

**BỘ GIÁO DỤC VÀ ĐÀO TẠO
TRƯỜNG ĐẠI HỌC MỎ-ĐỊA CHẤT**

BÁO CÁO HỌC THUẬT

**MỘT SỐ KHÁI NIỆM CƠ BẢN CỦA THỐNG KÊ VỚI
PHẦN MỀM R TRONG DẠY HỌC MÔN XÁC SUẤT THỐNG KÊ
TẠI TRƯỜNG ĐẠI HỌC**

ThS. Nguyễn Thu Hằng

Hà nội, tháng 12 năm 2021

BỘ GIÁO DỤC VÀ ĐÀO TẠO
TRƯỜNG ĐẠI HỌC MỎ-ĐỊA CHẤT

BÁO CÁO HỌC THUẬT

MỘT SỐ KHÁI NIỆM CƠ BẢN CỦA THỐNG KÊ VỚI
PHẦN MỀM R TRONG DẠY HỌC MÔN XÁC SUẤT THỐNG KÊ
TẠI TRƯỜNG ĐẠI HỌC

Xác nhận của bộ môn

Hà nội, tháng 12 năm 2021

MỤC LỤC

Lời giới thiệu	1
1. Hướng dẫn cài đặt và giới thiệu tổng quan	2
1.1. Hướng dẫn cài đặt	2
1.2. Giới thiệu tổng quan	2
2. Dữ liệu và một số thao tác tiền xử lý số liệu	6
2.1. Phép toán và hàm tính toán cơ bản	6
2.2. Nhập dữ liệu	8
2.3. Các thao tác cơ bản với dữ liệu	9
2.4. Bảng dữ liệu và truy cập vào bảng dữ liệu	11

LỜI GIỚI THIỆU

Xác suất thống kê là một môn học quan trọng trong chương trình giảng dạy đại học và có nhiều ứng dụng trong thực tiễn. Việc giảng dạy Xác suất thống kê cho sinh viên cũng ngày càng đổi mới theo hướng thực hành, tiếp cận với nhu cầu của nhiều bài toán thực tế. Các phần mềm phục vụ tính toán thống kê cũng theo đó được giới thiệu và giảng dạy ở nhiều trường trong và ngoài nước. Một trong những phần mềm rất phổ biến là phần mềm thống kê R. Nhằm xây dựng, đổi mới giáo án theo xu hướng mới phù hợp, tôi chọn làm báo cáo học thuật “Một số khái niệm cơ bản của thống kê với phần mềm R trong giảng dạy môn Xác suất thống kê tại trường đại học”, trình bày những vấn đề đơn giản nhất khi làm việc với R.

Báo cáo học thuật chia làm ba phần

Phần 1: Hướng dẫn cài đặt và giới thiệu tổng quan

Phần 2: Dữ liệu và một số thao tác tiền xử lý số liệu

1. Hướng dẫn cài đặt và giới thiệu tổng quan

R được bổ sung và phát triển từ ngôn ngữ lập trình S. Ngôn ngữ S được tạo ra bởi John Chambers ở Bell Labs vào cuối thập niên 80. Ngôn ngữ R được sáng tạo bởi Ross Ihaka và Robert Gentleman ở đại học Auckland – New Zealand vào năm 1995 và sau đó được phát triển bởi nhóm phát triển R (R Development Core Team), trong đó có Chambers là thành viên. R được đặt tên một phần theo chữ cái đầu tiên trong họ của hai tác giả và một phần dựa theo tên của S. R chạy được trên nhiều hệ điều hành: Windows, Mac OS,...

1.1. Hướng dẫn cài đặt

Để tải bộ cài R về máy ta thực hiện các bước sau

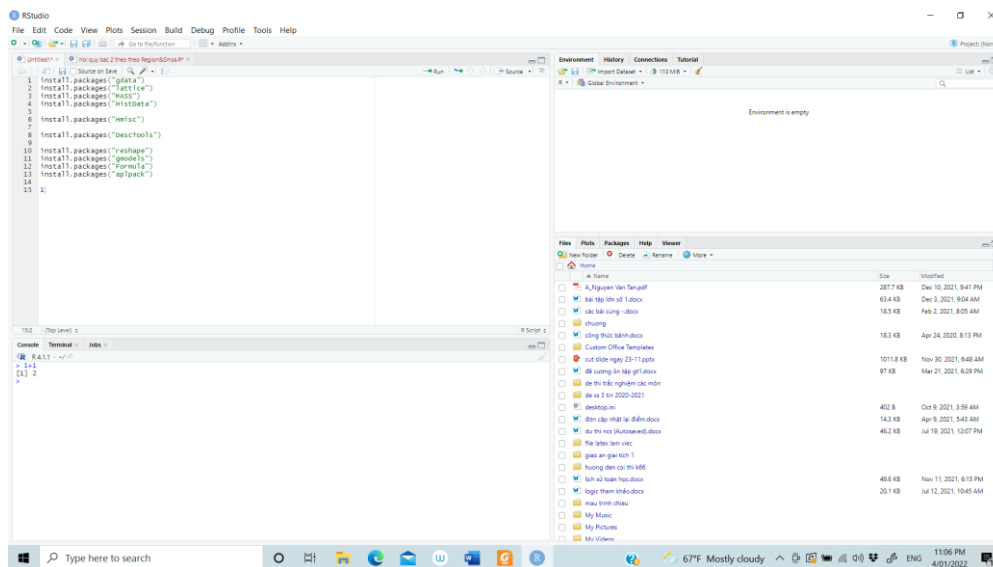
- Kết nối internet
- Vào đường link: <http://cran.r-project.org/>
- Chọn download R for Windows
- Chọn “Base” → Download R for the first time
- Chọn download R3.5.1 for Windows và bắt đầu cài đặt
- Kết thúc cài đặt, Icon R sẽ xuất hiện trên desktop

