Scholars Journal of Physics, Mathematics and Statistics

OPEN ACCESS

Abbreviated Key Title: Sch J Phys Math Stat
ISSN 2393-8056 (Print) | ISSN 2393-8064 (Online)
Journal homepage: https://saspublishers.com/sjpms/

Teaching Practice of R Language in a Quick Way

Sheng-Kun Li*

College of Applied Mathematics, Chengdu University of Information Technology, Chengdu 610225, P.R. China

DOI: 10.36347/sjpms.2020.v07i07.001 | **Received:** 23.06.2020 | **Accepted:** 01.07.2020 | **Published:** 08.07.2020

***Corresponding author:** Sheng-Kun Li

Abstract **Review Article**

The statistical software R language can be used as a powerful tool to apply the theory of applied mathematics to practice. The purpose of this paper is to discuss how to guide the students to R language quickly with less teaching hours and more exercises. In order to enable students to realize the basic calculation function of solving mathematical model with R language, this paper introduces in detail the basic teaching content arrangement and the attention details in the teaching process required by the entry-level students. This teaching practice embodies the teaching characteristics of less time investment, simplified teaching content, quick learning for students and easy to stimulate learning enthusiasm.

Keywords: Statistical software, R language, Applied mathematics.

INTRODUCTION

In addition to mastering the solid mathematical foundation, the undergraduates of applied mathematics should be good at using mathematical tools to establish mathematical models for practical problems, and have the ability to use software to solve the established mathematical models. There is no doubt that mathematics plays a decisive role in the development of modern science and technology. Ren Zhengfei mentioned mathematics six times in an exclusive interview of CCTV's face to face program broadcast at 21:30 on January 20, 2019. There are more than 700 mathematicians engaged in basic research in Huawei [1]. Statistics and mathematics are inseparable. Mathematics is the most useful tool of statistics [2], and vice versa. The R language of statistical software should also be one of the software skills that the undergraduates of applied mathematics must master. The 2019 dice technology salary report shows that R and python programmers belong to the high paid professions in the United States [3], and mastering R and Python is of great practical value to students. Wang Liqun wrote an article about 86 years old professor Chen Changlin's moving experience in learning python. Young undergraduates should learn R and python well. With our teaching experience, it is relatively easy to learn R than python. The following describes the practical experience of teaching R language in our school. Our school uses 8-hour teaching and 24-hour computer practising to quickly guide students to learn R language.

PREPARATORY WORK

We put lectures and computer courses in the laboratory so that students can experience the operation process in person from the beginning to the end. The R language is completely free of charge, powerful, and the installation file is small. The R 3.6.1 installation file for Windows operating system is only 80.7mb, which can be copied to the U disk for standby at any time. stay https://www.r-project.org/ Click "download R" on the homepage to display "crane mirrors", Click any of the links, such as 0-cloud https://cloud.r-project.org/ , the page shows that you can select a matching one according to your computer's operating system type, click the "base" for initial installation, and finally click "download R 3.6.1 for Windows (take the current version of windows that is suitable for the operating system as an example), you can download the installation file R 3.6.1- win.exe , download and install in the default "next" mode. After successful installation, start the display operation interface (R console) and run one line of command to display the corresponding results.

To learn R, we need to choose appropriate textbooks. We have tried "R software operation introduction" [5], "R language and application statistical analysis experimental guidance" [6], and the effect is very good. The latest R-based application statistics

published by China Statistics Publishing House in 2019 [7] enables students to master R's basic skills quickly in one chapter, and the following chapters are supplemented by a large number of statistical methods and R language implementation of statistical models to improve skills, which is in line with our teaching style of less lecture time and more computer hours. In addition, students with better English can also use the operation manual in R language. Click "help" → "manual (PDF)" → "an introduction to R" on the toolbar of "R console" interface to open an operation manual named "R- intro. pdf" Copy the commands in the file to practice faster. For example, copy "x < - C (10.4, 5.6, 3.1, 6.4, 21.7)" to the red prompt sign ">" on the operation interface, and enter "X" to enter, which will display the following operation results: [1] 10.45.63.16.4 21.7. By now, students have successfully experienced using function c () to generate a set of numbers into a vector named X in R. In the future, all commands of R language can be input in this way and run out the results immediately. However, this is straightforward but sometimes inconvenient. Although you can use the keyboard to turn up the arrow "↑" and down the arrow "↓" to find the previously run code, it is more convenient to use the method of creating program script when there are more codes. Click "file" → "new program script" "A dialog box will pop up, where you can enter commands separated by semicolons ";, or line breaks. Select the code to run, right-click → run current line or selected code, and the run result will be displayed in R console. The program script can be saved as a file with Suffix ". R". Next time you want to edit and run these codes, click "file" → "open program script" to reopen.

