JARAMOGI OGINGA ODINGA UNIVERSITY OF SCIENCE AND TECHNOLOGY
SCHOOL OF MATHEMATICS AND ACTUARIAL SCIENCE
UNIVERSITY EXAMINATION FOR DEGREE OF BACHELOR OF SCIENCE ACTUARIAL

4th YEAR 1st SEMESTER 2018/2019 ACADEMIC YEAR

MAIN REGULAR

COURSE CODE: SAS 415

COURSE TITLE: SURVIVAL ANALYSIS

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

DATE: **EXAM SESSION: SEP-DEC 2018**

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) All computations and data analysis to be done with STATA statistical software.

(v) STATA codes used to be shown beside the answer to every question.

Examination Data set

This examination uses Framingham Heart Study (FHS)) teaching dataset to answer questions. One goal of the Framingham Heart Study (FHS) was to observe the incidence of Coronary Heart Disease (CHD) among all participants and among participants with certain risk factor characterisics (e.g. smokers, non-smokers,..). The teaching data set for this class contains 4,434 participants who attended the 1956 biennial examination. However, 194 of them were previously diagnosed with CHD, leaving 4,240 participants who are at risk for developing a first CHD event during the 24 years of follow-up. The following results were observed for these 4,240 subjects:

- 1,406 died from any cause.
- 88,389.45 person-years were observed before subjects died or the study ended.
- 824 died from non-CHD causes (no longer at risk of developing CHD),
- 32 lost-to-follow-up and had not developed CHD at their last contact,
- 2338 completed 24 years of follow-up and did not develop CHD,
- 1046 developed CHD during the 24 years of follow-up (including 582 deaths).
- 80,925.16 person-years were observed before subjects were lost-follow-up, died, developed CHD, or the study ended.

QUESTION ONE (30 marks)

BMI AND CHD prevalence. The following table uses data from the NHLBI teaching data set and displays categories of body mass index for 4,415 participants in the Framingham Heart study attending an examination in 1956 with non-missing values for body mass index. For each body Disease (CHD) at that exam (**prevchd1=1)**

Body Mass Index Category		**Number of Subjects at 1956 Exam**	**Cases of CHD Diagnosed Prior to 1956**
Under Weight	BMI<18.5	57	0
Normal Weight	18.5 ≤ BMI <30	1936	66
Overweight	25 ≤ BMI <30	1848	90
Obese	BMI ≥ 30	574	38
TOTAL		**4415**	**194**

a) What is the prevalence of obesity among the 4415 participants at the 1956 exam? (2 mks)

b) What is the prevalence of CHD at the 1956 exam among the 4415 participants at the 1956 exam? (2 mks)

c) What is the prevalence of CHD at the 1956 exam for each of the body mass index classes?

i) Under Weight Participants (2 mks)

ii) Normal Weight Participants (2 mks)

iii) Over Weight Participants (2 mks)

iv) Obese Participants (2 mks)

Diabetes prevalence. Use Stata and the NHLBI data set to calculate the prevalence of diabetes among participants who attended and had non-missing data on diabetes at all three examinations. (**Hint: There were 3,206 such participants.)**

d) What is the prevalence of diabetes at the first exam (diabetes1=1)? (2 mks)

e) What is the prevalence of diabetes at the second exam (diabetes2=1)? (2 mks)

f) What is the prevalence of diabetes at the third exam (diabetes3=1)? (2 mks)

BMI and Hypertension prevalence. Use Stata and the BMI1 variable in the NHLBI data set to create the four categories of body mass index as defined in the first question.

g) What is the prevalence of hypertension (**prevhyp1-1)** at the 1956 exam for each of the body mass index classes?

i) Under Weight Participants (1 mks)

ii) Normal Weight Participants (1 mks)

iii) Over Weight Participants (1 mks)

iv) Obese Participants (1 mks)

The following tables show the code and sex-specific results from a prospective short study that examines the association between a binary exposure(E) and the development of a disease (D) during 20 years of follow-up.

h)

Full Data

	D +	D -	Total
E +	1123	8877	10000
E -	1008	8992	10000
Total	2131	17869	20000

Sex-specific data

Males

	D +	D -	Total
E +	259	1741	2000
E -	648	5352	6000
Total	907	7093	8000

Females

	D +	D -	Total
E +	864	7136	8000
E -	360	3640	4000
Total	1224	10776	12000

a) What is the value for the Crude Risk ratio, comparing exposed subjects to non-exposed subjects? (2 mks)

b) Using the Mendel-Haenszel formula, what is the value for the sex-adjusted Risk ratio, comparing exposed subjects to non-exposed subjects? (2 mks)

c) Using the total data as standard population, what is the value for the standardized Risk ratio? (2 mks)

d) Using the risk ratio as a measure of association, is sex an effect modifies in this study? (2 mks)

QUESTION TWO (20mks)

Hypertension and high blood pressure. Use Stata to create a binary variable (highbp1) to represent the presence/absence of high blood pressure at the 1956 examination.

```
generate highbp1=.
replace highbp1=1 if (sysbp1>=140 | diabpl >=90)
replace highbp1=0 if (sysbp1>140  & diabpl  < 90)
```

(Note: There are no missing data on sysbp1 and diabp1. If data were missing on both sysbp1 then should also be missing for highbp1. If data were missing on diabp1 only and sysbp1 $\geq$ 140 then highbp1 =1, otherwise highbp1 should be missing. Similarly, if data were missing in sysbp1 only and diabp1 $\geq$90 then highbp1 =1, otherwise highbp1 should be missing.)

a) What is the prevalence of CHD (prevchd1 =1) at the 1956 exam for participants with high blood pressure at the 1956 exam (highbp1=1)? (2 mks)

b) What is the prevalence of CHD (prevchd1 =1) at the 1956 exam for participants without high blood pressure at the 1956 exam (highbp1=0)? (2 mks)

Number of Survivors out of 100,000 Live Births.

