INSTITUTE OF ACTUARIES OF INDIA

EXAMINATIONS

8th December 2022

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) The Gym instructor participated in global survey of weights used for biceps curl machine. You are required to help him analyse the data to share the results.

i) Load the data "weights.csv" into R and print the summary of the data. (2)

ii) Obtain a 90% confidence interval for the standard deviation of weights lifted by Males (M). (3)

iii) Assuming that the weights follow normal distribution with mean and standard deviation of data provided for Males.
Test the Hypothesis that μ=12 kgs in R and obtain the p-value using pnorm or otherwise. (3)

iv) Further, assuming that the weights follow normal distribution with mean and standard deviation of data provided for Females (F).
Test the Hypothesis that μ=7 kgs in R and obtain the p-value using pnorm or otherwise. (4)

v) Additionally, generate datasets for Males and Females of size 10. Use *set.seed(2022)* to set seed value for generating sample data.

Perform paired t test to test the hypothesis that average weights used by Females are 5 kgs less than weights used by Males.
Comment whether average weights used by Females is less than 5kgs or not. (4)
[16]

Q. 2) An investor wants to invest in a portfolio comprising of two stocks. He is evaluating four stocks for the purpose of his investment *viz* ABC Oil, XYZ Airways, PQR Realty and LMN Bank.

Annual returns (%) from these stocks over the past 10 years have been tabulated below –

Year	ABC Oil	XYZ Airways	PQR Realty	LMN Bank
1	-6	8	-20	14
2	-10	0	19	4
3	3	-4	3	5
4	18	10	13	15
5	-10	20	20	19
6	1	8	7	9
7	3	0	-3	6
8	-13	10	13	10
9	-14	5	20	7
10	13	-19	1	16

i) Calculate the mean and standard deviation of the annual returns from each stock using appropriate functions in R. (5)

ii) Calculate the sample correlation coefficient between annual returns from various pairs of stocks viz (ABC, XYZ), (ABC, PQR), (ABC, LMN), (XYZ, PQR), (XYZ, LMN), (PQR, LMN). Use Pearson's method. (3)

iii) The investor feels that in between the following pairs of stocks, there should exist a perfect negative correlation –

a) Changes in oil prices affect the prices of aviation fuel thus affecting aviation profits. Thus, the investor feels that returns from ABC Oil and XYZ Airways should have perfect negative correlation.

b) He also feels that changes in interest rates lead to equal and opposite effects on the profits of the real estate sector and financial services sector. Thus, he is also contemplating a perfect negative correlation between returns from PQR Realty and LMN Bank.

Calculate a 95% two-sided confidence interval for the population correlation coefficient for the above two pairs using Pearson's Method and hence test the validity of the investor's premonitions. [**Hint:** Use **cor.test** function in R] (4)

The investor is considering two alternative investment strategies for setting up his portfolio as follows –

Strategy A: Investing 50% of his funds in ABC Oil and balance 50% in XYZ Airways.

Strategy B: Investing 25% of his funds in LMN Bank and balance 75% in PQR Realty.

He has asked to develop the following metrics to assist with the final choice between Strategy A or Strategy B:

Metric 1: Strategy with highest Mean Returns would be selected.

Metric 2: Strategy with lowest Risk (Lowest Standard Deviation) would be selected.

Metric 3: Strategy with highest "Risk-Adjusted Returns" would be selected.

Risk-Adjusted Return = (Mean Return)/(Standard Deviation of Return)

iv) Calculate the mean return for the above two strategies. Which strategy will be chosen based on Metric 1 i.e. highest Mean Returns? (2)

v) Investor feels that both the strategies are equally risky. Using **var.test** function in R (two-sided), check whether variance of returns from Strategy A is equal to the variance of returns from Strategy B at 10% level of significance. Your answer should clearly mention the null and alternate hypothesis, the p-value and the outcome of the test. [**Hint:** Create two new vectors for strategies A and B using the proportions given earlier.] (4)

Investor has come up with an alternative formula to calculate the standard deviation of both the strategies which is dependent on the correlation coefficients calculated in part (iii).

Using this formula, Standard deviation for a portfolio of two stocks 1 and 2 is to be calculated using the following method:

$$SD(1,2) = (P_1^2 * SD_1^2 + 2\ Corr(1,2) * P_1 * P_2 * SD_1 * SD_2 + P_2^2 * SD_2^2)^{½}$$

Where – P_1 and P_2 denote the proportion of invested funds in stocks 1 and 2 respectively.

vi) Using the confidence interval for population correlation coefficient calculated in part (iii), calculate interval for the standard deviation of returns for Strategy A and Strategy B using the above formula. Which strategy will you choose based on Metric 2 i.e. Lowest Standard Deviation of Returns? (4)

vii) Which strategy outperforms the other in terms of Metric 3 i.e. "Higher Risk-Adjusted Returns"? Use means calculated in part (iv) and interval of standard deviations calculated in part (vi) to calculate interval for risk-adjusted returns.

Also, compare your choices from part (iv), (vi) and (vii) and comment on the same. (2)

[24]

Q. 3) A company wants to launch a new product PA to provide special health cover. Under this health cover, the company will pay customer INR 5000 for each claim. They have asked an Actuary to derive price for the product for different segment of customers.

The Actuary has collected a sample data, named "*PA_Data",* for 20 customers showing number of claims (*Claim*) made by each customer along with the *Gender, Age* and *Health condition* (Diabetic or not) of the customer.

