韓國數學敎育學會誌 시리즈 E <數學敎育 論文集>
http://dx.doi.org/10.7468/jksmee.2015.29.4.573
제 29집 제 4호, 2015. 11. 573-588

J. Korea Soc. Math. Ed. Ser. E:
Communications of Mathematical Education
Vol. 29, No. 4, Nov. 2015. 573-588

# R을 활용한 '대화형 통계학 입문 실습실' 개발과 활용

이 상 구 (성균관대학교)†
이 긍 희 (한국방송통신대학교)
최 용 석 (부산대학교)
이 재 화 (한림대학교)
이 지 영 (서울대학교)

본 연구에서는 최근 통계 분야에서 활용도가 급격히 높아지고 있는 중요한 컴퓨터 언어이자 오픈 소스 통계 프로그램인 R을 활용하는 '대화형(interactive) 통계학 입문 실습실'의 개발 과정과 내용 및 활용을 다룬다. 최근에 개발을 마치고 2014 서울 세계수학자대회에서 소개된 후, 통계 강좌 등에 실제 사용되는 웹상의 R을 활용한 본 대화형 통계학 입문 실습실의 특징은 웹상에서 본문과 예제의 설명 및 풀이과정과 함께 대응하는 R 명령어 코드들을 함께 박스 안에 제공하여, 실습 때마다 일일이 컴퓨터 명령어 코드들을 입력해야하는 번거로움을 없앴다. 또한, 명령어의 실행을 위하여 프로그램을 설치하지 않고, 명령어 상자 아래 제공한 [클릭-실행] 버튼을 누르기만 하면, 클라우드 컴퓨팅으로 그 결과값과 그래픽을 동시에 바로 같은 화면에서 확인하면서, 시뮬레이션 및 실습을 할 수 있고, 더 나아가 그와 유사한 다른 문제에 함수와 조건만을 수정하여 바로 사용할 수 있는 편리함이 추가되었다. 그 결과 대화형 통계학 입문 실습실에서는 R 명령어를 이해하는 데 필요한 시간과 노력이 대폭 줄어들 뿐 아니라, 초보자에게 통계학 입문 과목을 지도하기에 적절하며, 그밖에 다양한 Java 시각화 도구와 이미지 및 통계 자료를 사용하여 사용자 맞춤형 강의실 개발이 가능하여 통계학입문 강의를 수강하는 학생들의 관심과 흥미를 유도할 수 있도록 하였다. 본 연구에서는 본 실습실을 통계입문 강좌의 효과적인 실습실 모델의 하나로 소개한다.

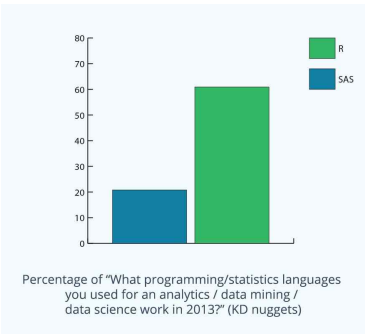
## Ⅰ. 서 론

통계적 지식과 활용은 정보화 사회가 가속화되는 21세기에 점점 더 중요해 지고 있다. 이에 세계 여러 대학에서는 통계 교육을 보다 효과적으로 진행하기 위하여 대화형 통계 실습실[1]을 개발하여 활용하고 있다. 대화형 통계 실습실은 클라우드 컴퓨팅을 추구하는 세계적인 추세에도 부합할 뿐만 아니라, 프로그래밍 언어를 학습하는 시간을 줄일 수 있기 때문에,[2] 학생들이 통계적 지식에 접근하기 쉽게 해 준다. 또한 통계적 이론과 개념을 학습하는 시간이 상대적으로 늘어나게 되어 학생들의 통계에 대한 이해력을 높이는데 적절한 환경을 제공한다. 한국에서는 부산대와 성균관대를 포함한 여러 학교에서 이미 공학적 도구의 필요성에 의해 개발된 통계실습실을 활용하기 시작하였다[3].

* 접수일(2015년 6월 30일), 심사(수정)일(1차: 2015년 7월 24일, 2차: 2015년 8월 22일), 게재 확정일(2015년 8월 27일)
* ZDM 분류 : M85, G45, M15
* MSC2000 분류 : 97C80, 97U70
* 주제어 : 통계학, 실습실, 모델, 세이지(Sage), R
† 교신저자 : sglee@skku.edu
1) http://wise.cgu.edu http://www.socr.ucla.edu
2) http://m.wolframalpha.com/examples/Statistics.html
3) http://matrix.skku.ac.kr/2015-R-Statistics/R-Sage-Statistics-Lab-1.htm

대학의 통계학 강좌에 활용되는 많은 소프트웨어 중에 본 연구에서 주목한 것은 현재 대학뿐만 아니라 구글과 페이스북, 뉴욕타임즈 등이 사용하는 R[4]이다. 2014년 6월 3일 권위 있는 데이터 캠프 블로그에 실린 'What is the best statistical programming language?'[5] 라는 제목의 글에서, 통계학의 가장 보편적인 소프트웨어 3개가 SAS, SPSS, R임을 확인하였으며, 특히 2013년 당신이 사용한 통계학 소프트웨어 부분 설문조사에서 R은 SAS나 SPSS를 압도적으로 추월하면서 1등으로 선정되었다.

R은 다양한 통계 계산 기법과 그래픽스 및 수치 해석 기법을 지원하는 오픈 소스 프로그래밍 언어이자 소프트웨어 환경이다(Bloomfield, 2014;Stowell, 2014). R은 특히 수학 기호를 포함할 수 있는 출판물 수준의 훌륭한 그래픽 기능 때문에, 통계학자들 사이에서 통계 소프트웨어 개발과 자료 분석에 널리 사용되고 있다. R은 최초 설치시 함께 설치되는 핵심적인 패키지이외에, 사용자가 직접 제작한 패키지들을 쉽게 추가하여 기능을 확장할 수 있도록 설계되었다. 윈도우, 맥 OS 및 리눅스를 포함한 UNIX 플랫폼에서 이용 가능한 R은 행렬 계산을 위한 도구로서도 사용될 수 있으며 이 부분에서 GNU Octave나 MATLAB에 견줄 만한 결과를 보여준다. 실제 R은 S의 상용판인 S-plus보다 많은 경우 처리 속도가 빠를 뿐만 아니라, 범용 행렬계 언어의 표준이라고도 하는 MATLAB과 그 파생어인 GNU Octave, Scilab보다도 종합적으로 빠르다고 평가된다. 초보자의 경우 R의 사용법은 웹 주소[6]를 이용할 수 있다.




