MERU UNIVERSITY OF SCIENCE AND TECHNOLOGY

P.O. Box 972-60200 – Meru-Kenya.
Tel: +254(0) 799 529 958, +254(0) 799 529 959, +254 (0)712 524 293
Website: www.must.ac.ke Email: info@mucst.ac.ke

UNIVERSITY EXAMINATIONS 2022/2023

FOURTH YEAR, SECOND SEMESTER SPECIAL SUPPLEMENTARY EXAMINATION FOR DEGREE OF BACHELOR OF SCIENCE IN STATISTICS

SMS 3457: REGRESSION MODELING II

DATE: JANUARY 2023 **TIME: 2 HOURS**

INSTRUCTIONS: Answer Question ONE and any other TWO questions.

QUESTION ONE (30 MARKS)

a) Explain the following as used in modelling; (6 Marks)

i. Model

ii. Factor analysis

iii. Normality test of a model

b) Model the properties of kernels as used in regression modelling. (4 Marks)

c) Given that $Y_i = a + bX_i + \in$ such that $\in \sim NIID(0, \sigma^2)$. Find $E(Y_i/X_i)$. (5 Marks)

d) Giving relevant examples, explain discriminant analysis as used in research. (5 Marks)

e) Given the model $Y_i = \beta_0 + \beta_1 X_i + \in_{i;}\ i = 1,2,\ldots,n$, show that $\hat{\beta}_1 = \dfrac{\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})}{\sum_{i=1}^{n}(x_i - \bar{x})^2}$

(6 Marks)

f) Describe a k-nearest neighbor as used in simple non-parametric regression. (4 Marks)

QUESTION TWO (20 MARKS)

a) Define a simple linear regression model. (2 Marks)

b) Briefly discuss heteroscedasticity as used in regression modelling. (4 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 the variables are related according to the simple linear regression model.

i. Calculate the least squares estimates for the slope and the intercept. (6 Marks)

ii. Use the equation of the fitted line to predict what root length would be observed when the moisture content is $x = 4.3$. (2 Marks)

iii. Given a point estimate of the mean root length when moisture content is $x = 3.7$. (2 Marks)

iv. Suppose that the observed value of moisture content at $x = 3.7\ is\ y = 46.1$. Calculate the value of the corresponding residual. (4 Marks)

QUESTION THREE (20 MARKS)

a) Given the model $Y_i = \beta_0 + \beta_1 x_i + \varepsilon_i$ with $\varepsilon_i \sim NIID(0, \sigma^2)$ show that $E\left(\hat{\beta}_0\right) = \beta_0$. (4 Marks)

b) Giving relevant examples, discern between non-linear and linear regression models. (3 Marks)

c) Given that $g(X, \alpha) = \alpha_0 \alpha_1 exp(\alpha_2 X)]$ is intrinsically linear, write it in the form $g(X, \alpha) = a_0 + b_1 X + \varepsilon$. (4 Marks)

d) A hospital administrator wished to develop a regression model for predicting the degree of long term recovery after discharge from the hospital for severly injured patients. The predictor variable to be utilized is number of days of hospitalization (X), and the response variable is a prognostic index for long term recovery (Y), with large values of the index reflecting a good prognosis. It was decided to investigate the appropriateness of the two-parameter nonlinear exponential model $Y_i = \gamma_0 exp(\gamma_1 X_i) + \in_i$. Determine the estimates of γ_0 and γ_1. (9 Marks)

QUESTION FOUR (20 MARKS)

a) Explain the hurdles to non-parametric regression. (4 Marks)

b) Given the data

X	20	30	40	15	25	28
Y	45.6	35.3	40.3	20.0	44.5	43.2

Using a rectangular kernel function, find the k-nearest neighbor estimate of the density $P(X)$ and regression function $m(X)$ where x is 29 and $k = 3$. (6 Marks)

c) Given that the training set consist of the following input vectors $X^{(1)} = \begin{pmatrix}1\\1\\0\end{pmatrix}, X^{(2)} = \begin{pmatrix}1\\0\\1\end{pmatrix}, X^{(3)} = \begin{pmatrix}1\\1\\1\end{pmatrix}, X^{(4)} = \begin{pmatrix}1\\0\\0\end{pmatrix}$ with correct classification $Z^{(1)} = Z^{(2)} = Z^{(3)} = 1, z^{(4)} = 0$. The perceptron with weights w_0, w_1, w_2 classifies an object as 1 if $w_0x_0 + w_1X_1 + w_2X_2 > 0$ and as 0 otherwise. Taking $w = (0\ 0\ t)^T$ as the initial weight and $\eta = 1$, train the perception and find the weight that achievers correct classification for all input vectors in the training set. (10 Marks)

QUESTION FIVE (20 MARKS)

a) Explain the equation $X = QF + \mu$ where X is data matrix as used in factor analysis.

(3 Marks)

b) Let Y be a random variable with the distributions $\pi_1: g_1(x) = p(x = 0) = p(x = 1) = \frac{1}{2}$ and $\pi_2: g_2(x) = p(x = 0) = \frac{1}{3}; p(x = 1) = \frac{2}{3}$, construct the decision making criterion.

(6 Marks)

c) Let $X \sim N(\mu, \Sigma)$ be tri-variate normal random vector with $\Sigma = \begin{bmatrix} 1 & 0 & 0.6 \\ 0 & 4 & 0 \\ 0.6 & 0 & 9 \end{bmatrix}$ find;

i. The eigenvalues and the eigenvector corresponding to the first eigenvalue.

(6 Marks)

ii. The first principal component. (5 Marks)