UNIVERSITY OF MORATUWA

MSC/POSTGRADUATE DIPLOMA IN OPERATIONAL RESEARCH

MA(5001) INTRODUCTION TO STATISTICS
THREE HOURS **AUGUST 2007**

Answer **FIVE** questions and **NO MORE**.

ADDITIONAL MATERIAL:

Statistical Tables will be available.

INSTRUCTIONS TO CANDIDATES:

This paper contains 8 questions and 8 pages.

Answer **FIVE** questions and NO MORE.

This is a closed book examination.

This examination accounts for 70% of the module assessment

Assume reasonable values for any data not given in or with the examination paper. Clearly State such assumptions made on the script.

If you have any doubt as to the interpretation of the wording of a question, make your own decision, but clearly state it on the script.

Continued……

NO QUESTIONS

Question 1

(a) New York advertisement agency, indicates that matters of taste cannot be ignored in television advertising. Based on a mail survey of 3440 people, 40 % indicated that they found TV commercials to be in poor taste; 55 % say that they avoid products whose commercials were judged to be in poor taste; and, of this latter group, only 20 % ever complained to a TV station or an advertiser about their dissatisfaction.

(i) Find a 95 % confidence interval for the percentage of TV viewers who find TV commercials to be in poor taste.

(ii) Find a 95 % confidence interval for the percentage of TV viewers who avoid products which use TV commercials that they consider to be in poor taste. .

(iii) Find a 95 % confidence interval for the percentage of those who avoid products who have complained to the TV station or the advertiser about poor taste in a TV commercial.

(b) The administrators for a hospital wished to estimate the average number of days required for treatment of patients between the ages of 25 and 34. A random sample of 500 hospital patients between these ages produced a mean and standard deviation equal to 5.4 and 3.1 days, respectively. Construct a 95 % confidence interval for the mean length of stay for the population of patients from which the sample was drawn. Use a confidence coefficient of .95.

(a) The joint probability function of X and Y is tabulated below.

		Y			
		0	1	2	3
X	0	$\frac{10}{220}$	$\frac{40}{220}$	$\frac{30}{220}$	$\frac{4}{220}$
	1	$\frac{30}{220}$	$\frac{60}{220}$	$\frac{18}{220}$	0
	2	$\frac{15}{220}$	$\frac{12}{220}$	0	0
	3	$\frac{1}{220}$	0	0	0

(i) State the marginal probability functions of X and Y.

(ii) Find the expected values of X and Y.

(iii) Calculate the covariance of X and Y.

(iv) Are X and Y independent? If not, why?

(v) Calculate $P(XY \geq 2)$.

(b) Suppose that a radioactive particle is randomly located in a square with sides of unit length. That is, if two regions of equal area are considered, the particle is equally likely to be in either. Let Y_1 and Y_2 have the joint probability density function given by,

$$f(y_1, y_2) = \begin{cases} 1, & 0 \leq y_1 \leq 1,\ 0 \leq y_2 \leq 1, \\ 0, & \text{elsewhere} \end{cases}$$

(i) Sketch the probability density surface.

(ii) Find F(.2, .4).

(iii) $P(0.1 \leq y_1 \leq 0.3, 0 \leq y_2 \leq 0.5)$.

Question 2

In answering a question on a multiple choice test, a student either knows the answer or guess. If the student guesses, they will be correct with probability 1/5, as there arte five possible answers for each question. Assume that, on average, the student guesses the answer to 60% of multiple choice questions of this type.

(i) what is the probability that the student answered a question correctly,

(ii) Given that the student answered correctly, what is the probability that they actually knew the answer?

(b) At an agricultural station it was desired to test the effect of a given fertiliser on wheat production. To accomplish this, 24 plots of land having equal areas were chosen, half of these were treated with the fertiliser and the other half were untreated (control group). Otherwise the conditions were the same. The mean yield of wheat on the untreated plots was 131 kilograms with a standard deviation of 11 kilograms, while the mean yield on the treated plots was 139 kilograms with a standard deviation of 10 kilograms. Can we conclude that there is a significant improvement in wheat production because of the fertiliser if a significance level of 5% is used?

Question 3

(a) A manufacturer recorded the number of times that each of his ninety employees was late over the period of a year. The following table shows the data.

No of times late	0	1	2	3	4	5	6+
Frequency	23	29	20	10	4	4	0

Test whether it is reasonable to assume that the data follow a Poisson distribution

(i) with mean 1,

(ii) with mean estimated from the data.

(b) In a study of heart disease researchers classified a random sample of 356 people according to their socioeconomic status (high, middle or low) and their smoking habits (current smokers, former smokers, never smoked). The following table summarizes the data.

	Socioeconomic status		
Smoking	High	Middle	Low
Current	51	22	43
Former	92	21	28
Never	68	9	22

Calculate the table of expected frequencies under the hypothesis that socioeconomic status and smoking habits are independent. Carefully justify the formula you use.

Test the hypothesis that socioeconomic status and smoking habits are independent.

