SMA2204

NATIONAL UNIVERSITY OF SCIENCE AND TECHNOLOGY

DEPARTMENT OF APPLIED MATHEMATICS

B.Sc. PART II HONOURS SUPPLEMENTARY EXAMINATIONS 2005

**SMA 2204 – APPLIED STATISTICS**

**JULY 2005**

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. **(10 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? **(4 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? **(5 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 total of 210 emphysema patients entering a clinic over a one-year period were treated with one of two drugs (either the **standard drug, A**, or an experimental **compound, B**) for a period of one week. After this period, each patient's condition was rated as **greatly improved**, **improved**, or **no change**. The sample results are shown below:

| | **Patient's Condition** | | |
|---|---|---|---|
| **Therapy** | No Change | Improved | Greatly Improved |
| Standard A | 20 | 35 | 45 |
| Experimental B | 15 | 45 | 50 |

Test, at the 5% level of significance, whether patient's condition is independent of therapy. **(8 marks)**

(b) Health authorities wish to compare the incubation period (in days) of a certain childhood disease in four different areas of a city. They take a random sample of six patients in each of the four areas and obtain the following incubation periods (in days)

| Area 1 | Area 2 | Area 3 | Area 4 |
|---|---|---|---|
| 13.0 | 12.5 | 7.9 | 8.6 |
| 12.5 | 8.4 | 8.5 | 10.7 |
| 18.0 | 9.7 | 10.2 | 9.3 |
| 15.2 | 10.6 | 11.8 | 13.4 |
| 17.8 | 16.1 | 12.0 | 11.1 |
| 19.0 | 14.0 | 15.1 | 12.4 |

Determine, at the 5% level of significance, whether there are any significant differences in the average incubation periods in the four different areas of the city. **(16 marks)**

3. (a) An experiment was conducted to compare the mean number of tapeworms in the stomachs of sheep that had been treated for worms against the mean number in those that were untreated. A sample of 14 worm-infected lambs was randomly divided into two groups. Seven were injected with the drug and the other seven were left untreated. After a six-month period, the lambs were slaughtered and the following worm counts were recorded:

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| **Drug-treated sheep** | 18 | 43 | 28 | 50 | 16 | 32 | 13 |
| **Untreated sheep** | 40 | 54 | 26 | 63 | 21 | 37 | 39 |

Make the relevant assumptions and carry out the appropriate test at the 5% level of significance to test the hypothesis that there is no significant difference in the mean number of worms between treated and untreated lambs. Assume that the drug cannot increase the number of worms and hence use the alternative hypothesis that the mean for treated lambs is less than the mean for untreated lambs. **(12 marks)**

(b) Two analysts, supposedly of identical abilities, each measure the parts per million of a certain type of chemical impurity in drinking water. It is claimed that **Analyst 1** tends to give higher readings than **Analyst 2**. To test this claim, each of six water samples is divided and analysed by both analysts separately. The following results (in ppm) are recorded:

| **Water Sample** | 1 | 2 | 3 | 4 | 5 | 6 |
|---|---|---|---|---|---|---|
| **Analyst 1** | 31.4 | 37.0 | 44.0 | 28.8 | 59.9 | 37.6 |
| **Analyst 2** | 28.1 | 37.1 | 40.6 | 27.3 | 58.4 | 38.9 |

State the underlying assumptions in this experimental design and hence test whether there is any evidence to indicate that **Analyst 1** reads higher on the average than **Analyst 2**. Use the 5% level of significance. **(12 marks)**

4. The laboratory of a hospital participating in the clinical trial of an antibiotic drug had to be validated to see that the laboratory personnel could accurately assay blood samples "spiked" with fixed amounts of the antibiotic. The validation consisted of the following experiment. Ten spiked samples (with amounts known only to the study investigator) were sent to the laboratory to be assayed for the amount of antibiotic present. The results of the validation experiment are shown below. (Note: The spiked samples with known amounts added, **x**, in micrograms per millilitre, were supplied in a blinded fashion to the laboratory.) The amounts found, **y**, in micrograms per millilitre, are the assay results found by the hospital laboratory.

| **y** | 0 | 4.5 | 5.0 | 4.8 | 8.9 | 8.9 | 17.0 | 18.2 | 32.6 | 31.5 |
|---|---|---|---|---|---|---|---|---|---|---|
| **x** | 0 | 5 | 5 | 5 | 10 | 10 | 20 | 20 | 40 | 40 |

(a) Draw a scatter diagram of y against x. Fit the least squares straight line to the 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 amount of antibiotic found by the hospital laboratory when the actual amount present in the blood is 15 micrograms per millilitre. 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 conducted to determine the effects of four different pesticides on the yield of fruit from three different varieties ($V_1$, $V_2$, $V_3$) of citrus. Eight trees from each variety were randomly selected from an orchard. The four pesticides were each randomly assigned to two trees of a particular variety and applications were made according to recommended levels. Yields of fruit, in bushels per tree, were obtained after the test period and are shown in the following table:

| | **PESTICIDE (P)** | | | |
|---|---|---|---|---|
| **VARIETY (V)** | **1** | **2** | **3** | **4** |
| **1** | 49<br>39 | 50<br>55 | 43<br>38 | 53<br>48 |
| **2** | 55<br>41 | 67<br>58 | 53<br>42 | 85<br>73 |
| **3** | 66<br>68 | 85<br>92 | 69<br>62 | 85<br>99 |

(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. **(8 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 the comparisons. **(4 marks)**

6. The Celtics were a vigorous race of people who once populated part of England. It is not entirely clear whether they simply died out or merged with other people who were the ancestors of those who live in England today. A study to determine whether modern Englishmen are of genetic stock comparable to the Celtics was based on the comparison of maximum head breadths (in millimetres) made on unearthed Celtic skulls and a comparable number of skulls of modern-day Englishmen. The following measurements were recorded:

| | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| **English:** | 141 | 148 | 132 | 138 | 154 | 142 | 150 | 146 | 155 |
| | 158 | 150 | 140 | 147 | 148 | 144 | 150 | 149 | 145 |
| **Celtic:** | 133 | 138 | 130 | 138 | 134 | 127 | 128 | 138 | 136 |
| | 131 | 126 | 120 | 124 | 132 | 132 | 125 | | |

(a) Construct the back-to-back **stem and leaf plots** of the data. **(5 marks)**

(b) Construct the **box plots** of the data on the same scale. **(5 marks)**

(c) Based on your plots in (a) and (b) above is there evidence to suggest that the modern Englishmen are of similar genetic stock to the Celtics? **(4 marks)**

(d) Carry out an appropriate statistical test to test whether there is any evidence of a significant difference in the mean breadth of skulls of the two groups. **(10 marks)**

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