BINDURA UNIVERSITY OF SCIENCE EDUCATION

BSc Statistics and Financial Mathematics

SFM 421: General Linear Models

Time: 3 hours

SECTION A (40 Marks)

Candidates may attempt ALL questions being careful to number them Q1 to Q5

1. (a) What are the advantages of general linear models over traditional (OLS) regression. [3]

   (b) State and explain three components of general linear models. [6]

2. Let $Y = X\beta + e$ where X is an $n \times p$ matrix of full rank, β is a $p \times 1$ vector of unknown parameters and e is an $n \times 1$ normally distributed random vector with mean 0 and variance $\sigma^2 I$

   (a) Derive the maximum likelihood estimator of β and σ^2. [10]

   (b) Show that $\frac{SS_{res}}{n}$ is a biased estimator of σ^2. [10]

3. Explain what you undestand by 80:20 training testing ration. [2]

4. Let $X_1, X_2, ...X_n$ denote a random sample from Poisson distribution with parameter $\lambda > 0$. Find the sufficient estrimator for λ. [4]

5. Prove that $E(Y^{'}AY) = tr(AV) + \mu^{'}A\mu$. [5]

SECTION B (60 Marks)

Candidates may attempt **TWO** questions being careful to number them Q6 to Q8

6. A data processing system uses three types of structural elements:files, flows and processes. Files are permanent records,flows are data interfaces, and processes are logical manipulations of the data. The costs of developing software for the system are based on the number of these three elements. A study is conducted with the following results:

Cost(Y)	Files(X_1)	Flows (X_2)	Processes (X_3)
22.6	4	44	18
15	2	33	15
78.1	20	80	80
28	6	24	21
80.5	6	227	50
24.5	3	20	18
20.5	4	41	13

(a) Find $\hat{\beta}$ [10]

(b) Construct the ANOVA table using the uncorrected sum total. [10]

(c) Test if the model is adequate. [5]

(d) Calculate $R(\beta_1/\beta_0)$. [5]

7. (a) It is known that advertising (x) influences sales (Y). Data on X and Y is presented in the table below.

X	15	25	19	32	33	41	54	49	52	21
Y	12	15.1	13	18	17	22	23	25	22.1	14

i. Find the 90% joint confidence interval for β assuming $\beta_0$and β_1 are significant. [10]

ii. Find the 90% confidence interval for the mean sales with advertising expenditure of 60. [5]

(b) Let $Y = X\beta + e$ where X is an $n \times p$ matrix of full rank, β is a $p \times 1$ vector of unknown parameters and e is an $n \times 1$ normally distributed random vector with mean 0 and variance $\sigma^2 I$. Show that $t'\hat{\beta}$ is BLUE for $t'\beta$ [15]

8. Let $Y = XB + e$ where X is an $n \times p$ matrix of full rank, β is a $p \times 1$ vector of unknown parameters and e is an $n \times 1$ normally distributed random vector with mean 0 and variance $\sigma^2 I$. Assume that $t'\beta$ is estimable where t is a $1 \times p$ non zero vector of real numbers and that $\hat{\beta}$ denotes any solutions to the normal equations.

(a) Show that $E(t'\hat{\beta}) = t'\beta$. [6]

(b) Prove that $Var(t'\hat{\beta}) = t'(X'X)^{-1}t\sigma^2$ [6]

(c) Deduce that the random variable $\frac{t'\hat{\beta}-t'\beta}{s\sqrt{t'(X'X)^{-1}t}}$ follows student t distribution with $n-p$ degrees of freedom. [10]

(d) Deduce that the random variable $\frac{t'\hat{\beta}-t'\beta}{\sigma\sqrt{t'(X'X)^{-1}t}}$ follows standard normal distribution. [8]

P