

The Teaching of Statistics using Excel VBA

Hyun Seok Choi¹⁾

Abstract

We introduce a program that enhances the interest and understanding of students in Statistics. This program explains various statistical concepts and procedures by showing detailed steps of calculations with graphs and simulations. This program utilizes a readily accessible Excel VBA.

Keywords : Macro, VBA

1. 서론

오늘날 통계학은 과학의 모든 분야 뿐 아니라 인문학, 사회과학 분야에서도 많이 이용되고 있으며 더욱 증가하는 추세에 있다. 따라서 많은 학생들은 여러 분야에서 통계학의 기본개념과 원리를 접하게 되며 이를 다루어야 할 필요가 있다.

이와 같이 통계에 대한 사회적 관심이 점차적으로 높아지고 있는 현 시점에서 통계학 교육의 중요성이 대두되고 있지만 대학에서의 통계학 교육은 수리적인 면에 치우쳐 있어서 실세계현상에 대한 응용력 부족이 우려되고 학생들의 통계학에 대한 관심이 저하되는 것이 사실이다. 통계학이 이해하기 어려운 학문으로 인식되어 학생들이 배우기를 꺼려하는 경향이 있다.

수학적인 기초가 부족한 학생들을 대상으로 수리적인 접근방법을 고집함으로써 교수와 학생이 모두 만족하지 못하는 결과를 산출하는 경우가 다반사였다. 통계학의 기본적인 개념들을 고등수학의 배경 없이도 쉽게 이해하여 다양한 응용분야에 적용할 수 있는 능력을 기르도록 해야 한다. 교육환경의 주체가 기존의 교수자 중심 교육에서 학습자의 요구와 학습자 중심 교육으로 변화하여야 한다. 흥미로운 학습이 가장 효율적이라는 사실에 바탕을 두고 통계학습의 흥미를 높이고 이해능력을 향상시키기 위해서는 학습자들이 적극적으로 활용할 수 있는 통계교육에 있다고 할 수 있다. 주어진 자료를 가지고 단계적 절차를 거쳐 결과가 나오기까지 과정을 나타내어 주고 많은 자료를 빠르게 처리하기보다는 통계 자료 분석의 단계적 과정들을 설명, 수식, 그

1) Full Time Instructor, Department of Statistics, Keimyung University, Daegu, 704-701, Korea
E-Mail : chsuk1@kmu.ac.kr

래프 등을 통해 보여줌으로 적극적인 학습이 되게 한다.

본 연구는 통계교육에 있어서 학습자 스스로 이해와 분석능력을 습득할 수 있고 능동적인 학습을 유도할 수 있는 프로그램을 개발하여 소개하고자 한다. 2절에서는 연구방법에 대하여 기술하고, 3절에서는 프로그램 구성 및 설명에 대하여 기술하고 4절에서 결론을 맺는다.

2. 연구방법

본 연구는 단순한 자료 분석보다는 학습자의 통계전반에 관한 개념의 이해를 돕는 역할에 초점을 맞추었다. 동적 통계 프로그램으로 다양한 그래프, 모의실험 등을 제공하면 통계학습의 효과를 극대화시킬 수 있을 것으로 기대된다.

본 연구의 특징은 다음과 같다.

첫째, 통계 개념설명을 설명, 수식, 그래프 등을 통해 보여준다.

둘째, 자료를 분석할 때에도 자료입력 후 결과가 바로 나오는 것이 아니라 중간과정, 설명, 수식 등이 제공되기 때문에 현재의 통계 프로그램과는 다른 독창성을 가진다.

셋째, 매크로 기능을 사용하여 동적인 엑셀 프로그램으로 구현한다.

위와 같은 연구 목적을 달성하기 위하여 프로그램 개발에 사용하는 도구는 계산업무 전반에 이용할 수 있는 엑셀을 기반으로 하여 스프레드시트(sheet)와 양식도구, 매크로, VBA(Visual Basic for Application)를 사용한다.