[그림 Ⅰ-1] 2013년 당신이 사용한 통계 소프트웨어 설문조사


[그림 Ⅰ-2] 통계 소프트웨어 비교

R은 기존의 상업용 통계소프트웨어를 대체하면서 통계학 강좌에 사용되는 기본 도구로 빠르게 시장 점유율을 높여가고 있다. 다만 프로그래밍에 대한 지식이 초보적인 사용자의 경우 프로그래밍하여 만든 스크립트(script)로 R을 실행하기가 쉽지는 않다. 더구나 최용석(2014a)과 같은 초급 수준의 통계학 입문서는 많지만 빅 데이터 분석과 데이터 마이닝의 근간이 되는 다변량 통계학과 같은 중, 상급 이상의 방법론을 다룬 안내서(최용석 · 정광모, 2003;최용석, 2014b;허명회, 2014)는 드물다. 그러나 R은 다양한 기본 시스템과 전문가들이 기여한 공개 패키지로 구성되어 있으며 현재 R의 패키지 수는 약 5천개 정도가 될 정도로 다양하다. 기존 타 메뉴 방식의 통계 소프트웨어와는 달리 오브젝트 기반으로 작동하므로 분석 결과를

4) R은 GNU("GNU's Not UNIX"의 약자) GPL(General Public License)하에 무료로 배포되는 S 프로그래밍 언어의 구현으로 GNU S라고도 한다. 공식 홈페이지는 http://www.r-project.org

5) http://blog.datacamp.com/statistical-language-wars-the-infograph

6) http://cran.r-project.org/doc/contrib/Park-BeginnersRcourse.pdf

나중 작업에 사용할 수 있으며 엑셀(Excel)의 자료를 쉽게 불러들이고 R에서 작업한 자료를 엑셀로 내 보낼 수 있는 장점이 있다.

이미 세계 최고 수준의 모바일 인프라를 갖추고 있는 한국은 시간과 장소에 구애받지 않고 웹(Web) 도구를 이용하여 누구나 무료로 수학 연산 및 시뮬레이션 실습을 할 수 있는 잠재력을 갖추고 있다. 본 연구진은 이러한 강점을 활용하여 수 년 동안 공개 소프트웨어인 Sage[7]를 기반으로 자체적으로 구축한 연산서버를 통해 다양한 시도를 하였다. 특히 미적분학, 선형대수학 및 공학수학 교육을 위해 개발한 웹 콘텐츠와 모바일 콘텐츠 및 계산도구는 교재에 반영되어 교육에 꾸준히 활용되고 있다(고래영 외, 2009;김경원・이상구, 2013;이상구・신준국・김경원, 2014;이상구・이재화・김경원, 2014; 이상구・이재화・김덕선, 2012;이상구・장지은・김경원, 2013;Kim et al, 2013;Lee et al, 2014; Lee et al, 2013). 본 연구에서는 그간 축적된 Sage 활용의 기술력을 통계학에 적용하고 급속도로 많은 관심을 받고 있는 R을 접목시켜 통계 입문반에서 활용할 수 있는 대화형 통계학 입문 실습실 모델의 개발 내용을 소개한다.

과거에는 강의에서 배운 이론을 실습하기 위해 고가의 각종 통계 패키지를 구입하고 이의 사용법을 숙지하여야 했다. 본 연구에서 소개하는 통계 대화형 실습실에서는 프로그램을 설치할 필요 없이 인터넷만 연결되어 있으면 모바일을 통해 홈페이지에 접속하여 웹주소를 클릭하는 것만으로도 각종 통계 패키지의 사용과 연산 및 시뮬레이션 실습이 가능하다. 또한 수업에서 배우는 통계적 지식을 바탕으로 제시된 R 명령어를 통해 다양한 예제를 다룰 수 있다. 이런 접근은 통계학에 대한 학생들의 접근성을 최대화 하여 통계학 교육의 기반을 혁신적으로 확대할 수 있다.

## Ⅱ. 본론

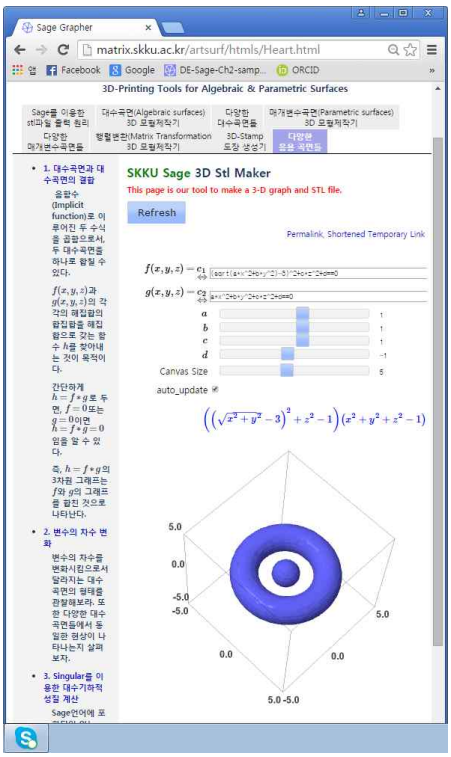
본 연구는 대학에서 배우는 통계학 지식의 습득과 실습을 시간과 장소에 구애받지 않고 누구나 무료로 활용할 수 있도록 하는 환경을 제공하는데 그 목적이 있다. 지금까지 대중적으로 활용되고 있는 상업용 통계 소프트웨어는 사용자에게 비용 부담이 될 뿐 아니라 설치한 소프트웨어의 유지와 보수 그리고 프로그래밍 언어 학습 등의 여러 가지 제한점을 가지고 있다. 따라서 이러한 한계를 극복하고 모바일 시대에 맞도록 모바일 기기를 이용하여 언제 어디서든 쉽게 통계 프로그래밍을 실행할 수 있는 환경을 제공하기 위하여, 한국에 맞게 자체적으로 개발한 오픈소스 소프트웨어인 Sage 기반의 서버를 이용하여 통계학 입문에 필요한 콘텐츠와 웹(Web) 도구를 개발하였다. 그 결과 기존의 상업용 통계 소프트웨어를 사용할 때 제기되는 문제점들을 해결할 수 있을 뿐 아니라, 아래 [그림 Ⅱ-1] 에서 보는 '스크롤바(Scroll Bar)'를 이용하여 마우스의 간단한 조작만으로도 다양한 시뮬레이션이 가능하도록 하였다. 그리고 최근에 '공유와 협력의 교과서 만들기 운동본부'의 빅북(Bigbook) 운동을 통하여 발간된 부산대 최용석 교수의 무료 전자 교재 '(빅북) R과 함께하는 통계학의 이해[8]'를 다운 받아 함께 사용하면 R 코드를 단순히 복사만하여 본 실습실에서 활용할 수 있으므로 그 효과를 극대화 할 수 있다. 사용방법은 빅북 전자교재 '선형대수학[9]'에서 사용한 방식과 일치한다.

---

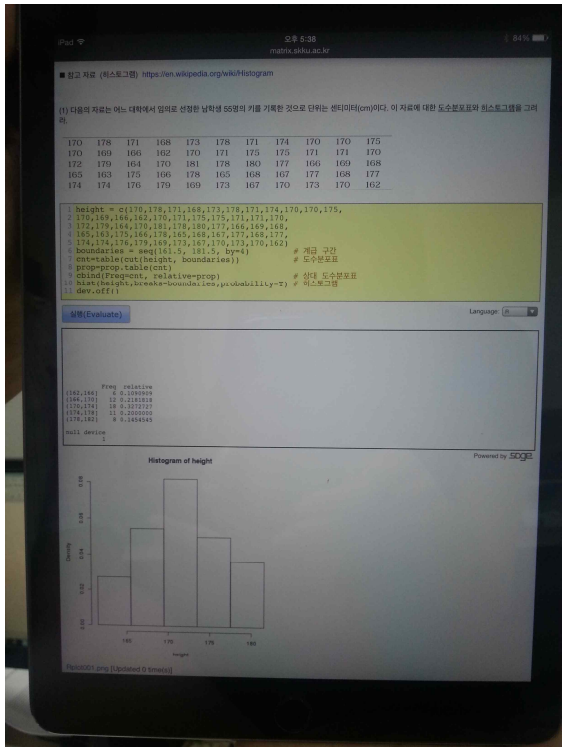
7) http://www.sagemath.org/ Firefox 및 크롬, 사파리 등의 브라우저에서 무료로 모든 실습이 실시간으로 가능하다.

