NATIONAL UNIVERSITY OF SCIENCE AND TECHNOLOGY

FACULTY OF APPLIED SCIENCE

DEPARTMENT OF STATISTICS AND OPERATIONS RESEARCH

B.Sc. APPLIED MATHEMATICS PART IV

SORS4106 – EXPERIMENTAL DESIGN AND MULTIPLE REGRESSION

DECEMBER 2024 EXAMINATION

Time: 3 Hours
Total Marks: 100

Candidates may attempt **ALL** questions in Section **A** and at most **THREE** questions in Section **B**. Show all working where necessary. You may use a calculator. Graph paper and Statistical Tables will be provided.

SECTION A: Attempt ALL questions in this section [40 marks].

Two plant pathologists, Pathologist A and Pathologist B, in a Tobacco Research Institute were asked to each independently design an experiment to determine if there is a significant difference in the mean number of lesions from two preparations of mosaic virus.

A1. *Pathologist A's design:*
A single leaf is taken from 11 different tobacco plants. Each leaf is divided in half, and each half is given one of two preparations of mosaic virus, Prep1 or Prep2, and the results inTable1are obtained.

Table1

Plant	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**11**
Prep1	18	20	9	14	38	26	15	10	25	7	13
Prep2	14	15	6	12	32	30	9	2	18	3	6

(i) Carry out an appropriate t-test at the 5% level of significance to determine if there is a significant difference in the mean number of lesions from the two preparations of mosaic virus. **[8]**

(ii) Carry out an appropriate Analysis of Variance to drtermine if there is a significant difference in the mean number of lesions from the two preparations of mosaic virus. Use the 5% level of significance. **[8]**

(iii) Indicate the similarities between tests (i) and (ii). [4]

A2. *Pathologist B's design:*
A single leaf is taken from 22 different tobacco plants. Eleven of the leaves are selected at random and treated with Prep1 and the remaining 11 are treated with Prep2. The results in Table 2 are obtained.

Table 2

Prep1	7	26	9	14	13	25	15	10	38	18	20
Prep2	6	15	3	12	32	30	9	2	18	6	14

(i) Carry out an appropriate t-test at the 5% level of significance to determine if there is a significant difference in the mean number of lesions from the two preparations of mosaic virus. **[8]**

(ii) Carry out an appropriate Analysis of Variance to determine if there is a significant difference in the mean number of lesions from the two preparations of mosaic virus. Use the 5% level of significance. **[8]**

(iii) Indicate the similarities between tests (i) and (ii). **[4]**

SECTION B: Attempt any **THREE** questions from this section **[60 marks]**.

B3. Presented in Table 3 are data on age (x) and blood plasma level of a polyamine (y) for 14 healthy children where age 0 refers to newborn.

Table 3

x	0	0	1	1	2	2	3	3	4	4	5	5	6	6
y	17.0	11.2	9.2	12.6	7.4	10.5	8.3	5.8	4.6	6.5	5.3	3.8	3.2	4.5

(a) Fit the straight line regression model, $y = \beta_0 + \beta_1 x + \varepsilon$, to the data and test for the significance of the slope by the Analysis of Variance. **[6]**

(b) Compute the fitted values and residuals. Plot the residuals against the fitted values and comment on the adequacy of the straight line model based on your plot. **[6]**

(c) Test for the lack of fit of the straight line model. **[8]**

B4. A horticulturist conducted an experiment to compare two varieties of roses; variety **WR** and variety **RR** on yield of marketable roses (y). The plots were not of the same size, hence the horticulturist wished to use plot size (x) as the concomitant variable. Six replicates were made for each treatment. The results in Table 4 were obtained.

Table 4

Variety WR		Variety RR	
y	x	y	x
98	15	55	4
60	4	60	5
77	7	75	8
80	9	65	7
95	14	87	13
64	5	78	11

(a) Ignore the variable (x) and carry out a One-Way Analysis of Variance to determine whether there is a significant difference in mean marketable yield of the two varieties at the 5% level of significance. **[6]**

(b) Considering the variable (x), perform the appropriate analysis to test whether there is a significant difference in mean marketable yield of the two varieties. Use the 5% level of significance. Compare your conclusion here with that in (a) and explain the difference if any. **[14]**

B5. The yield of a chemical process is related to the concentration (in ppm) of a reactant, (x_1) and the operating temperature (in degrees C), (x_2). The data in Table 5 were obtained in one such experiment.

Table 5

Chemical Yield (y)	**Concentration (x_1)**	Temperature (x_2)
81	1.5	110
89	1.5	140
83	3.0	110
91	3.0	140
79	1.5	110
87	1.5	140
84	3.0	110
90	3.0	140

(a) Fit a multiple linear regression model $y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \varepsilon$ to the data. **[6]**

(b) Partition the regression sum of squares into single-degree-of-freedom components attributable to x_1 and x_2 and hence test the significance of each regressor variable by the analysis of variance. **[10]**

(c) State the **X** matrix. Find $\mathbf{X}^T\mathbf{X}$ and $(\mathbf{X}^T\mathbf{X})^{-1}$ and hence compute the standard errors of your estimates of regression coefficients in (a). **[4]**

B6. The data for a study of the relation of body fat (y) to triceps skin fold thickness (x_1) and thigh circumference (x_2), based on a sample of 20 healthy females 25-34 years old were collected. Body fat (y) was regressed on (x_1), (x_2), and (x_1,x_2) respectively and the regression models and their analysis of variance tables, Table 6a, Table 6b, and Table 6c were obtained.

Table 6a $\hat{y} = -1.496 + 0.8572x_1$

Source	*df*	*SS*	*MS*	*F*
Regression			352.27	44.31
Residual		143.12	7.95	
Total	19	495.39		

Table 6b $\hat{y} = -23.634 + 0.8566x_2$

Source	*df*	*SS*	*MS*	*F*
Regression		381.97		60.63
Residual			6.30	
Total	19	495.39		

Table 6c $\hat{y} = -19.174 + 0.2224x_1 + 0.6594x_2$

Source	*df*	*SS*	*MS*	*F*
Regression		385.44	192.72	29.79
Residual			6.47	
Total	19	495.39		

(i) Copy and fill in the missing values in the three ANOVA tables, Table 6a, Table 6b, and Table 6c. [3]

Use the following to select the 'best' model:

(ii) The C_p and S_p Statistics, [9]

(iii) Forward Selection, [5]

(iv) Backward Elimination. [3]

B7. A bacteriologist is interested in the effect of two different culture mediums and two different times on the growth of a particular virus. He performs six replicates of a 2^2 factorial experiment, making the runs in random order. The results in Table 7 are obtained.

Table 7

	Culture Medium (Factor B)			
Time (Factor A)	**1**		**2**	
20 hr	21 23 20	22 28 26	25 24 29	26 25 27
30 hr	37 38 35	39 38 36	31 29 30	34 33 35

(a) Make a table of signs of contrasts showing treatment combinations in standard order and their totals. **[8]**

(b) Compute the sums of squares for both main effects and the two factor interaction by the contrast method. **[6]**

(c) Construct an ANOVA table and test for the significance of main effects and interaction. **[6]**

END OF QUESTION PAPER