

Statistica Basic Language를 이용한 통계 교육

박 동 준 (부경대학교)

수학 통계용 그래픽 계산기와 통계 자료 분석용 소프트웨어를 활용하는 외국과 국내의 통계 교육의 추세를 소개한다. 그리고 통계 자료 분석용 소프트웨어인 Statistica의 특징을 요약한 다음 통계학을 효과적으로 교육하기 위하여 Statistica에서 제공되는 Statistica Basic Language로 프로그램한 세 가지 모듈을 제시한다. 프로그램을 입력하는 과정을 간단히 설명한 후 각 모듈을 소개한다. 처음 모듈은 숫자 값으로 표현되는 계량 자료에 대한 단순 통계량들을 구하기 위하여 분석하려는 원시 자료를 Spreadsheet에 입력한 다음 Statistica Basic Language로 프로그램한 모듈을 실행시킴으로써 한 번에 입력 자료에 대한 단순 통계량의 결과들을 볼 수 있도록 하였다. 둘째로 1에서 100까지 숫자들을 단순 랜덤 비복원 추출하는 과정을 프로그램한 모듈을 제시하였다. 마지막으로 프로그램한 모듈에서 제시되는 윈도우에 균일 분포에서 단순 랜덤 복원 추출하기 위한 표본의 크기와 샘플링 반복 횟수를 입력하면 표본 평균의 도수분포표와 도수분포도가 작성되어 표본 평균의 분포가 중심극한정리를 따르는가를 확인하였다.

I. 서론

기초 통계학 또는 통계학 개론으로 불리는 통계학을 처음 배우는 학부 생들에게 통계학에 대한 흥미를 유발시키고 통계 이론을 보다 효율적으로 설명하기 위하여 주로 계산기와 개인용 컴퓨터가 교육 보조 기구들로 사용되고 있다. 계산기로서는 통계학 강의시 휴대할 수 있을 뿐 아니라 간단한 명령문만 입력하면 손쉽게 예제들을 해결할 수 있는 수학 통계용 그래픽 계산기(TI-82, TI-83, TI-92 등)들이 있고 여러 종류의 통계 자료를 분석할 수 있는 다양한 형태의 소프트웨어들이 있다.

1. 외국의 통계 교육의 추세

Francis(1997)는 모수에 따라 변하는 이항 분포와 포아송 분포의 특성을 이해하고 확률을 계산하기 위해 TI-82 그래픽 계산기에 입력하는 간단한 프로그램을 소개했다. Short(1997)는 모분산의 동일성에 대한 F-검정과 모평균의 t-검정 등을 시행하기 위한 TI-83에 입력하는 프로그램을 제시하였고 TI-83과 함께 사용할 수 있는 Graph Link Software를 활용하면 TI-83과 개인용 컴퓨터를 연결하여 Texas Instrument Site (<ftp://ftp.ti.com/pub/grap-ti/cals-apps/83/>)에서 필요한 프로그램을 가져와서 TI-83에서 사용할 수 있다고 강조했다.

Trumbo(1994)는 강의용 또는 탐험적 자료 분석을 위하여 대수의 법칙, 확률의 dots적인 정의, 포아송 분포 등을 설명하기 위하여 Microsoft Quick Basic으로 기능키를 사용한 2단계의 메뉴 시스템 형태로 프로그램을 작성하였다. Laviolette(1994)는 회귀 분석의 모형설정, 랜덤오차, 회귀계수, 영향관측값, 이상점등에 대한 통계적 개념과 의미를 설명하기 위하여 통계 자료 분석용 소프트웨어인 Macintosh용 Systat 5.2의 사용 방법을 소개하였다. Ferris와 Hardaway(1994)는 통계학 교재와 강의의 내용을 담은 Teacher 2000이라는 멀티미디어 소프트웨어를 개발하였다. 이 소프트웨어를 실행시키면 화면이 세 부분으로 나뉘어지는데 첫째 화면에는 강사의 강의를 비디오로 진행되고 둘째 화면에는 교재의 내용, 그래프, 그림, 표, 동영상 등이 보여지고 마지막 화면에는 나중에 복습과 함께 설명이 되는 노트를 적을 수 있는 화면이 전개된다. Ferrall(1995)은 통계 소프트웨어인 Stata를 사용하여 몬테칼로 모의실험, 중심극한정리, 조건부 평균과 분산 등을 컴퓨터 사용자가 컴퓨터와 대화형식으로 사용할 수 있는 모듈을 개발하였다. Anderson & Dayton(1995)은 Xlisp-Stat 프로그램을 이용하여 회귀 표면과 추정 회귀 계수의 표본 분포를 연습할 수 있는 모듈을 작성하였다. Whigham(1998)은 Microsoft Excel에서 정규 분포의 모양을 그려내고 필요한 확률값을 계산하기 위해서 Spreadsheet 안에 입력하는 명령문들을 소개하고 다른 확률 분포에도 적절한 명령문을 사용하면 사용자가 편리하게 확률값을 계산할 수 있음을 강조했다.

