INSTITUTE OF ACTUARIES OF INDIA

EXAMINATIONS

25th May 2023

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. *All the detailed guidelines are available on exam screen.*
5. *Do save your work in solution template on a regular basis.*
6. *If Any, Data set file(s) accompanying the question paper is available for download on the exam screen.*
7. *You need to import the same into R studio as soon as you begin the exam.*
8. *Ensure to copy and paste R codes and output at regular intervals onto the solution template.*
9. *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 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.

Q. 1)

i) Use the following code to generate 150 random numbers from uniform [0,1] distribution. Show that mean of the sample ~= 0.52 at 2 decimal places.

```
set.seed(052023)
u<-runif(150)
```

(2)

ii) Using the random numbers generated in part (i), simulate sample with sample size = 150 from chi-square distribution with 2 degrees of freedom.
No need to print the sample. Store the sample.
[**Hint:** useful functions : `qchisq`] (2)

iii) Using the random numbers generated in part (i) simulate another sample with sample size = 150 from Gamma (1, ½) distribution.

Show and Explain why samples generated in part (ii) and (iii) are same by establishing link between the chisquare and gamma distributions. (3)

iv)

a) Plot histogram of sample generated in part (ii) and comment on the shape of the distribution. (3)

b) Compute mean & median of sample and explain why mean is greater than median. (2)

v) Simulate 1000 values of sum of samples of size 150 from chi-square distribution with 2 degrees of freedom.

Store the value of sum of samples. Also, make sure to set.seed(052023) again before generating samples.
No need to print the data. (4)

vi) Plot histogram of 1000 samples sum generated in part (v) and comment on the shape of the distribution in the context of central limit theorem. (3)
[19]

Q. 2) An investment firm is planning to invest in a new start-up GPT. To assess the valuations, investment firm asked you to study the sales of GPT.

SalesData.csv contains the sample sales data of 60 repeat customers having the following information:

Order: Number of orders made by customers in last month
Value: Average order value per order of customers
Device: Device used to order
Age: Age of customers
City: Location of customers (grouped into 2 segments)

i) Using read.csv load the Sales data and compute sample mean of value. (2)

ii) Compute

a) Kendall correlation between Value and City. (1)

b) Covariance between Value and Age. (2)

iii)

a) Create a scatterplot between Value and Age. Without performing any test, state whether the hypothesis that correlation between Value and City equals 0 can be accepted or not. Provide reason for the same. (3)

b) Compute the confidence interval of correlation coefficient between Value and Age and test the assertion that correlation coefficient is 0. (3)

iv) Project manager asked you to fit a multiple linear regression in which Value is modelled as response and customer details (Device, Age and City) as explanatory variables. However, due to time constraints only 2 models can be tested.

a) Using part (iii), suggest, including reason, which explanatory variable (customer detail) should not be included while fitting the regression model. (2)

b) Fit regression model with Value as response and customer details (Device, Age and City) as explanatory variables.
Comment on the significance of the parameters including linkage with your suggestion in above part iv(a). (4)

c) Fit simple regression model Value ~Device and compare with the above model using ANOVA.
Explain the result of ANOVA including the hypothesis considered, degrees of freedom shown in output and inference drawn. (5)

d) Write down the regression equation for the above model, fitted in part iv(c).
Please use the parameter values and not the symbols. (3)

v) Model Value ~Device, fitted in part iv(c), is selected for further use.

Calculate a 95% confidence interval for

a) β , the true underlying slope parameter. (2)

b) σ^2 , the true underlying error variance. (5)

Hint: Some useful functions : qchisq ,qt, confint

vi) Investor wants to test if number of orders follows Poisson distribution with mean μ.

a) Estimate μ using the Sales data. (2)

b) Using Sales data, create a frequency table showing number of customers by orders. (2)

c) Perform goodness of test to assess Order follows Poisson distribution.
You can combine orders equal or more than 5 to ensure actual frequency is at least 5.

You can use *as.numeric* and apply on frequency table to store the frequency of customers in a vector.
Ensure that total sum of expected probabilities equals 1. (8)

Hint : Useful function : chisq.test

vii)

a) Assuming the number of orders follow Poisson distribution, fit a poisson Generalised Linear Model (GLM) specified as Order~Device and print its summary.

Make sure to specify the family as Poisson and appropriate link function. (3)

b) State the link function you have used in above model and provide reason why it is appropriate for poisson distribution. (2)

viii)

a) Project manager has asked you to predict the average number of order for following 2 customers using this fitted poisson model.

Device	Age	City
Mobile	18	1
Mobile	28	2

[**Hint:** type="response" provides prediction on the scale of the response variable. The default is on the scale of the linear predictors]. (4)

b) After observing the predicted average number of orders for both customers, project manager doubts if there is any error in the model. Explain to project manager that model is correct and predicted values are as expected. (2)

ix) Investment firm wants to close one of the channels – Website (Laptop) or App (Mobile). Project manager asks you to provide total value of these channels.

Total value is defined as Average order value x Average number of order.

Predict total value for laptop & mobile and suggest which channel to close.

Use models Value ~Device, fitted in part iv(c) and Order ~ Device fitted in part vii(a) for predicting total value. (6)

[61]

Q. 3) Total sales per month on a particular cloud kitchen follow a normal distribution with unknown mean θ and variance 20^2. Sales x_1, x_2,…..,x_n are observed over n months. Prior beliefs that θ follows normal distribution with mean 60 and variance 5^2.

Sales for last 5 months are extracted for analysis. Total sales for last 5 months were 340.

i) Mθ(t) denotes MGF of θ. State and compute M'θ (0) and M" θ (0).
Use m1t0 for M'θ (0) and m2t0 for M" θ (0). (3)

ii)

a) State the posterior distribution of θ and compute parameters of the distribution using the sales of last 5 months. (6)

b) After 50 months same analysis is performed using the 50 months sales data. Total sales for last 50 months were 3400.

Compute the parameters of posterior distribution of θ using 50 months sales data. (3)

iii)

a) Use the below code to plot the prior distribution of θ

```
x<-60+seq(-3,3,by=0.2)*5
y<-dnorm(x,mean=60,sd=5)
plot(x,y,ylim=c(0,.02))
```

(1)

b) Add the line to overlay the posterior distribution of θ derived in part iii(a). (2)

c) Add another line to overlay the posterior distribution of θ derived in part iii(b). (2)

d) Place the final graph and comment on it. (3)

Hint: useful function : lines

[20]

****