

Computer Program Development for Probability Distribution

Hyun Seok Choi¹⁾ · Gyu Moon Song²⁾

Abstract

The purpose of this thesis is to develop and introduce Add-in program which we can systematically, visually and dynamically study discrete probability distribution of binomial distribution, poisson distribution and hypergeometric distribution, and continuous probability distribution of normal distribution, exponential distribution, and the definition and characteristics of t distribution, F distribution and χ^2 distribution to be driven from normal distribution, and graphs, the computation process of probability by using VBA which is the device of Excel.

Keywords : 엑셀 매크로, Add-in 프로그램, VBA

1. 서론

통계학의 이론은 확률에 근거하여 설명되어지며, 모든 확률변수들은 특정한 확률분포를 갖는다. 확률분포들은 매우 다양한 형태로 나타나게 되는데, 다양한 형태로부터 여러 특징을 파악하는 것이 중요하다. 이산확률분포와 연속확률분포에서 관심 있는 확률이 어떻게 계산되며, 확률분포의 특징을 나타내는 모수(parameter)의 역할, 확률분포의 모양이나, 특징을 이해하는 것이 필요하다. 특히 정규분포는 통계추정방법에서 중심적인 역할을 하고, 응용범위가 넓으므로 정규분포의 특징을 잘 파악하여야 한다.

정규모집단의 모분산 σ^2 에 대한 통계적 추론, 범주형 자료의 분석 등에 활용되는 χ^2 분포, 모분산 σ^2 이 알려져 있지 않은 경우 모평균 μ 의 추정에 사용되는 t분포, 정규모집단의 모분산들의 통계적 추론, 분산분석 등에 활용되는 F분포는 통계추정, 가설검정 등 제 분야의 기초적인 분포들이므로 특징 및 확률계산, 자유도의 변화에 따라 표본분포의 형태가 어떻게 달라지는지를 이해하여야 한다.

1) First Author : Lecturer, Department of Statistics, Keimyung University, Daegu, 704-701, Korea

E-Mail : chsuk1@kmu.ac.kr

2) Professor, Department of Statistics, Keimyung University, Daegu, 704-701, Korea

확률분포와 표본분포의 성질을 이해하기 위해서는 단계적인 개념설명, 동적으로 변화하는 그래프의 움직임 등으로 분포의 특성을 파악하는 것이 내용이해를 돕는데 효율적일 것이다. 즉 평균, 표준편차, 자유도 등에 따라 분포의 변화하는 모양을 정적인 변화가 아닌 동적으로 보는 것이 이론적 내용을 효과적으로 이해하는데 도움을 줄 것이다.

다양한 확률분포와 표본분포의 특징을 올바르게 파악하기 위해서는 통계패키지 등의 소프트웨어 사용이 필수적이다. 현재 사용되고 있는 통계패키지는 많은 분석방법들을 포함하고 있으나 사용하기에는 까다롭고 사용 비용 또한 많이 든다. 분석기능에 초점을 맞춘 것으로 개념 설명은 없고 결과만 제공해준다.

효율적인 사용을 위해서는 기존의 통계분석용 패키지보다는 현재 업무 등에서 가장 많이 이용되고 있는 엑셀을 이용하는 것이 편리하다. 엑셀의 분석도구는 매우 제한적이어서 엑셀을 이용한 일반통계분석에는 한계가 있으나 엑셀 기능의 차원을 높일 수 있는 양식도구(Forms)와 VBA(Visual Basic Application)에 기초한 매크로를 이용한 통계교육은 통계학의 개념과 이론을 이해하는데 많은 도움을 줄 수 있다.

VBA를 이용한 Add-in 프로그램을 개발하여 새로운 메뉴와 툴바(Toolbar)를 생성한 후 단계적인 절차로 분포들의 특성을 파악하는데 활용하고자 한다.