8) http://www.bigbook.or.kr/bbs/bbs/board.php?bo_table=bo16&wr_id=5

9) http://www.bigbook.or.kr/bbs/bbs/board.php?bo_table=bo16&wr_id=2




[그림 II-1] 추가된 스크롤바



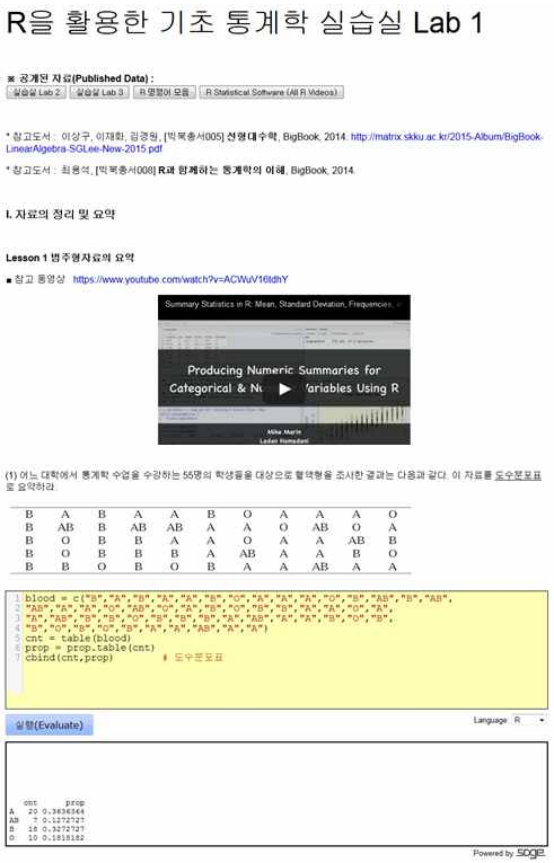

[그림 II-2] 모바일 기기에서 실습실 활용

서론에서 언급한 바와 같이 Sage는 이미 미적분학, 선형대수학, 공학수학 실습실에서 활발히 이용되고 있다. 본 연구에서 Sage를 활용하여 개발한 통계 실습실은 python 기반의 Sage 언어 대신에 통계학에서 최근 그 사용이 빠르게 확산되는 R 언어를 이용하여 모든 실습이 가능하도록 만들었다. 그리고 예제와 설명 아래에, R 명령어를 미리 입력하여 두어, 실습 때마다 매번 명령어를 입력해야하는 번거로움 없이 바로 실행하기를 클릭하면서 실습하고, 그와 유사한 다른 문제에 함수와 조건만 수정하여 바로 사용할 수 있도록 하였다. 더구나 R 소프트웨어를 다운 받아 사용해야 하는 불편함을 완전히 제거하였을 뿐만 아니라, 모바일 기기의 저장용량에도 전혀 영향을 받지 않는다([그림 II-2] 참조).

통계 입문자들에게는 필수적으로 기본적인 컴퓨터 코딩 지식이 요구되지만 통계학 입문 교육은 보통 컴퓨터 언어 교육 기회 보다 먼저 제공되는 추세이다. 따라서 통계학 입문자에게 컴퓨터 코딩 지식을 통계학 입문 교육 안에서 가장 효과적으로 학습 시키는 방법에 대한 고려와 대안 제시는 필수적이다. 본 연구와 개발 결과는 통계학 입문에 필요한 코딩 지식을 포함한 R 명령어를 이해하는 데 필요한 시간과 노력을 대폭 줄여 줄 뿐 아니라, 초보자에게 통계학 입문 과목을 지도하기에도 적절하다. 게다가 다양한 동영상 및 학습 자료의 사용으로 실제 학생들의 관심과 흥미를 유도할 수 있다. 이 실습실은 통계입문 강좌의 효과적인 실습실의 모델이 될 수 있다고 판단한다. 이제 본 연구진이 개발한 R을 활용한 대화형 통계학 입문 실습실 모델의 구조와 활용 및 개발에 대하여 논한다.

### 1. 통계 실습실의 기본 구조

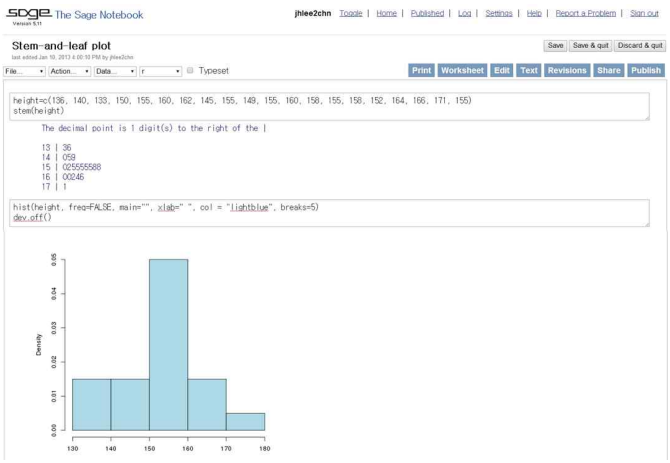
통계 실습실은 하나의 html 파일에 기초통계학 강좌에 적합한 예제 문제와 R 명령어를 실행할 수 있는 Sage 셀, 통계 관련 웹페이지 주소 및 관련 강의 녹화 파일을 통합하여 구성되어 있다. 따라서 기본적인 Sage 명령어를 이용한 통계 및 확률 계산, R 명령어를 자유롭게 사용할 수 있다([그림 Ⅱ-3]).



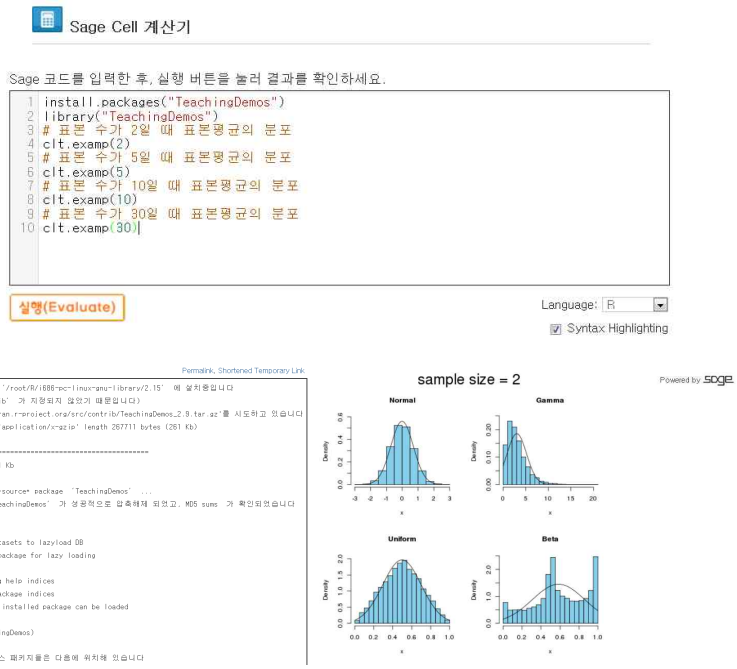

[그림 Ⅱ-3] 통계학 실습실 모델

물론 기존의 Sage 노트북과 Sage 셀에서도 아래 [그림 Ⅱ-4], [그림 Ⅱ-5]와 같이 명령어를 R로 설정하면 아무런 제약 없이 R 명령어를 사용할 수 있게 되어있으며, 다양한 패키지(package)[10]도 사용가능하다. 그러나 Sage 노트북은 로그인을 해야 하는 번거로움이 있으므로 본 실습실에서는 Sage 셀을 연산에 활용하였다.