Question 4

(a) A manufacturer of electric motors tests insulation at a high temperature (250°C) and records the number of hours until the insulation fails. The data for 5 specimens are

300 324 372 372 444.

Engineering experience suggests that the logarithm (to base e) of the failure time will have a normal distribution. Take the logarithms of the 5 observations and find a 95% confidence interval for the mean of the log failure time for insulation of this type.

(b) A study was conducted to determine whether there is a linear relationship between breaking

strength y, of wooden beams and the specific gravity, x of the wood. Ten randomly selected beams of the same cross-sectional dimensions were stressed until they broke. The breaking strengths and the density of the wood are shown for each of the ten beams.

Beam	Specific Gravity, x	Strength, y
1	0.499	11.14
2	0.558	12.74
3	0.604	13.13
4	0.441	11.51
5	0.550	12.38
6	0.528	12.60
7	0.418	11.13
8	0.480	11.70
9	0.406	11.02
10	0.467	11.41

(i) Calculate the correlation coefficient between x and y.

(ii) Do the data provide sufficient evidence to indicate that the strength and specific gravity are independent?

(iii) Fit the model $y = \alpha + \beta x + e$

(iv)Test $H_0 : \beta = 0$ against the alternative hypotheses, $H_1 : \beta \neq 0$, using 5% level of significance.

Question 5

The continuous random variables X and Y have joint density function

$$f_{X,Y}(x,y) = \begin{cases} kx^2 & 0 < x < 2y, \ \ 0 < y < 1 \\ 0 & \text{otherwise} \end{cases}$$

(i) Sketch the region where $f_{X,Y}(x,y)$ is positive.

(ii) Find k.

(iii) Find the marginal density function of X and Y.

(iv) The conditional mean of X given Y =1/2.

(v) Find the conditional mean of Y given X = 1.

Question 6

A publisher is considering marketing a new IT magazine. From the past experience the publisher

expects the following profits, in thousands of pounds, under three possible levels of buyer response.

Buyer Response	Profit ($ thousands)
Good (R_1)	3000
Fair (R_2)	400
Poor (R_3)	-2500

If the publisher abandons the project now there will be no costs. The publisher estimates the probabilities of buyer response as

$P(R_1) = 0.35 \qquad P(R_2) = 0.20 \qquad P(R_3) = 0.45$

(i) Draw a suitable decision tree and advise the publisher on the best course of action to maximise her expected profit.

(ii) The publisher could also carry out market research. The market research will indicate either a favourable reaction (F) or an unfavourable one (U). From past experience the publisher estimates the following conditional probabilities:

$P(F \mid R_1) = 0.8, P(F \mid R_2) = 0.6, P(F \mid R_3) = 0.2.$

Find P(F), P(U), $P(R_i \mid F)$ and $P(R_i \mid U)$ for i = 1, 2, 3

(iii) Draw a new decision tree and advise the publisher the maximum she should be prepared to pay for the market research.

(iv) Comment on the use of expected profit as the decision criterion.

Question 7

(a)The logarithmic distribution with parameter $(0 \le \theta \le 1)$, has discrete probability density function $f(x) = \frac{k\theta^x}{x}$, $x = 1,2,3....$, and X is a random variable with this distribution.

(i) Find the constant k as function of θ

(ii) Show that the mean of this distribution is $\frac{k\theta}{1-\theta}$.

(iii)By finding $E[X(X-1)]$ or otherwise, show that $E[X^2]=\frac{k\theta}{(1-\theta)^2}$ and hence find the variance of X.

(b)Define the moment generating function $M_X(t)$ of a random variable.

(i) Assume that X_i $(i=1,\ 2,\ 3....n)$ are independent and identically distributed, each with moment generating function $M_X(t)$. Define $Y=a+b\sum_{i-}^{n}X_i$ show that the moment generating function of Y is $M_Y(t)=e^{at}[M(bt)]^n$

(ii) If X has an exp(2)distribution, find the moment generating function of a random variable X.

Question 8

(a) A major oil company is considering the inclusion of an additive for their petrol in order to increase the miles per liter obtainable by a certain type of car. Under test at the moment are two additives, Apex and Bemax which are available from different suppliers to test the performance of the two additives an independent agency tests five cars with Apex additive and six cars with Bemax additives, all cars being of the same make, typed and size The results obtained presented in the bellow.

Miles Per Litre						
Apex	10.4	10.7	10.2	10.5	9.8	
Bemax	10.7	11.3	9.9	10.8	10.1	10.4

Perform a test to discover whether there is a significant difference between the mean petrol consumption given by the two additives

(b) A cement manufacturer claimed that concrete prepared from his product would possess a relatively stable compressive strength and that the strength measured in kilograms per square centimeter would lie within a range of 40 kilograms per square centimeter. A sample of n = 10 measurements produced a mean and variance equal to $\bar{x}=312$ and s^2 = 195, respectively, Do these data present sufficient evidence to reject the manufacturer's claim? Use 5% level of significance.