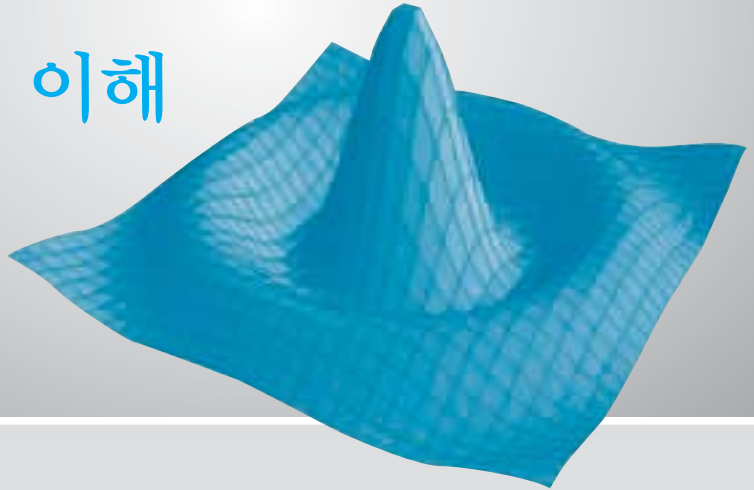


제1장

R 프로그램의 이해



제1절 R 둘러보기

제2절 R의 설치 및 실행

제3절 메뉴 표시줄의 기능

제 1 절

R 둘러보기



사회가 점점 더 복잡하고 다양화되어짐에 따라 의사결정 과정에서의 리스크 및 불확실성(uncertainty) 또한 날로 증가되고 있으며, 우리가 당면하고 있는 경제문제들 역시 이러한 불확실성의 문제로부터 자유롭지 못하다. 그러므로 미래에 대한 예측이 불투명한 상황에서 보다 정확한 의사결정을 내리기 위해서는 다양한 **정보**(information)를 고려해야 할 필요가 있으며, 우리는 이 과정에서 필연적으로 방대한 양의 **자료**(data)를 수집하고 분석해야만 한다. 따라서 계량분석을 용이하게 수행하기 위한 컴퓨터의 사용은 이제 필수적인 요소로 자리매김하고 있으며, 오늘날 다양한 정보의 요구는 이러한 필요성을 더욱 확고히 하고 있다.

다행히 컴퓨터 하드웨어의 지속적인 발전 덕분에 개인 컴퓨터(Personal Computer, PC)로도 경제·경영과 관련된 방대한 자료를 이용한 통계분석이 가능하게 되었으며, 컴퓨터를 이용한 계량분석 기법 역시 다양하고 심도 있게 발전하고 있다. 그리고 분석자가 더욱 쉽고 편리하게 컴퓨터를 사용하여 계량분석을 수행할 수 있도록 복잡하고 방대한 계량기법을 정리하고 프로그램화하여 사용하게 되었는데, 이것을 **통계 혹은 계량 패키지**(package)라고 한다. 경제 통계분석을 위한 대표적인 패키지로는 SPSS, SAS, STATA, RATS, Micro TSP, SHAZAM, EViews, GAUSS, MATLAB 등이 있다. 하지만 이런 프로그램들은 복잡한 계량분석을 쉽게 수행할 수 있다는 장점을 가짐에도 불구하고 고가의 구매비용 때문에 접근이 쉽지 않은 것이 현실이다.¹⁾ 따라서 저자는 **프로그래밍 언어**(programming language)²⁾로 구성된 통계 분석 도구(analysis tool)로, 다양하고 풍부한 분석기능을 갖춘 통계 패키지임에도 불구하고 사용자들에게 무료로 제공되고 있는 **R** 프로그램을 소개하고자 한다. 특히 R 프로그램은 사용이나 조작 및 배포(distribution)가 자유롭다는 이점으로 인해 전 세계 경제·경영 분야의 학자와 전문가들을 포함한 다양한 분야의 사용자들이 자발적으로 프로그램의 기능을 개선하는 데 공헌하고 있어 앞으로도 무한한 발전이 기대되는 프로그램이기도 하다. 이에 제1장에서는 R의 생성 배경과 장단점 등에