Để tải Rstudio ta thực hiện các bước sau

- Kết nối internet
- Vào đường link: <https://www.rstudio.com/products/rstudio/download/>
- Chọn “down load Rstudio for windows” và cài đặt

1.2. Giới thiệu tổng quan

➤ *Giao diện của Rstudio*



Giao diện của nó được chia thành 4 phần và phần dưới cùng bên trái chính là nơi chúng ta sẽ gõ các dòng lệnh. Tuy nhiên ta cũng có thể gõ lệnh ở cửa sổ phía trên bên trái rồi chọn “run” để chạy lệnh. Việc gõ lệnh trên cửa sổ phía trên bên trái sẽ giúp chúng ta lưu được các lệnh này lâu dài. Các dòng lệnh sẽ là màu xanh và kết quả hiện lên có chữ màu đen. Hình ảnh (nếu có) sẽ được hiện ở khu vực phía dưới bên phải của màn hình. Chẳng hạn để tính 1+1 tại cửa sổ lệnh của Rstudio (bên dưới góc trái màn hình) các bạn gõ 1+1 rồi enter. Nhớ rằng Rstudio là thư hỗ trợ chúng ta sử dụng R thuận tiện hơn. Ngoài ra Rstudio còn có thể “biến” các phân tích của bạn thành một bài trên web site.

➤ *Cài đặt thêm các gọi lệnh mới*

Mỗi khi cài đặt R thì chỉ có một số hàm, chẳng hạn hàm **mean()** để tính trung bình, là luôn sẵn có để chúng ta sử dụng. Tuy nhiên, nếu chúng ta cần một phân tích đặc biệt nào đó như phân tích dữ liệu bảng thì chúng ta cần phải cài đặt một hoặc một số gói (Packages) cần thiết tương ứng.

Cho đến thời điểm tài liệu này được viết đã có hơn 9000 gói được viết cho R. Có 3 loại gói lệnh chính

- Thứ nhất là các gói lệnh cơ bản cung cấp các hàm cơ bản của R. Các gói lệnh này chủ yếu được viết bởi chính nhóm tạo ra phần mềm R.
- Thứ hai là các gói lệnh cũng được mặc định cài đặt trên R như “MASS”, “class”, “foreign”... bao gồm các thủ tục thống kê phức tạp hơn.
- Thứ ba là các gói lệnh được đóng góp bởi người dùng, được lưu tại kho và bạn có thể download các gói lệnh này về máy mình để dùng

Đối với nghiên cứu thống kê – kinh tế lượng, dưới đây là một số gói quan trọng mà chúng ta cần đến:

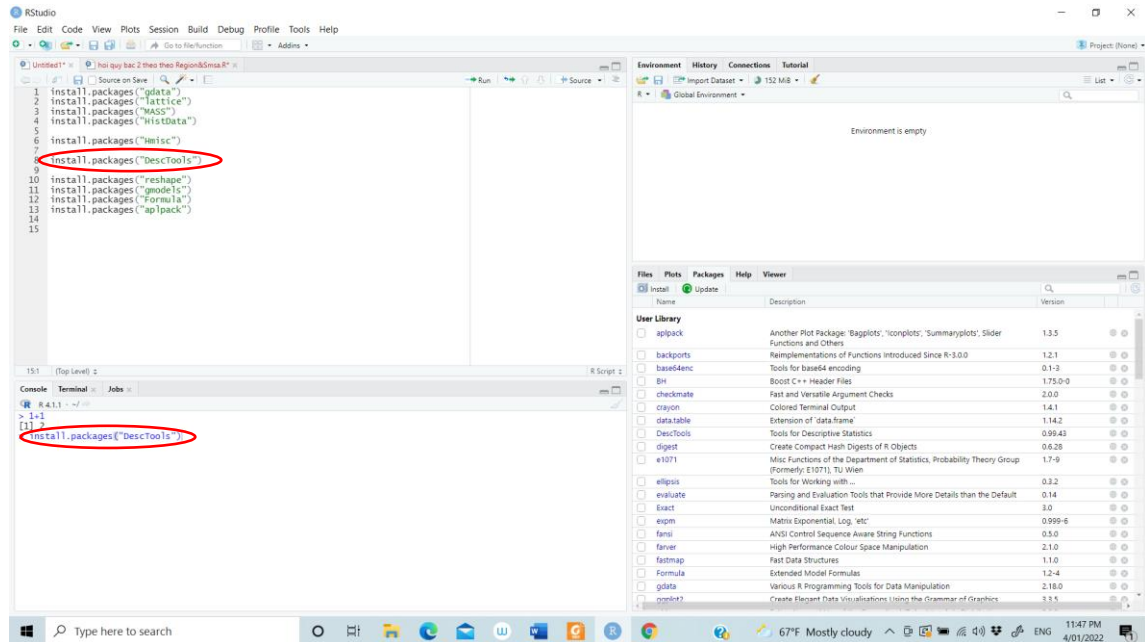
Tên gói	Công dụng
xlsx	Đọc data đuôi .xlsx (file Excel)
gdata	Đọc data đuôi .xls (file Excel)
foreign	Đọc data đuôi .sav (file SPSS), .dta (file Stata)
hexView	Đọc data đuôi .wfl và .WF1(file Eviews)
AER	Gói thực hiện nhiều phân tích kinh tế lượng cơ bản
ggplot2	Vẽ Graph với chất lượng hình ảnh cao
lattice	Dùng để vẽ đồ thị và làm cho đồ thị đẹp hơn

