

Basic R



최 용 석 (부산대학교 통계학과)

[yschoi.pusan.ac.kr](http://yschoi.pusan.ac.kr)

2019.08.21.

부산대학교 통계연구소 하계 워크숍

## 주제 2 : R 기초 및 사용법

### 1. R의 설치

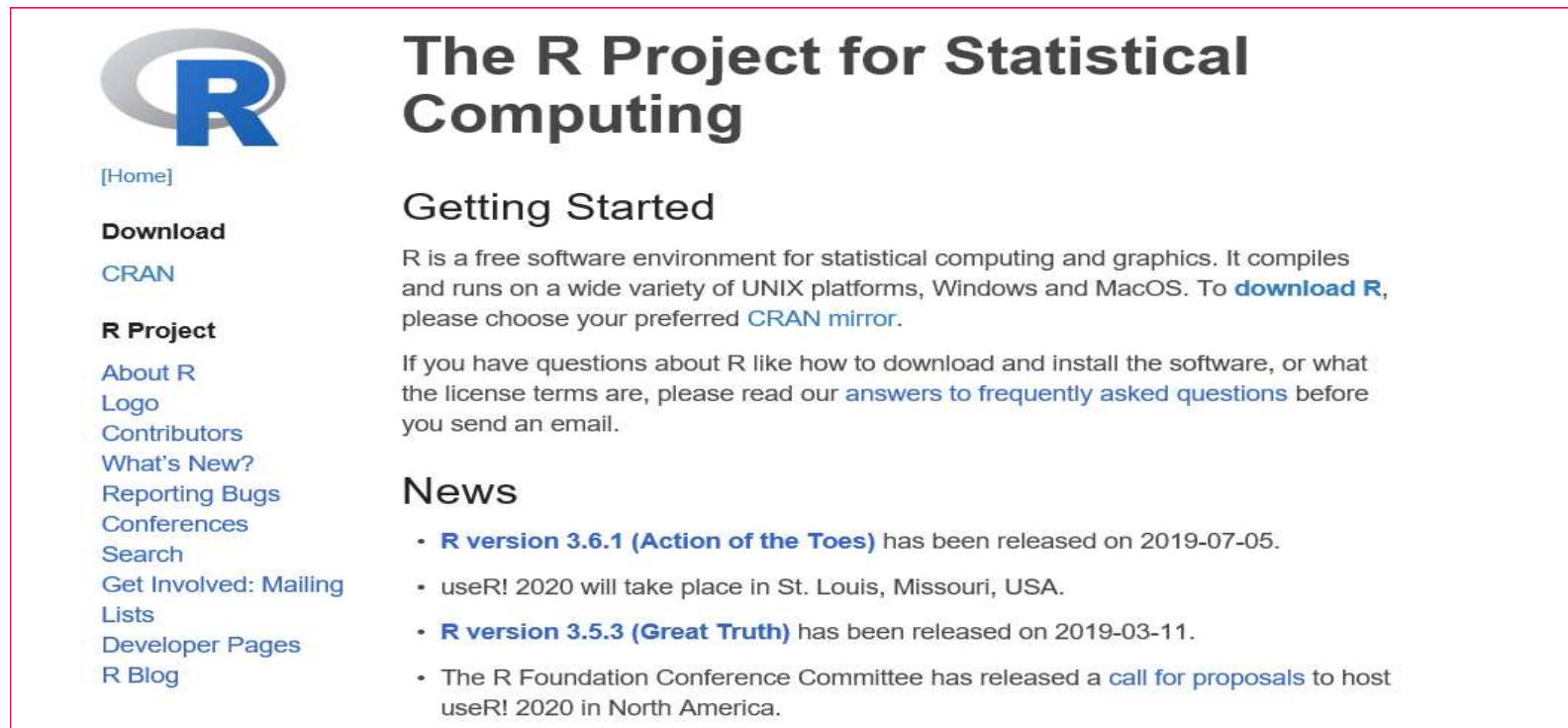
### 2. R의 기본 사용법

### 3. R 프로그램

- (1) 기초 : R 명령문의 입력 및 실행 방법
- (2) 작업 폴더(working directory) 설정하기
- (3) R에서 자료 입력(직접입력 방식)
- (4) R에서 자료 불러오기(텍스트파일)
- (5) R에서 자료 불러오기(엑셀파일)
- (6) R에서 자료 내보내기(엑셀파일)

## 1. R의 설치

- (1) 인터넷에서 [<http://www.r-project.org/>]로 접속한다.
- (2) 화면 왼쪽의 메뉴에서 [Download]의 [CRAN]을 클릭한다.



[그림 1.1] R 프로그램 홈페이지

- (3) 제시된 목록에서 [Korea]를 찾아 [http://cran.biodisk.org] 또는 [http://healthstat.snu.ac.kr/CRAN/] 중 하나를 클릭한다.

### CRAN Mirrors

The Comprehensive R Archive Network is available at the following URLs, please choose a location close to you. Some statistics on the status of the mirrors can be found here: [main page](#), [windows release](#), [windows old release](#).

If you want to host a new mirror at your institution, please have a look at the [CRAN Mirror HOWTO](#).

**0-Cloud**

<a href="https://cloud.r-project.org/">https://cloud.r-project.org/</a>	Automatic redirection to servers worldwide, currently sponsored by Rstudio
<a href="http://cloud.r-project.org/">http://cloud.r-project.org/</a>	Automatic redirection to servers worldwide, currently sponsored by Rstudio

