NATIONAL UNIVERSITY OF SCIENCE AND TECHNOLOGY

DEPARTMENT OF APPLIED MATHEMATICS

B.Sc. PART IV HONOURS EXAMINATIONS 2001

SMA 4105 – STATISTICS IV

NOVEMBER/ DECEMBER 2001

3 HOURS (100 Marks) This paper has 4 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) The amounts of a chemical compound y, which were dissolved in 100gm of water at various temperatures, x, were recorded as follows:

x (^{0}C)	y (grams)		
0	8	6	8
15	12	10	14
30	25	21	24
45	31	33	28
60	44	39	42
75	48	51	44

Test for the significance of the slope and hence carry out the lack-of-fit test. **(18 marks)**

(b) In multiple linear regression the least squares estimator of the vector β of the regression coefficients is given by

$$\hat{\beta} = (X^TX)^{-1} X^Ty$$

(i) Show that $\hat{\beta}$ is an unbiased estimator of β. **(5 marks)**

(ii) Show that Var $(\hat{\beta}) = \sigma^2 (X^TX)^{-1}$. **(5 marks)**

SECTION B (ANSWER ANY THREE QUESTIONS)

2. A small experiment was conducted to fit a multiple linear regression equation relating the yield (y) to temperature (x_1), reaction time (x_2), and concentration of one of the reactants (x_3). Two levels of each regressor variable were chosen and measurements corresponding to the coded independent variables were recorded as follows:

y	x_1	x_2	x_3
7.6	-1	-1	-1
8.4	1	-1	-1
9.2	-1	1	-1
10.3	-1	-1	1
9.8	1	1	1
11.1	1	-1	-1
10.2	-1	1	1
12.6	1	1	1

(a) Using the coded regressor variables, estimate (fit) the multiple linear regression equation.

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \varepsilon.$$

(10 marks)

(b) Partition the regression sum of squares into three single-degree-of-freedom components attributable to x_1, x_2, and x_3, respectively. Construct an analysis of variance table indicating significance tests on each regressor variable. Draw your conclusion.

(14 marks)

3. A bacteriologist is interested in the effect of two different culture mediums and two different times on the growth of a particular virus. She performs six replicates of a 2^2 factorial experiment, making the runs in random order and obtains the following data:

Time	Culture Medium 1		Culture Medium 2	
12 hr	21	22	25	26
	23	28	24	25
	20	26	29	27
18 hr	37	39	31	34
	38	38	29	33
	35	36	30	35

(a) Ignoring any differences among the six replicates, compute the sum of squares of the main effects and interaction by the contrast method. **(16 marks)**

(b) Construct an analysis of variance table and test for the significance of main effects and interaction. **(8 marks)**

4. An engineer is studying the effect of cutting speed on the rate of metal removal in a machining operation. However, the rate of metal removal is also related to the hardness of the test specimen. Five observations are taken at each cutting speed. The amount of metal removed (y) and the hardness of the specimen (x) are shown in the following table.

CUTTING SPEED (rpm)					
1000 (A)		1200 (B)		1400 (C)	
y	x	y	x	y	x
68	120	112	165	118	175
90	140	94	140	82	132
98	150	65	120	73	124
77	125	74	125	92	141
88	136	85	133	80	130
421	**671**	**430**	**683**	**445**	**702**

$\Sigma y = 1296$ $\Sigma x = 2056$ $\Sigma xy = 180946$
$\Sigma y^2 = 115148$ $\Sigma x^2 = 285366$

Note: The absence of subscripts indicates the sum of all x's, y's, and xy products.

(a) Ignore the covariate (x) and test for the significant differences in the amount of metal removed (y) at three cutting speeds by the analysis of variance (ANOVA). **(8 marks)**

(b) Consider the covariate (x) and perform the analysis of covariance (ANCOVA) on the data. **(12 marks)**

(c) Compare the conclusion drawn in (a) with that drawn in (b) and comment. **(4 marks)**

5. In an effort to try to predict pollution levels from prevailing meteorological conditions, hourly recordings of levels of a pollutant and various meteorological conditions were made. These were; maximum levels of an oxidant, (a photochemical pollutant), (y), and the morning averages of four meteorological variables: wind speed (x_1), temperature (x_2), humidity (x_3), and insolation (x_4). A total of thirty (30) observations of (y) were recorded at various levels of x_1, x_2, x_3 and x_4.

All possible linear regression models which included a constant term were examined and their residual sum of squares $RSS_{(p)}$ are as follows:

Regressors Included	p	$RSS_{(p)}$	C_p	S_p
Constant		1 054.80		
x_1		436.35		
x_2		447.22		
x_3		923.98		
x_4		785.65		
x_1 x_2		234.90		
x_1 x_3		431.09		
x_1 x_4		357.25		
x_2 x_3		441.77		
x_2 x_4		438.49		
x_3 x_4		712.22		
x_1 x_2 x_3		214.81		
x_1 x_2 x_4		230.32		
x_1 x_3 x_4		354.51		
x_2 x_3 x_4		435.13		
x_1 x_2 x_3 x_4		213.09		

(a) Complete the table by filling in the p value and computing the C_p and S_p values. **(6marks)**

(b) With the help of graphs use the C_p and S_p statistics to select the 'best' linear regression model. **(2 marks)**

(c) Use Forward Selection to select the 'best' model. **(6 marks)**

(d) Use Backward Elimination to select the 'best' model. **(6marks)**

(e) Use Stepwise Regression to select the 'best' model. **(4 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 *** ----------