

Introduction to R

❖ R 프로그램의 소개

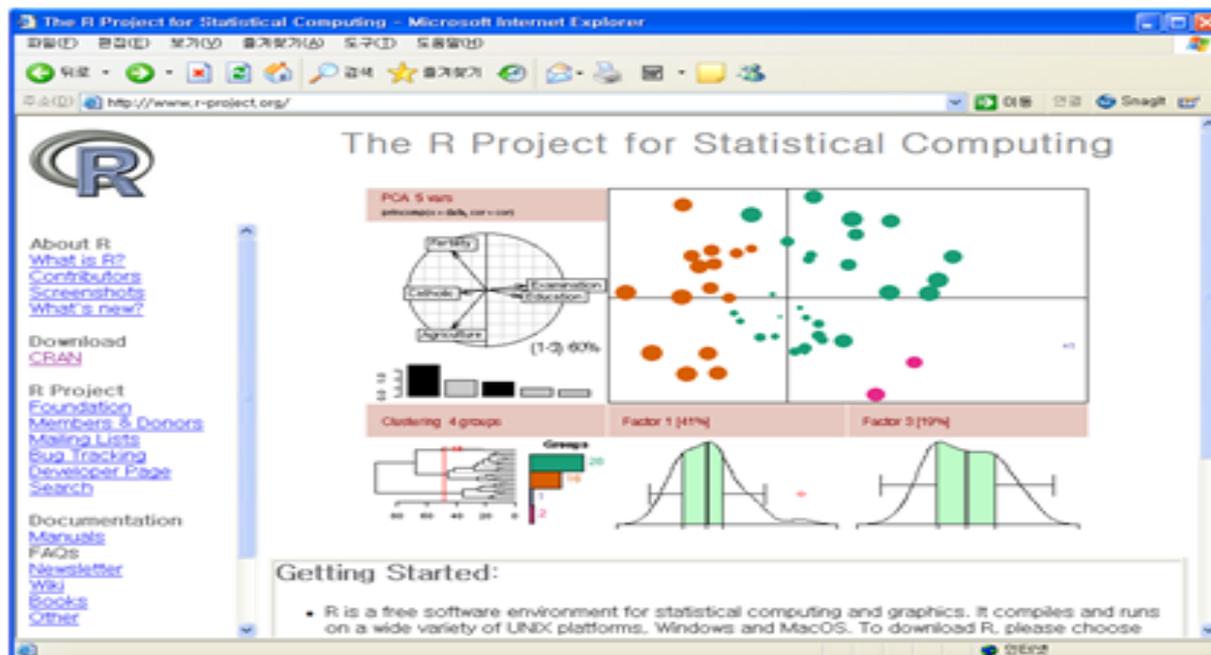
- 통계 패키지로써의 R
 - 통계 데이터 수정 및 분석에 사용되는 스크립트 언어
 - 오픈 소스 소프트웨어(open source software)
 - Windows / Mac / Linux 등의 여러 운영체제에서 사용 가능
 - 함수형 프로그래밍 언어와 객체지향 언어의 결합형
 - 그래픽 측면이 상용 프로그램들보다 월등히 우수함
 - 무료 다운로드 및 설치 가능 (www.r-project.org)

Introduction to R

❖ R 프로그램의 설치

R 프로그램 다운받기

R 홈페이지(<http://www.r-project.org>)에 접속하여 [그림 1-1]과 같이 왼쪽 부분의 메뉴의 다운로드에서 R을 다운로드 받을 수 있는 네트워크를 의미하는 CRAN을 클릭한다.



