



**AMREF INTERNATIONAL UNIVERSITY
SCHOOL OF PUBLIC HEALTH
DEPARTMENT OF COMMUNITY HEALTH
MASTERS IN PUBLIC HEALTH**

END OF SEPTEMBER-DECEMBER TRIMESTER 2022 EXAMINATIONS

MAP 716: BIostatistics II AND COMPUTING

DATE: 9th December, 2022

TIME: **Start:** 1600 Hours **Finish** 1900 Hours

INSTRUCTIONS

1. This exam is marked out of 100 marks
2. This Examination comprises TWO Sections
Section A: Compulsory Question (25 marks)
Section B: Long Answer Questions (75 marks)
3. All questions in Section A are compulsory and Answer any THREE questions in Section B
4. This online exam shall take 3 Hours
5. Late submission of the answers will not be accepted
6. Ensure your web-camera is on at all times during the examination period
7. No movement is allowed during the examination
8. Idling of your machine for 5 min or more will lead to lock out from the exam
9. The Learning Management System (LMS) has inbuilt integrity checks to detect cheating
10. Any aspect of cheating detected during and or after the exam administration will lead to nullification of your exam
11. In case you have any questions call the invigilator for this exam on Tel. 0722840012 and or the Head of Department on Tel +254720573449
12. For adverse incidences please write an email to: amiu.examinations@amref.ac.ke

SECTION A: COMPULSORY

Question 1

A public health research student is interested in the effects of sex and dietary group on systolic blood pressure (SBP) The "dietary group" factor includes three groups: Intervention A, intervention B and normal diet. Sex and dietary group may be independent or they may interact with each other. The results from Two-way ANOVA model predicting mean SBP level are tabulated in the tables below. Use the information provided to answer this question.

Table 1 Descriptive statistics

Variable 1	Variable 2	N	Mean	SD
Male	Intervention A	30	117.25	10.63
Male	Intervention B	25	119.01	8.49
Male	Normal diet	45	120.78	8.82
Female	Intervention A	25	110.93	10.37
Female	Intervention B	30	110.85	9.57
Female	Normal diet	45	117.78	9.35

Table 2: Two Way ANOVA

	DF	Sum of squares	Mean square	F statistics	P value
Sex	1	1594.39725	1594.39725	17.64529	Blank ?
Diet category	2	1137.86004	?? blank ?	6.29634	Blank ?
Interaction	2	245.48621	Blank ?	1.3584	0.2595
Model	5	2794.86329	Blank ?	6.19	Blank
Error	194	17529.50212	90.35826		
Total	199	20324.36541			

- a) Make a summary of the descriptive statistics (Table 1) (2 marks)
- b) What are the assumptions of Two Way ANOVA technique (3 marks)
- c) Complete the blanks (?) in the Two Way ANOVA table (3marks)
- d) State the hypotheses associated with the Two way ANOVA in this study (6 marks)
- e) Interpret the results from the ANOVA table in line with your stated hypotheses in (d) above (6 marks)
- f) Write a summary of the findings (5 marks)

SECTION B: SELECT ANY THREE (3) QUESTIONS -75 MARKS

Question 2

In a class of eighteen (18) MPH students, a biostatistics lecturer instructor compared male and females on the duration of time it take them to tackle a biostatistics problem on binomial distribution. The results are recorded in the table below. We would like to determine if there is a difference in time it takes male and female students to tackle a binomial distribution problem:

<i>Gender</i>	<i>01</i>	<i>02</i>	<i>03</i>	<i>04</i>	<i>05</i>	<i>06</i>	<i>07</i>	<i>08</i>	<i>09</i>	<i>10</i>
<i>Male</i>	18.5	16.0	22.7	17.6	18.9	25.5	16.5	24.2		
<i>Female</i>	12.6	14.5	20.5	10.7	15.9	19.6	12.0	15.5	11.9	14.9

- Develop a null and alternative hypotheses for this data (4 marks)
- Use the Mann-Whitney U test to test your hypotheses (12 marks)
- What conclusions do make from your analysis? (5 marks) :
- Why was the Mann-Whitney U-test preferred for this data (4 marks)

Question 3

A researcher analyzed data on retention to HIV/AIDS care up to 24 months. At multivariate level the researcher used logistic regression to obtain predictors of retention to care. He also applied Kaplan-Meir survival curves. Study the table below and Kaplan-Meir survival curve and answer the following questions

- Complete the column on P by indicating either $P <$ or $P >$ value (4 marks)
- Distinguish between crude OR and adjusted OR (3 marks)
- Interpret the results based on WHO stage (5 marks)
- Interpret the results based on CD level (5 marks)
- Interpret the results based on marital status (5 marks)
- Make a statement to summary the survival curve (3 marks)