Hmisc	Một số phương pháp mô hình dữ liệu của F. Harrell

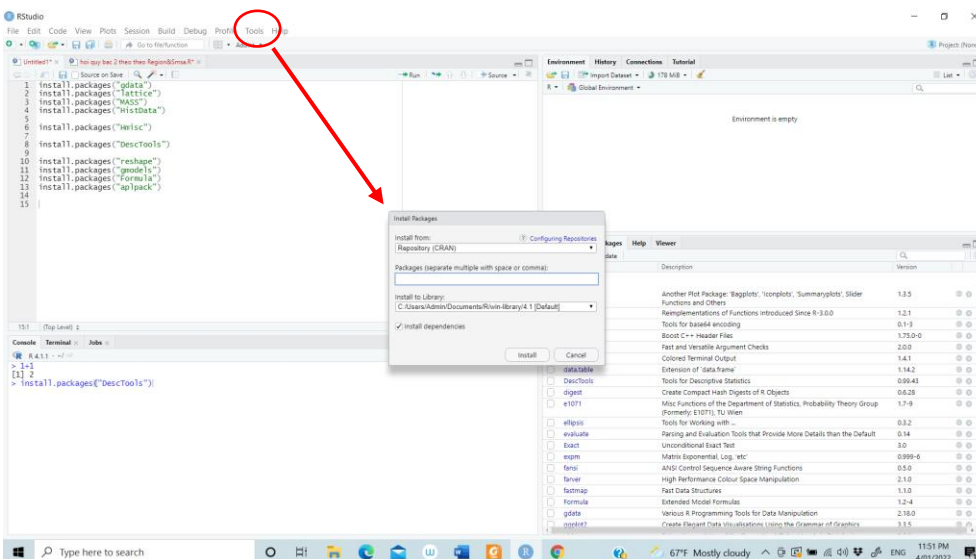
Muốn cài đặt một gói lệnh nào đó, chẳng hạn như gói “DescTools”, ta có thể làm theo một trong những cách sau đây

- Cách 1: Dùng hàm `install.packages` bằng cách gõ lệnh sau

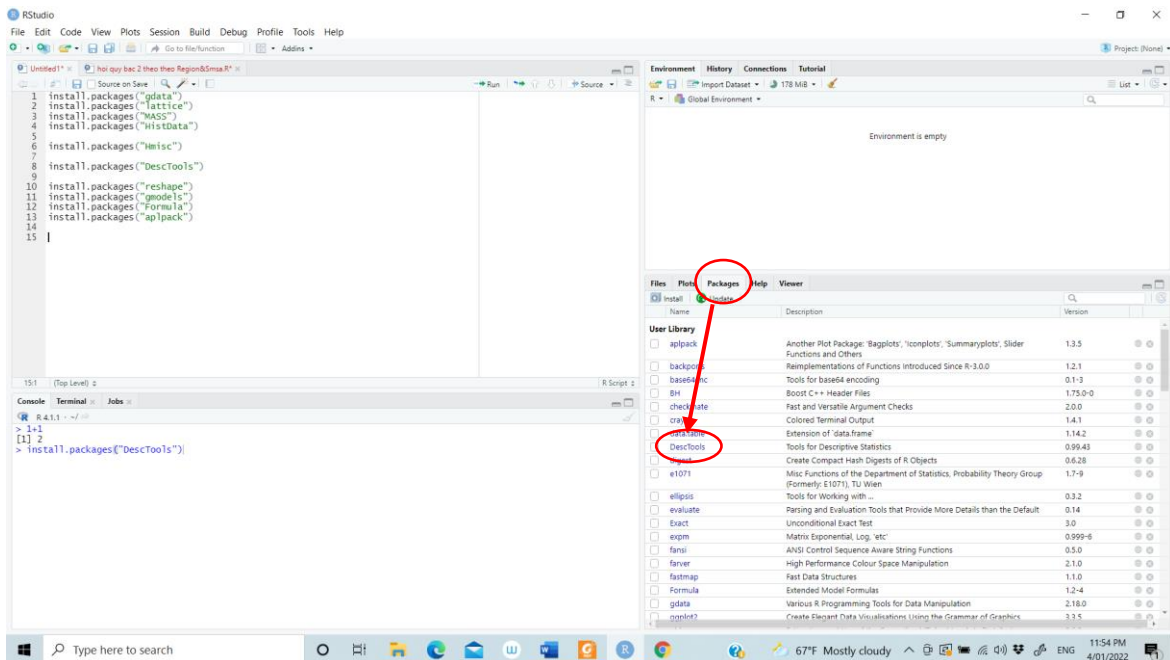
> `install.packages ("DescTools ")`



- Cách 2: Cài từ mục Tools trên thanh công cụ hoặc từ mục Packages TRên thanh công cụ chọn Tools → Install Packages → nhập tên Packages → Install



- Cách 3. Từ cửa sổ Plots and Files:



➤ **Trợ giúp trong R**

Muốn tìm hiểu về một hàm nào đó ta dùng lệnh **help**. Chẳng hạn muốn tìm hiểu về hàm “seq” ta gõ

```
> help(seq)
```

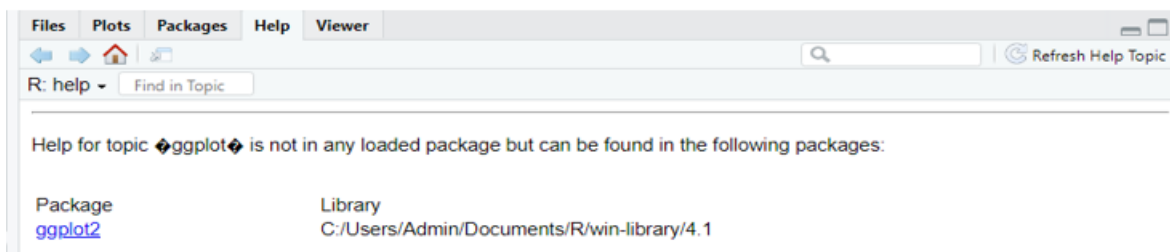
Hoặc

```
? seq
```

R sẽ cung cấp đầy đủ chi tiết từ mô tả chung đến ý, cách dùng khác cũng như ví dụ minh họa. Tuy nhiên hàm **help** chỉ được thực hiện khi ta nhớ chính xác tên hàm. Trong trường hợp không nhớ chính xác, ta có thể dùng hàm **apropos**. Hàm **apropos** liệt kê tất cả các hàm chứa cụm từ ta đang tìm kiếm.

Để biết được một hàm nào đó, chẳng hạn như “ggplot”, nằm trong gói lệnh nào ta gõ

```
> help("ggplot",try.all.packages = TRUE)
```



Sau khi biết được ggplot nằm trong gói ggplot2, thông tin về hàm ggplot được tìm hiểu như sau

```
> library(ggplot2)
```

```
> help(ggplot)
```

Tình huống nếu chúng ta muốn tra cứu chi tiết hơn nữa cách sử dụng, thậm chí là một bài báo nào đó về sử dụng một hàm cụ thể, chẳng hạn, với lệnh `ggplot()`: `help.search("ggplot")`

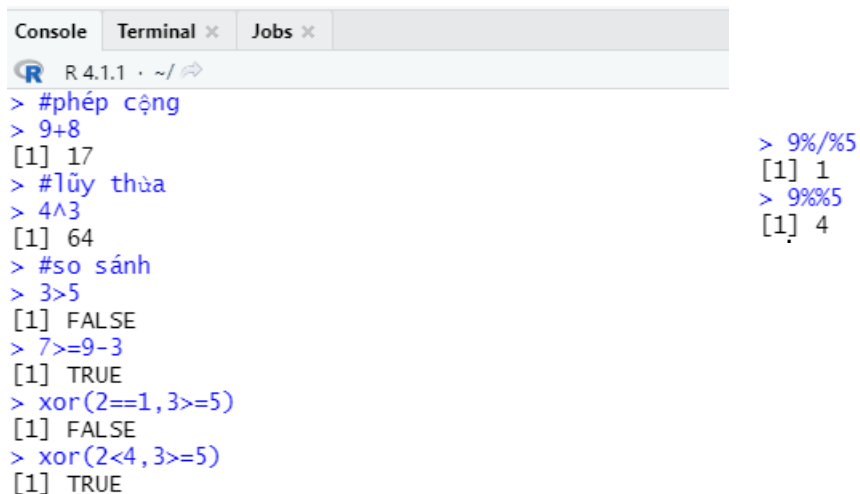
2. Dữ liệu và một số thao tác tiền xử lý dữ liệu

2.1. Phép toán và hàm tính toán cơ bản

➤ Phép toán trong R

<u>Phép toán số học</u>	<u>Phép toán so sánh</u>	<u>Phép toán Logic</u>
+ cộng - trừ * nhân / chia ^ lũy thừa %% nhân hai ma trận %% lấy phần nguyên %% lấy phần dư	< nhỏ hơn > lớn hơn <= nhỏ hơn hoặc bằng >= lớn hơn hoặc bằng == bằng	! phủ định & và hoặc xor tuyển loại

Ví dụ



```
Console Terminal x Jobs x
R 4.1.1 ~ /
> #phép cộng
> 9+8
[1] 17
> #lũy thừa
> 4^3
[1] 64
> #so sánh
> 3>5
[1] FALSE
> 7>=9-3
[1] TRUE
> xor(2==1, 3>=5)
[1] FALSE
> xor(2<4, 3>=5)
[1] TRUE

> 9%/%5
[1] 1
> 9%%5
[1] 4
```

Các phép toán trên còn có thể thực hiện với các vector

➤ Một số hàm cơ bản trong R

Hàm	Công dụng
<code>abs (x)</code>	Tính giá trị tuyệt đối của x
<code>sqrt (x)</code>	Tính căn bậc hai của x
<code>round(x,digits=k)</code>	Làm tròn x đến k chữ số thập phân
<code>t (A)</code>	Hàm chuyển vị ma trận A
<code>exp (x)</code>	e^x
<code>log (x,base=a)</code>	Logarit cơ số a, nếu ko chọn base thì mặc định là e
<code>sin (x) , cos (x) , tan (x)</code>	Tính sinx, cosx, tanx
<code>choose (n, k)</code>	Tính tổ hợp chập k của n
<code>sum (x₁,x₂,...)</code>	Tính tổng các số
<code>prod (x₁,x₂,...)</code>	Tính tích các số
<code>min (x₁,x₂,...)</code> , <code>max (x₁,x₂,...)</code>	Đưa ra số bé nhất, lớn nhất
<code>range (x₁,x₂,...)</code>	Đưa ra số bé nhất và lớn nhất
<code>cumsum (x)</code>	Tổng tích lũy các số trong vector x
<code>cumprod (x)</code>	Tích tích lũy các số trong x

