NATIONAL UNIVERSITY OF SCIENCE AND TECHNOLOGY

B.COMM (HONOURS) DEGREE ACTUARIAL SCIENCE

ACTUARIAL STATISTICS (CIN 2211)

JUNE 2004 SECOND SEMESTER EXAMINATION

DURATION : 3 HOURS

Instructions to Candidates

1. Answer all questions

Requirements

1. Statistical Tables
2. Scientific Calculator

1. The numbers of policies with a certain insurance company held by a random sample of 200 policyholders are given in the following frequency distribution:

Number of policies	1	2	3	4	5
Number of policyholders	125	57	13	3	2

Find the mean and standard deviation of the number of policies held by a single policyholder. **[6 marks]**

2. Consider $x_1, x_2, \ldots x_n$ independent and identically distributed random variables from $U(O, \theta)$ distribution.

   (a) Find the probability density function (pdf) of $x_{(n)}$ (where $x_{(n)}$ is the largest of the x_i's) **[3 marks]**

   (b) Show that the MLE for θ is biased and find its variance. **[5 marks]**

   (c) Find the MME for θ and show that both the MME and $\frac{n+1}{n}x_{(n)}$ are consistent estimators of θ. **[6 marks]**

   (d) Which of these two estimators is more efficient? **[3 marks]**

3. KBC insurance company is interested in the variation in the amounts paid out on claims under a certain class of policy. A random sample of 10 such claims (in $) is taken and the sample standard deviation is s = 55. Assuming that the claim amounts are normally distributed, find a 95% confidence interval for the standard deviation of the claim amount distribution. **[7 marks]**

4. Let $x_1, x_2, x_3 \ldots\ldots\ldots x_{10}$ be independent random variables, each with Poisson distribution with mean1. Let $S = x_1 + x_2 + \ldots. + x_{10}$.

   (a) Write down the exact distribution of S and use it to obtain $\Pr(8 \leq S \leq 12)$ **[4 marks]**

   (b) Use the Central Limit Theorem to obtain an approximation to $\Pr(8 \leq S \leq 12)$ **[5 marks]**

5. A random sample of size 16 is taken from a normal distribution with unknown mean μ and variance $\sigma^2 = 100$. It is required to test $H_o : \mu = 50$ vs $H_i : \mu > 50$. The value of the sample mean of these data is $\bar{x} = 55$. Find the p-value of this observed sample mean. **[5 marks]**

6. The following data are sizes of claims (in $) for a random sample of 20 recent claims submitted to an insurance company:

   174, 214, 264, 298, 335, 368, 381, 395, 402, 442, 487, 490, 564, 644, 686, 807, 1092, 1328, 1655, 2272. Calculate the median and interquartile range for this sample of claim sizes. **[5 marks]**

7. A random sample x_1, x_2, x_3 is taken from $N(\mu, 1)$ where μ is unknown. It is required to estimate $\theta = \mu^3$.

   (a) State (without deriving from first principles) the maximum likelihood estimator, $\hat{\theta} = \theta$. **[2 marks]**

   (b) Find an expression for the mean square error of the estimator

   $\theta^* = x_1, x_2, x_3$ as an estimator of θ. **[4 marks]**

8. The number of claims N made by a policyholder in a year has the following distribution

Number of claims	0	1	2
Probability	0.4	0.4	0.2

   While the size of a single claim x has the following distribution:

Size of claim	1	2
Probability	0.7	0.3

It can be assumed that the occurrence of a claim and the size of a claim are independent. Furthermore, claims made in any year are also independent. Let S be the total claim size in a year.

(a) Find the probability function of S and hence find E(S) **[5 marks]**

(b) Find E[S/N = n] for n = 0, 1, 2 and hence use the formula $E(S) = \sum Pr(N = n)\, E(S/N = n)$ to calculate E(S). **[3 marks]**

(c) Use the formula $E(S) = \sum Pr(N = n)\, E(S/N = n)$ to find an expression for E(s) in terms of E(N) and E(x). **[2 marks]**

9. The number of claims arising under a certain type of policy is thought to have a Poisson distribution with mean $\lambda = 1$. A random sample of n such policies is examined and a total r claims is found to have been on these policies.

   Use the Neyman-Pearson Lenna to find the form of the best test of $H_o : \lambda = 1$ vs $H_i : \lambda = \lambda_1$ where $\lambda_1 > 1$. **[4 marks]**

10. The university transport section asks a random sample of 800 of its workers (who use university bus) whether they are for or against a ban on smoking in all areas of its buses. Each person was classified by sex and by whether or not he/she was a smoker. The results of this sample survey are detailed below:

	For ban		**Against ban**	
	MEN	**WOMEN**	**MEN**	**WOMEN**
Smokers	177	83	123	67
Non-smokers	73	137	77	63

Investigate whether or not there is an association between smoking habit and opinion on the ban on smoking among:

(a) male workers **[3 marks]**

(b) female workers **[3 marks]**

(c) all workers **[3 marks]**

(d) comment briefly on your results (a), (b) and (c). **[2 marks]**

11. A random sample of 12 Tonga families with father and an adult son both living was taken. The heights of father and son (in centimeter) were as follows:

Father's height, x	67.1	64.2	59.8	72.3	62.5	71.3	68.8	70.3	65.1	73.8	67.9	66.2
Son's height, y	67.2	66.2	63.4	70.4	65.7	71.1	68.1	69.2	65.6	71.2	67.3	67.1

For these data $\sum_{i=1}^{12} x_i = 809.3$, $\sum_{i=1}^{12} x_i^2 = 54770.15$

$\sum_{i=1}^{12} y_i = 812.5$, $\sum_{i=1}^{12} y^2 = 55076.45$ and $\sum_{i=1}^{12} x_i y_i = 54902.93$

(a) Fit the linear regression model to these data i.e.
$y = \alpha + \beta(x - \bar{x}) + \varepsilon = a + \beta x + \varepsilon$ **[4 marks]**

(b) Calculate the coefficient of determination R^2. **[3 marks]**

(c) Test $H_o : \beta = 1$ vs $H_1 : \beta < 1$ **[4 marks]**

(d) Calculate a two-sided 95% confidence interval for the mean height of sons whose father's height is 69 centimeters. **[5 marks]**

12. Four groups of sales people for an Insurance Company were subjected to different sales training programmes. At the end of the training programs each person was randomly assigned a sales area from a group of sales areas that were judged to have equivalent sales potentials. The number of sales made by each person in each of the four groups of sales people during the first week after completing the training programme is given in the table below:

Training Group

1	2	3	4
65	75	59	94
87	69	78	89
73	83	67	80
79	81	62	88
81	72	83	
69	79	76	
	90		

For these data; $Y_{..} = 1779$, $\sum Y_{ij}^2 = 139511$

(a) Do the data present sufficient evidence to indicate a difference in the mean achievement for the four training programmes? **[4 marks]**

(b) What statistical assumptions are implicit in the analysis of variance? **[2 marks]**

(c) Find 95% confidence interval for the difference in mean sales for training programmes 1 and 4, using information from all the data. **[3 marks]**

END OF EXAMINATION PAPER!!