

NATIONAL UNIVERSITY OF SCIENCE AND TECHNOLOGY

DEPARTMENT OF APPLIED MATHEMATICS

B.Sc. PART II HONOURS EXAMINATIONS 2003

SMA 2215 – INTRODUCTION TO APPLIED STATISTICS

MAY 2003

3 HOURS (100 Marks)

This paper has 4 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) 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 at least two multiple births? **(4 marks)**
 - (ii) Suppose that 30% of the people in a given community church have university degrees. An education committee of 10 people is named. What is the probability that 4 out of the 10 will hold degrees? **(4 marks)**
 - (iii) Among the 120 applicants for a job, only 80 are actually qualified. If five of the applicants are randomly selected for an in-depth interview, find the probability that only two of the five will be qualified for the job. **(3 marks)**
 - (iv) A certain kind of appliance requires repairs on average once every two years. What is the probability that such an appliance will work for at least three years without requiring repairs? **(5 marks)**
 - (v) A textile producer has established that a spinning machine stops randomly due to the thread breakages at an average rate of 5 stoppages per hour. What is the probability that in a given hour, 3 stoppages will occur on this spinning machine? **(3 marks)**

LIBRARY USE ONLY

- (b) You have just purchased a new radio guaranteed to have an average life of two years with no trouble. Trouble-free lives of radios are normally distributed with a standard deviation of 0.5 years.
- (i) What is the probability that you will have no trouble with the radio for at least 3.5 years? **(2 marks)**
- (ii) What percentage of buyers of these radios would need to have them repaired within the first 15 months? **(4 marks)**
- (c) If X and Y are random variables with a joint probability distribution such that, $\text{Var}(X) = 9$, $\text{Var}(Y) = 4$, and $\text{Cov}(XY) = 3$;
- Find
- (i) the variance of the random variable $Z = 3X - 4Y + 8$ **(1 mark)**
- (ii) the covariance between the random variables $Z = 3X - 4Y$ and $W = 3X + 2Y$ **(2 marks)**

SECTION B (Answer any THREE QUESTIONS)

2. (a) A motor vehicle distributor wishes to find out if the size of car bought is in any way related to the age of a buyer. From sales invoices over the past two years, a sample of 300 buyers were classified by size of car bought and buyer's age. The following contingency table was constructed:

BUYER'S AGE	CAR SIZE		
	Small	Medium	Large
Under 30	10	22	34
30 - 45	24	42	48
Over 45	52	32	36

Test, at the 1% level of significance, whether car size bought and buyer's age are independent. Interpret your findings. **(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 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)

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

4. A study was conducted to examine the quality of fish after seven days in ice storage. Ten raw fish of the same kind and approximately the same size were caught and prepared for ice storage. Two of the fish were placed in storage immediately after being caught, and two each were placed in storage at 6, 9, and 12 hours after being caught. Let y denote a measurement of fish quality (on a 10-point scale) after the seven days of storage, and x denote the time after being caught that the fish were placed in ice packing. The sample data are as follows:

y	8.5	8.4	7.9	8.1	7.8	7.6	7.3	7.0	6.8	6.7
x	0	0	3	3	6	6	9	9	12	12

- (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 fish quality when the fish is placed in ice storage 7 days after being caught. Compute the standard error of the predicted value and hence find the 95% confidence interval of the predicted value.
(6 marks)

LIBRARY USE ONLY

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:

VARIETY (V)	PESTICIDE (P)			
	1	2	3	4
1	49	50	43	53
	39	55	38	48
2	55	67	53	85
	41	58	42	73
3	66	85	69	85
	68	92	62	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. List the ADVANTAGES and DISADVANTAGES of each of the following experimental designs:
- (a) The Completely Randomized Design. (6 marks)
- (b) The Randomized Complete Block Design. (6 marks)
- (c) The Latin Square Design. (6 marks)
- (d) The Factorial Experiment (6 marks)

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

LIBRARY USE ONLY