

INSTITUTE OF ACTUARIES OF INDIA

EXAMINATIONS

November 2025

Subject CS1B - Actuarial Statistics (Paper B)

Time allowed: 1 Hours 45 Minutes

Total Marks: 100

- Q. 1)** A gambler walks into a casino to play a simple gambling game. At each turn, he can either lose ₹400 with probability 0.32, breakeven (win or lose nothing) with probability 0.48, win ₹100 with probability 0.15, or win ₹800 with probability 0.05. Let the discrete random variable X be defined as the “amount earned” at each turn.
- i) Store the vectors of the possible outcomes and corresponding probabilities as `outcomes` and `prob` respectively. (2)
 - ii) Plot a labelled bar graph for the above data. (2)
 - iii) Calculate and print the cumulative probabilities as `cumul`. (2)
 - iv) Plot a suitable labelled line graph for the cumulative probabilities. (2)
 - v) Calculate and print the mean and standard deviation X . (3)
- Every Saturday, at the same time, he stands by the side of a road and tallies the number of cars going by within a 120-minute window. Based on previous knowledge, he believes that the mean number of cars going by during this time is exactly 107. Let Y represent the appropriate Poisson random variable of the number of cars passing his position in each Saturday session.
- vi) What is the probability that more than 100 cars pass him on any given Saturday? (2)
 - vii) Determine the probability that no cars pass. (1)
 - viii) Plot the relevant Poisson mass function over the values in $60 \leq y \leq 150$. (2)
 - ix) Set seed 2025 and simulate 260 results from this distribution (about five years of weekly Saturday monitoring sessions). (1)
 - x) Plot the histogram of simulated results and set the horizontal limits from 60 to 150. Overlay the histogram to the shape of your mass function from (viii). (3)
 - xi) Compare the two plots (viii) and (x) and provide the necessary interpretation. (1)

[21]

- Q. 2)** The annual number of claims X from a particular risk has a Poisson distribution with mean μ . The prior distribution for μ has a gamma distribution with $\alpha=2$ and $\lambda=5$.

Claim numbers over the last 8 years have been recorded. $X = 1, 6, 0, 3, 5, 2, 1, 0$. The posterior for μ distribution is Gamma ($\alpha + \sum x_i, n + \lambda$).

- i) Determine the parameters of the posterior distribution and Bayesian estimate for μ under:
 - a) Squared error loss
 - b) All-or-nothing loss (3)
- ii) Calculate a 95% equal-tailed credible interval for μ by performing χ^2 -Gamma mapping and using a χ^2 distribution. (2)
- iii) Carry out a simulation of $N = 2000$ posterior samples for the parameter μ by setting seed 1234. Print the last 6 simulated values (tail). (2)

- iv) Plot the histogram of the posterior distribution of μ with appropriate labelling. (2)
- v) Two possible values of parameter μ are $\mu=5$ and $\mu=1$. Comment on the feasibility of these values based on the posterior distribution of μ . (3)
- [12]

Q. 3) A dermatologist is interested in the successes in treating a common skin affliction. His records shows 355 patients have been treated at his clinic using one of four possible treatments—a course of tablets, a series of injections, a laser treatment, and an herbal-based remedy. The level of success in curing the affliction is also recorded—none, partial success, and full success and is given below:

	None	Partial	Full
Injection	20	9	16
Tablet	32	72	64
Laser	8	8	30
Herbal	52	32	12

- i) Enter and print the data in the form of a 4x3 matrix assigning appropriate row and column names. (3)
- ii) Perform chi-square test to test if there is any dependence between the level of success and the method of treatment.

You should clearly:

- State the null and alternative hypothesis
 - State the p-value
 - Reject or fail to reject the null hypothesis
 - Make an inference
- (4)

The dermatologist contacts the manufacturer of the tablets to produce a packet that would contain 80grams of the tablet. Assume the manufacturer knows that the total net weight in its package, X , is normally distributed with a mean of 80.2 grams and a standard deviation of 1.1 grams.

- iii) What is the probability that a packet is found to be weigh between 80.5 grams and 81.5 grams? (1.5)
- iv) What is the weight below which the lightest 20% of the packets lie? (1.5)

The dermatologist precisely weighed the contents of 40 randomly selected 80-gram packs from different stores and recorded the weights. He claims that his data cannot have arisen from a distribution with mean $\mu = 80$, so the true mean weight must be less than 80.

