MODEL PRACTICAL EXAMINATION

Course Code	ITA0480	Course Title	Statistics with R Programming for Health Care Applications	
Branch	**B.Tech- CSE/IT/AI&DS/AI&ML/BI**	**Year\Academic year**	**I, II, III, IV Year/2023-24**	
Date		**Time:**	**Marks**	**100**

SET-1

1. a. Create multi regression model to find a weight of the chicken , by "Time" and "Diet" as as predictor variables

    b. Predict weight for Time=10 and Diet=1

    c. Find the error in model for same

2. (i)Get the Summary Statistics of air quality dataset

    (ii) Melt airquality data set and display as a long – format data?

    (iii) Melt airquality data and specify month and day to be "ID variables"?

    (iv)Cast the molten airquality data set with respect to month and date features

    (v) Use cast function appropriately and compute the average of Ozone, Solar.R , Wind and temperature per month?

3. Randomly Sample the iris dataset such as 80% data for training and 20% for test and create Logistics regression with train data, use species as target and petals width and length as feature variables , Predict the probability of the model using test data, Create Confusion matrix for above test model.

4. Write a R program to create an array using four given columns, three given rows, and two given tables and display the content of the array.

Course Code	ITA0480	Course Title	Statistics with R Programming for Health Care Applications		
Branch	**B.Tech- CSE/IT/AI&DS/AI&ML/BI**	**Year\Academic year**	**I, II, III, IV Year/2023-24**		
Date		**Time:**		**Marks**	**100**

SET-2

1. Create below data frame exam_data = data.frame(name = c('Anastasia', 'Dima', 'Katherine', 'James', 'Emily', 'Michael', 'Matthew', 'Laura', 'Kevin', 'Jonas'), score = c(12.5, 9, 16.5, 12, 9, 20, 14.5, 13.5, 8, 19), attempts = c(1, 3, 2, 3, 2, 3, 1, 1, 2, 1), qualify = c('yes', 'no', 'yes', 'no', 'no', 'yes', 'yes', 'no', 'no', 'yes'))

2. Write a R program to create a list of elements using vectors, matrices and a function. Print content of the list.

3. Explore the USArrests dataset, contains the number of arrests for murder, assault, and rape for each of the 50 states in 1973. It also contains the percentage of people in the state who live in an urban area.

   (i)   a. Explore the summary of Data set, like number of Features and its type. Find the number of records for each feature. Print the statistical feature of data b. Print the state which saw the largest total number of rape c. Print the states with the max & min crime rates for murder

   (ii)   a. Find the correlation among the features
           b. Print the states which have assault arrests more than median of the country
           c. Print the states are in the bottom 25% of murder

   (iii). a. Create a histogram and density plot of murder arrests by US stat
           b. Create the plot that shows the relationship between murder arrest rate and proportion of the population that is urbanized by state. Then enrich the chart by adding assault arrest rates (by coloring the points from blue (low) to red (high)).

           c. Draw a bar graph to show the murder rate for each of the 50 states

4.   Write a R program to read the .csv file and display the content

Course Code	**ITA0480**	**Course Title**	**Statistics with R Programming for Health Care Applications**		
Branch	**B.Tech- CSE/IT/AI&DS/AI&ML/BI**	**Year\Academic year**	**I, II, III, IV Year/2023-24**		
Date		**Time:**	**Marks**	**100**	

SET-3

1. Write a R program to create three vectors numeric data, character data and logical data.Display the content of the vectors and their type

2. Create below data frame exam_data = data.frame(name = c('Anastasia', 'Dima', 'Katherine','James', 'Emily', 'Michael', 'Matthew', 'Laura', 'Kevin', 'Jonas'), score = c(12.5, 9, 16.5, 12, 9, 20, 14.5, 13.5, 8, 19), attempts = c(1, 3, 2, 3, 2, 3, 1, 1, 2, 1), qualify = c('yes', 'no', 'yes', 'no','no', 'yes', 'yes', 'no', 'no', 'yes'))
a. Write a R program to extract 3rd and 5th rows with 1st and 3rd columns from a given dataframe
b. Write a R program to add a new column named country in a given data frame Country<-
c("USA","USA","USA","USA","UK","USA","USA","India","USA","USA")
c. Write a R program to add new row(s) to an existing data frame new_exam_data = data.frame(name = c('Robert', 'Sophia'),score = c(10.5, 9), attempts = c(1, 3),qualify = c('yes','no'))
d. Write a R program to sort a given data frame by name and score e. Write a R program tosave the information of a data frame in a file and display the information of the file.