**Algeria**

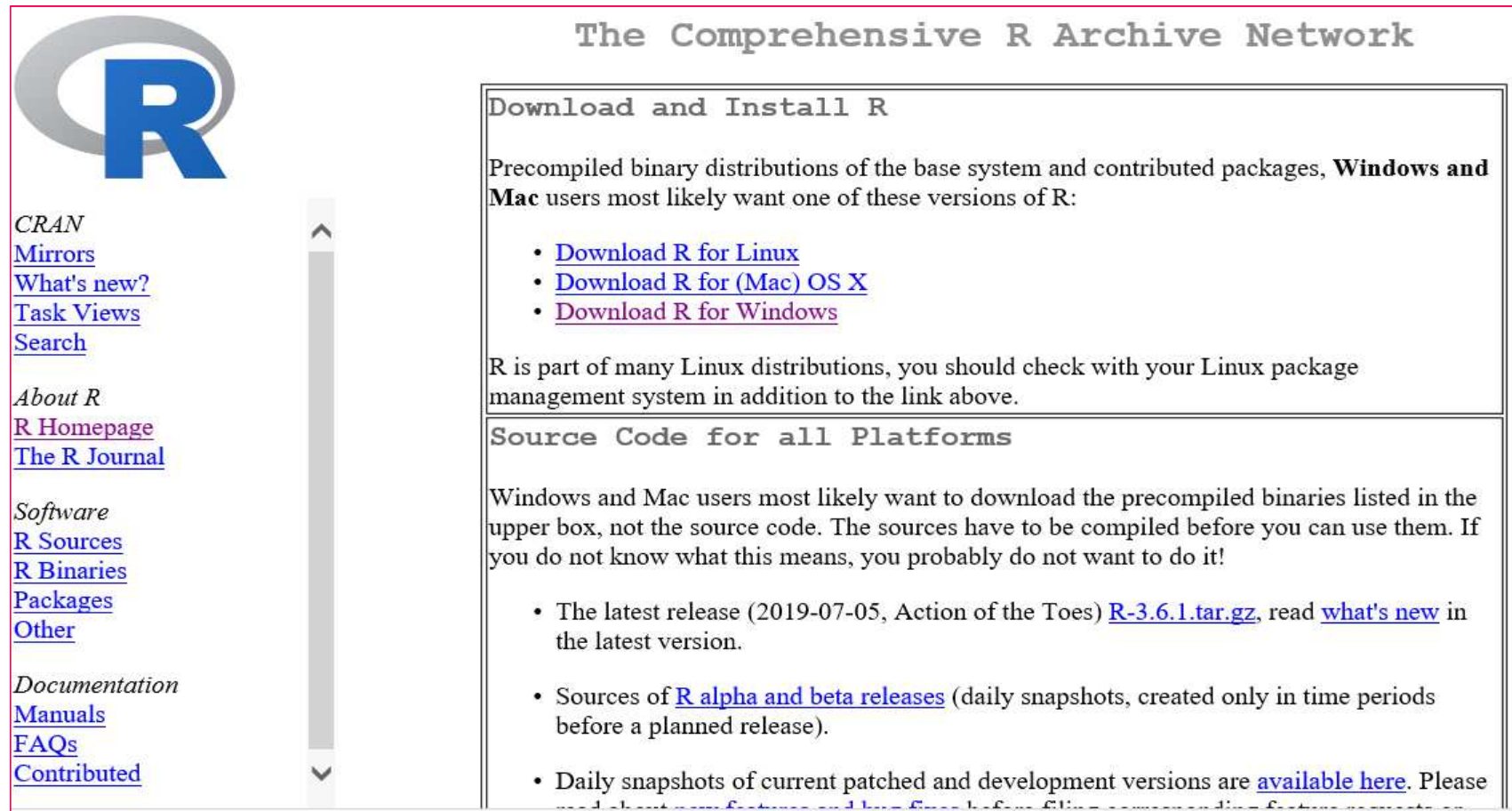
<a href="https://cran.usthb.dz/">https://cran.usthb.dz/</a>	University of Science and Technology Houari Boumediene
<a href="http://cran.usthb.dz/">http://cran.usthb.dz/</a>	University of Science and Technology Houari Boumediene

**Korea**

<a href="https://ftp.harukasan.org/CRAN/">https://ftp.harukasan.org/CRAN/</a>	Information and Database Systems Laboratory, Pukyong National University
<a href="https://cran.yu.ac.kr/">https://cran.yu.ac.kr/</a>	Yeungnam University
<a href="https://cran.seoul.go.kr/">https://cran.seoul.go.kr/</a>	Bigdata Campus, Seoul Metropolitan Government
<a href="http://healthstat.snu.ac.kr/CRAN/">http://healthstat.snu.ac.kr/CRAN/</a>	Graduate School of Public Health, Seoul National University, Seoul
<a href="https://cran.biodisk.org/">https://cran.biodisk.org/</a>	The Genome Institute of UNIST (Ulsan National Institute of Science and Technology)

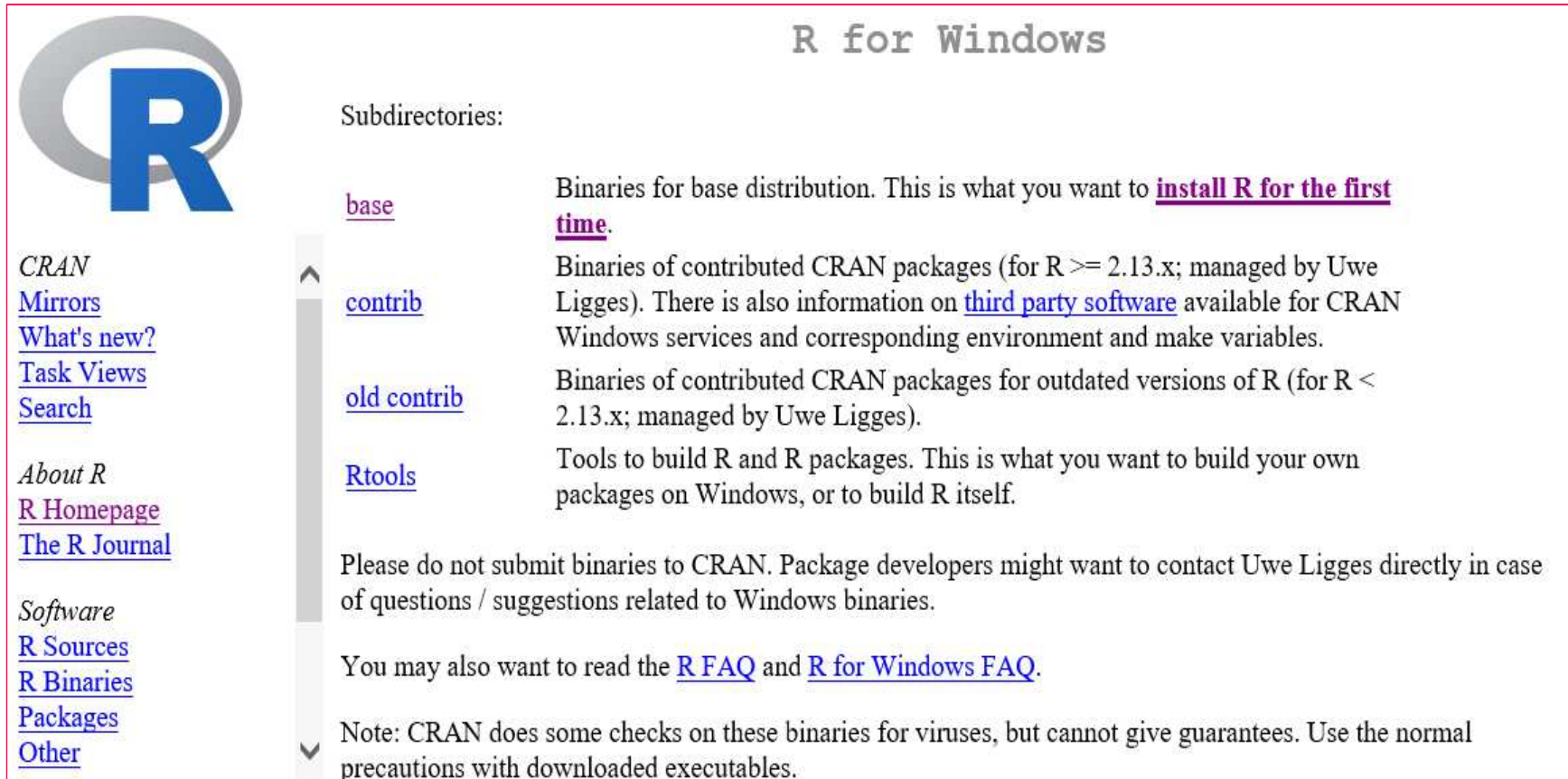
[그림 1.2] 프로그램을 다운 받을 수 있는 여러 나라 사이트

(4) 화면에서 [Download R for Windows]를 클릭한다.



[그림 1.3] OS 환경에 맞게 R 프로그램 다운 받기

(5) 화면에서 [base]를 클릭한다.



The screenshot shows the 'R for Windows' website. On the left is a sidebar with the R logo and various links. The main content area is titled 'R for Windows' and lists subdirectories. The 'base' subdirectory is selected and expanded, showing its description and links to other resources.

