10

(*University of Choice*)

**MASINDE MULIRO UNIVERSITY OF SCIENCE AND TECHNOLOGY (MMUST)**

**UNIVERSITY REGULAR EXAMINATIONS 2022/2023 ACADEMIC YEAR**

**FIRST YEAR SECOND SEMESTER REGULAR EXAMINATIONS FOR THE MASTERS OF SCIENCE (STATISTICS)**

**COURSE CODE:** **STA 806**

**COURSE TITLE:** **THEORY OF LINEAR MODELS**

**DATE: 27/04/2023** **TIME: 8 A.M – 11 A.M**

**INSTRUCTIONS TO CANDIDATES**

- Answer any THREE questions out of the five questions given

Time: 3 hours

## QUESTION ONE (20 MKS)

**a.** Show that if $E(y) = X\beta$ *and* $Cov(y) = \sigma^2 I$ the Least squares estimators $\beta_j$ , $j = 0,1,...,k$ have minimum variance among all linear unbiased estimators.

(6 marks)

**b.** For the data given below, obtain ANOVA table and test the hypothesis that;

$H_0 : \beta_1 = \beta_2 = 0$ *versus*

$H_1 : at \quad least \quad one \quad of \quad the \quad \beta j's \neq 0, i = 1,2$

*Use* $\alpha = 0.05$

| Price in thousands of dollars (y) | Age in years ($x_1$) | Square footage in thousands of square feet ($x_2$) |
|---|---|---|
| 65 | 9 | 10 |
| 70 | 9 | 8 |
| 80 | 8 | 8 |
| 90 | 9 | 8 |
| 100 | 9 | 7 |

(14 marks)

## QUESTION TWO (20 MKS)

a. An important practical feature of Generalized linear models is that they can all be fit to data using the same algorithm, a form of iteratively re-weighted least squares. Describe this algorithm.

(11 marks)

b. A study used logistic regression to determine characteristics associated with Y= whether a cancer patient achieved remission (1=yes). The most important explanatory variable was a labeling index (LI) that measures proliferative activity of cells after a patient receives an injection of tritiated thymidine. It represents the percentage of cells after a patient receives an injection of tritiated thymidine. It represents the percentage of cells that are "Labeled". Table (1) shows the data while Table (2) presents R glm results for $P(Y = 1/X = x)$.

Table (1)

| LI | 8 | 8 | 18 | 28 | 10 | 10 | 20 | 20 | 20 | 32 | 12 | 12 | 12 | 22 | 22 | 34 | 14 | 14 | 14 | 24 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Remission | 0 | 0 | 1 | 1 | 0 | 0 | 1 | 1 | 1 | 0 | 0 | 0 | 0 | 1 | 1 | 1 | 0 | 0 | 0 | 0 |

Table (2)

| | Estimate | Std. Error | z-value | Pr(>\|z\|) |
|---|---|---|---|---|
| (Intercept) | -4.2266 | 1.8551 | -2.28 | 0.0227 |
| LI | 0.2111 | 0.0975 | 2.16 | 0.0304 |

i. Write down the fitted model (2 marks)

ii. Is LI significant in explaining remission at 5% level of significance? (2 marks)

iii. Interpret the effect of LI on the odds of remission. (2 marks)

iv. Determine $\pi = P(Y = 1/X = 26)$ (3 marks)

**QUESTION THREE (20 MKS)**

a. A study is conducted to estimate the demand for housing based on current interest rate and the rate of unemployment. The data is given as follows;

| Units sold | 65 | 59 | 80 | 90 | 100 | 105 |
|---|---|---|---|---|---|---|
| Interest rate in (%) | 9.0 | 9.3 | 8.9 | 9.1 | 9.0 | 8.7 |
| Unemployment rate at (%) | 10 | 8 | 8.2 | 7.7 | 7.1 | 7.2 |

1.) Fit multiple regression model of the form $y = \beta_0 + \beta_1 x_1 + \beta_2 x_2$ to the above data. (7 marks)

2.) Test for $\beta_0$ , $\beta_1$ *and* $\beta_2$ (7 marks)

b. State and explain the components of a generalized linear model (6 marks)

## QUESTION FOUR (20 MKS)

a. Define Ridge estimator. Hence, show that in the orthonormal case, a ridge estimator is proportional to the least square estimator. (7 marks)

b. Determine the expectation and variance of a ridge estimator for $\beta$ hence state under what condition the ridge estimator is unbiased for $\beta$ (13 marks)

## QUESTION FIVE (20 MKS)

a. What is an exponential family? (3 marks)

b. Show that $f(y_i) = \frac{e^{-\lambda i}\lambda_i{}^{yi}}{y_i!}$ belongs to the exponential family. (6 marks)

c. Using the data below, fit the non linear regression model and predict the number of bacteria when dose is 16.

| Dose | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 |
|---|---|---|---|---|---|---|---|---|
| Bacteria | 35500 | 21100 | 19700 | 16600 | 14200 | 10600 | 10400 | 6000 |

| Dose | 9 | 10 | 11 | 12 | 13 | 14 | 15 |
|---|---|---|---|---|---|---|---|
| Bacteria | 5600 | 3800 | 3600 | 3200 | 2100 | 1900 | 1500 |

(11 marks)