

응용논문

기초통계교육을 위한 통계소프트웨어의 개발*

-Excel에 기초한-

조신섭 · 송문섭 · 이윤모 · 성병찬 · 윤영주 · 이현부

서울대학교 자연과학대학 통계학과

Development of Statistical Software for the Education of Statistics at the Introductory Level based on Excel

Sinsup Cho · Moonsep song · Yoonmo Lee · Byeongchan Seong ·

Youngjoo Yoon · Hyunbu Lee

Dept. of Statistics, Seoul National University

Abstract

In this paper we compare several statistical packages and propose basic requirements of the package for the efficient education of statistics. We develop a statistical software, KESS, based on Excel. KESS provides all the input menus and output results in Korean.

1. 서론

통계에 대한 인식이 확산되기 시작하면서 통계자료 분석은 이제 거의 모든 분야에서 아주 중요한 위상을 차지하고 있다. 이에 부응하여 대학에서의 기초통계교육의 중요성이 더욱 강조되고 있다. 효율적인 통계교육을 위해서는 이에 적합한 통계패키지를 이용하여 자료분석을 하는 것이 필수적이므로 현재 대부분의 통계교육은 MINITAB과 SAS, SPSS, S-PLUS 등의 통계패키지들을 이용한 실습이 이론통계교

* 본 연구는 1998년 서울대학교 발전기금(포철)학술연구비 지원에 의해 이루어진 것임.

육과 병행되어 이루어지고 있는 실정이다. 그러나 기초 통계교육에 주로 이용되어 온 MINITAB 또는 S-PLUS는 일반 사회에서는 별로 이용이 되지 않는 관계로 대학에서 배운 내용을 실제 사회업무에 적용하기 위해서는 또 다른 자료분석용 패키지를 숙지해야 한다는 문제점을 지니고 있다. 이에 비해 SAS와 SPSS는 범용성을 지니고 있어 교육뿐만 아니라 일반업무에도 광범위하게 활용이 되고 있으며, 따라서 이들 패키지는 한번 숙지하면 또 다른 통계패키지를 배워야 하는 문제점은 별로 없다고 할 수 있다. 그러나 범용성을 강조한 나머지 프로그램 자체가 너무 커지고, 이의 효율적인 사용을 위해선 상당한 수준의 통계이론을 필요로 하며 가격 또한 비싸진 관계로 일반 사회에서는 사용을 기피하는 경향이 있다.

특히 최근에 들어서는 많은 분야에서 기본적인 통계처리를 자체적으로 해결하려는 움직임이 확산되고 있으며 이를 위해 현재 업무에서 가장 많이 이용되고 있는 Excel을 자료분석에 이용하고 있다. 그러나 Excel은 본래 회계처리를 위한 Spreadsheet의 기능을 중심으로 만들어진 프로그램인 관계로 비록 여러 가지 자료처리기능을 가지고 있다고는 하나 통계교육을 담당하는 전문가의 입장에서 볼 때는 이론적으로 미흡하고 분석기능이 부족하다는 문제점을 지니고 있다.

물론 통계적 자료 분석을 위해서 사용되고 있는 통계 소프트웨어에는 이들 외에도 더 많은 종류가 출시되어 있다. 특정 통계적 방법만을 지원하는 것에서부터 모든 통계적 기법들을 총망라하는 대규모 패키지에 이르기까지 아주 다양하다. 그러나 대부분의 패키지들이 사용이 어렵고 분석화면과 결과화면이 한글화되어 있지 않은 점등은 일반 비전문가들이 사용하는데 있어서 가장 큰 문제점으로 지적되고 있다. 더 나아가 이러한 문제점이 일반인들로 하여금 통계란 어려운 학문이라는 인식을 주어 통계로부터 멀어지게 하여 과학적 사고에서 차지하는 통계의 중요성을 인식하지 못하게 할 우려가 있다. 따라서 통계적 분석을 보다 쉽게 일반인들이 받아들일도록 하는 것이 통계인의 중요한 역할이 될 것이다.

통계패키지의 사용과 관련된 국내에서의 논문으로는 통계패키지의 선택과 관련된 연구로 김병천(1987)과 허명희와 정진환(1990), 김성수와 한경수(1993), 김철웅과 서한손(1993), 성내경(1993), 조신섭 외(1994), 조신섭과 신봉섭(1997)을 들 수 있다.

한글화된 통계패키지를 만들려는 노력으로는 이정진 외(1992)와 이정진과 강근석(1994), 이정진 외(1995), 조신섭 외(1996), 유종영 외(1997)을 들 수 있다. 이외에도 효율적인 통계교육을 위한 컴퓨터의 활용과 관련된 논문으로 한경수와 안정용(1994)과 신봉섭(1995), 한경수와 안정용(1996), 안기수와 허문열(1996), 이태림 외(1997)를 들 수 있다. 특히, 한경수와 안정용(1996)은 전통적인 강의 방식에서 탈피하여 보조적인 교육용 소프트웨어를 사용하고 있는 국내외의 사례를 들고 있다. 이외에도 허문열(1991)과 전홍석(1992), 강근석(1993)은 통계계산과 관련된 연구 결과를 발표하고 있다.