**R for Windows**

Subdirectories:

- [base](#): Binaries for base distribution. This is what you want to [install R for the first time](#).
- [contrib](#): Binaries of contributed CRAN packages (for R  $\geq$  2.13.x; managed by Uwe Ligges). There is also information on [third party software](#) available for CRAN Windows services and corresponding environment and make variables.
- [old contrib](#): Binaries of contributed CRAN packages for outdated versions of R (for R  $<$  2.13.x; managed by Uwe Ligges).
- [Rtools](#): Tools to build R and R packages. This is what you want to build your own packages on Windows, or to build R itself.

Please do not submit binaries to CRAN. Package developers might want to contact Uwe Ligges directly in case of questions / suggestions related to Windows binaries.

You may also want to read the [R FAQ](#) and [R for Windows FAQ](#).

Note: CRAN does some checks on these binaries for viruses, but cannot give guarantees. Use the normal precautions with downloaded executables.

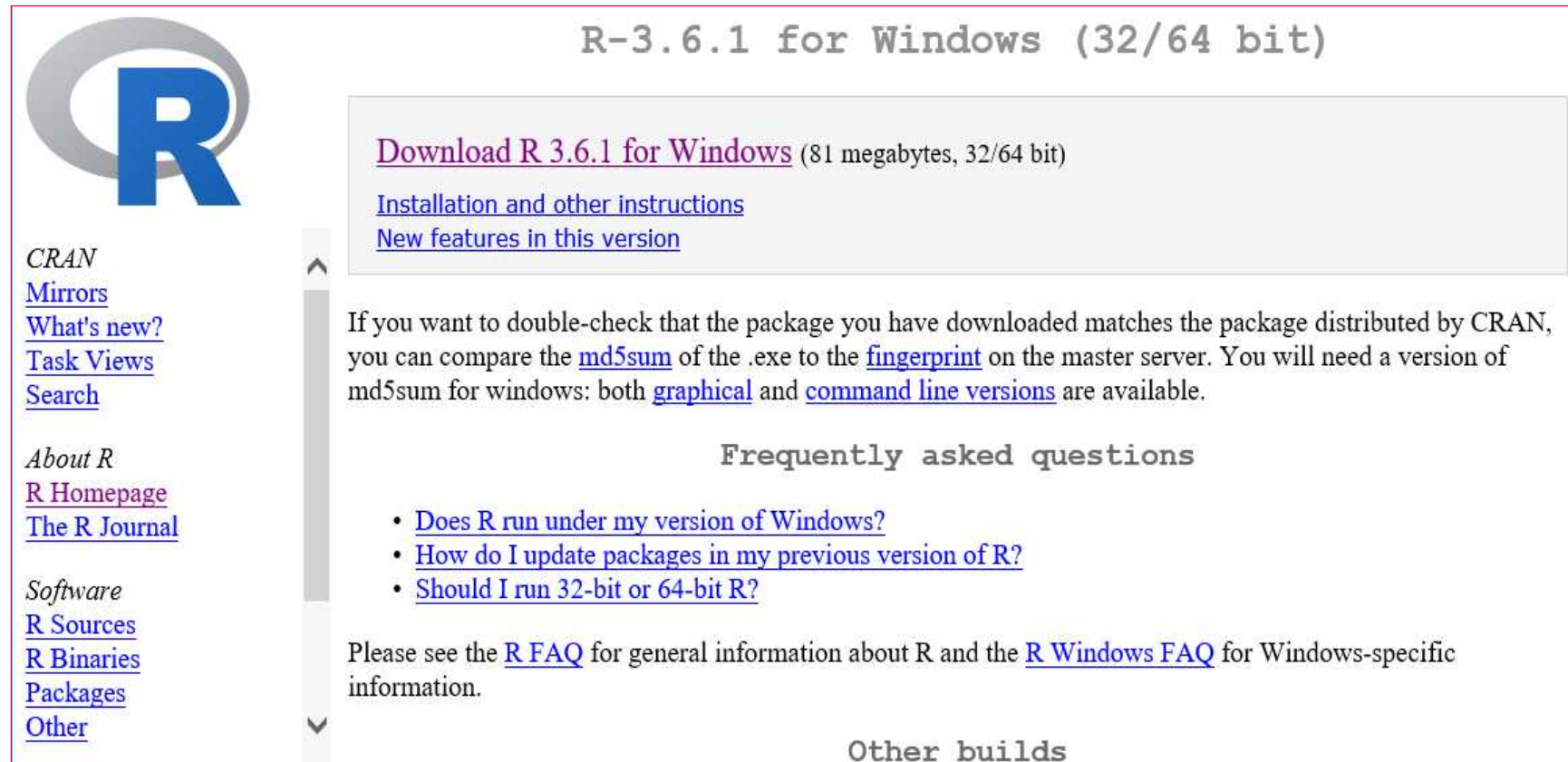
**Left Sidebar Links:**

- CRAN
- [Mirrors](#)
- [What's new?](#)
- [Task Views](#)
- [Search](#)
- About R
- [R Homepage](#)
- [The R Journal](#)
- Software
- [R Sources](#)
- [R Binaries](#)
- [Packages](#)
- [Other](#)

[그림 1.4] base를 가져 올 화면



(6) [Download R 3.6.1 for Windows]를 클릭하고, R 3.1.0 버전의 설치 파일인 [R-3.1.0-win.exe]를 다운로드하여 실행하고 설치 언어 [한국어]를 선택 설치하면 된다.



**R-3.6.1 for Windows (32/64 bit)**

[Download R 3.6.1 for Windows](#) (81 megabytes, 32/64 bit)

[Installation and other instructions](#)

[New features in this version](#)

If you want to double-check that the package you have downloaded matches the package distributed by CRAN, you can compare the [md5sum](#) of the .exe to the [fingerprint](#) on the master server. You will need a version of md5sum for windows: both [graphical](#) and [command line versions](#) are available.

**Frequently asked questions**

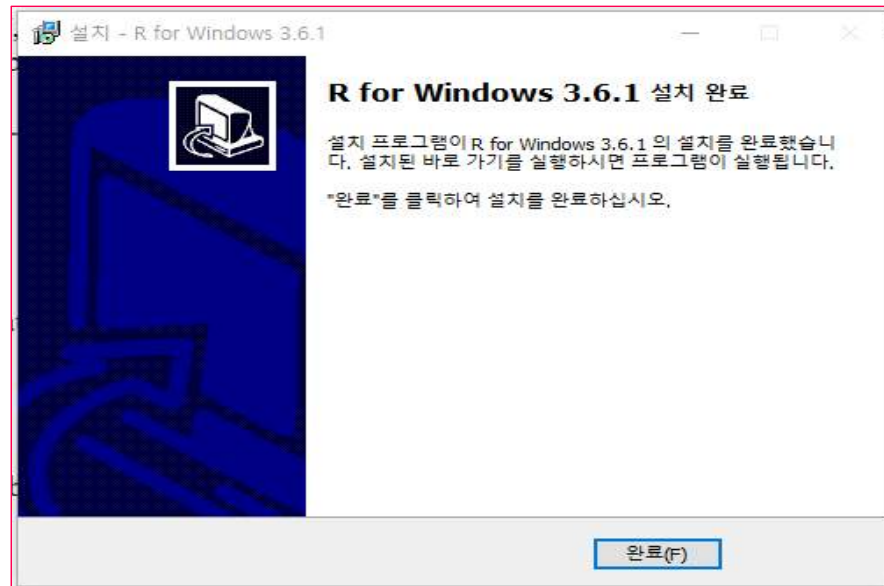
- [Does R run under my version of Windows?](#)
- [How do I update packages in my previous version of R?](#)
- [Should I run 32-bit or 64-bit R?](#)

Please see the [R FAQ](#) for general information about R and the [R Windows FAQ](#) for Windows-specific information.

