INSTITUTE OF ACTUARIES OF INDIA

EXAMINATIONS

November 2026

CS2B - Risk Modelling and Survival Analysis

Time allowed: 1 Hour 45 Minutes

Total Marks: 100

Q.1) Claim Severity Modelling and Reinsurance

Data file: Q1_claim_severity.csv contains 1,200 individual claim amounts, with variables `ClaimID` and `ClaimAmount`. Claim amounts are expressed in ₹'000. Use R throughout and retain full precision in intermediate calculations.

i) Import Q1_claim_severity.csv into R and define a suitable data object for the claim amounts. Calculate the sample mean, standard deviation and coefficient of skewness of the claim amounts. Produce:

- a histogram of claim amounts; and
- a histogram of log claim amounts.

Comment briefly on the shape and skewness of the claim severity distribution. **[4]**

ii) Fit Lognormal and Weibull distributions to the observed claim amounts using maximum likelihood estimation in R. Report the estimated parameters and Akaike Information Criterion (AIC) for each fitted distribution. **[5]**

iii) Assess the adequacy of the fitted Lognormal and Weibull distributions using the following graphical diagnostics in R:

1. **Density comparison:** Plot the empirical density of the observed claim amounts together with the fitted probability density functions of the Lognormal and Weibull distributions.
2. **Q–Q plots:** Produce separate quantile–quantile (Q–Q) plots comparing the observed claim amounts with the corresponding fitted quantiles from the Lognormal and Weibull distributions.

Comment on how well each distribution fits the observed data, paying particular attention to deviations in the upper tail of the claim severity distribution. Compare the evidence from both graphical diagnostics with the AIC results obtained in part (ii). Select the distribution to be used for subsequent calculations and justify your choice briefly. **[5]**

iv) Using the selected fitted distribution, calculate:

1. the expected claim payment per claim after applying an ordinary deductible of ₹10,000; and
2. the expected recovery per claim from an excess-of-loss reinsurance layer of ₹100,000 excess of ₹50,000.

Present the calculations and results using appropriate R coding. **[5]**

v) Briefly explain how the deductible and excess-of-loss reinsurance layer affect the insurer's retained claim severity and exposure to large claims. **[2]**

vi) Using the selected fitted Lognormal distribution, calculate the 99th percentile of the claim severity distribution. Report the result in ₹ and comment briefly on its interpretation as a high-severity claim threshold. **[4]**

[25 Marks]

Q.2) Time Series: Trend, Seasonality and Forecasting

Data file: Q2_monthly_demand.csv contains monthly demand observations. Use R throughout and retain full precision in intermediate calculations.

i) Import Q2_monthly_demand.csv and construct an appropriate monthly time-series object in R. Plot the observed series and use suitable graphical diagnostics to assess the presence of trend and seasonality. State briefly what the plot indicates. **[3]**

ii) Decompose the time series into its trend, seasonal and irregular components using a suitable decomposition method in R, such as `decompose()` or `stl()`. Extract and plot the estimated seasonal component and comment briefly on the nature and consistency of the seasonal pattern over time. **[5]**

iii) Create appropriate time and month variables in R and fit a linear regression model with a deterministic linear trend and monthly seasonal effects. Report the estimated trend coefficient, its standard error and p-value, and assess whether there is statistically significant evidence of a linear trend in the monthly demand. **[6]**

iv) Using the regression model fitted in part (iii), construct an appropriate six-month future data frame in R and generate forecasts for the next six months with 95% prediction intervals. Present the forecast values and corresponding prediction intervals in a suitable table and plot the historical observations together with the forecasts and prediction intervals. **[5]**

v) Assess the adequacy of the regression model fitted in part (iii). Using R, produce:

- a Ljung–Box test for residual autocorrelation.

Report the test statistic and p-value and comment briefly on whether the residuals provide evidence of remaining serial dependence. **[4]**

vi) Explain why a forecasting model based on a deterministic trend and fixed monthly seasonal effects may become unreliable when the underlying claims or demand environment changes. **[2]**

[25 Marks]

Q.3) Poisson Process and Simulation

Data file: Q3_poisson_events.csv contains the observed event times over a 24-month observation period. Use R throughout and retain full precision in intermediate calculations.

i) Import Q3_poisson_events.csv and estimate the homogeneous Poisson process intensity, λ, per month from the observed event times. Using the fitted intensity, calculate the probability of exactly 5 events occurring during the subsequent three-month period. **[4]**

ii) Using `set.seed(123)` and `rpois()` in R, simulate 10,000 independent three-month event counts using the fitted Poisson intensity. Calculate the simulated mean and variance and compare these with their theoretical values. Present the results in a suitable R object or table. **[4]**

iii) Using the fitted intensity, simulate 20 successive inter-arrival times using the appropriate exponential distribution in R. Convert these inter-arrival times into cumulative event times and produce a step plot showing the cumulative number of events against time. **[4]**

iv) Using R, calculate the observed inter-arrival times from the data. Assess the suitability of the homogeneous Poisson process assumption using appropriate graphical and numerical diagnostics **(****calculate mean, sd, cv, ACF of inter-arrival times, etc.)**, including a suitable plot (Histogram only) and summary statistics. State one brief conclusion regarding the adequacy of the assumption. **[7]**

v) State one suitable insurance application of a homogeneous Poisson process and describe one situation in which the Poisson process assumptions may be inappropriate. **[2]**

vi) Using R, divide the 24-month observation period into four consecutive six-month intervals. Calculate the number of observed events in each interval and present the results in a suitable table and bar plot. Calculate the mean number of events per interval and comment briefly on whether the observed counts appear reasonably stable across the observation period. **[4]**

[25 Marks]

Q.4) Mortality Estimation and Graduation

Data file: Q4_mortality.csv contains Age, Deaths, Exposure and Standard_qx. *Exposure* denotes central exposed-to-risk. Use R throughout and retain full precision in intermediate calculations.