10) Sage 노트북과 Sage 셀에서도 대부분의 R 패키지를 설치하여 이용할 수 있다. Sage에 이미 설치되어 있는 패키지는 library() 명령어를 이용하여 확인할 수 있다.




[그림 II-4] Sage 노트북에서 R 명령어 실행모습




[그림 II-5] Sage 셀에서 R 명령어 실행모습(패키지 사용)

그리고 R 명령어의 계산은 다음 <표 Ⅱ-1>과 같은 계산 서버를 이용하여 수행한다.

<표 II-1> 계산 서버

| Sage 셀 (http://sage.skku.edu/) | Sage 노트북 (http://math1.skku.ac.kr/) |
|---|---|
|  |  |





## 2. 통계 실습실의 주소

대화형 통계학 입문 실습실은 대학 기초통계학 강좌에 맞는 대부분의 콘텐츠를 모두 포함하여 구성할 수 있다. 아래 <표 II-2>는 이 중 일부를 나타낸다. 물론 학생의 수준과 강좌의 특징에 맞추어 내용의 일부를 선별적으로 선택하여 실습하는 것도 가능하다.

<표 II-2> 장별 통계학 실습 내용과 실습실 웹사이트 주소

| 장 별 | 실습 내용 | 실습실 |
|---|---|---|
| 1 | 통계학의 이해 | 없음 |
| 2 | 자료의 정리 및 요약 | http://matrix.skku.ac.kr/2015-R-Statistics/R-Sage-Statistics-Lab-1.htm |
| 3 | 이산 확률변수 및 분포 | http://matrix.skku.ac.kr/2015-R-Statistics/R-Sage-Statistics-Lab-2.htm |
| 4 | 연속 확률변수 및 분포 | |
| 5 | 표집분포와 중심극한정리 | |
| 6 | 추정 | |

이렇게 웹 주소를 이용하여 통계를 지도하는 실습실의 장점은 휴대용 기기의 사용이 증가하고 있는 추세에 맞추어 각 장 별 실습내용을 시간과 장소에 불문하고 쉽게 프로그래밍을 학습, 수정 및 실행이 가능하므로 실습 효과를 높일 수 있다는 것이다. 대부분 통계학 개론의 강의는 이론과 실습을 병행하여 이루어지며, 컴퓨터 실습을 위해 SAS, SPSS, MATLAB 등 고가의 프로그램이 필요하다. 그러나 정작 실습수업을 수강하는 학생들은 이런 통계 패키지를 자유롭게 이용하지 못하고 있는 실정이다. R의 이용은 기존의 엑셀 또는 SAS, SPSS의 장점을 모두 가지고 있으며, 새로운 언어를 학습해야 하는 번거로움을 줄이면서, 데이터 처리(불러오기, 저장하기 등) 및 다른 프로그램과 호환이 가능하다. 따라서 본 연구에서는 이러한 현실적 어려움을 해결하기 위하여 통계 실습실의 주소를 사용한다면 많은 학생들이 손쉽게 이용할 수 있을 것이라는 아이디어를 적용하였다.

특히 각 장 별 웹 주소에는 다양한 예제와 프로그램이 제공되어 있고, 또 약간의 프로그램 수정으로 시뮬레이션이 가능하므로 학생들의 학습효과도 높을 것으로 기대된다. 교수자는 학습자의 실습내용 이해도를 다양한 예제를 통하여 반추할 수 있고, 또한 학습자의 학습 능력에 따라 실습내용을 유연하게 편성할 수 있다. 학습자는

본인의 학습능력에 맞추어 예습과 복습을 자유롭게 할 수 있는 것도 큰 장점이라고 할 수 있다.

### 3. 통계 실습실의 개발과 활용

본 연구진이 개발한 통계학 입문 실습실 모델을 어떻게 효율적으로 활용 할 수 있는지에 대하여 설명한다. 먼저 교수자는 자신이 강의하고자 하는 내용과 실습의 범위에 따라 기존에 이미 만들어진 실습실과는 차별화된 강의 맞춤형 실습실을 구성할 수 있다. 예를 들어 교수자가 직접 작성한 R 명령어를 Sage 셀에 추가하여 html 파일에 포함시키려면 다음 [그림 Ⅱ-6]과 같이 html 소스 파일의 해당 위치에 추가하여 만들 수 있고, 필요한 통계 관련 웹 주소 링크도 다음 [그림 Ⅱ-7]과 같이 html 소스 파일에 추가할 수 있다.

```
<div id="cell_outer_1" class="cell_visible">
    <div id="cell_1" class="cell_evaluated">

        <div class="cell_input_print"><script type="text/code">
s_A = c(3, 4, 3, 4, 5, 3, 3, 4, 5, 4, 4)
s_B = c(2, 4, 4, 3, 2, 3, 4, 5, 4, 1, 3)
s_C = c(3, 2, 4, 2, 1, 3, 3, 4, 4, 3, 2)
scores = data.frame(s_A,s_B,s_C)
boxplot(scores)
dev.off()</script></div>

    </div>
</div>
```

[그림 Ⅱ-6] R 명령어를 html 파일에 추가하는 예시

```
<iframe width="960" height="800"
src="http://www.internettrend.co.kr/trendForward.tsp"
frameborder="0"></iframe>
```

[그림 Ⅱ-7] 웹 주소 링크를 html 파일에 추가하는 예시

이렇게 개발된 대화형 통계학 입문 실습실을 통해 학생들은 직접 코드와 데이터를 바꿔가면서 새로운 문제들에 대한 다양한 실습과 시뮬레이션을 할 수 있으므로, 과거 텍스트북에만 의존한 학습보다 다양한 차별화된 효과를 기대할 수 있다. 매주 학생들에게 숙제로 다른 교재나 인터넷을 사용하여 문제를 찾은 후 학습한 코딩을 변형시켜 코딩하여 얻은 결과를 정리하여 제출하게 하면, 학생들은 동일한 유형의 문제 해결 및 코딩에 자신감을 갖게 된다.