기초통계교육에 엑셀을 이용한 예는 교육용 프로그램개발([4],[5],[6],[7],[8],[9]), 교재출판([1],[2],[3],[10]), 웹사이트([13],[14],[15],[16])등 많이 활용되고 있다. 즉 엑셀을 이용한 다양한 분석방법을 제공하고 있으며 학교나 직장에서 자신의 학습과 업무에 활용할 수 있는 엑셀을 이용한 교육이 보편화되고 있다. 그러나 대부분이 함수식을 나열하여 계산의 편리함을 추구하거나, 자료입력 후 결과가 바로 출력되는 분석위주의 프로그램들이다.

엑셀의 도구인 VBA를 이용하여 이항분포, 포아송분포, 초기하분포의 이산확률분포와 정규분포, 지수분포의 연속확률분포, 정규분포로부터 유도된 t분포, F분포, χ^2 분포의 정의와 특성, 확률계산 과정 등을 단계적, 시각적, 동적(Dynamic)으로 학습할 수 있는 Add-in 프로그램을 개발하여 소개하고자 한다.

2. 프로그램 개발

본 프로그램은 통상 사용되는 통계패키지들이 일반인들에게 사용하기 까다롭고 비용 또한 많이 드는 단점을 극복하고자 일반 사회업무와 밀접한 관련이 있는 엑셀을 사용하여 개발하였다.

프로그램 개발에 사용하는 도구는 계산업무 전반에 이용할 수 있는 엑셀을 기반으로 하여 스프레드시트(spreadsheet)와 양식도구, 매크로, VBA를 사용한다. 여기서 스프레드시트는 자동계산 기능, 문서작성 기능, 차트작성 기능, 데이터베이스 관리 기능, 작업 자동화 기능을 수행하며 양식도구, 매크로, VBA는 프로그램 제어 및 함수 사용, 설명, 그래프 등을 위한 작업을 수행한다. 매크로는 엑셀에서 실행한 작업 내용을 비주얼베이직(Visual Basic: VB)이라는 프로그래밍언어로 기록하여 기록된 비주얼베이직 언어의 내용을 추후 자동으로 실행하게 된다. 매크로는 비주얼베이직 코드로 기록하고 마이크로소프트 오피스에서 사용되는 비주얼베이직을 VBA이라 하며 엑셀에서 사용되는 VBA 코드를 '엑셀 VBA'라고 한다([11],[12]).

양식도구는 Dialog Sheet에서 대화상자를 사용자가 직접 작성할 때 사용하는 것으로 프로그램에서는 회전자(Spinner), 명령단추(CommandButton), 확인란(CheckBox), 옵션단추(OptionButton), 그룹상자(GroupBox) 등을 사용한다.



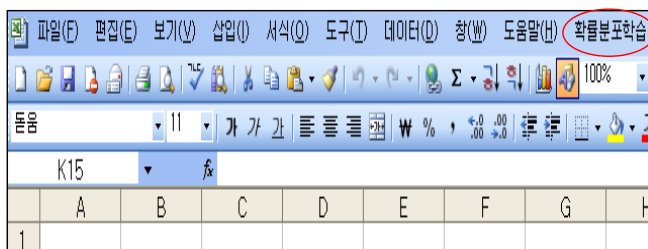
[그림 1] 엑셀 매크로와 비주얼베이직의 프로시저

VBA Project의 모듈(Module) 창에는 다음과 같이 코드를 작성하였다.
 첫째, Sub 문을 사용하여 일반프로시저로 작성하였다.
 둘째, 엑셀 자체에서 제공되는 분석기능과 VBA 등으로 작성된 프로그램을 연결하였다.
 셋째, 설명, 수식, 그래프 등은 Rectangle, Object, Group, Chartobjects 등으로 작성하여 ActiveSheet를 사용하여 활성화하였다.
 넷째, 양식도구를 사용하여 단추를 누르면 바로 매크로가 실행되게 하였다.

3. 프로그램 설명

엑셀 Add-in 프로그램으로 KESS와 XLSTAT가 있지만 자료분석에 대한 결과만 제공할 뿐이다([8]). 이외의 다른 통계패키지들은 통계학 전반에 걸쳐서 많은 분석과 자료분석에 대한 결과만 제공할 뿐이지만, 본 프로그램은 확률분포와 표본분포에 대하여 다른 통계패키지에는 없는 단계적 절차, 수식, 그림 등을 함께 제공하여 기본 개념을 터득하게 하는 학습용 또는 교육용 프로그램이다.