Teaching of realizing computing function

Basic concept of data structure

In order to facilitate students' understanding, it is very useful to use popular expressions rather than strict computer terms, especially for students who lack computer foundation.

To calculate, first of all, we need to let the data used in the calculation enter into the software in two ways: either manually input from the keyboard, or automatically import in batches from files such as excel. The common basic formats of data are vector, matrix, array, data frame and list. As shown in Figure 1, the relationship between several data structures is displayed intuitively.

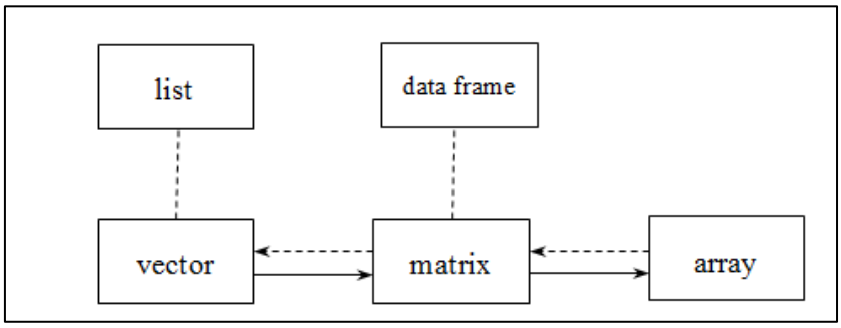

Fig-1: Data structure relationship

Take vector and matrix as the core to clarify the relationship between these data structures. The concept of matrix in R language is consistent with that in linear algebra. The vector of R language covers a wider range than that of linear algebra. There is only one vector of numerical type in linear algebra. In addition to the numerical type, the vector of R language has character type (element is character) and logical type (element is logical value true and false). Next, we will talk about the relationship between vector and matrix, which refers to the vector of numerical type. A matrix is also called a 2-dimensional array because it involves two dimensions, row and column. In particular, when the number of rows of a matrix M = 1 or the number of columns n = 1, the matrix is reduced to a vector. The former is called a row vector, and the latter is called a column vector. If the number of rows and columns are reduced to 1 at the same time, the matrix becomes a number. On the basis of the matrix, increase the dimension to become the array. The elements in the array are required to be numbers. Intuitively, for example, a vector is a row (or a column) on a page; a matrix is a page, or a 2-dimensional array; a page is folded into a book, or a 3-dimensional array; a book is arranged, or a 4-dimensional array; the concept of array dimension is understood by analogy. With the intuitive understanding of arrays, it is very useful for programming how to select the data in them. Students often don't know how to nest when there are multiple loops.

According to the intuitive understanding of arrays, we need to get a certain number of arrays. We need to determine which book, which page, which row and column are the first to get the number, so who is the outermost layer when nesting It's clear. Students usually use 1-dimensional vector, 2-dimensional matrix and 3-dimensional array (also known as data cube) for beginners. It is not difficult to skillfully use these three kinds of arrays first and then deal with high-dimensional ones. The structure of data frame is the same as that of matrix. The difference is that matrix

requires all elements to be numbers, and the elements in data frame can also be more general characters. For example, students' score table is a typical data frame, and each column of score table can record person name, student number, score, etc. A list is a generalization of a vector whose elements can be any data object.