이제 대표적인 콘텐츠의 내용에 대하여 설명한다. 먼저 각 실습실에는 다른 실습실로 연결되는 링크를 주어 자유자재로 실습실을 옮길 수 있도록 하였고, 사용된 “R 명령어 모음”[11]과 R 관련 동영상[12] 및 관련 개념을 출처와 함께 첨부하여 추가정보에 접속할 수 있도록 하였다. 특히 통계에서는 데이터를 실질적으로 다룰 수 있기 위해 다른 주소의 데이터를 불러오는 방법을 익히는 것이 필수적인데, “R 명령어 모음”의 “데이터 불러오기” 예제에서는 데이터 자체를 코드로 바꾸어 입력하는 방식이 아닌 .xls, .txt, .data, .csv와 같은 파일을 웹 서버로부터 R 또는 Sage 코드로 읽어오는 read.csv('http://matrix.skku.ac.kr/2015-R-Statistics/wine.data', header = FALSE) 이라는 명령어를 사용하였다. ( [그림 Ⅱ-8]참조)

11) http://matrix.skku.ac.kr/2015-R-Statistics/R-Sage-commands-examples.htm
12) https://www.youtube.com/playlist?list=PLqzoL9-eJTNBDdKgJgJzaQcY6OXmsXAHU

<예제> 데이터 불러오기

HTML로부터 데이터 불러오기에 대한 R 코드

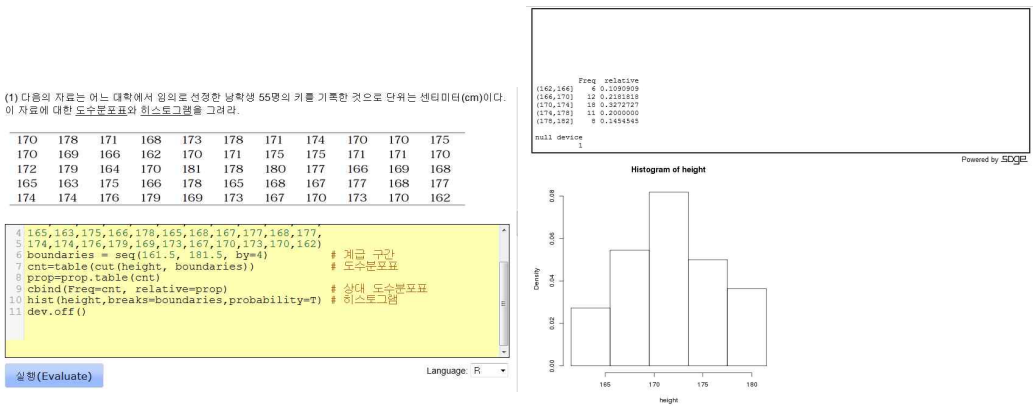
```
## 데이터 불러오기(HTML)

read.csv('http://archive.ics.uci.edu/ml/machine-learning-databases
/wine/wine.data',header=FALSE)
```

| | V1 | V2 | V3 | V4 | V5 | V6 | V7 | V8 | V9 | V10 | V11 | V12 | V13 | V14 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 1 | 1 | 14.23 | 1.71 | 2.43 | 15.6 | 127 | 2.80 | 3.06 | 0.28 | 2.29 | 5.640000 | 1.040 | 3.92 | 1065 |
| 2 | 1 | 13.20 | 1.78 | 2.14 | 11.2 | 100 | 2.65 | 2.76 | 0.26 | 1.28 | 4.380000 | 1.050 | 3.40 | 1050 |
| 3 | 1 | 13.16 | 2.36 | 2.67 | 18.6 | 101 | 2.80 | 3.24 | 0.30 | 2.81 | 5.680000 | 1.030 | 3.17 | 1185 |
| 4 | 1 | 14.37 | 1.95 | 2.50 | 16.8 | 113 | 3.85 | 3.49 | 0.24 | 2.18 | 7.800000 | 0.860 | 3.45 | 1480 |
| ..... | | | | | | | | | | | | | | |
| 176 | 3 | 13.27 | 4.28 | 2.26 | 20.0 | 120 | 1.59 | 0.69 | 0.43 | 1.35 | 10.200000 | 0.590 | 1.56 | 835 |
| 177 | 3 | 13.17 | 2.59 | 2.37 | 20.0 | 120 | 1.65 | 0.68 | 0.53 | 1.46 | 9.300000 | 0.600 | 1.62 | 840 |
| 178 | 3 | 14.13 | 4.10 | 2.74 | 24.5 | 96 | 2.05 | 0.76 | 0.56 | 1.35 | 9.200000 | 0.610 | 1.60 | 560 |

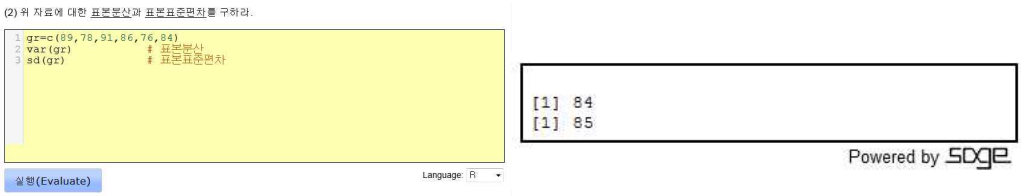
[그림 II-8] 데이터 불러오기 예시

통계자료에 대한 도수 분포표와 히스토그램은 [그림 II-9]와 같이 처리된다.




[그림 II-9] 도수분포표와 히스토그램

표본 분산과 표본표준편차는 [그림 II-10]과 같이 처리된다.

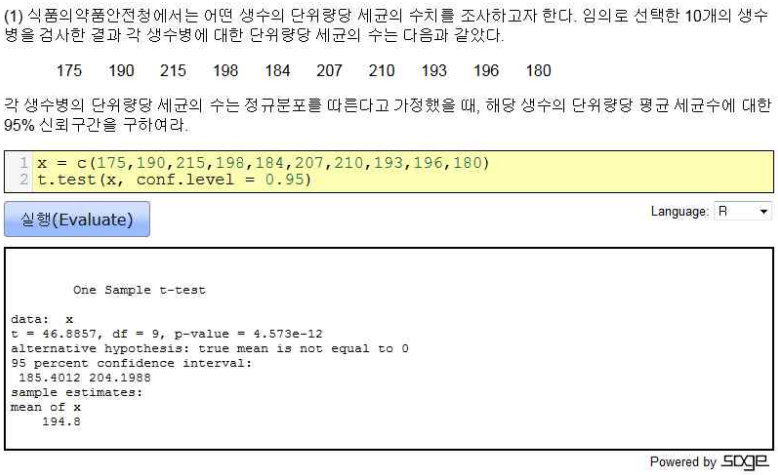



[그림 II-10] 표본분산과 표본표준편차

조건부 확률은 [그림 II-11]과 같이 처리된다.


[그림 II-11] 조건부 확률

모평균과 모비율에 대한 신뢰구간은 [그림 II-12]와 같이 처리된다. 그 외에 실습실의 각 예제마다 각 단원의 중요한 정의나 정리 등을 간단하게 클릭하여 확인해 보면서 실습을 할 수 있도록 하여 사이버 실습실의 장점을 살렸다.

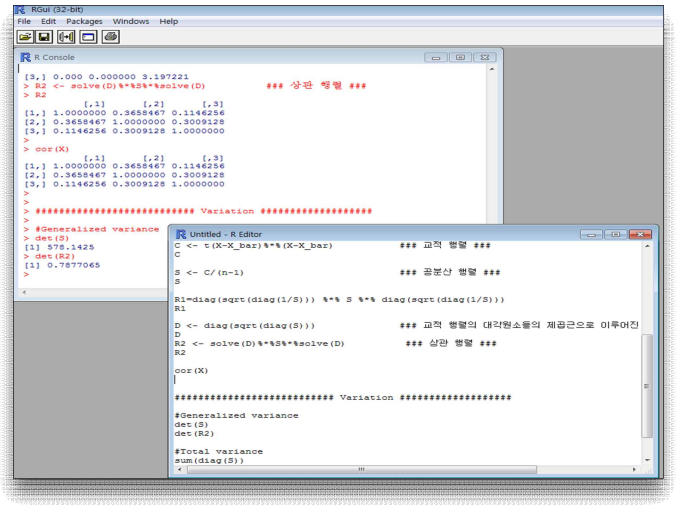



[그림 II-12] 모평균의 신뢰구간

여기까지 대화형 통계학 입문 실습실이 다루는 내용을 몇 가지 예를 통해 살펴보았다. 현재 실제 수업에 적용할 수 있는 디자인, 셀의 크기, 다양한 흥미로운 소스의 사용 등에 대하여 필드 테스트가 진행되고 있다.

