MASINDE MULIRO UNIVERSITY OF SCIENCE AND TECHNOLOGY (MMUST)

MAIN CAMPUS

UNIVERSITY EXAMINATIONS
MAIN EXAM

2023/2024 ACADEMIC YEAR

FIRST YEARFIRST SEMESTER EXAMINATION
FOR THE MASTER OF PUBLIC HEALTH

COURSE CODE: PHC 815

COURSE TITLE: BIOSTATISTICS

DATE: 11ᵀᴴ DECEMBER 2023   **TIME: 2.00PM- 5.00PM**

INSTRUCTIONS TO CANDIDATES:
Answer Question one and any other three questions
TIME: 3 Hours

Question One (Compulsory) (25 MARKS)

a. The following data represents the withdrawals and the balances(in '000Ksh) of some Bank customers:

Balance	51	59	64	76	93	106	125	149	171	199
Withdrawal	5	1	6	4	10	12	14	12	15	10

Consider the following output

$ANOVA^b$					
Model	Sun of squares	df	Mean square	F	Sig.
Regression	12635.723	1	12635.723	9.714	0.014^a
Residual	10406.377	8	1300.797		
Total	23042.100	9			
a:Predictors(constant), withdrawal					
b:Dependent variable: Balance					

Table 1 output

$Coefficients^a$						
	Unstandardized Coefficients		Unstandardized Coefficients		Collinearity statistics	
Model	B	Std.Error	Beta	t	Tolerance	VIF
Constant	37.639	25.666		1.466		
Withdrawal	8.052	2.583	.471	3.117	1.000	1.000
a:Dependent variable: Balance						

Table 2 output

   i.   Write down the fitted model. (1 marks)

   ii.   Is the overall regression model significant at 5% level of significance? Explain. (3 marks)

   iii.   Write down a well commented R program that could have given the above output. (6 marks)

b. Fit least square regression model to the data below. (15 marks)
(x,y,z); (2,3,4), (6,5,10), (1,8,5), (3,4,12), (8,5,9)

Question Two (25 MARKS)

The following data represents the marks scored by students in a class.

Class	200-249	250-299	300-349	350-399	400-449	450-499	500-549
frequency	7	19	27	15	12	12	8

Determine:

a. Mean using short cut method (6 marks)

b. Mode (4 marks)
c. Median (4 marks)
d. Variance (6 marks)
e. Standard deviation (2 marks)
f. Coefficient of variation (3 marks)

Question Three (25 MARKS)

a. Define Analysis of variance hence state 4 assumptions involved in its use in statistics.

(5 marks)

b. A company appoints 4 sales men A, B,C and D and observes their sales performance in three seasons; summer, winter and monsoon. The figures are in Ksh,000;

	A	B	C	D
Summer	13	16	16	14
Winter	17	16	17	16
Monsoon	13	14	15	15

Carry out an analysis of variance and discuss your results exhaustively as a consultant to this company. (20 marks)

Question Four (25 MARKS)

a. Consider two series of data X and Y such that;

$n_1 = n_2 = 9,\quad \overline{X} = 35.22,\quad \sum x^2 = 11361,\quad \overline{y} = 31.56\quad and\quad \sum y^2 = 9122$

Estimate the confidence interval for the difference between means $\mu_1 - \mu_2$, *at 5% level of significance.*

(12 marks)

b. The following is a cohort study to investigate the association between recent use of oral contraceptives and myocardial infarction.

Recent oral contraceptive use	Myocardial infarction.	Controls
Yes	29	135
No	205	1607

    i. Calculate the relative risk (5 marks)
    ii. Calculate the 95% confidence interval for the relative risk (3 marks)
    iii. Calculate the odds ratio (2 marks)
    iv. Calculate the 95% confidence interval for the odds ratio (3 marks)

Question Five (25 MARKS)

a. Outline the steps used in carrying out Chi-square test for independence in a 2x2 contingency table. (5 marks)

b. A hypothetical cohort study in which 5000 women who used oral contraceptives and the same number who did not were followed for 10 years. The number of deaths due to myocardial infarction (Heart disease) in e ach group was recorded. 200 oral contraceptive users were lost during the followup period due to migration and other causes.

	Death from HD	
OC Use	Yes	No
Yes	7	4793
No	2	4823

Calculate the Chi-square statistic, test the hypothesis and interpret your results. (10 marks)

c. The following data relate to household income, expenditure and savings;

Household	1	2	3	4	5	6	7
Income	200	300	120	70	256	180	150
Expenditure	50	150	60	30	59	60	80
Savings	120	130	40	15	130	80	40

Determine the correlation coefficient between;
i. Income and expenditure keeping savings constant
ii. Multiple correlation of income on expenditure and savings.

(10 marks)