Age	**1950-1952**	**1990-1992**
0	100,000	100,000
20	73,412	96,902
40	56,884	92,638
70	31,744	79,873

c) What is the probability of surviving from birth to age 20 in 1950-1952? (2 mks)

d) What is the probability of surviving from age 40 to age 70 in 1990 -1992? (2 mks)

e) Define the absolute survival increase over the 40 year span as p_1-p_2, where p_1 is the chance of surviving from age x to age x^{+n} in 1990-1992 and p_2 is the chance of surviving from age x to age x^{+n} in 1950-1952. Which age group has the greatest absolute survival increase?

i) 0-20 (2 mks)
ii) 20-40 (2 mks)
iii) 40-70 (2 mks)

f) Define the relative survival increase over the 40 year span as (p_1-p_2)/ p_2, where p_1 is the chance of surviving from age x to age x^{+n} in 1990-1992 and p_2 is the chance of surviving age x to age x^{+n} in 1950-1952. Which age group has the greatest relative survival increase?

i) 0-20 (2 mks)
ii) 20-40 (2 mks)
iii) 40-70 (2 mks)

QUESTION THREE (20mks)

Time to death and Systolic Blood Pressure.

For this problem set, we will return to the Framingham data set. We will examine time to death (in years), timedth, where the variable death, is the censoring indicator.

Out of the 500 people, how many died? (2 mks)

Suppose we want to look at the effect of systolic blood pressure on time to death. Lets classify those people with a systolic blood pressure greater than 140mmHg at exam 1 as having high systolic blood pressure.

a) Plot the Kaplan Meier estimates of the survival function for those do and do not have high blood pressure. (2 mks)

b) What is the probability of surviving beyond 2 years in the group without high blood pressure? (2 mks)

c) Conduct a log rank test to determine if the distributions of survival time differ between those with and without high systolic blood pressure. Use a 0.05 level of significance.

i) What is your test statistic? (2 mks)

ii) True or false: the test statistic is a random variable that follows a chi-square distribution with one degree of freedom under the null hypothesis. (2 mks)

iii) Is your p-value less than 0.05? (2 mks)

iv) This analysis tells us that we can significantly improve someone's life expectancy if we lower their systolic blood pressure from above 140 mmHg to below 140 mmHg. True or false? (2 mks)

QUESTION FOUR (20mks)

Time to Death, Age, and systolic Blood Pressure. For this reason, use the full Framingham dataset (fhs.dta), posted on the webpage. Age is likely a confounder in the relationship between systolic blood pressure and time to death. In this problem, we will conduct two logrank tests-for older and younger subsets of the Framingham cohort-and compare the results.

a) First let's examine some descriptive statistics, focusing on patients age 35-40.

i) How many participants were between age 35-40 at baseline? (2 mks)

ii) How many of those participants died during follow up? (1 mks)

iii) Among participants age 35-40 at baseline, how many had high blood pressure. What percent of patients with high blood pressure died? (2 mks)

iv) What percent of patients without high blood pressure died? (1 mks)

v) How many participants were between age 65-70 at baseline? (2 mks)

vi) How many of those participants died during follow up? (2 mks)

b) Now, look at patients age 65-70. (2 mks)

i) Among participants age 65-70 at baseline, how many had high blood pressure? (2 mks)

ii) What percent of patients with high blood pressure died? (2 mks)

iii) What percent of patients without high blood pressure died? (2 mks)

iv) Plot the survival curves for those age 35-40 and age 65-70. True or false: when stratifying by age, the difference between the survival curves is smaller than in the previous analysis, when we did not stratify. (2 mks)

QUESTION FIVE (20mks)

a) Conduct a log rank test to determine if the distributions of survival time differ between those with and without high systolic blood pressure among **the subset of participants ages 35-40 at baseline**. Use a 0.05 level of significance.

i) What is your test statistic? (2 mks)

ii) What is the null distribution of your test statistic? (2 mks)

iii) Is your p-value less than 0.05? (2 mks)

iv) What do you conclude? (2 mks)

b) conduct a log rank test to determine if the distributions of survival time differ between those with and without systolic blood pressure among **the subset of participants ages 65-70 at baseline**. Use a 0.05 level of significance.

i) What is your test statistic? (2 mks)

ii) What is the null distribution of your test statistic? (2 mks)

iii) Is your p-value less than 0.05. (2 mks)

iv) What do you conclude? (2 mks)

c) Conduct Cox-proportional regression involving death time to death (timedth) during the 24 years of follow-up. The regressor variables to include age1, sex1, cursmoke1, bmi1, and sysbpi. Write the regression equation. (2 mks)

d) Interpret the cox-regression parameters and slopes (2 mks)