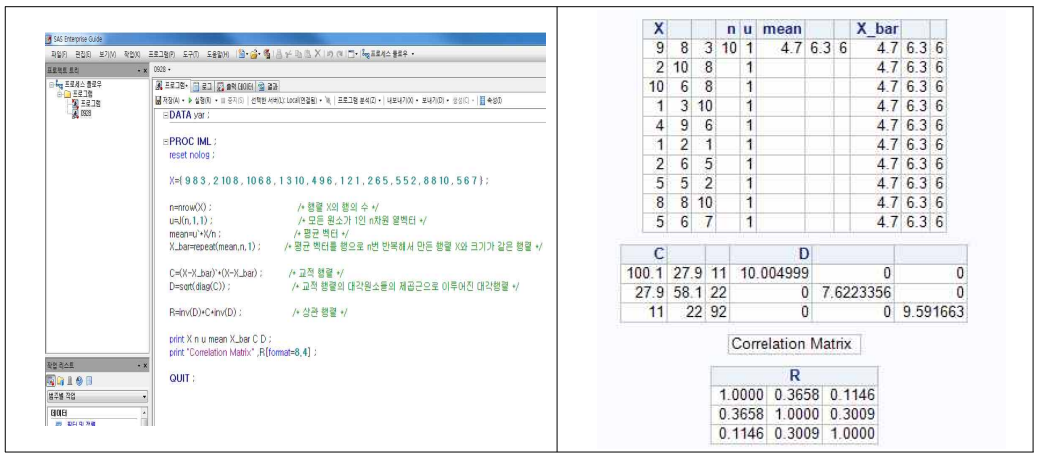
### 4. 대화형 통계학 실습실의 장점

통계학에서 R의 활용은 학부나 대학원에서 다변량 통계학, 회귀분석, 분산분석 그리고 데이터마이닝과 같이 행렬대수(matrix algebra)에 대한 기초적인 지식이 요구되는 수업에서 매우 유용하다. 행렬연산은 크기가 작은 행렬이나 벡터의 연산의 경우 손으로 또는 계산기를 활용할 수는 있으나 크기가 큰 경우에는 불가능에 가깝다. 특히, 행렬의 교차곱(cross product)과 역행렬(inverse matrix) 계산이나 이를 활용하는 회귀분석모형 (regression analysis model) 또는 분산분석모형(ANOVA model)의 계수벡터를 추정하는 문제에서 R은 진가를 발휘한다. 기존의 통계패키지에서 이들 모형에 대한 계수벡터에 대한 추정을 해결하는 방법도 있지만 R을 활용하면 모형을 통해 계수벡터를 추정하는 알고리즘을 이해하고 이에 따라 단계별로 프로그램을 만들어 실습하게 된다. 이는 지루하고 복잡한 알고리즘을 계산을 통해 이해하고 R을 활용하는 동안 집중력도 높일 수 있는 이중의 장점이 있다. 더군다나 기존의 통계패키지 SAS나 Minitab에서도 행렬연산을 제공하는 SAS/IML또는 Matrix가 있지만 고액과 큰 용량의 기존 패키지의 시스템을 구입하고 설치해야 한다는 어려움이 있다. 이에 비해 R은 무료로 다운받아 쉽게 설치하는 과정을 거치므로 누구든지 접근이 용이하며 학부수준에서 매우 유용한 프로그램으로 여겨진다. 학부 및 대학원 수업에서 어느 특정한 알고리즘에 바탕을 둔 통계적 방법과 관련된 R 프로그램이 필요할 때, Google에서 통계적 방법론의 중심어만 입력해도 해당하는 R 코드를 쉽게 찾을 수 있다.

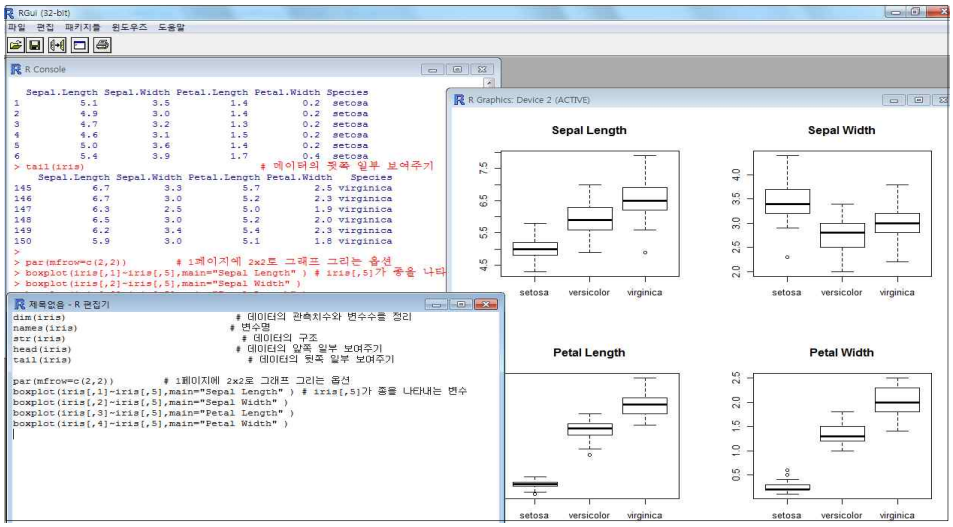



[그림 II-13] R에서 구현한 행렬연산

[그림 II-13]은 실제 행렬대수 과목의 실습 시간에서 입력행렬로 다변량 통계학에서 많이 활용되는 공분산행렬과 상관행렬 그리고 변동을 구하는 과제를 R을 활용한 프로그램의 실행과 그 결과를 보여주고 있다. 물론 Sage 셀을 활용해도 같은 결과를 얻을 수 있다. [그림 II-14]는 SAS/IML로 같은 결과를 얻는 과정이다. R 프로그램으로 실행한 [그림 II-13]보다는 [그림 II-14]의 프로그램을 작성하는 창, 로그, 출력 데이터, 결과 등 메뉴의 구성이 매우 복잡하다. 이는 SAS 뿐만 아니라 Minitab과 SPSS 등은 많은 통계적 기법을 수행 할 수 있는 다 기능 통계 패키지이기 때문이다. 특히, SAS의 경우 IML을 모듈로 추가해서 시스템 상에서 구현하는 것으로 쉽지 않은 SAS 프로그램 사용법과 문법을 전반적으로 습득해야 한다.




