

NATIONAL UNIVERSITY OF SCIENCE AND TECHNOLOGY

SMA 2115

FACULTY OF APPLIED SCIENCES  
BACHELOR OF SCIENCE HONOURS DEGREE EXAMINATIONS  
DEPARTMENT OF APPLIED MATHEMATICS  
SMA 2115: INTRODUCTION TO APPLIED STATISTICS

DECEMBER 2004

3 HOURS (100 MARKS)

INSTRUCTIONS

This paper has TWO sections and THREE pages. Candidates should attempt ALL questions from Sections A and ANY THREE questions from Section B. Statistical tables and graph papers are provided.

SECTION A

Answer ALL questions in this section (40 marks)

- A1. Explain what each of the following terms mean:
- a) random sample (1 mark)
  - b) discrete random variable (1 mark)
  - c) probability distribution (1 mark)
- A2. The exam scores of 10 randomly selected students from Applied Statistics are as follows:  
30; 80; 40; 30; 90; 90; 40; 80; 60; 80;
- Find:
- a) the mean score (1 mark)
  - b) the modal score (1 mark)
  - c) the median score (1 mark)
  - d) the standard deviation (6 marks)
- A3. a) State two advantages of using the mean as a measure of centrality. (2 marks)  
b) State two disadvantages of using the median as a measure of centrality. (2 marks)
- A4. If the count of a certain species of insect per square metre of a football pitch is a random variable having the Poisson distribution with  $\theta = 0.9$ , find the probability that;
- a) 9 insects are found in 5 square metres. (5 marks)
  - b) 3 or less insects are found in 11 square metres. (5 marks)
- A5. The probability that a patient recovers from a rare blood disease is 0.2. If 20 people are known to have contracted this disease, what is the probability that;
- a) exactly 5 survive (2 marks)
  - b) at least 3 survive (2 marks)
  - c) from 4 to 8 survive (3 marks)

A6. Tutorials in a university are supposed to last 40 minutes. However, a Mathematics tutor finds that students are usually late in arriving for his tutorials, and that the actual length for his tutoring time available can be modeled by a normal distribution with mean 34.8 minutes and standard deviation 1.6 minutes.

- (i) Find the probability that the length of tutoring time available will be less than 37.0 minutes. (3 marks)
- (ii) The probability that the length of tutoring time available exceeds  $m$  minutes is 0.75. Find  $m$ . (4 marks)

**SECTION B**

Answer any THREE questions (60 marks)

- B7. Explain clearly with appropriate examples, how randomization is done in each of the following:
- a Completely Randomised Design (4 marks)
  - a Randomised Complete Block Design (6 marks)
  - a Latin Square Design. (10 marks)
- B8. a) The mean time to mark a certain set of examination papers is estimated by the examination board to be 12 minutes per paper. A random sample of 150 examination papers gave  $\sum x = 2130$  and  $\sum x^2 = 37\,746$ , where  $x$  is the time in minutes to mark an examination paper.
- Calculate unbiased estimates of the population mean and variance. (3 marks)
  - Stating the null hypothesis and alternative hypothesis, use a 10% significance level to test whether the examination board's estimated time is consistent with the data. (6 marks)
- b) For the data shown in the following table, test for independence between a person's ability in mathematics and his or her interest in statistics. Use the 0.01 level of significance.

INTEREST IN STATISTICS	ABILITY IN MATHEMATICS		
	LOW	AVERAGE	HIGH
LOW	63	42	15
AVERAGE	58	61	31
HIGH	14	47	29

(11 marks)

- B9. 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. (11 marks)
- b) A food nutritionist at NUST, however, suspects that there may be systematic difference 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. (9 marks)
- B10. Suppose you are interested in using past expenditure on development for various Sports Clubs. You obtain the following data by taking a random sample of clubs, where  $X$  is the amount spent on development (in ten million of dollars) 2 years ago and  $Y$  is the amount spent on development (in ten million of dollars) in the current year.

X: 3    5    2    8    1    2    2    4  
 Y: 5    8    3    11    2    2    4    5

- a) Draw a scatter plot of  $y$  against  $x$ . Fit the least squares straight line to these data and draw your fitted line on your scatter plot. (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, calculate the coefficient of determination,  $r^2$ , and interpret it. (8 marks)
- c) Predict  $y$  when  $x = 4$ . (2 marks).

END OF EXAMINATION QUESTION PAPER