NATIONAL UNIVERSITY OF SCIENCE AND TECHNOLOGY

Faculty of Applied Sciences

SMA 5151

DEPARTMENT OF APPLIED MATHEMATICS

SMA 5151: APPLIED STATISTICS

DECEMBER 2004
Time : 3 hours

Candidates should attempt **ALL** questions from Section A and **ANY THREE** questions from Section B.

SECTION A: Answer ALL questions in this section (56 Marks)

A1. The following are cholesterol contents, in milligrams per package, which four laboratories obtained for 20 gram packages of three very similar diet foods.

	Diet Food A	Diet Food B	Diet Food C
Laboratory 1	3.4	2.6	2.8
Laboratory 2	3.0	2.7	3.1
Laboratory 3	3.3	3.0	3.4
Laboratory 4	3.5	3.1	3.7

Perform a two way analysis of variance and test the null hypotheses concerning the diet foods and the laboratories. [16]

A2. Explain the term 'Lack of Fit' (you can use graphs to help your explanation) and under what conditions is it possible to calculate it. [5]

A3. Show that in linear regression, the hat matrix,

$$\mathbf{H} = \mathbf{X}(\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T \text{ and the matrix } \quad \mathbf{Q} = (\mathbf{I} - \mathbf{H}) \text{ are}$$

(i) idempotent and [3,4]

(ii) symmetric. [3,4]

A4. Consider the simple linear regression model $Y_i = \beta_0 + \beta_1 x_i + e_i$;

(a) show that $\widehat{\beta}_1 = \frac{s_{xy}}{s_{xx}}$ is an unbiased estimator of β_1 and that $\widehat{\beta}_0 = \bar{y} - \widehat{\beta}_1 \bar{x}$ is an unbiased estimator of β_0. [4,4]

(b) Derive expressions for the variances of $\widehat{\beta}_0$ and $\widehat{\beta}_1$. [4,4]

A5. Show that $S_{xy} = \sum_1^n (x_i - \bar{x})(y_i - \bar{y}) = \sum_1^n (x_i - \bar{x}) y_i$. [5]

SECTION B: Answer ANY TWO Questios in this section (44 Marks)

B6. Suppose four basketbal players were selected at random in order to test their accuracy levels (in % shots in) in ball-shooting. The heights (in metres) and the hand spans (in cm)were recorded as in the table below.

player	1	2	3	4
Accuracy(%)	80	90	60	82
Height(m)	1.78	1.92	1.60	1.81
Span(cm)	18	23	20	17

Suppose you categorise the heights in the two groups; (0 to 1.80m) and above 1.80m. Suppose you also categorise the palm spans into the two groups; (above 20 cm inclussive) and below 20 cm.

(a) Write an appropriate two-factor model and carry out a two-way ANOVA for testing the effects of the two treatments. [8]

(b) Estimate the parameters of the model in (a). [6]

(c) Carry out one-way ANOVA for each of the two treatments. [8]

B7. The following are the exam grades of samples from three groups of students who were taught a certain language using three methods: method 1 was lectures and laboratory practicals; method 2 was only lectures; method 3 was laboratory practicals only.

Method 1: 94, 88, 91, 74, 87, 97
Method 2: 85, 82, 79, 84, 61, 72, 80
Method 3: 89, 67, 72, 76, 69

(a) Use the Kruskal Wallis test at the 0.05 significance level to test for hypothesis that the three methods are equally effective.The Kruskal Wallis test statistic is

$$H = \left(\frac{12}{n(n+1)} \sum_{i=1}^{k} \frac{R_i^2}{n_i} - 3(n+1) \right) \sim \chi^2(k-1) \text{ where } k = \text{number of groups.}$$ [12]

(b) Show that the Kruskal Wallis test statistic is equivalent to

$$H = \frac{12}{n(n+1)} \sum_{i=1}^{k} n_i \left[\frac{R_i}{n_i} - \frac{n+1}{2}\right]^2$$ [10]

B8. Consider the following data:

x:	16	17	26	24	22	21	32	18	30	20
y:	22	26	45	37	28	50	56	34	60	40

The regression model suggested for this data is $Y_i = \alpha + \beta(x_i - \bar{x}) + e_i$.

(a) Make a scatter plot of Y against X. Calculate the values of $(x_i - \bar{x})$ and plot these values against Y on the graph above. Comment on the plots. [2]

(b) Obtain the least squares estimates of α and β. [8]

(c) Obtain the residuals $(y_i - \widehat{y}_i)$ and hence calculate the variance of Y. [6]

(d) Obtain the values of $(h_{ii},\ i = 1, 2\ldots, 7)$ using the estimation $h_{ii} = \frac{1}{n} + \frac{(x_i - \bar{x})^2}{s_{xx}}$. Comment on the values of the h_{ii}'s and what they imply. [4]

(e) Test for the **influence** of the third observation using the Cook's statistic. [2]

(Cook's statistic is given by $Q_i = \frac{\widehat{e_i}^2 h_{ii}}{p\widehat{\sigma}^2(1 - h_{ii})^2} \sim F_{(p,n-p)}(\alpha)$).

END OF QUESTION PAPER