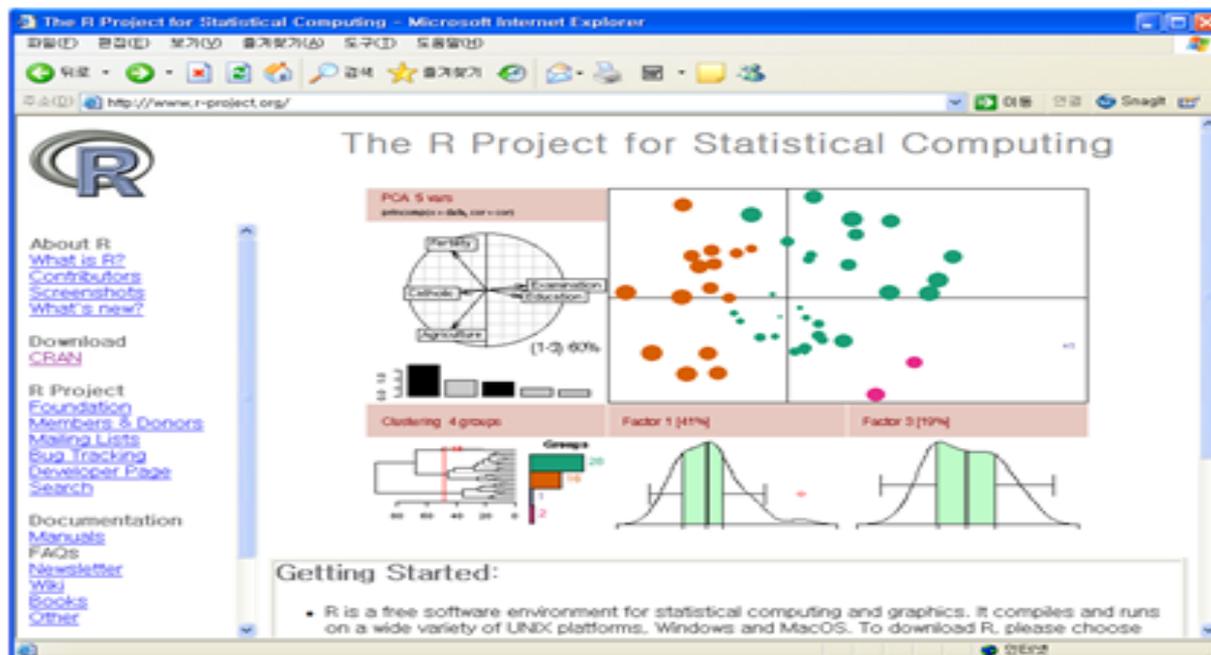
[그림 1-1] R 프로그램 홈페이지

Introduction to R

❖ R 프로그램의 설치

R 프로그램 다운받기

R 홈페이지(<http://www.r-project.org>)에 접속하여 [그림 1-1]과 같이 왼쪽 부분의 메뉴의 다운로드에서 R을 다운로드 받을 수 있는 네트워크를 의미하는 CRAN을 클릭한다.