Data manual input

Manual input of data can only teach vector, matrix, 3D array and data frame (1). The form of numerical vector is x < - c (10.4, 5.6, 3.1, 6.4, 21.7) or c (10.4, 5.6, 3.1, 6.4, 21.7) - > x or x = c (10.4, 5.6, 3.1, 6.4, 21.7). Where x is the vector name, case sensitive. The elements of numerical vector are all numbers, separated by "," comma "," and then the space can be more or less, which is not very strict with the format requirements like python. This is also the reason why students are recommended to use R language as a starting point. Debugging code reports fewer errors, which is easy to stimulate learning interest. Emphasize to the students that we must remember to name the vector, and do the same with the matrix, array, data frame, etc. in the future, because we need to use the name when we call the elements, such as the vector x just now, we need to call the third component, just use the code x [3], and so on. Running code c (10.4, 5.6, 3.1, 6.4, 21.7) is consistent with running code "x < - c(10.4, 5.6, 3.1, 6.4, 21.7); X" shows the same effect, but the number in the former vector can be seen but cannot be used. If the result of running code x ^ 2 is: 108.1631.369.6140.96 470.89, this is to get a new vector for each number square of vector X. if you want to call the number 9.61, students often mistake code x ^ 2 [3] for implementation, and the correct code is: y = x ^ 2; y [3]. It is very useful for students to learn to program in the later stage to remember to create new objects in time and name them in advance. Also remind students that the input code should be in the English punctuation mode of the keyboard, and many code operation errors are caused by the input of punctuation in the Chinese punctuation mode. When programming, regular numerical vectors are used more often. Data satisfying specific rules (such as equal difference sequence, obeying a certain probability distribution, etc.) are called regular data. The function seq (), rep (), rnorm () is introduced to students. The function seq() is used to generate the sequence of equal difference numbers. It is widely used in numerical simulation and drawing. The code A1 = seq (from = 2.1, to = 6, by = 0.5); the result of A1 operation is: 2.12.63.13.6 4.14.6 5.15.6. The starting point of the generated vector is from = 2.1, the end point is to = 6, and the step is by = 0.5 (that is, the tolerance of the sequence of equal difference numbers). Note that the last number is 5.6 instead of 6. Code A3 = seq (from = 2, to = 6, length = 4); A3 operation result is: 2.000000 3.333333 4.666667 6.000000, parameter length = 4 indicates the number of components, the start and end points of the generated sequence are exactly the values specified by the parameters from and to, without specifying the step length. Code a4=seq (from=2, by=0.2, length=4); A4 running results are: 2 2.2 2.4 2.6, starting from 2, step size 0.2, generating 4 numbers. A special sequence of equal difference numbers with step size of 1 is generated with the symbol: "generate, code n = 1:100; n generates an integer vector named n from 1 to 100, similar vectors are commonly used in programming. Function rep () is used to generate the repeated vectors of elements, code R1 = Rep (0, 5); R1 runs with the result of zero vector 0 0 0 0 0 0 0. RN1 = rnorm (10); the result of RN1 operation is to generate a vector named RN1 with 10 random numbers following the standard normal distribution. Code RN2 = rnorm (10, mean = 12, sd = 3); RN2 operation result is to generate a vector named RN2, the component is 10 random numbers that obey the normal distribution with the mean value of 12 and the standard deviation of 3. The operation of character type vector and logic type vector is completely similar, but the components are different. The components are any characters with double quotation marks", and the components of logic vector are true or false.

(2) The function matrix() to generate the matrix. Code M1 = matrix (1:6, nr = 2, nc = 3); M1 runs the result to generate a matrix named M1 with 2 rows and 3 columns. Its elements are filled in columns by vectors generated by 1:6. To fill in rows, add the parameter byrow = true. Demonstration code a = c(1, 3, 4, 2); m3 = matrix (a, nr = 2, nc = 3, byrow = true); the operation result of M3 enables students to understand the meaning of cyclic complement when the vector length is not enough. Zero matrixes, unit matrix and all 1 matrix can be demonstrated. Code M4 = matrix (0, nr = 2, nc = 3); M4 generates a 2-row 3-column matrix named M4 with all elements of 0. Code M5 = matrix (data = Na, nr = 2, nc = 3); M5 generates a two row three column empty matrix named M5. Code M6 = diag (4); M6 is generated into a 4th order unit matrix named M6. Pay attention to naming the matrix, and its benefits will be revealed quietly when students program.

(3) Generate array function array (). Code A1 = array (1:24, dim = C (3,4,2)); A1 generates an array named A1, which contains two matrices of three rows and four columns. As a starting point, just practice 3D array.

(4) Generate data box function data.frame () take student's score table as an example to demonstrate that generating data frame can make students get intuitive experience. For example, the code is as follows: student number = c (20190701, 20190702, 20190703); name = c ("Zhao Yi", "Qian Er", "Sun San"); language = c (89, 97, 88); mathematics = c (120, 145, 112); score sheet<- data frame (student number, name, Chinese, Mathematics); score sheet. Run the above code to generate a simple score sheet. Note that the punctuation in the above code should be input in English punctuation mode.

