



Strathmore  
UNIVERSITY

**STRATHMORE INSTITUTE OF MATHEMATICAL SCIENCES**  
**MASTER OF SCIENCE IN STATISTICAL SCIENCES**  
**END OF SEMESTER EXAMINATION**  
**STA 8103: LINEAR MODELS**

Date: 4<sup>th</sup> December 2023

Time: 2.5 Hours

**INSTRUCTIONS: ANSWER QUESTION ONE AND ANY OTHER TWO QUESTIONS**

**Question 1 [30 Marks]**

- a. Briefly outline the following variable selection techniques for linear regression model.
  - a. AIC (2 marks)
  - b. BIC (2 marks)
  - c. LRT (2 marks)
  - d. F-Test (2 marks)
- b. Outline any two approaches you would use to determine if interaction exists between two variables (4 marks).
- c. Penalized regression is one way of dealing with too many variables in the model. This is achieved by employing a penalty term  $\lambda$ . Discuss any 3 shrinkage methods (6 marks)
- d. Consider the following linear regression output.

```
Call:
lm(formula = height ~ age, data = ageandheight)

Residuals:
    Min       1Q   Median       3Q      Max
-0.27238 -0.24248 -0.02762  0.16014  0.47238

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  64.9283    0.5084  127.71 < 2e-16 ***
age           0.6350    0.0214   29.66 4.43e-11 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.256 on 10 degrees of freedom
Multiple R-squared:  0.9888,    Adjusted R-squared:  0.9876
F-statistic:  880 on 1 and 10 DF,  p-value: 4.428e-11
```

- Exhaustively discuss the regression model (6 marks)
- e. The following is a regression function:

$$y = f(.) + \epsilon_i$$

Discuss any three ways you can estimate the unknown regression function  $f(.)$  (6 marks)

**Question 2 [15 Marks]**

- a. Non-linear models require specialized algorithms to estimate model parameters. These algorithms require good initial estimates. Outline any 5 methods you could use to determine good initial estimates (5 marks)
- b. Consider the following non-linear model:

$$y = \frac{\hat{\beta}_0 + \hat{\beta}_1 x}{1 + \hat{\beta}_2 x}$$

Discuss the Gauss-Newton algorithm for estimating  $\hat{\beta}_0$  and  $\hat{\beta}_1$  (10 marks)

**Question 3 [15 Marks]**

Consider one way ANOVA with 2 groups  $n_1$  and  $n_2$  such that  $n_1 + n_2 = n$

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

Obtain the solution  $\hat{\beta}$  to the normal equations using the generalized inverse approach (15 marks).

**Question 4 [15 Marks]**

The following is data on sales vs advertisement expenditure.

Sales (100k)	Advertisement (100k)
50	25
67	45
23	10
49	20

- a. Obtain the R-Squared and the adjusted R-squared for the model and interpret your results (7 marks)
- b. Test at 5% level of significance whether  $\beta_0$  and  $\beta_1$  are significant. (6 marks)
- c. Obtain the 95% confidence interval for  $\beta_0$  and  $\beta_1$  and comment on your result. (2 marks)

**Question 5 [15 Marks]**

a. Discuss the following statistical approaches:

- a. Penalized Splines approximation (3 marks)
- b. LOWESS (4 marks)

b. Residuals are useful in model diagnostics. Discuss any 4 model diagnostics that you could carry out using residuals and suggest any remedies should there be a violation of regression assumptions (8 marks)