SMA2204

NATIONAL UNIVERSITY OF SCIENCE AND TECHNOLOGY

DEPARTMENT OF APPLIED MATHEMATICS

B.Sc. PART II HONOURS EXAMINATIONS 2003

SMA 2204 – APPLIED STATISTICS

MAY 2003

3 HOURS (100 Marks) This paper has 5 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) Two educators hypothesised that there was a relationship between students' home conditions and their reading ability. Under "home conditions" they included factors such as whether or not the home contained books and magazines, whether or not the parents were readers, and whether or not the children were provided with appropriate reading material. These factors were then rated as excellent, average, or poor. A sample of 200 produced the following results where reading ability is rated as A (excellent), B (good), C (satisfactory), and D (poor).

| | Reading Ability | | | | |
|---|---|---|---|---|---|
| Home Conditions | A | B | C | D | **Total** |
| Excellent | 15 | 30 | 20 | 5 | 70 |
| Average | 20 | 15 | 19 | 13 | 67 |
| Poor | 10 | 13 | 18 | 22 | 63 |
| **Total** | 45 | 58 | 57 | 40 | 200 |

Test at the 5% the level of significance whether reading ability and home conditions are statistically independent. **(8 marks)**

(b) The average red-cell count of a population of healthy persons is 4.5 [in millions per cubic millimetre]. A random sample of 36 people in an area in which there are low socio-economic conditions results in a mean of 3.5 and a standard deviation of 0.6. At a significance level of 0.01, can we conclude that in this area the mean red-cell count is lower than the overall population average? **(5 marks)**

(c) Two groups of occupational therapy patients, taught to assemble a simple mechanism by two different methods, A and B, are timed to find out how long it takes to complete the process of assembly. The times of both groups are normally distributed. A random sample of 12 is taken from the group taught by Method A. It produced a mean of 7.2 minutes with a standard deviation of 2.3 minutes. From group B, a random sample of 10 patients assembled the mechanism in a mean time of 5.9 minutes with a standard deviation of 1.8 minutes.

(i) At a significance level of 0.01, can we conclude that the two methods produce different results? **(6 marks)**

(ii) What assumption have you made in (i) about the two population variances from which the two samples were drawn? Carry out a test to determine whether this assumption holds or not. **(5 marks)**

(d) If X and Y are random variables with a joint distribution such that Var (X) = 2, Var (Y) = 4, and Cov (XY) = -2; find

(i) Var (Z), where $Z = 3X - 4Y + 8$ **(1 mark)**

(ii) Cov (P,Q), where $P = X - 2Y$, and $Q = 3X + Y$. **(3 marks)**

**SECTION B (Answer any THREE QUESTIONS)**

2. (a) A manufacturer of lamps wanted to know whether there was any difference in the results of advertising in newspapers, magazines, or on TV. She took a random sample of sales for five weeks and noted the number of sales each week attributable to each type of advertising. The random variable is the amount of sales in thousands of dollars. She obtained the following results.

| Newspapers | Magazines | TV |
|---|---|---|
| 15 | 10 | 14 |
| 18 | 7 | 27 |
| 19 | 12 | 30 |
| 20 | 9 | 19 |
| 25 | 19 | 21 |

(i) Test at the 5% level of significance, whether the advertising medium makes a difference in sales. **(8 marks)**

(ii) If over the years the traditional medium of advertising for the manufacturers has been the newspaper; compare the other two media with it at the 5% level of significance. **(6 marks)**

(b) A manufacturing company produces wire by four different methods. Quality Control

Department wishes to know whether there are any differences in the tensile strength of wire produced by these different methods. Researchers take an independent sample from each machine at each of five locations and obtain the following data:

| | METHOD | | | |
|---|---|---|---|---|
| LOCATION | I | II | III | IV |
| Bulawayo | 56 | 64 | 45 | 42 |
| Gweru | 55 | 61 | 46 | 39 |
| Mutare | 62 | 50 | 45 | 45 |
| Kwekwe | 59 | 55 | 39 | 43 |
| Harare | 60 | 56 | 43 | 41 |

Test the null hypothesis of no significant differences in tensile strength among the four methods. Use the 0.01 level of significance. **(10 marks)**

3. (a) It is claimed that a new diet will reduce a person's weight by 4.5kilograms on the average in a period of two weeks. The weights of seven women who followed this diet were recorded before and after a two-week period.

| | Woman | | | | | | |
|---|---|---|---|---|---|---|---|
| | 1 | 2 | 3 | 4 | 5 | 6 | 7 |
| Weight before | 58.5 | 60.3 | 61.7 | 69.0 | 64.0 | 62.6 | 56.7 |
| Weight after | 60.0 | 54.9 | 58.1 | 62.1 | 58.5 | 59.9 | 54.4 |