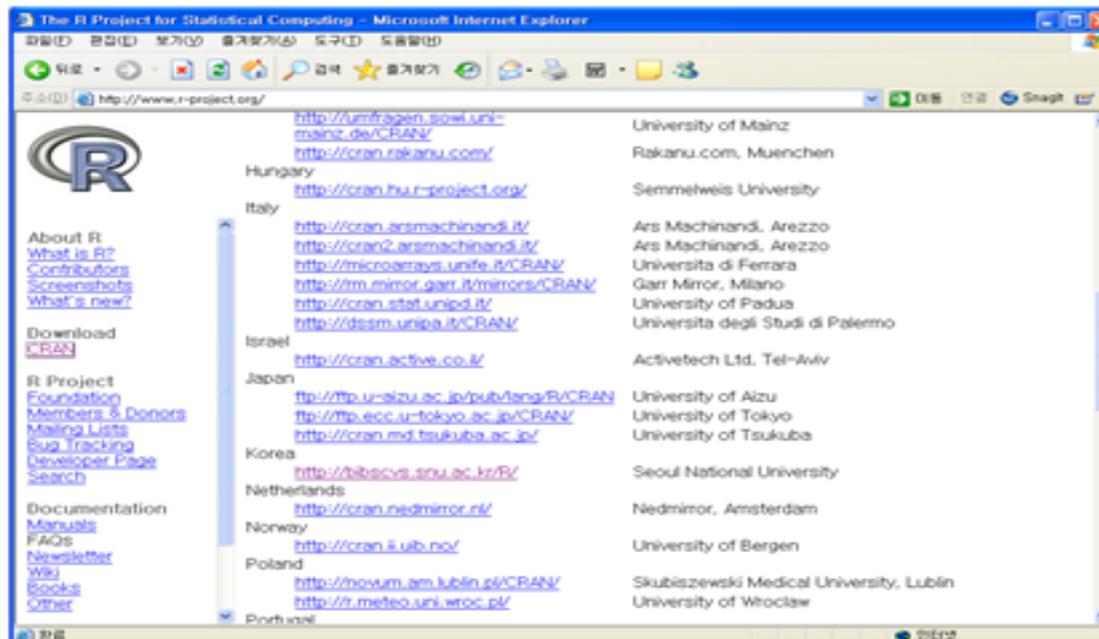


[그림 1-1] R 프로그램 홈페이지

Introduction to R

❖ R 프로그램의 설치

[그림 1-2]와 같이 R은 다른 나라의 많은 홈페이지를 통해 내려받을 수 있다. 여기서 접속이 빠른 홈페이지 주소를 선택하여 내려받으면 된다. 여기서는 한국 홈페이지인 <http://bibscvs.snu.ac.kr/R/> 클릭한다.

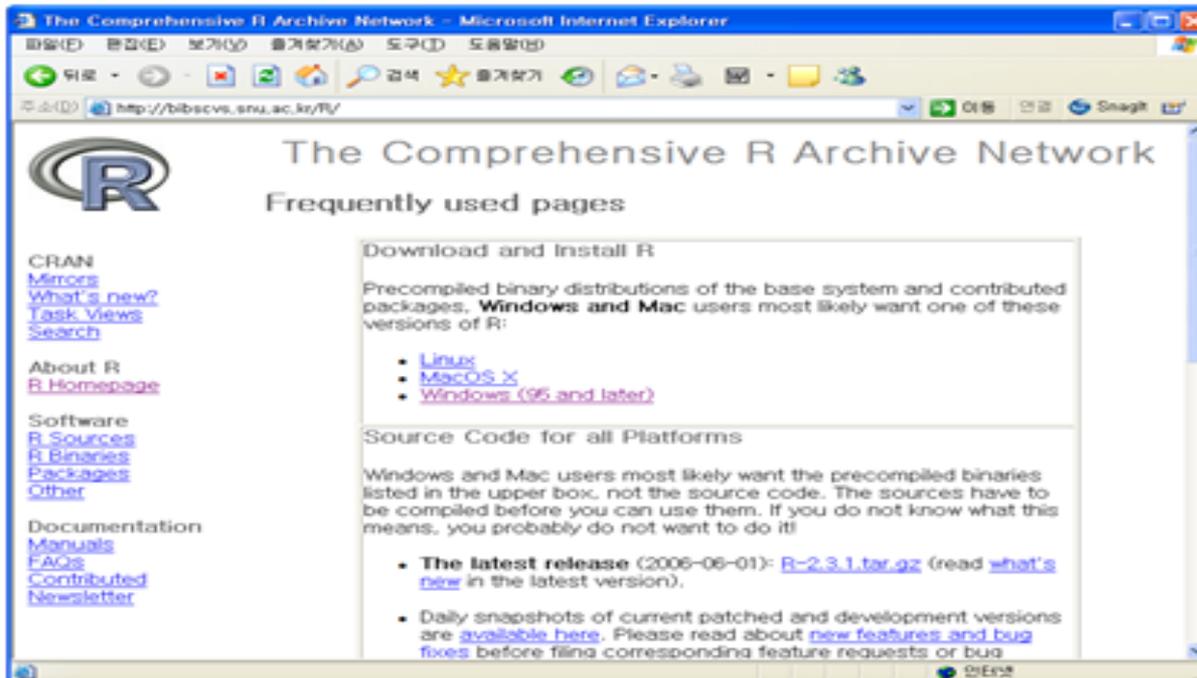


[그림 1-2] 프로그램을 받을 수 있는 여러나라 사이트

Introduction to R

❖ R 프로그램의 설치

컴퓨터 운영체제별로 R 프로그램을 다운로드 받을 수 있는 화면이 [그림 1-3]과 같이 나타난다. 윈도우용 R 프로그램을 다운로드 받기 위해 Windows(95 and later)를 클릭한다.