- 1) 최근에는 상경계 분야는 물론 다양한 분야의 학생이나 전문가들이 통계분석을 위해 비교적 접근이 쉬운 마이크로소프트(Microsoft)의 **엑셀**(Excel) 소프트웨어를 이용하고 있으며, 저자 역시 이러한 요구에 부응하여 엑셀의 분석 도구로는 분석이 어려운 계량경제, 금융 및 재무관리 분석을 사용자가 엑셀에서 쉽게 수행할 수 있도록 **ETEX**(Econometric Toolbox for Excel) 프로그램을 개발하였다. 그러나 엑셀 프로그램 역시 구매비용이 수반되며, ETEX 프로그램을 이용하여도 고급계량분석에는 한계를 지닌다. 참고로 박범조의 *경영경제통계학*, 4판(2012)을 통해 ETEX을 이용한 통계분석 방법을 보다 자세히 살펴볼 수 있다.
- 2) 프로그래밍 언어는 컴퓨터에 명령을 내리는 데 필요한 '컴퓨터의 언어'로 요약될 수 있다. 프로그래밍 언어는 사용 목적에 따라 수십 가지가 존재하는데, 대표적으로 **Basic, Cobol, Pascal, Fortran, C, C++** 등이 있다.

대해 알아보고, 프로그램의 설치 방법과 더불어 전반적인 기능들을 소개함으로써 독자들로 하여금 앞으로 이 책에서 수행하게 될 계량분석 과정에서 사용할 R에 대한 이해와 흥미를 불러일으키고자 한다.

컴퓨터로 통계 및 계량분석이 가능하도록 계산 과정을 정리해놓은 프로그램을 **통계 패키지**라고 한다. R은 **프로그래밍 언어**로 구성된 통계분석 도구로, 다양하고 풍부한 분석 기능을 갖춘 통계 패키지임에도 불구하고 대부분의 통계 패키지와는 달리 무료로 제공되고 있어 전 세계적으로 많은 분석가들이 사용하고 있다.

1. S 언어와 R

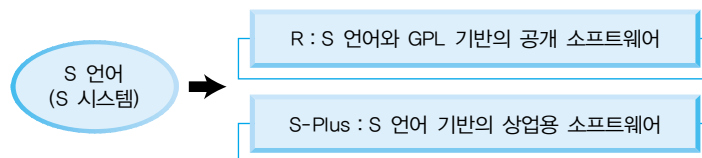


R의 기능과 장단점에 대한 본격적인 논의에 앞서 R의 역사와 생성 배경을 알아보기로 하자. 앞서 R은 프로그래밍 언어로 구성된 통계분석 도구라고 정의 내린 바 있다. 그러므로 R을 정확하게 이해하기 위해서는 우선 R 프로그래밍 언어의 기초를 이루는 **S 언어**에 대해 살펴볼 필요가 있다.

S 언어는 미국 최대의 전화 통신 회사인 AT&T의 벨연구소(Bell Laboratories)에서 존 챔버스(John Chambers), 릭 베커(Rick Becker), 앨런 윌크스(Allan Wilks)에 의해 개발된 프로그래밍 언어이다. 앞서 프로그래밍 언어는 사용 목적에 따라 수십 가지가 존재한다고 하였는데, S 언어는 특별히 통계 연산의 연구 목적으로 개발된 프로그래밍 언어이다. S 언어 개발에 참여했던 존 챔버스가 “S 언어의 목적은 아이디어를 빠르고 충실하게 소프트웨어로 전환하는 것이다.”³⁾라고 하였을 정도로 S 언어는 매우 유연한 자료 조작 및 변환 환경을 제공하며, 매우 뛰어난 그래픽 구현능력도 가지고 있다. 참고로 우리에게 너무나도 잘 알려져 있는 유닉스(UNIX) 운영체제와 프로그래밍 언어인 C, C++ 또한 벨연구소에서 개발된 것들이다.

S 언어(S 시스템)는 이후 **R**과 **S-Plus**로 분화되었는데, R과 S-Plus는 모두 S 언어를 기반으로 작성된 프로그램이라는 점에서 공통점을 갖는다. 따라서 두 프로그램은 상당 부분 상호 호환이 가능하며 R에서 사용하는 대부분의 기능들과 구문들이 S-Plus에도 적용 가능하다. 그러나 S-Plus는 상업용 소프트웨어로 그 저작권(license)이 현재 TIBCO Software 사에 있는 반면, R은 **GNU**(GNU' Not Unix) **프로젝트**의 **GPL**(General Public License)을 따르는 공개 소스 소프트웨어로 발전하였다는 데 차이가 있다.

그림 1-1
S 언어의 분화



3) 위키피디아(<http://www.wikipedia.org>)에서 인용.

