SA 0 – Questions and Solutions

Q1. Data structures in R programming are tools for holding multiple values. Which of the following does NOT produce a data structure in R?

a) list()
b) factor()
c) matrix()
d) struct()

ans. d) struct()

Q2. If X, Y and Z are random variables. X and Y are positively correlated, and Y and Z are positively correlated. What is the relationship between X and Z?

a) negatively correlated
b) positively correlated
c) not correlated
d) can't say

ans. d) can't say

Q.3 Which function in R is used to combine vectors of the same size by rows? It also works for matrices and even data frames

a) cbind()
b) rbind()
c) dim()
d) list()

ans. b) rbind()

Q.4 If my_vec is a vector in R, what is the output of my_vec[-1] ?

a) Returns last item in my_vec
b) Returns everything except the last item in my_vec
c) Returns everything except the first item in my_vec
d) Returns first item in my_vec

ans. c) Returns everything except the first item in my_vec

Q.5 Suppose that U is the variable made of the 10 dots on the blue line with equation (y = 0.3 x +1) , and V is the variable made of the 10 dots on the red line with equation (y = 1.2 x + 2) in the graph below. What is the correlation of U and V?

a) 0 (uncorrelated)
b) 0.3 (partially correlated)

c) 0.6 (partially correlated)
d) 1 (perfectly correlated)

ans. d) 1 (perfectly correlated)

Q.6 A factor object in R can

a) Store only numeric values
b) Store only character values
c) Store only integer values
d) Store both numeric and character values

ans. d) Store both numeric and character values

Q.7 which of the following creates string variables from numeric variables?

a) as.character()
b) character()
c) as.string()
d) str()

ans. a) as.character()

Q.8 Which of the following is a supervised learning method?
a) Linear Regression
b) Cluster Analysis
c) Logistic Regression
d) Classification using K-nearest neighbors

ans. a,c,d

Q.9 lapply function in R is used to?

a) Apply a function over a list, vector and dataframe and returns a vector
b) Apply a function over a list, vector and dataframe and returns a list
c) Apply a function over a dataframe or a matrix and returns a list or a vector
d) Infer from the data, apply a model to it, and return a predicted value

Ans. b) Apply a function over a list, vector and dataframe and returns a list

Q.10 We collect a set of data on the top 500 firms in the US. For each firm we record sales, profit, number of employees, industry, and the CEO salary. We'd like to understand which factors affect CEO salaries.

a) Classification problem. Inference. n = 500 x number of employees. p = five
b) Regression problem. Inference. n = 500. p = four

c) Classification problem. Prediction. n = 500 x number of industries. p = five
d) Regression problem. Prediction. n = 500. p = four

Ans. b) Regression problem. Inference. n = 500. p = four

Q.11 The marketing division of a company is considering launching a new product and wishes to know whether it will be a success or a failure. It collects data on 20 similar products that were previously launched. For each product, it records whether it was a success or failure, price charged for the product, length of time on the market, marketing budget, volume of sales, and ten other variables.

a) Classification. Prediction. n = 20, p = 14
b) Classification. Prediction. n = 14, p = 20
c) Regression. Inference. n = 20, p = 14
d) Regression. Prediction. n = 20, p = 14

Ans. a) Classification. Prediction. n = 20, p = 14

Q.12 We would like to know the percentage change in the USD/Euro exchange rate in relation to the weekly changes in the world stock markets. Hence we collect weekly data for all of 2019. For each week, we record the percentage change in the USD/Euro, the % change in the US stock market, the % change in the British market, the % change in the French market, and the % change in the German market.

a) Classification. Inference. n= 365. p = five
b) Classification. Prediction. n= 52. p = five
c) Regression. Inference. n= 50. p = four
d) Regression. Prediction. n= 52. p = four

Ans. d) Regression. Prediction. n= 52. p = four