본 논문에서는 통계전문가들을 위한 실험도구의 기능을 하는 패키지보다는 일반 사용자들이 보다 손쉽게 접근할 수 있고 기초통계교육에 적합한 패키지가 지녀야 할 특성에 대해 알아보았다. 현재 많이 사용되고 있는 MINITAB과 SAS, SPSS, S-PLUS

에서의 여러 가지 기능 및 사용되는 수식, 알고리즘들을 비교한 후 기초통계교육에 적합한 장점들을 취합하였으며 이를 토대로 하여 한글로 구현된 새로운 형태의 기초통계교육용 패키지를 개발하였다. 개발에 사용된 도구는 상용화되어 사용중인 Spreadsheet 제품군중 가장 범용성 있고, 광범위하게 사용중인 마이크로소프트사의 Excel을 기반으로 하며, 패키지의 제어와 계산 함수의 사용 등을 위해서는 VBA (visual basic for application)를 사용하였다.

기존의 통계분석용 패키지들과는 달리 사용자들에게 아주 친숙한 Excel을 기반으로 함으로써 통계에 대한 인식을 확산시킬 수 있는 계기가 될 뿐만 아니라, 통계를 연구하는 많은 통계전문가들에게도 자신들의 새로운 기법을 보다 쉽게 일반 사용자들에게 소개할 수 있는 기회가 될 것으로 기대한다. 외국에서의 Excel을 이용한 통계패키지의 개발과 관련된 내용을 알아보려면 Internet 상의 <http://www.unistat.com>을 활용할 수 있다.

2. 교육용패키지의 특성

기초통계교육에서 일반적으로 다루어지고 있는 내용과 분석에 필요한 통계량 및 그림들을 항목별로 분류해 보면 다음과 같다.

(1) 기초통계분석

평균과 중앙값, 최빈값 등의 위치와 관련된 대표값

분산과 범위 등의 산포와 관련된 통계량 및 왜도와 첨도 등의 기초통계량

정규성검정을 위한 통계량 및 정규확률그림

(2) 탐색적 자료의 분석

히스토그램과 상자그림, 줄기-잎-그림, 산점도 등의 그림 기능

IQR 등의 탐색적 자료분석을 위한 통계량

(3) 통계학습

이항분포와 *t*-분포, χ^2 -분포, *F*-분포 등의 표본분포

중심극한정리와 *t*-통계량의 경험분포의 모의실험

(4) 모집단에 대한 추론

일표본과 이표본의 모수의 비교를 위한 *t*-검정과 *F*-검정

일표본과 이표본의 모비율검정

(5) 분산분석

일원배치법과 이원배치법

(6) 범주형자료의 분석

분할표, 적합도검정, 독립성 또는 동일성검정

(7) 회귀분석

상관분석과 선형회귀모형 및 회귀진단을 위한 여러 가지 도구

이러한 통계적인 측면 이외에도 통계라는 학문이 지니고 있는 여러 가지 특징들을 잘 파악하고 그 특징들을 효과적으로 구현하기 위한 조건을 고려해야 할 것이다. 통계라는 학문이 자료와 밀접한 관련성을 맺고 있으며, 또한 그 자료를 다른 일반인들에게 쉽게 설명하는 도구라는 점을 감안할 때 통계패키지 개발 시에 다음과 같은 기본조건들을 고려하여야 할 것이다.

- (1) 자료파일을 불러오거나 자료를 직접 입력할 수 있어야 한다.
- (2) 다양한 차트 출력으로 사용자가 자료의 구조를 쉽게 파악할 수 있어야 한다.
- (3) 분석에 필요한 여러 수학, 통계함수의 구현이 쉬워야 한다.
- (4) 분석결과의 보고서화와 프리젠테이션을 쉽게 할 수 있어야 한다.

즉, 자료 파일을 불러오거나 자료를 직접 입력할 수 있어야 하며, 자료 파일을 불러오거나 입력하는 방법 외에 자료 파일의 크기가 대용량일 경우에는 외부 자료 파일을 응용 프로그램에 로딩시키지 않고 그대로 연결만 시켜서 사용하는 방법도 가능하게 해야 할 것이다. 이런 방법은 메모리 절약과 개발들의 수행 속도와 밀접한 관련을 맺고 있다.

개발에 사용된 Excel에서는 위의 3가지 방법이 모두 가능하여 자료의 직접 입력을 위해서는 워크시트를, 자료 파일을 읽기 위해서는 Excel의 오픈함수를 사용하였다. 또한 Excel은 운영체제에 설치된 ODBC라는 드라이버를 써서 외부자료파일과 정보를 주고받을 수 있으므로 자료파일의 규모가 아무리 커져도 그와 상관없이 파일을 이용할 수가 있다.