미국의 저널 The American Statistician의 Statistical Computing Software Reviews Section에서는 기존의 통계 소프트웨어들에 관한 기술적인 부분들과 지원 가능한 자료 분석에 관한 부분들을 구체적으로 면밀히 비교 분석하고 있다. 뿐만 아니라 컴퓨터의 발달과 전자 통신 기술의 진전과 더불어 인터넷을 이용한 통계 교육의 연구도 진행중이다. 전산망을 통하여 인터넷에 접속한다면 통계교육과 관련된 여러 가지 정보를 찾아볼 수 있다. 예를 들면 미국의 카네기멜론대학의 통계학과의 홈페이지인 Statlib이라고 불리는 웹사이트(<http://lib.stat.cmu.edu>)에는 통계소프트웨어, 데이터셀 등을 비롯한 기타 통계 관련정보를 안내하고 있다.

2. 국내의 통계 교육의 추세

국내의 통계학 입문 과정의 일부 교재들(김우철·김재주·박병우·박성현·송문섭·이영조·전중우·조신섭(1998A, 1998B), 김동희·김충락·손건태·정광모·정윤식·최용석·홍창곤(1998), 김달호·박진표·이우동·이재만·장덕준·조교영·조길호·차영준(1995))은 주로 Minitab 또는 Sas등을 실습용 소프트웨어로 활용하여 각 단원의 끝 부분에 실습 예제를 제시하고 있다. 안기수·허문열(1997)은 통계와 관련된 자원을 지원하는 교육용 통계 서버와 통계 컨설팅을 위한 통계 서버들을 목적과 그래픽의 지원 여부에 따라 분류한 웹사이트의 주소들을 제공하였다.

국내의 통계 교육 관련 주요 학회나 연구회에서는 Mathematica, Excel, Xlisp-Stat 등에서 제공되는 여러 기능들을 이용하여 자체 제작한 모듈들을 활용한 연구가 진행 중이다. 외국에서 제작된 수많은 소프트웨어들 가운데 국내에서 흔히 사용되는 통계 자료 분석용 소프트웨어로서는 Sas, Spss, Minitab, Excel, Xlisp-Stat, Statgraphics, S-Plus 등이 있다. 전홍석·전성해(1997)는 저가이고 비주얼 베이직이란 언어를 사용한 Ms-Excel을 활용하여 실험 계획법의 자료 분석의 장점들을 다른 기존의 소프트웨어를 사용한 실험계획법의 자료 분석과 비교하였다. 신봉섭(1997)은 Mathematica의 장점인 애니메이션을 이용하여 확률 분포와 관련된 예제 위주의 문제들을 제시하였다. 그리고 신봉섭(1995)은 통계교육의 효과를 증대시키기 위하여 문제 중심의 Minitab Macro를 작성하였다. 안기수·허문열(1996)은 하이퍼미디어를 이용한 통계학습환경을 제안하고 회귀분석내용을 중심으로 한 시스템을 소개했다. 이태립·김성수·강근석·이정진(1997)은 통계학 보조수업 코스웨어를 개발하고 대학 수준에서의 코스웨어 효과를 분석하여 통계학 계열 과목들의 교육 소프트웨어 개발 연구에 대한 지표를 제시했다. 최숙희(1997)는 통계학의 기초 개념들을 쉽게 배울 수 있게 하기 위하여 유의성 검증시 필요한 p-값의 의미를 정확히 이해하고 적용할 수 있는 멀티미디어 프로그램을 개발하였다. 송문섭·조신섭·이윤모·성병찬·윤영주·이현부(1998)는 Excel을 이용하여 분산분석, 회귀분석, 범주형 자료 분석을 할 수 있는 한글로 처리한 소프트웨어를 개발하였다.

