

STAT 3008
APPLIED REGRESSION ANALYSIS
TUTORIAL 1: Introduction

LAI Chun Hei*

Department of Statistics, The Chinese University of Hong Kong

1 Mathematical Background

In this courses, some theorems from elementary linear algebra, calculus and statistics would be applied. This section is a reminder of some concepts in those topics that are frequently applied in this course.

1.1 Linear Algebra

Linear Independent Vectors

Definition 1.1. A set of vectors $\{\underline{v}_1, \underline{v}_2, \dots, \underline{v}_k\}$, where $\underline{v}_i \in \mathbb{R}^n$, $i=1,2,\dots,k$, is linearly independent if the system of linear equation

$$x_1\underline{v}_1 + x_2\underline{v}_2 + \dots + x_k\underline{v}_k = \mathbf{0}$$

has only the trivial solution

$$x_1 = x_2 = \dots = x_n = 0$$

i.e. no vector in the set is a linear combination of other vectors.

Inverse of Matrix

Definition 1.2. Let $A, B \in \mathbb{R}^{n \times n}$, if

$$AB = BA = I_n$$

where I is the identity matrix, B is called the inverse of A , we write $B = A^{-1}$, and vice versa. In this case, we say A and B are non-singular.

Theorem 1.1. The following statements are equivalent for $A \in \mathbb{R}^{n \times n}$:

1. A is non-singular.
2. $\det(A) \neq 0$
3. The column vectors of A are linearly independent.
4. The row vectors of A are linearly independent.

Theorem 1.2. If $A \in \mathbb{R}^{n \times n}$ is non-singular, then the inverse of A is given by

$$A^{-1} = \frac{1}{\det(A)} \text{adj}(A)$$

where $\text{adj}(A) \triangleq C^T$, i.e. the transpose of the cofactor matrix of A whose (i,j) entry is the (i,j) cofactor of A .

*s1155002282@sta.cuhk.edu.hk

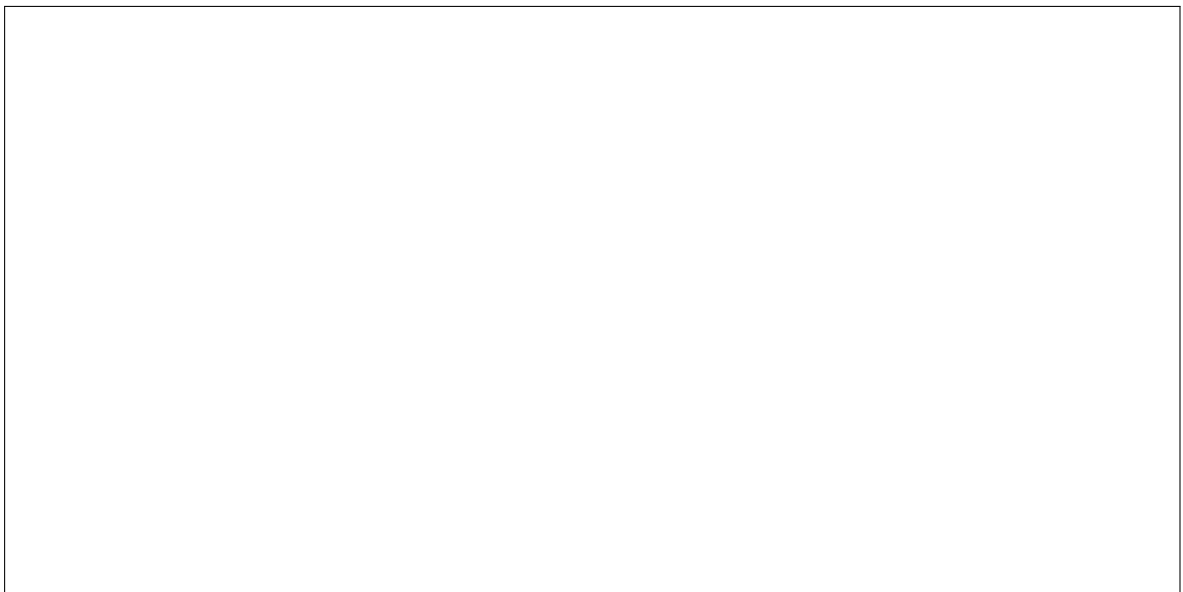
Exercise 1.1. Find the inverse of the following matrices.

