INSTITUTE OF ACTUARIES OF INDIA

EXAMINATIONS

10th May 2010

Subject CT3 – Probability & Mathematical Statistics

Time allowed: Three Hours (15.00 – 18.00 Hrs)

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.*

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

Q. 1) The following amounts are the sizes of claims (in Rs.) on house insurance policies for a certain type of repair.

199, 240, 215, 209, 230, 210, 218, 215, 203, 210, 218, 201, 206, 216, 212

(a) Two statisticians, A and B find the inter quartile range as 10.75 and 12.00 respectively. Comment on whether statistician A or B or both are correct.   (3)

(b) Determine the sample mean and sample variance of the data.   (3)

(c) Assuming the data to be normally distributed, find the probability of a claim amount exceeding Rs. 230   (2)

(d) Empirically find the probability of a claim amount exceeding Rs. 230 and hence comment on your observation when compared with the answer in (c)   (1)

[9]

Q. 2) The following is true about the masses of elephants from a sample of 12:
5 elephants weigh 900 kilograms each; 4 elephants weigh 1,000 kilograms each; 3 elephants weigh 1,396 kilograms each.

Show that the empirical skewness coefficient for this sample is 1   **[3]**

Q. 3) A test to detect H1N1 virus shows 5% positive result for those not having H1N1 and 100% positive results for those having H1N1. The probability of occurrence of H1N1 is (1/1000).

What is the probability that a person found to have a positive result actually has H1N1 assuming that you know nothing about the person's symptoms?   **[4]**

Q.4) Let $Y = e^{2X} + 5$.
Find $F_Y(y)$, expressed in terms of $F_X[g(y)]$, where $g(y)$ is some function solely of the variable y.   **[2]**

Q.5) Show that $C_X''(0) = \text{Var}[X]$, where $C_X(t)$ is the cumulant generating function of random variable X and 'denotes differentiating with respect to t.   **[4]**

Q.6) There are two continuous probability distributions:
A is an exponential distribution with mean $\theta = 7$
B is a distribution that is uniform on the interval from 0 to 8, and thereafter proportional to A.
Show that the probability of a random variable that follows distribution B and lying between 0 and 8 is 8/15   **[5]**

Q. 7)   You are analyzing population data with the following probabilities:

$Pr(X = 2) = 0.23$
$Pr(X = 5) = 0.34$
$Pr(X = 7) = 0.43$.

A sample of this population is taken. A method of moments procedure is used to estimate the mean μ using the sample mean $\overline{X}$ and the population variance σ^2 by the statistic ${S'_n}^2 = \frac{\sum(X_i - \overline{X})^2}{n}$, rather than using sample estimator S_n^2, where n is the sample size.

Find the bias of ${S'_n}^2$, when n = 13.   **[3]**

Q. 8)   You have the following shifted exponential distribution such that

$f(x) = \frac{1}{\theta}\, e^{-\frac{(x-\beta)}{\theta}}$, for $\beta < x$

(a) Determine the $E(X)$ and $E(X^2)$ of the above distribution. Hence, or otherwise show that s.d. of X is θ.   (5)

A sample of 50 values is given, such that $\sum x_i = 8{,}200$ and $\sum x_i^2 = 2{,}000{,}000$. You are modeling the data using the above shifted exponential distribution.

(b) Use a method of moments procedure to estimate the value of β and θ.   (2)
                                                                                                        [7]

Q. 9)   A single round of a game is played as follows:

You roll a fair 8-sided dice with faces numbered 1 through 8, whose outcome determines the value of N.

Based on the above outcome of N, you then roll fair 8-sided dice of the same sort N times and receive an amount of money equal to the sum of their faces.

Find the variance of your payoff in each individual round.   **[4]**

Q. 10) Suppose that the number of accidents an individual insured suffers in a year follows a Poisson distribution with mean 0.1. The magnitude of loss from each accident has mean 500 and standard deviation 1,000. Frequency and magnitude of losses are independent.

At least how many insureds must the company have in order that the probability of losses exceeding 1.5 times the expected amount is no greater than 0.05?

Use a normal approximation and the central limit theorem to obtain your answer.   **[5]**

Q. 11) The random variable X has the exponential distribution with mean θ.

Calculate the mean-squared error of X^2 as an estimator of θ^2.   **[3]**

Q.12) Claim frequency (N) follows a negative binomial distribution with mean 5,667 and variance 46,090. Claim severity (X) has the following probabilities:

Pr(X = 350) = 0.4;
Pr(X = 460) = 0.3;
Pr(X = 900) = 0.2;
Pr(X = 3400) = 0.1.