통계패키지에서, 시각화에 의한 자료의 설명을 위해서 다양한 차트의 출력은 상당히 큰 비중을 차지하고 있다. Excel은 이러한 통계적 요구를 잘 충족시킬 수 있다. 특히, Excel이 제공하고 있는 차트 마법사를 이용하면 다양한 2차원 차트와 3차원 차트를 아주 쉽게 그릴 수가 있다. 또한 VBA를 이용하면 차트 마법사에서 제공하는 차트 외에 다른 종류의 차트들도 매크로로 추가할 수가 있어 편리하다.

비용절약이라는 면은 프로그램 개발 시에 매우 중요한 요건으로서 DLL(dynamic linking library)이라는 기법의 사용은 비용절약 차원에서뿐만 아니라 여러 가지 이유에서 통계패키지 개발에 반드시 필요한 수단이다. 인터넷상에서나, 기존의 서적들을 통해서 많은 수학 및 통계함수들을 접할 수가 있는데, 이들은 대부분 FORTRAN, C, C++으로 개발되어 있다. 이런 코드들을 DLL을 이용하여 재활용한다면 프로그램 개발이 더욱 쉬울 것이다. 특히 Excel에서는 이런 DLL을 이용할 수가 있어서 많은 통계함수들을 다시 프로그램 할 필요가 없다. 또한 DLL은 FORTRAN, C, C++등으로 제작할 수 있으며 그 방법은 아주 간단하다.

아무리 좋은 통계량과 좋은 모델을 구했다 하더라도 그 분석 결과를 일반 비전문가들에게 효과적으로 설명하지 못한다면, 통계의 역할은 축소될 수밖에 없을 것이다. 컴

퓨터의 발달과 더불어 일반인들에게 시각적 요소가 중요해지면서 분석 결과의 보고서화와 프리젠테이션은 아주 중요한 위치에 있게 되었다. Excel은 워드프로세서인 MS-WORD, 프리젠테이션을 효과적으로 할 수 있게 해 주는 도구인 MS-POWERPOINT와 쉽게 상호 호환이 된다. 따라서 OLE기법을 사용한다면 분석결과와 보고서화와 프리젠테이션화는 아주 간편하게 이루어질 수가 있다.

3. KESS의 특징 및 다른 통계패키지의 비교

기초통계교육에 Excel을 이용한 예로는 조신섭 외(1998)와 안윤기 외(1998)와 신봉섭(1998)을 들 수 있다. 이들은 Excel의 도구메뉴 중 [데이터 분석]에서 제공되는 분석기능을 이용하여 자료분석을 하는 방법에 대해 설명하고 있으나 Excel에서 제공되는 기능들만으로는 통계교육에 부족한 면이 있다. 또한 출력되는 결과와 통계량들의 계산방법이 명확하지 않고 특히 회귀분석의 경우에는 표준잔차의 출력값이 다른 패키지들에서 출력되는 값들과 비교해 볼 때 무엇을 의미하는지 알 수 없다. 이러한 문제점들을 해결하기 위해 송문섭과 조신섭(1999)은 Excel을 기본 도구로 사용하여 전문가가 아닌 일반인들이 사용하기 쉽고, 기초통계교육에 유용한 통계패키지인 KESS (Korean Educational Statistics Software)를 개발하였다. KESS에서는 Excel 자체에서 제공되는 분석기능을 사용하지 않고 VBA 등으로 작성된 프로그램을 연결하여 사용하고 있다[Jacobson, 1997].

사용자의 편의를 위해 <그림 1>과 같이 원래의 엑셀메뉴에 [통계분석] 메뉴를 추가하여 풀-다운 메뉴 형태로 원하는 항목을 선택하여 이용할 수 있도록 하였다.

나이	수속월	이관월	통해수	수속월
1	42	110	96	291
2	53	130	72	278
3	53	120	90	342
4	48	120	80	239
5	53	118	74	243
6	58	122	72	210
7	48	130	90	219
8	60	124	80	203
9	59	180	100	289
10	40	120	80	185
11	58	115	80	260
12	58	140	90	312
13	64	135	95	185
14	57	110	78	282
15	32	112	70	254
16	59	140	90	303
17	48	130	80	271
18	47	115	84	304
19	47	130	80	334
20	28	120	88	322
21	37	95	55	226
22	54	141	100	383
23	36	130	80	399

< 그림 1 > KESS의 기본 메뉴화면

통계분석의 각 항목별로 KESS와 다른 통계패키지들의 기능 및 차이점을 설명하기로 한다.

