MURANG'A UNIVERSITY OF TECHNOLOGY

SCHOOL OF PURE AND APPLIED SCIENCE

DEPARTMENT OF APPLIED SCIENCE

UNIVERSITY ORDINARY EXAMINATION

2017/2018 ACADEMIC YEAR

THIRD YEAR **SECOND** SEMESTER EXAMINATION FOR BACHELOR OF SCIENCE IN APPLIED STATISTICS WITH PROGRAMMING

AMS320 – STATISTICAL MODELLING

DURATION: 2 HOURS

DATE:

TIME:

Instructions to Candidates:

1. Answer **Section A** and **Any Other Two** questions in **Section B**.
2. Mobile phones are not allowed in the examination room.
3. You are not allowed to write on this examination question paper.

SECTION A – ANSWER ALL QUESTIONS IN THIS SECTION (30 Marks)

QUESTION ONE

a) i) What is a linear regression model? State a simple linear regression model and explain all the symbols that you have used. (3 Marks)

ii) Explain R^2 as used in regression analysis. (2 Marks)

iii) What is a logistic regression? (1 Mark)

iv) What is the difference between a linear and a non-linear regression? (2 Marks)

b) i) Use the data below to fit a linear regression model using least square method.

x	25	16	28	20	22	23	16	18
y	125	79	140	103	111	115	80	91

(5 Marks)

ii) Use the model obtained above to predict the value of x given $y = 88$. (2 Marks)

c) i) Consider the relationship $y_i = \propto + \beta X_i + e_i$ and $e_i \sim N(0, \delta^2)$

Show that under ordinary least square;

i. $E(\hat{\beta}) = \beta$ (3 Marks)

ii. $E(\hat{\propto}) = \propto$ (2 Marks)

ii) An experiment involving five observations is conducted to determine the probability of a relationship between the percentage of a certain drug in the blood stream (y) and the length of time (x) that it takes ti react to a stimulus. An incomplete ANOVA table for this is provided below

Source	Regression Fr.	S.S	M.S.S	Fraction
Regression		8.1667		
Error				0.611
Total	4			

Complete the table. (4 Marks)

d) Assuming the normal error model $y = a + bx_i + e_i$ and given the data below

x	20	30	30	40	50	60	60	60	70	80
y	50	73	69	87	108	135	132	128	148	170

Test the following hypothesis at $\propto = 0.05$

i. $H_0: b = 0 \quad vs\ H_1: b \neq 0$ (3 Marks)

ii. $H_0: b = 1.9\ vs\ H_1 \neq 1.9$ (3 Marks)

SECTION B – ANSWER ANY TWO QUESTIONS IN THIS SECTION

QUESTION TWO (20 MARKS)

a) i) Give any three (3) differences between simple regression and multiple regressions. (3 Marks)

ii) Suppose $y_i = \beta_i + \beta_0 x_i + e_i\,; i = 1,2,\cdots n$ is a multiple linear regression. Find the least square estimator of β_0 *and* β_i and state their variance and covariance equation. (7 Marks)

b) In a spring balance, when weights were added to the scale pan, the string stretches. The following table shows the results obtained when different load x were applied a random order and the length of the spring y.

x	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	1.0
y	10.7	11.3	12.0	12.4	13.0	13.7	14.5	15.1	15.6	16.0

i. Find the relationship between x and y in the form $y = a + bx_i + e_i$ where $e_i \sim N(0, \delta^2)$ random variables by method of least square. (4 Marks)

ii. Estimate total sum of squares (TSS) and regression sum of squares (RSS). (6 Marks)

QUESTION THREE (20 MARKS)

a) i) State the four (4) phases involved in model building process. (4 Marks)

ii) Explain briefly what we mean by generalized linear model (GLM). (2 Marks)

b) The effectiveness of a tablet containing X_1gm of drug 1 and X_2gm of drug 2 was being tested in trials. The following data results were obtained.

X_1	6	7	7	8	10	10	15
X_2	4	20	20	10	10	2	1
% effectiveness Y_i	49	55	50	41	17	26	16

i. Write down the predictor matrix X and the response vector Y. (2 Marks)

ii. Compute $X^T X$ *and* $X^T Y$ (3 Marks)

iii. Obtain the estimated regression equation. (4 Marks)

iv. Construct the ANOVA table. (5 Marks)

QUESTION FOUR (20 MARKS)

a) Show that the binomial distribution belongs to the exponential family and identify the link function to be used in binomial regression. (10 Marks)

b) In multiple regression $y_i = \beta_0 + \beta_1 Xi \quad + e_i$ prove that the variance of $\hat{\beta}_0 \; and \; \hat{\beta}_1$ are

$$Var(\hat{\beta}_0) = \frac{\delta^2 \sum_{i=1}^{n} X^2}{\sum_{i=1}^{n}(X_i - \bar{X})^2}$$

and

$$Var(\hat{\beta}_1) = \frac{\delta^2}{\sum_{i=1}^{n}(X_i - \bar{X})^2}$$

(10 Marks)