INSTITUTE OF ACTUARIES OF INDIA

EXAMINATIONS

21st November 2019

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) A discrete random variable X has the following probability density function:

$P(X = x) = p(1 - \text{p})^{\text{x}-1}$ where $0 < p < 1$ and x>0

i) Derive the Moment Generating Function (MGF) and the Cumulant Generating Function (CGF) for the above distribution. (4)

ii) Using either the MGF or the CGF determine E(X). (2)

[6]

Q. 2) i) Define the variable t_k used in the t-test for sampling distribution of sample mean describing all the symbols used. (2)

ii) State the mean and variance of t_k for k>2. (1)

iii) A sample of 10 numbers from normal population has sample mean and sample variance as 50 and 48.667 respectively.

Determine the confidence interval for the population mean at 99% confidence level-

a) Using the t-test tables (2)
b) Assuming a Normal distribution with parameters as the results of part (ii) above (4)

[9]

Q. 3) The number of claims X, on an insurance policy over a year follows a Poison distribution with unknown parameter θ. The number of claims observed in the previous n years are $x_1, x_2 \dots x_n$

Prior distribution for θ has gamma distribution with parameters α and λ, as defined in the actuarial tables.

i) Derive the posterior distribution of θ given $x_1, x_2 \dots x_n$ (4)

ii) Show that the Bayesian estimate of θ under quadratic loss is equal to $\frac{\alpha+\sum x_i}{\lambda+n}$ (3)

iii) Show that the mean of the posterior distribution can be written in the form $Z * (sample\ mean) + (1 - Z) * (mean\ of\ prior\ distribution)$, defining Z as appropriate. (3)

[10]

Q. 4) The investment department of a life insurance company feels that the chances of the economy moving into a high level of financial stress over the next month are 80%.

Based on an alternate study of macro-economic variables, it has been found that the relative position of credit spreads vis-à-vis their long-term historical average at the beginning of a month is a *leading indicator* of the level of financial stress in the economy over the following month. The studies indicate that the economy may face high levels of stress when there is a spike in credit spreads.

Based on data gathered for the past 10 years, it has been observed that:

- 75% of the times when the economy ends up in high level of financial stress, it is preceded by high credit spreads; and
- 40% of the times when the economy ends up in a low level of financial stress, it is preceded by high credit spread.

Compute the following:

i) Prior probability of the economy moving into a high level of financial stress. (1)

ii) Find the conditional probability of the credit spreads being high in the beginning of the month given that the level of financial stress was high over the following month. (2)

iii) Calculate the posterior probability that the financial stress index will be high given that the credit spreads are high in the beginning of the month. (3)

[6]

Q. 5) You are given a coin. Let p be the probability of getting heads if the coin is flipped. You have not been told if coin is fair or biased. To start with you assume that p can take any value uniformly over the range 0 and 1. Next you decide to perform an experiment. You start flipping the coin until you get the head for the first time.

i) Derive the posterior distribution of p if you get the head for the first time after m flips. (4)

ii) Derive the posterior distribution of p if you get the head for the second time after a further n flips. (4)

[8]

Q. 6) A statistics student was observing the number of passengers sitting (other than the driver) inside five-seat capacity cars entering the city centre in City A and City B respectively over a particular period and summarised his observations as below:

Number of seats occupied apart from driver in the observed car	0	1	2	3	4
Number of cars (City A)	70	120	201	80	29
Number of cars (City B)	40	100	160	170	70

It is proposed that a Binomial model with parameters n and p can be fitted for city A (where p is the probability that a seat is occupied with $0 < p < 1$).

i) Determine the maximum likelihood estimate of p for City A. (5)

ii) θ_1 and θ_2 are the proportions of cars plying with less than 2 passengers in cities A and B respectively.

a) Estimate θ_1 *and* θ_2 using the sample data provided. (2)

b) Calculate the symmetrical 95% confidence interval for the difference in proportion of cars plying with less than 2 passengers in both the cities. (4)

c) Comment on your answer in part b. (1)

[12]

Q. 7) Number of claims in a year on an insurance policy is believed to follow a Poisson distribution. Claims on portfolio of 1000 such policies were observed for one year. It was suggested that the value of Poisson parameter is 3. If the observed number of claims in that one year is less than 3100 then the suggested value of 3 for the Poisson parameter is accepted else rejected.

