INSTITUTE OF ACTUARIES OF INDIA

EXAMINATIONS

13th May 2008

Subject CT3 – Probability and Mathematical Statistics

Time allowed: Three Hours (10.00 – 13.00 Hrs)

Total Marks: 100

INSTRUCTIONS TO THE CANDIDATES

1. Do not write your name anywhere on the answer sheet/s. You have only to write your Candidate's Number on each answer sheet/s.
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. Fasten your answer sheets together in numerical order of questions. This, you may complete immediately after expiry of the examination time.
5. In addition to this paper you should have available graph paper, Actuarial Tables and an electronic calculator.

Professional Conduct:

"It is brought to your notice that in accordance with provisions contained in the Professional Conduct Standards, If any candidate is found copying or involved in any other form of malpractice, during or in connection with the examination, Disciplinary action will be taken against the candidate which may include expulsion or suspension from the membership of IAI."

Candidates are advised that a reasonable standard of handwriting legibility is expected by the examiners and that candidates may be penalized if undue effort is required by the examiners to interpret scripts.

AT THE END OF THE EXAMINATION

Please return your answer scripts and this question paper to the supervisor separately.

Q1) Consider the following sample of number of motor accidents reported to an insurance company in the 12 months of a year

179 121 116 125 119 136 149 239 131 123 113 134

(a) Calculate the mean and median of the above data. (2)

(b) Trimmed-mean is defined as the mean of the sample after removing the first few and last few observations from the sorted data. Calculate the trimmed mean after removing the least two and the largest two sample points. (2)

(c) Comment on the effect of the likely outliers in the data on the mean, median and trimmed-mean calculated above. (2)

(d) If the above sample is from a Normal distribution with variance 1300, what is the probability that the sample variance (S^2) is more than 1300? (2)

[8]

Q2) For the events A_1, A_2 and A_3, let $P(A_2/A_3) = 0.6$, $P(A_3/A_2) = 0.8$ and $P(A_1/A_2 \cap A_3) = 0.4$. Compute $P(A_1 \cap A_3/A_2)$ **[2]**

Q3) Two dice are tossed. Define the events A,B and C as follows.

A = {second die shows 1,2 or 5}
B = {second dies shows 4,5 or 6}
C = {the sum of the number on the faces of the two dice is 9}.
Are A,B and C mutually independent? **[4]**

Q4) Find the *pgf* of the distribution having *pmf*

$$p(x) = \frac{1}{2} pq^{|x|-1} \quad ; x=\pm 1, \pm 2\ldots, \quad 0<p,q<1, \quad p+q=1$$

[3]

Q5) In a certain chemical plant, the concentration of pollutants (X-in parts per million) has lognormal distribution with parameter μ=3.2 and σ=1

(a) write the *pdf* of X (1)

(b) compute the mean and variance of X (1)

(c) compute the probability that the concentration exceeds 8 parts per million (2)

[4]

Q6) Out of the 90 tosses of a coin, 50 tosses turn out to be heads.

(a) Let N denote the total number of heads in 90 tosses, what is the most suitable distribution of N? Estimate the mean and variance of N. (2)

(b) Find out the approximate probability that $N > 50$, assuming that the coin is a fair coin (2)

(c) Let the distribution of N be Binomial (90, p). Test the hypothesis that H_o: $p = 0.5$ *vs* H_1: p>0.5 at the significance level of 5% using the probability value calculated in (b) above. (1)

(d) For what values of N in the above test the null hypothesis would have been rejected? (2)

[7]

Q7) Let $X_1, X_2, \ldots, X_{20}$ denote a random sample of size 20 from $U(0,1)$. Let $Y = X_1 + X_2 + \ldots + X_{20}$
Using the central limit theorem, compute

(a) $P(Y \le 9.1)$ (1)

(b) $P(8.5 \le Y \le 11.7)$ (2)

[3]

Q8) In a railway station, passenger cabs wait until they have either acquired four passengers or a period of ten members has passed since the first passenger stepped into the cab. Passengers arrive according to a Poisson process with an average of one passenger every three minutes.