[그림 II-14] SAS/IML에서 구현한 행렬연산



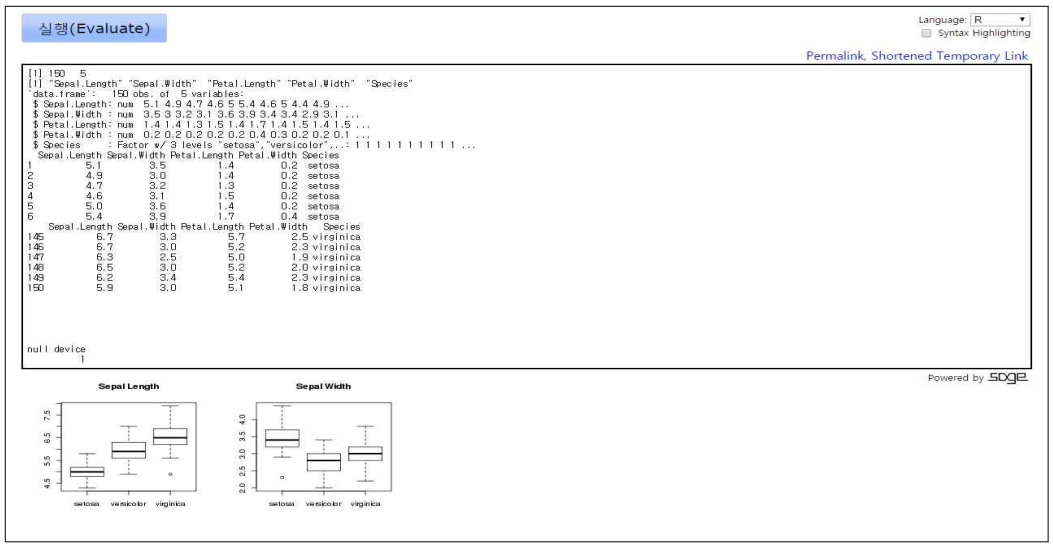

[그림 II-15] R에서 구현한 데이터 마이닝의 일부

[그림 II-15]와 [그림 II-16]에 데이터 마이닝(http://matrix.skku.ac.kr/2015-R-Statistics/DataMining.htm)과 관련된 실습 내용 중 일부를 R 프로그램과 Sage 셀에서 각각 구현한 결과가 주어져 있다. 특히, 이 자료는 영국의 통계학자 피셔(R.A. Fisher)의 판별분석(discriminant analysis)에 관한 그의 연구에서 인용된 것으로 다변량 통계학에서 매우 잘 알려진 붓꽃(iris flower)자료이다. 여기서 3품종(Setosa, Versicolor, Virginica) 각각 50포기씩 꽃받침길이와 폭, 꽃잎 길이와 폭을 측정한 것이다(최용석 · 정광모, 2003, 6장). 먼저 [그림 II-15] 은 R 프로그램에서 직접 구현한 것으로 크게 R Code 창, R 편집기 창, R Graphics 창으로 이루어져 있다. 프로그램은 R 코드를 편집기 창에서 작성하거나 복사해 와서 실행하면 Code 창에서 그 과정과 결과를 보여주며 특히, 그림에

관한 것은 Graphics 창에서 얻게 되며 각 품종별 4분위수를 기준으로 분포를 보여주는 상자그림(box plot)이 나타나 있다. 이 그림은 마우스 우측 버튼을 클릭하면 메타파일 형식으로 복사 및 저장할 수 있어 그림의 크기를 편집하고자 하는 문서에 맞게 조절할 수 있고 해상도가 좋다.

```
dim(iris)                                    # 데이터의 관측치수와 변수수를 정리
names(iris)                                  # 변수명
str(iris)                                    # 데이터의 구조
head(iris)                                   # 데이터의 앞쪽 일부 보여주기
tail(iris)                                   # 데이터의 뒷쪽 일부 보여주기

par(mfrow=c(2,2))                            # 1페이지에 2x2로 그래프 그리는 옵션
boxplot(iris[,1]~iris[,5],main="Sepal Length" ) # iris[,5]가 종을 나타내는 변수
boxplot(iris[,2]~iris[,5],main="Sepal Width" )
boxplot(iris[,3]~iris[,5],main="Petal Length" )
boxplot(iris[,4]~iris[,5],main="Petal Width" )
```




[그림 II-16] Sage 셀에서 구현 한 데이터 마이닝의 일부

반면에 [그림 Ⅱ-16]은 Sage 셀의 편집 창에서 R 코드를 작성하거나 복사한 후 우측의 설정된 언어(Language) R에 의해서 실행(Evaluate)을 하면 얻게 되는 결과이다. 특히, 그림에 대한 결과가 맨 하단에 나타나는 데 [그림 Ⅱ-15]와는 다르게 그림을 다른 문서에 편집하기 위해선 그림판을 이용해야 하고 해상도가 다소 떨어지는 결함이 있게 된다. 그러나 <표 Ⅱ-2>의 통계학 개론 수준의 강의라면 Sage 셀을 이용하는 것이 R 프로그램을 다운 받아 컴퓨터에 인스톨하고 그 사용법을 익혀야 하는 번거로움을 피할 수 있어 대화형 통계학 실습실을 구축하는 데 다소 장점이 있다고 여겨진다. 더군다나 만약에 하드 보안이 되어 있는 경우라면 매번 실습 때마다 R 프로그램을 다운 받아 인스톨해야 하는 어려움이 따른다. 다만 다변량 통계학이나 다차원척도법과 같은 교과목에서는 경우에 따라 3차원 그림을 회전해서 살펴 볼 경우가 있는 데 R 프로그램의 [그림 Ⅱ-15]에서는 마우스를 이용하여 매우 간단하게 작동할 수 있는 장점이 있다.

## Ⅲ. 결론

현대사회에서 많은 학생들이 모바일기기를 사용하는 시간이 늘어났다. 모바일 기기를 이용하여 수업자료를 내려 받고, 수업을 듣는다. 하지만 컴퓨터 프로그래밍을 사용하는 수업은 휴대용기기를 통해 사용 할 수 없었던 것이 학생들에게 가장 큰 불편함이었다. 유료 프로그램은 반드시 학교 컴퓨터를 사용해 공부를 해야만 했다. 이러한 문제점을 극복하기 위하여 개발된 통계실습실은 Sage와 R을 휴대폰에서도 자유롭게 사용할 수 있도록 구현했다. C++ 언어를 대체하면서 21세기 현재 가장 주목 받는 오픈소스 소프트웨어인 Python 언어도 사용가능하다. 정의와 정리에 대한 설명, 흥미를 유발시키는 참고자료, 코드를 사용한 다양한 예시를 통해 학생들이 휴대용 기기를 사용해 공부할 수 있는 걸어 다니는 강의실 및 학습실을 실제로 구현하였다. R 또는 Sage의 소용량 버전이 아니라 웹사이트에 연결만 하면 바로 실행할 수 있기에 휴대용 기기의 용량을 차지하지 않는다. 이에 더하여, Sage 뿐만 아니라 R 언어도 사용할 수 있도록 설정해 놓은 것이 큰 장점이다. 본 연구에서는 그렇게 개발된 통계학 대화형 실습실의 개발 과정과 그 활용을 다루었다.

