UNIVERSITY OF FORT HARE

DEPARTMENT OF STATISTICS

DEGREE EXAMINATIONS

SURVIVAL ANALYSIS

STE 702

JUNE/JULY 2023

Time: 3 hours **Marks: 100**

This paper consists of 7 pages including the cover page

Internal Examiners

Dr Ruffin Mutambayi

Ms S. Zungu

External Examiner

Dr Oliver Bodhlyera

INSTRUCTIONS

- **Answer ALL questions**
- **Non-programmable Calculators may be used**
- **Submit your answer sheets through Blackboard**

Question 1 **[20 marks]**

Please circle your choice among the given choices of the following questions.

1.1 In a survival analysis, the outcome variable is dichotomous. **[1]**

a. True
b. False

1.2 In a survival analysis, the event is usually described by a (0, 1) variable. **[1]**

a. True
b. False

1.3 If the study ends before an individual has gotten the event, then his or her survival time I censored. **[1]**

a. True
b. False

1.4 If, for a given individual, the event occurs before the person is lost to follow-up or withdraws from the study, then this person's survival time is censored. **[1]**

a. True
b. False

1.5 $S(t) = P(T > t)$ is called the hazard function. **[1]**

a. True
b. False

1.6 The hazard function is a probability. **[1]**

a. True
b. False

1.7 Theoretically, the graph of a survivor function is a smooth curve that decreases from $S(t) = 0$ at $t = 0$, and $S(t) = 0$ when t approaches ∞ **[1]**

a. True
b. False

1.8 The survivor function at time t gives the instantaneous potential per unit for a failure to occur, given survival up to time t. **[1]**

a. True

b. False

1.9 The formula for a hazard function involves a conditional probability as one of its components. **[1]**

a. True
b. False

1.10 The outcome variable used in survival analysis is different from that used in linear or logistic modeling. **[1]**

a. True
b. False

1.11 The survivor function ranges between 0 and ∞. **[1]**

a. True
b. False

1.12 The concept of instantaneous potential is illustrated by velocity. **[1]**

a. True
b. False

1.13 A hazard rate of one per day is equivalent to seven per week. **[1]**

a. True
b. False

1.14 If you know the form of a hazard function, then you can determine the corresponding survivor curve, and vice versa. **[1]**

a. True
b. False

1.15 One use of a hazard function is to gain insight about conditional failure rates. **[1]**

a. True
b. False

1.16 If the survival curve for group 1 lies completely above the survival curve for group 2, then the median survival time for group 2 is longer than that for group 1. **[1]**

a. True
b. False

1.17 The risk set at 6 weeks is the set of individuals whose survival times are less than or equal to 6 weeks. **[1]**

a. True
b. False

1.18 If the risk set at 6 weeks consists of 22 persons, and 4 persons fail and 3 persons are censored by the 7^{th} week, then the risk set at 7 weeks consists of 18 persons. **[1]**

a. True
b. False

1.19 The measure of effect used in survival analysis is an odds ratio. **[1]**

a. True
b. False

1.20 If a hazard ratio comparing group 1 relative to group 2 equals 10, then the potential for failure is ten times higher in group 1 than in group 2. **[1]**