Assuming the distribution of weights to be approximately normal

(i) Test the manufacturer's claim that the new diet will reduce a person's weight by 4.5kg on average. Use the 5% level of significance. **(8 marks)**

(ii) Find the 95% confidence interval for the true mean difference in the weight. Draw your conclusion to the manufacturer's claim based on the 95% confidence interval. **(4 marks)**

(b) An industrial engineer is investigating the effect of four assembly methods (A, B, C, D) on the assembly time for a colour television component. Four operators are selected for the study. Furthermore, the engineer knows that each assembly method produces such fatigue that the time required for the last assembly may be greater than the time required for the first, regardless of method. That is, a trend develops in the required assembly time. The engineer takes into account this source of variability in his experimental design and obtains the following results.

| **Order of Assembly** | **Operator** | | | |
|---|---|---|---|---|
| | 1 | 2 | 3 | 4 |
| 1 | 10 (C) | 14 (D) | 7 (A) | 8 (B) |
| 2 | 7 (B) | 18 (C) | 11 (D) | 8 (A) |
| 3 | 5 (A) | 10 (B) | 11 (C) | 9 (D) |
| 4 | 10 (D) | 10 (A) | 12 (B) | 14 (C) |

Test at the 5% level of significance whether there are any significant differences in assembly time among the four assembly methods. **(12 marks)**

4. The maximum volume of oxygen uptake ($VO_2$ max) has been used as a measure of cardiac status in healthy individuals as well as in persons suffering from cardiac-related illnesses (such as congestive heart failure). The $VO_2$ max readings for 12 healthy adult males following strenuous exercise are shown in the following table. In general, $VO_2$ max decreases with any increase in activity level.

| Individual | $VO_2$ max<br>y | Duration of exercise (minutes)<br>x |
|---|---|---|
| 1 | 82 | 10.0 |
| 2 | 73 | 9.5 |
| 3 | 68 | 10.2 |
| 4 | 74 | 10.5 |
| 5 | 66 | 11.0 |
| 6 | 63 | 11.3 |
| 7 | 58 | 11.6 |
| 8 | 54 | 12.0 |
| 9 | 56 | 12.1 |
| 10 | 51 | 12.5 |
| 11 | 55 | 12.8 |
| 12 | 44 | 13.0 |

(a) Draw a scatter diagram of y against x. Fit the least squares straight line to these data and draw your fitted line on your scatter diagram. **(10 marks)**

(b) Carry out an analysis of variance (ANOVA) to test at the 5% level of significance whether the slope is significantly different from zero. From the ANOVA table, compute the coefficient of determination, $r^2$, and interpret it. **(8 marks)**

(c) Predict the maximum volume of oxygen uptake ($VO_2$ max) when the individual's duration of exercise is 10.8 minutes. Compute the standard error of the predicted value and hence find the 95% confidence interval of the predicted value. **(6 marks)**

5. An experiment was set up to compare the effect of different soil pH and calcium additives on the increase in trunk diameters for orange trees. Annual applications of elemental sulphur, gypsum, soda ash, and other ingredients were applied to provide pH value levels of 4, 5, 6, and 7. Three levels of a calcium supplement (100, 200, and 300 kg per ha) were also applied. All factor-level combinations of these two factors were used in the experiment. At the end of a two-year period, three diameters were examined at each factor-level combination and the following data recorded.

| | Calcium | | |
|---|---|---|---|
| **pH value** | **100** | **200** | **300** |
| **4.0** | 5.2<br>5.9<br>6.3 | 7.4<br>7.0<br>7.6 | 6.3<br>6.7<br>6.1 |
| **5.0** | 7.1<br>7.4<br>7.5 | 7.4<br>7.3<br>7.1 | 7.3<br>7.5<br>7.2 |
| **6.0** | 7.6<br>7.2<br>7.4 | 7.6<br>7.5<br>7.8 | 7.2<br>7.3<br>7.0 |
| **7.0** | 7.2<br>7.5<br>7.2 | 7.4<br>7.0<br>6.9 | 6.8<br>6.6<br>6.4 |

(a) Construct a two-way table of totals. **(2 marks)**

(b) Carry out a preliminary ANOVA to test for the significance of treatment combinations at the 5% level of significance. **(6 marks)**

(c) Carry out a complete ANOVA to test for the significance of the main effects and interaction at the 5% level of significance. **(10 marks)**

(d) Give an appropriate table of means and appropriate standard errors for carrying out statistical comparisons. **(6 marks)**

6. Explain clearly how randomisation is done in each of the following:

(a) A Completely Randomized Design with four treatments and 16 experimental units . **(4 marks)**

(b) A Randomized Complete Block Design with four treatments in four blocks of size four experimental units each. **(8 marks)**

(c) A 4 x 4 Latin Square Design. **(12 marks)**

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