**Other builds**

[그림 1.5] 최신 버전의 R 다운 받기

(7) 설치 완료 및 바탕화면 아이콘

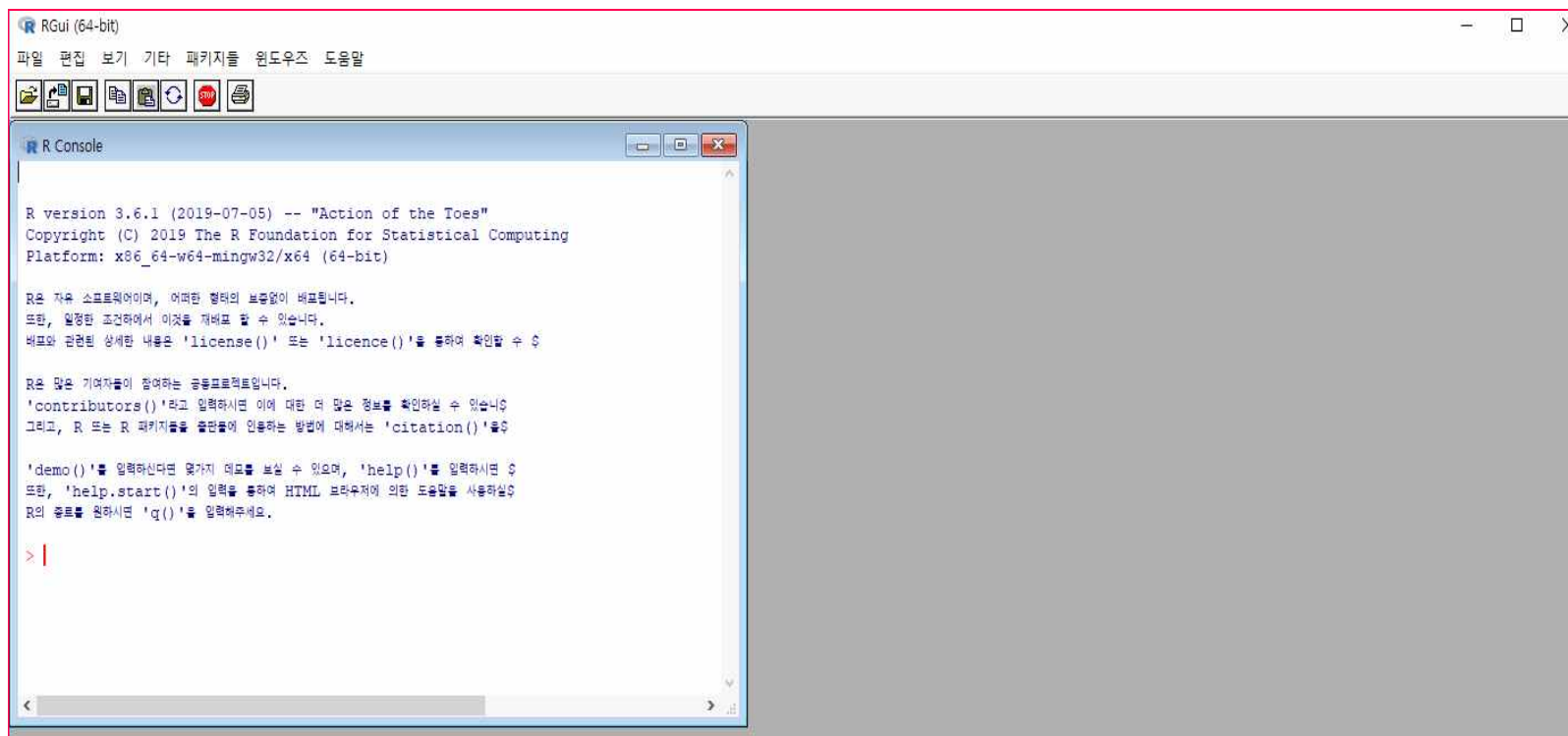


[그림 1.6] R 프로그램 설치 및 종료 및 아이콘



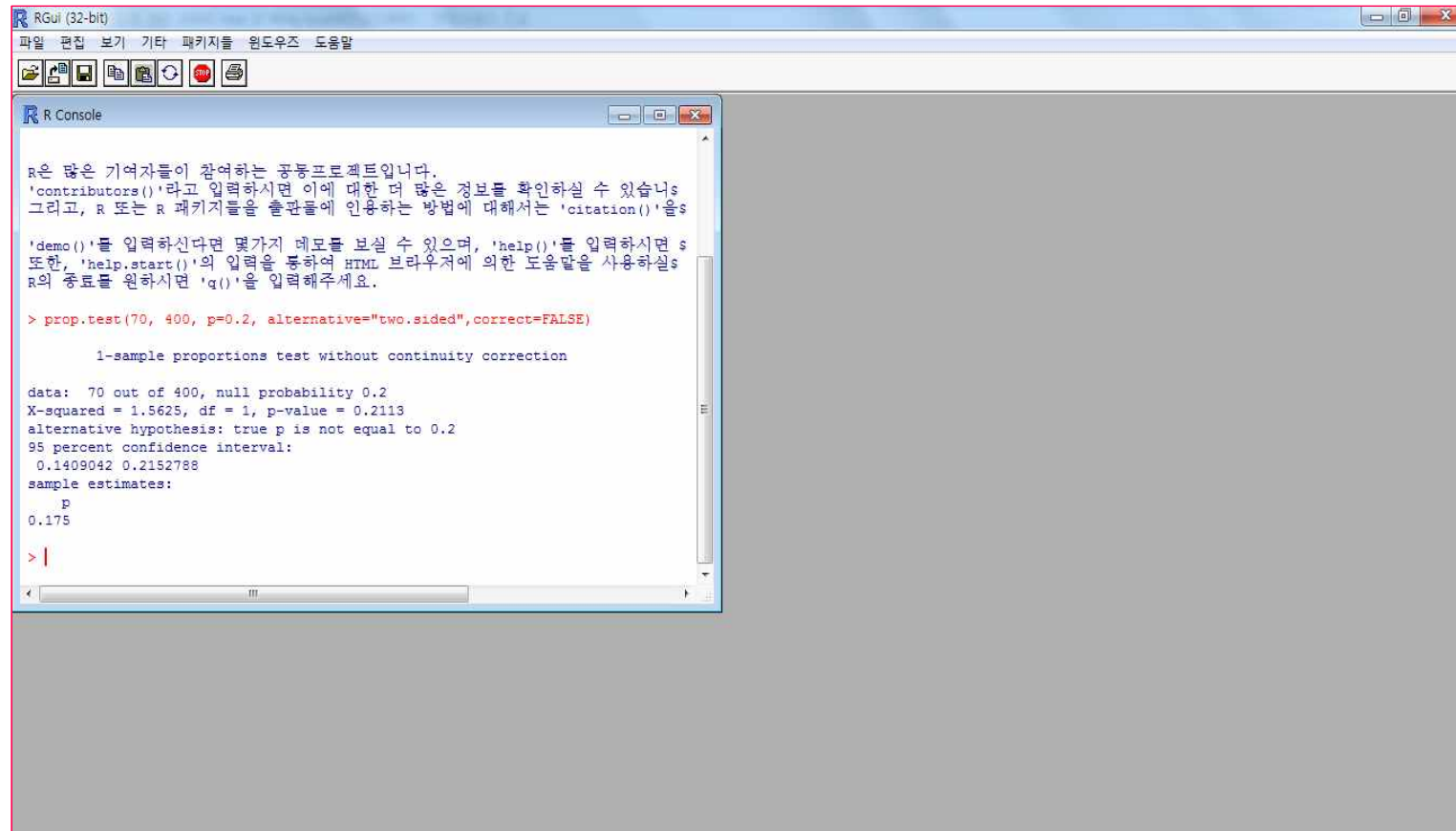
## 2. R의 기본 사용법

(1) R을 실행하면 다음과 같이 [R Console] 창이 기본적으로 열린다.



[그림 2.1] R Console 창

(2) [R Console] 창에 명령문을 입력하고 [Enter]를 누르면 다음과 같이 명령문이 실행되어 그 결과가 출력된다.



```
RGui (32-bit)
파일 편집 보기 기타 패키지들 윈도우즈 도움말

R Console

R은 많은 기여자들이 참여하는 공동프로젝트입니다.
'contributors()'라고 입력하시면 이에 대한 더 많은 정보를 확인하실 수 있습니다.
그리고, R 또는 R 패키지들을 출판물에 인용하는 방법에 대해서는 'citation()'을
'demo()'를 입력하신다면 몇가지 데모를 보실 수 있으며, 'help()'를 입력하시면
또한, 'help.start()'의 입력을 통하여 HTML 브라우저에 의한 도움말을 이용하실
R의 종료를 원하시면 'q()'을 입력해주세요.