a. True
b. False

Question 2 **[50 marks]**

2.1 Suppose that the estimated survivor curve of a time to event dataset is as follows: **[3]**

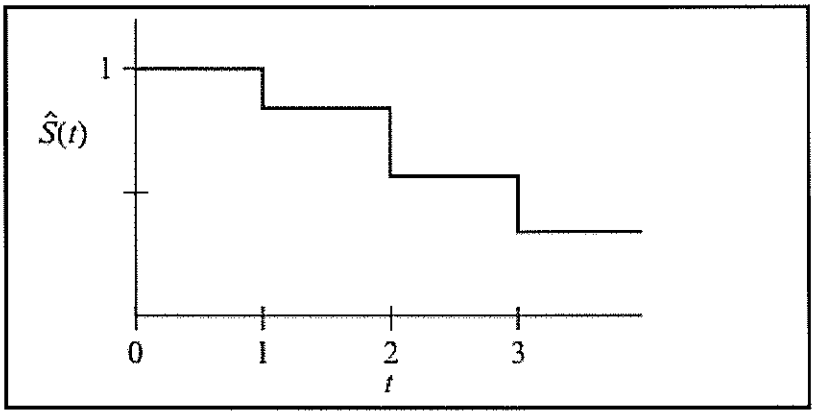

Figure 1

Using information from figure 1, determine the value of:

a. The median, and interpret your result. **[2, 1]**

b. The third quartile, and interpret your result. **[2, 1]**

2.2 Consider the comparison of the following two survivor curves.

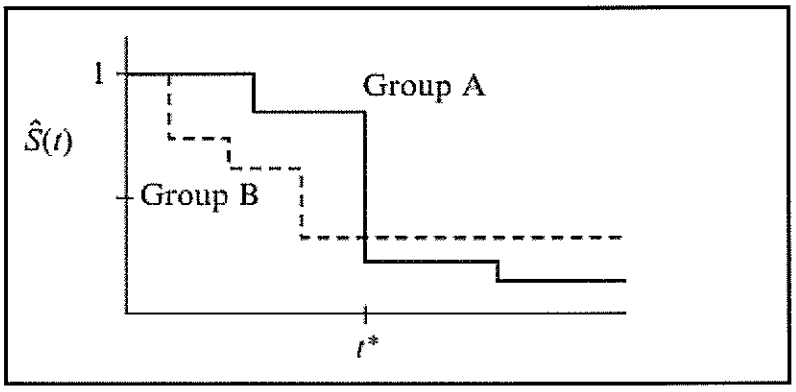

Figure 2

Using information from figure 2:

a. Determine the group that has a longer median survival time. **[3]**

b. What could be your conclusion if you were to run a log-rank test? **[3]**

2.3 Consider the time (in weeks) to recover from patients who were randomly assigned to a treatment of an undisclosed disease.

Treatment: 6, 6, 6, 7, 10, 13, 13, 16, 22, 23, 6^+, 9^+, 10^+, 11^+, 17^+, 19^+, 20^+, 25^+, 32^+, 32^+, 34^+, 35^+

Placebo: 1, 1, 2, 2, 3, 4, 4, 5, 5, 8, 8, 8, 8, 11, 11, 12, 12, 15, 17, 22, 23

N.B: $^+$ denotes censored value.

a. Using the Kaplan Meier technique, estimate the survival function values of patients who were assigned to placebo, and plot the survival curve. **[13]**

b. Formulate a hypothesis of a log-rank test over the groups of patients. **[4]**

c. Write down a SAS program, relating to your log-rank test from (b). **[5]**

2.4 Consider the following SAS output

Test of Equality over Strata			
Test	Chi-Square	DF	Pr > Chi-Square
Log-Rank	0.0878	1	0.7670
Wilcoxon	0.6163	1	0.4324
-2Log(LR)	0.7327	1	0.3920

