# INSTITUTE OF ACTUARIES OF INDIA

## EXAMINATIONS

### 23rd March 2022

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

### Time allowed: 2 Hours (14.30 - 16.30 Hours)

### Total Marks: 100

**Q. 1)**   An Actuarial student fits following simple regression model to the data
$y_i$ = alpha + beta * $x_i$ + $e_i$ ;  i =1 to 12
where $e_i$ are independent normal random variables with mean 0 and variance sigma$^2$

Use following 12 data points for x and y
where y is response variable while x is explanatory variable

x = c(5,10,15,20,25,30,35,40,45,50,55,60)
y = c(15,12,25,23,35,36,33,38,43,45,50,53)

**Note: Do not use standard 'model fitting related' R codes - lm, glm, fitted, residuals, predict, anova - to answer parts of this question.**

   **i)**   Calculate $S_{xx}$, $S_{xy}$ and $S_{yy.}$                                                                (7)

   **ii)**   Calculate alpha, beta and sigma$^2$ using results in part (i).                            (3)

   **iii)**   Calculate fitted values of y using results in part (ii).                                    (1)

   **iv)**   Calculate residuals using results of earlier parts and Calculate mean and variance of residuals and comment on the result.                                               (3)

   **v)**   Calculate 95% confidence interval for beta and comment if we can conclude that beta is not zero stating the Null and alternate hypothesis. Calculate 95% confidence interval for sigma$^2$.
          **Note: Candidates are expected to find tabulated value using R.**                (7)

   **vi)**   State $SS_{TOT}$ and Calculate $SS_{REG}$ and $SS_{RES.}$ Hence calculate the proportion of variability explained by the model using $SS_{REG}$ and $SS_{RES}$ and comment on the result. Calculate adjusted $R^2$ and compare it with R^2 to explain result.        (5)

   **vii)** Calculate mean predicted response when x=52 and 95% confidence interval for the same. **Note: Candidates are expected to find tabulated value using R.**        (4)
                                                                                                                      **[30]**

**Q. 2)**   Policy and claims information (PolicyData.csv) of 650 policies is provided to you. The data contains following fields:

Policy : Policy Number
Claim : Number of claims corresponding to each policy
Cust_Exp: Policyholder's experience (VS = Very satisfied, SA= Satisfied, DS= Disappointed, VD= Very Disappointed) at the end of the policy tenure.
Amount: Claim amount per policy. Note that amount is set 0 if there is no claim.

   **i)**   Create a frequency table of claim and share how many policies don't have any claim.        (2)

   **ii)**   Plot a histogram of claim count and suggest 2 distributions that can be a good fit.        (2)

   **iii)** Given that claim count follows Poisson distribution with following two possible values for Poisson parameter:

       a)  0.35 and
       b)  0.30

Compute confidence interval at 95% confidence level to assess which value is more suitable for the given data. (3)

**iv)** Compute mean, variance and median of log of claim amount. (Name it: log amount). Make sure to exclude policies with no claim. (4)

**v)** Obtain histogram and Normal QQ Plot of log amount. Add a line to the QQ plot for normal distribution (3)

**vi)** Indicate which distribution the claim amount might be following using evidence from (iv) and (v). (3)

**vii)** Assuming the log amount (i.e. log of claim amount) follows a Normal distribution, test if mean of log amount is greater than 10 at 90% level of confidence. State the hypothesis and conclusion clearly. (4)

**viii)** Assess whether the policyholder experience (i.e. Cust_Exp) changes with more number of claims. Create contingency table and perform test to check the above assertion. State the hypothesis clearly. (3)

**ix)** Please explain why warning message appears while performing the above test. (1)

**x)** Again, perform the above test by combining

a. Very Satisfied and Satisfied customers
b. Disappointed and very disappointed customers
and
c. 2 or more claims.

Please provide your conclusion on association of policyholder's experience with number of claims. (5)

**xi)** Amount is defined to be large if the amount is greater than 100,000. Calculate 95% confidence interval for proportion of large claim, and comment on the likelihood if more than 25% of claims are large. (5)

**[35]**

**Q. 3)** A General Insurance company is trying to analyse the two-wheeler motor insurance claims reported over last one quarter.

The data is provided herewith the file MotorClaim.csv which contains the following fields

POLICY: Policy Number
CLAIM : Amount of claim reported for a policy

Insurance Company is interested to find out an appropriate distribution to fit the "CLAIM" data. You are being asked to find out the appropriateness of the following distributions based on method of moments:

1. Normal distribution
2. Lognormal distribution
3. Exponential distribution
4. Gamma distribution.

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

**ii)** Plot a histogram of "CLAIM" data with 35 equal class intervals. Superimpose the histogram with the probability density function of the above four distributions using their estimated parameters as obtained in part (i). Mark each plot distinctly using appropriate legend. (8)

**iii)** Compute the 5th percentile, 1st quartile, median, 3rd quartile and 95th percentile of both the actual claim paid as well as the fitted distributions. (5)

**iv)** Using the results from (ii) and (iii) comment on goodness of fit of the models to the data. (4)

**v)** Assuming Gamma distribution to be the right fit to the data, simulate 20,000 values of claim amounts using the Gamma distribution based on the parameter estimates obtained in part (i) and print first 10 values of claim amounts. (Set seed to 2022) (2)

**vi)** Generate 700 different random samples of size 400 from the simulated data obtained in part (v) and compute sample mean for each of the samples. (Set seed to 2022) (5)

**vii)** Plot the histogram of sample means generated from part (vi) and comment on the distribution of the sample means from the point of view of central limit theorem. (3)

**[35]**

\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*