GNU 프로젝트란 **무료 소프트웨어 재단**(Free Software Foundation, FSF)에서 유닉스 호환의 소프트웨어를 개발하여 무료로 배포하는 프로젝트를 의미하며, GPL은 GNU 프로젝트가 부여하는 일반 공중 사용 허가서이다. 따라서 GPL이 부여된 소프트웨어는 무료이며, 프로그램의 사용이나 수정(modification), 재배포에 제약이 없을 뿐만 아니라 수정이나 재배포된 프로그램은 다시 GPL을 부여받게 된다. 그렇기 때문에 GPL을 따르는 R을 **GNU R**이라고도 하며, R은 수많은 프로그래밍 기술자들과 수학자 및 통계학자, 그리고 다양한 분야의 전문가들에 의해 그 기능이 개선되고 무료로 재배포되면서 지속적으로 발전하고 있다.⁴⁾

S 언어는 특별히 통계 연산의 연구 목적으로 개발된 프로그래밍 언어이다. **R과 S-Plus**는 모두 S 언어를 기반으로 작성된 프로그램이라는 점에서 공통점을 가지나, S-Plus는 상업용 소프트웨어인 반면, R은 **GNU 프로젝트의 GPL**을 따르는 공개 소스 소프트웨어이다.

2. 통계 및 계량분석 프로그램으로서의 R



출처 : r-project.org

R은 오클랜드대학교(University of Auckland)의 로버트 젠틀맨(Robert Gentleman)과 로스 이하카(Ross Ihaka)에 의해 1995년 처음 개발되었는데, 현재는 1997년에 설립된 R 코어 팀(R core team)이 GNU R 프로젝트를 이끌어어나가고 있으며, 벨연구소에서 S 언어 개발에 공헌했던 존 챔버스도 이 팀의 구성원으로 프로젝트 발전에 기여하고 있다. 물론 20여 명으로 구성된 이 R 코어 팀은 중심점 역할을 하는 주체일 뿐, 지금의 R은 수많은 사용자들의 공헌과 노력의 산물이라 할 수 있다.

R은 데이터의 조작(manipulation)과 연산(calculation), 그리고 그래픽 표현(graphical display)을 통합하는 통계 패키지로서 현재 여러 분야에서 널리 사용되고 있다. 특히 금융공학(Financial Engineering)이나 생명공학(Biotechnology)과 같은 전문 분야에서 그 활용도가 높으며, 유럽의 여러 국가들과 미국에서는 대학교 통계분석 관련 수업에서도 R이 활발하게 사용되고 있다. 그 이유는 R이 다음과 같은 장점을 지니기 때문이다.

- S 언어가 가지고 있는 강력한 확장성 덕분에 R은 Linux, UNIX, MAC OS X(Macintosh), Windows 등 모든 운영 체제에서 실행 가능하며, 각종 DBMS(Database Management System)⁵⁾ 데이터에도 쉽게 접근할 수 있다. 또한 별도의 패키지(package)⁶⁾를 이용

4) GNU 프로젝트에 대해 관심이 있거나 더 많은 정보를 알고 싶다면 GNU 웹페이지(<http://www.gnu.org>)를 방문해보기 바란다.
 5) 다수의 사용자가 데이터베이스에 접근하여 데이터를 사용할 수 있도록 하는 소프트웨어 도구를 말하며, 대표적으로 오라클(Oracle), 사이베이스(Sybase), 인포믹스(Infomix) 등이 널리 사용된다.
 6) 패키지에 대한 내용은 제3장에서 자세히 다룰 것이다.

하면 R의 소스(source)를 Java, perl, python, C, C++ 등의 언어와 호환하여 사용할 수도 있다.

- R은 **객체 지향적**(object oriented)인 프로그램이기 때문에 데이터의 조작 및 연산 과정에서 사용자는 보다 쉽고 빠르게 원하는 결과를 얻을 수 있다. 예컨대 방대한 양의 함수나 조건식, 데이터를 모두 “function”, “conditional”, “data” 등과 같은 객체로 전환하여 사용하면 연산 과정이 단순화되고 오류가 줄어 생산성이 향상된다.
- R은 GPL을 따르는 공개 소프트웨어이기 때문에 모든 소스가 공개되며 자유롭게 수정 및 변경이 가능하다. 즉, 기존의 통계 패키지들은 정해진 분석 기법만을 제공하는 반면, R은 정해진 분석 방법에 구애받지 않고 분석자가 자유롭게 분석 기법을 변형하여 사용할 수 있기 때문에 좀 더 다양하고 정밀한 분석을 수행할 수 있다.
- 경제학의 계량분석 방법뿐만 아니라 행정학, 의학, 생물학, 환경학 등 다양한 학문분야에서 사용되는 수천 가지의 통계분석 방법이 **패키지**의 형태로 공개되므로, 사용자는 복잡한 계산식을 일일이 입력하여 분석해야 하는 수고를 덜 수 있다. 예를 들어, plm 패키지를 이용하면 수집된 자료에 단순한 명령어를 입력하는 것만으로 패널 회귀분석(panel regression analysis)을 수행할 수 있다.