3.1 기초통계분석 및 탐색적 자료분석

기초통계분석, 탐색적 자료 분석에 사용되는 통계량들이 각 패키지에서 어떠한 알고리즘과 수식에 의해 구현되는지를 정리 비교하였다. Excel이 제공하는 여러 가지 통계량들이 교육에 가장 적합한 것은 아닐 수 있으나 KESS의 경우에는 가능하다면 Excel이 제공하는 통계량들을 이용하였으며 필요한 경우에는 도움말 기능을 이용하거나 새로운 알고리즘을 개발하여 적합한 통계량을 구현하고자 노력하였다. <표 1>에는 기초통계분석, 탐색적 자료 분석에 사용되는 통계량들이 각 패키지에서 어떠한 알고리즘에 의해 구현되는지가 정리되어 있다.

< 표 1 > 기초통계분석 및 탐색적자료분석에 사용되는 통계량들의 비교

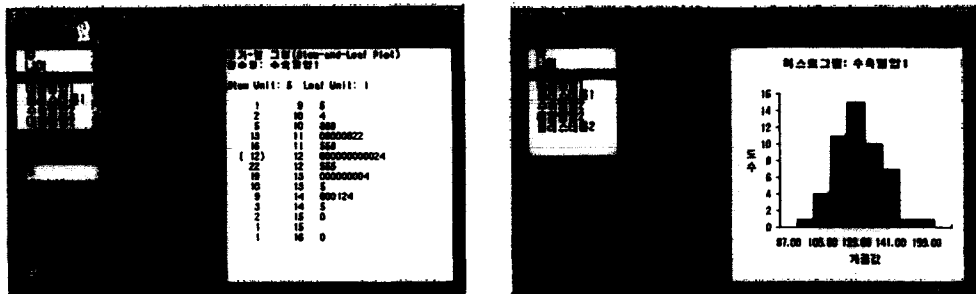
	SAS	SPSS	MINITAB	KESS
정규성 검정법	Kolmogorov-Smirnov Shapiro-Wilks	Kolmogorov-Smirnov Shapiro-Wilks	Anderson-Darling Ryan-Jonier Kolmogorov-Smirnov	Shapiro-Wilks
통계량	Default : 표본수, 평균, 표준편차, 최대값, 최소값 Option : 분산, 평균의 표준 오차, 합계, 변동 계수, 왜도, 첨도, USS, CSS, Q1, Q3, Med Range, 최빈값, Q3-Q1, Extremes 1)	Default: 평균, 표준 편차, 최소값, 최대값 Option: 합계, 분산 범위, 평균의 표준오차, 첨도, 왜도	Default : 표본수, 평균, 중앙값, 절사평균, 표준편차, 평균의 표준 오차, 최대, 최소, Q1, Q3 Option: Store Menu에 있는 통계량들	기초통계량: 평균, 합계, 개수, 중앙값, 표준편차, 평균의 표준오차 추가통계량: 최빈값, 분산, 사분위수 범위, 범위, 변동계수, 왜도, 첨도, 정규성 검정
백분위수	Hinge	$\frac{q}{100} (n+1)th$	$\frac{q}{100} (n+1)th$	$\frac{q}{100} (n-1) + 1th$
t-검정	일표본, 이표본, 대응표본	일표, 이표본, 대응표본	일표본, 이표본, 대응표 본 2)	일표본, 이표본, 대응표본 2)
그림	산점도, 히스토그램 줄기-잎-그림, 상자그림 3)	산점본도, 히스토그램 줄기-잎-그림, 상자그림	산점도, 히스토그램 줄기-잎-그림, 상자그림	산점도, 히스토그램 줄기-잎-그림, 상자그림 정규확률그림
비고	SAS 6.12	SPSS 7.5	Minitab 12.1(demo)	KESS(version 1.0)

- 1) SAS의 경우는 Means 절차에 의한 출력결과이다. Univariate 절차에서는 대부분의 기초통계량과 일표본 t-검정 및 비모수검정을 시행하고 그 결과를 출력한다.
- 2) 이표본 t-검정의 경우 MINITAB과 KESS는 수준별로 분류가 되어 있지 않은 자료인 경우에도 독립표본의 검정을 시행할 수 있다.
- 3) SAS의 경우는 Univariate 절차에서 줄기-잎-그림, 상자그림 및 정규확률그림을 제공한다.

비교에 사용된 모든 패키지들이 교육에 필요한 대부분의 기초통계량들과 t -검정 및 기본적인 그림들을 제공하고 있다. 그러나 패키지들이 동일한 분석 또는 통계량의 계산을 위해 다른 알고리즘 또는 수식을 사용하고 있다는 점에 유의하여야 할 것이다. 예를 들어, 사분위수인 Q_1 과 Q_3 을 계산할 때에도 SPSS와 MINITAB은 $(a/100) \times (n+1)$ 번째 관측값을 보간법으로 계산하고, Excel은 $(a/100) \times (n-1) + 1$ 번째 관측값을 계산하며, SAS의 경우에는 hinge를 구하는 식을 이용하여 구하는 등의 차이점이 있다. 사용자들의 입장에서는 패키지가 어떠한 알고리즘 또는 공식을 사용하고 있는지를 명확하게 알고 있어야만 해석 시에 혼란이 없을 것이므로, 도움말 등을 통해 이를 명확하게 밝혀야 할 것이다.