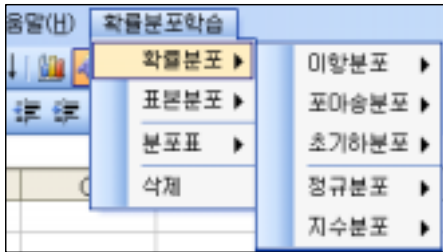
본 프로그램의 일부분을 설명하면 다음과 같다.
 엑셀을 구동하면 새로운 메뉴(확률분포학습)가 생성되어([그림 2]) 대화상자를 활성화하여 실행할 수 있다([그림 3]~[그림 7]).



[그림 2] 확률분포학습의 메뉴 창



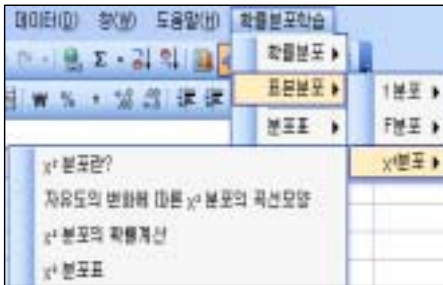
[그림 3] 확률분포학습



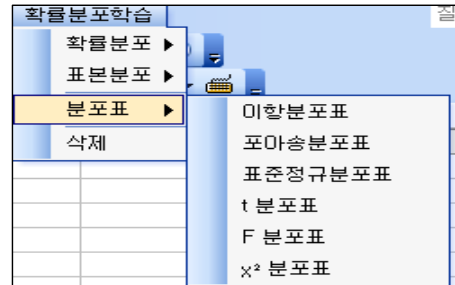
[그림 4] 확률분포의 내용



[그림 5] 이항분포의 내용

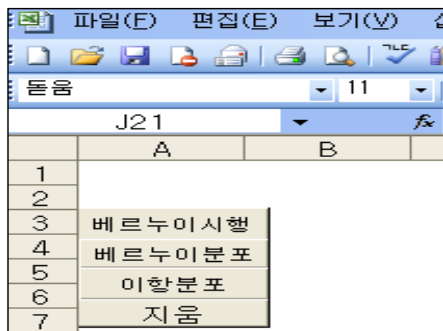


[그림 6] x^2 분포 내용

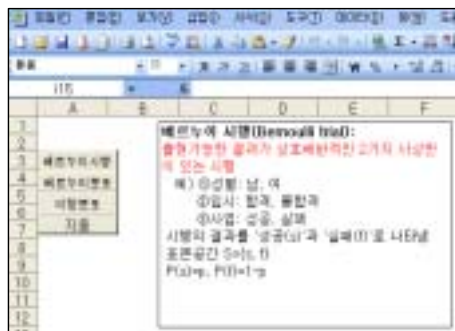


[그림 7] 분포표의 내용

[그림 5]에서 '이항분포란?'을 클릭하면 [그림 8]이 나타나고 [그림 8]에서 베르누이 시행 단추를 클릭하면 [그림 9]와 같이 베르누이 시행에 대한 설명 화면이 주어진다.



[그림 8] 이항분포 개념 설명 단추 창



[그림 9] 베르누이 시행에 대한 설명화면

[그림 5]의 이항분포에서 'p의 변화에 따른 이항분포 모양'을 클릭하면 [그림 10]과 같다. 회전자로 확률을 변화시키면 $0 < p < 0.5$, $p = 0.5$, $0.5 < p < 1$ 에 따라 분포모양과 설명 화면이 자동으로 변화하는 것을 확인할 수 있다.



