

SMA2115

NATIONAL UNIVERSITY OF SCIENCE AND TECHNOLOGY
DEPARTMENT OF APPLIED MATHEMATICS
SUPPLEMENTARY EXAMINATIONS 2003

SMA 2115 – INTRODUCTION TO APPLIED STATISTICS

MARCH 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)

LIBRARY USE ONLY

1. (a) The probability that a patient recovers from a rare blood disease is 0.3. If 20 people are known to have contracted this disease, what is the probability that
- i) at least 10 survive **(3 marks)**
 - ii) from 3 to 8 survive **(3 marks)**
 - iii) exactly 5 survive **(2 marks)**
- (b) As part of an air pollution survey, an inspector decides to examine the exhaust of six of a company's 24 trucks. If four of the company's trucks emit excessive amounts of pollutants, what is the probability that none of the four will be included in the inspector's sample? **(4 marks)**
- (c) Records show that the probability is 0.00005 that a car will have a flat tyre while crossing a certain bridge. What is the probability that among 10 000 cars that cross the bridge
- i) exactly two will have a flat tyre? **(2 marks)**
 - ii) at most two will have a flat tyre? **(2 marks)**
- (d) The mean level of prothrombin in the normal population is known to be approximately 20 mg/100ml of plasma. A sample of 625 patients showing a vitamin K deficiency has

a mean prothrombin level of 18.50 mg/100ml. The sample standard deviation is 4 mg/100ml. Do patients with vitamin K deficiency have a significantly lower prothrombin level than that of the general population? **(5 marks)**

- (e) An electrical firm manufactures light bulbs that have a length of life that is normally distributed with mean equal to 800 hours and standard deviation of 40 hours.
- (i) Find the probability that a bulb burns between 778 and 834 hours. **(3 marks)**
 - (ii) Find the probability that a random sample of 16 bulbs drawn from this normal population will have an average life of less than 775 hours. **(4 marks)**

LIBRARY USE ONLY

SECTION B (Answer any THREE questions)

2. (a) The incomplete table below displays the distribution of the HIV status of a population in some region X.

GENDER	STATUS		Total
	HIV ⁺	HIV ⁻	
Female		125	180
Male		75	
Total	90		300

- (i) Complete the table. **(2 marks)**
 - (ii) Carry out an appropriate test at the 5% level of significance to test whether there is any association between HIV status and gender. **(5 marks)**
- (b) Two protein diets were allocated randomly to male chicks of the same age and breed. The weight (in grams) of chicks after four weeks are as follows:

DIET	WEIGHT (in gm)					
	High protein	150	136	144	168	152
Low protein	121	109	115	106	99	

- (i) Carry out a two-sample (pooled) t-test at the 5% level of significance to test whether there is a significant difference between the two protein diets. **(7 marks)**
- (ii) Carry out a one-way analysis of variance (ANOVA) to test whether there is a significant difference between the two protein diets at the 5% level of significance. **(8 marks)**
- (iii) Explain how the test in (i) is related to the test in (ii). **(2 marks)**

3. (a) Five specimens of a ferrous-type substance are to be used to determine if there is a significant difference between laboratory chemical analysis and X-Ray fluorescence analysis of the iron content. Each specimen was split into two sub-specimens and the two types of analysis were applied. The following are the results of the iron content analysis:

Analysis	Specimen				
	1	2	3	4	5
X-Ray	2.0	2.0	2.3	2.1	2.4
Chemical	2.2	1.9	2.5	2.3	2.4

State your assumptions and hence test at the 5% level of significance whether the two methods of analysis give, on average, the same result. **(9 marks)**

- (b) You are now told that ten specimens were available for the experiment and that five specimens were chosen at random and a chemical analysis applied. An X-Ray fluorescence analysis was applied on the other five specimens and the results were as shown in (a). State your assumptions and hence test at the 5% level of significance whether the two methods of analysis give, on average, the same result. **(10 marks)**
- (c) What are the advantages and disadvantages of the test in (a) over that in (b)? **(5 marks)**
4. 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 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)**

5. Determining small quantities of calcium in the presence of magnesium is a difficult problem for the analytical biochemist. Direct precipitation is not feasible. One of the procedures proposed involves the use of alcohol as a solvent. The data that follow present the results of applying the alcohol method to ten mixtures containing known quantities of CaO, x . The second row gives, in each instance, the amount of CaO recovered, y :

CaO present (x)	4.0	8.0	12.5	16.0	20.0	25.0	31.0	36.0	40.0	40.0
CaO recovered (y)	3.7	7.8	12.1	15.6	19.8	24.5	31.1	35.5	39.4	39.5

$$\sum_{i=1}^{10} x_i = 232.5; \quad \sum_{i=1}^{10} y_i = 229.0; \quad \sum_{i=1}^{10} x_i^2 = 6974.2; \quad \sum_{i=1}^{10} y_i^2 = 6796.7$$

$$\sum_{i=1}^{10} x_i y_i = 6884.65;$$

- (a) Draw a scatter diagram of y against x on graph paper. 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 0.05 level of significance whether the slope is significantly different from zero. From your ANOVA table, compute the coefficient of determination, r^2 , and interpret it. (8 marks)
- (c) Predict the amount of CaO recovered when the CaO present in the mixture is 10.5. Compute the standard error of the predicted value and hence construct the 95% confidence interval of this prediction. (6 marks)
6. The yield of a chemical process is being studied. The two most important factors are thought to be the pressure and temperature. Three levels of each factor are selected and a factorial experiment with two replicates performed. The yield data are as follows:

TEMPERATURE	PRESSURE		
	200	215	230
Low	90.4	90.7	90.2
	90.2	90.6	90.4
Medium	90.1	90.5	89.9
	90.3	90.6	90.1
High	90.5	90.8	90.4
	90.7	90.9	90.1

- (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 comparing any two means. (4 marks)

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