(Knowledge for Development)

KIBABII UNIVERSITY

UNIVERSITY EXAMINATIONS

2021/2022 ACADEMIC YEAR

THIRD YEAR SECOND SEMESTER

MAIN EXAMINATION

FOR THE DEGREE OFBACHELOR OF SCIENCE

COURSE CODE: STA 322

COURSE TITLE: REGRESSION ANALYSIS AND ANALYSIS OF VARIANCE

DATE: 05/09/2022   **TIME:** 2:00 PM – 4:00 PM

INSTRUCTIONS TO CANDIDATES

Answer Question One and Any other TWO Questions

TIME: 2 Hours

QUESTION ONE (30 MARKS)

(a) Let X and Y be linearly related by a model of the form

$$Y_i = \beta_0 + \beta_1 X_i + \varepsilon_i$$

Where ε_i is the model error and β_0, β_1 are the intercept and regression coefficient respectively

Show that i) $\hat{\beta}_0 = \bar{Y} - \hat{\beta}_1 \bar{X}$ (5marks)

ii) $E(\hat{\beta}_0) = \beta_0$, if $\hat{\beta}_0$ is an estimator of β_0 (5marks)

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

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

Where $X = \begin{pmatrix} 1, x_{11}, x_{12}, ...x_{1k} \\ 1, x_{21}, x_{22}, ...x_{2k} \\ \\ \\ 1, x_{n1}, x_{n2}, ...x_{nk} \end{pmatrix}$, $Y = \begin{pmatrix} y_1 \\ y_2 \\ . \\ . \\ y_n \end{pmatrix}$ 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).   (5 marks)

(c) 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.   (5 marks)

QUESTION TWO (20 MARKS)

Consider the following data on the number of hours, 10 students studied for a Statistics test and their scores in that test:

Hours studied (X)	Test score (Y)
9	58
10	65
14	73
4	37
7	44
12	60
22	91
1	21
17	84

(a) Find the equation of the least squares line that approximates the regression of the test scores on the number of hours studied   (10 marks)

(b) Predict the average test score of a student who ;
  i)   Studied 14 hours for the test,   (3 marks)
  ii)   Did not study at all for the test.   (4 marks)

(c) Give an interpretation of your results obtained in (a) above   (3 marks)

QUESTION THREE (20 MARKS)

Using the data in Question 2 above pertaining to the amount of time that 10 students studied for the statistics test and the scores that they obtained,

  i)   Test the null hypothesis $\beta = 3$ against the alternative hypothesis $\beta > 3$ at the 0.01 level of significance,   (6 marks)

  ii)   Test for the statistical significance of the regression coefficient at 5% level of significance and comment on the results so obtained.   (7 marks)

Construct a 95% confidence interval for β. Using the confidence interval obtained here, comment on the statistical significance of the regression coefficient β. If it turns out that β is statistically insignificant, what would be the implication to relationship between Y and X?
(4+2+1 marks)

QUESTION FOUR (20 MARKS)

If a set of paired data gives the indication that the regression equation is of the form $y = \alpha\beta^x$, it is customary to estimate α and β by fitting the line,

$$\log \hat{y} = \log \hat{\alpha} + x.\log \hat{\beta}$$ to the points $(x_i, \log y_i)$; $i = 1,2,3,\ldots\ldots n$ by the

method of least squares. Use this technique to fit an exponential curve of the form $\hat{y} = \hat{\alpha}.\hat{\beta}^x$ to the following data on the growth of Cactus grafts under controlled environmental conditions:

Weeks after grafting(x)	Height in inches(y)
1	2.0
2	2.4
4	5.1
5	7.3
6	9.4
8	18.3

(20 marks)

QUESTION FIVE (20 MARKS)

(a) In an experiment, welding fluxes with differing chemical compositions were prepared. Several welds using each flux were made. The results of hardness measurements of five welds using each of the four fluxes are presented below.

Flux	Sample values	Sample mean
A	250 264 256 260 239	253.8(μ_1)
B	263 254 267 265 267	263.2(μ_2)
C	257 279 269 273 277	271.0(μ_3)
D	253 258 262 264 273	262.0(μ_4)

Can we conclude that there are differences in the population means among the four flux types?

(8 marks)

(b) 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

i) Construct an ANOVA table. What is the F-value in this case?   (7+2 marks)

ii) Can you conclude that the treatment methods differ in their rates of removal?

(3 marks)