INSTITUTE OF ACTUARIES OF INDIA

EXAMINATIONS

3rd April 2021

Subject CS1B – Actuarial Statistics (Paper B)

Time allowed: 2 Hour (15.30 – 17.30 Hours)

Total Marks: 100

Q. 1) The following amounts are the sizes of claims (in INR) on house insurance policies for a certain type of repair.

1990, 2400, 2150, 2090, 2300, 2100, 2180, 2150, 2030, 2100, 2180, 2010, 2060, 2160, 2120

i) Enter data in R. (1)

ii) Calculate Q1, Q2, Q3 and Inter-quartile range. (4)

iii) Determine the sample mean and variance of the data. (2)

iv) Test the hypothesis whether the mean claim amount is equal to INR 2000 and comment on the results. (5)

v) Assuming the data to be normally distributed, calculate the probability of a claim amount exceeding INR 2300. (4)

vi) Calculate the revised mean and median after removing the largest two values from the dataset. Comment on the result. (6)

[22]

Q. 2) The prior and posterior distribution for values of systolic blood pressure follows Normal distribution. Prior distribution of systolic blood pressure (x) has a mean of 120 and standard deviation of 10.

i) Generate range of values of x in the interval [80,160] and using len = 100. (3)

ii) Plot the posterior probability density function of x using answer from (i). (6)

iii) Perform a simulation of 1000 posterior samples for the parameter x. (2)

iv) Plot a histogram of the posterior distribution of x. (3)

v) What is the mean and standard deviation of posterior distribution of x? (2)

vi) Calculate the 95% confidence interval for systolic blood pressure using the posterior distribution. (4)

[20]

Q. 3) An agency has collected data on the number of COVID19 cases of two cities in order to analyse the similarities & differences between them. Below is the data for two cities on monthly basis.

Month	City A	City B
1	9150	8919
2	9418	9095
3	9218	9046
4	9539	9321
5	9179	9719
6	8907	9704
7	9472	9107
8	8921	9275

i) Enter data in R. (1)

ii) Test at 5% level with clearly mentioning the hypothesis, if there is a difference in the mean of the two sample data assuming equal & unknown variance. (8)

iii) Test whether the variances are equal at 5% level and comment on the results. (5)

iv) Calculate the 95% confidence interval for the difference in means. (3)

v) Comment on your findings in part (ii) and part (iv). (2)

[19]

Q. 4) Five years of marketing spend and company sales by month

i) Construct a scatterplot of the data. Comment on the relationship between the Sales & Spend based on the plot. (4)

ii) Calculate Pearson's correlation coefficient between Sales and Spend of the company. (2)

iii) Perform a hypothesis test for the null hypothesis that Pearson's population correlation coefficient is equal to zero, against the alternative that it is positive. You should report the p-value of the test and a clear conclusion. (5)

iv) Perform a simple linear regression analysis on the data. Your answer should report the estimate of parameter sigma. (6)

v) Plot the fitted line on the data scatterplot. (2)

vi) State the proportion of the total variability of the responses explained by the model based on your output in (iv). (1)

vii) Plot a graph of the residuals of the model fitted in (iv) against the explanatory variable. (2)

viii) Obtain a 99% confidence interval for parameter sigma. (4)

ix) Comment on the validity of the model based on results in part (vii) and part (viii). (2)

x) Calculate the p-value of a hypothesis test for this suggestion (slope equal to 10), by creating a suitable test statistic. (7)

xi) Comment on the suggestion in point (x). (2)

xii) Calculate the predicted amount of sales when the marketing spend is INR 4500. (2)

[39]

****