> prop.test(70, 400, p=0.2, alternative="two.sided", correct=FALSE)

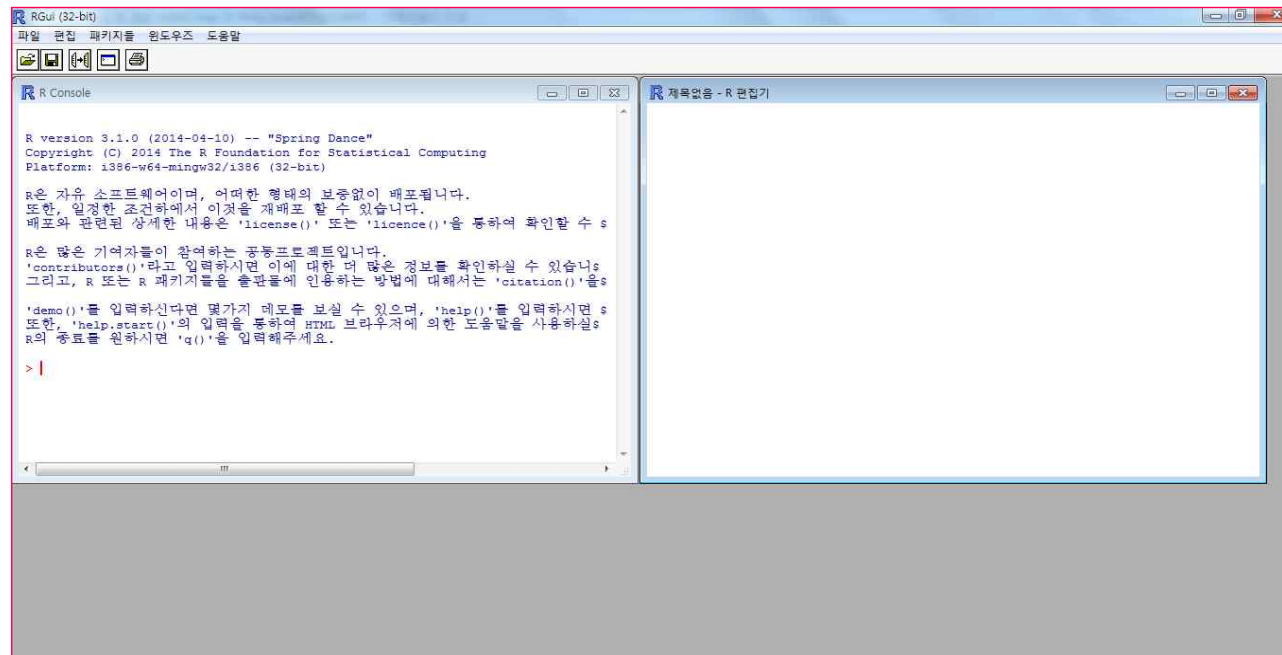
1-sample proportions test without continuity correction

data: 70 out of 400, null probability 0.2
X-squared = 1.5625, df = 1, p-value = 0.2113
alternative hypothesis: true p is not equal to 0.2
95 percent confidence interval:
 0.1409042 0.2152788
sample estimates:
p
0.175

> |
```

[그림 2.2] Console 창에 명령문 입력 및 실행

(3) 명령문을 입력하는 또 다른 방법은 [파일]-[새 스크립트]를 클릭하면 열리는 [R 편집기]를 활용하는 것이다. 다음의 화면에서 오른쪽이 [R 편집기] 창이다.



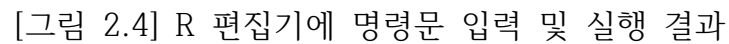
[그림 2.3] R 편집기

[1단계] [R 편집기]에 명령문을 입력

[2단계] 실행하고자 하는 명령문을 드래그 선택

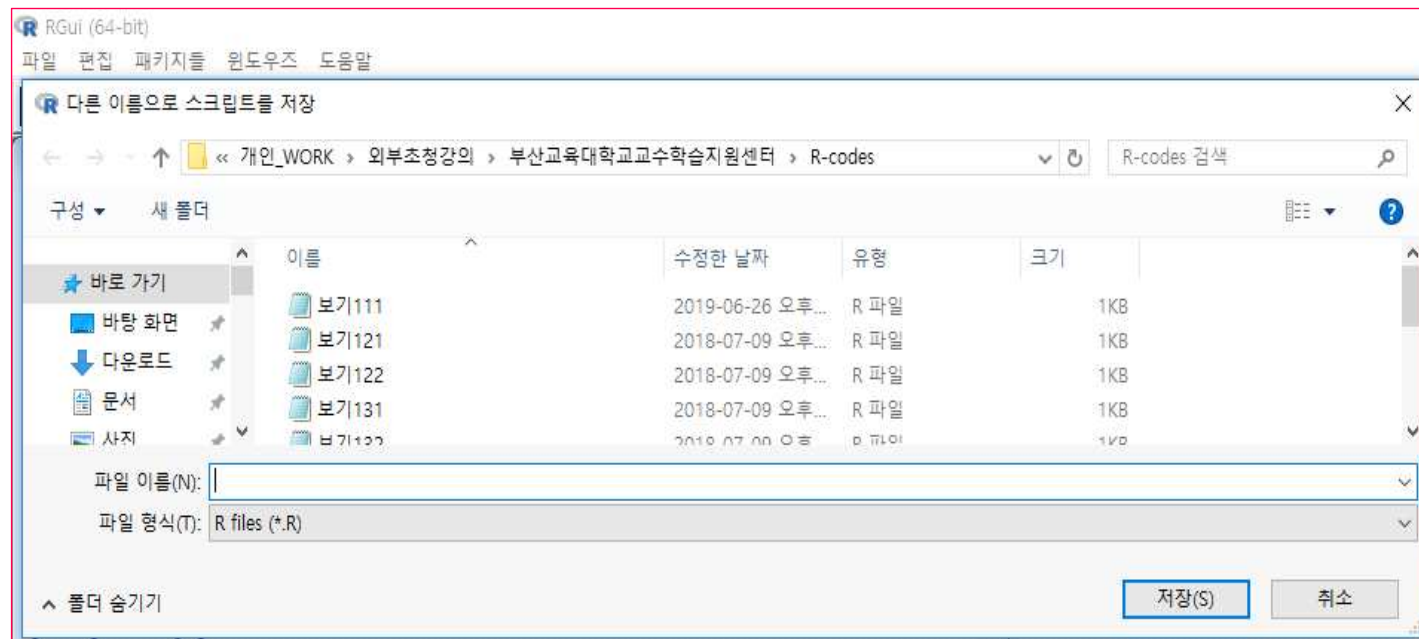
[3단계] 도구모음의 실행 버튼()을 클릭하거나, [F5] 또는 [Ctrl+R]을 누른다.