[그림 10] $X \sim B(6, 0.35)$

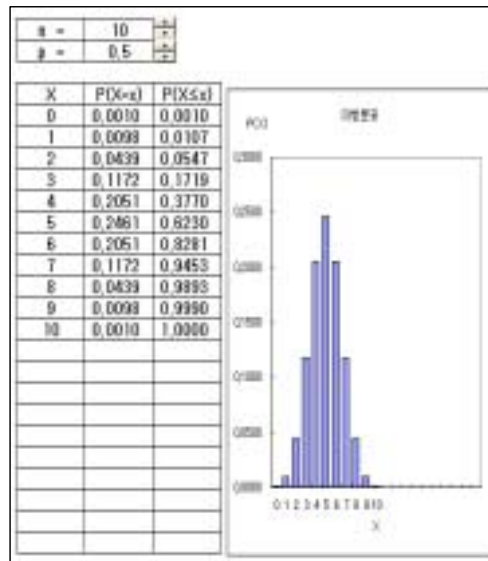


[그림 11] $X \sim B(6, 0.75)$

[그림 5]에서 '이항분포의 확률계산'을 클릭하면 [그림 12]가 주어지서 다양한 확률 변수에 대하여 확률을 계산할 수 있다. [그림 13]은 이항분포표로서 회전자로 n과 p를 변화시키면 동적으로 확률과 그래프가 변하는 것을 확인할 수 있어 수표를 읽는 방법과 그래프의 특징을 이해할 수 있다.



[그림 12] 이항분포 확률계산



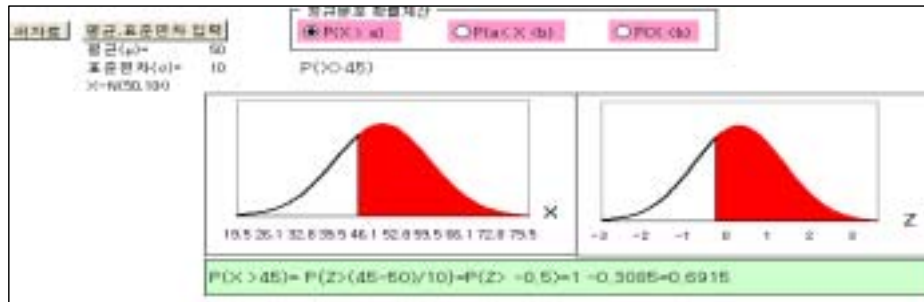
[그림 13] 이항분포표

정규분포 문제에서 평균, 표준편차, 구하고자 하는 확률변수의 값을 입력하면 기존의 통계패키지는 확률값만 출력된다. 그러나 본 프로그램은 [그림 15]와 같이 해당 단추를 클릭하여 Inputbox에 값을 입력하면 정규분포곡선, 표준정규분포곡선, 표준화과정의 단계적 절차를 거쳐서 확률값을 출력하므로 내용이해를 쉽게 할 수 있다. 입력

값에 따라 정규분포와 표준정규분포 곡선이 동적으로 움직인다.

