MERU UNIVERSITY COLLEGE
OF SCIENCE & TECHNOLOGY
P.O. Box 972-60200 Meru - Kenya. Tel: 020-2092048, 020 2069349
Fax: 020-8027449

**University Examinations 2011/2012**

FIRST YEAR, FIRST SEMESTER EXAMINATION FOR THE DEGREE OF MASTER OF SCIENCE IN APPLIED STATISTICS

**STA 3102: PARAMETRIC REGRESSION ANALYSIS**

**DATE: JANUARY 2012** **TIME: 3 HOURS**

**INSTRUCTIONS:** *Answer Question **one** and any other **two** questions*

**QUESTION ONE (30 MARKS)**

a) Show that the Best Unbiased Linear Estimator (BLUE) is the Generalised Least Squares (GLS) estimator. (4 Marks)

b) Let the relationship between dependent variable y and k independent variables be given by $y = \beta_o + \beta_1 x_1 + \beta_2 x_2 + \cdots + \beta_k x_k + \varepsilon$. Find expressions for estimating the regression coefficients by the method of least squares. (5 Marks)

c) Data on moisture content (x) in the soil and root (y) length of an experimental plant gave the following summaries: $n = 14, \sum y_i = 572, \sum y_i^2 = 23{,}530, \sum x_i = 43, \sum x_i^2 = 157.42$ and $\sum x_i y_i = 1697.80$. Assume that the variables are related according to the simple linear regression model.

   i. Calculate the least squares estimates for the slope and the intercept.(4 Marks)

   ii. Use the equation of the fitted line to predict what root length would be observed when the moisture content is $x = 4.3$. (1 Mark)

   iii. Give a point estimate of the mean root length when moisture content is $x = 3.7$. (1 Mark)

   iv. Suppose that the observed value of moisture content at x=3.7 is y=46.1. Calculate the value of the corresponding residual. (1 Mark)

d) Explain how a multivariate regression model with interaction effects can be as a multiple linear regression model. (3 Marks)

e) A regression model is to be developed for predicting the ability of soil to absorb chemical contaminants. Ten observations have been taken on a soil absorption index (y) and two

regressors: $x_1$ = amount of extractable iron ore and $x_2$ = amount of bauxite. We wish to fit the model $Y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \varepsilon$. Some necessary quantities are;

$$(X^1X)^{-1} = \begin{bmatrix} 0.214653 & & \\ -0.007491 & 0.001617 & \\ -0.000340 & -0.000019 & 0.0000015 \end{bmatrix} \text{ and } X^1y = \begin{bmatrix} 725.82 \\ 8{,}008.37 \\ 274{,}811.31 \end{bmatrix}$$

i. Estimate the regression coefficients of the model specified above. (5 Marks)

ii. Predict the soil absorption ability if $x_{11} = 3$ and$x_{12} = 50$. (3 Marks)

iii. If the observed value $y_1 = 9.95$, calculate its corresponding residual. (3 Marks)

**QUESTION TWO (20 MARKS)**

a) The following data were obtained from twenty four specimens of paper that were made from five levels of hardwood concentration.

| | Observations | | | | | | |
|---|---|---|---|---|---|---|---|
| Hardwood Concentration (%) | 1 | 2 | 3 | 4 | 5 | 6 | **Total** |
| 5 | 7 | 8 | 15 | 11 | 9 | 10 | 60 |
| 10 | 12 | 17 | 13 | 18 | 19 | 15 | 94 |
| 15 | 14 | 18 | 19 | 17 | 16 | 18 | 102 |
| 20 | 19 | 25 | 22 | 23 | 18 | 20 | 127 |
| | | | | | | | **383** |

Describe the linear statistical model that can be used to test for differences between the hardwood concentration levels. (3 Marks)

By, first, formulating the appropriate hypothesis for testing the different hardwood concentrations' effect on the quality of paper produced, construct the ANOVA table for the data and test the hypotheses so formulated at $\propto = 0.05$. (17 Marks)

**QUESTION THREE (20 MARKS)**

a) Consider a multiple regression model with k regressors. Describe how to estimate the regression parameters by the Matrix approach. (10 Marks)

b) The following quantities were summarized from an experiment:
$n = 25, \sum y_i = 725.8, \sum x_{i1} = 206, \sum x_{i2} = 8284, \sum x_{i1}^2 = 2396,$
$\sum x_{i2}^2 = 3531848, \sum x_{i1}x_{i2} = 77177, \sum x_{i1}y_i = 8008.37, \sum x_{i2}y_i = 274811.31$

Compute the fitted regression line equation. (10 Marks)

**QUESTION FOUR (20 MARKS)**

Describe the extra sum of squares method for testing the contribution of an individual regressor variable to a multivariate linear regression model. (20 Marks)