MANICALAND STATE UNIVERSITY OF APPLIED SCIENCES

FACULTY OF ENGINEERING, APPLIED SCIENCES & TECHNOLOGY

DEPARTMENT OF APPLIED STATISTICS

MODULE: REGRESSION AND ANOVA II

CODE: HAST 423

SESSIONAL EXAMINATIONS
APRIL 2023

DURATION: 3 HOURS

EXAMINER: NYAKUAMBA T

INSTRUCTIONS

1. *Answer **All** in Section A*
2. *Answer **three** questions in Section B.*
3. *Start a new question on a fresh page*
4. *Total marks 100*

Additional material(s): Non-programmable electronic scientific calculator, Statistical Tables

SECTION A (40 MARKS)

ANSWER ALL QUESTIONS

A1 (a) What is meant by regression analysis? [3]

(b) What justifies the inclusion of a disturbance or error term in regression analysis? [3]

A2. What do you understand by the following terms as used in regression analysis?

(a) autocorrelation [3]
(b) responds variable [3]
(c) laverage point [3]
(d) predictor [3]
(e) coefficient of determination [3]

A3 Suppose you have a regression model of the form

$$Y = a + bXi + ei \qquad \{ i = 1,2, \dots n\}$$

(a) state the assumption underlying this model [5]

(b) derive the least square estimates of the two parameters and show that both are unbiased. [6, 8]

SECTION B (60 marks)

Attempt any 3 Questions being careful to number them from B4 to B7

B4. Draw a scatter plot diagram of two variables that

(a). are typically almost un corelated [2]

(b). have a correlation coefficient close to 1
[2]

(c). have a correlation coefficient close to -1 [2]

(d). Prove that, if two random variables X and Y are such that $y = a + bX$

(Where a and b are constants), then $\rho xy = +1 \; or - 1$ [4]

(e). Let two random variables X and Y have correlation coefficient ρxy,

Prove that $-1 \leq \rho xy \leq 1$ [10]

B5.

A scientist collects experimental data on the radius of a propellant grain Y, as a function of powder temperature $X_1$and extension rate X_2.

He took 5 observations of each variable and the data is as follows

Y	25	30	28	32	40
X1	40	20	15	32	28
X2	20	15	15	24	20

Suppose the data can be described by the model

$Y_i = b_0 + b_1 X_i + b_2 X_2 + e_i$

Where e_i f0llows $N(0, \sigma^2)$ and $cov(e_i, e_j) = 0 \; if \; i \neq j$

(a). express the model in matrix form [2]

(b). obtain the decision matrix and calculate

$X'Y$ and $X'X$ [1, 3,3]

(c). calculate the least squares estimate of b given that

$$(X'X)^{-1} = \begin{bmatrix} 1.165 & -0.134 & -0.180 \\ -0.134 & 0.0041 & 0.0049 \\ -0.180 & 0.0049 & 0.0156 \end{bmatrix}$$

(i). calculate the hat matrix H and explain its importance in regression

[3, 4]

(ii).test the hypothesis that $H_0 : \beta_1 = 0 \ versus \ H_1 : \beta_1 \neq 0$

use $\alpha = 0.10$ and MSE = 7.815 [4]

B6. Use the following data to answer the following questions:

X	22	68	108	137	255	315	390	405	685	700	1100
y	0.756	2.4	3.2	4.7	9.3	12.0	13.4	14.4	24.5	26.00	38.0

I. Plot the data above and comment. [5]

II. Fit the model above [4]

III. Do you think the above model is the best?

Support your answer [4]

IV. State the assumptions satisfied by this model [3]

V. Estimate the error variance [4]

B7. Discuss the problems of

I. Multicollinearlity in multiple regression.

(a) What it is [2]

(b) How it affects regression analysis [5]

(c) What could be done about it [3]

ii. (a) Distinguish between step-wise regression and forward selection approach to regression [3]

(b) Explain the use of step wise regression [3] (c)Outline what is involved in the analysis of residuals. [4]

END OF EXAMINATION PAPER