INSTITUTE OF ACTUARIES OF INDIA

EXAMINATIONS

21st May 2012

Subject CT3 – Probability & Mathematical Statistics

Time allowed: Three Hours (15.00 – 18.00)

Total Marks: 100

INSTRUCTIONS TO THE CANDIDATES

1. ***Please read the instructions on the front page of answer booklet and instructions to examinees sent along with hall ticket carefully and follow without exception***

2. ***Mark allocations are shown in brackets.***

3. ***Attempt all questions, beginning your answer to each question on a separate sheet. However, answers to objective type questions could be written on the same sheet.***

4. ***In addition to this paper you will be provided with graph paper, if required.***

5. ***Please check if you have received complete Question Paper and no page is missing. If so kindly get new set of Question Paper from the Invigilator***

AT THE END OF THE EXAMINATION
Please return your answer book and this question paper to the supervisor separately.

Q. 1) A cricket coach records the number of runs the players in his squad scored in a tournament. Each player got a chance to bat at least once. He presents the data in a stem and leaf diagram:

	KEY: **2 | 7** means 27 runs scored			
0	1	1	2	7
1	2	5	5	
2	3	7		
3	6			
4	0			
5	0	9		

a) What is the range of the data? (1)

b) What is the median number of runs scored? (1)

c) Compute the mean number of runs scored. (2)

[4]

Q. 2) The observed mean (and standard deviation) of the number of claims and the individual losses over a given period are 7.6 (3.2) and 197,742 (52,414), respectively. Assume the variable for number of losses is independent of the variable for individual loss sizes. Determine the mean and standard deviation of aggregate claims.

[5]

Q. 3) The random variables ϵ_i , $i = 1, 2, 3$ are independent and normally distributed with mean 0 and variance 1. Let α be an unknown parameter.

Suppose you are given observations y_1 and y_2 such that:

- $y_1 = \alpha + \epsilon_1$
- $y_2 = 3\alpha + \epsilon_2$

a) Write down the regression model. (1)

b) Derive the expression for the least square estimator $\hat{\alpha}$ of α. (3)

You are given that the values of y_1 and y_2 are 0.6 and 1.8 respectively.

c) Calculate the value of the estimator $\hat{\alpha}$. (1)

d) Suppose now $y_3 = \alpha + \epsilon_3$. Find $E(\widehat{y_3}), Var(\widehat{y_3})$ and hence provide a 95% prediction interval for y_3 (7)

[12]

Q. 4) Based on a Normal random sample of size 100, a 90% confidence interval for the population mean turned out to be (20, 40). Find a 95% confidence interval for the population mean based on this information.

[5]

Q. 5) Let N be a Poisson random variable with mean λ.
Define a random variable Y: $Y = \alpha + \beta N$ where α, β > 0 are given constants.

An actuarial student decided to construct a random variable X based on Y by playing the following game:

- He tosses an unbiased coin
- If it turns up 'Head', he assigns X = Y
- If it turns up 'Tail', he assigns X = 0

a) Show that the moment generating function $M_X(t)$ of X can be expressed in terms of the moment generating function $M_Y(t)$ of Y as below:

$$M_X(t) = \frac{1}{2}.[1 + M_Y(t)]$$ (2)

b) Hence or otherwise, show that the probability distribution of X can be expressed as:

Value	Probability
0	0.5
$\alpha + \beta.(x-1)$	$\frac{0.5\, e^{-\lambda}\lambda^{x-1}}{(x-1)\,!}$ for $x = 1, 2, 3 \ldots$

(7)

c) Consider two such independent random variables X_1 and X_2 (constructed similar to X above) with the following values of the parameters:

Variables	α	β	λ
X_1	1	2	1.0
X_2	2	3	1.5

Compute $P(X_1 + X_2 > 5)$. (8)

[Hint: $P(X_1 + X_2 = c) = \sum_{\{(x_1,x_2):\, x_1+x_2 = c\}} P(X_1 = x_1).P(X_2 = x_2)$]

[17]

Q. 6) A university runs a 3-year B.Sc degree course in Statistics. The course is divided over 6 semesters each consisting of 5 credit papers over the three year period. Each credit paper is assessed on a maximum possible 100 marks and is recorded as integers.

At the end of the course, the university ranks the students based on a measure called "grade point average" which is the average of marks obtained over all credit papers examined over 3 years.