스프레드시트는 자동계산 기능, 문서작성 기능, 차트작성 기능, 데이터베이스 관리 기능, 작업 자동화 기능을 수행하며 양식도구, 매크로, VBA 등은 프로그램 제어 및 함수 사용, 설명, 그래프 등을 위한 작업을 수행한다. 양식도구는 Dialog Sheet에서 대화상자를 사용자가 직접 작성할 때 사용하는 것으로 프로그램에서는 명령단추(CommandButton), 확인란(CheckBox), 옵션단추(OptionButton), 그룹상자(GroupBox) 등을 사용한다. 매크로는 엑셀에서 실행한 작업 내용을 비주얼베이직(Visual Basic)이라는 프로그래밍언어로 기록하여 기록된 비주얼베이직 언어의 내용을 추후 자동으로 실행하게 된다. 매크로는 비주얼베이직 코드로 기록되는데 마이크로소프트 오피스에서 사용되는 비주얼베이직을 VBA라 하며 엑셀에서 사용되는 VBA 코드를 '엑셀 VBA'이라고 한다(Jacobson, 2002; Walkenbach, 2004).

엑셀은 학교나 직장에서 학습과 업무에 쉽게 활용할 수 있는 도구이기 때문에 엑셀을 이용한 다양한 통계 교육 혹은 분석 도구가 개발되고 있다. 조신섭 등(1999)은 엑셀의 매크로를 엑셀을 이용하여 구현한 통계교육용 프로그램을 개발하였고, 엑셀을 사용하여 통계자료분석을 소개하는 문헌으로는 송문섭과 조신섭(2002), 신봉섭(1998) 등이 있으며 이외에도 엑셀로 자료분석을 다룬 많은 문헌들이 출판되고 있다. 또한 웹사이트 www.unistat.com, www.xlstat.com에서도 엑셀을 이용한 다양한 통계 분석 방법을 제공하고 있다. 이러한 도구들의 공통적인 단점은 분석 기능에 초점을 맞춘 것으로 사용자에게 패키지의 작동 단계를 동적으로 보여주지 않기 때문에 학습효과 개선기능은 부족한 것으로 판단된다.

엑셀의 통계분석도구는 다른 통계패키지보다 매우 제한적인 것은 사실이나 통계학 개론에서 다루는 통계기법들을 실습하는 데는 별로 어려움이 없다. 특히 최현석과 김

태운(2005), 최현석과 송규문(2005), 최현석(2006)은 매크로(Macro)로 프로그램을 작성하여 고급통계기법들을 사용할 수 있도록 하였다(http://home.kmu.ac.kr/~statexe).

VBA Project의 모듈 창에는 다음과 같이 코드를 작성하였다.

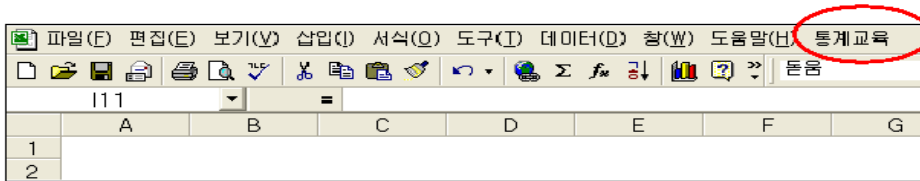
첫째, Sub 문을 사용하여 셀을 지정하여 입력받은 값을 수식을 이용하여 기록하는 프로시저를 작성하였다.

둘째, 엑셀 자체에서 제공되는 분석기능과 VBA 등으로 작성된 프로그램을 연결하였다(Jacobson, 2002; Walkenbach, 2004).

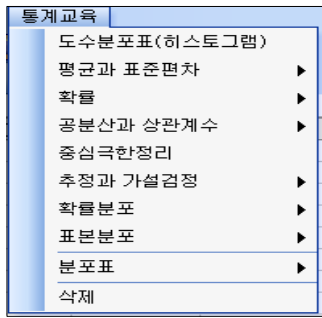
셋째, 명령단추를 사용하여 단추를 누르면 바로 매크로가 실행되게 하였다.