[4단계] 선택된 명령문의 결과가 [R Console] 창 또는 [R Graphics] 창에 출력



(4) [R 편집기]에 입력된 명령문은 수정 및 편집이 쉽게 가능하고 텍스트 파일 로 저장할 수 있어 [R Console] 창보다 편리하다.

[R 편집기]에 입력된 명령문을 다음에 다시 사용하기 위해서는 [파일]-[다른 이름으로 저장]에서 저장하면 된다. 다음과 같이 저장 위치를 지정한 다음, 파일 이름을 [파일명.txt]로 입력하고 파일 유형은 [All files (\*.\*)]를 선택한다.

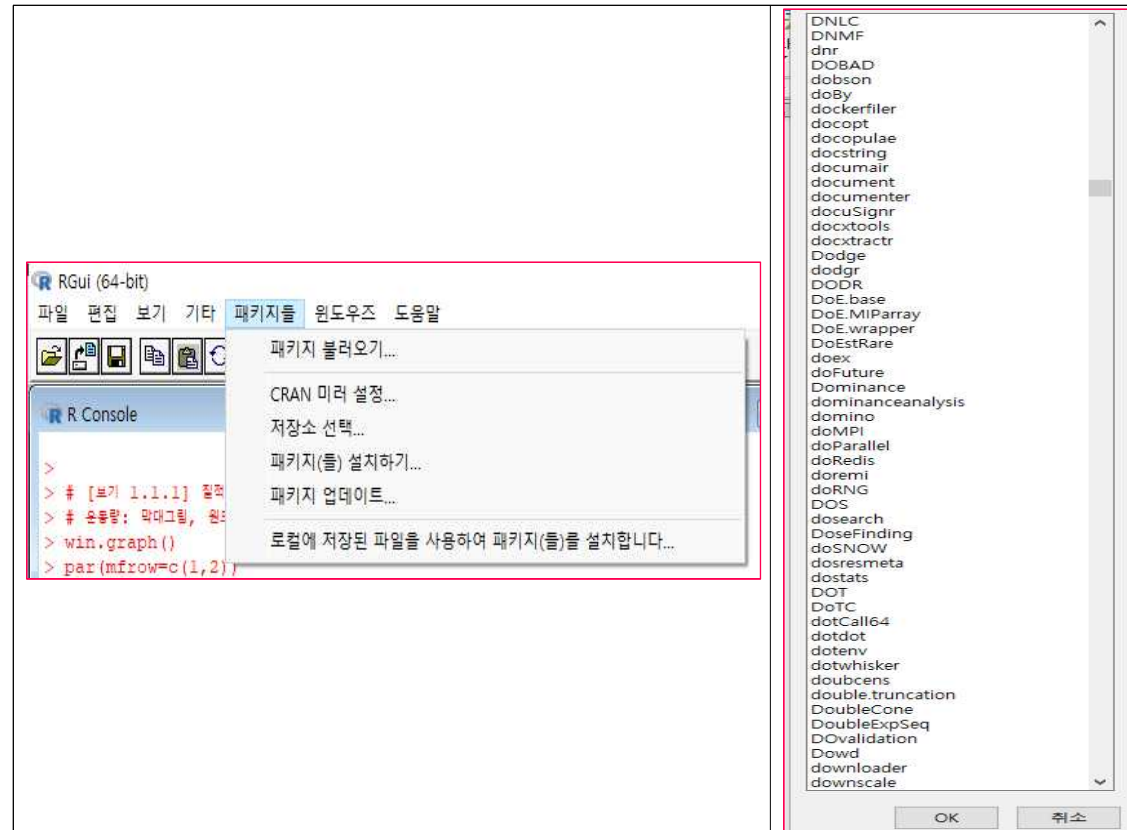


[그림 2.6] 편집기(스크립트)의 명령문 R-코드로 저장

이렇게 저장된 파일은 메모장에서 바로 열어볼 수도 있으며, [파일]-[스크립트 열기]에서 다음에 다시 열어 사용할 수 있다.

(5) 추가 패키지 설치 : base 포함되지 않은 함수를 가진 라이브러리

기초통계량 summaryBy()는 라이브러리 [doBy]에 포함되어 있어 [패키지들] 메뉴에서 [패키지(들) 설치하기]-[doBy]-[OK] 설치한다. 한번 설치로 지속적 사용 가능하다.



[그림 2.7] 추가 패키지 설치 및 라이브러리 지정

### 3. R 프로그래밍

#### What is R?

- High-level computer language and environment for statistics and graphics.
  - Performs a variety of simple and advanced statistical methods.
  - Produces high quality graphics.
  - Makes new functions based on the algorithms for R's users.
- Free open source software maintained by several contributor, including an "R Core Team" of around 20 programmers responsible for modifying the R source codes.
- R home page: <http://www.r-project.org>
- R system can be installed on Windows, Mac or Linux.





