

INSTITUTE OF ACTUARIES OF INDIA

EXAMINATIONS

26th July 2022

Subject CS1B – Actuarial Statistics (Paper B)

Time allowed: 2 Hours (14.30 - 16.30 Hours)

Total Marks: 100

Q. 1) Identify the probability distribution that best fits the below questions. Calculate the following using R functions:

- i) Assume golf balls from the driving range next door lands in your yard at an average rate of 3 balls per hour during the day. What is the probability that 10 or fewer golf balls will land in your yard during the afternoon, assuming the afternoon is 5 hours long? (4)
- ii) You are surveying people exiting from a polling booth and asking them if they voted independently. The probability that a person voted independently is 20%. What is the probability that 70 people must be asked before you can find 5 people who voted independently? (5)
- iii) Assume you flip a fair coin 100 times. What is the number N such that, 90% of the time, the number of heads is less than or equal to N ? (4)
- iv) A researcher is waiting outside of a library to ask people if they support a certain law. The probability that a given person supports the law is $p = 0.2$. What is the probability that the fourth person the researcher talks to is the first person to support the law? (4)
- v) Assume that a light bulb has a mean lifetime of 1000 hours. What is the probability that the light bulb survives to 2000 hours? (5)
- vi) Assume a random variable Z is distributed according to the normal distribution with mean 6 and standard deviation 4. What is the probability that Z takes on a value between -1 and 3? (4)

[26]

Q. 2) In a study done at the National Institute of Science and Technology, asbestos fibers on filters were counted as part of a project to develop measurement standards for asbestos concentration.

An operator counted the number of fibers in each of 23 grid squares, yielding the following counts:

31,29,19,18,31,28, 34,27,34,30,16,18, 26,27,27,18,24,22, 28,24,21,17,24

Assume that the Poisson distribution with unknown parameter λ describes the variability from each of the grid squares.

- i) Calculate Q_1 , Q_3 and Inter-quartile range. (4)
- ii) Plot histogram of sample data and label it appropriately. (3)
- iii) Use the method of maximum likelihood to estimate the parameter λ . (3)
- iv) Test the hypothesis whether the mean fiber count is equal to 25. Comment on the results. (5)
- v) Calculate the standard error of parameter λ . (2)
- vi) Calculate the 90% confidence interval for standard error. (3)

- vii) Calculate the probability of fiber count exceeding 30, with the help of Central Limit Theorem.

(4)

[24]

- Q. 3)** An analysis was carried out to investigate the annual average rainfall of two countries. The data is as below:

Iran: 128,125,133,104,146,132,125,118,129,124

Belgium: 160,128,169,105,151,164,162,177,185,150,182,158,156,123,141,176,162,172

- i) Perform a suitable test to determine whether the rainfall in both countries has equal variance or not, at the 5% confidence level.

(6)

- ii) Test whether the mean rainfall in both countries is equal or not, at 5% confidence level.

(6)

- iii) Calculate the 95% confidence interval for difference in means.

(2)

- iv) Comment on the results in part (ii) and (iii).

(3)

[17]

- Q. 4)** The *marketing* dataset contains the impact of three advertising medias (**youtube**, **facebook** and **newspaper**) on sales. The first three columns are the advertising budget in thousands of dollars along with the fourth column as sales. The advertising experiment has been repeated 200 times.

The marketing data is provided in a csv file “data.csv”.

- i) Plot the data. Analyze the trend of how sales varies with the advertising budget of all 3 advertising medias and comment on the same.

(5)

- ii) Perform a simple linear regression analysis on the data. Your answer should include summary of the data.

(6)

- iii) Comment on the significance of the parameters of the model and justify your observations from point (i).

(3)

- iv) Calculate the correlation between independent and dependent variables.

(3)

- v) Fit an improved model for the model in part (ii), using your answer in part (iv). State the linear regression formula clearly explaining all parameters.

(4)

- vi) Which model is better between part (ii) and (v) and why?

(3)

- vii) What is the maximum sales generated?

(2)

- viii) Based on the linear regression model fitted in question (v), what is the predicted value for maximum sales generated in question (vii) .

(4)

- ix) What is the relative error between the estimated (prediction calculated in question (viii)) and the actual sales computed in question (vii)?

(3)

[33]