3. 프로그램의 구성 및 설명

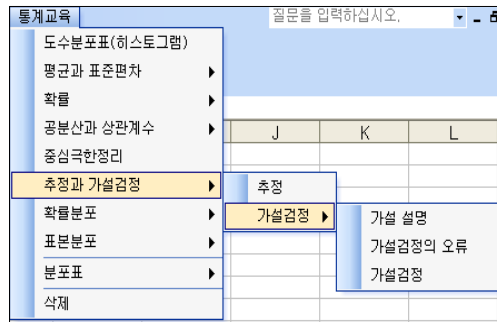
본 프로그램의 일부분을 설명하면 다음과 같다. Add-in software Excel VBA를 이용하여 개발하였다. 엑셀을 구동하면 새로운 메뉴(통계교육)이 생성되어 대화상자를 활성화하여 사용할 수 있다(<그림 1>~<그림 3>).



<그림 1> 초기화면



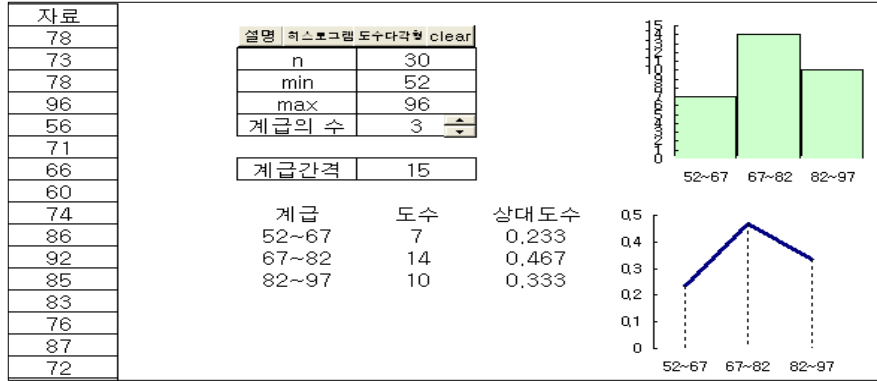
<그림 2> 통계교육 메뉴



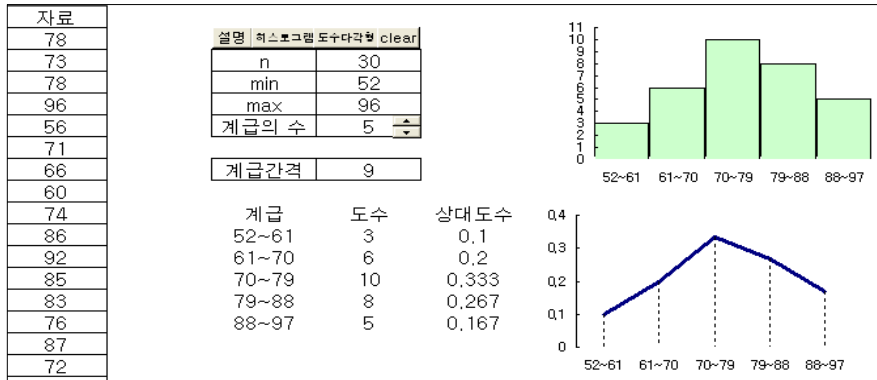
<그림 3> 통계교육 메뉴

도수분포표를 작성할 때 계급(class)의 수는 자료의 특성이 잘 반영되도록 정하여야 한다. 마우스를 회전자(spinner)를 움직여 계급의 수를 변화시켜 변화 상태를 파악하여 분포의 형태를 파악하도록 하였다(<그림 4>~<그림 6>).