하지만 위와 같이 많은 장점이 존재함에도 불구하고, R을 처음 접하거나 프로그램을 다루는 데 능숙하지 않은 사용자는 R을 사용하면서, 혹은 사용하기에 앞서 필연적으로 몇 가지 어려움을 경험하게 된다.

- R은 간단한 명령어만으로도 복잡한 계산을 수행할 수 있는 프로그램이기 때문에 분석을 빠르게 수행할 수 있다는 장점을 가진다. 그러나 대부분의 사용자들은 프로그래밍 언어 자체가 생소할 수 있기 때문에 R을 사용하기 위해서 별도로 R 언어를 배워야 하는 어려움이 따른다.
- R은 전 세계 수많은 사용자들의 상호작용 속에서 프로그램의 오류가 개선되거나, 새로운 기능들이 빠르게 추가된다는 장점을 가진다. 그러나 이는 사용자가 빠르게 변화하는 프로그램 환경에 적응하기 위해 계속해서 새로운 기능들을 습득해야 함을 의미하기도 한다.
- 마지막으로 현재 국내에서는 R을 사용하는 인구가 그리 많지 않기 때문에, R을 처음 공부하는 사용자는 R과 관련된 참고 서적을 구하기가 쉽지 않다. 물론 R의 기본적인 사용 방법이나 프로그래밍 및 그래픽 구현 관련 서적은 다수 존재하지만, R을 이용하여 계량분석을 수행하는 데 필요한 지식을 담은 서적은 그리 많지 않다. 저자 역시 이러한 문제점에 착안하여 이 책을 집필하게 되었음을 밝혀둔다.

R을 배우고 익히는 데 있어 위와 같은 어려움이 존재함에도 불구하고, 계량분석 과정에서 R의 사용이 매력적인 이유는 프로그래밍 언어 기반의 통계 패키지가 가지는

장점들이 단점들을 충분히 상쇄하고도 남기 때문이다. 특히 간단한 명령어만으로 분석을 수행할 수 있어 보다 빠르고 정확한 분석이 가능하다는 점과, 개방된 사용 환경에서 다양한 프로그래밍 소스 코드를 공유할 수 있다는 점이 그것이다. 또한 R이 다른 유료 통계패키지들만큼이나 풍부한 분석 기능을 갖춘 유용한 통계 패키지임에도 불구하고 누구나 자유롭게 무료로 설치하여 사용할 수 있다는 점도 우리가 계량분석 과정에서 R을 사용해야 하는 중요한 이유 중 하나이다.

R은 확장성이 뛰어나며, 객체 지향적인 프로그램이기 때문에 데이터의 조작 및 연산 과정에서 사용자는 보다 쉽고 빠르게 원하는 결과를 얻을 수 있다. 또한 분석자가 프로그램을 자유롭게 변형하여 사용할 수 있기 때문에 좀 더 다양하고 정밀한 분석을 수행할 수 있으며, 다양한 패키지가 제공되기 때문에 사용자는 복잡한 계산식을 일일이 입력하여 분석해야 하는 수고를 덜 수 있다.

3. RStudio



출처 : rstudio.org

RStudio는 R의 **통합개발환경**(Integrated Development Environment, **IDE**) 소프트웨어이다. 통합개발환경이란 소프트웨어 개발 과정에서 필요한 코딩(coding), 디버깅(debugging), 컴파일(compile)의 과정을 하나로 패키지화한 소프트웨어를 의미하는데,⁷⁾ 좀 더 쉽게 설명하면 프로그램을 개발하는 과정에서 사용되는 일련의 과정들을 하나로 통합한 환경을 말한다. 이처럼 RStudio는 기존의 R 개발환경에 새로운 기능들이 추가되어 사용자 효용을 높인 **유틸리티 소프트웨어**(utility software)로 다음과 같은 장점을 가진다.

- RStudio는 모든 R 버전과 호환이 가능하다. 또한 Linux와 UNIX, MAC OS, Windows 등 모든 운영체제에서 실행 가능하며, **RStudio 서버**(RStudio server) 버전을 이용하면 인터넷에 접속할 수 있는 곳이면 어디에서든 RStudio 서버에 접속하여 작업환경을 실행할 수 있다. 물론 R과 마찬가지로 RStudio와 RStudio 서버 역시 **AGPLv3**⁸⁾를 따르기 때문에 자유롭게 사용·수정 및 배포가 가능하다.
- 코딩작업에 필요한 **콘솔**(console)과 신속한 디버깅 작업을 가능하게 하는 **소스 에디터**(source editor), 그리고 **데이터 뷰어**(data viewer)와 **도표 이력**(plot history) 등 통합 개발환경을 구성하는 주요 요소들이 잘 통합되어 있어, 보다 편리하고 신속한 작업 환경이 제공된다.
- 표시되는 구문을 종류별로 구분하고(예를 들어 입력문과 출력문, 함수 등) 여러 가