- v) Using the vector `weights1` provided in `codes.docx`, conduct a hypothesis test using a significance level of $\alpha = 0.05$ to investigate the dermatologist's claim.

You should clearly:

- State the null and alternative hypothesis
- State the p-value

- c. Reject or fail to reject the test
- d. Make an inference

(4)

- vi) Will your inference change if the test is performed for $\alpha = 0.01$?

(1)

After collecting sample packs from the original manufacturer, he goes out and collects 30 randomly selected 80-gram packs from a rival manufacturer. He is now interested in testing whether there is statistical evidence to support the claim that mean weight of the second pack is greater than mean weight of the first pack.

- vii) The data of the second sample is stored in `weights2` provided in `codes.docx`. Compute the mean of the second sample and compare it with first sample's mean

(2)

- viii) To investigate this claim, conduct a hypothesis test using a significance level of $\alpha = 0.05$.

You should clearly:

- a. State the null and alternative hypothesis
- b. State the p-value
- c. Reject or fail to reject the null hypothesis
- d. Make an inference

(4)

The dermatologist also wants to study the variability in skin thickness due to environmental factors. The following data represent the thickness (in micrometres) of the epidermal layer measured in skin biopsy samples from patients at four different dermatology clinics across India (Clinic I, Clinic II, Clinic III, and Clinic IV)

Clinic I	Clinic II	Clinic III	Clinic IV
93	85	100	96
120	45	75	58
65	80	65	95
105	28	40	90
115	75	73	65
82	70	65	80
99	65	50	85
87	55	30	95
100	50	45	82
90	40	50	
78		45	
95		55	
93			
88			
110			

For ease, these vectors can be found in the file `codes.docx`

- ix) Store these data in your R workspace, with one vector containing skin thickness and the other vector containing the clinic of each observation.

(3)

- x) Produce side-by-side boxplots of the thickness split by group, and use additional points to mark the locations of the sample means.

(5)

xi) Assuming independence, execute diagnostic checks and comment for

a) normality based on QQ plot of the residuals and a normality indicating line. (5)

b) equality of variances. (3)

[Hint: Equality of variance can be assumed if the ratio of the largest sample standard deviation to the smallest is less than 2]

xii) Perform and conclude a one-way ANOVA test for evidence of a difference between the means.

You should:

- State the null and alternative hypothesis
- Print the ANOVA table
- State the p-value
- Reject or fail to reject the null hypothesis
- Make an inference

(5)

[42]

Q. 4) A city planning department wants to analyse how community safety outcome (crime rate) varies with law enforcement resources and socioeconomic conditions.

The department collects:

- Crime Rate: Number of crimes per 100,000 population
- Unemployment: Unemployed people per 100,000 population
- Arms Permit: Firearms licenses issued as a % of overall population
- Urbandensity: Number of residents per square kilometer in urban areas and is given below:

Crimerate	Unemployment	Armspermit	Urbandensity
8.6	260.35	11	178.15
8.9	269.8	7	156.41
8.52	272.04	5.2	198.02
8.89	272.96	4.3	222.1
13.07	272.51	3.5	301.92
14.57	261.34	3.2	391.22
21.36	268.89	4.1	665.56
28.03	295.99	3.9	1131.21
31.49	319.87	3.6	837.6
37.39	341.43	7.1	794.9
46.26	356.59	8.4	817.74
47.24	376.69	7.7	583.17
52.33	390.19	6.3	709.59

i)

a) Load the data `planning.csv` in R containing the above data. (1)

b) Produce a scatterplot matrix. (2)

- c) Calculate and print the correlation between crime rate and other variables. (3)
 - d) Which of the variables appears to be most strongly related to the crime rate? (1)
 - ii) a) Fit a multiple linear regression model using the Crimerate as the response and all other variables as predictors. (2)
 - b) Write down the model equation and interpret the coefficients. (3)
 - c) Identify the amount of variation in the response attributed to the joint effect of the three explanatory variables and comment on the significance of the parameters. (3)
 - iii) a) Refit the model excluding the most nonsignificant predictor and state the equation (3)
 - b) Compare the new coefficient of determination with that of the previous model. (2)
 - iv) a) Use your model from (iii) to predict the mean crime rate per 100,000 residents, with 300 unemployed people per 100,000 residents and 500 residents in sq km area and provide a 99% CI for the prediction. (3)
 - b) What is the number of unemployed people per 100,000 residents so that the mean crime rate is zero ? (2)
- [25]**