Ví dụ

```
> abs(-1)
[1] 1
> choose(10,2)
[1] 45
> range(1,2,3,4,5,6,7)
[1] 1 7
> cumsum(1:10)
[1] 1 3 6 10 15 21 28 36 45 55
```

Một số hàm khác liên quan đến việc tạo dữ liệu

Hàm	Công dụng
<code>seq ()</code>	Tạo ra một dãy số trong R
<code>rep ()</code>	Tạo ra véc tơ trong R bằng cách lặp lại giá trị của phần tử
<code>sample ()</code>	để lấy mẫu ngẫu nhiên có hoàn lại

```

> #tạo dãy số từ 5 đến 10 với khoảng cách là 0.5
> x = seq(5, 10, by = 0.5)
> x
[1] 5.0 5.5 6.0 6.5 7.0 7.5 8.0 8.5 9.0 9.5 10.0
> #tạo dãy 10 số 1
> y = rep(1, 10)
> y
[1] 1 1 1 1 1 1 1 1 1 1
> #lấy mẫu ngẫu nhiên có hoàn lại 10 số nguyên có giá trị từ 0 đến 9
> z = sample(0:9, 10, replace = TRUE)
> z
[1] 1 9 8 1 4 0 6 7 0 5

```

2.2. Nhập dữ liệu

Có 2 cách nhập dữ liệu là nhập trực tiếp hoặc nhập từ file có trước. Chú ý việc đặt tên một đối tượng (object) hay một biến số (variable) trong R khá linh hoạt, vì R không có nhiều giới hạn như các phần mềm khác. Tên một object phải được viết liền nhau (tức không được cách rời bằng một khoảng trống). Chẳng hạn như R chấp nhận myobject nhưng không chấp nhận my object. Một điều quan trọng cần lưu ý là R phân biệt mẫu tự viết hoa và viết thường. Cho nên My.object khác với my.object.

➤ Nhập dữ liệu dạng vector bằng hàm c hoặc hàm scan

```

> so.loi=c(2,3,1,0,3,4,1,5,2,0,1)
> so.loi
[1] 2 3 1 0 3 4 1 5 2 0 1
> SoLoi=scan()
1: 1 2 3 5 0 1 2 5 4 1 0 2 1
14:
Read 13 items
> #enter 2 lần để kết thúc việc nhập dữ liệu. vector SoLoi có 13 phần tử
> SoLoi
[1] 1 2 3 5 0 1 2 5 4 1 0 2 1

```

Sự khác biệt giữa hai cách nhập là c có thể nhập dữ liệu dạng kí tự nhưng scan chỉ nhập dữ liệu dạng số. Tuy nhiên, hàm scan đôi khi có thể giúp ta nhập dữ liệu từ tệp có sẵn như Excel...

```

> giới.tính=c("nam","nu","nu","nam","nu")
> giới.tính
[1] "nam" "nu" "nu" "nam" "nu"
> GiớiTinh=scan()
1: nam nu nu nam nu
Error in scan() : scan() expected 'a real', got 'nam'

```

➤ Nhập dữ liệu dạng ma trận

`matrix(data,nrow,ncol,byrow)`

trong đó

data: vector dữ liệu cho ma trận

nrow, ncol: số hàng, số cột

byrow: tham số logic xác định việc sắp xếp theo hàng hay cột, mặc định là `byrow = FALSE`, tức là theo cột.

```
> x=c(1,2,3,4,5,6,7,8,9)
> A=matrix(x,nrow=3)
> A
     [,1] [,2] [,3]
[1,]    1    4    7
[2,]    2    5    8
[3,]    3    6    9
> B=matrix(x,nrow=3,byrow=T)
> B
     [,1] [,2] [,3]
[1,]    1    2    3
[2,]    4    5    6
[3,]    7    8    9
```

➤ Nhập dữ liệu từ file Excel có sẵn

Khi nhập dữ liệu vào từ file Excel, thông thường, tệp chứa dữ liệu có đuôi `.xls` hoặc `.xlsx`. Để đọc được những tệp này từ R ta làm như sau

B1. Lưu tệp dưới dạng đuôi `.csv` (Vào save as chọn CSV(Comma delimited))

B2. Trong R, truy cập vào thư mục chứa dữ liệu

B2. Đọc tệp đuôi `.csv` bằng lệnh `read.csv`

Ví dụ.

```
> #truy cập vào thư mục chứa dữ liệu
> setwd("F:/hang/diem b,c/he 2020-2021")
> #đọc dữ liệu và lưu vào đối tượng diem
> diem=read.csv("diemlogic.csv")
> #hiển thị 5 dữ liệu đầu tiên
> head(diem,5)
```

