INSTITUTE OF ACTUARIES OF INDIA

EXAMINATIONS

25th November 2020

Subject CS1B – Actuarial Statistics (Paper B)

Time allowed: 2 Hour (14.30 – 16.30 Hours)

Total Marks: 100

Q. 1) Use the 'MotorClaims' dataset provided to answer the following questions: (MotorClaims.csv)

i) Fit Gamma distribution on the dataset provided by determining its scale and shape parameters. State clearly the distribution with the parameters. (8)

ii) Simulate 1000 values from the distribution obtained in question (i) and print the first six values. (Set seed to 100) (5)

iii) Calculate the mean and variance of simulated values. (2)

iv) Obtain a QQ plot for the simulations of 1000 values and a normal distribution. (4)

v) Add a line to the above plot to show the true position of normal distribution. (2)

vi) Comment on the shape of the distribution and how close it is to a normal distribution. (3)

[24]

Q. 2) The dataset 'mtcars' (built into R) consists of data on various models of car, taken from an American motoring magazine (1974 Motor Trend magazine). For each car, there are certain features expressed in varying units.

i) Load the 'mtcars' dataset which is built into R. How many observations and variables are there in this dataset? Your answer should include the R output. (4)

ii) Identify the categorical variables from the dataset 'mtcars' and create a dataset excluding the categorical variables. (5)

iii) How many observations and variables are there in the new dataset? Your answer should include the R output. (2)

iv) Carry out a principal component analysis on the new dataset of mtcars by passing two arguments, 'center' and 'scale' to be TRUE. Your answer should include a summary of the principal component analysis. (5)

v) How many components of the reduced data should be retained using the output derived in question (iv)? Also state the reason for the same. (3)

[19]

Q. 3) Data is provided for 'BMIClaims' of 150 policyholders and corresponding claim count. (BMIClaims.csv)

i) Obtain 95% confidence interval for the standard deviation of BMI (using qchisq). (6)

ii) Further test the standard deviation of BMI to be equal to 4 by obtaining p value. State your conclusion of the test. (6)

iii) If obese is defined to be BMI above 30, use binom.test to calculate 99% confidence interval for proportion of obese people and comment on the likelihood if more than 20 pc are obese. (6)

iv) Claim frequency can be calculated as claim count divided by number of policyholders. Test whether claim frequency is different between obese and others. (6)

[24]

Q. 4) The CDC and EISS detect influenza activity through clinical data including Influenza-like Illness (ILI) physician visits on weekly basis. The objective of this question is to estimate influenza-like illness (ILI) activity using Google web search logs.

The csv file FluTrain.csv aggregates this data from January 1, 2004 until December 31, 2011 as follows:

"Week" - The range of dates represented by this observation, in year/month/day format.

"ILI" - This column lists the percentage of ILI-related physician visits for the corresponding week.

"Queries" - This column lists the fraction of queries that are ILI-related for the corresponding week, adjusted to be between 0 and 1 (higher values correspond to more ILI-related search queries).

i) Plot a histogram of the dependent variable, ILI. Comment on the shape of the distribution. (3)

ii) Plot the natural logarithm of ILI versus Queries. What does the plot suggest? (4)

iii) Fit a linear regression model for dependent variable log(ILI). Summarize it. (6)

iv) State the formula of the model fitted in part (iii), explaining all the terms used. (3)

v) Calculate R-squared and the correlation between the independent and dependent variable. What is the relationship between the two values? (5)

vi) Looking at the time period 2004-2011, which week corresponds to the highest percentage of ILI-related physician visits? (4)

vii) Based on the linear regression model fitted in question (iii), what is the estimate for the percentage of ILI-related physician visits for the week computed in question (vi)? (4)

viii) What is the relative error between the estimate (prediction calculated in question (vii)) and the actual observed value for the week computed in question (vi)? (4)

[33]

CS1B_BMIClaims

http://www.actuariesindia.org/downloads/Year_2020_November_Exam/CS1B_BMIClaims.csv

CS1B_FluTrain

http://www.actuariesindia.org/downloads/Year_2020_November_Exam/CS1B_FluTrain.csv

CS1B_MotorClaims

http://www.actuariesindia.org/downloads/Year_2020_November_Exam/CS1B_MotorClaims.csv

****