7) 코딩은 프로그래밍 언어를 이용하여 구체적인 컴퓨터 프로그램을 만드는 기술, 디버깅은 코드상의 오류를 찾아내어 수정하는 과정을 말하며, 컴파일은 컴퓨터가 처리한 언어를 사람이 읽을 수 있는 언어나 그림으로 변환하는 프로그램을 말한다. [출처 : 위키피디아(<http://www.wikipedia.org>)]

8) AGPL(Affero General Public License)은 자유 소프트웨어 재단이 새롭게 발표한 라이선스로, GPL을 따르는 공개 소프트웨어가 서버 소프트웨어의 형태로 공개되는 경우에도 반드시 소스코드를 공개해야 한다는 규정을 담고 있다.

지 다른 색으로 강조하여 표시하는 **구문 강조**(syntax highlight) 기능과 함수 이름의 첫 글자로 함수를 자동으로 검색하거나 함수에 포함될 요소들을 표시해주는 **코드 완성**(code completion) 기능, 코드 입력 시 괄호나 따옴표가 자동으로 입력되는 기능 등이 추가되어 복잡한 수식 입력 과정에서 사용자의 혼동을 줄이고 편의를 증진시켰다.

위와 같은 장점들로 인해 RStudio는 R을 처음 접하는 학생들에게 더욱 편리하고 흥미로운 학습 도구가 될 것이라 생각한다. 따라서 이어지는 “R의 웹 페이지 소개”와 “제2절 : R의 설치 및 실행” 부분에서는 모두 R과 RStudio의 내용을 병행하여 설명할 것이며, 제2장부터는 모든 이론과 예제의 분석 과정에서 RStudio 프로그램을 사용하게 될 것이다.

RStudio는 R의 **통합개발환경** 소프트웨어이다. RStudio는 모든 R 버전과 호환이 가능하며 **RStudio 서버**를 이용하면 인터넷에 접속할 수 있는 곳이면 어디에서든 RStudio 서버에 접속하여 작업환경을 실행할 수 있다. 또한 통합개발환경을 구성하는 주요 요소들이 잘 통합되어 있어, 보다 편안하고 신속한 작업 환경이 제공되며, **구문 강조** 기능과 **코드 완성** 기능이 추가되어 복잡한 수식 입력 과정에서 사용자의 혼동을 줄이고 편의를 증진시켰다.

4. R의 웹페이지 소개

R을 배우는 과정에서 보다 더 많은 정보들을 필요로 한다면 메인 웹페이지(webpage)를 방문해보는 것도 좋은 방법이 될 것이다. 그 이유는 R의 경우 널리 보급되어 사용된 기간이 길지 않아 참고자료를 구하기가 쉽지 않은 데다가, 개방형 소프트웨어이기 때문에 웹페이지에서 제공되는 각종 매뉴얼(manual)들이 오히려 보다 다양하고 풍부한 내용을 담고 있기 때문이다. 여기에서는 R을 배우는 학생들에게 도움이 되는 R의 메인 웹페이지와 RStudio의 웹 페이지에 대해 간략히 소개하고자 한다.

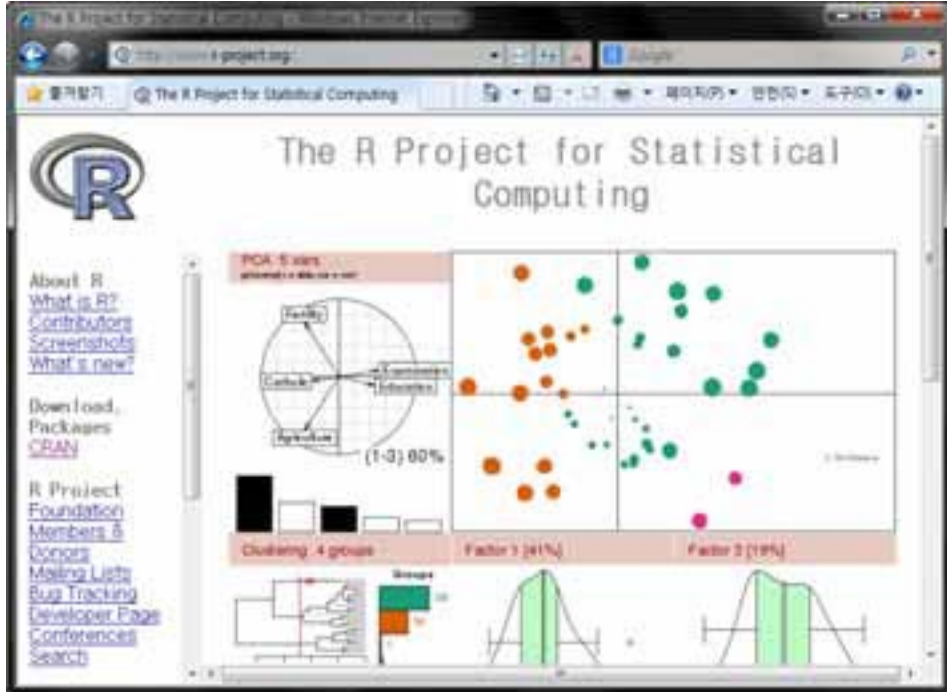
1) R 웹페이지(<http://www.r-project.org>)