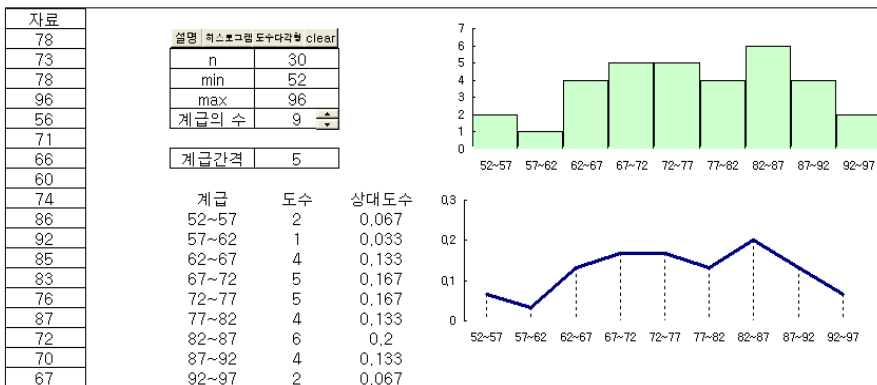
자료를 입력하면 n, min, max값이 계산되어지고 계급의 수에 있는 회전자(spinner)를 변화시키면 도수분포표와 히스토그램이 나타난다. 적당한 계급의 수를 정하면 그 자료가 갖고 있는 중심위치와 산포도 그리고 치우침을 한눈에 파악할 수 있는 히스토그램을 확인할 수 있을 것이다.



<그림 4> 계급의 수가 3일 때 도수분포표, 히스토그램, 도수다각형



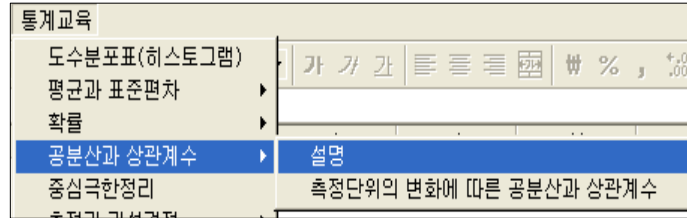
<그림 5> 계급의 수가 5일 때 도수분포표, 히스토그램, 도수다각형



<그림 6> 계급의 수가 9일 때 도수분포표, 히스토그램, 도수다각형

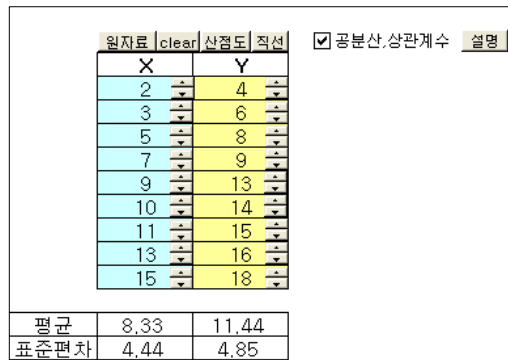
<그림 7>~<그림 11>은 두 변수의 연관성의 측도로 사용되는 공분산과 상관계수의 개념을 파악하는 것으로 확률변수 X의 증감에 따른 확률변수 Y의 증감의 경향에 대한 측도로서 $(X - \mu_X)(Y - \mu_Y)$ 의 기대값의 성질을 쉽게 이해하도록 하여 공분산과

상관계수의 의미를 파악하도록 하였다.