	STT	maSV	hoten	X	ns	lop	A	B1	B2	B	C	DGHP	xeploai
1	1	2024010849	pham phuong	Anh	21/11/2002	DCKTKD65_A2	0	0	0	NA	0	NA	NA
2	2	1621060988	nguyen van	chuc	13/06/1998	DCCDTD61_1	8	9	10	NA	10	NA	NA
3	3	1821060103	nguyen viet	Cuong	17/05/2000	DCCDTD63C	0	0	0	NA	0	NA	NA
4	4	1821050517	Trinh Viet	Cuong	4/08/2000	DCCTMM63A	8	9	10	NA	10	NA	NA
5	5	1821050824	Bui Quang	Duy	27/01/2000	DCCTHT63A	8	10	9	NA	10	NA	NA

2.3. Thao tác cơ bản với dữ liệu

➤ Truy cập dữ liệu

Mỗi một phần tử trong một vector được xác định bởi thứ tự của phần tử đó. Ta có một số lệnh sau để truy cập vào các phần tử trong vector

Lệnh	Công dụng
<code>length(x)</code>	Đưa ra số phần tử của x
<code>x[i]</code>	Đưa ra phần tử thứ i
<code>x[m:n]</code>	Đưa ra các phần tử từ vị trí thứ m đến n

<code>x[c(i,j,k)]</code>	Đưa ra các phần tử ở vị trí i, j, k của x
<code>x[x>a]</code>	Đưa ra những phần tử lớn hơn a
<code>sum(x>a)</code>	Tính tổng những phần tử lớn hơn a của x
<code>which(x==a)</code>	Đưa ra những phần tử bằng a
<code>which.max(x)</code>	Đưa ra vị trí đầu tiên của phần tử lớn nhất
<code>subset(x, subset)</code>	Đưa ra những phần tử của x thỏa mãn điều kiện nêu trong subset

```
> x=c(1,2,3,4,5,6,7,8,9)
> length(x)
[1] 9
> #đưa ra các phần tử từ vị trí thứ 2 đến thứ 4
> x[2:4]
[1] 2 3 4
> #lấy ra các phần tử ở vị trí 1, 3 và 5
> x[c(1,3,5)]
[1] 1 3 5

> #lấy ra các phần tử lớn hơn 3 và nhỏ hơn 5
> subset(x,x>3&x<5)
[1] 4
> which.max(x)
[1] 9
```

➤ Biên tập dữ liệu

Với các vector, ta cũng có thể thực hiện các công việc như sao chép, ghép nối, thay đổi giá trị...

```
> x
[1] 1 2 3 4 5 6 7 8 9
> y=x
> y
[1] 1 2 3 4 5 6 7 8 9
> #sửa số ở vị trí giữa thành 10
> y[length(y)%/%2+1]=10
> y
[1] 1 2 3 4 10 6 7 8 9
> #thêm vào y 6 số
> y[10:15]=c(15,14,13,12,11,10)
> #xóa bớt đi 3 phần tử đầu tiên
> y[-(1:3)]
[1] 4 10 6 7 8 9 15 14 13 12 11 10
> #ghép hai vector
> z=c(x,y)
> z
[1] 1 2 3 4 5 6 7 8 9 1 2 3 4 10 6 7 8 9 15 14 13 12 11 10
```

➤ Chuyển dữ liệu số thành dữ liệu định danh

Quay trở về bảng dữ liệu diem đã nhập từ file excel phía trên. Ta sẽ tính cột điểm B và cột DGHP theo công thức

$$B = \frac{B_1 + B_2}{2}; \quad DGHP = 0.6A + 0.3B + 0.1C$$

```
> diem["B"]=(diem["B1"]+diem["B2"])/2
> diem["DGHP"]=round(diem["A"]*0.6+diem["B"]*0.3+diem["C"]*0.1,1)
> head(diem,5)
```

STT	maSV	hoten	X	ns	lop	A	B1	B2	B	C	DGHP	xeploai
1	2024010849	pham phuong	Anh	21/11/2002	DCKTKD65_A2	0	0	0	0.0	0	0.0	NA
2	1621060988	nguyen van chuc	chuc	13/06/1998	DCCDTD61_1	8	9	10	9.5	10	8.7	NA
3	1821060103	nguyen viet Cuong	Cuong	17/05/2000	DCCDTD63C	0	0	0	0.0	0	0.0	NA
4	1821050517	Trinh Viet Cuong	Cuong	4/08/2000	DCCTMM63A	7	9	10	9.5	10	8.1	NA
5	1821050824	Bui Quang Duy	Duy	27/01/2000	DCCTHT63A	5	10	9	9.5	10	6.8	NA

Sau đó tạo một vecto `phanloai` để phân loại điểm DGHP theo thang F, D, C, B, A

```
> diemhp=t(diem["DGHP"])
> diemhp
```

	[,1]	[,2]	[,3]	[,4]	[,5]	[,6]	[,7]	[,8]	[,9]	[,10]	[,11]	[,12]	[,13]	[,14]	[,15]	[,16]	[,17]	[,18]
DGHP	0	8.7	0	8.1	6.8	7	7.7	7.2	6.4	6.8	9.4	0	7.7	8.7	5.5	7.9	6.4	4.8

```
[,19]
DGHP 6.7
> length(diemhp)
[1] 19
> phanloai=diemhp
> phanloai[diemhp>=8]="A"
> phanloai[diemhp<8&diemhp>=6.5]="B"
> phanloai[diemhp<6.5&diemhp>=5.5]="C"
> phanloai[diemhp<5.5 &diemhp>=4]="D"
> phanloai[diemhp<4]="F"
> phanloai
```

