(Knowledge for Development)

KIBABII UNIVERSITY

UNIVERSITY EXAMINATIONS

2020/2021 ACADEMIC YEAR

THIRD YEAR SECOND SEMESTER

SPECIAL/SUPPLIMENTARY EXAMINATION

FOR THE DEGREE OF BACHELOR OF SCIENCE

COURSE CODE: **STA 322**

COURSE TITLE: REGRESSION ANALYSIS AND ANOVA

DATE: 20/01/2022 **TIME**: 11:00 - 1:00 PM

INSTRUCTIONS TO CANDIDATES

Answer Question One and Any other TWO Questions

TIME: 2 Hours

This Paper Consists of 5 Printed Pages. Please Turn Over.

QUESTION 1: (30 marks)

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

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

Where,

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

If $\hat{\beta}_1$ is an estimator of β_1,the regression coefficient, then by the Least squares criterion show that

$$\hat{\beta}_1 = \frac{\sum_{i=1}^{n} x_i y_i - n\overline{xy}}{\sum_{i=1}^{n} x_i^{\,2} - n\overline{x}^2} \qquad \text{(8 marks)}$$

What implication would it have to the relationship between X and Y when $\hat{\beta}_1$=0 ?

(1 marks)

(b) 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)

(iii) Basing on the results so obtained in part b(ii) above, interpret your results

(2 marks)

(c) The following table gives the gain in body weight(kg) per heifer during four grazing treatments.

Heifer no.	Gain in body weight (kg) Treatments				Total
	T_1	T_2	T_3	T_4	
1	67.3	74.2	63.1	48.7	
2	36.9	42.2	32.9	49.0	
3	63.2	58.6	59.2	62.0	
4	26.8	36.6	42.4	38.8	
5	54.8	54.6	34.0	48.2	
6	64.2	81.8	65.6		
7	81.4				
	394.6	348.0	297.2	246.7	1286.5

Test the hypothesis that the mean gain in weight of Heifers under the four treatments is equal against the alternative that at least two means are different, that is,

H_0: $\mu_1 = \mu_2 = \mu_3 = \mu_4$ versus

H_1: at least two means are different at 5% level of significance

(9 marks)

QUESTION 2: (20 marks)

(a) Define correlation

(b) Suppose that we want to determine on the basis of the following data whether there is a relationship between the time, in minutes, it takes a secretary to complete a certain form in the morning and the late afternoon

Morning (x)	Afternoon (y)
8.2	8.7
9.6	9.6
7.0	6.9
9.4	8.5
10.9	11.3
7.1	7.6
9.0	9.2
6.6	6.3
8.4	8.4
10.5	12.3

(i) Compute and interpret the sample correlation coefficient (8+2 marks) (ii)Test the hypotheses,

$$H_0 : \rho = 0$$
$$H_1 : \rho \neq 0$$

at $\alpha = 0.01$ level of significance. Comment on the results so obtained

(8 + 2 marks)

QUESTION 3: (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)

QUESTION 4: (20 marks)

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

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

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

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

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)

The removal of ammoniacal 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)