MFDA 2021-2022

Problem statements

1. Create a dataframe using all the functionality of R programming for student database consisting of 10 files with attributes studid, studname, marks, gender. Display the summary and structure of student dataframe. Edit the student dataframe with one new file.
   a. Print only marks of student along with studid.

2. Create a .csv and .txt file with 10 records for student with Sr.No., Stdid, StdName. Import the files and print the data. Also explore the publicly available dataset. Download any one dataset as .csv file and export it in a new file.

3. For the given mtcars dataset, find different central tendencies of mpg,cly,gear and disp. Analyse the spread of mpg and cyl data and comment on it. Also implement the basic visualizations techniques for the mpg, cyl, gear and disp data.

4. For the given dataset mtcars:
   i. Find the range for attributes of dataset mtcars.
   ii. Find the interquartile range of mpg in the dataset mtcars.
   iii. What are the quantiles for attribute (wt)?
   iv. What is the 75th and 80th percentiles of the (wt) of the cars?
   v. Create a boxplot graph for the relation between mpg (miles per gallon) and cyl (number of cylinders) and answer the following:
      a. How many numbers of cylinders are needed for lowest mileage per gallon?
      b. If you used this data to predict the mileage per gallon for New Car for all three number of cylinders (4,6,8), which prediction would be you more confident in? Explain.

5. A. Consider a situation at Kerr Pharmacy, where employees are often late. Five workers are in the pharmacy. The owner has studied the situation over a period of time and has determined that there is a 0.4 chance of any one employee being late and that they arrive independently of one another. Draw a binomial probability distribution:
   i. illustrating the probabilities of 0, 1, 2 or 3 workers being late simultaneously

ii. analyse the shape of binomial distribution for:
    a. n=5, p=0.1,0.3,0.5,0.9 (constant n = 5 and varying p)
    b. p=0.4, n=5,10, 20(constant p and n is increased)
iii. illustrating the probabilities of at least 3 workers being late
iv. between 1 and 4 workers

B. Suppose that we are investigating the safety of a dangerous intersection. Past police records indicate a mean of five accidents per month at this intersection. The number of accidents is distributed according to a Poisson distribution, and the Highway Safety Division wants to calculate and draw the:

i. probability in any month of exactly 0, 1, 2, 3, or 4 accidents.
ii. the probability of having 7 or less accidents at intersection in a particular minute
iii. the probability of having 7 or more accidents at intersection in a minute
iv. the probability between 4 and 8 accidents
v. plot density of the Poisson distribution for x=1 to 50 and lambda=3. plot

6.

A. Create a sample of 50 numbers which are normally distributed.

B. Calculate the probability that a normal random variable with mean 22 and 29

i. lies between 16.2 and 27.5
ii. is less than 17
iii. is less than 15 or greater than 25

C. Dennis Hogan is the supervisor for the Conowingo Hydroelectric Dam. Mr. Hogan knows that the dam's turbines generate electricity at the peak rate only when at least 1,000,000 gallons of water pass through the dam each day. He also knows, from experience, that the daily flow is normally distributed, with the mean equal to the previous day's flow and a standard deviation of 200,000 gallons. Yesterday, 850,000 gallons flowed through the dam. What is the probability that the turbines will generate at peak rate today?

7.

Run R code online - RDRR.io
https://rdrr.io/snippets/

a. Implement cross product of any given two vectors and two matrices.

b. Implement dot products of two given vectors and 2D arrays.
c. Implement various distance metrics like Euclidian, Jaccard, Hamming, Manhattan for any given vectors.
d. Check whether the given matrix is orthogonal or not.

$$A= \begin{matrix} 0.68567 & 0.12975 & -0.71626 \\ 0.14807 & 0.93855 & 0.31176 \\ 0.71269 & -0.31982 & 0.62433 \end{matrix}$$

e. For any given square and symmetrical matrix, $A=\begin{matrix} 13 & -4 & 2 \\ -4 & 11 & -2 \\ 2 & -2 & 8 \end{matrix}$
get eigen values and eigen vectors. Also, implement and verify that:
   I. sum of squares of A = sum of squares of eigenvalues
   II. determinant of matrix = product of eigenvalues
   III. rank of a matrix= number of non-zero eigenvalues

8. Implement the sampling distribution for n=1000.
   i. Generate a sampling distribution
   ii. Visualize the sampling distribution
   iii. Calculate the mean and standard deviation of the sampling distribution
   iv. Calculate probability that sample mean is less than or equal to 6.

9.