(a) You are the first passenger to get into a cab. What is the probability that you will have to wait ten minutes before the cab gets underway? (3)

(b) You are the first passenger to get into a cab and you have been waiting there for five minutes. At this stage two other passengers have entered the cab. What is the probability that you will have to wait another five minutes before the cab gets underway? (2)

[5]

Q9) Let $X_1, X_2, \ldots, X_n$ be a random sample from the following density function

$$f(x;\theta) = \frac{kx}{\theta^2}; \quad 0 < x < \theta,\ \theta > 0$$

(a) Find k such that above is a valid density function (2)

(b) Find the MLE of θ, for the given sample (3)

[5]

Q10) When Ramesh was appointed as the laboratory assistant on 1st Jan 2008 to observe the life time of mice, there were 10 mice in the laboratory. His assignment was to observe the life time of the mice till 100 weeks and then estimate the expected remaining life time (in weeks) of mice as at 1st Jan 2008. 7 Mice died within the 100 weeks period at the following times (in weeks)

23, 27, 39, 52, 68, 89, 95

and 3 mice were alive at the end of 100th week. Assuming that the future life time (as at 1st Jan 2008) follows Exp(λ) with density function

$$f(x,\lambda) = \lambda e^{-\lambda x}; \lambda > 0, x > 0:$$

(a) Write down the likelihood function for the sample of 10 life times that Ramesh observed. (2)

(b) Compute the MLE of λ based on this likelihood. (3)

(c) What is the asymptotic variance of the MLE? (2)

(d) Write down the formula for 95% confidence interval for λ and construct it using the MLE in (b) and the asymptotic variance in (c) (2)

Ganesh (Ramesh's boss) is the laboratory in-charge. He knows that the experiment actually started 25 weeks before 1st Jan 2008 with 15 mice. 5 mice had died before 1st Jan 2008 (Ramesh didn't know about it). Even Ganesh did not have the exact weeks in which these 5 mice died. He only knows that they had died before 1st Jan 2008. Based on this new information Ramesh wanted to correct the likelihood function.

(e) What is the correct likelihood function for the life time (starting 25 weeks before 1st Jan 2008) of 15 mice? (3)

[12]

Q11) A study was done to find if the students who are good in high school, carry on to do well also in the college. The high school grade (X) and college grade (Y) of 15 randomly chosen students are given below.

S.No.	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
X	2.0	3.4	3.7	1.5	3.3	0.3	0.4	2.0	2.0	2.1	2.1	1.3	1.5	3.1	2.1
Y	2.0	2.6	3.8	1.1	3.0	0.1	1.4	1.5	1.4	4.0	1.5	1.3	1.9	3.1	1.9

(a) Draw a scatter plot of the above data with high school grades on X-axis and college grades on Y-axis. (3)

(b) Estimate the slope and intercept parameters of the linear regression. Also calculate the sample correlation coefficient between the high school grades and the college grades. (3)

(c) Can it be concluded that, in general, the expected performance in college is the same as that in high school? (2)

(d) Estimate the college grade of a student who has a grade of 2.1 in high school. Comment on the college grade of student number 10 in the above data. (3)

[11]

Q12) A farmer is testing the effect of 3 different fertilizers A, B and C on the level of productivity (in tons per acre). He conducts an experiment for one-way ANOVA analysis (assume usual notations) and observes the following

Fertilizer	Number of fields (n_i)	Average Productivity($\overline{Y_{i.}}$)
A	8	100
B	10	110
C	9	95

The corrected error (Residual) sum of squares SSE is 1075.

(a) Find out the estimates of μ, τ_A, τ_B, τ_C and σ^2 (3)

(b) Calculate sum of squares due to fertilizers and corrected Total SS (2)

(c) Test the null hypothesis that the all the fertilizers have the same effect on productivity. (3)

(d) Calculate the 95% confidence interval for $\mu + \tau_A$ (2)

[10]

Q13) A training manager wishes to see if there has been any alteration in the ability of his trainees after they have been on a course. The trainees take an aptitude test before they start the course and on equivalent one after they have completed it. The scores are given below.

Before Training (X)	42	35	37	46	53	38	44	40	43
After Training (Y)	47	28	26	54	42	17	44	31	44

(a) Compute the Pearson's correlation coefficient between the scores before and after training and test the significance of the calculated correlation coefficient. (4)

(b) Assume that the populations are independent test whether there is any significant difference between the variances of the scores before and after training. (3)

(c) Obtain 90% confidence interval for the ratio of population variances of scores before and after training. (2)

(d) In view of the conclusions in a),examine whether training has increased their aptitude (at 5% significant level) (3)

(e) Obtain 95% confidence interval for the difference in the population mean scores before and after training. (2)

[14]

Q14) The relationship between household expenditure*(X)* and net income of households*(Y)* in a certain metropolitan city is given by the joint *pdf*

$$f(x, y) = k(x-10)(y-10) \quad ; \quad 10 < x < y < 30$$
$$= 0 \textit{ otherwise}$$

(a) What is the expected value of the household expenditure X of a randomly chosen household given that the income of the household is Y? (4)

(b) What is the probability that the household expenditure of a household is more than 20 given that the income of the household is 25? (3)

[7]

Q15) Let the random variable X follow Binomial distribution with parameters n and ½ and let the compounding distribution of the parameter n be $f(n) = \frac{n}{3}$; $n = 1,2$.

Find the unconditional distribution of X and its mean and variance. **[5]**

****