SINGAPORE INSTITUTE OF MANAGEMENT

PRELIMINARY EXAM

PROGRAMME(S) : University of London Degree and Diploma Programmes
(Lead College: London School of Economics & Political Science)

SUBJECT : **04B STATISTICS 2**

DATE : Thursday, 5 March 2009

DURATION : 2 hrs

INSTRUCTIONS :-

DO NOT TURN OVER THIS QUESTION PAPER UNTIL YOU ARE TOLD TO DO SO.

Candidates should answer **all questions**

You are supplied with:

New Cambridge Statistical Tables

A hand held electronic calculator may be used when answering questions on this paper. The make and type of machine must be stated clearly on the front of the answer booklet.

Total number of pages : 7 (including this page)

1. a) For each one of the statements below say whether the statement is true or false explaining your answer.

    i. Mutually exclusive events with positive probability can not be independent.

    ii. If $P(A) + P(B) > 1$, A and B can not be independent.

    iii. If $P(A) + P(B) > 1$, A and B can not be mutually exclusive.

    iv. If T_1 and T_2 are unbiased estimators for a parameter θ, then $T_1 T_2$ is an unbiased estimator for θ^2.

    v. If T_1 and T_2 are unbiased estimators for a parameter θ, then $T_1 T_2$ is an unbiased estimator for θ^2 only if T_1 and T_2 are independent.

(10 marks)

b) An amateur gardener is trying to predict his water consumption (measured in litres) in terms of the air temperature (measured in ^0C) and the time he spends mowing the lawn (measured in hours). The following model was fitted:

$$y_i = a + bx_i + cz_i + \varepsilon_i,$$

where x_i represents the maximum temperature in day i, z_i represents the time he spent mowing the lawn in day i, y_i is his water consumption in day i and ε_i is the error which is normally distributed with mean 0 and unknown variance. The estimated values were -1.3 for a, 0.1 for b and 0.8 for c. Interpret the results.

(4 marks)

c) Suppose that you are given observations y_1, y_2, y_3 and y_4 that are such that

$$y_1 = \alpha + \beta + \varepsilon_1$$

$$y_2 = \beta + \varepsilon_2$$

$$y_3 = \alpha + \varepsilon_3$$

$$y_4 = \alpha - \beta + \varepsilon_4.$$

The variables $\varepsilon_i, \;\; i = 1, 2, 3, 4$ are normally distributed with mean 0. Find the least squares estimator for the parameter β and verify that it is unbiased.

(9 marks)

d) Let U and V be two independent and normally distributed random variables with mean 0 and variance 1. Find k that satisfies

$$P(kU - V > 6.198) = 0.025$$

(6 marks)

e) A man has a choice of three kinds of toothpaste. If he uses toothpaste A, there is a probability of 0.4 that he will develop a cavity in the next year. If he uses toothpaste B, the same probability is 0.5. If he uses toothpaste C, it is 0.3. If he gets no cavities during a particular year, he uses the same toothpaste the next year. Otherwise he chooses any of the other two toothpastes with equal probability. During year 1 he used toothpaste A.

i. What is the probability that he used toothpaste A in year 3?

(4 marks)

ii. Given that he used toothpaste A in year 3, what is the probability he developed no cavities in years 1, 2 and 3?

(7 marks)

2. The random variable X has a density function given by

$$f(x) = ax^2(x+1)$$

defined over the region $0 \le x \le 1$. Find a, $\Pr(X < 0.5 | X > 0.25)$, $E\left(\frac{1}{X}\right)$, $\mathrm{Var}\left(\frac{1}{X}\right)$ and $\mathrm{Cov}\left(\frac{1}{X}, \frac{1}{X+1}\right)$.

(19 marks)

3. A durian stall sells four varieties of durian and is open six days a week. The takings due to each variety of durian during each one of the six days of the same week were recorded. The total takings over the week measured in Singapore dollars for D99 durians were 844, for D24 durians 866, for D10 durians 838 and for D2 durians 849. The following is the calculated ANOVA table based on daily takings with some entries missing.