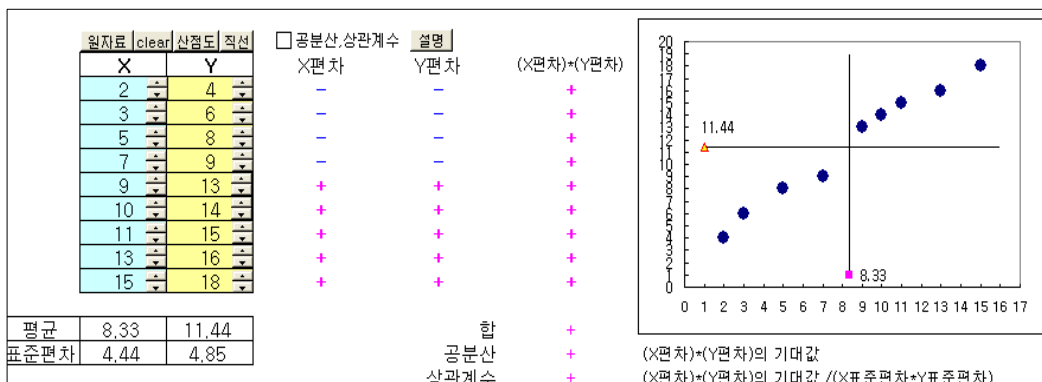


<그림 7> 공분산과 상관계수

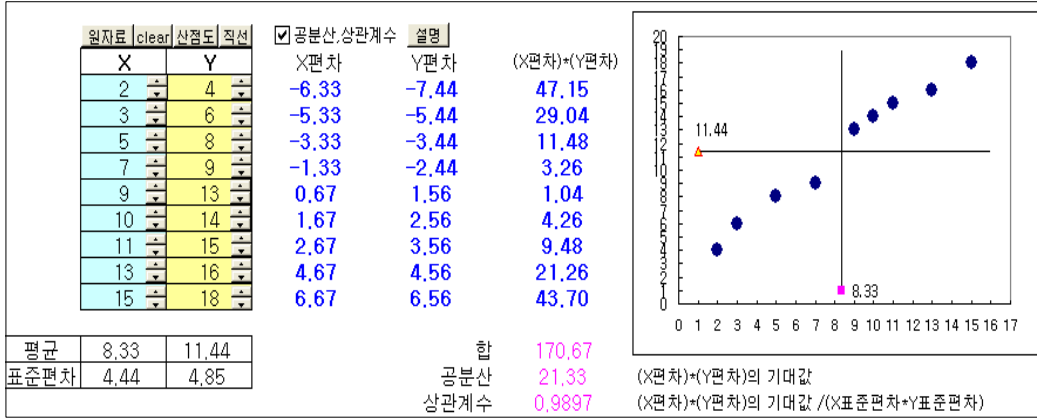
<그림 8>은 공분산과 상관계수 설명 초기화면으로 데이터를 입력해 놓았다 회전자를 움직여서 데이터를 변화시킬 수 있다. 각 버튼을 클릭하여 부호, 값, 산점도, 회귀직선의 변화 상태를 확인할 수 있다.



