

# INSTITUTE OF ACTUARIES OF INDIA

## EXAMINATIONS

18<sup>th</sup> June 2019

**Subject CS1B – Actuarial Statistics (Paper B)**

**Time allowed: 1 Hour 45 Minutes (10.15 – 12.00 Hours)**

**Total Marks: 100**

### *INSTRUCTIONS TO THE CANDIDATES*

- 1. Mark allocations are shown in brackets.*
- 2. Attempt all questions, beginning your answer to each question on a new page.*
- 3. Attempt all sub-parts of the question in one document only, unless otherwise instructed to do so.*
- 4. Data set file accompanying the question paper is located at:  
C:\Users\admin\Desktop\CS1B\filename*
- 5. You need to import the same into R studio as soon as you begin the exam.*
- 6. Ensure to copy and paste R codes and output at regular intervals onto the MS word file.*
- 7. 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.*
- 8. Ensure that you click the “Submit” button only when you have completed the question paper and final submission has to be made.*

#### AT THE END OF THE EXAMINATION

Please return this question paper to the supervisor separately. You are not allowed to carry the question paper in any form with you. You are requested to save and submit the work before leaving the examination premises.

The CSV file “AutoClaims.csv” accompanies this exam paper.

### Description of the given data sets

#### AutoClaims.csv

An insurance company is trying to analyse the paid and closed claims data of the past quarter.

The columns carry the following explanation

STATE: The state of registration of the vehicle

CLASS: Rating class of the insured based on marital status, use of vehicle and other attributes

GENDER: Gender of the Insured (Same as the Driver)

AGE: Age of the Insured (Same as the Driver)

PAID: Amount paid to settle and close a claim

**Q. 1)** With reference to the dataset “AutoClaims.csv”, answer the following questions.

- i)** Prepare a table with mean, standard deviation and coefficient of variance (CV) of the claims paid
- a)** for each STATE and identify the one with the least and the highest CV
- b)** for each CLASS and sort the table in the ascending order of the CV

*Hint: Coefficient of Variance (CV) can be computed as the ratio of standard deviation to the mean.* (10)

- ii)** By using a box plot, identify the STATE(s) and CLASS(es) which have no outlier values in terms of the claims paid. (5)

- iii)** Prepare a contingency table identifying the proportion of insured across different CLASS(es) for each gender. Test the null hypothesis “There is no relationship between the GENDER and the CLASS of the insured” against an alternate hypothesis of existence of significant relation at 95% confidence level. Perform an appropriate test and comment. (10)

- iv)** Use an appropriate test to test if the mean claim amount paid and variance of claim amounts paid is uniform between males and females. (5)

**[30]**

**Q. 2)** Refer to the dataset “AutoClaims.csv” and answer the following questions.

You were asked to fit an appropriate distribution to the “PAID” data. You decided to fit Normal distribution, Lognormal distribution, Exponential distribution and Gamma Distribution based on the method of moments.

- i)** Estimate the parameters of each of these distributions. (8)

- ii) Plot a histogram of “PAID” data with 30 equal class intervals. Superimpose the histogram with the Probability Density Functions of the above four distributions using their estimated parameters obtained in part (i). Mark each plot distinctly using an appropriate legend. (8)
- iii) Compute the 5<sup>th</sup> percentile, 1<sup>st</sup> quartile, median, 3<sup>rd</sup> quartile and 95<sup>th</sup> percentile of both the actual claims paid as well as from the fitted distributions. (5)
- iv) Using the results from (ii) and (iii), comment on goodness-of-fit of the models to the data. (4)
- [25]

**Q. 3)** Refer to the dataset “AutoClaims.csv” and answer the following questions.

- i) Fit a linear regression model to predict the “PAID” claim amount based on other variables (*Consider the AGE as a numerical variable and all others as categorical*). Provide your interpretation of the model by explaining R-Squared, Adjusted R-Squared, p-value of the model and p-value of each of the coefficients. Identify the significant variables in the prediction of “PAID” claims. (12)
- ii) Comment on the applicability of the linear regression model by plotting “Residuals vs. Fitted Values” and “QQ Plot of the residuals”. (10)
- iii) Your actuarial friend has suggested you to use natural logarithm of “PAID” claims instead of the actual “PAID” Claim amount because the  $\log_e(\text{PAID})$  is more closer to normal distribution than “PAID” Claims. Verify the statement made by your friend by comparing the Skewness and Excess Kurtosis of both the PAID claims as well as  $\log_e(\text{PAID})$ . Write appropriate custom functions to compute both of them. (7)
- iv) Repeat the model in (i) above by considering the suggestion in (iii). Identify and comment on the key differences between both the models. (6)
- v) Your Manager has suggested that the model can be improved by adding interaction effects between STATE and CLASS, STATE and GENDER, CLASS and GENDER as additional variables to the set of independent variables taken in (i). Evaluate the worthiness of this suggestion. [10]
- [45]

Link for Data Set:[http://www.actuariesindia.org/downloads/CS1B\\_AutoClaims.csv](http://www.actuariesindia.org/downloads/CS1B_AutoClaims.csv)

\*\*\*\*\*