NATIONAL UNIVERSITY OF SCIENCE AND TECHNOLOGY

DEPARTMENT OF APPLIED MATHEMATICS

BSc PART IV HONOURS EXAMINATIONS 2002

SMA 4105 - STATISTICS IV

NOV/DEC 2002

TIME 3 HOURS
Total = 100 Marks

This paper has 6 pages

Answer **FOUR** questions: Question 1 in **SECTION A** (28 Marks) and **THREE** from **SECTION B** (24 Marks Each). Where a question contains subdivisions, the mark value of each subdivision is indicated in brackets.

Candidates are expected to spend not more than one hour on Question 1. Calculators may be used. Statistical Tables and graph paper are provided, however, Statistical Tables should not be marked or taken out of the examination room. GOOD LUCK!

SECTION A (Compulsory)

1. (a) A manufacturer of laundry detergent was interested in testing a new product prior to market release. One area of concern was the relationship between the height of the detergent suds in a washing machine as a function of the amount of detergent added in the wash cycle. For a standard size washing machine tub filled to the full level, random assignments of amounts of detergent were made and tested on the washing machine and the following data obtained:

Amount (x)	Height (y)	
6	28.1	27.6
7	32.3	33.2
8	34.8	35.0
9	38.2	39.4
10	43.5	46.8

(i) Draw a scatterplot of the data on graph paper. Fit a simple linear regression model to the data and draw your fitted line on your scatterplot. **[4 marks]**

(ii) Test for the significance of the slope through the analysis of variance approach. **[4 marks]**

(iii) Conduct a test for lack of fit of the simple linear regression model. **[6 marks]**

(b) Assuming the following multiple linear regression model $y_i = \beta_1 x_{1i} + \beta_2 x_{2i} + \beta_3 x_{3i} + e_i$, for i = 1, 2, . . . , 53 with $Var(y_i) = \sigma^2$ and $Cov(y_i, y_j) = 0$, for $i \neq j$, an experiment was conducted and the following results were obtained from the data:

$$\mathbf{X}^T\mathbf{Y} = \begin{bmatrix} 200 \\ 49 \\ 10 \end{bmatrix}, \quad \mathbf{Y}^T\mathbf{Y} = 504, \quad \mathbf{X}^T\mathbf{X} = \begin{bmatrix} 100 & 0 & 0 \\ 0 & 49 & 0 \\ 0 & 0 & 20 \end{bmatrix}$$

(i) Find the vector of regression coefficients, $\hat{\boldsymbol{\beta}} = (\hat{\beta}_1 \ \hat{\beta}_2 \ \hat{\beta}_3)^T$. **[5 marks]**

(ii) Find the estimate of σ^2 and hence construct the variance - covariance matrix for the vector $\hat{\boldsymbol{\beta}}$. **(5 marks)**

(iii) Find a 95% confidence interval for β_2 **(4 marks)**

SECTION B (Answer Any Three Questions)

2.. The yield of a chemical process (y) is related to the concentration of the reactant (x_1), and the operating temperature (x_2). The following data were obtained in one such experiment:

Yield (y)	Concerntration (x_1)	Temperature (x_2)
81	1.00	150
89	1.00	180
83	2.00	150
91	2.00	180
79	1.00	150
87	1.00	180
84	2.00	150
90	2.00	180

(a) Fit a multiple linear regression model to the data. **[10 marks]**

(b) Partition the regression sum of squares into two single-degree-of-freedom components attributable to x_1 and x_2. Construct an ANOVA table to test for the significance of each regressor. Draw your conclusion. **[14 marks]**

3. An engineer is interested in the effect of cutting speed (A), tool geometry (B), and cutting angle (C) on the life of a machine tool. Two levels of each factor are chosen and three replicates of a 2^3 factorial experiment are run. The following results are obtained:

Treatment Combination	Replicate		
	I	II	III
(1)	22	31	25
a	32	43	29
b	35	34	50
c	44	45	38
ab	55	47	46
ac	40	37	36
bc	60	50	54
abc	39	41	47

(a) Compute the sums of squares of all main effects and interactions by the contrast method. **[16 marks]**

(b) Construct an analysis of variance (ANOVA) table and test for the significance of all main effects and interactions. Use the 5% level of significance. **(8 marks)**

4. Three diets, A, B, and C, were compared for their effect on blood cholesterol level in females. Age (X) was used as a control factor because of its apparent association with cholesterol level. Following a specified time, the blood cholesterol level (Y) (in mg/100ml) was determined. Thirty subjects were used, ten for each diet and the following data obtained:

Subject Number	**Diet A**		**Diet B**		**Diet C**	
	X_A	Y_A	X_B	Y_B	X_C	Y_C
1	40	190	41	201	41	201
2	47	205	30	187	32	192
3	28	178	58	226	57	215
4	51	215	48	222	49	202
5	50	202	57	220	36	197
6	48	208	26	184	30	195
7	27	181	50	208	34	194
8	47	201	44	199	42	200
9	42	192	39	196	61	248
10	38	193	29	185	40	198
Total	**418**	**1965**	**422**	**2028**	**422**	**2042**

$\sum_{i=1}^{30} X_i = 1262$ $\sum_{i=1}^{30} Y_i = 6035$ $\sum_{i=1}^{30} X_i Y_i = 257628$

$\sum_{i=1}^{30} X_i^2 = 55908$ $\sum_{i=1}^{30} Y_i^2 = 1220341$

(a) Ignore the control variable (X) and test for significant differences among the three diets by the analysis of variance approach. Use the 5% level of significance. **(8 marks)**

(b) Using the control variable (X), perform the appropriate analysis to test whether there are significant differences among the three diets. Use the 5% level of level of significance. **(12 marks)**

(c) Compare and comment on the results you obtained in (a) with those you obtained in (b). **(4 marks)**

5. The heat evolved in joules per gram of cement (y) is thought to be a function of the amount of each of four ingredients in the mix; triculcium aluminate (x_1), tricalcium silicate (x_2), tetracalcium alumino ferrite (x_3), and dicalcium silicate (x_4). Thirteen observations were obtained in the investigation of this functional relationship; that is, n = 13. All possible linear regression models which included a constant term were examined and their residual sums of squares (RSS) are as follows:

Regressors in model	p	RSS	Cp	Sp
Constant	1	2715.7635		
x_1	2	1265.6867		
x_2	2	906.3363		
x_3	2	1939.4005		
x_4	2	883.8669		
x_1 x_2	3	57.9045		
x_1 x_3	3	1227.0721		
x_1 x_4	3	74.7621		
x_2 x_3	3	415.4427		
x_2 x_4	3	868.8801		
x_3 x_4	3	175.7380		
x_1 x_2 x_3	4	48.1106		
x_1 x_2 x_4	4	47.9727		
x_1 x_3 x_4	4	50.8361		
x_2 x_3 x_4	4	73.8145		
x_1 x_2 x_3 x_4	5	47.8636		

Complete the table and use the following methods to select the 'best' linear regression model:

(a) The Cp and Sp statistics, **[8 marks]**

(b) Forward Selection, **[8 marks]**

(c) Backward Elimination. **[8 marks]**

6. Discuss clearly what is meant by each of the following ,giving an appropriate example in each case:

(a) Complete confounding, **[8 marks]**

(b) Partial confounding, **[8 marks]**

(c) Aliasing. **[8 marks]**

***** END OF QUESTION PAPER *****