Table 1: Multivariate logistic regression for factors associated with retention

Variable	Crude OR (95% CI)	Adjusted Odds ratio(95% CI)	P
WHO stage 1	3.943(1.033-15.043)	3.611 (1.020, 12.779)	?
WHO Stage 2	2.247(0.631-7.996)	2.191(0.656, 7.318)	?
WHO Stage 3	1.672(0.479-5.829)	1.590 (0.486, 5.205)	?
WHO Stage 4	1	1	
CD4 (below 200)	0.204(0.062-0.676)	0.304(0.0106-0.873)	
CD4 (201-400)	0.359 (0.107-1.201)	0.582(0.201-1.682)	0.32
CD4 (401- 600)	2.652(0.390-18.013)	2.333(0.482-11.297)	?
Over 600	1	1	
Single	0.153(0.014-1.669)	0.208 (0.020-2.144)	0.19
Married	0.466(0.046-4.773)	0.578(0.059-5.650)	?
Separated	0.823(0.056-12.011)	0.583(0.044- 7.661)	?
Divorced	0.191(0.015-2.471)	0.200(0.017-3.286)	0.20
Widowed	0.405(0.038-4.3338)	0.522(0.051-5.324)	?
Cohabiting	1	1	
Employed yes	0.705(0.364-1.366)	0.838 (0.462-1.522)	0.56
Employed No	1	1	

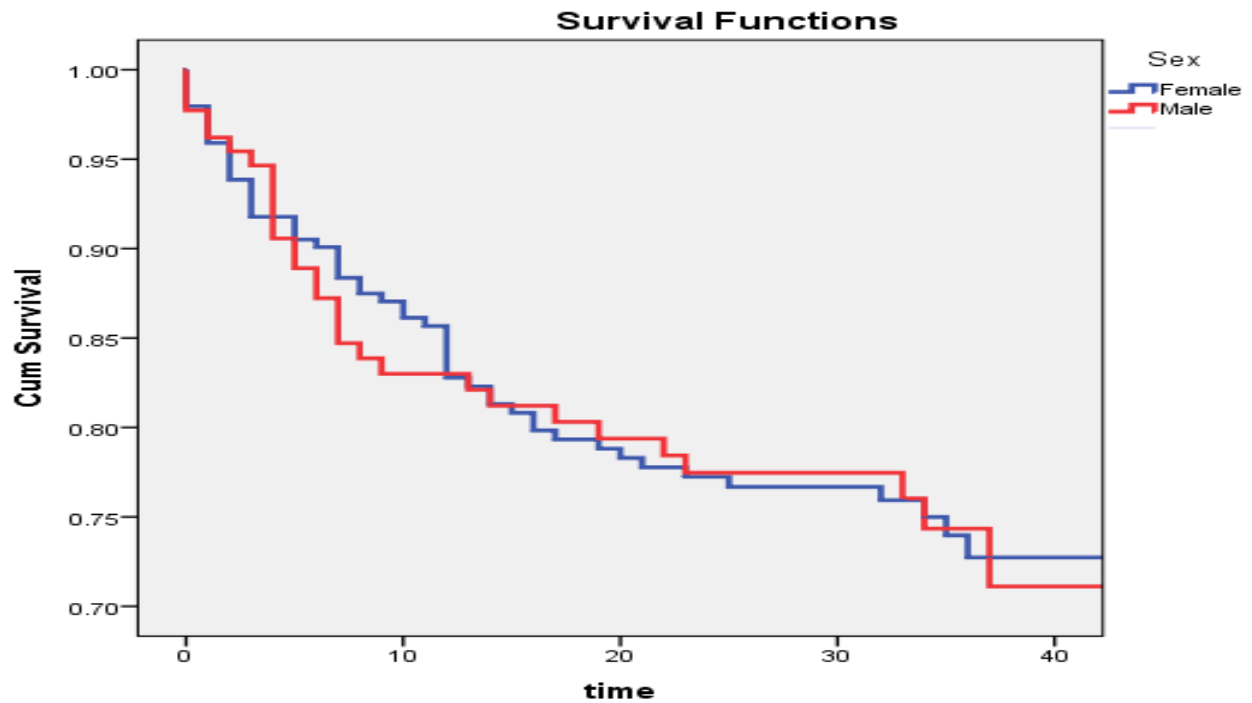


Figure 11 : Kaplan Meier probability of retention to care by sex

Question 4

Explain with appropriate examples the application of the following statistical techniques

- a) Simple linear regression (4 marks)
- b) Multiple linear regression (5 marks)
- c) One way ANOVA (6 marks)
- d) Independent samples T-test (5 marks)
- e) One Sample T-test (5 marks)

Question 5

a) Explain the application of survival analysis using an example of public health importance (6 marks)

b) In a study, time to cancer remission is of interest. An oncology clinic began record keeping in 2000. The dataset contains complete information of all their patients since 2000 (i.e the treatment date and date of remission for those who got into remission). There is also information on patients who were on treatment prior to Jan 2000 but got remission after January 2000. There is no information on patients who started treatment and got into remission before January 2000. The time unit in the study was year.

- I. What is the appropriate time scale for the study and why? (3 marks)
- II. Define these two terms as used in survival analysis
 - i. Truncation (2marks)
 - ii. Censoring (2 marks)
- III. For the four selected patients described below what type of censoring and truncation (if any) are represented and why?
 - i. A patient started on treatment in 2005 and had remission in 2007 (3 marks)
 - ii. A patient started on treatment in 2005 and has had no remission (3marks)
 - iii. A patient started on treatment in 1995 and had remission in 2001 (3marks)
 - iv. A patients started on treatment in 1990 and has had no remission (3 marks)

Question 6

- a) Explain the application of principal components approach. (10 marks)
- b) For most parametric tests, there is alternative non-parameter method. Explain at least five parametric tests and their alternative non parametric tests (15marks)