

SMA2115

NATIONAL UNIVERSITY OF SCIENCE AND TECHNOLOGY

DEPARTMENT OF APPLIED MATHEMATICS

SMA2115 INTRODUCTION TO APPLIED STATISTICS

NOVEMBER/DECEMBER 2002

TIME: 3 HOURS

Total = 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) A car hire firm rents out only BMW and Mazda cars. Experience has shown that one in four clients chooses a BMW. If 5 reservations are randomly selected from today's bookings, what is the probability that 2 will have requested a BMW? **(3 marks)**
 - (ii) A textile producer has established that a spinning machine stops randomly due to 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)**
 - (iii) To avoid detention at customs, a traveller has placed six narcotic tablets in a bottle containing nine vitamin pills that are similar in appearance. If the customs official selects three of the tablets at random for analysis, what is the probability that the traveller will be arrested for illegal possession of narcotics? **(6 marks)**
 - (iv) The number of seconds of continuous spray yielded by cans of a certain brand of (ozone friendly) deodorant spray is normally distributed with a mean of 260 seconds and a standard deviation of 15 seconds. What is the probability that a can selected at random will yield a continuous spray of duration between 245 and 275 seconds? **(4 marks)**

SMA2115

- (b) If X and Y are random variables with variances, $\text{var}(X) = 2$, $\text{var}(Y) = 4$ and covariance, $\text{cov}(X, Y) = -2$; find
- (i) the variance of the random variable $Z = 3X - 4Y + 8$. (2 marks)
- (ii) the covariance between the random variables $Z = 3X - 4Y$ and $W = 3X + 2Y$ (4 marks)

- (c) When auto tellers were first introduced, a building society conducted a public survey in Harare four months after it installed experimental auto tellers (ATM's) in three of its branches in the city. The building society wished to establish if public acceptance would be affected by income level. The results of the survey are shown in the following contingency table:

INCOME LEVEL

RESPONSE	Low	Middle	High
Approve of ATM's	30	45	23
Disapprove of ATM's	30	35	27

At the 1% significance level, establish if public acceptance of ATM's was conditional upon income level. Comment on your findings. (6 marks)

SECTION B (Answer any THREE QUESTIONS)

2. (a) By examining X-rays, two doctors determine the severity of ulcers in a group of 12 patients following two different treatment regimes. The doctors want to know whether the results of the two kinds of treatment are significantly different at the 5% level of significance. Make the relevant assumption and carry out the appropriate test on the following results obtained by the doctors: (10 marks)

Patient	1	2	3	4	5	6	7	8	9	10	11	12
Treatment A	3.2	7.8	4.9	7.1	3.8	5.4	2.7	7.3	4.2	4.8	7.0	5.9
Treatment B	2.8	7.3	4.2	5.7	3.4	2.9	5.6	4.9	7.1	6.4	6.2	6.9

- (b) Let us now suppose that the results in 2(a) were obtained from 24 ulcer patients 12 of which were randomly allocated treatment A and the other 12 treatment B. Make relevant assumption(s) and carry out the appropriate test under these circumstances to test whether there is any significant difference between the two treatment regimes at the 5% level of significance. **(10 marks)**

- (c) What are advantages and disadvantages of the experimental design in (a) over that in (b)? **(4 marks)**

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

DIET FOOD			
LABORATORY	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 this 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)**

4. Most of South Africa's power stations are coal-fired. A random sample of 10 power stations were selected and their coal usage and electricity generated for 1996 was obtained as follows:

Coal usage (10^6 tonnes)	15	6	10	18	9	7	14	11	5	8
Electricity generated (10^6 kilowatt hrs)	35	18	24	32	24	20	32	29	14	22

- (a) Draw a scatterplot of electricity generated (y) against coal usage (x). Fit the least squares straight line to these data and draw your fitted line on your scatterplot. **(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 electricity generated when coal usage is 12 million tonnes. Compute the standard error of the predicted value and hence find the 95% confidence interval of the predicted value. **(6 marks)**

5. 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 the 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)**
6. Explain clearly with appropriate examples, how randomisation is done in each of the following:
- (a) a Completely Randomised Design **(4 marks)**
- (b) a Randomised Complete Block Design. **(8 marks)**
- (c) a Latin Square Design. **(12 marks)**