3. Statistica의 특징

Statistica의 특징을 간단히 살펴보면 Excel등과 같은 Spreadsheet를 활용한 통계 자료 분석용 소프트웨어이다. Statistica는 프리젠테이션용의 품질을 갖는 여러 가지 그래프를 제공하는 자료 분석 소프트웨어로서 사용자의 편리에 따라 수정이 가능한 자동기능 도구상자(Auto Task Toolbars)와 메뉴가 윈도우에 제공된다. Statistica내에서 활용 가능한 통계 자료 분석 기법으로서는 기초 통계량, 탐색적 자료분석, 비모수 자료분석, 분포 적합, 회귀 분석, 일반 비선형 추정, 단계별 로짓/프로빗 자료분석, 일반 공분산 분석/다변량 공분산 분석, 단계별 판별분석, 로그-선형분석, 확정적/탐험적 요인분석, 군집분석, 정준 상관 분석, 신뢰성 분석, 대응 분석, 생존 분석, 시계열 모형화, 구조 방정식 모형화 등이 있고 이들 분석 기법을 하나로 통합한 통계 자료 분석 모듈이다. Spreadsheet를 활용한 광범위한 자료 관리 기능을 갖고 있고 행렬 조작과 2차원, 3차원의 100여가지의 그래프지원이 가능할 뿐 아니라 수학적 그래프를 제공하는 프로그램처럼 데이터 없이 수식을 이용하여 그래프를 그릴 수 있고 그래프와 자료가 실시간(real time)으로 연결되어 데이터가 바뀌면 동시에 그래프가 바뀌어 그려진다. Statistica Basic Language와 같은 자체 프로그래밍 언어의 기능을 통합한 Auto

Task Button이라는 하나의 툴 박스를 만들어서 다른 통계 패키지에 비해 좀 더 많은 유연성을 가지고 있다. 외부 프로그램과의 인터페이스가 가능하며 Scrollsheet형태의 출력이 가능하며 글자와 그래프를 편집하여 함께 출력할 수 있다. Statistica는 수치의 정확도, 포괄성(통계 분석 과정의 다양성), 그래프의 우수성, 가격 등의 여러 가지 면에서 다른 소프트웨어(Spss, S-Plus, Systat, Statgraphics, Winstat, and Sigmastat)들과 비교 평가하여 그 우수성을 자랑하고 있다. Statistica의 웹주소(1997)를 참고하면 Statistica의 특징에 관한 보다 구체적인 내용을 살펴볼 수 있다.

본 소고에는 두 가지 계량 자료에 대한 단순 통계량과 산점도를 볼 수 있도록 Statistica Basic Language로 프로그램한 모듈을 제시하였고 단순 랜덤 비복원 추출하는 과정을 프로그램한 모듈과 균일 분포에서 추출된 표본 평균의 분포의 히스토그램을 그려내는 모듈을 제시하였다.

II. Statistica Basic Language 모듈

1. 두 변수의 단순 통계량

숫자 값으로 표현되는 계량자료에 대한 자료의 중심과 퍼짐에 관한 측도 들을 포함하는 단순 통계량들을 구하기 위하여 Statistica를 실행시킨 뒤 나타나는 Spreadsheet의 column 2 과 column 3에 분석하고자 하는 원시 자료를 <그림 1>과 같이 입력시킨다. 그 다음 Statistica의 메뉴에서 작성한 Statistica Basic Language의 프로그램을 불러서 메뉴의 "Execute"를 누르면 <그림 2>와 같이 자료 분석 여부를 묻는 메시지 상자가 나타나고 자료 분석을 위해서 "예(Y)" 단추를 누르면 변수의 합, 평균, 표준편차, 분산, 표준오차, 최대값, 최소값, 중위수, 백분위수 등을 포함하는 단순 통계량과 두 변수의 자료에 대한 산점도, 두 변수간의 상관계수가 <그림 3>과 같이 화면에 출력된다.