You may use the result that probability distribution of summation of 'n' Poisson variables with parameter μ is $Poi(n\mu)$.

i) Define Type I error and estimate it for the above case. (4)

ii) Define Type II error. (1)

iii) Define power of a test and determine the power of test in terms of μ in above case. (2)

iv) If the actual observed number of claims is 2900, determine the confidence interval for the Poisson parameter at 99% confidence level. (2)

[9]

Q. 8) There are 2 fair dice coloured yellow and black respectively which are rolled simultaneously. Two random variables are defined as follows:

- X: Number rolled on the yellow dice
- Y: Number of times '3' appears when both dice are rolled once together

i) Determine the joint probability mass function (X, Y) by filling up the following table:

	X=1	X=2	X=3	X=4	X=5	X=6
Y=0						
Y=1						
Y=2						

(4)

ii) Compute the value var(X+Y) (6)

You are given $E(X) = \frac{21}{6}, E(Y) = \frac{1}{3}\ , (X^2) = \frac{91}{6}\ , E(Y^2) = \frac{14}{36}$

[10]

Q. 9) **i)** Consider a response variable Y which is related to the value of x by the following equation

$$y_i = \alpha + \beta . x_i + \varepsilon_i \qquad i = 1,2,3 \ldots , n$$

Here, ε_i are independent and identically distributed standard normal variables.

State the function which is to be minimized to arrive at the least squares estimates for α & β. (2)

ii) You have been given a sample set of size n of paired data (x, y) and it has been suggested to use a model of the form:

$$E(Y_i) = \gamma . e^{x_i}$$

Derive the least squares estimate of γ. (4)

iii) A weighted least squares regression is a regression where to arrive at the estimates of the regression coefficients, the weighted sum of squared errors are minimized instead of a simple sum of squared errors.

Now using the model prescribed in part (ii), derive the least squares estimate of γ by assigning a weight to each of the error terms where the error term is inversely proportional to the value of the independent variable x_i. For instance, the weight applied to ith squared error term will be $\frac{1}{x_i}$ (4)

[10]

Q. 10) You have been given the data set comprising of mortality rating and premium rates.

Mortality Rating (%)	25	50	75	100	135	180	215	245	300	365	400
Premium Rate	3.50	5.80	8.07	10.33	13.44	17.38	20.39	22.94	27.53	32.83	35.62

It has been specified that the premium rates can be expressed as a cubic function of mortality rating. You have been given the task of deriving a simple formula for calculation of premium rates at different mortality ratings.

Suppose, moving ahead in line with the proposed methodology, you decide to fit a simple linear regression with premium rates being regressed on cubic function of mortality rating. The transformed data is as below:

(Mortality Rating)^3 : (X)	0.02	0.13	0.42	1.00	2.46	5.83	9.94	14.71	27.00	48.63	64.00
Premium Rate: (Y)	3.50	5.80	8.07	10.33	13.44	17.38	20.39	22.94	27.53	32.83	35.62

You are given the following summary statistics:

$$\sum x = 174.13, \sum y = 197.84,$$

$$\sum x^2 = 7545.90, \sum y^2 = 4747.45,$$

$$\sum (x - \bar{x})(y - \bar{y}) = 2176.84$$

i) Derive the linear regression equation of premium rates Y on X. (3)

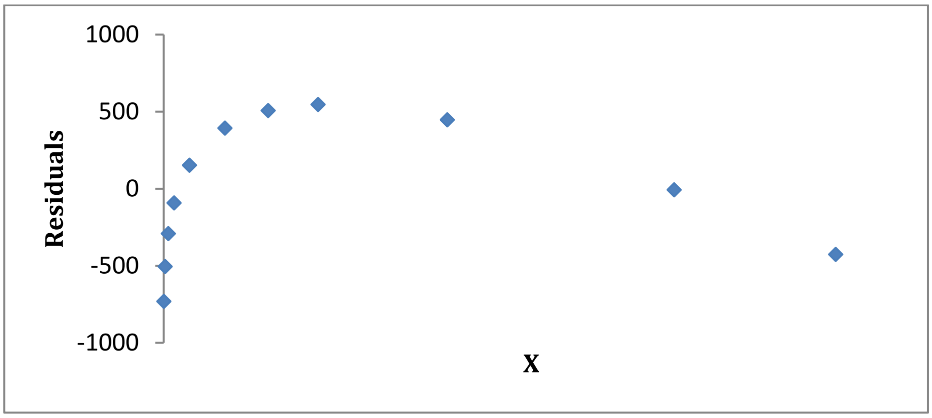
ii) Perform a statistical test to investigate the hypothesis that there is no linear relationship between X and Y.

State clearly all assumptions made. (6)

iii) Calculate the sample correlation coefficient. (2)

iv) Calculate the 95% confidence interval for the individual and mean responses corresponding to x^3 =25. (7)

v) Consider the residual plot of the fitted regression:



Comment on the fit of the model and any drawbacks of using the model rather than the full premium rate table. (2)

[20]

****