대부분의 통계패키지들이 도움말을 제공하고 있으나 한글화가 되어있지 않거나(예: SAS와 MINITAB), 번역이 충실하지 않아(예: SPSS), 사용자들에게 어려움을 주고 있음을 고려하여 KESS에서는 가능하면 개발에 사용된 이론, 공식 및 알고리즘들을 도움말을 통해 설명하고 있으며 또한 화면을 이용하여 메뉴의 사용법을 상세하게 설명하려고 노력하였다.

탐색적 자료분석 단계에서 매우 중요한 “그림”의 경우 KESS에서는 [기술통계] 메뉴에서 여러 가지 통계량들과 더불어 원하는 그림을 출력할 수 있도록 하였으며, “그림”만을 원하는 경우에는 [그래프] 메뉴에서 원하는 그림들만을 단독으로 선택하여 구현할 수 있도록 하였다. 특히 줄기-잎 그림과 히스토그램의 경우에는 <그림 2>에서 보는 바와 같이 오른쪽에 미리보기 화면을 제공하여 최종적으로 그림을 출력하기 전에 원하는 형태인지를 확인할 수 있도록 하였다. 만일 구현된 줄기-잎 그림이나 히스토그램이 적절하지 않은 경우에는 “줄기의 단위” 또는 “계급의 개수” 단추를 이용하여 단위와 개수를 선택할 수 있도록 하였다.

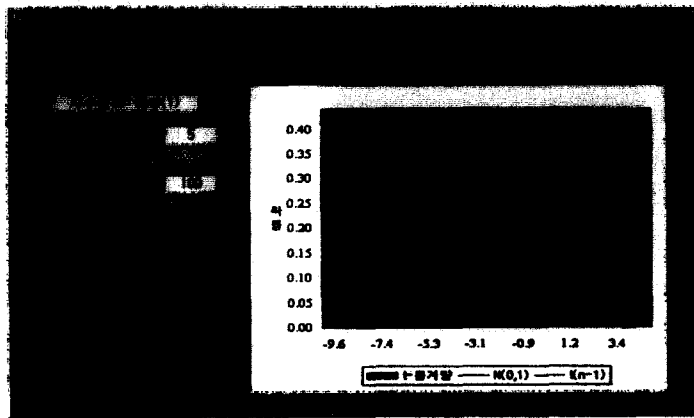


< 그림 2 > 줄기-잎 그림과 히스토그램 대화상자

3.2 통계학습

통계교육의 관점에서 볼 때 KESS와 기존의 패키지들과의 두드러진 차이는 중심극한 정리와 t -통계량의 경험분포를 모의실험을 통해 이해할 수 있도록 하였으며 자유

도의 변화에 따라 여러 가지 표본 분포의 형태가 어떻게 달라지는 지를 쉽게 확인할 수 있도록 한 점이다. 기초통계교육에서 이해하기 가장 어려운 내용 중의 하나가 중심극한정리이다. 수식으로만 설명하면 이해하기 어려운 중심극한정리를 쉽게 이해할 수 있도록 균등분포, 지수분포와 정규분포 등의 모집단에서 추출된 표본들의 평균의 경험분포가 표본의 크기 n 과 실험의 횟수자료가 변함에 따라 어떻게 정규분포의 형태로 접근하는지를 쉽게 볼 수 있도록 하였다. 또한 t -통계량의 경험분포가 표본의 크기, 실험의 횟수 및 모집단에 따라 얼마나 빨리 t -분포에 근사되는 지를 <그림 3>에서와 같이 모의실험을 통하여 그림으로 확인할 수 있도록 하였다. 즉, 표본크기가 작고 모집단이 정규분포가 아닌 경우에 t -통계량의 분포는 표준정규분포보다는 t -분포에 더 잘 근사됨을 확인할 수 있다.



< 그림 3 > t -통계량의 경험분포 대화상자

3.3 분산분석과 범주형 자료분석

<표 2>는 여러 통계패키지에서 제공하는 분산분석 기능 및 출력되는 내용들을 비교 정리한 것이다. 분산분석의 경우 SAS와 SPSS, MINITAB에서는 자료를 입력할 때 반응변수의 값과 함께 각 관측값이 속한 처리 또는 범주를 나타내는 요인변수를 입력하는 형식을 사용하고 있다. 그러나 대부분의 기초통계학 교재에서는 이들 분석에 사용되는 자료들이 분할표의 형태이므로 패키지에서 입력하는 자료와 다른 형태라는 점이 사용자의 입장에서 이해하기 어려운 점일 것이다. 분산분석의 경우 KESS에서는 두 가지 형태로 자료의 입력이 가능하도록 하였다. 먼저 “표준입력” 방식을 이용하여 <그림 4>와 같이 통계학 교재에서 사용하는 자료의 형태로 입력이 가능하도록 하였으며, 다른 패키지의 사용이 익숙하거나 다른 패키지와의 호환을 고려하여 “고급입력” 방식에 의해 <그림 5>와 같이 자료를 입력할 수 있도록 하였다. 자료분석 측면에서 볼 때, SAS와 SPSS, MINITAB은 다양한 형태의 모형을 분석할 수 있는 기능을 제공하고 있으나 기초통계교육에서는 일원배치법과 이원배치법 만을 다루고 있