본론에서 소개한 통계 입문 강좌용 모든 실습은 노트북 PC 뿐만 아니라, 스마트폰, 갤럭시 탭, 아이패드를 포함한 다양한 모바일 기기에서 가능하도록 디자인된 대화형 통계학 입문 실습실이다. 이는 현대 사회의 휴대용 기기를 사용이 증가하는 추세에 맞추어 언제 어디서든 쉽게 프로그래밍을 학습, 수정, 실행이 가능하다는 커다란 장점을 확인해 준다. 더구나 제공한 플랫폼은 다양한 컴퓨터 언어를 필요에 따라 설정할 수 있고, 다른 언어 또한 같은 플랫폼에서 무리 없이 적용될 수 있다는 점에서 잠재적 가능성도 크다. 또 예제에 적용하여 제시한 명령어를 따로 입력하지 않고 바로 클릭하면서 실습하고, R 명령어에 대한 이해를 다른 문제에 바로 적용하여 함수와 조건만 수정하여 사용하기 때문에 시간과 노력을 대폭 줄여 줄 뿐 아니라, 초보자에 통계학 입문 과목을 지도하기에 적절하다. 게다가 교수자가 강의 내용 및 필요에 따라 적절한 Java 도구와 이미지 및 통계 자료를 포함시켜 맞춤형 실습실을 구성하면 실제 학생들의 관심과 흥미를 유도할 수 있다. 이 실습실은 통계입문 강좌의 효과적인 실습실의 모델의 하나가 될 수 있다고 판단한다.

## 감사의 글:

원고의 초안을 검토해주신 우석대 안승철, 전남대 백장선 교수님과 심사과정에서 귀한 조언을 보태주신 익명의 심사자 및 본 연구결과를 활용한 수학 사이버실습실 구축을 지원해 주신 한국방송통신대학교 프라임칼리지 담당자에게 감사드립니다.

## 참 고 문 헌

고래영 · 김덕선 · 박진영 · 이상구 (2009). 모바일 환경에서의 Sage-Math의 개발과 선형대수학에서의 활용, 한국수학교육학회지 시리즈 E <수학교육 논문집>, **23(4)**, 1023-1041.

Ko, R.-Y., Kim, D.-S., Bak, J.-Y. & Lee, S.-G. (2009). Development of Mobile Sage-math and its use in Linear Algebra, *J. Korea Soc. Math. Ed. Ser. E: Communications of Mathematical Education*, **23(4)**, 1023-1041.

김경원 · 이상구 (2013). 모바일 선형대수학 스마트폰 콘텐츠 개발과 활용, 한국수학교육학회지 시리즈 E <수학교육 논문집>, **27(2)**, 121-134.

Kim, K.-W. & Lee, S.-G. (2013). Development of smart-phone contents for mobile linear algebra, *J. Korea Soc. Math. Ed. Ser. E: Communications of Mathematical Education*, **27(2)**, 121-134.

이상구 · 신준국 · 김경원 (2014). 스토리텔링 수학 교과서에서 공학적 도구의 활용과 미분적분학 단원에 대한 개발 사례, 한국수학교육학회지 시리즈 E <수학교육 논문집>, **28(1)**, 65-79.

Lee, S.-G., Shin, J. & Kim, K.-W. (2014). A Case Study of Perceptions on Storytelling Mathematics Textbooks with Computer Algebra System, *J. Korea Soc. Math. Ed. Ser. E: Communications of Mathematical Education*, **28(1)**, 65-79.

이상구 · 이재화 · 김경원 (2014). [빅북] 선형대수학, 교보출판사.

Lee, S.-G., Lee, J. H. & Kim, K.-W. (2014). *[BigBook] Linear Algebra*, Kyobo Book.

http://matrix.skku.ac.kr/2015-Album/BigBook-LinearAlgebra-SGLee-New-2015.pdf

이상구 · 이재화 · 김덕선 (2012). 현대선형대수학 with Sage, 경문사.

Lee, S.-G., Lee, J. H. & Kim, D.-S. (2012). *Contemporary Linear Algebra with Sage*, KyungMoonSa.

이상구 · 장지은 · 김경원 (2013). Sage와 GeoGebra를 이용한 선형대수학 개념의 Visual-Dynamic 자료 개발과 활용, 한국수학교육학회지 시리즈 E <수학교육 논문집>, **27(1)**, 1-17.

Lee, S.-G., Jang, J.-E. & Kim, K.-W. (2013). Visualization of Linear Algebra concepts with Sage and GeoGebra, *J. Korea Soc. Math. Ed. Ser. E: Communications of Mathematical Education*, **27(1)**, 1-17.

최용석 · 정광모 (2003). SAS를 활용한 다변량 분석 기법과 응용, 자유아카데미.

Choi, Y,-S. & Jeong, K.M(2003). *Methods and Applications of Multivariate Analysis Using SAS*, Free Academy.

최용석 (2014a). [빅북] R과 함께하는 통계학의 이해, 교보출판사.

Choi, Y.-S. (2014a). *[BigBook] Introduction to Statistics with R*, Kyobo Book.

최용석 (2014b). 다차원척도법의 산책, 자유아카데미.

Choi, Y.-S. (2014b). *Walk in Multidimensional Scaling*, Free Academy.

허명회 (2014). 응용데이터분석, 자유아카데미.

Huh, M.-H. (2014). *Applied Data Analysis Using R*, Free Academy.

Bloomfield, V. A. (2014). *Using R for Numerical Analysis in Science and Engineering*, Chapman & Hall/CRC.

Kim, K.-W., Lee, S.-G. & Sun, S. (2013). Modeling of Mobile Sage and Graphing Calculator, *Journal of Modern Education Review*, **3(12)**, 918-925.

Lee, S.-G., Kim, E.-K., Ham, Y., Kumar, A., Beezer. R., Vu, Q.-P., Simon, L. & Hwang, S.-G. (2014). *Calculus with Sage*, KyungMoonSa. http://matrix.skku.ac.kr/Cal-Book

Lee, S.-G., Kim, K.-W. & Lee, J. H. (2013). Sage matrix calculator and full Sage contents for linear algebra, *Korean J. Math.*, **21(4)**, 503-521.

Stowell, S. (2014). *Using R for Statistics*. Apress.

# Interactive Statistics Laboratory using R and Sage

**Sang-Gu Lee**†
Department of Mathematics, Sungkyunkwan University, Suwon 440-746, Korea
E-mail : sglee@skku.edu

**Geung-Hee Lee**
Department of Statistics, Korea National Open University, Seoul 110-791, Korea
E-mail : geunghee@knou.ac.kr

**Yong-Seok Choi**
Department of Statistics, Pusan National University, Busan 609-735, Korea
E-mail : yschoi@pusan.ac.kr

**Jae Hwa Lee**
Department of Mathematics, Hallym University, Chuncheon 200-702, Korea
E-mail : jhlee2chn@hallym.ac.kr

**Jenny Jyoung Lee**
Graduate School of Public Health, Seoul National University, Seoul 151-742, Korea
E-mail : jenny.lee@yale.edu

In this paper, we introduce development process and application of a simple and effective model of a statistics laboratory using open source software R, one of leading language and environment for statistical computing and graphics. This model consists of HTML files, including Sage cells, video lectures and enough internet resources. Users do not have to install statistical softwares to run their code. Clicking 'evaluate' button in the web page displays the result that is calculated through cloud-computing environment. Hence, with any type of mobile equipment and internet, learners can freely practice statistical concepts and theorems via various examples with sample R (or Sage) codes which were given, while instructors can easily design and modify it for his/her lectures, only gathering many existing resources and editing HTML file. This will be a resonable model of laboratory for studying statistics. This model with bunch of provided materials will reduce the time and effort needed for R-beginners to be acquainted with and understand R language and also stimulate beginners' interest in statistics. We introduce this interactive statistical laboratory as an useful model for beginners to learn basic statistical concepts and R.

* ZDM Classification : M85, G45, M15
* 2000 Mathematics Subject Classification : 97C80, 97U70
* Key Words : Statistics, Laboratory, Model, Sage, R
† Corresponding author