## (1) 기초 : R 명령문의 입력 및 실행 방법

- R은 대소문자를 구분하여 인식함
- 한 줄(line)에 둘 이상의 명령문 입력 시에는 세미콜론(;)으로 구분
- 주석(comment statement) 처리 시에는 샵 기호(#)를 사용

### R 실습 : 자료 입력 및 불러오기

```
## 자료 입력
x<-matrix(c(11,22,33,43,55,66,77,88,99,100), 10, 1, byrow=T)
colnames(x)<-c("수")
x
summary(x)

## txt 파일로 내보내기
write.table(x, "d:/sample.txt")

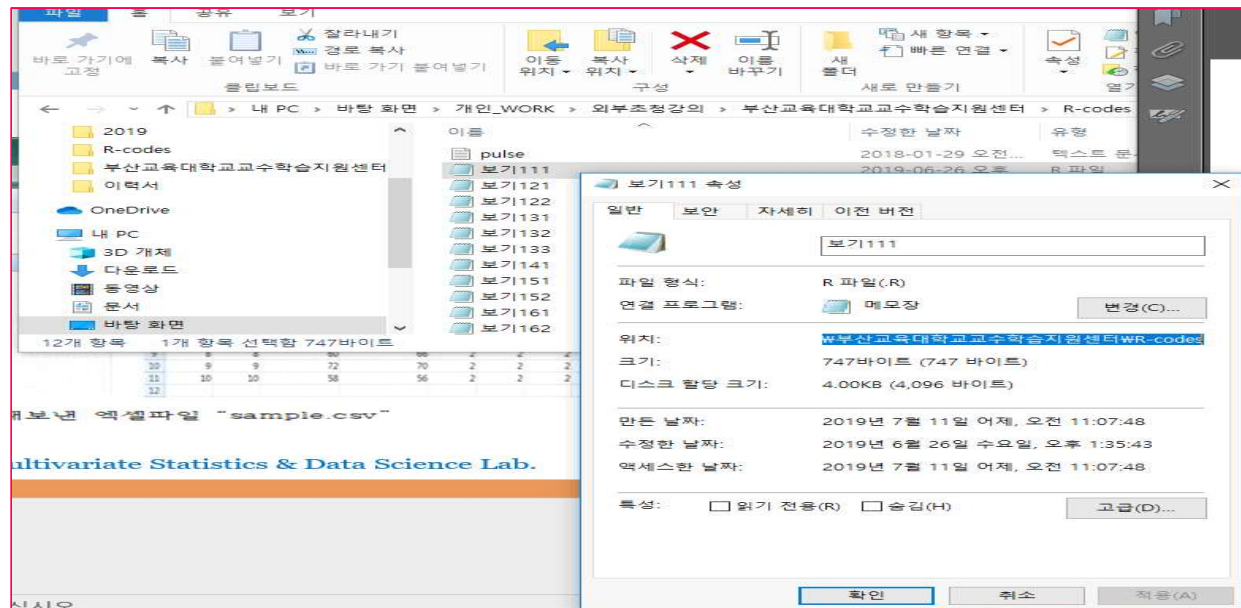
## txt 파일 불러오기
z=read.table("d:/sample.txt", header=T)
z

## Excel 파일로 내보내기
write.csv(x, "d:/sample.csv")

## Excel 파일 불러오기
y<-read.csv("d:/sample.csv", header=T)
y
summary(y)
```

## (2) 작업 폴더(working directory) 설정하기

- 작업 폴더 : 프로그램 및 외부 데이터의 저장 위치
- 작업 폴더 설정 : `setwd("directory")` 함수를 이용
- 작업 폴더 경로(directory) 지정 시 R 코드와 자료가 있는 폴더에서 [마우스(우측키)-(속성(R))]을 활용 [위치:]의 해당 경로를 복사해서 사용.
- **유의 사항** : 기본적인 컴퓨팅 경로는 역슬러시(\)를 사용하나 R에서는 슬러시(/)를 사용



[그림 3.1] 파일 속성의 경로 복사

### (3) R에서 자료 입력(직접입력 방식)

[프로그램 1]은 R에서 데이터를 직접 입력하기 위한 R-프로그램 이다. data를 직접입력하기 위해서는 matrix(data,nrow,ncol)를 이용하면 된다. byrow=T는 행순서로 데이터를 입력하는 옵션이다. 변수명을 입력하기 위해서는 colnames()를 이용하여 직접 입력가능하다.

**[프로그램 1]** [표1.1] R에서 데이터 직접 입력하기

```
T1 <- matrix(c(1,64,88,1,2,1,168,63.5,2,2,58,70,1,
2,1,183,65.8,2,3,62,76,1,1,1,185,72.6,3,4,66,78,1,1,
1,185,86,2,1,5,64,80,1,2,1,175,70.3,2,6,64,60,2,2,2,
168,81.6,3,7,94,92,2,1,2,157,82.1,2,8,60,66,2,2,2,157,
54.4,2,9,72,70,2,2,2,173,78.5,2,10,58,56,2,2,2,170,56.7,2),10,9,byrow=T)
colnames(T1)<-c("학생","처음심장박동수","나중심장박동수","달리기","흡연","성별","키","몸무게","운동량")
```

**[결과 1]** [프로그램 1]의 수행 결과

	학생	처음심장박동수	나중심장박동수	달리기	흡연	성별	키	몸무게	운동량
1	1	64	88	1	2	1	168	63.5	2
2	2	58	70	1	2	1	183	65.8	2
3	3	62	76	1	1	1	185	72.6	3
4	4	66	78	1	1	1	185	86.2	1
5	5	64	80	1	2	1	175	70.3	2
6	6	64	60	2	2	2	168	81.6	3
7	7	94	92	2	1	2	157	82.1	2
8	8	60	66	2	2	2	157	54.4	2
9	9	72	70	2	2	2	173	78.5	2
10	10	58	56	2	2	2	170	56.7	2

[그림 3.2] 자료 입력 및 결과

#### (4) R에서 자료 불러오기(텍스트파일)

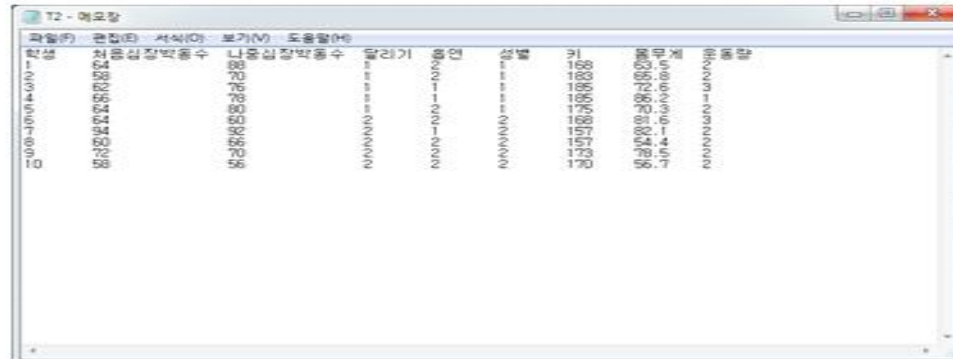
[프로그램 2]는 텍스트 파일을 불러오기 위한 R-프로그램이다.

텍스트 파일을 불러오기 위해서는 read.table()를 이용하면 되고, ()에 파일의 저장 경로를 입력해주면 된다. header=T는 변수명이 입력된 자료를 불러오기 위한 옵션이다.

**[프로그램 2]** [표1.1]데이터 텍스트 파일 불러오기

```
T2 <- read.table("E:/T2.txt",header=T)
```

[그림 1.11]은 텍스트 파일에 입력된 [표1.1] 자료이고, "T2.txt"로 저장되어 있다. 텍스트 파일의 데이터 구성을 보면, 첫행에 변수명이 입력되어 있다.



