INSTITUTE OF ACTUARIES OF INDIA

EXAMINATIONS

6th December 2022

Subject CS1A – Actuarial Statistics (Paper A)

Time allowed: 3 Hours 15 Minutes (10.15 – 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.***
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. You are not allowed to carry the question paper in any form with you.

Q. 1) To estimate the population variance σ^2 the statistic ${S'_n}^2 = \frac{\sum(X_i - \bar{X})^2}{n}$ is used rather than using estimator S_n^2, where n is the sample size.

Find the bias of ${S'_n}^2$, when n = 13, mean is 5.24 and $\sigma^2 = 3.4224$ **[3]**

Q. 2) Choose the correct option and provide a reason for your choice. No marks would be awarded if reason is not provided.

i) A multivariate model is fit with 10 explanatory variables and 100 observations. Due to IT restrictions the model can't be implemented. An Actuary decided to use Principal Component Analysis (PCA) for reducing the dimensionality of the data set. How many Components will be there after fitting PCA?

A. 3
B. 9
C. 10
D. 100 (2)

ii) Which of the following is not a Linear Predictor?

A. $Y = \alpha + \beta X^2$
B. $Y = \alpha + \beta^2 X$
C. $Y = \alpha + \beta \left(\frac{1}{X} \right)$
D. All of the above (2)

[4]

Q. 3) **i)** For a standard normal random variable Z, derive an expression for its Moment Generating Function (MGF) using first principles. (3)

ii) Using the results obtained in part (i), prove that a normal variable X with mean μ and variance δ^2, is perfectly symmetrical about its mean i.e. the coefficient of skew-ness of the normal variable X is equal to 0.

Hints:

a) Standard relationship between normal variable X and standard normal variable Z i.e. "$Z = (X - \mu) / \delta$" can be directly used without proof.

b) Use the fact that $E(Z^r)$ is the coefficient of the term $t^r / r!$ in the Taylor expansion. (3)

[6]

Q. 4) A health insurance company has recently launched a new one year health insurance product which pays a fixed sum assured on the incidence of Heart, Cancer and Liver related ailments in the next one year.

Even after a claim under one ailment, coverage continues for the other ailments. A fixed sum assured would be paid on the incidence of the pre-defined ailment and no further claims can then arise for that particular ailment.

It can be assumed that these three risks are independent.

Sum Assured for Heart, Cancer and Liver related ailments are INR 20 lakhs, INR 25 lakhs and INR 15 lakhs respectively.

The company estimates the probabilities of claim arising in the next year to be 0.01 for Heart related ailments, 0.02 for Cancer and 0.005 for Liver related ailments.

i) Determine, for a single policy, using suitable Bernoulli variables, the mean and standard deviation of the total claim amount to be paid over the next year. (2)

ii) You are informed that a claim has been reported under a policy. Given that there is a claim under the policy, calculate the expected pay-out on this claim. (3)

iii) Why does mean claim pay-out in part (i) differ from the expected claim pay-out in part (ii)? (2)

[7]

Q. 5) Out of the 85 tosses of a coin, 40 tosses turn out to be heads.

i) Let N denote the total number of heads in 85 tosses, what is the most suitable distribution of N? Estimate the mean and variance of N. (2)

ii) Find out the probability that N > 40 using approximate distribution. (2)

Let the distribution of N as specified in part (i)

iii) Test the hypothesis that:

H_0: probability of getting heads = 0.5 v. H_1: probability of getting heads > 0.5 at the significance level of 5% using the probability value calculated in part (ii) above. (1)

iv) Find X where P(N > X) is less than the significance level of 5% leading to rejecting the null hypothesis of the above test. (2)

[7]

Q. 6) A company offers an optional basic group term insurance policy to its employees, as well as an accidental death benefit rider. To be covered under the accidental death benefit rider, an employee needs to first opt for group term insurance policy.

Let X denote the proportion of employees who have opted for group term insurance cover. Let Y denote the proportion of employees who have opted for accidental death benefit rider.

Let X and Y have the following joint density function f (x,y) on the region where both X and Y are non-negative:

$$f(x,y) = 2(x + y)$$

i) Clearly specify the bounds on values of X and Y for which the above joint density function holds true. (1)

ii) Determine the marginal density function of X. (2)

iii) Given that 10% of the employees opt for the group term insurance policy, calculate the probability that less than 5% of the employees opt for the accidental death benefit rider. (4)

[7]

Q. 7) Let the random variable X have the Poisson distribution with probability function:

$f(x) = \frac{e^{-\lambda}\lambda^{x}}{x!}$, x = 0,1,2,...

i) Show that P(X = k+1) = $\frac{\lambda}{k+1}\, P\,(X = k)$, k = 0,1,2,… (2)

It is believed that the distribution of the number of claims which arise on insurance policies of a certain class is Poisson.

A random sample of 1,000 policies is taken from all the policies in this class which have been in force throughout the past year. The table shows the observed number of policies with 0, 1, 2, 3, 4, 5, 6, 7 and 8 or more claims during the year:

No. of Claims(k)	0	1	2	3	4	5	6	7	8+
No. of policies(f_k)	300	365	216	70	30	16	2	1	0

For these data the Maximum Likelihood Estimate (MLE) of the Poisson parameter λ is $\hat{\lambda}$ =1.186

ii) Calculate the expected number of policies with 0, 1, 2, 3, 4, 5, 6, 7 and 8 or more claims during the year under the Poisson model with parameter given by the MLE above, using the recurrence formula of part (i) (or otherwise). (3)

iii) Perform an appropriate statistical test to investigate the assumption that the numbers of claims arising from this particular class of policies follow a Poisson distribution. (5)

[10]

Q. 8) A new political party is investigating whether size of policemen impacts petty crimes. Following data of 9 zones of a state has been collected:

Zones	A	B	C	D	E	F	G	H	I
Policemen	178	161	140	106	171	153	142	124	37
Cases	171	114	62	46	184	149	99	70	39

i) Calculate Spearman's and Kendall's correlation coefficients. (4)

ii) After investigation, party concluded that more crimes are done where more policemen are deployed and suggesting reduction police force. Comment on their conclusion. (2)

iii) An expert suggested party to use Pearson's correlation coefficients instead. Compute Pearson's correlation coefficient and test whether there is no correlation between policemen and cases.

$$\sum x = 1212\ ,\sum y = 934\,,\sum x^2 = 178000\ \sum y^2 = 120476 \text{ and } \sum xy = 140790$$

(6)

[12]

Q. 9) Five years ago, an insurance company began to issue insurance policies covering medical expenses for dogs. The insurance company classifies dogs into three risk categories: large pedigree (category 1), small pedigree (category 2) and non-pedigree (category 3).

The number of claims n_{ij} in the i^{th} category in the j^{th} year is assumed to have a Poisson distribution with unknown parameter θ_i.

Data on the number of claims in each category over the last 5 years is set out as follows:

Category	Year					$\sum_{j=1}^{5} n_{ij}$	$\sum_{j=1}^{5} n_{ij}^2$
	1	2	3	4	5		
1	28	41	47	54	62	232	11434
2	35	48	55	57	65	260	14028
3	26	29	20	39	31	145	4399

Prior beliefs about θ_1 are given by a gamma distribution with mean 50 and variance 25.

i) Find the Bayes estimate of θ_1 under quadratic loss. (5)

ii) Calculate the expected claims for year 6 of each category under the assumptions of Empirical Bayes Credibility Theory Model 1. (6)

iii) Explain the main differences between the approach in part (i) and that in part (ii). (2)

iv) Explain why the assumption of a Poisson distribution with a constant parameter may not be appropriate and describe how each approach might be generalised. (2)

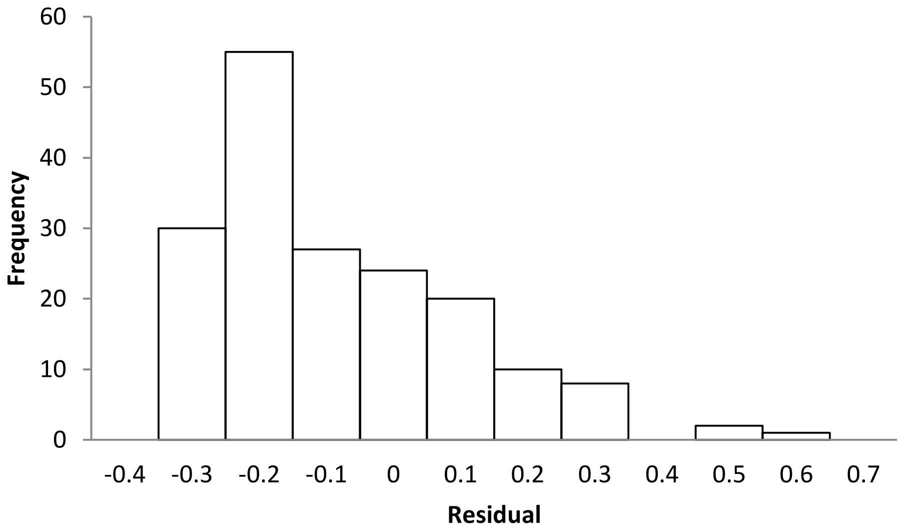
[15]

Q. 10) An actuarial trainee fit a simple linear model to detect bus cancellation charges (Y).

As an exploratory analysis, trainee determined that correlation coefficient between Y and X is 0.623 as a first step.

The slope and intercept of the model are b = -5 and a = 10.

Further, after fitting the model, histogram of residuals (shown below) was prepared as per manager's request:



i) Plot graph showing relationship between Y and X for X = -2 to 2

Note: Compute Y for each X (-2,-1, 0, 1, 2) and then plot a freehand graph. (2)

ii) Comment on the following:

a) Validity of assumption of the linear model. (2)

b) Possibility of error made by Actuarial Trainee. (2)

Trainee fits a generalised linear model to predict number of free bus cancellations for prime members. The following data for 20 prime members collected for 2 cities:

City I	2	2	0	0	1	0	0	0	1	0
City II	1	2	1	0	2	1	1	0	2	2

Trainee uses Poisson distribution to analyse bus cancellation charges and fits following models:

Model 1 : $\log \mu_i = \alpha$

Model 2 : $\log \mu_i = \alpha + \beta\, x_i \; where \;\; x_i = \begin{cases} 1 \;\; for\ City\ I \\ 0\ for\ City\ II \end{cases}$

iii) Show that Poisson distribution is a member of the exponential family of distributions. (2)

iv)

a) Calculate the maximum likelihood estimator for α and β under Model 1 and 2. (5)

b) Compute the probability of 3 cancellations for City I and II under Model 2. (2)

v) Compute Scaled deviance for Model 1 and 2.

*Note: y * log y = 0 for y=0 can be assumed.* (5)

vi) Suggest which model is better by using appropriate statistic (2)

Trainee also tried another model.

Model 3: $\log \mu_i = \begin{cases} \delta \ for\ City\ I \\ \gamma\ for\ City\ II \end{cases}$

vii) For Model 3, compute the following:

a) MLE for δ and γ (2)

b) AIC (Akaike's Information Criterion) (1)

viii) Compare Model 2 and Model 3 and comment. Are there any similarities between the models 2 and 3? (2)

ix) For review, Plot of Pearson residual is used. Write down its disadvantages for checking Poisson distribution. (2)

[29]

****