

NATIONAL UNIVERSITY OF SCIENCE AND TECHNOLOGY

FACULTY OF APPLIED SCIENCES

SMA 4105

DEPARTMENT OF APPLIED MATHEMATICS

EXPERIMENTAL DESIGN & MULTIPLE REGRESSION

NOV/DEC 2004

Time: 3 HOURS

ANSWER **ALL QUESTIONS**

SECTION A

1. The following data show a segment of the series on percentage wage change (Y) unemployment (X_1) and inflation (X_2)

Y	7	3	6	4	6	3	7	4	7	6
X_1	7	2	6	5	6	2	7	5	7	6
X_2	5	0	2	1	2	0	5	1	5	2

- (ii) State the model in matrix notation. Define all terms. [6 marks]

(iii) If $(X'X)^{-1} = \begin{pmatrix} 1.62051 & -0.39872 & 0.25769 \\ & 0.11571 & -0.09327 \\ & & 0.10288 \end{pmatrix}$

Find $\hat{\beta} = (X'X)^{-1} X'Y$. [6 marks]

- (iv) Construct the ANOVA TABLE for this data and hence test the hypothesis that $\beta_1 = \beta_2 = 0$ at $\alpha = 0,05$. [8 marks]

- (b) In multiple linear regression the least squares estimator β of the regression coefficient is given by $\hat{\beta} = (X'X)^{-1} X'Y$.

- (i) Show that $\hat{\beta}$ is an estimator of β . [4 marks]
(ii) Show that $\text{Var}(\hat{\beta}) = \sigma^2(X'X)^{-1}$. [4 marks]

SECTION B

2. A router is used to cut locating notches on a printed circuit board. The vibration level at the surface of the board as it is cut is considered to be a major source of dimensional variation in the notches. 3 factors are thought to influence vibration i.e. cutting speed (A) metal hardness (B) and the bit size (C). A 2^3 factorial experiment replicated 3 times was run and the following results were obtained.

Treatment Combination	REPLICATE		
	I	II	III
(I)	195	177	160
a	290	300	270
b	290	142	270
c	300	306	287
ab	354	350	320
ac	255	258	270
bc	360	355	390
abc	270	276	265

If SSTO=134588

- (i) Compute the sums of squares of all main effects and interactions by the contrast methods. [16 marks]
(ii) Construct an Analysis of Variance table and test for the significance of all main effects and interactions. [8 marks]
3. The yield of a chemical process 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)	Concentration (x_1)	Temperature (x_2)
82	3.00	140
90	3.00	170
84	4.00	140
92	4.00	170
80	3.00	140
88	3.00	170
85	4.00	140
91	4.00	170

- (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 analysis of variance table indicating significance on each regressor variable. Draw your conclusion. [14 marks]

4. An experiment is concluded to determine if there is a difference in the breaking strengths of a fibre produced by three machines. However the strength of the fibre is also affected by its thickness. A thicker fibre is stronger than a thin one. Therefore the diameter X of the fibre is used as a covariate. A random sample of five fibre specimens is selected from each machine, and the fibre breaking strength Y and the corresponding diameter X of each specimen measured. The data is as shown in the following table.

Specimen #	1		2		3	
	Y	X	Y	X	Y	X
1	36	20	40	22	35	21
2	41	25	48	28	37	23
3	39	24	39	22	42	26
4	42	25	45	30	34	21
5	49	32	44	28	32	15
Total	207	126	216	130	180	106

$$\begin{aligned} \sum Y &= 603 & \sum Y^2 &= 24587 & \sum XY &= 14835 \\ \sum X &= 362 & \sum X^2 &= 8998 & & \end{aligned}$$

NB 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 significance differences in the breaking strength Y on the three machines by the Analysis of Variance (ANOVA). [8 marks]
- (b) Using the covariate X, Perform the analysis of covariance (ANCOVA) on the data. [12 marks]
- (c) Compare the results obtained in (a) with those obtained in (b). [4 marks]
5. A nickel titanium alloy is used to make components for jet turbine aircraft engines. Cracking is a potentially serious problem in the final part as it can lead to non recoverable failure. A test is run at the parts producer to determine the effect of four variables on cracks. The four variables are pouring temperature (x_1) titanium content (x_2), heat treatment method (x_3) and amount of grain refiner used (x_4). A total of 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, were examined and the residual sums of squares RSS (p) are as follows.

Regression Included	RSS	P	C_p	S_p
Constant	2 715,7635			
x_1	1 265, 6867			
x_2	906,3363			
x_3	1 939,4005			
x_4	883,8669			
x_1, x_2	57,9045			
x_1, x_3	1 227,0721			
x_1, x_4	74,7621			
x_2, x_3	415,4427			
x_2, x_4	868,8801			
x_3, x_4	175,7380			
x_1, x_2, x_3	48,1106			
x_1, x_2, x_4	47,9727			
x_1, x_3, x_4	50,8361			
x_2, x_3, x_4	73,8145			
x_1, x_2, x_3, x_4	47,8636			

Complete the table and use the following methods to select the “best” linear regression model.

- (a) The c_p and s_p statistics. (8 marks)
 (b) Forward selection. (8 marks)
 (c) Backward elimination. (8 marks)

Show your working clearly and logically.

6. Discuss clearly what is meant by each of the following giving an appropriate example for each.
 (a) Complete confounding (8 marks)
 (b) Partial confounding (8 marks)
 (c) Aliasing. (8 marks)

END OF QUESTION PAPER