R의 웹페이지에 접속하면 실시간으로 제공되는 R에 대한 다양한 정보들을 확인해볼 수 있다. R 웹페이지의 메인 화면은 총 5개의 범주(category)로 분류되어 있는데 간략하게 소개하면 다음과 같다.

- About R : R과 R의 작업 환경에 대한 기본적인 소개와 더불어 R 코어 팀(R core team)과 그 밖의 기여자(contributor) 명단을 확인할 수 있고, 수정되거나 추가된 패키지의 내용 및 각종 오류의 수정 내역도 확인할 수 있다. 또한 주요 운영체제(Windows, Mac, Unix)에서 실행된 R의 화면과 각종 그래픽 예제를 그림으로 살펴볼 수 있다.
- Download, Packages : 인터넷 저장소 및 서버 역할을 하는 **CRAN**(Comprehensive

그림 1-2

R의 웹페이지(<http://www.r-project.org>)



R Archive Network)에 각 국가별로 약 90여 개의 **미러**(mirror) 사이트가 링크되어 있으며, 해당 사이트에 접속하여 R의 설치파일과 패키지를 다운로드할 수 있다.

- R Project : R 재단의 설립 취지와 프로젝트에 참여 중인 회원, 기관 및 후원자들을 살펴볼 수 있다. 또한 우편 목록과 버그 추적(Bug Tracking) 페이지에서는 여러 사용자들이 프로그램과 관련된 다양한 의견을 상호 교환할 수 있다.
- Documentation : R을 배우는 데 좀 더 폭넓은 지식을 쌓고 싶다면 이 링크(link)에 수록된 다양한 매뉴얼을 참고할 수 있다. 특히 「Writing R Extensions」과 「R Data Import/Export」에는 기본 소개서에서 자세하게 다루지 않은 프로그램의 제작과 데이터 입/출력에 대한 내용들이 다수 포함되어 있다.
- Misc : 세계 각지의 R 사용자 그룹 홈페이지들이 링크되어 있다. 참고로 한국의 사용자 그룹 홈페이지(<http://r-project.kr>)를 방문하면 보다 쉽게 R을 접할 수 있을 것이다.

2) RStudio 웹페이지(<http://www.rstudio.org>)

RStudio는 통합개발환경 소프트웨어인 만큼 다양하고 편리한 기능들을 많이 포함하고 있음에도 불구하고 시중에서 RStudio의 기능이나 사용법을 설명하는 책을 거의 찾아볼 수 없다. 따라서 RStudio에 포함된 보다 많은 기능들에 대해 자세히 알아보고 싶다면 웹페이지에 수록되어 있는 설명을 참조하는 것이 바람직할 것이다. RStudio 웹페이지에서는 RStudio IDE, Shiny, Training, Projects, Blog 등 다양한 카테고리(category)들이 제공

되는데, 여기서는 가장 유용한 RStudio IDE 카테고리에서 제공되는 하위 6개 메뉴들에 대해서 소개하고자 한다.

- About : RStudio에 대한 간략한 소개와 더불어 개선된 기능들과 특징에 대해 소개한다.
- Screenshots : 주요 운영체제(Windows, MAC 등) 및 웹에서 실행된 RStudio의 화면과, RStudio에 포함된 새로운 기능들을 직접 구현한 모습을 그림으로 살펴볼 수 있다.
- Download : RStudio와 RStudio 서버 버전을 다운로드할 수 있다. 단, RStudio는 어디까지나 R의 유틸리티 소프트웨어이므로 기존에 R이 설치되어 있지 않다면 RStudio를 설치하더라도 실행되지 않는다.
- Documentation : RStudio와 RStudio 서버를 사용하면서 유용하게 참고할 수 있는 매뉴얼을 다수 포함하고 있다. 또한 RStudio를 더욱 심도 있게 배우고자 하는 사용자는 고급형 주제(advanced topic)를 다루는 매뉴얼을 참고할 수 있다.
- Support : 각종 도움말 정보가 제공된다.
- Development : Github를 통하여 R의 소스코드를 공유·관리하는 페이지이다. Git는 프로그램의 소스 코드를 관리하는 **분산 버전 관리 시스템**(distributed version control system)⁹⁾이며 GNU 프로젝트의 GPL을 따르기 때문에 자유 소프트웨어이다. Github는 Git 사용자를 위한 웹 호스팅 서비스(web hosting service)이다.

