INSTITUTE OF ACTUARIES OF INDIA

EXAMINATIONS

14th March 2018

Subject CT3 – Probability & Mathematical Statistics

Time allowed: Three Hours (10.30 – 13.30 Hours)

Total Marks: 100

INSTRUCTIONS TO THE CANDIDATES

1. ***Please read the instructions inside the cover 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. ***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) Define the following for a discrete random variable X.

i) The kth moment. (0.5)

ii) The kth moment about α. (0.5)

iii) The kth central moment. (0.5)

iv) The coefficient of skewness. (0.5)
[2]

Q. 2) The mean and standard deviation of marks scored by students in an examination of two years are as follows:

	Number of students	Mean	Standard deviation
Year 1	190	53	19
Year 2	175	57	15

Calculate the overall mean and standard deviation of the combined sample. **[5]**

Q. 3) Let X and Y be two independent random variables following N $(15, 5^2)$ and N $(12, 10^2)$ respectively. Define $Z = X - Y + 5$.

i) Determine the distribution of Z and its mean and variance. (2)

ii) Calculate the probability that a sample of size 5 drawn from Z has mean greater than 10 and standard deviation greater than 20. (5)
[7]

Q. 4) A newly established life insurance company is analyzing the experience of policy withdrawal by the policyholders of a portfolio of 10,000 policies, on basis of the channel through which the policy was sold. The table below shows the split of the number of policies sold by the respective channels and the probability of a policyholder withdrawing the policy in a year.

Channel	Agency	Bank	Online
Probability (p)	0.05	0.08	0.14
Number of policies (n)	2000	3500	4500

It can be assumed that the withdrawal by any individual policyholder during any year is independent of withdrawal by other policyholders of same or different channel.

i) Calculate the probability that a randomly selected policyholder will withdraw in a particular year. (2)

ii) Calculate the probability that a randomly selected policyholder will withdraw in a particular year given that the policy was not sold through online channel. (2)

iii) Calculate the probability that a randomly selected policy was sold by Bank given that the policyholder withdrew last year. (2)

[6]

Q. 5) Let X and Y be *iid* random variables from an exponential distribution with mean 0.5.

i) Defining Z = Min (X ,Y), obtain the cumulative distribution function of Z. (3)

ii) Hence, find the mean of Z. (1)

[4]

Q. 6) An insurance company pays commission to its agents at a rate of 30% of first premiums on selling an insurance policy. The number of policies sold by an agent in a year follows a Poisson distribution with mean 20. The amount of first premiums for a policy follows a Lognormal distribution with parameters $\mu = 6$ and $\sigma = 2.5$.

i) Calculate the mean and variance of amount of commissions on the first premiums paid to an agent by the company in a year. (5)

ii) If there are 1000 agents working for the company, calculate (stating the assumption made) the probability that the amount of commissions paid on first premiums in a year exceeds INR 65,000,000. (4)

iii) The company intends to reduce the commissions and focus more on online sales. By considering revised commissions as a proportion of existing commissions, determine the revised commission rate as a percentage of first premiums, if the amount of commissions paid on first premiums in a year needs to be limited to INR 65,000,000 with 95% confidence. (5)

[14]

Q. 7) Let Y be a random variable such that $Y = \sum_{i=1}^{n} X_i$; where $X_i, i = 1,2, \ldots, n$ are independent random variables $\sim Exp(\theta)$. Let Z be a random variable such that the moment generating function of Z is $M_z(t) = \sqrt{M_Y(t)}$.

By comparing $M_z(t)$with the MGF of chi square distribution, determine the value of θ for which Z follows a chi square distribution with appropriate degrees of freedom. **[4]**

Q. 8) A discrete random variable X assumes values -1, 0, 1 each with non-zero cell probability as under:

X	-1	0	1
$P(X=x)$	$\frac{1}{6} + \alpha$	$\frac{1}{2} - 3\alpha$	$\frac{1}{3} + 2\alpha$

($-\frac{1}{6} < \alpha < \frac{1}{6}$, is a parameter).