[그림 1-3] R 프로그램을 받을 수 있는 화면

Introduction to R

❖ R 프로그램의 설치

설치가 완료되면 다음 [그림 1-5]과 같은 화면이 나타난다. 여기에서 완료를 클릭하면 설치가 끝난 것으로 바탕화면에 R을 수행할 수 있는 아이콘이 생성된다. 앞으로는 바탕화면의 아이콘을 선택하여 R을 수행할 수 있다. 이로써 R의 설치가 완료된다.



[그림 1-5] R 프로그램의 설치 종료

Introduction to R

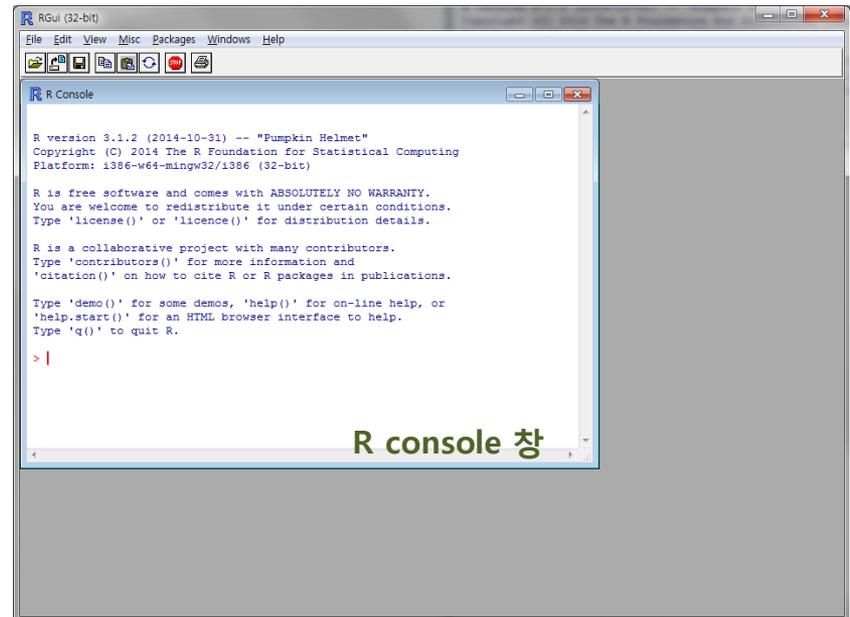
❖ R 프로그램의 소개

■ R 프로그램의 인터페이스

- 기본적으로 R console 창만 제공
- 프로그램 수행 과정의 로그와 수행 결과를 제공
- 명령어는 R console 창에 직접 입력가능
- R script 창을 이용하는 것이 프로그램 작성 및 수정에 용이함
- R script 창은 별도 저장 가능
- R script 창 생성 방법 :

『파일(File)』

- 『새 스크립트(New script)』



Introduction to R

❖ R 프로그램의 실행

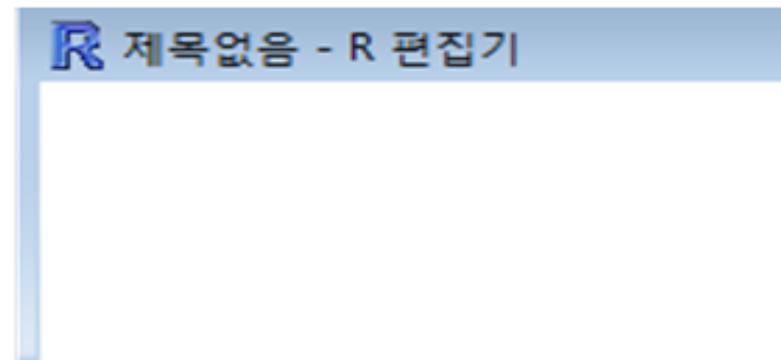
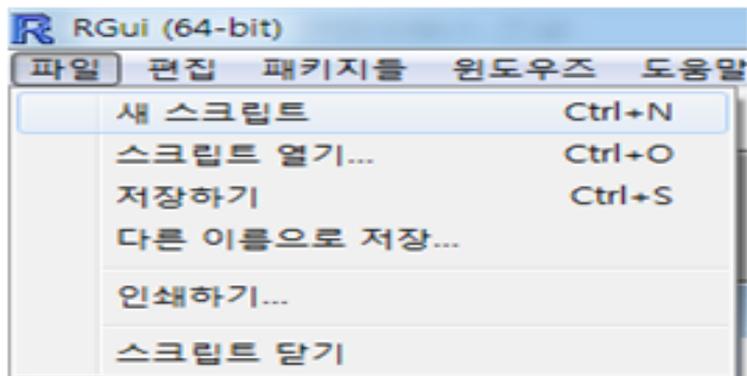
명령문 입력과 실행

