

# STAT 3008 Applied Regression Analysis

## Tutorial 1. **R** for STAT3008

XU Yongze @ 143

39431747

[colexyz18@gmail.com](mailto:colexyz18@gmail.com)

Sep 9 & 10& 12, 2013

### **Contents**

# 1 Download & Install R

1. R web-site at [www.r-project.org](http://www.r-project.org)
2. CRAN → Find China and click one of those links → download R for Windows → base → Download R 2.15.2 for Windows
3. Save and install

# 2 Quick start using R

The best way to learn a statistical package or computer language is to get started using it as soon as possible.

- **Simple arithmetic expression:**

```
> 3 + 7    #can be used as a simple calculator
```

```
[1] 10
```

```
> 3 * 21
```

```
[1] 63
```

- **Generate Sequence:**

```
> 1:7    #generate sequence from 1 to 7
```

```
[1] 1 2 3 4 5 6 7
```

- **Function for 'combining':**

```
>c(3,4,1,6)    #combining numbers into a vector
```

```
[1] 3 4 1 6
```

- **Assignment:**

```
>newvec = c(3,4,1,6)    #assign the vector to newvec
```

```
>newvec
```

```
[1] 3 4 1 6
```

- **Open and save your own script:**

File → New script/Open script

coding → select lines → right-click → Run line

**Attention:**

1. hash mark(#) is used to denote comments.
2. R is case sensitive

## 3 Data Objects

Data are stored in objects in R. This is an abstract term that can be assigned to a variable. There are many types of objects, e.g. vector, matrix ... etc. We will discuss the vector & matrix for our course.

### 3.1 Vector

```
>x = c(5:1)    # numeric vector
```

```
>x
```

```
[1] 5 4 3 2 1
```

```
>length(x)    # display the length of x
```

```
[1] 5
```

```
>z = c("a","b","c")    #character vector
```

```
>length(z)
```

```
[1] 3
```

```
>y = x    #assign value from another vector
```

```
>c(x,y)    #combine two vectors
```

```
[1]5 4 3 2 1 5 4 3 2 1
```

## 3.2 Matrix

Matrix is a collection of numbers in a rectangular form. A matrix with dimension n by m means that the matrix has n rows and m columns.

```
>m = matrix(1:12,nrow = 3, ncol = 4) # 3 by 4 matrix in columnwise
```

```
>m
```

```
  [,1] [,2] [,3] [,4]
[1,]  1   4   7  10
[2,]  2   5   8  11
[3,]  3   6   9  12
```

```
>matrix(1:12,nrow = 3, ncol = 4,byrow = T) # 3 by 4 matrix in rowwise
```

```
  [,1] [,2] [,3] [,4]
[1,]  1   2   3   4
[2,]  5   6   7   8
[3,]  9  10  11  12
```

```
>m[2,] # select the 2nd row
```

```
[1] 2 5 8 11
```

```
>m[,3] #select the 3rd column
```

```
[1] 7 8 9
```

```
>m[2,3] #select an element
```

```
[1] 8
```

```
>m[1:2,2:4] #select a submatrix
```

```
  [,1] [,2] [,3]
[1,]  4   7  10
[2,]  5   8  11
```

## 4 Expressions and Operators

In R, expressions and operators are very easy to follow, and operations can be performed on each component of the vector and matrix, which is called the **vectorized** operation.

```
>x = c(4:1)
```

```
>2*x #operate on each element of x
```

```
[1] 8 6 4 2
```

In our course, we need to perform some operations on vector and matrix. Here are some important operators:

Operator	Meaning
% * %	matrix multiply
* /	multiply, divide
+ -	add, subtract

## 5 Built-in Functions

There are many built-in function in R. Besides the common mathematical functions like sqrt, abs sin, cos, log, exp, ... etc. There are functions worth mentioning:

Name	Operations
ceiling	smallest integer greater than or equal to element
floor	largest integer less than or equal to element
trunc	ignore the decimal part
round	round up for positive and round down for negative
sort	sort the vector in ascending or descending order
sum, prod	sum and produce of a vector
cumsum, cumprod	cumulative sum and product
min, max	return the smallest and largest values
range	return a vector of length 2 containing the min and max
mean	return the sample mean of a vector
var	return the sample variance of a vector
sd	return the sample standard deviation of a vector
seq	generate a sequence of number
rep	replicate elements in a vector

In our course, we may need to get the transpose of vector and matrix, as well as the inverse of non-singular square matrix:

```
>t(x) #transpose of a matrix
```

```
>solve(x) #inverse of a matrix, You should make sure that x has inverse
```

```
>tr(x) #trace of a matrix
```

## 6 On-line Help(This is the most important!!!!)

A very important approach to learn R programming is to learn from its on-line help. The on-line help feature provides an easy and convenient way to learn the built-in functions. It comes with the syntax, explanation, reference, related functions and examples on the usage of that function. For example, see help on the function, *sort*, *solve*. Please, remember to use the on-line help when using R!