다. 따라서 KESS에서는 최소한의 기능으로 이들 두 모형만을 처리할 수 있도록 하였다. 출력의 경우에도 SAS와 SPSS, MINITAB은 요인들의 조합별로 기술통계량을 제공하고 있으나 KESS에서는 역시 분할표의 형태로 기술통계량들을 출력하여 한눈에 파악하기 쉽도록 하였다.

< 표 2 > 분산분석방법의 비교

	SAS	SPSS	MINITAB	KESS
입력	요인변수와 반응변수를 입력	요인변수와 반응변수를 입력	요인변수와 반응변수를 입력	표 형태로 입력
분석기능	일원배치법 이원배치법 다원배치법 가능	일원배치법 이원배치법 다원배치법 가능	일원배치법 이원배치법 다원배치법 가능	일원배치법 이원배치법 (반복수가 같은 경우)
기술통계	각 조합별로 출력 (관측수, 평균, 표준편차)	각 조합별로 출력 (관측수, 평균)	각 조합별로 출력 (관측수, 평균)	분할표 형태의 출력 (관측수, 평균, 표준편차)
그래프	따로 그려야 함	따로 그려야 함	일원배치의 경우 평균과 신뢰구간이 수준별로 표시됨	일원배치의 경우 상자그림, 이원배치의 경우 수준별로 평균을 표시
비고	SAS 6.12	SPSS 7.5	Minitab 12.1(demo)	KESS(version 1.0)

	A	B	C
지역1	42.8	52.3	48.2
지역2	38.6	43.5	40.3
지역3	50.2	58.7	53.5
지역4	48.2	50.8	51.2

< 그림 4 > 이원배치법에서의 자료의 "표준입력"화면

요인변수1	요인변수2	분석변수
지역1	A	42.8
지역1	B	52.3
지역1	C	48.2
지역2	A	38.6
지역2	B	43.5
지역2	C	40.3
지역3	A	50.2
지역3	B	58.7
지역3	C	53.5
지역4	A	48.2
지역4	B	50.8
지역4	C	51.2

< 그림 5 > 이원배치법에서의 자료의 "고급입력"화면

범주형 자료분석의 경우, <표 3>에서 보는 것처럼 SAS와 SPSS, MINITAB에서는 자료를 입력할 때 범주형 변수와 도수를 입력하여야 한다. 그러나 대부분의 기초통계 교재에서는 자료가 분할표의 형태로 제공되므로 KESS에서는 <그림 4>와 같이 분할표의 형태로 입력을 하도록 하였다. 출력 통계량의 경우, 기대값 이외에도 행비율과 열비율, 전체비율이 제공되고 있으나 이러한 비율들이 오히려 표를 복잡하게 만들어 해석에 도움이 되지 않는다고 판단되어 KESS에서는 제공하지 않는다.

독립성검정과 동질성검정, 적합도검정에서는 모든 경우에 카이제곱검정통계량과 유의확률이 제공되고 있다. SAS와 SPSS의 경우에는 추가로 연관성의 측도가 제공되고 있으나 기초통계교육에서는 일반적으로 다루지 않는 내용이므로 KESS에서는 제공하지 않는다. 기초통계에 관한 대부분의 교재들이 적합도검정을 다루고 있으나 SAS와 SPSS, MINITAB에서는 이런 기능이 제공되지 않는다. 그러나 KESS에서는 <그림 6>과 같이 간단한 적합도검정을 시행할 수 있는 기능을 설치하였다

< 표 3 > 분할표에 의한 범주형자료의 분석 비교

	SAS	SPSS	MINITAB	KESS
입력	범주형변수와 도수를 입력	범주형변수와 도수를 입력	범주형변수와 도수를 입력	분할표의 형태로 입력
출력	기대값, 행비율, 열비율, 전체비율	기대값, 행비율, 열비율, 전체비율	기대값, 행비율, 열비율, 전체비율	기대값
검정통계량	카이제곱검정통계량 각종 연관성의 측도	카이제곱검정통계량 각종 연관성의 측도	카이제곱검정통계량	카이제곱검정통계량
비고	SAS 6.12	SPSS 7.5	Minitab 12.1(demo)	KESS(version 1.0)