학생	최종심장박동수	나중심장박동수	알러기	종전	심박	키	몸무게	혈압	동맥
1	64	88		1	168	63	65	110	
2	56	70		1	183	83	72	110	
3	63	75		1	136	55	70	110	
4	56	78		1	105	46	65	110	
5	64	80		1	175	75	70	110	
6	64	80		1	166	68	70	110	
7	54	52		1	157	58	61	110	
8	50	56		1	157	54	64	110	
9	72	70		1	173	78	75	110	
10	56	56		1	170	56	61	110	

[그림 1.11] 변수명이 입력된 텍스트 파일

[그림 3.3] 텍스트 자료 파일 불러오기

## (5) R에서 자료 불러오기(엑셀파일)

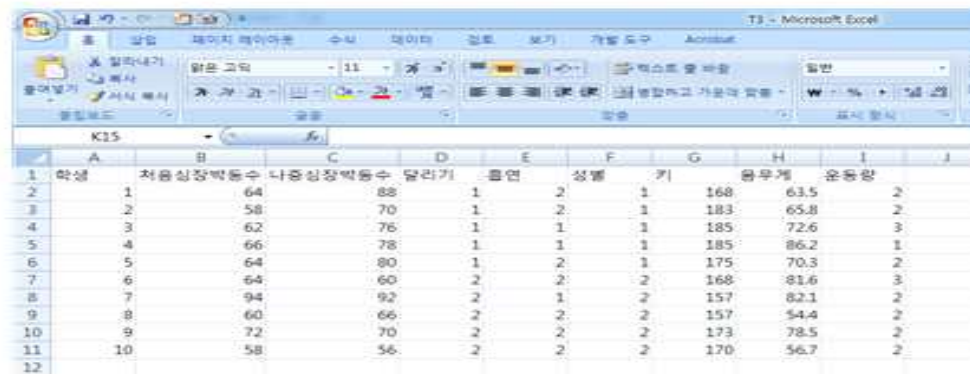
[프로그램 3]은 엑셀 파일을 불러오기 위한 R-프로그램이다.

엑셀 파일을 불러오기 위해서는 read.csv()를 이용하면 되고, ()에 파일의 저장 경로를 입력해 주면 된다. header=T는 변수명이 입력된 자료를 불러오기 위한 옵션이다.

**[프로그램 3]** [표1.1]데이터 엑셀 파일 불러오기

```
T3 <- read.csv("E:/T3.csv",header=T)
```

[그림 1.12]는 엑셀에 입력된 [표1.1] 자료이고, "T3.csv"로 저장되어있다. 엑셀 파일의 데이터 구성을 보면, 첫행에 변수명이 입력되어 있다.



	A	B	C	D	E	F	G	H	I	J
	학생	처음심장박동수	나중심장박동수	달리기	흡연	성별	키	몸무게	운동량	
1	1	64	88	1	2	1	168	63.5	2	
2	2	58	70	1	2	1	183	65.8	2	
3	3	62	76	1	1	1	185	72.6	3	
4	4	66	78	1	1	1	185	86.2	1	
5	5	64	80	1	2	1	175	70.3	2	
6	6	64	60	2	2	2	168	81.6	3	
7	7	94	92	2	1	2	157	82.1	2	
8	8	60	66	2	2	2	157	54.4	2	
9	9	72	70	2	2	2	173	78.5	2	
10	10	58	56	2	2	2	170	56.7	2	

[그림 1.12] 변수명이 입력된 엑셀 파일

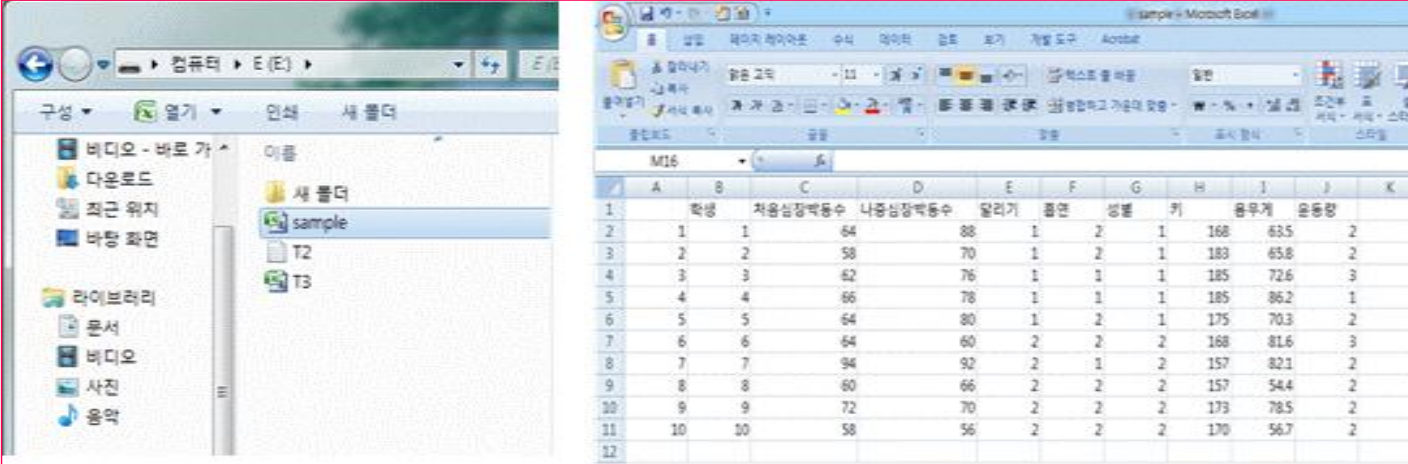
[그림 3.4] 엑셀 자료 파일 불러오기

## (6) R에서 자료 내보내기(엑셀파일)

[프로그램 4]는 R-프로그램을 이용하여 엑셀파일로 데이터를 내보내기 위한 프로그램이다. 먼저 "T3.csv"파일을 R에서 불러들여서 [결과 4]에 제시하였다. 그리고 write.csv를 이용하여 데이터를 엑셀파일로 저장하였다. 함수형식은 write.csv(내보내고자하는 데이터명,"저장하고자하는 경로 및 지정할 파일이름.csv")이다. 아래의 예제는 T3이라는 데이터를 E드라이브에 sample.csv라는 파일명으로 저장한것이다. 주의할점은 경로구분을 역슬래시가 아닌 슬래시(/)로 입력하여야 한다.

**[프로그램 4]** [표1.1]데이터 엑셀 파일로 내보내기

```
T3 <- read.csv("E:/T3.csv",header=T)
T3
write.csv(T3,"E:/sample.csv")
```



	A	B	C	D	E	F	G	H	I	J	K
1		학생	처음심장박동수	나중심장박동수	달리기	흡연	성별	키	몸무게	운동량	
2	1	1	64	88	1	2	1	168	63.5	2	
3	2	2	58	70	1	2	1	183	65.8	2	
4	3	3	62	76	1	1	1	185	72.6	3	
5	4	4	66	78	1	1	1	185	86.2	1	
6	5	5	64	80	1	2	1	175	70.3	2	
7	6	6	64	60	2	2	2	168	81.6	3	
8	7	7	94	92	2	1	2	157	82.1	2	
9	8	8	60	66	2	2	2	157	54.4	2	
10	9	9	72	70	2	2	2	173	78.5	2	
11	10	10	58	56	2	2	2	170	56.7	2	
12											

[그림 3.4] 엑셀 자료 파일로 내보내기

Multivariate Statistics :  
PCA, CA, DSCA, Biplots,  
MDS, etc.

Big Data :  
unstructured&structured  
data, data mining,  
visualization techniques.



**YONG-SEOK CHOI**  
Department of Statistics  
Pusan National University

Shape Analysis :  
Procrustese analysis, GPA

Categorical Data Analysis :  
SCA, MCA, logistic&loglinear  
models

Statistical Consulting,  
Statistical Survey,  
Comparative Study