

R Programming Assignment-7

1. In R programming, when should the mode() function be used?

ANS: Mode:

The mode is the value that has highest number of occurrences in a set of data. Unlike mean and median, mode can have both numeric and character data.

R does not have a standard in-built function to calculate mode. So we create a user function to calculate mode of a data set in R. This function takes the vector as input and gives the mode value as output.

Example:

```
# Create the function.
getmode <- function(v) {
  uniqv <- unique(v)
  uniqv[which.max(tabulate(match(v, uniqv)))]
}

# Create the vector with numbers.
v <- c(2,1,2,3,1,2,3,4,1,5,5,3,2,3)

# Calculate the mode using the user function.
result <- getmode(v)
print(result)

# Create the vector with characters.
charv <- c("o","it","the","it","it")

# Calculate the mode using the user function.
result <- getmode(charv)
print(result)
```

When we execute the above code, it produces the following result –

```
[1] 2  
[1] "it"
```

2. Why are factor variables useful?

ANS: Factors are the data objects which are used to categorize the data and store it as levels. They can store both strings and integers. They are useful in the columns which have a limited number of unique values. Like "Male", "Female" and True, False etc. They are useful in data analysis for statistical modeling

Factors are created using the **factor ()** function by taking a vector as input.

Generating Factor Levels:

We can generate factor levels by using the **gl()** function. It takes two integers as input which indicates how many levels and how many times each level.

Syntax:

```
gl(n, k, labels)
```

Following is the description of the parameters used –

- **n** is a integer giving the number of levels.
- **k** is a integer giving the number of replications.
- **labels** is a vector of labels for the resulting factor levels.

3. Is it possible to make a decision tree with R?

ANS: Decision tree is a graph to represent choices and their results in form of a tree. The nodes in the graph represent an event or choice and

the edges of the graph represent the decision rules or conditions. It is mostly used in Machine Learning and Data Mining applications using R.

Examples of use of decision tress is – predicting an email as spam or not spam, predicting of a tumor is cancerous or predicting a loan as a good or bad credit risk based on the factors in each of these. Generally, a model is created with observed data also called training data. Then a set of validation data is used to verify and improve the model. R has packages which are used to create and visualize decision trees. For new set of predictor variable, we use this model to arrive at a decision on the category (yes/No, spam/not spam) of the data.

The R package "**party**" is used to create decision trees.

4.Can you write an R programme to print the total and mean of a vector?

ANS:

R Programming Code :

```
nums = c(10, 20, 30)
print('Original vector:')
print(nums)
print(paste("Sum of vector elements:",sum(nums)))
print(paste("Mean of vector elements:",mean(nums)))
```

OUTPUT:

```
[1] "Original vector:"
[1] 10 20 30
[1] "Sum of vector elements: 60"
[1] "Mean of vector elements: 20"
```

5.How are Lists and Arrays different?

ANS:A **list** can hold items of different types and the list size can be increased on the fly. List contents can be accessed either by index (like `mylist[[1]]`) or by name (like `mylist$page`).

An **array** is a vector with one or more dimensions. So, an array with one dimension is (almost) the same as a vector. An array with two dimensions is (almost) the same as a matrix. An array with three or more dimensions is an n-dimensional array.

Similarities between Lists and Arrays:

- Both are used for storing data
- Both are mutable
- Both can be indexed and iterated through
- Both can be sliced

Differences:

The main difference between these two data types is the operation you can perform on them. Arrays are specially optimised for arithmetic computations so if you're going to perform similar operations you should consider using an array instead of a list.

Also lists are containers for elements having differing data types but arrays are used as containers for elements of the same data type.

The example below is the result of dividing an array by a certain number and doing the same for a list. When we try the same operation

(example: division) on a list, we get a `TypeError` because builtin python lists do not support the `__div__` protocol. It takes an extra step to perform this calculation on a list because then you'd have to loop over each item one after the other and save to another list.

6.What role does R play in predictive analytics?

ANS: Predictive analysis in R Language is a branch of analysis which uses statistics operations to analyze historical facts to make predict future events. It is a common term used in data mining and machine learning. Methods like time series analysis, non-linear least square, etc. are used in predictive analysis. Using predictive analytics can help many businesses as it finds out the relationship between the data collected and based on the relationship, the pattern is predicted. Thus, allowing businesses to create predictive intelligence.

In this article, we'll discuss the process, need and applications of predictive analysis with example codes.

Process of Predictive Analysis:

Predictive analysis consists of 7 processes as follows:

- Define project: Defining the project, scope, objectives and result.
- Data collection: Data is collected through data mining providing a complete view of customer interactions.
- Data Analysis: It is the process of cleaning, inspecting, transforming and modelling the data.
- Statistics: This process enables validating the assumptions and testing the statistical models.
- Modelling: Predictive models are generated using statistics and the most optimized model is used for the deployment.

- Deployment: The predictive model is deployed to automate the production of everyday decision-making results.
- Model monitoring: Keep monitoring the model to review performance which ensures expected results.

Need of Predictive Analysis:

Understanding customer behavior: Predictive analysis uses data mining feature which extracts attributes and behavior of customers. It also finds out the interests of the customers so that business can learn to represent those products which can increase the probability or likelihood of buying.

Gain competition in the market: With predictive analysis, businesses or companies can make their way to grow fast and stand out as a competition to other businesses by finding out their weakness and strengths.

Learn new opportunities to increase revenue: Companies can create new offers or discounts based on the pattern of the customers providing an increase in revenue.

Find areas of weakening: Using these methods, companies can gain back their lost customers by finding out the past actions taken by the company which customers didn't like.

7.Create a simple Data Frame using the software?

ANS:

R PRAGRAMMING CODE:

```
Name <- c("Jon", "Bill", "Maria", "Ben", "Tina")
```

```
Age <- c(23, 41, 32, 58, 26)
```

```
df <- data.frame(Name, Age)
```

```
print (df)
```

OUTPUT:

Name	Age
Jon	23
Bill	41
Maria	32
Ben	58
Tina	26