A fast way to search for the key words:

**Command "** >help.start() " → Search Engine & Keywords → input the keywords

## 7 Examples for STAT3008

#some hint code for assignments will be provided in later tutorial

### 7.1 Hat matrix in regression analysis

You may find the hat matrix  $H = X(X'X)^{-1}X'$  in the Revision Exercise.

Let's try to calculate the hat matrix of  $X = \begin{pmatrix} 1 & 2 \\ 3 & 4 \end{pmatrix}$

```
>X = matrix(c(1,2,3,4),ncol = 2, byrow = T)    #create matrix X
>inverse_in_middle = solve(t(X)% * % X)      #get the inverse of(X'X)
>H = X % * % inverse_in_middle % * % t(X)    #get the hat matrix
```

### 7.2 Example in linear regression

In this subsection, a full program of linear regression will be presented, the first part of which is in the last slide of chapter1

**you should download a package first:**

Package → Install Package → select CRAN mirror (USA(CA1)is the database in Berkeley)  
→ alr3 (click OK to install)

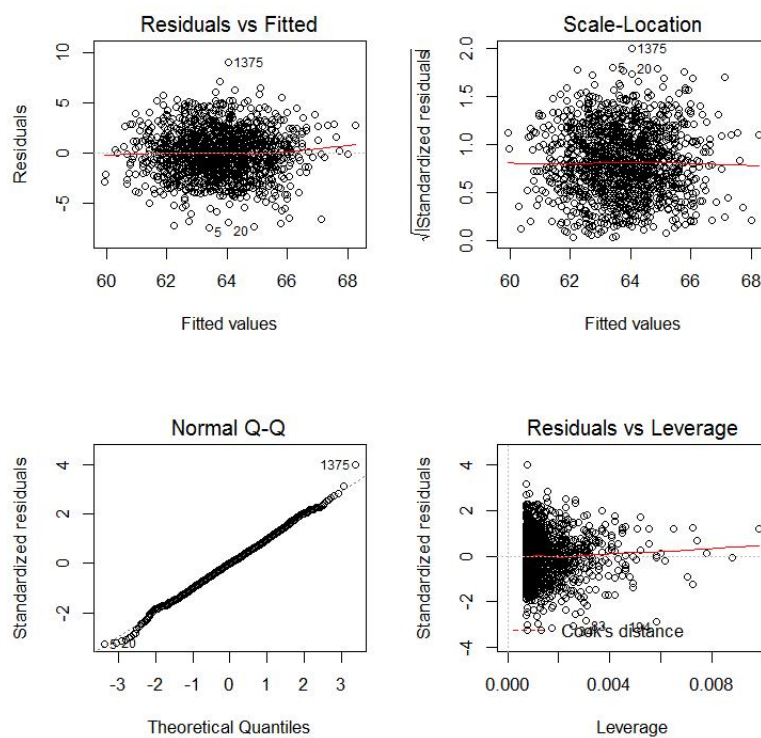
```
>library(alr3)    # to load the library
>data(heights)   # load the data
>x = heights     # x is a list variable. It contains two vectors. See reference notes or book to
know more about "list".
>plot(x$Mheight,x$Dheight)    # Scatter plot. $ is the operator to select component.
```

```

>fit = lm(x$Dheight ~ x$Mheight)    #fit the regression model
>abline(a = fit$coef[1],b = fit$coef[2])    #add fitted regression line in the scatter plot
>fit    # see what is in "fit"
>summary(fit)    #detail result of the fitted regression
>layout(matrix(1:4,2,2))    # set the layout of the plots we need.
>plot(fit)    #the first plot is the residual plot

```

The corresponding plot is:



### 7.3 Read data into R

In this course, we will only use simple and standard data file types.

When you want to read a data set, you have to put the file **under the direction** of R:  
file → change dir... (select file folder contains the data set)

Assume we have a data set "HtWt.txt":

```

> x = read.table("HtWt.txt",header = T)    #read the data set, and the first row of the data

```

set is the labels.

```
> x$Ht #component Ht in x
```

```
> x$Wt #component Wt in x
```

**# see the reference for more ways to read data**

## 8 Reference Book & Notes

- 1. STAT 2009 R Notes Prepared by Prof. Leung
- 2. Chan Ngai Hang, Leung Pui lam, An introduction to R. (in Chinese)