콘솔에서는 라인 타입 명령문 입력과 실행으로 작동하여 [Enter]키를 입력하지 않는 한 라인에 작업한 내용은 실행되지 않는다. 그러므로 상이한 명령문을 연속하여 동일 라인에 입력하고자 하면 명령문 사이에 세미콜론(:)을 사용하면 된다.

라인 에디터 작업환경은 다소 불편하므로 스크립트를 활용하는 것이 좋다.

[파일]-[새스크립트]를 이용하면 새로운 스크립트창을 만들 수 있다.

스크립트 작업이 끝나면 [파일]-[다른이름으로 저장] 을 이용하여 스크립트를 저장하고 다시 불러 사용하면 된다



[그림 1.8] R에서의 입력

Introduction to R

❖ R 프로그램의 실행

스크립트에 작성된 프로그램은 라인별로 실행할 수도 있으며 전체 프로그램을 한번에 실행할 수도 있다. 실행된 명령문은 콘솔에 명령문과 함께 결과가 출력된다.

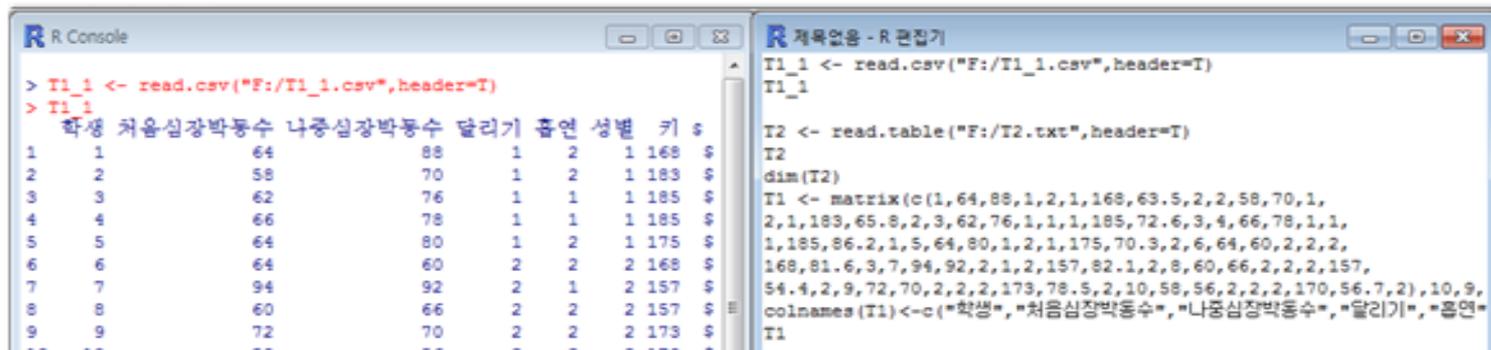
1. 라인별 실행

커서를 명령문 라인 아무곳에나 놓고 [F5]키를 누르면 라인 명령문이 실행되어 콘솔에 명령문과 결과가 출력되고 커서는 다음 라인 명령문으로 이동한다.

2. 전체 스크립트 실행

커서를 아무곳에나 놓고 [ctrl+A]키를 누르면 스크립트 명령문 전체가 선택된다. 스크립트 전체를 선택한 뒤 [F5] 키를 누르면 스크립트 전체가 실행되어 콘솔에 나타난다.

