Strathmore
UNIVERSITY

STRATHMORE INSTITUTE OF MATHEMATICAL SCIENCES
MASTER OF SCIENCE IN DATA SCIENCE AND ANALYTICS
END OF SEMESTER EXAMINATION
STA 8103: LINEAR MODELS

DATE: 17th July 2022 Time: 2.5 Hours

ANSWER QUESTION ONE AND ANY OTHER TWO QUESTIONS.

Question One.

a. Outline the following approaches to model selection.
    a. AIC. (2 marks)
    b. LRT. (2 marks)
    c. BIC. (2 marks)
b. Discuss any two regularization techniques in regression analysis outlining their disadvantages. (6 marks)
c. Ordinary least squares technique is one of the parameter estimation methods for linear models. Derive the least squares estimates for the regression parameter β_1 (6 marks)
d. Describe two approaches you would use to determine whether an interaction effect between two variables exit (4 marks)
e. Defining the Generalized linear model, outline its three components (4 marks).
f. Outline the properties of the hat matrix H (4 marks)

Question Two.

The following is data on percentage of taxes and revenue collection.

Percentage	20	25	33	40
Revenue (in billion $)	120	116	98	60

a. Determine R^2 and explain it (6 marks)
b. Test at 5% level of significance the relevant hypothesis for $\hat{\beta}_1$ (9 marks)

Question Three

a. A start-up company is conducting an aggressive marketing campaign on YouTube, Facebook, and newspaper. Their data scientist built a model for estimating sales based on the advertising budget invested on the three platforms and the results shown below.

```
model <- lm(sales ~ youtube + facebook + newspaper, data = marketing)
summary(model)
```

```
##
## Call:
## lm(formula = sales ~ youtube + facebook + newspaper, data = marketing)
##
## Residuals:
##    Min     1Q Median     3Q    Max
## -10.59  -1.07   0.29   1.43   3.40
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)  3.52667    0.37429    9.42   <2e-16 ***
## youtube      0.04576    0.00139   32.81   <2e-16 ***
## facebook     0.18853    0.00861   21.89   <2e-16 ***
## newspaper   -0.00104    0.00587   -0.18     0.86
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 2.02 on 196 degrees of freedom
## Multiple R-squared:  0.897,  Adjusted R-squared:  0.896
## F-statistic:  570 on 3 and 196 DF,  p-value: <2e-16
```

Provide a concise description of the results (8 marks).

b. Non-linear least squares regression extends linear least squares regression for use with much larger and more general class of functions. Outline the non-linear least squares optimization using the Gauss-Newton algorithm (7 marks)

Question Four

a. Discuss any 4 departures from the linear model assumption, how to determine them giving possible remedial measures (8 marks).
b. Local polynomial model is one of the ways of relaxing the non-linearity assumptions. Discuss the local polynomial regression for estimating a regression function at point x_0. (7 marks)

Question 5

Consider the following one-Way ANOVA with 2 groups with three observations each.

$$\begin{pmatrix} Y_{11} \\ \vdots \\ Y_{1n_1} \\ Y_{21} \\ \vdots \\ Y_{2n_2} \end{pmatrix} = \begin{pmatrix} 1 & 1 & 0 \\ \vdots & \vdots & \vdots \\ 1 & 1 & 0 \\ 1 & 0 & 1 \\ \vdots & \vdots & \vdots \\ 1 & 0 & 1 \end{pmatrix} \begin{pmatrix} \mu \\ \alpha_1 \\ \alpha_2 \end{pmatrix} + \begin{pmatrix} \varepsilon_{11} \\ \vdots \\ \varepsilon_{1n_1} \\ \varepsilon_{21} \\ \vdots \\ \varepsilon_{2n_2} \end{pmatrix}$$

The rank of X is 2. Reducing the model to one of full rank, obtain the estimates for α
(15 marks)