measured from birth: 0.1+, 0.3, 0.3, 0.5+, 0.6, 0.9, 0.9, 1+, 1.4 where numbers followed by a "+" indicate right-censored values.

i. Calculate the Nelson-Aalen estimate of the cumulative hazard function for these data, presenting the results in a table. Sketch a graph of the estimated function. **(6marks)**

ii. Calculate a linear 95% confidence interval for H(0.6). **(4marks)**

(b) The below data is on time to death for breast cancer patients who where classified as being on drug A and drug B.

Drug A: 15, 18, 19, 19, 20

Drug B: 16+, 18+, 20+, 23, 24

Test the hypothesis that there is no difference in survival between the two groups using the log-rank test. **(10marks)**

Question 4 (20 marks)

(a) Explain how the Cox-Snell residuals are used to assess the appropriateness of the Cox proportional hazards model for these data. ***(4 marks)***

(b) The models below were generated from a dataset which represents the death times of male laryngeal cancer patients contain times recorded in years between first treatment and either death, or the end of the study. Also recorded are the patient's age at diagnosis and the year of diagnosis.

Model 1: Cox regression using the 4-level stage variable

```
> fit <- coxph(Surv(time,delta) ~factor(stage), data=larynx)
```

```
> summary(fit)
Call:
coxph(formula = Surv(time, delta) ~ factor(stage), data = larynx)

  n= 90, number of events= 50

                   coef exp(coef) se(coef)     z Pr(>|z|)
factor(stage)2 0.06481   1.06696  0.45843 0.141   0.8876
factor(stage)3 0.61481   1.84930  0.35519 1.731   0.0835 .
factor(stage)4 1.73490   5.66838  0.41939 4.137 3.52e-05 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

               exp(coef) exp(-coef) lower .95 upper .95
factor(stage)2     1.067     0.9372    0.4344      2.62
factor(stage)3     1.849     0.5407    0.9219      3.71
factor(stage)4     5.668     0.1764    2.4916     12.90

Concordance= 0.668  (se = 0.043 )
Rsquare= 0.167   (max possible= 0.987 )
Likelihood ratio test= 16.49  on 3 df,   p=0.0009016
Wald test            = 19.24  on 3 df,   p=0.0002433
Score (logrank) test = 22.88  on 3 df,   p=4.284e-05
```

Model 2: Cox regression using the stage variable as a linear term in the model

```
> fit <- coxph(Surv(time,delta) ~ stage, data=larynx)
> summary(fit)
Call:
coxph(formula = Surv(time, delta) ~ stage, data = larynx)

  n= 90, number of events= 50

        coef exp(coef) se(coef)     z Pr(>|z|)
stage 0.5088    1.6633   0.1412 3.604 0.000313 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

      exp(coef) exp(-coef) lower .95 upper .95
stage     1.663     0.6012     1.261     2.193

Concordance= 0.668  (se = 0.043 )
Rsquare= 0.137   (max possible= 0.987 )
Likelihood ratio test= 13.26  on 1 df,   p=0.0002706
Wald test            = 12.99  on 1 df,   p=0.000313
Score (logrank) test = 13.82  on 1 df,   p=0.0002017
```

i. Interpret the exp(coef) from model 2. **(2 marks)**

ii. Based in the results in model 1, write a sentence description of the stage 4 result for a research section in a scientific paper. **(4 marks)**

iii. Calculate hazard ratio to compare the risk of death for those in stage 4 & stage 3 and interpret your answer. **(5 marks)**

iv. How would you evaluate whether there is any significant interaction in a model with age and stage as a linear predictor. State the null and alternative hypotheses that are being used. **(5 marks)**