Data import automatically

Importing data files of other formats into R software is a skill that students must master. Obviously, when there is a large amount of data, it is time-consuming and error prone to input one by one by keyboard, and the way of automatic import is fast and accurate. R can import many data formats, such as Excel data, SPSS data, SAS data, etc. It is not necessary to explain the import of these data one by one according to the textbook. Beginners only need to learn to import Excel data, because the original data is usually stored in Excel, and other software data can be converted into Exel. Import Excel data only need to be familiar with the command read.csv(), let's say that in the "R practice" folder of disk D, the name is“ chengji.xlsx ”The excel file that records the student's grade information, and saves it as a. CSV file in the original location“ chengji.csv ”, run the command: mydata= read.csv ("D: \\r practice\\ chengji.csv ", header = true), you can read the data into the object named mydata, The parameter header = true causes the first row of read data to be the header row, and header = false to be the opposite. Note that "\\" or "/" instead of "\" should be used to indicate the path.

Data retrieval

Data is stored in vectors, matrices, data frames and arrays, and the element is taken from it to specify the location of the element in the parentheses behind the object. The second component code of calling vector V is v [2], and the second component code of removing vector V is v [2]. The code of the second row and the third column of the call matrix M is m [2, 3], the code of the second row of the call matrix M is m [2], and the code of the third column of the call matrix M is [3] and the method of data frame is the same. Call the element code of the second matrix, the third row and the fourth column of a three-dimensional array a [3, 4, 2].

Basic operation and common built-in functions of data

The operation mode of built-in function is to perform the same operation on each element of vector, matrix and array, for example, matrix M, code m ^ 2, which means to perform square operation on each element of matrix M. It is worth noting that the running code m = matrix (1:6, nr = 2, nc = 3); the result of m ^ 2 is displayed as a matrix. To retrieve the element 9 in the second column of the first row of the matrix, it is often mistaken that the running code m ^ 2 [1,2] can be implemented, and the correct code is M1 = m ^ 2; M1 [1,2].

Loop statement

Beginners can master only for loop and while loop. Take the generation of Fibonacci sequence as an example: 1,1,2,3,5,8,13,21,34 The Fibonacci sequence satisfies the following conditions: F (1) = 1, F (2) = 1, F (n) = F (n-1) + F (n-2) (n > = 3, n ∈ n). The Fibonacci sequence code with 100 elements is shown in Table 1:

Table-1: Generating Fibonacci sequence with for loop

```
①F=NA
②F[1]=1; F[2]=1
③for (i in 3:100) {F[i]=F[i-1]+F[i-2]}
④F
```

Sentence ③ corresponds to for (var in seq) expr in the textbook. The cyclic variable var is taken from the sequence seq in turn, and each time var is taken, the corresponding expr is executed. Statement ③ means: i take value 3, execute F[3] = F [2] + F [1]; i take value 4, execute F [4] = F [3] + F [2]; until i is 100, execute F [100] = F [99] + F [98]. When expr is a single statement, curly braces {} may not be added. When expr is multiple statements, curly braces {} must be added. Semicolons are used to separate or wrap statements. It is suggested that students form the habit of adding curly braces to both single and multiple statements in expr, and the situation of error reporting often occurs when students do not add curly braces to the computer. Through this simple example, let students have a preliminary clear impression that the expr of a circular statement is an expression similar to a general term formula or a recurrence formula, and F[i] = F [i-1] + F [i-2] will change with the value of i. Sentence ① is often ignored by beginners, reminding students to remember that when new objects are needed, they are defined by assignment first.

The Fibonacci sequence code with mantissa no more than 10000 is generated as shown in Table 2.