통노	통능	모노	모능
315	108	101	32
9	3	3	1

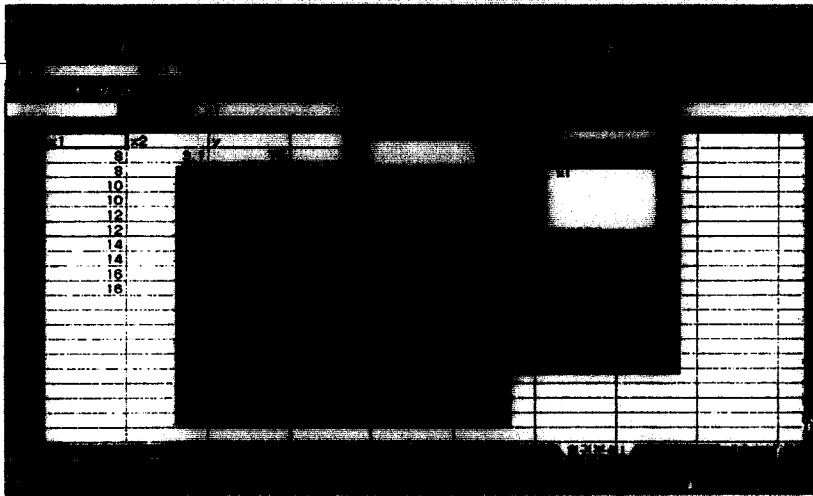
< 그림 6 > 적합도검정을 위한 자료의 입력화면

3.4 회귀분석

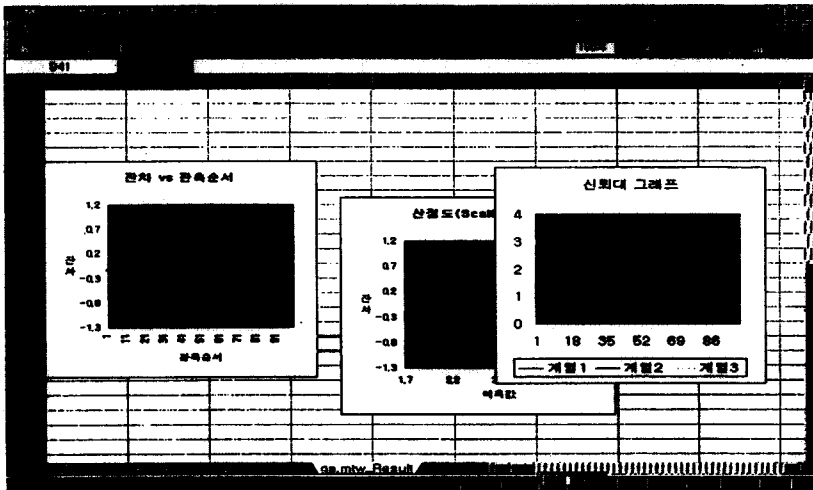
회귀분석의 경우에는 각 패키지마다 너무나도 다양한 형태의 기능과 출력물을 제공하고 있으나 기초통계교육에서는 선형회귀와 잔차분석 기능만을 제공하고 있다. 따라서 변수선택 및 회귀진단과 같은 기능은 KESS에서는 제공하지 않는다. 참고로 KESS의 출력으로는 <그림 7>과 같이 분석결과의 분산분석표와 모수의 추정결과를 출력하도록 하였다. 잔차분석을 위해서는 <그림 8>에서 보는 것처럼 여러 가지 그래프들을 사용자가 선택할 수 있도록 하였으며 이를 <그림 9>와 같이 출력하고 있다.

유형	자용도	제곱합	평균제곱	F	유의확률
회귀	2	7.51388	3.75694	14.92109	0
잔차	97	24.42272	0.25176		
계	99	31.9366			
Root MSE	0.50178				
관측계수	0.23527				
수정계수	0.2195				
변수명	수정치	표준오차	T	유의확률	
CONST	0.4708	0.54327	0.86624	0.3985	
GENERAL	0.00358	0.000731	4.847621	0	
MATH	0.00016	0.00085	0.18509	0.85355	

< 그림 7 > 회귀분석 결과 출력화면



< 그림 8 > 잔차분석을 위한 그래프의 선택화면



< 그림 9 > 여러 가지 그래프의 출력결과

3.5 KESS의 구성

KESS는 Kess.xla(948 KB)와 Kess1.xla(296 KB), Kess2.xla(316 KB), Kess3.xla(474 KB), Kess.hlp(954 KB) 및 Data.xls(56 KB)로 구성되어 있다. 모든 구성파일은 동일한 디렉토리에 위치해야 한다. Kess.hlp는 도움말파일이며 Data.xls는 sample 데이터파일이다. KESS는 송문섭과 조신섭(1999)에 디스켓의 형태로 제공되고 있다.