스크립트에서 명령문을 실행하지 않으려면 명령문 제일 앞에 '#' 표시를 하여 실행되지 않도록 한다.



The screenshot shows two windows from an R environment. The left window, titled 'R Console', displays the execution of R code. The code reads a CSV file 'T1_1.csv' with headers. The output shows a data frame with columns: 학생, 처음심장박동수, 나중심장박동수, 달리기, 흡연, 성별, 키, and \$. The data is presented in a table format with 9 rows and 7 columns of data. The right window, titled 'R 재목언용 - R 편집기', shows the source code for the script. The code includes: `T1_1 <- read.csv("F:/T1_1.csv",header=T)`, `T1_1`, `T2 <- read.table("F:/T2.txt",header=T)`, `T2`, `dim(T2)`, and a large `matrix` definition for `T1` with 10 columns and 9 rows, followed by `colnames(T1) <- c("학생", "처음심장박동수", "나중심장박동수", "달리기", "흡연", "성별", "키")` and `T1`.

[그림 1.9] R스크립트 실행

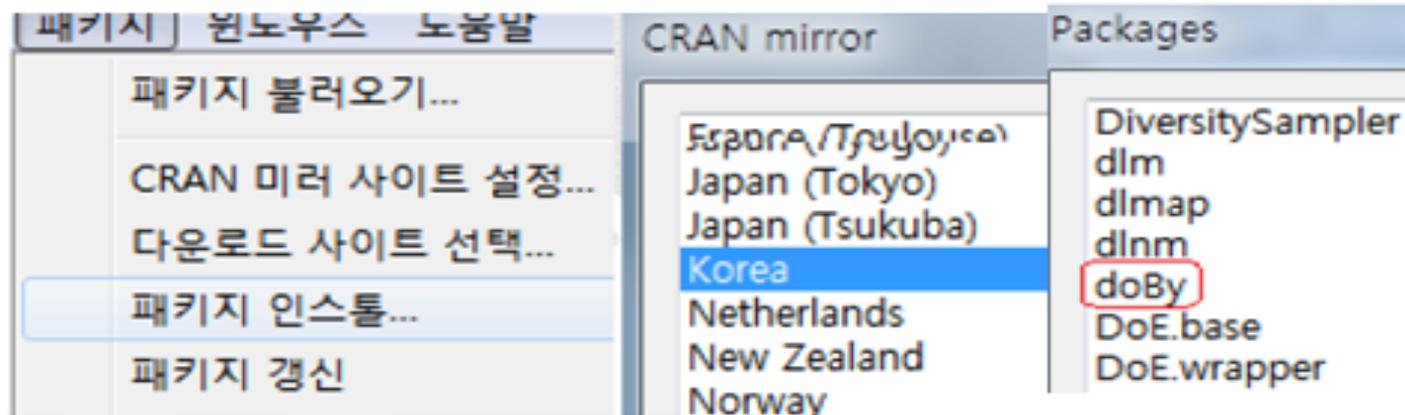
Introduction to R

❖ R 프로그램의 실행

패키지 설치

“base”에 포함되지 않는 함수가 포함된 라이브러리는 “패키지”메뉴에서 가능하다. 기초통계량을 구하는 함수 `summaryBy()`는 “base”에 포함되어 있지 않고 “doBy” 라이브러리에 있다. [그림 1.10]은 이를 설명하고 있다.

설치가 완료되면 콘솔 화면에 설치완료문구가 출력된다. 패키지는 한 번만 설치하면 지속적으로 사용할 수 있다.