A random sample of 25 observations gave respective frequencies of 7, 6 and 12.

i) Show that the log likelihood function for the given sample of observations is.

$$\ln L\,(\alpha) = 7\ln\left(\frac{1}{6}+\ \alpha\right) +\ 6\ln\left(\frac{1}{2}-\ 3\alpha\right) +\ 12\ln\left(\frac{1}{3}+\ 2\alpha\right)$$ (3)

ii) Determine the maximum likelihood estimate and explain why one of the two roots of the log likelihood equation is rejected as a possible estimate of α. (7)

iii) Using method of moments, determine the estimate of α. (2)

[12]

Q. 9) An automobile company plans to introduce fuel additive to increase the mileage in cars. In order to verify the claim that the additive increases the mileage, a random sample of 20 similar cars are used- ten with the additives and ten as control group.

The mileage outcomes over a test track under regulated conditions are as follows:

Control group	13.8	19.9	15.1	18.6	15.5	14.6	17.5	13.2	19.7	12.3
With additive	15.9	20.4	14.0	18.8	17.7	17.5	17.9	13.1	19.3	22.2

i) Calculate a 95% confidence interval for the difference between the mean mileages. State the assumptions made and comment on the result. (6)

ii) Calculate the minimum sample size so that the width of a 95% confidence interval for the difference between means is less than four, assuming that the samples have the same sizes and equal variances. (4)

iii) Can we apply a paired *t*-test in part **(i)**? Justify. (2)

[12]

Q. 10) A certain species of buffaloes are divided into three types based on the yield of milk per day.
The three types are:

- High yielding type: More than 10 litres.
- Medium yielding type: 5 - 10 litres.
- Low yielding type: Less than 5 litres.

These buffaloes have either curved or straight horns. For a random sample of 800 buffaloes, the distribution of yield and horn type is as under.

Horn \ Yield	Low	Medium	High
Curved	155	67	123
Straight	168	237	50

i) Test whether these data indicate any association between yield of milk and horn type at 5% level of significance. (5)

ii) A genetic model suggests that the proportion of each combination is as follows:

Horn \ Yield	Low	Medium	High
Curved	p	$\frac{p}{2}$	$\frac{1-p}{6}$
Straight	$\frac{7p}{6}$	$\frac{5p}{3}$	$\frac{5(1-5p)}{6}$

where $0 < p < 0.2$ is an unknown parameter. Test the goodness of fit for this model at 5% level, given that the MLE of $\hat{p} = 0.1846$. (5)

iii) Comment briefly on the conclusions of the above two tests. (2)

[12]

Q. 11) A finance student recorded Net Asset Values (NAVs) (in INR per unit) of a mutual fund company, at annual intervals over a 10-year period. The data are given in the table below:

Time (x)	0	1	2	3	4	5	6	7	8	9	10
NAV (y)	10	13	18	24	32	44	58	79	106	138	179

$$\sum_{i=0}^{10} x_i = 55; \sum_{i=0}^{10} y_i = 701; \sum_{i=0}^{10} x_i^2 = 385; \sum_{i=0}^{10} y_i^2 = 76{,}055; \sum_{i=0}^{10} x_i y_i = 5{,}250$$

i) Assuming the NAVs are normally distributed, fit the linear regression model to the above data:

$$Y_i = \alpha + \beta X_i + e_i ; i = 0, 1, 2, \dots, 10$$

where α and β are parameters and $\{e_i\}$ are *iid* normal random variables with mean zero and variance σ^2. (2)

ii) Calculate the "proportion of variability explained by the model" and comment. (2)

iii) Estimate NAV at time 11 using the fitted regression line and comment. (2)

iv) Complete the table of residuals (rounding to the nearest integer):

Time (X)	0	1	2	3	4	5	6	7	8	9	10
Residual $\{e\}$	26	13	2							11	36

(3)

v) Use a dot plot of the residuals to comment on the assumption of normality. (2)

vi) Plot the residuals against time and hence comment on the appropriateness of the linear model. (3)

[14]

Q. 12) A chemical plant is in the process of choosing the most efficient catalyst for a chemical reaction. The test runs yielded the following efficiencies for various catalysts:

Catalyst (C1)	Catalyst (C2)	Catalyst (C3)
0.81	0.84	0.94
0.76	0.94	0.89
0.89	0.88	0.91
0.79	0.88	0.88
0.89	0.89	0.90
	0.85	
Number of tests		
5	6	5
Sum of squares		
3.442	4.6526	4.0882

i) Test whether the efficiencies of the catalysts differ significantly at 5% level of significance. (5)

ii) The catalyst C3 is the most expensive amongst the three whereas the other two are similarly priced. Using the least significant approach at 10% level, recommend a catalyst amongst the three that the company should choose for the reaction. (3)

[8]

****