(Knowledge for Development)

KIBABII UNIVERSITY

UNIVERSITY EXAMINATIONS

2020/2021 ACADEMIC YEAR

THIRD YEAR SECOND SEMESTER

MAIN EXAMINATION

FOR THE DEGREE OF BACHELOR OF SCIENCE

COURSE CODE: STA 322

COURSE TITLE: REGRESSION ANALYSIS AND ANOVA

DATE: 7/10/2021 **TIME:** 2:00 pm - 4:00 pm

INSTRUCTIONS TO CANDIDATES

Answer Question One and Any other TWO Questions

TIME: 2 Hours

QUESTION 1:(30 marks)

(a) Let variables X and Y be related by a simple linear regression model of the form,

$$Y_i = \beta_0 + \beta_1 X_i + \varepsilon_i \qquad \text{for } i = 1,2,3,......n$$

Where,

ε_i is the model error and $\beta_0 and \beta_1$ are the intercept and Regression coefficient respectively.

Consider $\hat{\beta}_0$ and $\hat{\beta}_1$ to be the estimators of $\beta_0 and \beta_1$ respectively. By the least squares criterion, show that

i) $\hat{\beta}_1 = \dfrac{\sum_{i=1}^{n} x_i y_i - n\overline{xy}}{\sum_{i=1}^{n} x_i^2 - n\overline{x}^2}$ (6 marks)

ii) $\hat{\beta}_0 = \overline{y} - \hat{\beta}_1 \overline{x}$ (6 marks)

(b) Prove that,

(I) $E(\hat{\beta}_0) = \beta_0$, (4 marks)

ii) $E(\hat{\beta}_1) = \beta_1$ (4 marks)

and explain the significance of these results.

(c) A departmental store gives in-service training to salesmen followed by a test. It is experienced that the performance regarding sales of any salesman is linearly related to the scores secured by him or her. The following data give test scores and sales made by nine salesmen during fixed period.

Test score (X):	16	22	28	24	29	25	16	23	24
Sales (in Kshs), Y:	35	42	57	40	54	51	34	47	45

(i) Fit a simple linear regression model to these data (5 marks)

(ii) Formulate a hypothesis to test the statistical significance of the regression coefficient and hence perform the test at 0.05 level of significance (5 marks)

QUESTION 2: (20 marks)

The removal of ammoniac nitrogen is an important aspect of treatment of Leachate at landfill sites. The rate of removal (in percent per day) is recorded for several days for each of several treatment methods. The results are presented in the following table.

Treatment	Rate of Removal			
A	5.21	4.65		
B	5.59	2.69	7.57	5.16
C	6.24	5.94	6.41	
D	6.85	9.18	4.94	
E	4.04	3.29	4.52	3.75

(a) Construct an ANOVA table. What is the F-value in this case? (10+5 marks)

(b) Can you conclude that the treatment methods differ in their rates of removal? (5 marks)

QUESTION 3: (20 marks)

The ranks of 12 students according to their marks in mathematics and statistics were as follows:

Student no.	1	2	3	4	5	6	7	8	9	10	11	12
Mathematics	5	2	1	6	8	11	12	4	3	9	7	10
Statistics	4	3	2	7	6	9	10	5	1	11	8	12

(i) Obtain the rank correlation coefficient and hence comment on the students' performance in Mathematics and Statistics. (7 marks)

(ii) Fit a simple linear regression model to the data values. What do you conclude about the expected performance in mathematics in relation to that in Statistics? (8 marks)

(iii)Do the above results in (i) and (ii) lead to the same logical conclusion on the performance of Students in the two subjects? Discuss. (5 marks)

QUESTION 4: (20 marks)

(a) Prove that the Least squares estimates of the multiple regression coefficients are given by

$$B = (X'X)^{-1}X'Y$$

$$\text{Where} \quad X = \begin{pmatrix} 1, x_{11}, x_{12}, \ldots\ldots x_{1k} \\ 1, x_{21}, x_{22}, \ldots\ldots x_{2k} \\ \\ \\ 1, x_{n1}, x_{n2}, \ldots\ldots x_{nk} \end{pmatrix}, \quad Y = \begin{pmatrix} y_1 \\ y_2 \\ . \\ . \\ y_n \end{pmatrix} \text{ and}$$

$$B = \begin{pmatrix} \hat{\beta}_0 \\ \hat{\beta}_1 \\ . \\ . \\ \hat{\beta}_k \end{pmatrix}$$

X' is the transpose of X and $(X'X)^{-1}$ is the inverse of (X'X). (7 marks)

(b) In a given study, suppose the following were obtained

$$X'X = \begin{pmatrix} 8,,25,,16 \\ 25,,87,,55 \\ 16,,55,,36 \end{pmatrix} \text{ and } X'Y = \begin{pmatrix} 637000 \\ 2031100 \\ 1297700 \end{pmatrix}$$

Obtain the Least Squares estimates of the multiple regression coefficients.
(13 marks)

QUESTION 5 : (20 marks)

An experiment gives values of the pressure, P of a given mass of gas corresponding to various values of volume, V. According to thermodynamics principles, a relationship having the form $PV^{\gamma} = C$, where γ and C are constants, should exist between the variables.

Table

Volume, V (in^3)	54.3	61.8	72.4	88.7	118.6	194.0
Pressure, P (Ib/m^3)	61.2	49.5	37.6	28.4	19.2	10.1

(a) Find the values of γ and C (10 marks)
(b) Write an equation connecting P and V (4 marks)
(c) Estimate P when V= 100.00 in^3 (6 marks)