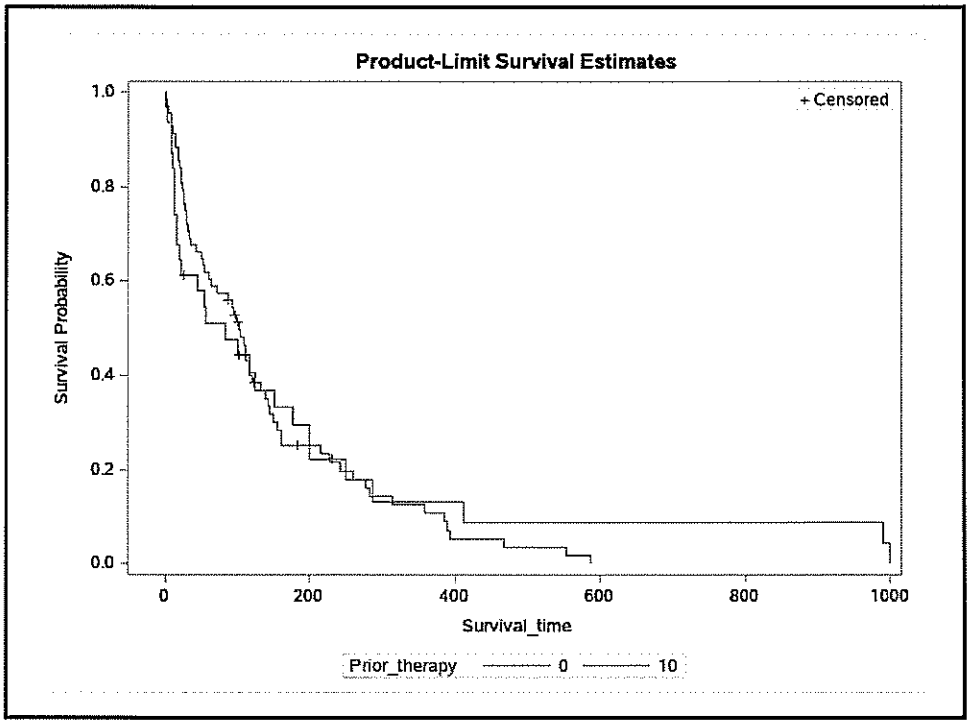

Figure 3

Using the information above,

a. State the null and alternative hypotheses in terms of the hazard ratio θ using both hazards and survival functions and state how θ affects the direction of effects. **[7]**

b. Using the p-value approach, what would your decision be? Provide a brief explanation. **[3]**

Question 3 [30 marks]

3.1 Derive the estimate of the survival function from the expression of the likelihood function of the hazard function up to time t_i. **[10]**

$$L\left(h_{j,j\le i} / d_{j,j\le i, n_j, j\le i}\right) = \prod_{j=1}^{i} h_j^{d_j} \left(1-h_j\right)^{n_j - d_j}$$

3.2 Consider the following output.

Testing Global Null Hypothesis: BETA=0

Test	Chi-Square	DF	Pr > ChiSq
Likelihood Ratio	465.6779	5	<.0001
Score	545.7160	5	<.0001
Wald	526.5441	5	<.0001

Type 3 Tests

Effect	DF	Wald Chi-Square	Pr > ChiSq
gender	1	5.5483	0.0185
AGE82	1	347.7227	<.0001
MARIED82	1	23.0021	<.0001
BMI82	1	10.5311	0.0012
CESD82	1	47.7229	<.0001

Analysis of Maximum Likelihood Estimates

Parameter		DF	Parameter Estimate	Standard Error	Chi-Square	Pr > ChiSq	Hazard Ratio	95% Hazard Ratio Confidence Limits	
gender	F	1	[redacted] A	0.08025	5.5483	0.0185	0.828	0.707	0.969
AGE82		1	0.09663	0.00518	347.7227	<.0001	1.101	1.090	1.113
MARIED82		1	-0.41698	0.08694	23.0021	<.0001	0.659	0.556	0.781
BMI82		1	-0.02727	0.00840	10.5311	0.0012	0.973	0.957	0.989
CESD82		1	0.02573	0.00372	47.7229	<.0001	[redacted] B	1.019	1.034

Using the information above.

a. Write down, the SAS program that was run to have the above results. **[5]**

b. Determine the values of:

i. A **[2]**

ii. B, and interpret it. **[2, 3]**

c. Formulate the hypothesis relating to goodness of fit of the model. **[4]**

d. Using the p-value approach, what would your decision be? Provide also a brief comment. **[4, 2]**

e. Write down the mathematical model, and explain it briefly. **[5, 5]**