JARAMOGI OGINGA ODINGA UNIVERSITY OF SCIENCE AND TECHNOLOGY
SCHOOL OF BIOLOGICAL, PHYSICAL, MATHEMATICS AND ACTUARIAL SCIENCES
UNIVERSITY EXAMINATION FOR DEGREE OF BACHELOR OF SCIENCE IN ACTUARIAL SCIENCE

WITH IT

4th YEAR 1st SEMESTER 2023/2024 ACADEMIC YEAR

MAIN REGULAR

COURSE CODE: WAB2413

COURSE TITLE: SURVIVAL ANALYSIS

EXAM VENUE: **STREAM: (Bsc. Actuarial Science)**

DATE: **EXAM SESSION: Sep-Dec 2023**

TIME: 2.00 HOURS

Instructions:

i. Answer questions one and any other two.
ii. Candidates are advised not to write on the question paper.
iii. Candidates must hand in their answer booklets to the invigilator while in the examination room.
iv. Statistical software R and the required packages to be used for data analysis

QUESTION ONE (30 marks)

a) Consider a simple example of five cancer patients who enter a clinical trial as illustrated in the following diagram:

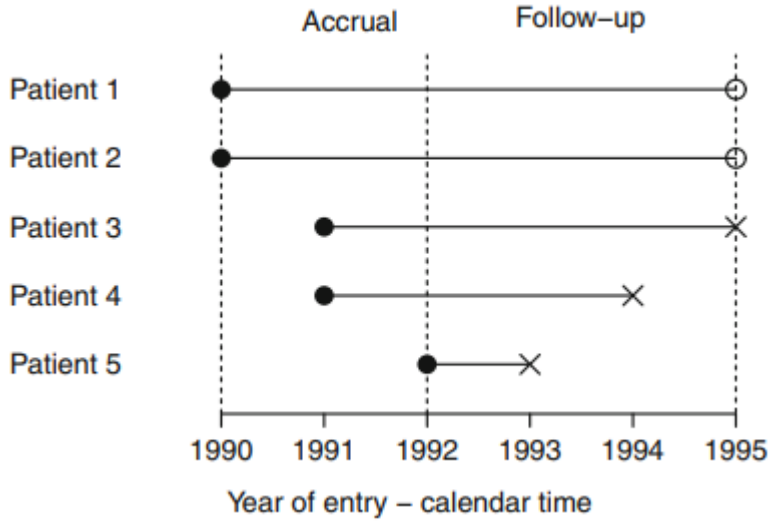

Figure 1: A simple example of five cancer patients who enter a clinical trial

i. Re-write these survival times in terms of patient time (2 marks)

ii. Create a simple data set listing the survival time and censoring indicator for each patient. (2 marks)

iii. How many patients died? (1 marks)

iv. How many person-years are there in this trial? (1 marks)

v. What is the death rate per person-year? (2 marks)

b) For the "gastricXelox" data set, use R to determine:

i. How many patients had the event (death or progression). (2 marks)

ii. The number of person-weeks of follow-up time. (2 marks)

iii. The event rate per person-week. (2 marks)

c) Consider a survival distribution with constant hazard $\lambda = 0.07$ from $t = 0$ until $t = 5$ and then hazard $\lambda = 0.14$ for $t > 5$. (This is known as a piecewise constant hazard.)

i. Plot this hazard function and the corresponding survival function for $0 < t < 10$. (2 marks)

ii. What is the median survival time? (2 marks)

d) Another parametric survival distribution is the log-normal distribution.

i. Use the density and cumulative distribution R functions "dlnorm" and "plnorm" to compute and plot the lognormal hazard functions with the parameter "meanlog" taking the values 0, 1, and 2, and with "sdlog" fixed at 0.25. (2 marks)

ii. Describe the risk profile a disease would have if it followed one of these hazard functions. (2 marks)

e) Using the survival function below,

i. Find the median survival, and a 95 % confidence interval for the median. (2 marks)

ii. Explain why the upper limit of the confidence interval is undefined (2 marks)