그림 1-3

RStudio 웹페이지(<http://www.rstudio.org>)



9) <http://git-scm.com>에서 인용.

제 2 절

R의 설치 및 실행

이번 절에서는 R과 RStudio의 설치와 실행 방법에 대해 알아본다. 제1절에서 설명했던 것처럼 R과 RStudio는 기존에 널리 사용되고 있는 상업용 통계패키지와는 달리 프로그램의 사용이나 수정 및 배포에 제약이 없는 공개 소프트웨어이기 때문에 별도의 라이선스 (license) 없이 설치하고 사용할 수 있으며, 프로그램의 설치파일 역시 웹페이지에 접속하여 자유롭게 다운로드할 수 있다. 앞서 이 책에서 배우게 될 모든 이론과 예제는 RStudio 프로그램을 이용하여 분석할 것이라고 하였는데, RStudio를 사용하기 위해서는 반드시 R이 설치되어 있어야 한다. 먼저 R의 설치 방법부터 알아보자.

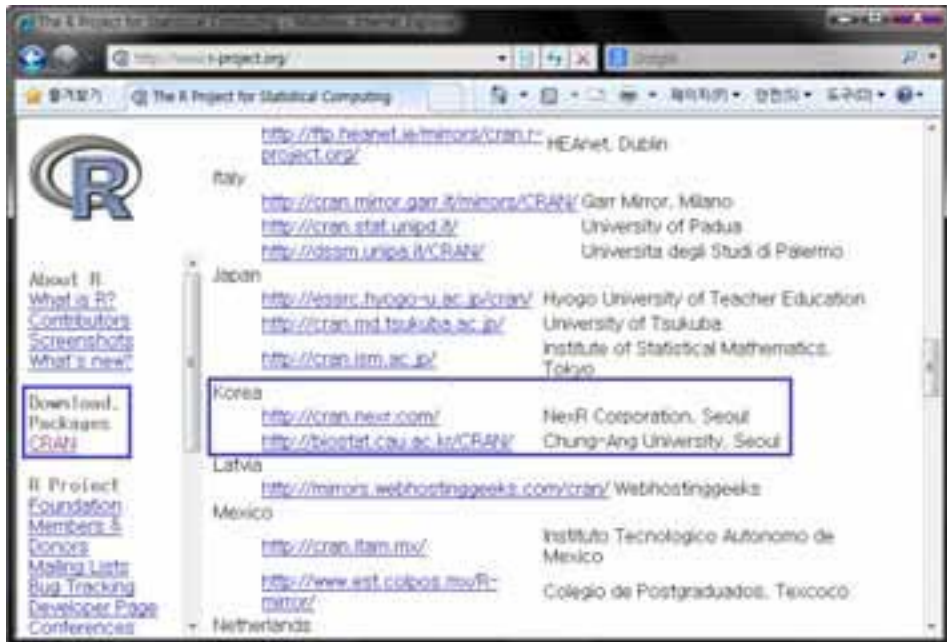
1. R의 설치 방법

R을 설치하기에 앞서, 먼저 R의 설치파일을 다운로드해보도록 하자. R의 설치파일은 R의 웹페이지(www.r-project.org)에 접속하여 다운로드할 수 있다.

➔ 먼저 R의 웹페이지에 접속한 후 그림 1-4와 같이 **CRAN** 링크를 클릭하고, 국가별 CRAN 페이지 목록에서 대한민국(Korea) 웹페이지 주소의 링크(<http://cran.nexr.com> 또는 <http://biostat.cau.ac.kr/CRAN>)를 선택한다.