3. Write a R program to create a factor corresponding to height of women data set , whichinbuild in R, contains height and weights for a sample of women.

4. Write a R program to create a 5 x 4 matrix , 3 x 3 matrix with labels and fill the matrix byrows and 2×2 matrix with labels and fill the matrix by columns

Course Code	ITA0480	Course Title	Statistics with R Programming for Health Care Applications		
Branch	B.Tech- CSE/IT/AI&DS/AI&ML/BI	**Year\Academic year**	I, II, III, IV Year/2023-24		
Date		**Time:**		**Marks**	**100**

SET-4

1. Write a R program to create an array, passing in a vector of values and a vector of dimensions. Also provide names for each dimension

2. Write a R program to call the (built-in) dataset airquality. Check whether it is a data frame or not? Order the entire data frame by the first and second column. remove the variables 'Solar.R' and 'Wind' and display the data frame

3. Load dataset named ChickWeight,

   (i). Order the data frame, in ascending order by feature name "weight" grouped by feature "diet" and Extract the last 6 records from order data frame.
   (ii).a Perform melting function based on "Chick", "Time", "Diet" features as ID variables
      b. Perform cast function to display the mean value of weight grouped by Diet
      c. Perform cast function to display the mode of weight grouped by Diet

4. Randomly Sample the iris dataset such as 80% data for training and 20% for test and create Logistics regression with train data, use species as target and petals width and length as feature variables , Predict the probability of the model using test data, Create Confusion matrix for above test model

Course Code	ITA0480	Course Title	Statistics with R Programming for Health Care Applications		
Branch	B.Tech- CSE/IT/AI&DS/AI&ML/BI	Year\Academic year		I, II, III, IV Year/2023-24	
Date		Time:		Marks	100

SET-5

Consider the following data present. Create this file using Windows Notepad. Save the file as input.csv using the save As All files(*.*) option in Notepad.

i.Use appropriate R commands to read the input.csv file.

ii. Analyze the CSV File and compute the following.

a. Get the maximum salary

b. Get the details of the person with max salary

c. Get all the people working in IT department

d. Get the persons in IT department whose salary is greater than 600

e. Get the people who joined on or after 2014

2.Write a R program to extract the five of the levels of factor created from a random sample from the LETTERS (Part of the base R distribution).

3.Write a R program to create a vector which contains 10 random integer values between - 50 and +50.

4.Write a R program to print the numbers from 1 to 100 and print "Fizz" for multiples of 3, print "Buzz" for multiples of 5, and print "FizzBuzz" for multiples of both.

Course Code	**ITA0480**	**Course Title**	**Statistics with R Programming for Health Care Applications**		
Branch	**B.Tech- CSE/IT/AI&DS/AI&ML/BI**	**Year\Academic year**		**I, II, III, IV Year/2023-24**	
Date		**Time:**		**Marks**	**100**

SET- 6

1. . Write a R program to find the factors of a given number

2. Write a R program to create a list of random numbers in normal distribution and count occurrences of each value.

3. For this exercise, use the (built-in) dataset Titanic.

   a. Draw a Bar chart to show details of "Survived" on the Titanic based on passenger Class
   b. Modify the above plot based on gender of people who survived
   c. Draw histogram plot to show distribution of feature "Age"

4. Write a R program to combine three arrays so that the first row of the first array is followed by the first row of the second array and then first row of the third array

Course Code	ITA0480	Course Title	Statistics with R Programming for Health Care Applications		
Branch	**B.Tech- CSE/IT/AI&DS/AI&ML/BI**	**Year\Academic year**	**I, II, III, IV Year/2023-24**		
Date		**Time:**		**Marks**	**100**

SET-7

1. Write a R program to create an array of two 3x3 matrices each with 3 rows and 3 columns from two given two vectors. Print the second row of the second matrix of the array and the element in the 3rd row and 3rd column of the 1st matrix.