	SP	RATE	
1	male	52.0	68.0
2	male	65.0	56.0
3	female	61.0	65.0
4	female	72.0	60.0
5	male	69.0	72.0
6	female	73.0	82.0
7	male	58.0	60.0
8	female	88.0	93.0
9	male	80.0	73.0
10	female	98.0	92.0
11	male	58.0	60.0
12	male	62.0	58.0
13	male	69.0	65.0

<그림 1> Spreadsheet에 입력된 점수값

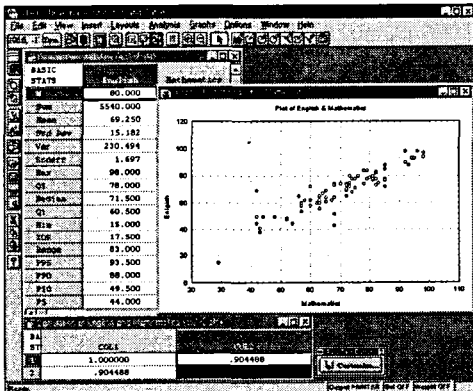
```

DisplayMessageBox (MB_YESNO+ MB_ICONQUESTION,
  "Descriptive Statistics: 'Would you like to analyze the data?' ) = idno then stop.

Dim result(17,2);
Dim quart(3);
Dim quart2(3);

begin
  result(1)=ValCount (Q2, 1, Ncases);
  ValSum (Q2, 1, Ncases, result(+1,1));
  ValMean (Q2, 1, Ncases, result(+2,1));
  ValSDDeviation (Q2, 1, Ncases, result(+3,1));
  ValVariance (Q2, 1, Ncases, result(+4,1));
  ValSEError (Q2, 1, Ncases, result(+5,1));
  ValMin (Q2, 1, Ncases, result(+6,1));
  ValMedian (Q2, 1, Ncases, result(+8,1));
  ValQuantiles (Q2, 1, Ncases, result(+9,1));
  result(+7,1)=quart(3);
  result(+8,1)=quart(1);
  ValMin (Q2, 1, Ncases, result(+10,1));
  result(+11,1)=quart(2)-quart(1);
  result(+12,1)=result(+6,1)-result(+10,1);
  ValPercentile (Q2, 1, Ncases, 95, result(+13,1));
  ValPercentile (Q2, 1, Ncases, 90, result(+14,1));
  ValPercentile (Q2, 1, Ncases, 10, result(+15,1));
  ValPercentile (Q2, 1, Ncases, 5, result(+16,1));
  end
  
```

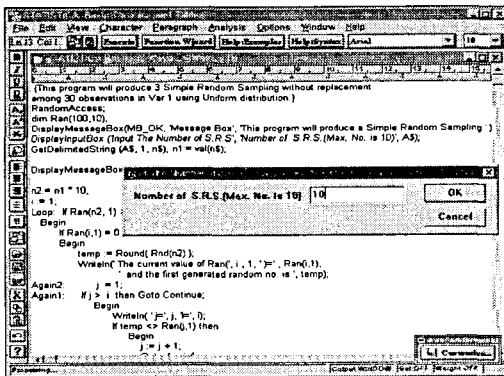
<그림 2> 단순통계량 계산을 위한 프로그램



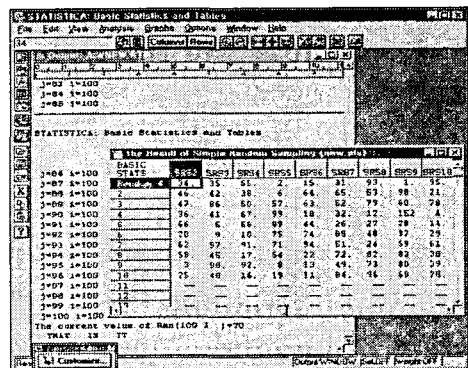
<그림 3> 실행 후 출력 결과

2. 단순 랜덤 비복원 추출

단순 랜덤 비복원 추출을 하기 위하여 주사위, 난수표 등을 이용할 수 있다. 그러나 Statistica Basic Language를 이용하여 1부터 시작하여 최대 100까지 숫자들을 단순 랜덤 비복원 추출하는 과정을 프로그램한 모듈을 준비하였다. <그림 4>와 같이 프로그램을 불러 실행시키면 몇 개를 표본 추출 할 것인가를 묻는 메시지 상자가 나타나고 추출하려는 최대 숫자를 입력하면 샘플링하는 과정과 비복원 추출된 숫자의 결과가 한 개의 column에 10개씩 정리되어 <그림 5>와 같이 출력된다.