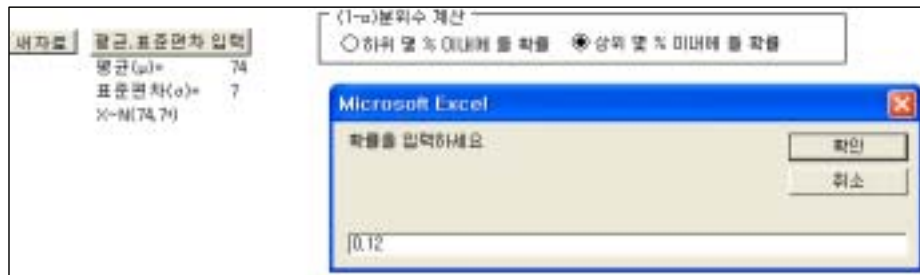


[그림 14] 정규분포 확률계산 초기화면

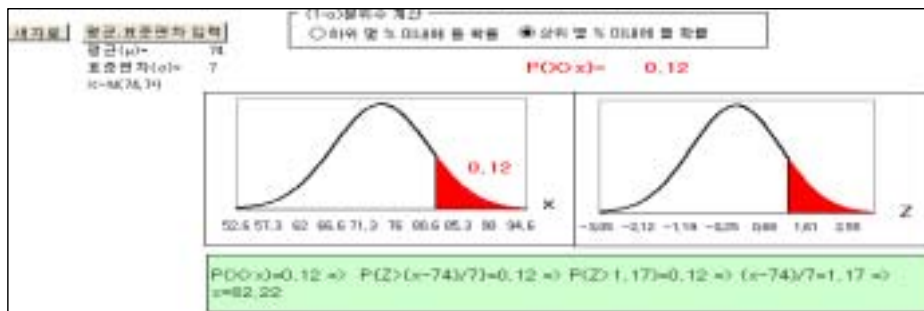


[그림 15] 정규분포 확률계산 화면

[그림 17]은 평균과 표준편차 입력 후 ‘상위 몇 % 이내에 들 확률’을 클릭하여 확률을 입력한 후 나타난 결과화면이다.

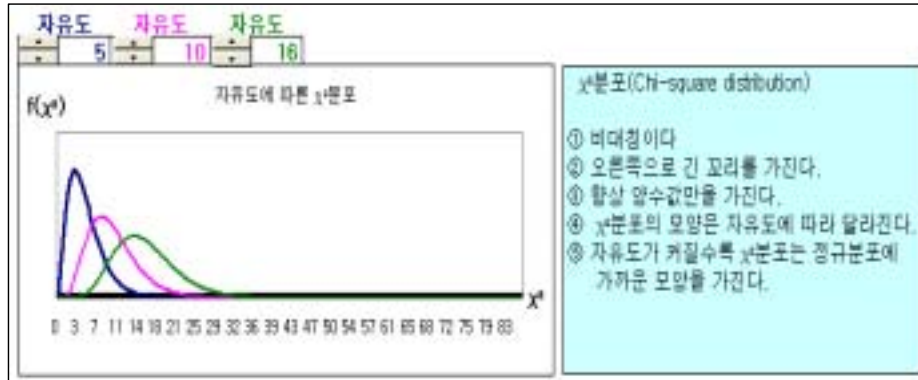


[그림 16] (1-α)분위수 계산 화면



[그림 17] (1-α)분위수 계산 결과 화면

[그림 6]에서 '자유도의 변화에 따른 χ^2 분포의 곡선모양'을 클릭하면 [그림 18]이 나타난다. 회전자를 변화시키면 자유도에 따라 동적으로 곡선이 움직여 χ^2 분포의 특성을 파악할 수 있다.



[그림 18] 자유도의 변화에 따른 χ^2 분포

4. 프로그램의 특징과 구성

본 프로그램의 특징은 다음과 같다.

첫째, 프로그램을 실행하면 Add-in 프로그램으로 새로운 메뉴와 툴바(Toolbar)가 생성된다.

둘째, 개념설명, 수식, 그래프 표현 등의 전자교재의 형태를 갖는다.

셋째, 그래프의 움직임을 마우스 버튼만으로 동적으로 변화시켜 분포의 특징을 쉽게 파악할 수 있다.

넷째, 단계적 절차, 그래프 등으로 내용을 체계적으로 이해할 수 있다.

다섯째, 결과가 나오는 과정을 GUI(Graphic User Interface)환경으로 버튼을 누름으로써 진행이 되도록 하였다.

본 프로그램은 자주 사용되고 실제 현상에서도 많이 나타나는 대표적인 분포들에 대하여 개념의 특성을 단계적이고 시각적인 기법을 잘 활용하여 학습자들의 이해를 돕는 프로그램이다. 학습하고자 하는 단원의 단추를 누르면 해당단원으로 넘어가는 학습자가 단지 바라보는 것이 아닌 직접 실행하는 학습이 되어 학습의욕을 고취시킬 수 있다.

본 프로그램은 <표 1>의 경우에 대하여 학습할 수 있다.