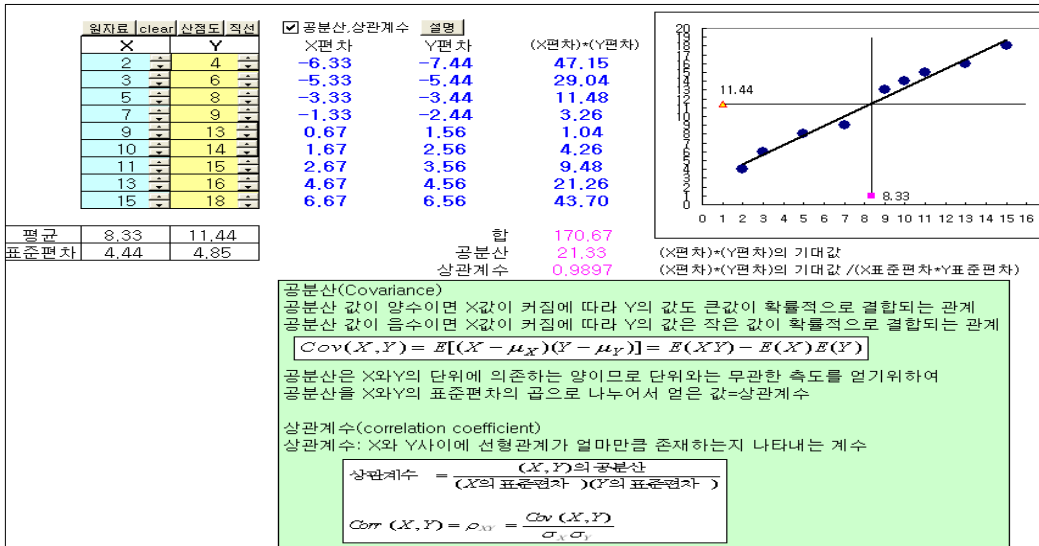
<그림 8> 공분산과 상관계수 설명 초기화면



<그림 9> 확률변수 X의 증감에 따른 확률변수 Y의 증감의 경향에 대한 부호와 산점도

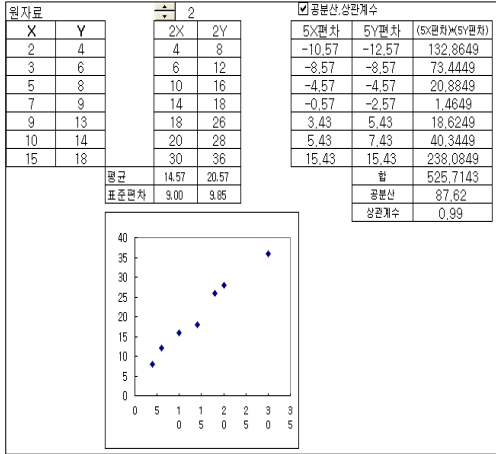


<그림 10> 확률변수 X의 증감에 따른 확률변수 Y의 증감의 경향에 대한 값과 산점도

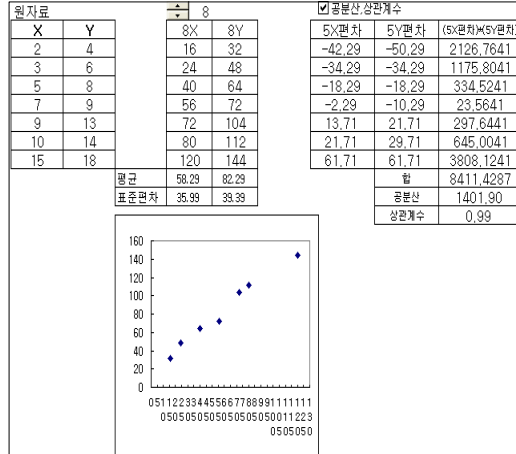


<그림 11> 공분산과 상관계수 설명화면

<그림 12>와 <그림 13>은 공분산은 X와 Y의 단위에 의존하고, 상관계수는 단위에 무관한 측도라는 것을 이해하는 내용이다. 데이터가 단위에 따라 변할 때의 공분산, 상관계수와 통계량의 변화 및 산점도의 관계 등을 관찰함으로써 학습효과를 높일 수 있게 하였다. 회전자로 증가시켰을 때 공분산은 값이 변하나 상관계수는 변하지 않는 것을 확인할 수 있다.

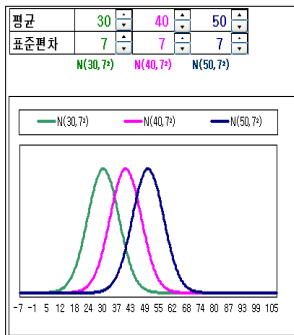


<그림 12> 원자료를 2배 증가했을 때의 공분산과 상관계수

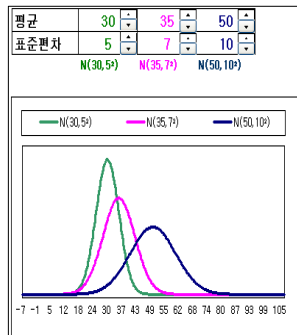


<그림 13> 원자료를 8배 증가했을 때의 공분산과 상관계수

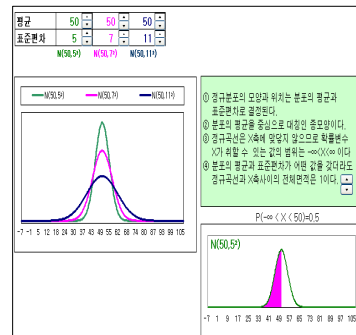
<그림 14>~<그림 16>에서 정규분포는 평균과 분산에 의해 분포의 모양이 결정된다는 것을 확인할 수 있다. 평균은 다르나 분산이 같은 경우, 평균은 같으나 분산이 다른 경우, 그리고 평균과 분산이 모두 다른 경우의 정규분포 밀도함수를 회전자를 움직여서 비교할 수 있다.