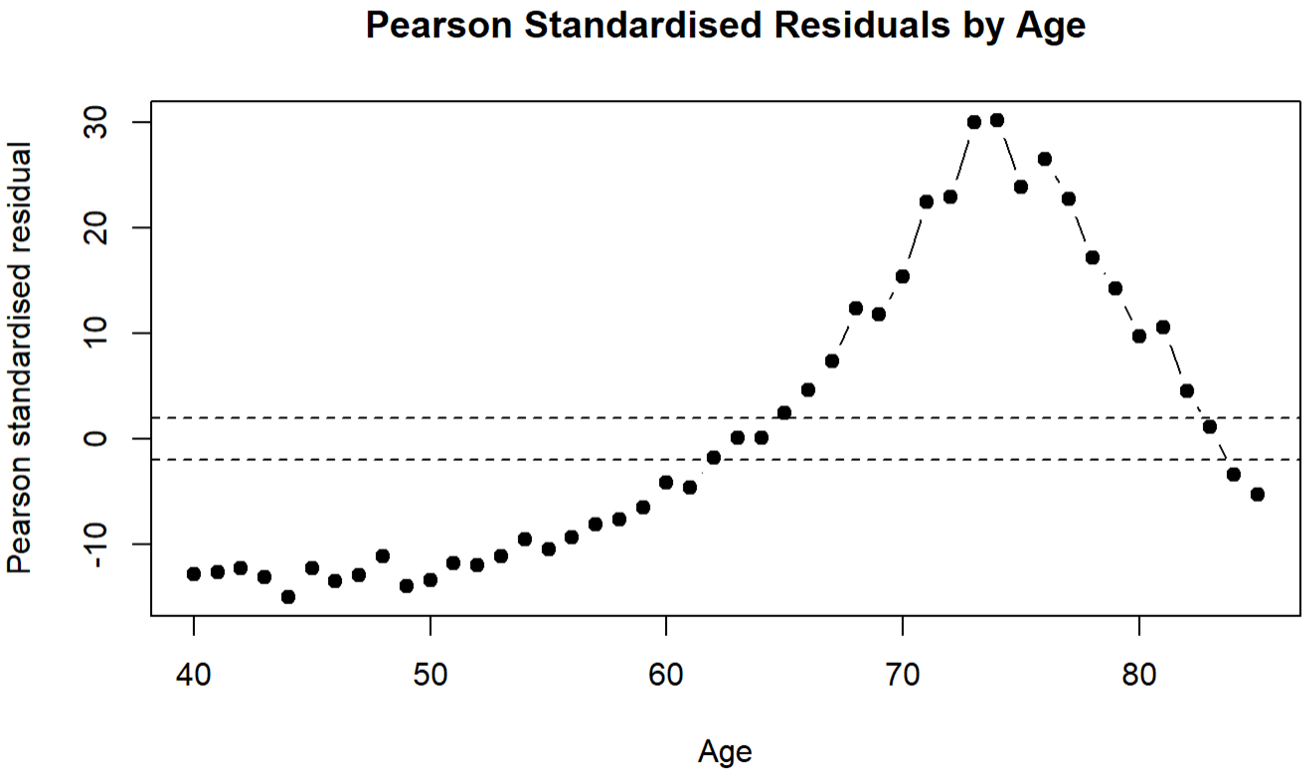
i) Import Q4_mortality.csv and calculate the crude central mortality rate $m_x = Deaths / Exposure$. Assuming deaths are uniformly distributed over each one-year age interval, calculate $q_x = m_x / (1 + 0.5\, m_x)$. Plot the crude q_x values together with Standard_qx. **[5]**

ii) Fit the Gompertz graduation $q_x = B\, c^x$ by nonlinear regression using *nls()* in R. Report the fitted values of B and c, calculate the graduated q_x values for all observed ages, and superimpose the graduated curve on the crude q_x values. **[6]**

iii) Using R, create a comparison table containing the crude, graduated and standard q_x values at ages 60, 70 and 80. Comment briefly on the differences. **[4]**

iv) Using the graduated mortality probabilities (q_x) obtained from the fitted Gompertz model and the appropriate exposure data, calculate the expected deaths at each age. Then calculate the Pearson chi-square goodness-of-fit statistic *sum[(Observed - Expected)^2 / Expected]*, the degrees of freedom allowing for the two fitted Gompertz parameters, and the corresponding p-value. State the null and alternative hypotheses and interpret the test result briefly. **[5]**

v) Using the expected deaths from part (iv), calculate Pearson standardised residuals *(Observed - Expected) / sqrt(Expected)*. Plot of the Pearson standardized residuals against age for the graduated mortality model is provided below, Comment briefly on the adequacy of the graduation, referring to the magnitude and pattern of the residuals and identifying any evidence of systematic departures from zero. **[3]**



vi) Explain why a mortality graduation based only on the age relationship $q_x = B\,c^x$ may be unreliable for projecting mortality rates 20 calendar years into the future. Name one alternative mortality projection model from the syllabus that allows mortality to vary over calendar time. **[2]**

[25 Marks]

****