<그림 4> 단순 랜덤 비복원 추출 프로그램

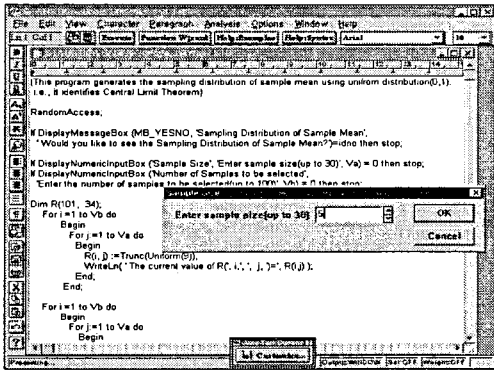


<그림 5> 실행 후 출력 결과

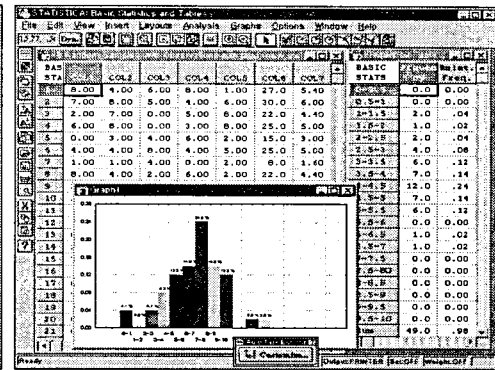
3. 균일 분포에서 추출된 표본평균의 분포

균일 분포에서 확률 표본의 크기(최대 30)를 <그림 6>과 같이 메시지 상자에 입력하고 표본 추출 반복 횟수(최대 100)를 입력한다. 그러면 추출된 표본들의 값들과 그

표본에서 계산된 합과 표본 평균들에 대한 dots분포표와 도수분포도가 <그림 7>과 같이 작성되어 표본 평균의 분포가 중심극한정리에 따라 정규분포의 모습에 가깝게 나타나는 것을 확인할 수 있다.



<그림 6> 표본평균의 분포



<그림 7> 실행 후 출력 결과

II. 결론

처음 소개한 두 변수의 단순 통계량을 구하기 위해서는 Statistica의 Spreadsheet에 자료를 입력하는 것이 필요하다. 자료의 입력이 끝나면 Statistica Basic Language로 프로그램한 모듈을 불러서 실행시킴으로써 입력 자료에 대한 단순 통계량의 값들을 볼 수 있도록 하였다. 이와 같이 프로그램한 모듈을 불러서 실행하면 한 번에 원하는 결과를 볼 수 있으므로 Statistica의 초기 화면에 나타나는 여러 단계의 복잡한 풀다운 메뉴의 옵션들을 거치지 않고 편리하게 결과를 볼 수 있다는 장점이 있다. 단순 랜덤 비복원 추출과 표본 평균의 분포는 Spreadsheet에 자료를 입력할 필요 없이 Statistica에서 제공되는 Basic Language로 작성한 Macro를 불러서 실행만 시키면 그 진행 과정과 함께 출력결과를 화면에 볼 수 있었다. 통계학에 대한 흥미를 유발시키고 교육 효과를 증진시키기 위해서 여기 제시한 Statistica Basic Language로 작성한 프로그램의 실행 도중 나타나는 메시지 상자에 허용된 범위내의 여러 가지 다른 값들을 입력하면 어떻게 그 결과가 변화하는가를 살펴볼 수 있다. 더 다양한 종류의 Scrollsheet나 Graph를 출력하기를 원한다면 여기 제시한 Macro에 Basic Language에서 지원 가능한 명령문들을 추가하거나 사용자의 편리에 따라 수정한다면 보다 다양한 출력 결과를 구할 수 있겠다.