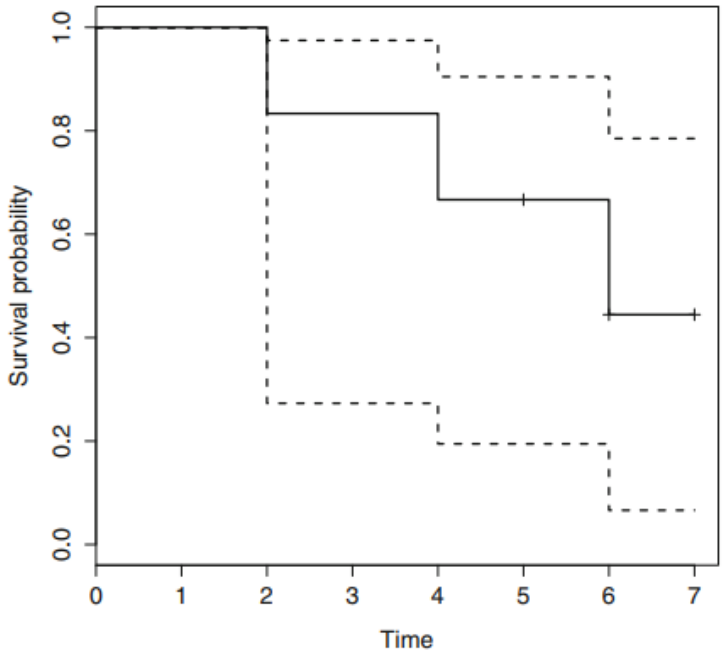

Figure 2: Kaplan-Maier survival curve estimate with 95 % confidence intervals

f) Use the pharmacoSmoking data to answer the following questions:
   a) Compare the two treatments using the Prentice modification of the Gehan test, and compare your results to those from the log-rank test. (2 marks)
   b) Carry out a log-rank test comparing the two treatments stratifying on employment status. (2 marks)

QUESTION TWO (20mks)

a) Consider the data set "aml", which is included in the "survival" package. This is a study of whether or not maintenance therapy increases survival among patients with acute myelogenous leukemia, with survival time measured in weeks. The basic Cox model may be fitted as follows:
result <- coxph(Surv(time, status) ~ x, data=aml)
   i. Provide the cox regression summary for the slope parameter estimates and their confidence intervals (3 marks)
   ii. Create a coarser time variable by expressing it in months instead of weeks as follows: time.months <- cut(aml$time, breaks=seq(0,161,4), labels=F) and provide the variables summary statistic. (2 marks)
   iii. Now re-fit the model, modeling ties using the Breslow, Efron, and exact methods. Which approximate method gives a result closest to that from the exact method? (3 marks)

b) The data set "hepatocelluar" is in the "asaur" package. It contains 17 clinical and biomarker measurements on 227 patients, as well as overall survival and time to recurrence, both recorded in months [42, 43]. There are three measures of CXCL17 activity, CXCL17T (intratumoral), CXCL17P (peritumoral), and CXCL17N (nontumoral). There is a particular interest in whether they are related to overall and also recurrence-free survival. Which of the three is most strongly related for each survival outcome? For the one most strongly related with survival, fit a spline model and plot it. Does this suggest that categorizing CXCL17 would be appropriate? (4 marks)

c) For the covariates with complete data (in Columns 1–22), use stepwise regression with AIC to identify the best model for (a) overall survival, and (b) recurrence-free survival. (4 marks)

d) Consider the CXCL17 model you fitted above. Check the functional form using martingale residuals, and use case-deletion residuals to identify any outlying points. Also use Schoenfeld residuals to check the proportional hazards assumption. (4 marks)

QUESTION THREE (20mks)

a) Consider the following data from two groups X and Y

X	3	5	7	9+	18
Y	12	19	20	20+	33+

Test:

H0: $S_x(t) = S_y(t)$ verses H1: $S_x(t) \neq S_y(t)$ (10 marks)

b) The following table refers to the survival times (in years) of the patients in a certain clinical trial. An asterisk indicates that the observation was censored. It is assumed that censoring was random, and that patients were randomly assigned to the treatment or control arms. (10 marks)

Control Arm	3	6*	7	8*	8*	8*	9	11	13
Treatment Arm	4*	5	8*	8*	8*	8*	10	12*	14*

Assuming a proportional hazards model, find the estimate of the log- hazard β using the Cox proportional hazards method.

QUESTION FOUR (20mks)

A group of 15 laboratory rats are injected with a new drug. They are observed over the next 30 days. The following events occur.

Day	Event
3	Rat 4 dies from the effect of drug
4	Rat 13 dies from the effect of the drug
6	Rat 7 gnaws through bars of the cage and escapes
11	Rat 6 and 9 die from the effect of the drug
17	Rat 1 is killed by other rats
21	Rat 10 dies due to the effect of the drug
24	Rat 8 is freed during raid by animal liberation activists
25	Rat 12 is accidentally freed by journalist reporting earlier raid
26	Rat 5 dies from the effect of the drug
30	Investigation closes and remaining rats hold a street party

(a) How would this information be represented notationwise? (8 marks)

(b) Estimate the Survival function S(t) (4 marks)

(c) Estimate the variance of S(t) (4 marks)

(d) Obtain the 95% confidence interval for S(t). (4 marks)

QUESTION FIVE (20mks)

a) Given ${}_tP_x = 1 - \left(\frac{t}{100}\right)^{1.5}$ for x=60 and 0<t<100.

Calculate $E(T(x))$ (10 marks)

b) Given $\mu_{x+t} = \frac{1}{85-t} + \frac{3}{105-t}$ for $0 \leq t < 85$

Calculate ${}_{20}P_x$ (10 marks)