	[,1]	[,2]	[,3]	[,4]	[,5]	[,6]	[,7]	[,8]	[,9]	[,10]	[,11]	[,12]	[,13]	[,14]	[,15]	[,16]	[,17]	[,18]
DGHP	"F"	"A"	"F"	"A"	"B"	"B"	"B"	"B"	"C"	"B"	"A"	"F"	"B"	"A"	"C"	"B"	"C"	"D"

```
[,19]
DGHP "B"
```

Sau đó gán vector `xeploai` trong bảng dữ liệu bằng vector `phanloai` vừa tạo

```
> rownames(phanloai)=c("xeploai")
> diem["xeploai"]=t(phanloai)
> head(diem,5)
```

STT	maSV	hoten	X	ns	lop	A	B1	B2	B	C	DGHP	xeploai
1	2024010849	pham phuong	Anh	21/11/2002	DCKTKD65_A2	0	0	0	0.0	0	0.0	F
2	1621060988	nguyen van chuc	chuc	13/06/1998	DCCDTD61_1	8	9	10	9.5	10	8.7	A
3	1821060103	nguyen viet Cuong	Cuong	17/05/2000	DCCDTD63C	0	0	0	0.0	0	0.0	F
4	1821050517	Trinh Viet Cuong	Cuong	4/08/2000	DCCTMM63A	7	9	10	9.5	10	8.1	A
5	1821050824	Bui Quang Duy	Duy	27/01/2000	DCCTHT63A	5	10	9	9.5	10	6.8	B

2.4. Bảng dữ liệu và truy cập vào bảng dữ liệu

Ta có thể ghép các vector để tạo thành bảng dữ liệu bằng lệnh

```
data.frame(c1,c2,...,cn,)
```

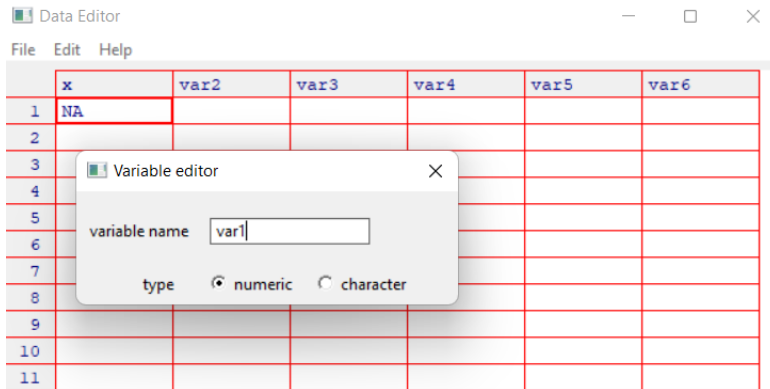
Trong đó, `c1, c2,..., cn`: là các vector cần ghép

```
> L3 <- LETTERS[1:3]
> fac <- sample(L3, 5, replace = TRUE)
> (d <- data.frame(x = 1, y = 1:5, fac = fac))
```

x	y	fac
1	1	C
2	1	B
3	1	B
4	1	A
5	1	C

Ngoài ra ta cũng có thể nhập dữ liệu cho bảng tên là `dulieu` bằng cách dùng hàm `edit` theo lệnh sau

```
dulieu=edit(data.frame())
```



Một số thao tác khác như đổi tên cho các cột, đặt tên cho dòng, thêm vào bảng một số cột, dòng, hay chỉnh sửa dữ liệu ở một vị trí nào đó trong bảng

Ví dụ. Tạo một bảng

```
> chieucac=c(170, 165, 155, 157,166)
> canngang=c(65,55,45,50,60)
> gioitinh=c("nam","nam","nu","nu","nam")
> Dulieu=data.frame(chieucac, canngang, gioitinh)
> Dulieu
```

```
  chieucac canngang gioitinh
1      170      65      nam
2      165      55      nam
3      155      45       nu
4      157      50       nu
5      166      60      nam
```

```
> tuoi=c(23,25,24,30,27)
> #ghép thêm cột tuổi
> Dulieu=data.frame(Dulieu,tuoi)
> Dulieu
```

```
  chieucac canngang gioitinh tuoi
1      170      65      nam   23
2      165      55      nam   25
3      155      45       nu   24
4      157      50       nu   30
5      166      60      nam   27
```

```
> #đổi tên cột tuổi thành age
> colnames(Dulieu)[4]="age"
> Dulieu
```

```
  chieucac canngang gioitinh age
1      170      65      nam   23
2      165      55      nam   25
3      155      45       nu   24
4      157      50       nu   30
5      166      60      nam   27
```

Ta cũng có thể dùng hàm `cbind` để ghép thêm cột và `rbind` để ghép thêm dòng.

```
> chieucac=c(170)
> canngang=c(62)
> gioitinh=c("nam")
> age=c(22)
> dulieuthem=data.frame(chieucac, canngang, gioitinh, age)
> dulieuthem
  chieucac canngang gioitinh age
1      170      62      nam   22
```

```
> gop=rbind(Dulieu,dulieuthem)
> gop
  chieucao cannang gioitinh age
1      170      65      nam  23
2      165      55      nam  25
3      155      45       nu  24
4      157      50       nu  30
5      166      60      nam  27
6      170      62      nam  22
```