Actuary assumes that the response variable – number of claims (*Claim*) follows Poisson distribution.

i) Load the *"PA_Data"* datset and fit the Poisson Generalised Linear Model (GLM) to the dataset, specified by the following parameter of R code for GLM:
*Claim ~ Gender*Health + Age*
Include summary of the fitted model and label the model as Model1. (4)

ii) Write down the linear predictor for the model given in part (i) (i.e. Model1). Explain all the terms clearly. (3)

iii) Comment on the significance of the parameters of the fitted model in part (i) (i.e. Model1). Also state the scaled deviance and compute Log Likelihood of the model. (4)

iv) Fit another model (labelled as Model2) specified by following R code:
Claim~Health+Age
Demonstrate that Model2 outperforms the model defined in part (i). (2)

v) Examine the significance of Model2 and fit an improved model. Label this model as Model3. Also justify why Model3 is an improvement. (5)

vi) Compute the price for the product for below customers using Model2.
Note that Price is equal to expected aggregate claim amount per customer. (In other words, expected claim amount * expected number of claims)

Customer	Age	Health	Gender
Student1	30	Diabetic	Male
Actuary2	50	Non-Diabetic	Female

Which customer price will be higher under the models. Provide the reason for the same. (7)

Company is planning modification to the product to make it affordable and attractive. Under this modification, INR 6000 will be paid for 1st claim and then INR 4000 for 2nd claim onwards.

Hint: Under this product, essentially INR 4000 amount is paid for each claim plus additional INR 2000 amount if there are claim(s).

vii)

a) Compute the price under this modification for Student1 and Actuary2 (as per details given in part (vi) for Model3. (6)

b) Compare and explain why price for Student1 is lower for modified product whereas for Actuary2 is higher. (2)

[33]

Q. 4) Government of "Actuaria" is operating a popular jackpot lottery called "Power Ball". The game consists of two draws as given below –

Draw 1 = Drawing 5 White Balls from a set of 69 White Balls.

Draw 2 = Drawing 1 Red Ball (Power Ball) from a set of 26 Red Balls (Power Balls).

After drawing 5 white balls and the power ball, a ticket is issued for INR 0.50. A specimen ticket is given below –

GOVERNMENT OF ACTUARIA

"POWER BALL LOTTERY"

Draw 1 – White Balls

05	18	29	37	61

Draw 2 – Power Ball

21

Ticket Price - INR 0.50

The Government of Actuaria publishes its draw of 5 White Balls and the Power Ball. If all the 5 White Balls and the Power Ball drawn by the ticketholder exactly match with the balls drawn by the Government, then the ticketholder is entitled to a jackpot. The amount of jackpot is INR 2 crore. Kindly note that for white balls, the order is not considered relevant while matching.

There are other prizes available depending on the white/power balls matched (notwithstanding the order for white balls). The prize money for various matches can be found in the following table as per the official website of Government of Actuaria: -

Case	White Balls Matched	Power Ball Matched	Prize Money in INR
A	5	1	Jackpot
B	5	0	10,00,000
C	4	1	50,000
D	4	0	100
E	3	1	100
F	3	0	7
G	2	1	7
H	1	1	4
I	0	1	4

Let X denote number of matching draws for White Balls and Y denote number of matching draws for Power Ball.

Under the given circumstances, hyper-geometric distribution is considered suitable to calculate probabilities.

In R, probability of matching white balls / power ball using hyper-geometric distribution is determined using the **dhyper** function.

For instance, under Case E, probability of matching 3 white balls is determined using the following code:

dhyper(3,m=5,n=64,k=5)

Here the arguments of **dhyper** function can be explained as below –
- x denotes the value of white balls to be matched.
- m denotes the maximum number of successes.
- n denotes the maximum number of failures.
- k denotes the total number of draws.

i)

a) Create two separate vectors **w_draw** and **r_draw** for the white balls and power ball to be matched under cases A to I. (2)

b) State the probability mass function of hyper-geometric distribution in terms of the arguments x, m, n and k as defined earlier. (2)

c) Using the vectors created in part (i)(a) and the **dhyper** function as illustrated earlier, create two separate result vectors **prob_w_draw** and **prob_r_draw** showing the probabilities of matching white balls and power ball under Cases A to I. (4)

Probability of matching white balls is independent of the probability of matching the power ball.

d) Calculate the joint probabilities of matching white balls and power balls using result vectors **prob_w_draw** and **prob_r_draw** created in part (i)(c). (2)

e) Create a vector **prize** showing the prize money under cases A to I. (1)

f) Determine the total expected prize money pay-out using the joint probabilities determined in part (i)(d) and the prize money vector created in part (i)(e). (2)

g) Assuming that there are no other costs to the Government in operating the lottery expect the prize money pay-out, calculate the profits, if any, implicit in the ticket price. (1)

ii) If instead of using the hyper-geometric distribution, we decided to use the binomial distribution for modelling the probabilities of matching white balls and power ball.

a) State the probability mass function of binomial distribution in terms of arguments x, m, n and k as defined earlier. (2)

b) Re-determine the joint probabilities using the **dbinom** function and arrive at the total expected prize money pay-out. (4)

iii) Point out the difference, if any, in the outcomes under part (i) and (ii). (2)

iv) Determine the mean and variance of both the binomial distribution and hyper-geometric distribution for random variable X (matching white balls) and comment on the differences in the two. (3)

v) Which distribution is more suitable to model these matching probabilities – hyper-geometric or binomial? Substantiate with reasons. (2)

[27]

****