<표 1> 학습범위

이항 분포·	☞ 이항분포 개념 설명	지수 분포	☞ 지수분포 개념 설명
	☞ p 의 변화에 따른 이항분포 모양	t분포	☞ 지수분포 확률계산
	☞ 이항분포의 확률계산		☞ t분포 개념 설명
	☞ 이항분포표		☞ 자유도의 변화에 따른 t분포의 곡선 모양
	☞ t분포의 확률계산		
포아송 분포	☞ 포아송분포 개념 설명	F분포	☞ t분포표
	☞ 포아송분포의 확률계산		☞ F분포 개념 설명
	☞ 이항분포의 포아송분포 수렴		☞ 자유도의 변화에 따른 F분포의 곡선 모양
	☞ 포아송분포표		☞ F분포의 확률계산
초기하 분포	☞ 초기하분포 개념 설명	카이제곱 분포	☞ F분포표
	☞ 초기하분포의 확률계산		☞ χ^2 분포 개념 설명
	☞ 초기하분포의 이항분포 수렴		☞ 자유도의 변화에 따른 χ^2 분포 의 곡선모양
정규 분포	☞ 정규분포 개념 설명	분포표	☞ χ^2 분포의 확률계산
	☞ μ 와 σ 의 변화에 따른 정규 분포 곡선모양		☞ χ^2 분포표
	☞ 정규분포의 확률계산		☞ 이항분포표
	☞ 이항분포의 정규분포 수렴		☞ 포아송분포표
	☞ 표준정규분포표		☞ 표준정규분포표
			☞ t분포표
			☞ F분포표
			☞ χ^2 분포표

5. 결론

본 연구에서는 엑셀매크로를 사용하여 확률분포를 학습할 때 체계적이고 효율적인 통계학습을 할 수 있는 프로그램을 개발하여 내용 및 방법을 소개하였다. 다른 통계 패키지들은 결과가 나타나기까지의 과정이 없지만, 본 프로그램은 분포학습에 대하여 동적으로 변하는 그래프 및 설명, 계산과정 등이 주어지므로 학습효과를 최대화 할 수 있으며, 엑셀 프로그램만 있으면 바로 실행하여 원리와 과정을 학습할 수 있는 장점이 있다.

참고문헌

1. 송문섭, 조신섭(2002). 엑셀에 기초한 통계학 입문, 자유아카데미, 서울.
2. 안봉근(2004), 사례중심의 통계학, 삼영사, 서울.
3. 윤상운, 이태섭(2000). Excel을 이용한 실용통계학 입문, 자유아카데미, 서울.
4. 조신섭, 송문섭, 이윤모, 성병찬, 윤영주, 이현부(1998). 기초통계교육을 위한

통계패키지의 비교 연구 및 엑셀을 이용한 한글 통계패키지의 구현.
 한국통계학회 춘계학술발표회 논문집, 75-79.

5. 조신섭, 송문섭, 이운모, 성병찬, 윤영주, 이현부(1999). 기초통계교육을 위한 통계 소프트웨어의 개발 -Excel에 기초한-. 품질경영학회지, 제27권 제2호, 277-290.
6. 최현석(2004a). The Development of Program for Teaching on Statistical Inference at One Population, *Journal of the Korean Data & Information Science Society*, Vol 15, No 3, 543-554.
7. 최현석(2004b). The Program for Teaching on Type I error and Type II error, *한국데이터정보과학회, 추계학술발표회논문집*, 17-23.
8. 최현석, 최성우, 김태운(2005). Computer Program Development for Two Populations Inference, *Journal of the Korean Data & Information Science Society*, Vol. 16, No 2. pp.185~193.
9. 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.
10. 한상태, 강현철(2002). EXCEL을 활용한 통계자료분석, 자유아카데미, 서울.
11. Jacobson, R.(2002). *Microsoft Excel 2002 Visual Basic Step by Step*. Redmond, W.A., Microsoft Press.
12. Walkenbach, J.(2004). *Excel 2003 Power Programming With VBA*, Wiley Publishing.
13. <http://home.kmu.ac.kr/~statexe>
14. <http://stats.snu.ac.kr/time/tslab.htm>
15. <http://www.xlstat.com>
16. <http://www.unistat.com>

[2005년 7월 접수, 2005년 8월 채택]