Cycle 1 (R LAB)

1. The file 'smoker.csv' contains results of a survey conducted on 356 persons in a city. Each record contains the ID of the person, smoking habit (current-current smoker, former-former smoker, never-never a smoker) and his economic status(Low,middle,high). Write an R program to read the data and construct a two-way table of smoking status versus economic status. Also calculate the proportion of people belonging to each category. You may use the function `table()` to accomplish this. Write the output to a text file.

2. Simulate the outcome of tossing a fair die 10000 times and create the probability distribution of the outcomes. You may use the built-in function `sample()` to do the simulation. Plot the probability function and check whether the die is fair.

3. write an R program to find the sum of the first n terms of the Fourier series
$$\frac{\pi}{2} + 2\left[\frac{\sin x}{1} + \frac{\sin 3x}{3} + \frac{\sin 5x}{5} + \cdots\right]$$
The program should include a function to evaluate the sum with n as an argument. Also plot the graphs of the sum for $n = 3, 10, 20$ and 50.

4. File 'sales.txt' contains data of price, discount factors and sales volume (number of units sold) of 3 items in three different Indian States. Read the data and compute the total sales bill of each item in each state. write the output with details in an output file

5. Simulate a random list of 1000 patients, each belonging to one of the age groups 40-45,45-0,50-55,50-55,55-60, and each having BP level one of "low","med", or "high" using the data prepare a table indicating the percentage of people in each category.

6. The data file "price.txt" contains the price (in Rs.) of 3 items for 100 consecutive days. Read the data into a data frame.

   (a) Calculate the mean, standard deviation, quartiles and quartile deviation of the prices of each item and append these information to the end of the data frame. You should use the built-in function `apply()` to do the calculations. write the output to a file.

(b) Plot the graphs of the time series of the prices of each item (plot of value versus time).

7. The following time series shows the sales of a particular product over the past 12 months

Month	1	2	3	4	5	6	7	8	9	10	11	12
Sales	105	135	120	105	90	120	145	140	100	80	100	110

(a) Draw a scatter plot of month verses sales

(b) Calculate the third order moving averages of the above data. The method of calculation of moving averages is explained below

The third order moving averages of a time series

Month	t_1	t_2	t_3	t_4	$\cdots$	t_n
Sales	y_1	y_2	y_3	y_4	$\cdots$	y_n

are calculated by

Month	t_1	t_2	t_3	t_4	$\cdots$	t_{n-1}	t_n
Moving Average	y_1	$\frac{y_1+y_2+y_3}{3}$	$\frac{y_2+y_3+y_4}{3}$	$\frac{y_3+y_4+y_5}{3}$	$\cdots$	$\frac{y_{n-2}+y_{n-1}+y_n}{3}$	y_n

(c) Plot the moving averages as a line graph to display the smoothed data in the scatter plot drawn in part (a)

8. The following data gives the height (x in cm) and weight (y in kg.) of 10 persons from a locality.

Height (x):	151	174	138	186	128	136	179	163	152	131
Weight (y):	63	81	56	91	47	57	76	72	62	48

(a) Use the `lm()` function to construct a regression model line of y on x. print the summary of the relationship.

(b) Find the predicted weights according to the fitted model

(c) Find the residuals of the regression model

(d) Using the `predict()` function evaluate the weights when the height is 170 cm and 180 cm

(e) Draw a scatter plot of the data along with the regression line.

9. Consider the data set "mtcars" available in the R environment. It gives a comparison between different car models in terms of mileage per gallon (mpg), cylinder displacement("disp"), horse power("hp"), weight of the car("wt") and some more parameters.

(a) Construct a multiple regression model of the form $y = a_0 + a_1 x_1 + a_2 x_2 + a_3 x_3$ with "mpg" as a response variable y and "disp","hp" and "wt" as predictor variables x_1, x_2 and x_3.

(b) What are the values of the coefficients $a_0, \ldots, a_3$ in the fitted model ?

(c) Find the values of "mpg" predicted by the fitted model

(d) Find the residuals of the regression model.

(e) Using the `predict()` function or otherwise evaluate the predicted mileage for a car with disp $= 221$, hp $= 102$ and wt $= 2.91$