참 고 문 헌

- 김달호 · 박진표 · 이우동 · 이재만 · 장덕준 · 조교영 · 조길호 · 차영준 (1995). *미니탭을 이용한 통계학개론*, 서울: 자유아카데미.
- 김동희 · 김충락 · 손건태 · 정광모 · 정윤식 · 최용석 · 홍창곤 (1998). *통계학*, 서울: 자유아카데미.
- 김우철 · 김재주 · 박병옥 · 박성현 · 송문섭 · 이영조 · 전중우 · 조신섭 (1998A). *일반통계학*, 영지문화사.
- 김우철 · 김재주 · 박병옥 · 박성현 · 송문섭 · 이영조 · 전중우 · 조신섭 (1998B). *통계학개론*, 영지문화사.
- 송문섭 · 조신섭 · 이윤모 · 성병찬 · 윤영주 · 이현부 (1998). 기초통계교육을 위한 여러 통계패키지의 비교 연구 및 엑셀을 이용한 한글 통계패키지의 구현II, *한국통계학회 추계학술발표회 논문집*, pp.82-86.
- 신봉섭 (1995), 통계교육을 위한 Minitab Macro, *한국통계학회 추계학술발표회 논문집*, pp.67-73.
- 신봉섭 (1997). Mathematica의 활용, *한국통계학회 통계계산연구회 추계학술발표회 논문집*, pp.22-31.
- 안기수 · 허문열 (1996). 하이퍼미디어를 이용한 통계 학습 시스템, *한국통계학회 추계학술발표회 논문집*, pp.177-180.
- 안기수 · 허문열 (1997). 인터넷을 이용한 통계교육과 컨설팅의 현황, *한국통계학회 논문집*, 제4권 2호, pp.473-489.
- 이태림 · 김성수 · 강근석 · 이정진 (1997). 원격교육에 적합한 통계학 개론 코스웨어 개발 연구, *한국통계학회 춘계 학술 발표회 논문집*, pp.184-190.
- 전홍석 · 전성해 (1997). Ms-Excel을 이용한 실험 계획법, *한국통계학회 통계계산연구회 추계 학술 발표회 논문집*, pp.6-13.
- 최숙희 (1997), P-값을 이해하기 위한 멀티미디어 프로그램의 개발, *한국통계학회 논문집*, Vol.4 No.3, pp.807-816.
- Anderson, J.E. & Dayton, J.D. (1995). Instructional Regression Modules Using Xlip-Stat, *Journal of Statistics Education*(Online <http://www.stat.ncsu.edu/info/jse/v3n1/anderson.html>), Vol.3, No.1.
- Ferrall, C. (1995). Interactive Statistics Tutorials in Stata, *Journal of Statistics Education*(Online <http://www.stat.ncsu.edu/info/jse/v3n3/ferrall.html>), Vol.3, No.3.
- Ferris, M. & Hardaway, D. (1994). Teacher 2000: A New Tool For Multimedia Teaching of Introductory Business Statistics, *Journal of Statistics Education*

- (Online <http://www.stat.ncsu.edu/info/jse/v2n1/ferris.html>), Vol. 2, No. 1.
- Francis, B. (1997). Graphs of Binomial and Poisson Distributions on a Graphical Calculator, *Teaching Statistics*, Vol. 19, No. 1. pp.24-25.
- Laviolette, M. (1994). Linear Regression: The Computer as a Teaching Tool, *Journal of Statistics Education*(Online <http://www.stat.ncsu.edu/info/jse/v2n2/laviolette.html>.), Vol. 2, No. 2.
- Short, T. (1997). Statistics in the Classroom: A Quick Introduction to the Statistical Capabilities of the TI-83 Calculator, *The Statistics Teacher Network* (<http://www.amstat.org/education/STN/winter97/Short.html>.), Winter 1997.
- Statistica (1997). Release 5, <http://www.statsoft.com>.
- TI-82, Texas Instruments-82, The math power of a computer with the independence of a graphing calculator.
- TI-92, Texas Instruments-92, The math power of a computer with the independence of a graphing calculator.
- Trumbo, B.E. (1994). Some Demonstration Programs for Use in Teaching Elementary Probability: Part1 and 2, *Journal of Statistical Education*, (Online <http://www.stat.ncsu.edu/info/jse/v2n2/trumbo.html>), Vol. 2, and No. 2.
- Whigham, D. (1998). How to Excel with Normal Distributions, *Teaching Statistics*, Vol. 20, No. 3. pp.84-85.