(Knowledge for Development)

KIBABII UNIVERSITY

UNIVERSITY EXAMINATIONS

2017/2018 ACADEMIC YEAR

THIRD YEAR SECOND SEMESTER

MAIN EXAMINATION

FOR THE DEGREE OF BACHELOR OF SCIENCE (MATHEMATICS)

COURSE CODE: STA 344

COURSE TITLE: REGRESSION ANALYSIS AND ANOVA

DATE: 02/08/18 **TIME:** 2 PM -4 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,

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

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, $E(\hat{\beta}_1) = \beta_1$ and explain the significance of these result. (4 marks)

(c) Consider a random sample X1, X2, X3, ...,Xn from a distribution having probability density function , f(x;ϑ) , ϑϵ .Find an estimator ϑ of ϑ which maximizes the likelihood function, L of the random sample, that is, obtain the maximum likelihood estimator of ϑ (6 marks)

(c) 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?. That is, test

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

H_1: not all the means are equal at 10% level of significance. (9 marks)

QUESTION 2: (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 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 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 modulation study on the yield of a certain plant gave the following data

Dry weight of plants(mg), Y	Root length(cm), X_1	Shoot length (cm), X_2
412	28.7	21.5
226	13.4	11.7
292	14.6	12.9
323	18.0	14.8
233	12.1	11.0
368	23.4	19.2
239	12.6	11.4
382	30.2	22.6
218	11.6	10.8
222	12.0	10.2
214	12.4	10.1

(a) Fit a linear regression equation of Y on X_1 and X_2. (12 marks)

(b) Test the statistical significance of each partial regression coefficient (8 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? (15marks)

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