그림 1-4

CRAN mirror 선택

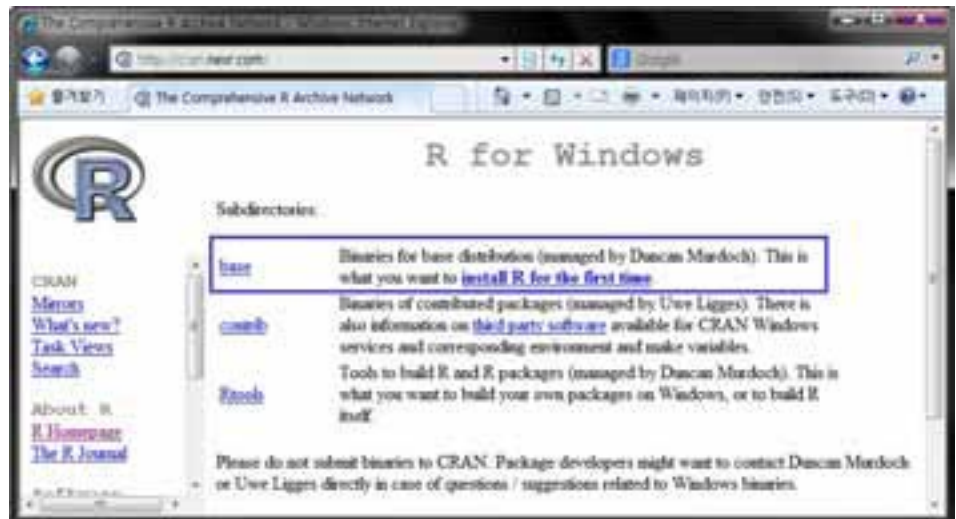


➔ 선택이 완료되면 운영체제 선택메뉴가 나타나는데, 현재 사용자의 운영체제에 맞추어 선택한다. 저자는 Windows 운영체제를 사용 중이기 때문에 **Download R for Windows** 메뉴를 선택하였다.

➔ 선택이 완료되면 그림 1-5와 같이 세 가지 메뉴가 나타나는데, **base** 메뉴를 선택한다. base 메뉴는 R 사용자를 위한 기본 배포 버전을, **contrib** 메뉴는 실시간으로 제공되는 R의 패키지와 소스코드를, 그리고 **Rtools** 메뉴는 R과 R 패키지를 자유롭게 제작·수정할 수 있는 개발도구를 각각 포함하고 있다. 참고로 R의 패키지는 contrib 메뉴를 이용하여 설치할 수도 있지만, R 프로그램 내에서 필요한 패키지를 쉽게 설치할 수 있으므로 후자의 방법을 이용하는 것이 좀 더 편리하다.

그림 1-5

다양한 R의 개발 도구



➔ 그림 1-6과 같은 화면이 나타나면 마지막으로 **Download R 2.15.3 for Windows**를 클릭하여 설치파일을 다운로드한다. 참고로 2.15.3은 R의 버전(version)을 나타내며, 자유 소프트웨어의 특성상 새로운 버전으로 자주 업데이트되는 경향이 있으므로 사용 중인 R의 버전을 꼼꼼히 확인하여 업그레이드해주는 것도 좋을 것이다. 물론 R의 버전 업그레이드도 실행 중인 프로그램 안에서 수행할 수 있다.

➔ 이제 설치파일을 이용하여 R을 설치해보자. 다운로드한 설치파일을 실행하면 **설치 언어 선택** 대화상자가 나타난다. 아직까지 한국어 번역이 매끄럽지 못한 부분이 있어 영문 버전 사용을 권장하나, 한국어 버전도 사용에 큰 지장은 없다.

➔ **확인** 버튼을 클릭하면 설치 초기화면이 나타나며 **다음** 버튼을 클릭하여 설치를 진행한다.

그림 1-6

R의 설치프로그램 다운로드



➔ 다음 단계에서 정보 대화상자에 GPL 라이선스에 대한 내용이 나타나는데, 라이선스의 내용에 동의하면 다음 버튼을 선택하고 설치할 위치 선택 대화상자가 나타나면 다음 버튼을 선택한다. 물론 설치 경로는 사용자 임의로 변경 가능하며, R의 작업 환경을 직접 수정·개발하거나 방대한 양의 데이터를 다루는 과정에서 R의 설치 경로를 콘솔에 직접 입력하여 사용하는 경우가 발생하므로, 이에 해당되는 사용자라면 C:\R과 같이 설치경로의 입력 범위를 단축하는 것도 좋다.

➔ 그림 1-7과 같이 구성 요소 설치 대화상자가 나타나면 구성 요소 선택 옵션이 32비트 표준 설치(32-bit user installation)¹⁰로 지정되어 있는 것을 볼 수 있는데, 이것은 현재 저자가 32비트 운영체제를 사용하고 있어 자동으로 선택된 사항이다.

그림 1-7

구성 요소 설치 옵션



10) 여기서 32-bit는 프로세서와 운영체제의 연산 단위를 나타내는데, 만약 64bit 운영체제를 사용하고 있다면 64-bit로 나타난다.