AMREF INTERNATIONAL UNIVERSITY

SCHOOL OF PUBLIC HEALTH

DEPARTMENT OF COMMUNITY HEALTH

MASTER OF PUBLIC HEALTH

END OF SEMESTER EXAMINATION APRIL 2023

MAP 716: BIOSTATISTICS II AND COMPUTING
DATE: 28th April 2023,
TIME: Three Hours **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

The following table shows the result of a multivariate logistic regression analysis from a study of 5209 participants where 9 covariates were measured at baseline. The dependent variable was whether coronary heart disease (CHD) was present (coded as 'one') or absent (coded as 'zero') after 10 years. Study the table and answer the questions below

Variable (x_i)	Definition	b_i	OR	Confidence Interval for OR
Sex	M = 0, F = 1	-1.588	0.20	0.14, 0.29
Age	years	0.081	1.08	1.07, 1.10
Height	inches	-0.053	0.95	0.95, 1.00
SBP*	mm Hg	0.009	1.01	1.00, 1.02
DBP**	mm Hg	0.006	1.01	1.01, 1.02
Cholesterol	mg/ml	0.007	1.01	1.00, 1.01
ECG abnormal	Y = 1, N = 0	0.854	2.35	1.67, 3.31
Relative weight	***	1.359	3.89	1.89, 8.00
Alcohol consumption	oz/month	-0.059	0.94	0.88, 1.01
Constant term	*a* = -5.370			

*SBP = systolic blood pressure **DBP = diastolic blood pressure ***100wt/(median wt)% Based on the table above.

a) Based on the results, which group is more likely to suffer from CHD is it men or women? Explain your answer (3 marks)
b) Based on the results which group is more likely to suffer CHD when you compare persons with an ECG? Explain your answer (3 marks)
c) State the regression equation (5 marks)
d) What a detailed summary of the results (10 marks)
e) State your overall conclusions focusing on key points (4 marks)

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

Question 2

a) Why is it necessary to carry out univariate and bivariate analysis before multivariate analysis? explain (5 marks)
b) With examples explain four statistical techniques that are applicable at bivariate level with appropriate examples for each (20 marks)

Question 3

a) Explain the applications of the following non parametric tests
   - **i.** Median test (3.5 marks):
   - **ii.** Wilcoxon sum rank test (3.5 marks)
   - **iii.** Mann Whitney U-test (4 marks)
   - **iv.** Kruskal Wallis (4 marks):

b) Explain the advantages and disadvantages of non-parametric methods (10 marks)

Question 4

a) Explain the application of analysis of covariance (ANCOVA) (3 marks)

b) With nan appropriate example, explain the term interaction between variables (4 marks)

c) Explain the assessment of multicollinearity in regression analysis (4 marks)

d) The table below shows an output of ANCOVA from a study on the effect of chewing gum on a child DMFT (decayed, missing filled teeth). The research compared two groups (children who were exposed to chewing gum and those who were not) An interaction term was included in the model.
   i. State the various hypotheses for this test (6 marks)
   ii. Interpret results from the table and write your detailed conclusions (8 marks)

Source	Df	SSS.	M.S.	F	P
Group	2	278.974	139.487	11.25	?
Baseline DMFS	1	277.894	277.894	22.411	?
Group * baseline DMFS	2	243.900	121.9501	9.834	?
Error	94	1165.592	12.399		

Total					

Question 5

The table below shows the results from a multiple regression analysis carried out 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-value's **(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)**

Question 6

a) Describe your understanding of the following concepts
   i. Statistical model (4 marks)
   ii. Cox proportional hazards model (4 marks)
   iii. Multicollinearity (4 marks)
   iv. Residual analysis (4 marks)

b) Factor analysis allows for selection of factors that mainly explain the interrelationship. The process involves finding estimates of factor loading and their communalities. Describe the approach used in achieving loadings and communalities (9 marks)