

INSTITUTE OF ACTUARIES OF INDIA

EXAMINATIONS

26th November 2020

**Subject CS2B – Risk Modelling and Survival Analysis
(Paper B)**

Time allowed: 2 Hour (14.30 – 16.30 Hours)

Total Marks: 100

Paste your R Code, Outputs and plots/ graphs in your answer sheet.

- Q. 1)** Use the data file “Mortality_Investigation.csv” to answer the below questions. This table contains the details of mortality investigation performed on 100 different male lives between 1-Jan-2016 and 31-Dec-2018.
- i)** Print first 6 rows of the data ensuring correct data formats for all the columns and calculate the proportion of the total sample died, survived and withdrew from the observation period. (4)
- ii)** Compute two new columns ‘Age_At_Entry’ and ‘Age_At_Exit’ for each person (you can assume a completed year as 365.25 days and you can compute the age as difference between the dates in decimals like 35.234 years) and print the last 6 rows of the newly formed table. (5)
- iii)** Compute the average age at entry and the average age at exit of the people for whom “Death” was the reason of Exit. (3)
- If the age label used is “Age last Birthday”, then calculate
- iv)** Number of lives out of total 100 lives who have contributed the full year towards the central exposed to risk E_{37}^c . (4)
- v)** Number of lives who did not contribute at all towards the central exposed to risk E_{37}^c . (4)
- vi)** Total number of years of contribution to E_{37}^c by all the lives together (Round your answer to 2 digits after the decimal). (7)
- [27]
- Q. 2)**
- i)** Simulate a time series of length 200 observations in a line chart using the following R function and comment on stationarity. (Set a seed value of 100)
- ```
arima.sim(list(order = c(1,1,1), ar = 0.7, ma = 0.3), n = 200)
```
- (6)
- ii)** Plot the auto correlation function and partial auto correlation function of the data if (i) is stationarity, otherwise take the first difference of the observations and plot the above graphs. Comment on the possible ARMA model based on the graphs. (5)
- iii)** Fit AR(1), AR(2), MA(1) and ARMA(1,1) models to the observations from (ii) and comment on the best fitting model using AIC or any other metric. (7)
- iv)** Using the best fitted model from (iii) create a forecast of the observations for the next 3 periods. (4)
- [22]
- Q. 3)** The number of claims in a year on an individual policy follows a Poisson distribution with a parameter of 0.75. Individual claim amounts follow a Gamma distribution with parameters (shape parameter  $\alpha = 2$  and rate parameter  $\lambda = 20,000$ ). Set seed value of 100.
- i)** Simulate claims for 10,000 policies and create a frequency distribution of the number of claims made on a policy per year. (4)

- ii) Simulate the claim amounts for each claim and compute the aggregate claim amounts for each policy. Create a histogram of the aggregated claim amounts with proper labelling and legends. (6)
- iii) Compute the mean and variance of the aggregated claim amounts by using the method of moments of the compound Poisson distribution. (5)

An insurer plans to go for an aggregate excess of loss reinsurance policy. The insurer wants to evaluate the expected cost for the retention limits from 50,000 to 100,000 in steps of 5,000.

- iv) Compute expected costs for the insurer and the reinsurer for each of the retention limits. (6)
- v) The insurer decides to set a maximum retention limit where the mean cost to the insurer is not more than 75% of the mean aggregated claims. Identify the maximum possible retention limit based on the computations of (iv) and the proportion of total claims to be handled by the reinsurer. (5)
- vi) Based on that limit, compute the standard deviation and coefficient of skewness of the costs to the reinsurer. (4)
- [30]**

**Q. 4)** Consider the data set ‘Covid\_2019.csv’ (name it covid19) where the first row of the csv file contains the headings for the columns.

- i) Print the number of missing values in each of the columns and create a new data set ‘covid19\_1’ by removing all the missing values. (4)

From covid19\_1, use the columns from Population Density (8th column) to Life Expectancy (17th column) to answer the following questions.

- ii) Create a new data frame “Covid\_Cluster” containing only the above mentioned columns. Normalize all the columns of the data frame using the scale function. (3)
- iii) Classify the countries into five groups by using the values obtained from (ii) applying K-Means clustering algorithm. It is mandatory to set a seed value of 100 before executing the algorithm. Print the number of countries in each cluster. (4)
- iv) What proportion of total countries in each cluster are severe with respect to COVID-19? You can use the “Severe” column from the original dataset. (5)
- v) Print the total number of cases and total number of deaths for each cluster. (5)

**[21]**

CS2B\_COVID\_2019

[http://www.actuariesindia.org/downloads/Year\\_2020\\_November\\_Exam/CS2B\\_COVID\\_2019.csv](http://www.actuariesindia.org/downloads/Year_2020_November_Exam/CS2B_COVID_2019.csv)

CS2B\_Mortality\_Investigation

[http://www.actuariesindia.org/downloads/Year\\_2020\\_November\\_Exam/CS2B\\_Mortality\\_Investigation.csv](http://www.actuariesindia.org/downloads/Year_2020_November_Exam/CS2B_Mortality_Investigation.csv)

\*\*\*\*\*