Table-2: Generating Fibonacci sequence with while loop

```
⑤F=NA
⑥F[1]=1; F[2]=1
⑦F[3]=F[2]+F[1]
⑧i=3
⑨while(F[i]<=10000) {
⑩i=i+1
⑪F[i]=F[i-1]+F[i-2] }
⑫L=length(F)
⑬ F=F[-L]
⑭ F
```

Statements ⑨-⑪ correspond to the grammatical structure of the while loop in the textbook: while (cond) {expr}. As long as the condition cond is satisfied, execute the statement expr, and cycle to check whether the cond is satisfied, then continue to execute expr until the cond is not satisfied. In Table 2, statement ⑤-⑦ defines a vector named F and assigns values to the first three components. Before entering the while statement, statement ⑧ needs to be used to assign values to i. otherwise, an error is reported in statement ⑨ that "where true / false values are required, missing values cannot be used". In the loop body, statement ⑩ needs to be used to ensure that the conditions in statement ⑨ change, otherwise, the statement cannot be

run the expected correct result will also cause the program to fall into a dead cycle. Since the loop statement stops executing when F [i] > 10000 is checked in statement ⑨, the last number greater than 10000 in statement F needs to be removed in statement ⑫ ⑬ because the generated mantissa of the topic does not exceed 10000.

If we know the single loop and the data structure such as matrix and array clearly, the double loop is not difficult. Take the generation of 10 order Hilbert matrix as an example, $H = (h_{ij})_{10\times10}$, $h_{ij} = \frac{1}{i+j-1}$, The code is shown in Table 3.

Table-3: Double cycle generating 10 order Hilbert matrix

```
①n=10
②H=matrix(data=Na, nr=n, nc=n)
③for (i in 1:n) {
④for (j in 1:n) {
⑤H[i,j]=1/(i+j-1)
⑥}
⑦}
⑧H
```

It's similar to writing on a notepad to do a double loop operation i n a matrix. In Table 3, determine 1-10 rows in turn with the loop variable i and then determine 1-10 columns in turn with the loop variable j after determining a row. Statement ③-⑥ is an outer for loop structure for (I in 1: n) {expr1}, and statement ④-⑥ is expr1. The inner and outer loops are nested.

Conditional statements

Students should be familiar with if else structure. The syntax structure if (cond) {expr} indicates that the statement expr is executed when the condition cond is satisfied. If (cond) {expr1} else {expr2} indicates that the statement expr1 is executed when the condition cond is met, otherwise the statement expr2 is executed. The piecewise function can be used as a typical example. There are piecewise functions as follows

$$y = \begin{cases} x^2, & x \le 0; \\ \sin(x), & x > 0. \end{cases} \quad (1)$$

The R code of formula (1) is shown in Table 4

Table-4: The if-else structure calculation for subsection function

```
① x=  # Enter the value of the argument x
② if (x<=0) {y=x^2} else {y=sin(x)}
③ y
```

In order to give students a good practical experience, we often use the tax rate table to practice the conditional sentence, and combine the tax planning problem to establish a mathematical model and solve it with R language [8].

DISCUSSION

Many R language courses are a little complicated for beginners. If beginners learn the content page by page in turn, the number of times of encountering code errors will easily affect their enthusiasm to continue learning. In the teaching, we follow the principle of simplicity, select the necessary teaching content, focus on explaining the details that students are prone to make mistakes, require students to synchronize with the teacher's explanation from the beginning to the end, run through the ideas that students should be able to use if they want to make students interested, and achieve better teaching results. Many students consciously study other content by their selves after we teach the basic introductory content. We provide students with the necessary foundation and motivation to continue learning.

REFERENCE

1. Ren, Z. (2019). CCTV network: face to face of Huawei [EB / OL]. Update time: January 20, 2019, 23:17 http://tv.cctv.com/2019/01/22/VIDEp6XrXlJEFI7vQ5sO79En190122.shtml [reference time: August 31, 2019].
2. Zhang, H. (2019). What is statistics [J]. China Statistics, 4: 68.
3. CSDN headline. (2019). The highest paid technical skills in the United States [EB/OL]. March 4, 2019 13:19:50 https://blog.csdn.net/rx3oyuyi/article/details/88111010[reference time: August 31, 2019].
4. Wang, L. (2019). Learn Python together [J]. China Statistics, 4: 49-50.
5. Chen, Y. (2010). R software operation introduction [M]. Beijing: China Statistics Press.
6. Qin, Yi, N. (2017). Experimental guidance of R language and applied statistical analysis [M]. Beijing: China Statistics Press.
7. Xu, D. (2019). Application statistics based on R [M]. Beijing: China Statistics Press.\
8. Zhao, Y., Chen, Y. (2013). Annual lump sum bonus and the planning model of salary and tax payment in the 12th month [J]. Journal of Finance and accounting, 8, 104-106.