<그림 14>
 $N(30, 7^2)$, $N(40, 7^2)$, $N(50, 7^2)$

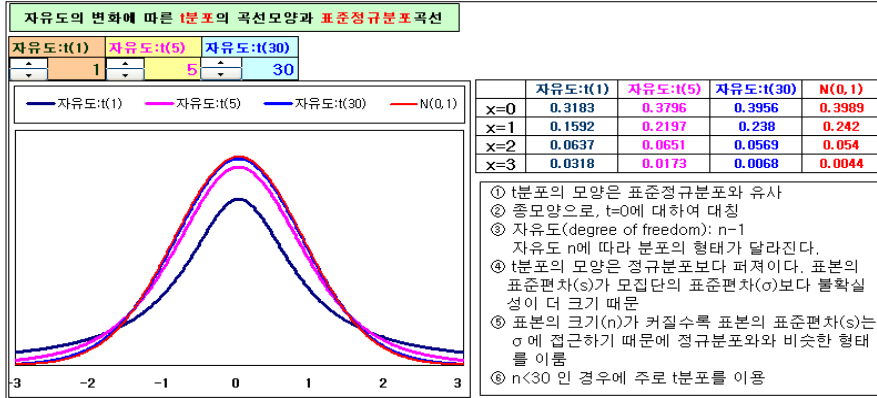


<그림 15>
 $N(30, 5^2)$, $N(30, 7^2)$, $N(30, 10^2)$



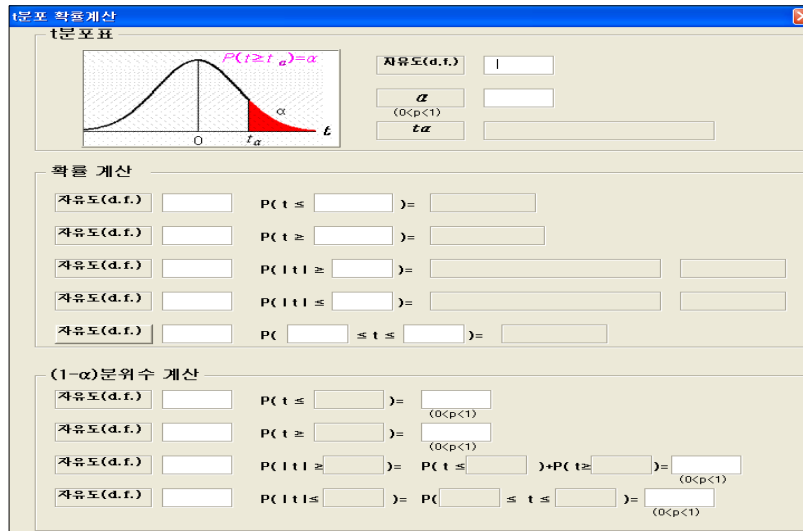
<그림 16>
 $N(50, 5^2)$, $N(50, 7^2)$, $N(50, 11^2)$

<그림 17>은 정규모집단에서 표본의 크기가 작을 때 유용하게 쓰이는 t분포를 나타내는 것으로 t분포의 곡선은 표준정규분포의 곡선과 $x=0$ 을 대칭축으로 하여 완전대칭이지만, 곡선모양이 표준정규분포의 곡선에 비하여 완만하고 두터운 꼬리를 가지고 있다는 것을 확인할 수 있다. 또 회전자를 변화시켜 자유도가 증가함에 따라 분포모양을 $N(0,1)$ 과 비교할 수 있다.



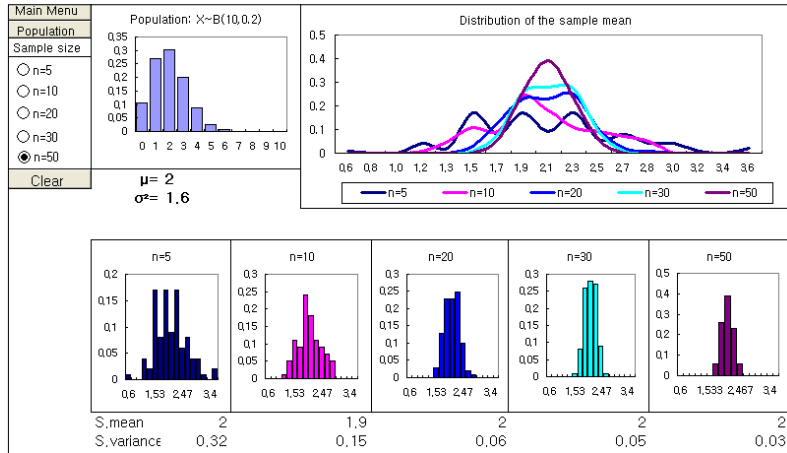
<그림 17> 자유도 변화에 따른 t분포 곡선과 표준정규분포곡선

<그림 18>에서 여러 가지 자유도에 관한 t분포에 대한 확률을 계산할 수 있고, 자유도가 n인 t분포의 (1- α)백분위수인 $P[T \geq t_\alpha(n)] = \alpha$ 를 구할 수 있다.



