Day 5- Model

1. 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  
  
  
 2.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  
(iii)a.  Create Box plot for “weight” grouped by “Diet”  
b. Create a Histogram for “weight” features belong to Diet- 1 category  
c.  Create Scatter plot for “ weight” vs “Time” grouped by Diet  
(iv)  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

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 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 .    
  
  
4. Using linear regression analysis establish a relationship between height and weight of a person using the input vector given below.  
# Values of height  
151, 174, 138, 186, 128, 136, 179, 163, 152, 131  
# Values of weight.  
63, 81, 56, 91, 47, 57, 76, 72, 62, 48  
  
Predict the weight of a person with height 170 and Visualize the regression graphically

5.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”  
  
  
   
  
  
6 a. Create a data frame based on below table. Month

|  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| 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 |

b. 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

7.a. Create a 6 × 10 matrix of random integers chosen in the range of from 1:10  
b.  Find the number of entries in each row which are greater than 4.  
c.  Which rows contain exactly two occurrences of the number 7?

8.Suppose you track your commute times for two weeks (10 days) and you find the following  
times in minutes 17 16 20 24 22 15 21 15 17 22 Enter this into R  
a. create function “maxi” to find the longest commute time, the function “avger” to find the average and the function “mini” to find the minimum.  
b. Oops, the 24 was a mistake. It should have been 18. How can you fix this? Do so, and  
then find the new average.  
c. How many times was your commute 20 minutes or more?

9.a. Create a 3x4 matrix with 12 random numbers between 1-100; have the matrix be filled our row-by-row, instead  of column-by-column. Name the columns of the matrix uno, dos, tres, cuatro, and the  rows x, y,  z. Scale the matrix by 10 and save the result.  
b. Extract the column called “uno” as a vector from the original matrix and save the result

c. Extract the row called ‘y’ as a vector from the original matrix and print thesum of the vector.  
  
  
   
  
  
10. a. Write suitable R code to compute the mean, median ,mode of the following values  
 c(90, 50, 70, 80, 70, 60, 20, 30, 80, 90, 20,75,70,10,60,70,85,95,55,15)        
 b. Write R code to find 2nd  highest and  4th Lowest value of above problem.  
  
  
11. a. Write a program for creating a pie-chart in R using the input vector (21,62,10,53). Provide labels for the chart as ‘London’, ‘New York’, ‘Singapore’, ‘Mumbai’. Add a  
title to the chart as ‘city pie-chart’ and add a legend at the top right corner of the chart.

b. Write a program for creating a bar chart using the vectors H=c(7,12,28,3,41) and M=c(“mar”, “apr”, “may”, “jun”, “jul”). Add a title to the chart as “Revenue chart”  
  
  
 c. Creat histogram plot for  data below with suitable bin-width  
v <- c(19, 23, 11, 5, 16, 21, 32, 14, 19, 27, 39, 120, 40, 70, 90)  
  
  
12. The table above shows one year of marketing spend and company sales by month.

|  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| 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 to show the amount of sales(Sales) based on the how much the company spends (Spends) in advertising. Predict the Sales if Spend=13500.