$$A = \begin{bmatrix} 4 & 3 \\ 3 & 2 \end{bmatrix} \quad \text{and} \quad B = \begin{bmatrix} 1 & 2 & 3 \\ 0 & 1 & 4 \\ 5 & 6 & 0 \end{bmatrix}$$



Exercise 1.2. Find the determinant of the following matrix.

$$A = \begin{bmatrix} 4 & 0 & 0 & 0 \\ 0 & 0 & 2 & 0 \\ 0 & 1 & 2 & 0 \\ 1 & 0 & 0 & 1 \end{bmatrix}$$



1.2 Basic Statistics

In this course, the mostly used concepts from your basic statistic courses are the results of transforming random variables. They are listed below.

Theorem 1.3. Let $X \sim \mathcal{N}(\mu, \sigma^2)$, then

$$\left(\frac{X - \mu}{\sigma}\right)^2 \sim \chi^2(1)$$

Theorem 1.4. Let V_1, V_2, \dots, V_n be i.i.d. random variables where $V_i \sim \chi^2(k_i), i = 1, 2, \dots, n$, then

$$V_1 + V_2 + \dots + V_n \sim \chi^2\left(\sum_{i=1}^n k_i\right)$$

Theorem 1.5. Let $Z \sim (N)(0, 1)$ and $V \sim \chi^2(k)$ be independent, then

$$T = \frac{Z}{\sqrt{V/k}} \sim t(k)$$

Theorem 1.6. Let $V_1 \sim \chi^2(k_1)$ and $V_2 \sim \chi^2(k_2)$, then

$$F = \frac{V_1/k_1}{V_2/k_2} \sim F(k_1, k_2)$$

In statistics, tests and estimations are often based on the above theorems, and STAT3008 is no exception. You may find plenty of application of them in the construction of confidence intervals and tests very soon.

Exercise 1.3. Prove Theorem 1.4.

Exercise 1.4. Provide an example of hypothesis testing and explain the philosophy of a test.

2 R-Programming

There are plenty of statistics oriented programming languages, to name but a few, R, S-Plus and SAS. Among all of them, the most commonly used one is R, thanks to its open-source structure and that it is free.

R is an interpreted programming language, as posed to compiled languages. Users are free to develop statistical packages which targets different applications, such as linear regression, survival analysis and time series analysis. In this course, the knowledge of R definitely makes doing homework easier and greatly reduce the time of finishing your assignments.

