

PSG COLLEGE OF ARTS & SCIENCE  
(AUTONOMOUS)

BSc DEGREE EXAMINATION MAY 2022  
(Sixth Semester)

Branch – COMPUTER SCIENCE

DATA SCIENCE WITH R PROGRAMMING

Time: Three Hours

Maximum: 75 Marks

SECTION-A (10 Marks)

Answer ALL questions

ALL questions carry EQUAL marks (10 x 1 = 10)

- 1 R is technically much closer to the Scheme language than it is to the original \_\_\_\_\_ language.  
(i) B (ii) C (iii) C++ (iv) S
- 2 The R-help and \_\_\_\_\_ mailing lists have been highly active for over a decade now.  
(i) R-mail (ii) R- devel (iii) R-dev (iv) R-del
- 3 The most convenient way to use R is at a graphics workstation running a \_\_\_\_\_ system.  
(i) windowing (ii) running (iii) interfacing (iv) matrix
- 4 The entities that R creates and manipulates are known as \_\_\_\_\_.  
(i) objects (ii) task (iii) container (iv) packages
- 5 Numbers in R are generally treated as \_\_\_\_\_ precision real numbers.  
(i) single (ii) double (iii) real (iv) imaginary
- 6 The \_\_\_\_\_ function takes an arbitrary number of arguments and concatenates them one by one into character strings.  
(i) copy() (ii) paste() (iii) bind() (iv) del()
- 7 Which of the following statement can be used to explicitly control looping?  
(i) if (ii) while (iii) break (iv) for
- 8 Which of the following is an example of a valid graphics device in R?  
(i) A socket connection (ii) A Microsoft Word document  
(iii) A PDF file (iv) A file folder
- 9 \_\_\_\_\_ is used to view all packages installed.  
(i) library() (ii) search() (iii) .libPaths() (iv) stringr()
- 10 Which level plotting commands generate figures?  
(i) Low (ii) High  
(iii) Both high and low (iv) No levels

Cont...

**SECTION - B (25 Marks)**

Answer ALL questions

ALL questions carry EQUAL Marks (5 x 5 = 25)

- 11 a What is data science and bring out its benefits?  
OR  
b Explain role and stages in data science?
- 12 a What are the goals of data science ?  
OR  
b What all are the main packages used in Python for Data science and Machine Learning?
- 13 a What are the population, sample, training set, design set, validation set, and test set?  
OR  
b State the various types of functions in R to support linear regression?
- 14 a State the steps used to evaluate the data model.  
OR  
b What is Graphical Data Analysis with R? Explain it.
- 15 a How to create (i) Scatter Plot (ii) Box Plot (iii) Area chart  
OR  
b Explain vectorized operations in R Language?

**SECTION -C (40 Marks)**

Answer ALL questions

ALL questions carry EQUAL Marks (5 x 8 = 40)

- 16 a What are input and output in R language? Explain with examples.  
OR  
b Explain different data structures in R.
- 17 a Explain different types of operations in R.  
OR  
b Write about control statements in R.
- 18 a What is cumulative sum, product, min , max? Write R function used for this purpose.  
OR  
b Name some functions available in “dplyr” package? Give some examples.
- 19 a Write R program to perform the following:  
a. Find the correlation matrix of iris dataset.  
b. Plot the correlation plot on dataset and visualize giving an overview of relationships among data on iris data.  
c. Perform analysis of covariance.  
OR  
b Elaborate supervised and unsupervised learning.
- 20 a Generalize the graphical analysis in data analysis? List the various plots in R and explain in detail.  
OR  
b Write R script to plot a data frame having: {df1: {red,green,blue,pink,black}  
df2: {3,5,8,10,34}} using relevant plot.