[그림 1.10] R 패키지 설치

Introduction to R

❖ 기초 R 프로그래밍

- 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 script에 기록된 명령문 실행 방법 :
 - 『편집(Edit)』 - 『줄 또는 선택영역 실행(Run line or selection)』
 -  아이콘 클릭
 - 『Ctrl』 + 『R』

Introduction to R

❖ 기초 R 프로그래밍

- 작업 폴더(working directory) 설정하기
 - 작업 폴더 : 프로그램 및 외부 데이터의 저장 위치
 - 작업 폴더 설정 : `setwd()` 함수를 이용

```
setwd("directory ")
```

- 작업 폴더 경로(directory) 지정 시 유의 사항 :
기본적인 컴퓨팅 경로는 역슬러시(\)를 사용하지만
R에서는 슬러시(/)를 사용

Introduction to R

❖ 기초 R 프로그래밍

- 외부 데이터 불러오기

- 구분자(separator)가 빈칸(space)이나 탭(tab)으로 구성된 정형화 데이터

```
read.table("file name ",header = TRUE )
```

- 구분자(separator)가 콤마(,)나 세미콜론(;)으로 구성된 정형화 데이터

```
read.csv("file name ",header = TRUE )
```

- **header** 옵션에는 파일의 첫 번째 줄에 변수명의 기재 여부에 따라 논리값(TRUE/FALSE) 지정
- **read.table()** 함수와 **read.csv()** 함수에 의한 결과는 행은 개체, 열은 변수를 나타내는 데이터프레임(data frame)의 형태를 가짐

Introduction to R

❖ R에서 데이터 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

Introduction to R

❖ R에서 데이터 불러오기(텍스트 파일 불러오기)

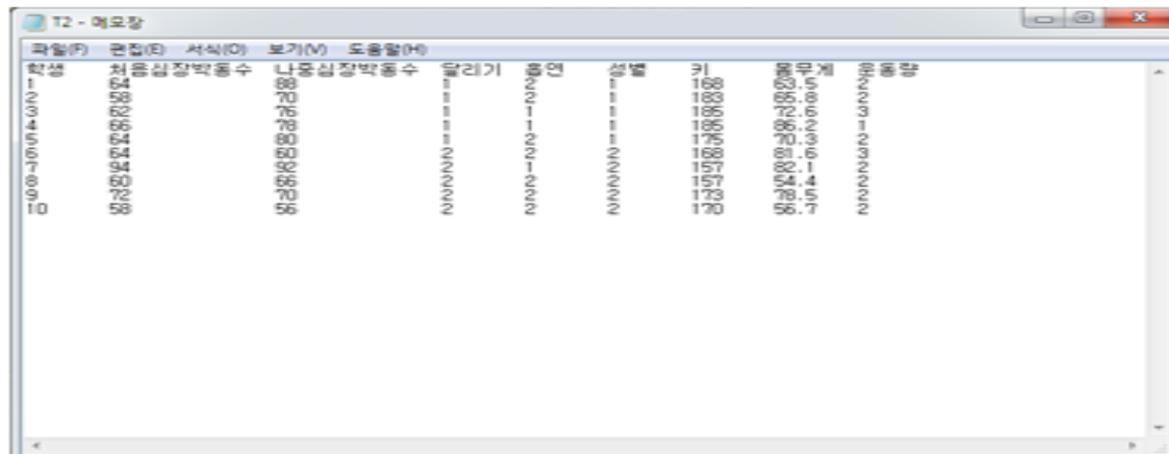
[프로그램 2]는 텍스트 파일을 불러오기 위한 R-프로그램이다.

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	처음심장박동수	나중심장박동수	알리기	흡연	성별	키	몸무게	운동량
64	88	88	N	N	N	168	63.5	N
58	70	70	N	N	N	168	62.0	N
62	76	76	N	N	N	165	62.0	N
66	78	78	N	N	N	175	70.3	N
64	80	80	N	N	N	168	61.6	N
94	92	92	N	N	N	157	54.1	N
60	66	66	N	N	N	157	54.4	N
72	56	56	N	N	N	173	78.5	N
10	58	58	N	N	N	170	56.7	N

[그림 1.11] 변수명이 입력된 텍스트 파일

Introduction to R

❖ R에서 데이터 불러오기(엑셀 파일 불러오기)