Source	degrees of freedom	sum of squares	mean square	F - value
Durian				
Day			27.54	
Error				
Total		345.96		

.

a) Complete the table using the information given above.

(7 marks)

b) Is there a significant difference between the daily takings due to each variety of durian?

(4 marks)

c) Construct a 90% confidence interval for the difference in daily takings between D24 and D10 durians? Would you say there is a difference?

(5 marks)

d) Construct simultaneous 90% confidence intervals for the difference in daily takings between D24 and D10 and the difference between D99 and D2.

(5 marks)

4. Consider two random variables X and Y. They both take the values -1, 0 and 1. The joint probabilities for each pair are given by the following table.

	$X=-1$	$X=0$	$X=1$
$Y=-1$	0	0.2	0.1
$Y=0$	0.1	0.2	0
$Y=1$	0.1	0.05	0.25

a) Calculate the marginal distributions and expected values of X and Y.

(6 marks)

b) Calculate $E(X|Y=1)$ and $E(X|X+Y=0)$.

(8 marks)

c) Let $W=\min(X,Y)$, Calculate $\text{Cov}(X,W)$.

(6 marks)

Formulae for Statistics

Discrete Distributions

Binomial Distribution

The probability of x successes in n trials is

$$\binom{n}{x} \pi^x (1-\pi)^{n-x}$$

for $x = 0, 1, \ldots, n$ The mean number of successes is $n\pi$ and the variance is $n\pi(1-\pi)$.

Poisson Distribution

The probability of x is

$$e^{-\mu}\frac{\mu^x}{x!}.$$

The mean number of successes is μ and the variance is μ.

Hypergeometric Distribution

The probability of x successes in a sample of size n from a population of size N with M successes is

$$\binom{M}{x}\binom{N-M}{n-x} \Big/ \binom{N}{n}.$$

The mean number of successes is nM/N and the variance is
$n(M/N)(1-M/N)(N-n)/(N-1)$.

Sample Quantities

Sample Variance	$s^2 = \sum(x_i - \bar{x})^2/(n-1) = (\sum x_i^2 - n\bar{x}^2)/(n-1)$
Sample Covariance	$\sum(x_i - \bar{x})(y_i - \bar{y})/(n-1) = (\sum x_i y_i - n\bar{x}\bar{y})/(n-1)$
Sample Correlation	$(\sum x_i y_i - n\bar{x}\bar{y})/\sqrt{(\sum y_i^2 - n\bar{y}^2)(\sum x_i^2 - n\bar{x}^2)}$

Inference

Variance of Sample Mean	σ^2/n
One-sample t statistic	$\sqrt{n}(\bar{x}-\mu)/s$ with $(n-1)$ degrees of freedom
Two-sample t statistic	$\dfrac{(\bar{x_1}-\bar{x_2})-(\mu_1-\mu_2)}{\sqrt{[1/n_1+1/n_2]\{[(n_1-1)s_1^2+(n_2-1)s_2^2]/(n_1+n_2-2)\}}}$
Variances for differences of binomial proportions	
Pooled	$\left[\dfrac{(n_1p_1+n_2p_2)}{(n_1+n_2)}\right]\left[1-\dfrac{(n_1p_1+n_2p_2)}{(n_1+n_2)}\right]\left[\dfrac{1}{n_1}+\dfrac{1}{n_2}\right]$
Separate	$p_1(1-p_1)/n_1+p_2(1-p_2)/n_2$
Least Squares	Estimates for $y=\alpha+\beta x$ fitted to (y_i,x_i) for $i=1,2,\ldots,n$ are $a=\bar{y}-b\bar{x}$ and $$b=\sum(x_i-\bar{x})(y_i-\bar{y})/\sum(x_i-\bar{x})^2.$$ The estimate of variance is $$[\sum(y_i-\bar{y})^2-b^2\sum(x_i-\bar{x})^2]/(n-2).$$ The variance of b is $\sigma^2/\sum(x_i-\bar{x})^2$
Chi-squared Statistic	$\sum$(Observed − Expected)^2/Expected, with degrees of freedom depending on the hypothesis tested.

END OF PAPER