

SMA2115

NATIONAL UNIVERSITY OF SCIENCE AND TECHNOLOGY
DEPARTMENT OF APPLIED MATHEMATICS
BSc HONOURS PART II EXAMINATIONS 2005

SMA 2115 – INTRODUCTION TO APPLIED STATISTICS

NOVEMBER/ DECEMBER 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) In each case identify the appropriate probability distribution and compute the required probability:
- (i) The probability of a female birth in a certain area is 0.45. What is the probability that a couple who have five children will have at least three girls? **(4 marks)**
 - (ii) In a certain hospital, the probability of a multiple birth is 0.01. Suppose that you take a random sample of 20 deliveries. What is the probability of no more than one multiple birth? **(4 marks)**
 - (iii) It is known from experience that 1.4 percent of the calls received by the NUST switchboard are wrong numbers. What is the probability that among 150 calls received by the NUST switchboard two are wrong numbers? **(4 marks)**
 - (iv) Among the 16 applicants for a job, ten have university degrees. If three of the applicants are randomly chosen for interviews, what is the probability that all three have university degrees. **(4 marks)**
- (b) A certain kind of appliance requires repairs on the average once every two years. Assuming that the times between repairs are exponentially distributed, what is the probability that such an appliance will work at least three years without requiring repairs? **(4 marks)**

(c) In a photographic process, the developing time of prints may be looked upon as a random variable having a normal distribution with a mean of 15.40 seconds and a standard deviation of 0.48 second. Find the probabilities that the time it takes to develop one of the prints will be:

(i) at least 16.00 seconds, (1 mark)

(ii) at most 14.20 seconds, (1 mark)

(iii) anywhere from 15.00 to 15.80 seconds. (2 marks)

(d) If X and Y are random variables with a joint probability distribution such that $\text{Var}(X) = 4$, $\text{Var}(Y) = 3$ and $\text{Cov}(X,Y) = 2$

(i) find $\text{Var}(Z)$, where $Z = 3X - 2Y + 5$ (1 mark)

(ii) find $\text{Cov}(P,Q)$, where $P = 3X + 4Y$, and $Q = X - 2Y$ (3 marks)

SECTION B (Answer any THREE questions)

2. In a study to investigate the therapeutic value of ascorbic acid (vitamin C) for treating the common cold, 279 students at a certain boarding school in Mangwe district of Matebeleland South province were involved during two periods of 5 to 7 days in the month of June 2005. The study was a double-blind with one group of 140 students receiving a placebo (*an inert substance with no therapeutic value*) while a second group of 139 students received one gram of ascorbic acid per day. The following frequencies were observed:

THERAPY	CONDITION		Total
	Cold	No Cold	
Placebo	31	109	140
Ascorbic Acid	17	122	139
Total	48	231	279

(i) Test at the 5% level of significance whether the proportion of those with cold in the placebo group is the same as the proportion of those with cold in the ascorbic acid treated group. (12 marks)

(ii) Test at the 5% level of significance whether there is any association between the occurrence of the common cold and the therapeutic treatment administered. (12 Marks)

3. The following data pertain to the chlorine residual in a swimming pool at various times after it has been treated with chemicals;

Number of hours, x	2	4	6	8	10	12
Chlorine residual, y (ppm)	1.8	1.5	1.4	1.1	1.1	0.9

- (a) Draw a scatter diagram of y against x. Fit a least squares 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. Compute the coefficient of determination, r^2 , and interpret it. **(8 marks)**
- (c) Predict the chlorine residual in the pool 5 hours after it has been treated with chemicals. Compute the standard error of the predicted value and hence find the 95% confidence interval for the predicted value. **(6 marks)**
4. 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 obtained:

pH Value	Calcium		
	100	200	300
4.0	5.2	7.4	6.3
	5.9	7.0	6.7
	6.3	7.6	6.1
5.0	7.1	7.4	7.3
	7.4	7.3	7.5
	7.5	7.1	7.2
6.0	7.6	7.6	7.2
	7.2	7.5	7.3
	7.4	7.8	7.0
7.0	7.2	7.4	6.8
	7.5	7.0	6.6
	7.2	6.9	6.4

- (a) Construct a two-way table of totals. **(2 marks)**
- (b) Carry out a preliminary ANOVA to test for the significant differences among the twelve treatment combinations at the 5% level of significance. **(8 marks)**

SMA2115

(c) Carry out a complete ANOVA to test for the significance of 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)

5(a) An investigator performs serum iron analyses on a random sample of 12 healthy subjects by both methods M and Q. The investigator wants to know whether there is sufficient evidence to conclude that method M produces a larger mean serum iron value than method Q. The following results are obtained:

Subject	1	2	3	4	5	6	7	8	9	10	11	12
Method M	100	118	115	110	117	108	114	116	115	122	114	115
Method Q	104	111	111	113	112	110	109	108	120	113	113	117

Make the relevant assumption and carry out the appropriate test at the 5% level of significance to test whether method M produces a significantly larger mean serum iron value than method Q. (10 marks)

(b) You are now told that a random sample of 12 subjects were investigated using method M and another random sample of 12 subjects were investigated using method Q and the results obtained are as in (a). Make the relevant assumptions and carry out the appropriate test at the 5% level of significance to test whether method M produces a significantly larger mean serum iron value than method Q. (10 marks)

(c) What are the advantages and disadvantages of the approach in (a) over that in (b) (4 marks)

6 The following are the cholesterol contents (in milligrams per package), which four laboratories obtained for packages of three very similar diet foods:

LABORATORY	DIET FOOD		
	A	B	C
1	3.4	2.6	2.8
2	3.0	2.7	3.1
3	3.3	3.0	3.4
4	3.5	3.1	3.7

(a) A food nutritionist at the University of Zimbabwe decides that there are no systematic differences among the four laboratories and she therefore ignores the effect of

SMA2115

laboratory in her analysis to determine whether significant differences exist in cholesterol content of the three diet foods. Carry out this analysis at the 5% level of significance. (10 marks)

(b) A food nutritionist at NUST, however, suspects that there may be systematic differences among the four laboratories and she, therefore, includes this source of variation in her analysis to determine whether there are any significant differences in cholesterol content of the three diet foods. Carry out an analysis at the 5% level of significance. (12 marks)

(c) Of the two approaches, (a) and (b), which one is more precise and why? (2 marks)

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