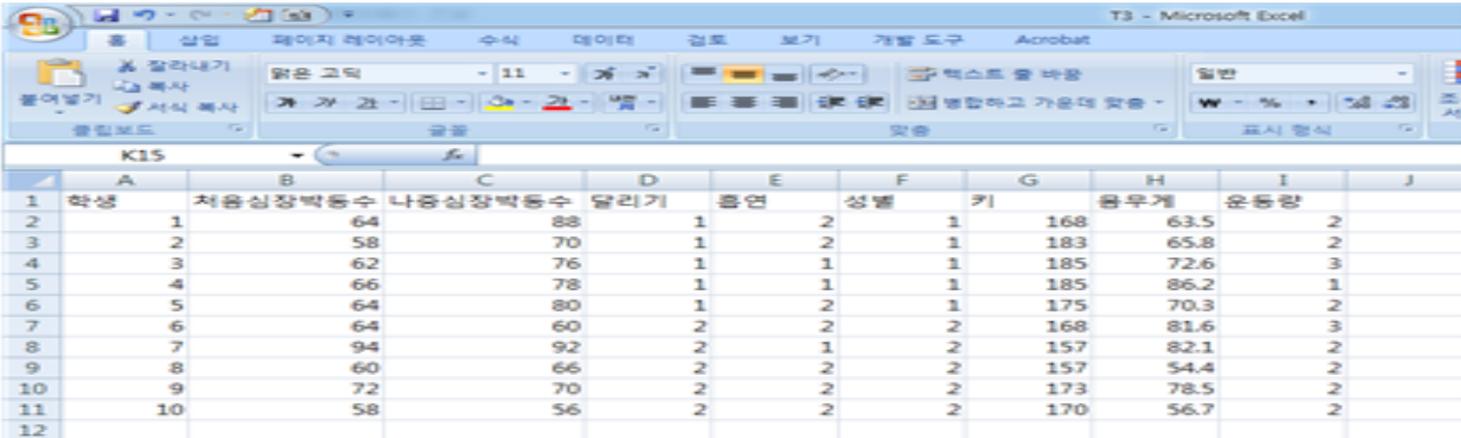
[프로그램 3]은 엑셀 파일을 불러오기 위한 R-프로그램이다.

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 저장되어 있다. 엑셀 파일의 데이터 구성을 보면, 첫행에 변수명이 입력되어 있다.



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

[그림 1.12] 변수명이 입력된 엑셀 파일

Introduction to R

❖ R에서 데이터 내보내기(엑셀 파일로 내보내기)

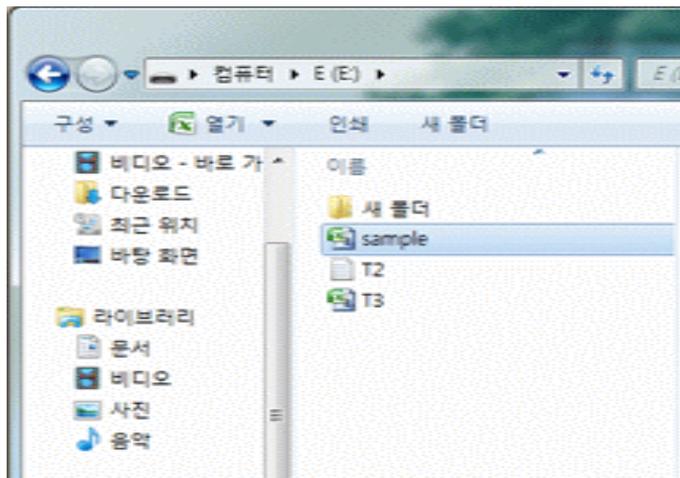
[프로그램 4]는 R-프로그램을 이용하여 엑셀파일로 데이터를 내보내기 위한 프로그램이다. 먼저 "T3.csv"파일을 R에서 불러들여서 [결과 4]에 제시하였다.

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 screenshot of a Microsoft Excel spreadsheet. The spreadsheet has 12 rows and 11 columns. The first row is a header row with the following text: '학생', '처음심장박동수', '나중심장박동수', '달리기', '휴면', '성별', '키', '몸무게', '운동량'. The data rows contain numerical values for each of these categories. The spreadsheet is titled 'sample - Microsoft Excel'.

[그림 1.13] R에서 내보낸 엑셀파일 "sample.csv"