AMREF INTERNATIONAL UNIVERSITY

SCHOOL OF PUBLIC HEALTH

DEPARTMENT OF COMMUNITY HEALTH

MASTER OF PUBLIC HEALTH

END OF SEMESTER EXAMINATION DECEMBER 2024

MAP 716: BIOSTATISTICS II AND COMPUTING

DATE: WEDNESDAY 11TH DECEMBER 2024

TIME: THREE HOURS **START:** 5.00PM HOURS **FINISH** 8.00PM HOURS

INSTRUCTIONS

1. This exam will be graded 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 your machine for 5 minutes or more will lead to a lockout from the exam.
9. The Virtual Assessment System (VAS) has built-in integrity checks to detect cheating.
10. Any aspect of cheating detected during and or after the exam administration will lead to disciplinary measures.
11. In case you have any questions, call the invigilator for this exam Dr. Alice Lakati on Tel. +254722840012 and or the Head of Department Dr. Faith Muhonja on Tel +254 723 742 370 or ICT related questions Mr. Patrick Njine on +254725835496
12. For adverse incidences, please write an email to amiu.examinations@amref.ac.ke

SECTION A: COMPULSORY (25 MARKS)

Question 1

The table below shows the results from a multiple regression analysis to assess what factors can predict barriers to mammography screening in women. The predictors were Age in years, Education (years of formal education) and Race (1=African, 0= Other).

Overall model: F = 16.58, df = 3 and 1035, p < .0001

R2= .046 Adjusted R2= .043

Least squares estimates of parameters

Variable	df	Regression Coefficient	Standard Error	t-value	Pr>|t|
Intercept	1	26.63	2.277	11.69	<.0001
Age	1	0.087	0.026	?	0.0024
Education	1	-0.265	0.088	?	0.0011
Race	1	3.218	0.559	?	<.0001

---

a) What hypothesis is being tested by the F-value? (2 marks)

b) What is the conclusion from this test? (3 marks)

c) What is the interpretation of the R-square (3 marks)

d) What is the difference between the R-square and the adjusted R square (2 marks)

e) Fill in the missing t-values (3 marks)

f) State the null and alternative hypothesis being tested by the t-value associated with Age (3 marks)

g) What is the conclusion for this test? (4 marks)

h) What is the interpretation of the intercept? (2.5 marks)

i) What is the interpretation of the coefficient associated with education? (2.5 marks)

SECTION B: ANSWER ANY THREE (3) QUESTIONS (75 MARKS)

Question 2

Amref International University surveyed student satisfaction levels across three service platforms: **Phone Support**, **Email Support**, and **Live Chat**. Descriptive and inferential statistical analyses were conducted; the results are provided below.

Study the output carefully, and answer the following questions:

a) State the possible null and alternative hypotheses for comparing satisfaction levels across the three service platforms. (10 marks)

b) Based on the results of the inferential test, interpret whether there are statistically significant differences in satisfaction levels among the three platforms. Support your interpretation with details from the output. (10 marks)

c) Explain why the Kruskal-Wallis test was used for this analysis instead of a parametric test. (5 marks)

NPar Tests

Descriptive Statistics

	N	Mean	Std. Deviation	Minimum	Maximum
satisfaction	90	5.3667	2.64129	1.00	9.00
group	90	2.00	.821	1	3

Kruskal-Wallis Test

Ranks

	group	N	Mean Rank
satisfaction	Email Support	30	40.17
	Live Chat	30	51.87
	Phone Support	30	44.47
	Total	90	

Test Statistics[a,b]

	satisfaction
Kruskal-Wallis H	3.125
df	2
Asymp. Sig.	.210

a. Kruskal Wallis Test

b. Grouping Variable: group

Question 3

a) Explain the following concepts in the context of statistical analysis:

i) statistical model (5 marks)

ii) Multicollinearity (5 marks)

iii) Residual analysis (5 marks)

b) Use an appropriate example to describe the application of **Principal Components Analysis (PCA)** in a public health context. (10 marks)

Question 4

A public health research student is interested in the effects of sex and dietary groups on systolic blood pressure (SBP). The "dietary group" factor includes Intervention A, intervention B and regular diet. Sex and dietary groups may be independent or interact with each other. The results from the 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 (3 marks)

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)

Question 5

A researcher analyzed data on retention in HIV/AIDS care over 24 months. At the multivariate level, logistic regression was used to identify predictors of care retention and Kaplan-Meier survival curves were applied to illustrate retention trends over time. Refer to the table and Kaplan-Meier survival curve provided, and answer the following questions:

a) Complete the missing values in the P-value column. (4 marks)

b) Explain the difference between crude odds ratio (OR) and adjusted odds ratio (AOR). (3 marks)

c) Interpret the logistic regression results for **WHO stage** as a predictor of retention in care. (5 marks)

d) Interpret the logistic regression results for **CD4 level** as a predictor of retention in care. (5 marks)

e) Interpret the logistic regression results for **marital status** as a predictor of retention in care. (5 marks)

f) Provide a summary statement describing the retention trends shown in the Kaplan-Meier 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	

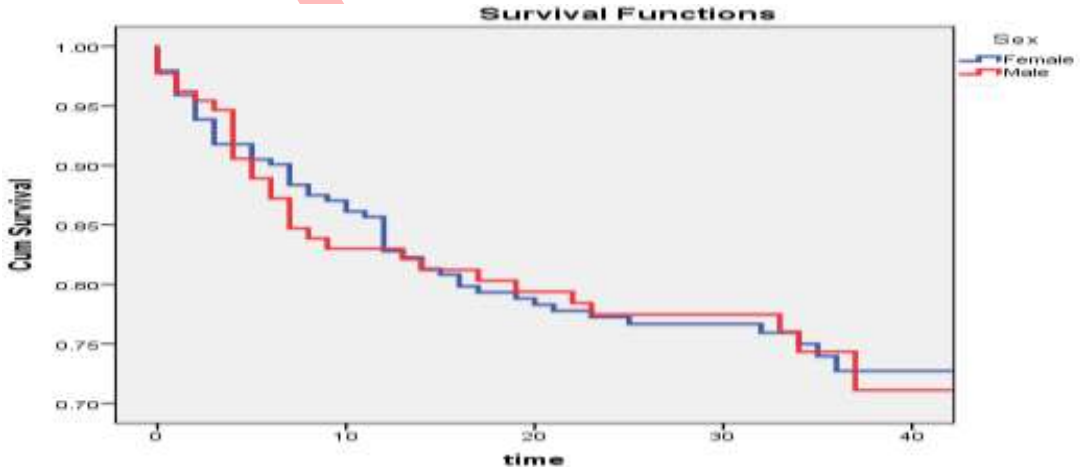

Figure 11 : Kaplan Meier probability of retention to care by sex

Question 6

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 a 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 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. Patients started on treatment in 1990 and have had no remission (3 marks)