Ngoài ra ta vẫn có thể dùng hàm `edit` để trực tiếp chỉnh sửa dữ liệu. Một số lệnh thường dùng để truy cập vào bảng dữ liệu như sau

Lệnh	Công dụng
<code>Dim(A)</code>	Cho biết số dòng, số cột của A
<code>Nrow(A), ncol(A)</code>	Cho biết số dòng, số cột của A
<code>Names(A)</code>	Cho biết tên các cột của A
<code>A[i,j]</code>	Cho biết giá trị của phần tử ở dòng i, cột j của A
<code>A[i,]</code>	Cho biết thông tin dòng i
<code>A[, j]</code>	Cho biết thông tin cột j
<code>A[c(i,j),]</code>	Cho biết thông tin dòng i, j của A
<code>A[, c(i,j)]</code>	Cho biết thông tin cột i và j của A
<code>Subset(A, dieukien)</code>	Đưa ra những phần tử trong A thỏa mãn điều kiện
<code>Cut(m:n,c(a,b))</code>	Lấy dữ liệu từ dòng m đến n ở cột a và b

```
> dim(Dulieu)
[1] 5 4
> names(Dulieu)
[1] "chieucao" "cannang" "gioitinh" "age"
> subset(Dulieu,gioitinh=="nu")
  chieucao cannang gioitinh age
3      155      45       nu  24
4      157      50       nu  30
> #Lấy dữ liệu ở cột 1 và 3 của 3 dòng đầu tiên
> cut=Dulieu[1:3,c(1,3)]
> cut
  chieucao gioitinh
1      170      nam
2      165      nam
3      155       nu
```

Tiếp theo ta thực hành một ví dụ đọc một file điểm từ trong một thư mục trong máy, lọc tất cả những sinh viên vắng mặt vào bảng “vang”, lọc tất cả những sinh viên tham gia thi vào bảng “dsthi” sau đó tính điểm DGHP cho những sinh viên này.


```

> #truy cập vào thư mục chứa dữ liệu
> setwd("F:/hang/diem b,c/he 2020-2021")
> #đọc dữ liệu và lưu vào đối tượng diem
> diemhe=read.csv("test.csv")
> vang=subset(diemhe,diemhe["A"]=="V")
> head(vang,3)
  STT      MSV      hoten      X      ngaysinh      lop      A      B1      B2      B3      C1      DGHP
20  20  2021060387  Nguy?n V?n D?ng  24/02/2002  DCCDTD65  V    9    8    8    9    NA
25  25  2021060392  Hoàng V?n  ??i  15/07/2002  DCCDTD65  V    0    0    0    0    NA
33  33  2021060531  ?âm Hoàng  ??c  17/11/2002  DCCDTD65  V    9   10   10   10    NA
> dsthi=subset(diemhe,diemhe["A"]!="V")
> head(dsthi,3)
  STT      MSV      hoten      X      ngaysinh      lop      A      B1      B2      B3      C1      DGHP
1   1  2021060367  Nguy?n Duy   An  25/12/2002  DCCDTD65  3.00  10    9    9   10    NA
2   2  2021060368  D??ng Tu?n  Anh  01/05/2002  DCCDTD65  2.50   6    6    6    8    NA
3   3  2021060370  Lê V?n Hoàng Anh  26/09/2002  DCCDTD65  0.50   8    7    7    9    NA
> dsthi[,7]=as.numeric(dsthi[,7])
> dsthi["DGHP"]=(dsthi["A"])*0.3+(dsthi["B1"]+dsthi["B2"])*0.3+dsthi["C1"]*0.1
> head(dsthi,5)
  STT      MSV      hoten      X      ngaysinh      lop      A      B1      B2      B3      C1      DGHP
1   1  2021060367  Nguy?n Duy   An  25/12/2002  DCCDTD65  3.0  10    9    9   10  7.60
2   2  2021060368  D??ng Tu?n  Anh  01/05/2002  DCCDTD65  2.5   6    6    6    8  5.15
3   3  2021060370  Lê V?n Hoàng Anh  26/09/2002  DCCDTD65  0.5   8    7    7    9  5.55
4   4  2021060371  Nguy?n Minh  Anh  02/10/2002  DCCDTD65  3.0   9    9    9   10  7.30
5   5  2021060373   Ph?m Ng?c  Anh  17/12/2002  DCCDTD65  4.0   9    8    8    9  7.20

```

KẾT LUẬN

Báo cáo đã trình bày một số vấn đề cơ bản nhất cho người bắt đầu tìm hiểu về R như cài đặt, các khái niệm và cách khai báo số, truy cập vào số liệu. Với mục đích xây dựng giáo án và cung cấp tài liệu tự học cho sinh viên, báo cáo trình bày ngắn gọn, lấy ví dụ đơn giản nhất. Trong thời gian sắp tới, tác giả hi vọng sẽ cùng các đồng nghiệp xây dựng giáo án xác suất thống kê với phần mềm R hoàn thiện hơn và tìm hiểu thêm những ứng dụng chuyên sâu của R.

TÀI LIỆU THAM KHẢO

1. Nguyễn Thanh Hồng, Nguyễn Thị Nhung, *Thống kê ứng dụng hướng dẫn thực hành trên phần mềm R*, NXB Thống kê, 2015.
2. Đặng Hùng Thắng, Trần Mạnh Cường, *Thống kê cho Khoa học xã hội và Khoa học sự sống (với phần mềm R)*, NXB Đại học Quốc gia Hà Nội, 2019.