4. 결론

본 논문에서 소개된 KESS는 범용 통계패키지는 아니며, 기초를 효율적으로 교육시키고 통계 비전문가가 통계 학습을 쉽게 할 수 있도록 개발된 패키지이다. 따라서 여러 가지 표본분포의 모양을 쉽게 그려볼 수도 있고, 중심극한정리와 t -통계량의 분포를 모의실험을 통하여 확인할 수 있도록 구성하였다. 특히 이와 같은 통계학습을 Excel에서 실행할 수 있도록 함으로써, 대학에서의 통계교육을 일반 사회업무와 연계시켰다는 장점을 갖고 있다. 이러한 시도는 기존의 통계패키지에서는 찾아보기 힘든 것으로서, 이와 같이 사회업무와 연계되는 통계패키지의 개발 및 통계교육은 앞으로도 계속 시도되어야 할 과제라고 생각된다.

5. 참고문헌

- [1] 강근석(1993) “비선형 회귀분석을 위한 소프트웨어 NLIN90의 소개,” 「응용통계연구」, 6권 1호, pp. 163-172.
- [2] 김성수, 한경수(1993) “S-PLUS의 소개 및 SAS와의 그래픽 비교,” 「응용통계연구」, 6권 1호, pp. 1-12.
- [3] 김수화, 김승희, 조신섭(1994) “통계패키지에서의 시계열 분석방법의 비교연구,” 「한국통계학회논문집」, 1권 1호, pp. 119-130.
- [4] 김철용, 서한손(1993) “S-PLUS와 XLISP-STAT의 다이내믹그래픽 기능,” 「응용통계연구」, 6권 1호, pp. 23-28.
- [5] 성내경(1993) “SAS/GRAPH의 성능과 한계- S-PLUS의 기능과 대비하여,” 「응용통계연구」, 6권 1호, pp. 13-22.
- [6] 송문섭, 조신섭(1999) 「엑셀에 기초한 통계학 입문」, 자유아카데미.
- [7] 신봉섭(1995) “통계교육을 위한 MINITAB Macro,” 「한국통계학회 추계 학술발표회 논문집」, pp. 67-73.
- [8] 신봉섭(1998) 「엑셀 활용과 통계자료분석」, 교우사.
- [9] 안기수, 허문열(1996) “하이퍼미디어를 이용한 통계 학습 시스템(틀북과 XLISP-STAT을 이용하여),” 「한국통계학회 추계 학술발표회 논문집」, pp. 177-180.
- [10] 안기수, 허문열(1997) “인터넷을 이용한 통계 교육과 컨설팅의 현황,” 「한국통계학회 논문집」, 4권 2호, pp. 473-489.
- [11] 안윤기 외 9인(1998) 「Excel 97 데이터 처리」, 자유아카데미.
- [12] 유종영, 안기수, 허문열(1997) “동적그래픽스에 의한 회귀진단시스템(REDS)의 구현,” 「응용통계연구」, 10권 2호, pp. 241-251.
- [13] 이정진, 강근석(1994) “한국형 통계패키지 개발 연구,” 「응용통계연구」, 7권 2호, pp. 279-288.

- [14] 이정진, 강근석, 이윤오(1992) "통계학 교육용 한글 소프트웨어의 개발연구," 「응용통계연구」, 5권 1호, pp. 81-91.
- [15] 이정진, 강근석, 이윤오, 김지현, 이창수, 김성철(1995) "전문가용 한국형 통계패키지 개발연구 I," 「한국통계학회 논문집」, 2권 2호, pp. 434-444.
- [16] 이태림, 김성수, 강근석, 이정진(1997) "원격교육에 적합한 통계학 개론 코스웨어 개발연구," 한국통계학회 춘계 학술 논문 발표회.
- [17] 전홍석(1992) "통계계산에서의 갱신 알고리즘에 관한 연구," 「응용통계연구」, 5권 2호, pp. 283-291.
- [18] 조신섭, 김성수, 이기재(1998) 「통계적 수치해석」, 한국방송대학교출판부.
- [19] 조신섭, 성병찬, 강병국(1996) "Excel을 이용한 시계열 분석틀의 개발," 「한국통계학회 추계학술발표회 논문집」, pp. 187-188.
- [20] 조신섭, 신봉섭(1997) "통계적 공정관리를 위한 주요통계패키지의 비교," 「응용통계연구」, 10권 1호, pp. 29-36.
- [21] 한경수, 안정용(1994) "통계 기본 개념을 위한 통계소프트웨어 개발에 관하여," 한국통계학회 추계학술논문발표회.
- [22] 한경수, 안정용(1996) "저작도구를 이용한 통계교육용 소프트웨어 개발 연구-주사위 게임과 카드게임," 「응용통계연구」, 9권 1호, pp. 73-82.
- [23] 한경수, 안정용, 강윤비(1996) "인터넷을 이용한 통계교육," 「한국통계학회 추계학술발표회 논문집」, pp. 181-182.
- [24] 허문열(1991) "통계전문가 시스템의 특성과 개발도구의 선택," 「응용통계연구」, 4권 1호, pp. 85-92.
- [25] Jacobson R.(1997) *Microsoft Excel 97 Visual Basic Step by Step*, Microsoft Press. Redmond, W.A.