Next year, severities are expected to decrease by 40%, and a deductible of 200 will be imposed for each claim (i.e. claims will be payable after deducting 200).

Find the insurance company's expected payout under next year's conditions.   **[3]**

Q.13) For a group of policies, losses follow the distribution function
$$F(x) = 1 - \frac{\theta}{x}, \quad \theta < x < \infty$$
A sample of 20 losses resulted in the following:

Interval	Number of Losses
x <= 10	9
10 < x <= 25	6
x > 25	5

a) Show that the maximum likelihood estimate (MLE) of θ is 5.5   (6)

b) Find the Cramer-Rao lower bound (CRlb) using MLE of θ   (2)

c) Use the asymptotic distribution for the MLE of θ to obtain 95% confidence interval for θ   (2)

[10]

Q.14) 1,000 workers insured were observed for one year. The number of work days missed by them in a year is given below:

Number of days work missed	Number of Workers
0	818
1	153
2	25
3 or more	4
Total	1,000

The total number of days missed by all the workers is 230

The chi-square goodness-of-fit test is used to test the hypothesis that the number of work days missed follows a Poisson distribution, where the Poisson parameter is estimated by the average number of work days missed.

Determine the results of the test at 5% level of significance (it is not required to combine any cell due to less observations in any cell)

[5]

Q. 15) In an investigation into the comparison of claim amounts between three different regions A, B and C, a random sample of 10 independent claim amounts was taken from each region and an analysis of variance was performed. The resulting ANOVA table is given below with some entries omitted:

Source of Variation	d.f. (degree of Freedom)	SS (sum of Squares)	MSS (Mean Sum of Squares)
Between Regions	*	4,352.8	*
Residual	*	*	*
Total	*	10,345.6	

 (i) Calculate the missing entries in this table and perform the appropriate F-test to determine whether there are significant differences between the mean claim amounts for the three regions. (3)

(ii) The three sample means were given by:

Region	Sample Mean
A	155.48
B	164.68
C	120.90

Given that it was of particular interest to compare regions A and B, calculate a 95% confidence interval for the difference in their means and comment on your answer in the light of the F-test performed in part (i). (4)

[7]

Q.16) An actuary recorded the length of time, y minutes, taken to travel to office when leaving home x minutes after 7 am on seven selected mornings. The results are as follows:

x	0	10	20	30	40	50	60
y	16	27	28	39	39	48	51

(a) Plot the data on a scatter diagram.   (1)

(b) (i) Calculate the equation of the least squares regression line of y on x, writing your answer in the form y = a + bx.   (4)

    (ii) Draw the regression line on your scatter diagram.   (2)

(c) (i) Calculate a two-sided 95% confidence interval for b, the slope of the regression line.   (3)

    (ii) Hence or otherwise, test the hypothesis $H_0 : b = 1$ vs $H_1 : b \neq 1$ at 5% level of significance.   (1)

(d) The actuary needs to arrive at office no later than 8:40 am. The number of minutes by which the actuary arrives early at office, when leaving home x minutes after 7 am, is denoted by z.

    (i) Deduce that z can be estimated by, z = 81.5 – 1.564x   (2)

    (ii) Hence estimate, to the nearest minute, the latest time that the actuary can leave home without then arriving late at office.   (1)

          [14]

Q.17) **(a)** I.Q test were administered to 6 persons before and after the training. The results are given below:

Candidate	I.Q before training	I.Q after training
1	110	115
2	105	117
3	111	109
4	125	124
5	120	122
6	115	114

Test whether there is any change in I.Q. after the training programme at 5% level of significance.   (5)

(b) A sample of 52 pairs on a bivariate normal distribution gave correlation coefficient equal to 0.7.

Obtain a two-sided confidence interval for population correlation coefficient, ρ with 95 % confidence coefficient $\hspace{5cm}$ (4)

(c) X and Y are two Random variables having the joint density function

$f(x, y) = \frac{1}{27}(2x + y),$ where X and Y can assume only the integer values 0, 1 and 2. Prepare a table for the conditional distribution of X given Y = y, for all values of x and y.

$\hspace{13cm}$ (3)

$\hspace{13cm}$ **[12]**

****