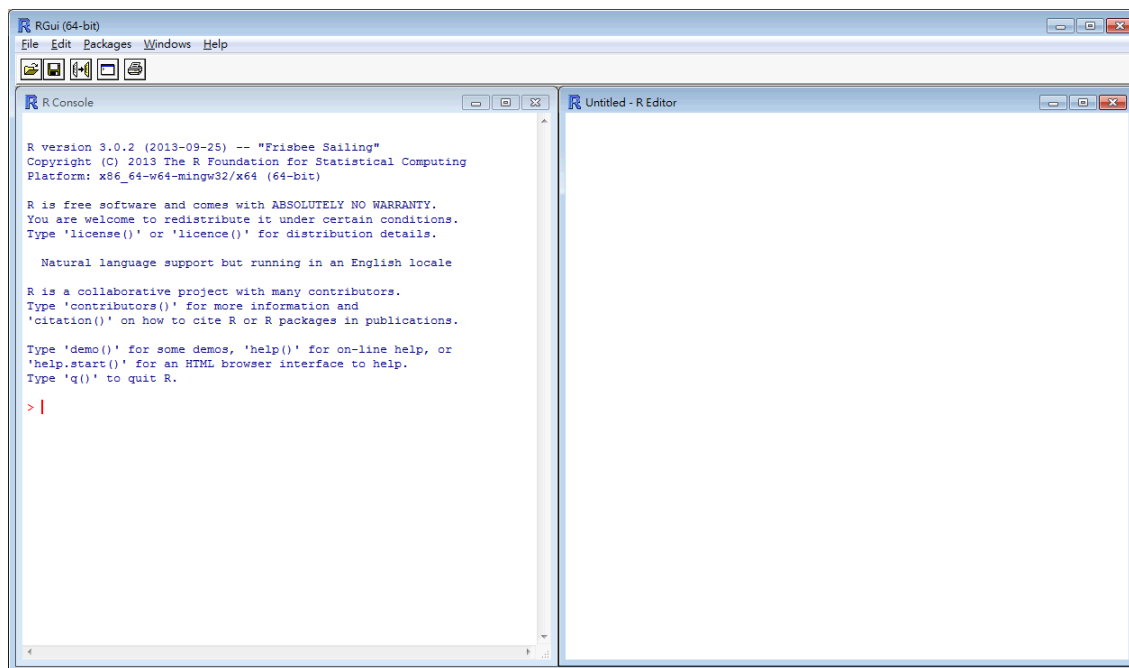
2.1 Installation

The installation executable can be found in the following address. You are free to choose any mirror which fits you. Set-up files for both Windows and Mac are both available. For Linux, usually the distribution comes with R. Yet, the Linux version can also be found.

<http://cran.r-project.org/mirrors.html>

2.2 Interface

The basic graphical user interface of R is shown below.



On the left, it is the R console where the codes would be run upon pressing the enter key. On the right, the R Editor is the place where you can first construct the codes, without running them line by line when you write them.

It is highly recommend that you use the R editor to develop codes. This helps prevent the situation that the value of a variable is overwritten too many times that you lose track of what it is. Also, for extra long programmes, this reduce the development time as you do not have to run the codes whenever you have a new line.

2.3 Basic Syntax

Below, some syntax for basic computation is provided. You may find them helpful during finishing your homework.

Construction of Lists and Vectors

The following codes construct a list, namely `x`, which contains the number 3, 10, 20, 40. Then the second part makes it into a vector, `y`.

```
x<-c(3,10,20,40) #construction of list, method 1
y<-as.vector(x) #makes x into a vector y
#Equivalently, you can directly do the following
y<-as.vector(c(3,10,20,40))
```

Construction of Matrix from List The easiest way to construct a matrix in R is to use a list. The elements of the matrix should be listed from first left to right, then up to down. The following codes construct the matrix

$$\begin{bmatrix} 1 & 8 & 3 \\ 6 & 4.3 & 4 \\ 5 & 6 & -1 \end{bmatrix}$$

and assign it to the variable `A`.

```
Aelem<-c(1,8,3,6,4.3,4,5,6,-1)
A<-matrix(Aelem,nrow=3,ncol=3,byrow=TRUE)
#The "byrow" argument is bydefault FALSE,
#i.e. the matrix is filled from up to down then left to right.
```

Matrix Computation If `A` and `B` are matrices that are already defined, then the following codes returns different values associated with the matrix.

```
A*B      #elementwise multiplication
A%%B     #matrix multiplication
det(A)   #determinant of A
solve(A) #inverse of A
dim(A)   #dimension of A
```

The above codes will be useful in Chapter 3, multiple linear regression. However, it is highly recommended that you familiarise yourself with matrix computation by hand, as you will *NOT* be provided access to a computer during midterm and examination.

Exercise 2.1. Use R to compute the determinant of the matrix in Exercise 1.2.

Answer. ¹

3 Appendix

For more reference on linear algebra and statistics, you may refer to the following text books.

References

- [1] STEVEN J. LEON (2006). *Linear Algebra with Applications*, Pearson.
- [2] ROBERT V. HOGG AND ALLEN T. CRAIG *Introduction to Mathematical Statistics*, Pearson.

¹-8.