MURANG'A UNIVERSITY OF TECHNOLOGY

SCHOOL OF PURE AND APPLIED SCIENCES

DEPARTMENT OF APPLIED SCIENCES

UNIVERSITY ORDINARY EXAMINATION

2017/2018 ACADEMIC YEAR

EXAMINATION FOR MASTER OF SCIENCE IN STATISTICS

AMS 602: THEORY OF LINEAR MODELS

DURATION: 3 HOURS

DATE: 22ND AUGUST, 2018

TIME: 9.00 A.M. – 12.00 NOON

Instructions to Candidates:

1. Answer **Any Four** questions.
2. Mobile phones are not allowed in the examination room.
3. You are not allowed to write on this examination question paper.

QUESTION ONE (25 Marks)

a) Distinguish between the following sets of terms

i. Random and fixed effects model (4 marks)

ii. Maximum likelihood estimation and Restricted maximum likelihood estimation (4 marks)

b) State and prove the Gauss – Markov theorem (5 marks)

c) A study was performed on wear of a bearing y and its relationship to x_1 = oil viscosity and x_2 = load. The following data were obtained:

y	x_1	x_2
1.93	1.6	8.5
2.30	1.8	8.2
1.72	2.2	10.6
0.91	4.3	12.1
1.13	3.3	13.3
1.25	4.2	11.2
2.25	2.3	9.6
1.45	3.4	10.5
1.94	3.1	11.7
1.52	3.8	9.3

i. Fit a multiple linear regression model to the data (5 marks)

ii. Test for significance of regression. Use α = 0.05. (4 marks)

iii. Compute t statistics for each model parameter. What conclusions can you draw? Use α = 0.05. (3 marks)

QUESTION TWO (25 Marks)

a) Describe the centred form of the multiple linear regression model (5 marks)

b) With the normality assumption, obtain maximum likelihood estimators of $\boldsymbol{\beta}$ and $\boldsymbol{\sigma^2}$ (5 marks)

c) Illustrate the properties of $\hat{\boldsymbol{\beta}}$ and $\hat{\boldsymbol{\sigma}}^2$ under the normal model (5 marks)

d) Explain how to perform regression model diagnostics (5 marks)

e) For the data below, perform a test of lack of fit on the simple linear regression model. Use α = 0.05. (5 marks)

y	1.5	8.9	3.9	7.2	7.1	6.9	1.9	1.5	7.8	6.5
x	3.1	3.2	2.8	2.7	2.5	3.2	3.3	3.5	2.8	2.9

QUESTION THREE (25 Marks)

a) Consider the model: $\boldsymbol{y} = \boldsymbol{X\beta} + \boldsymbol{\varepsilon}$, $\quad E(\boldsymbol{y}) = \boldsymbol{X\beta}$, $\quad cov(\boldsymbol{y}) = \boldsymbol{\Sigma} = \boldsymbol{\sigma^2 V}$ where $\boldsymbol{X}$ is full-rank and $\boldsymbol{V}$ is a known positive definite matrix, derive the estimators for $\boldsymbol{\beta}$ and $\boldsymbol{\sigma^2}$ (8 marks)

b) Given $\boldsymbol{X} = \begin{bmatrix} 1.0 & 1.0 & 1.3 \\ 1.0 & 1.1 & 2.8 \\ 1.0 & 1.2 & 3.3 \end{bmatrix}, \boldsymbol{y} = \begin{bmatrix} 4.5 \\ 6.8 \\ 2.4 \end{bmatrix}$ and $\boldsymbol{V} = \begin{bmatrix} 2 & -1 & 0 \\ -1 & 2 & -1 \\ 0 & -1 & 2 \end{bmatrix}$, compute $\widehat{\boldsymbol{\beta}}$ and $\text{cov}(\widehat{\boldsymbol{\beta}})$

(8 marks)

c) Describe in detail how to perform hypothesis testing in multiple regression (9 marks)

QUESTION FOUR (25 marks)

a) Explain the generalized linear model (2 marks)

b) Describe how to conduct multiple linear regression in R explaining the associated outputs

(10 marks

c) Given the following data, fit the second-order polynomial regression model

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_{11} x_1^2 + \beta_{12} x_1 x_2 + \beta_{22} x_2^2 + \varepsilon$$

(13 marks)

y	x_1	x_2
26	1.0	1.0
24	1.0	1.0
175	1.5	4.0
160	1.5	4.0
163	1.5	4.0
55	0.5	2.0
62	1.5	2.0
100	0.5	3.0
26	1.0	1.5
30	0.5	1.5
70	1.0	2.5
71	0.5	2.5

QUESTION FIVE (25 marks)

a) Show that if $\boldsymbol{y} \sim N_n(\boldsymbol{X\beta}, \sigma^2 \boldsymbol{I})$, then $\widehat{\boldsymbol{\beta}}$ and s^2 have minimum variance among all unbiased estimators. (7 marks)

b) Illustrate the general linear hypothesis $H_0\colon \boldsymbol{C\beta} = \boldsymbol{0}$, where $\boldsymbol{C}$ is a known coefficient matrix of rank $q \leq k + 1$ (8 marks)

c) Consider the model $\boldsymbol{y} = \boldsymbol{X\beta} + \sum_{i=1}^{m} \boldsymbol{Z_i a_i} + \boldsymbol{\varepsilon}$, where $\boldsymbol{X}$ is a known $n \times p$ matrix, the $\boldsymbol{Z_i}$'s are known $n \times r_i$ full rank matrices, $\boldsymbol{\beta}$ is $p \times 1$ a vector of unknown parameters, $\boldsymbol{\varepsilon}$ is an $n \times 1$ unknown random vector such that $E(\boldsymbol{\varepsilon}) = \boldsymbol{0}$ and $cov(\boldsymbol{\varepsilon}) = \sigma^2 \boldsymbol{I_n}$, and the $\boldsymbol{a_i}$'s are $r_i \times 1$ unknown random vectors such that $E(\boldsymbol{a_i}) = \boldsymbol{0}$ and $cov(\boldsymbol{a_i}) = \sigma_i^2 \boldsymbol{I}_{r_i}$. Furthermore, $cov(\boldsymbol{a_i}, \boldsymbol{a_j}) = \boldsymbol{0}$ for $i \neq j$, where $\boldsymbol{O}$ is $r_i \times r_j$, and $cov(\boldsymbol{a_i}, \boldsymbol{\varepsilon}) = \boldsymbol{0}$ for all i, where $\boldsymbol{O}$ is $r_i \times n$. Show that $E(\boldsymbol{y}) = \boldsymbol{X\beta}$ and $cov(\boldsymbol{y}) = \boldsymbol{\Sigma} = \sum_{i=1}^{m} \sigma_i^2 \boldsymbol{Z_i Z_i'} + \sigma^2 \boldsymbol{I_n}$ (10 marks)