<그림 18> t분포 확률계산

<그림 19>는 이항분포에 대하여 중심극한정리를 실험하는 화면이다. 표본 크기의 옵션단추를 클릭함으로써 표본크기 증가에 따라 표본평균의 분포가 정규분포에 근사됨을 동적으로 확인할 수 있게 된다. $\bar{X}_n = S_n/n$ 의 점근 정규성(asymptotic normality)에 초점을 맞추어서 n이 클 경우 $\bar{X}_n \approx N(\mu, \sigma^2/n)$ 이 확인하였다.



<그림 19> 중심극한정리

4. 결론

본 연구에서는 통계학 전반의 내용을 엑셀을 이용하여 실행하는 매크로로서 엑셀에서 바로 구동이 가능한 Add-in 방식으로 스스로 체계적이고 효율적인 통계학을 할 수 있는 프로그램을 소개하였다. 본 연구는 다음과 같은 특징을 가짐으로써 통계학 수업에 꼭 필요한 프로그램이 될 것이다.

첫째, 통계이론과 자료 분석의 단계적 과정들을 설명, 수식, 그래프 등을 통해 보여주어 개념에 대한 이해를 돕게 하였다.

둘째, 확률과 중심극한정리에서 통계적 실험을 수없이 많이 반복 시행하는 모의실험의 결과를 보여주어 효과적인 학습이 되도록 하였다. 통계이론의 가장 중추적인 역할을 하는 각종 점근이론(예를 들면 중심극한정리 등)들의 표본크기 변화에 따라 달라지는 모습을 제공하여 내용이해를 돕게 하였다.

셋째, 매크로 기능을 사용하여 동적인 엑셀프로그램으로 구현하였다. 확률분포에서 평균과 표준편차, 자유도 등의 변화를 회전자 기능으로 제어하여 다양한 분포들의 모습을 보여준다.

현실세계에서 가장 많이 사용하고 있는 엑셀을 이용하여 흥미 있고 응용성이 뛰어난 프로그램을 작성하면 수리적인 접근을 배제한 통계학을 이해하고 활용하여 문제해결능력향상과 실제 업무적용능력을 증대시킬 수 있을 것이다.

참고문헌

1. 송문섭, 조신섭 (2002). 엑셀에 기초한 통계학 입문, 자유아카데미, 서울.
2. 신봉섭 (1998). 엑셀 활용과 통계자료분석, 자유아카데미, 서울.
3. 조신섭, 송문섭, 이운모, 성병찬, 윤영주, 이현부 (1999). 기초통계교육을 위한 통계소프트웨어의 개발 -Excel에 기초한-. 품질경영학회지, 제27권

- 제2호, 277-290.
4. 최현석, 김태운 (2005). Computer Simulation Program for Central Limit Theorem, *Journal of the Korean Data & Information Science Society*, Vol. 16, No 2. pp.359~369.
 5. 최현석, 송규문 (2005). Computer Program Development for Probability Distribution, *Journal of the Korean Data & Information Science Society*, Vol. 16, No 3. pp.581~589.
 6. 최현석 (2006). A Program for Teaching Type I and Type II Errors, *Journal of the Korean Data & Information Science Society*, Vol. 17, No 1. pp.19~30.
 7. Jacobson, R. (2002). *Microsoft Excel 2002 Visual Basic Step by Step*. Redmond, W.A., Microsoft Press.
 8. Walkenbach, J.(2004). *Excel 2003 Power Programming With VBA*, Wiley Publishing.
 9. <http://home.kmu.ac.kr/~statexe>
 10. <http://www.xlstat.com>
 11. <http://www.unistat.com>

[2006년 6월 접수, 2006년 8월 채택]