Assume that in each subject, the instructor makes an error of quantum k in awarding marks with probability $\frac{1}{20.|k|}$ where $k = \pm1, \pm2, \pm3, \pm4, \pm5$. Assume that these errors occur independently.

a) Show that the probability of no error is $\frac{463}{600}$. (2)

b) State the approximate distribution of quantum of error in a given student's final grade point average using the Central Limit Theorem. (3)

c) Hence show that there is only a 17.7% chance that his final grade point average is accurate to within ±0.05. (3)

[8]

Q. 7) Suppose that $X_1, X_2, \ldots, X_n$ are independent and identically distributed Poisson(λ) random variables.

a) Find the maximum likelihood estimator of λ. (4)

b) Suppose that rather than observing the random variables precisely, only the events "$X_i = 0$" *or* "$X_i > 0$" for i =1, 2 … n are observed. Find the maximum likelihood estimator of λ under the new observation scheme. (6)

[10]

Q. 8) Let X_1 and X_2 constitute a random sample of size 2 from the $N(\theta, 1)$population.

For testing H_0: θ = 0 versus H_1: θ > 0, we have two competing tests:

Test 1: $Reject\ H_0\ if\ X_1 > 0.95$

Test 2: $Reject\ H_0\ if\ X_1 + X_2 > C$

a) Find the value of C so that Test 2 has the same P (Type I error) as that of Test 1. (2)

b) Compute the P (Type II error) of each test for a given value of $\theta = \theta_1 > 0$. (3)

c) Comment on your results as obtained in part (b)? (1)

[6]

Q. 9) Suppose 2,000 finished products both from Factory A and Factory B were chosen at random by the CEO of the Company and verified if the same have any defects. Following are the results:

	Factory A	**Factory B**
Non-defective	1,816	1,986
Defective	184	14

Perform a chi-square test on this contingency table to show that there is overwhelming evidence against the hypothesis that there is no association between the factory and whether or not the product is defective. State your level of significance.

[5]

Q. 10) An actuarial student fits the following linear regression model to a given data:

$$y_i = \alpha + \beta x_i + \epsilon_i \text{for } i = 1, 2, \dots n$$

Here ε_i's are independent, identically distributed random variables, each with a normal distribution with mean 0 and unknown varianceσ^2.

The following information is available:

- n = 7
- $\sum(x_i - \bar{x})^2 = 280{,}000$
- $\sum(x_i - \bar{x})(y_i - \bar{y}) = 16{,}500$
- 95% confidence interval for β : $(0.030,\ 0.088)$

Calculate what portion of the total variability of the responses is explained by the model.

[5]

Q. 11) To measure the effect of a fitness campaign, a gym instructor devised two types of sampling design:

Design 1: Here he randomly sampled 5 members before the campaign and measured their weights (X_1), and another 5 after the campaign (X_2). The results (along with some summary statistics) are as follows:

	Weights					**ΣX_k**	**$\Sigma X_k * X_k$**	**$\Sigma X_1 * X_2$**
Before: X_1	168	195	155	183	169	870	152,324	148,265
After: X_2	183	177	148	162	180	850	145,366	

Design 2: Here he decided to measure the weights of the same people after(X_3), as before the campaign. The results (along with some summary statistics) are as follows:

	Weights					**ΣX_k**	**$\Sigma X_k * X_k$**	**$\Sigma X_1 * X_3$**
Before: X_1	168	195	155	183	169	870	152,324	149,032
After: X_3	160	197	150	180	163	850	145,878	

a) Is it appropriate to assume that under each of the two respective sampling design schemes the two samples of data obtained constitute two independent random samples? Explain. (1)

b) Calculate a 95% confidence interval for the mean weight loss during the campaign on the basis of results obtained under each of the two respective sampling design schemes. (8)

c) What can you conclude about the effectiveness of the campaign from the two confidence intervals obtained in part (b)? Comment on the relative width of the two intervals as well? (2)

[11]

Q. 12) Many businesses have music piped into the work areas to improve the environment. At a company an experiment is performed to compare different types of music. Three types of music – country, rock and classical – are tried, each on four randomly selected days. Each day the productivity, measured by the number of items produced, is recorded. The results appear below:

Music Type	**Productivity (y)**				**Σy**	**Σy2**
Country	857	801	795	842	3,295	2,717,039
Rock	791	753	781	776	3,101	2,404,827
Classical	824	847	881	865	3,417	2,920,771
					9,813	8,042,637

[Draw all your statistical inferences at 5% significance level]

a) Perform an analysis of variance to show that the mean number of items produced differs for at least two of the three types of music. (5)

b) Show that the mean number of items produced in rock music is significantly worse than those produced in the other two. (4)

c) Show that it is statistically difficult to ascertain which music has the best effect in terms of the mean number of items produced. (3)

[12]

****