2. Write a R program to create an array with three columns, three rows, and two "tables", taking two vectors as input to the array. Print the array

3. Explore the airquality dataset. It contains daily air quality measurements from New York during a period of five months:
   • Ozone: mean ozone concentration (ppb),
   • Solar.R: solar radiation (Langley),
   • Wind: average wind speed (mph),
   • Temp: maximum daily temperature in degrees Fahrenheit,
   • Month: numeric month (May=5, June=6, and so on),
   • Day: numeric day of the month (1- 31).
   i.   Compute the mean temperature(don't use build in function)
   ii. Extract the first five rows from airquality.
   iii. Extract all columns from airquality except Temp and Wind
   iv. Which was the coldest day during the period?
   v.   How many days was the wind speed greater than 17 mph?

4. Write a R program to sort a given data frame by name and score.

Course Code	ITA0480	Course Title	Statistics with R Programming for Health Care Applications	
Branch	B.Tech- CSE/IT/AI&DS/AI&ML/BI	Year\Academic year	I, II, III, IV Year/2023-24	
Date		Time:	Marks	100

SET-8

1. Write a R program to draw an empty plot and an empty plot specify the axes limits of the graphic

2. Write a R program to create an array, passing in a vector of values and a vector of dimensions. Also provide names for each dimension.

3. (i)   Get the Summary Statistics of air quality dataset

   (iii)   Melt airquality data set and display as a long – format data?

   (iv)   Melt airquality data and specify month and day to be "ID variables"?

   (v)   Cast the molten airquality data set with respect to month and date features

   (vi)   Use cast function appropriately and compute the average of Ozone, Solar.R , Wind and temperature per month?

4. . a. Create a data frame based on below table

Month	1	2	3	4	5	6	7	8	9	10	11	12
Spends	1000	4000	5000	4500	3000	4000	9000	11000	15000	12000	7000	3000
Sales	9914	40487	54324	50044	34719	42551	94871	118914	158484	131348	78504	36284

Create a regression model for that data frame table to show the amount of sales(Sales) based on the how much the company spends (Spends) in advertising c. Predict the Sales if Spend=13500

Course Code	ITA0480	Course Title	Statistics with R Programming for Health Care Applications		
Branch	B.Tech- CSE/IT/AI&DS/AI&ML/BI	Year\Academic year	I, II, III, IV Year/2023-24		
Date		Time:	Marks	100	

SET-9

1. Write a R program to create a factor corresponding to height of women data set , which inbuild in R, contains height and weights for a sample of women.

2. Write a R program to extract the five of the levels of factor created from a random sample from the LETTERS (Part of the base R distribution.

3. Write a R program to create an array of two 3x3 matrices each with 3 rows and 3 columns from two given two vectors. Print the second row of the second matrix of the array and the element in the 3rd row and 3rd column of the 1st matrix.

4. For this exercise, use the (built-in) dataset Titanic. a. Draw a Bar chart to show details of "Survived" on the Titanic based on passenger Class b. Modify the above plot based on gender of people who survived c. Draw histogram plot to show distribution of feature "Age

MODEL PRACTICAL EXAMINATION

Course Code	**ITA0480**	**Course Title**	**Statistics with R Programming for Health Care Applications**	
Branch	**B.Tech- CSE/IT/AI&DS/AI&ML/BI**	**Year\Academic year**	**I, II, III, IV Year/2023-24**	
Date		**Time:**	**Marks**	**100**

SET-10

1. Explore the USArrests dataset, contains the number of arrests for murder, assault, and rape for each of the 50 states in 1973. It also contains the percentage of people in the state who live in an urban area.

(i) a. Explore the summary of Data set, like number of Features and its type. Find the number of records for each feature. Print the statistical feature of data

b. Print the state which saw the largest total number of rape c. Print the states with the max & min crime rates for murder

(ii).   a. Find the correlation among the features

   b. Print the states which have assault arrests more than median of the country

   c. Print the states are in the bottom 25% of murder

(iii).   a. Create a histogram and density plot of murder arrests by US stat

   b. Create the plot that shows the relationship between murder arrest rate and proportion of the population that is urbanised by state. Then enrich the chart by adding assault arrest rates (by colouring the points from blue (low) to red (high)).

   c. Draw a bar graph to show the murder rate for each of the 50 states .

 2. Randomly Sample the iris dataset such as 80% data for training and 20% for test and create Logistics regression with train data, use species as target and petals width and length as feature variables, Predict the probability of the model using test data, Create Confusion matrix for above test model

3. Write a R program to save the information of a data frame in a file and display the information of the file.

4. Write a R program to call the (built-in) dataset airquality. Check whether it is a data frame or not? Order the entire data frame by the first and second column. remove the variables 'Solar.R' and 'Wind' and display the data frame