

Computer Program Development for Two Populations Inference

Hyun Seok Choi¹⁾ · Sung Woo Choi²⁾ · Tae Yoon Kim³⁾

Abstract

This study develops and introduces the program for letting learners study statistics in a systematic and efficient way by using Excel tools such as VBA and Macro, when they study statistical inference at two populations. This study helps learners understand the steps on statistical inference at two populations by utilizing the systematic and visual techniques.

Keywords : 양식도구, 엑셀 매크로, 추정과 검정, VBA

1. 서론

여러 분야에서 두 모집단의 데이터를 비교하여 분석해야 할 상황들이 실제로 많이 존재한다. 즉 두 모집단간의 유사성이나 차이점을 찾아내기 위해 두 모집단을 비교연구하기를 원한다. 두 모수의 통계적 추론은 두 모집단의 평균, 분산, 비율 등을 비교하여 보다 나은 방법을 찾고자 할 때 이용되는 것이다.

기존에 있는 대부분의 프로그램이 두 모집단의 추론 부분에서 대규모의 자료를 분석하기 위한 것이지 통계전공자가 아닌 학습자들이 사용하기에는 학습용이나 교육용으로는 미흡한 점이 있다. SAS나 SPSS등의 통계분석프로그램을 사용하지 않고 학교 뿐만 아니라 기업체에서 가장 보편적으로 사용되는 엑셀프로그램을 이용하여 학생들이 통계학의 기본개념을 쉽게 이해하고 습득할 수 있도록 엑셀을 사용하여 통계자료 분석을 다룬 문헌들이 대량 출판되고 있는 실정이다. 그러나 기존의 엑셀을 이용한

1) First Author : Lecturer, Department of Statistics, Keimyung University, Daegu, 704-701, Korea
E-Mail : chsuk1@kmu.ac.kr

2) Associate Professor, Department of Environmental Science & Engineering, Keimyung University, Daegu, 704-701, Korea
E-Mail : tswchoi@kmu.ac.kr

3) Professor, Department of Statistics, Keimyung University, Daegu, 704-701, Korea
E-Mail : tykim@kmu.ac.kr

통계학 교재들은 통계분석을 위한 엑셀 명령어를 습득하여 단순히 엑셀프로그램을 활용하는데 주안점을 두었다.

엑셀의 분석도구는 매우 제한적이어서 엑셀을 이용한 일반통계분석에는 한계가 있으나 엑셀 이용의 차원을 높일 수 있는 양식도구와 Visual Basic Application(VBA)에 기초한 매크로를 이용한 통계교육은 통계학의 개념과 이론을 이해하는데 많은 도움을 줄 수 있다.

기초통계교육에 엑셀을 이용한 예는 송문섭과 조신섭(1999)등은 엑셀을 이용하여 구현한 통계교육용 도구(KESS)를 개발하였고, 최현석은 한 모집단에서 통계적 추론에 관한 프로그램(2004a), 가설에 대한 설명, 제1종 오류와 제2종 오류에 대한 교육용 프로그램을 개발하였다(2004b). 엑셀 Add-in 프로그램인 KESS와 XLSTAT도 데이터 분석 및 통계솔루션의 일종으로 많이 사용되고 있다. 웹사이트 <http://www.unistat.com>에서도 엑셀을 이용한 다양한 분석방법을 제공하고 있으며 학교나 직장에서 자신의 학습과 업무에 활용할 수 있는 엑셀을 이용한 교육이 보편화되고 있다.

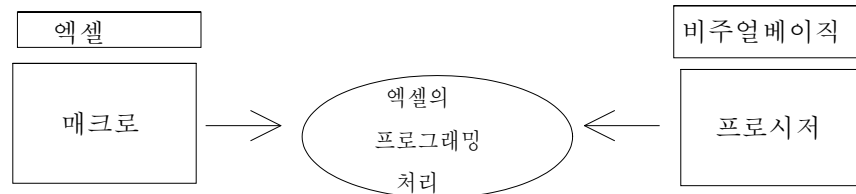
엑셀의 도구인 VBA, 매크로와 양식도구를 이용하여 두 모집단의 통계적 추론에 관한 과정을 단계적, 시각적으로 학습할 수 있는 프로그램을 개발하여 소개하고자한다.

2. 프로그램 개발

엑셀은 가장 일반적으로 사용되는 업무용 스프레드시트(Spread Sheet)프로그램이다. 자료를 기록, 계산, 통계처리 하는데 뛰어난 기능을 가지고 있을 뿐만 아니라 데이터베이스, 그래프, 자동계산 기능, 문서작성 기능, 차트작성 기능, 작업 자동화 기능을 프로그램으로 계산업무 전반에 이용할 수 있는 프로그램이다.

엑셀에서 기본적으로 제공되는 내장함수(Built-in function)를 이용하여 자료를 처리하거나 시각화할 수 있었다. 그러나 자료 분석을 위한 내장함수가 제공되지 않거나 여러 함수를 동시에 다루어야 하는 경우가 발생한다. 매크로 또는 VBA는 엑셀에서 이러한 것을 해결할 수 있도록 도와주는 하나의 프로그래밍언어이다.

매크로는 이용자의 작업 처리를 코드로 기록해 두었다가 나중에 이 코드로 작업을 자동으로 수행하기 위한 명령어로 엑셀매크로는 반복적으로 실행되는 명령어를 모아서 한 번에 실행하는 것 외에 비주얼베이직이라는 프로그래밍언어를 내장하여 단순한 명령어 나열이 아니라 명령어를 이용한 프로그래밍 기능이 추가되어 있다.



<그림 1> 엑셀 매크로와 비주얼베이직의 프로시저

양식도구는 Dialog Sheet에서 대화상자를 사용자가 직접 작성할 때 사용하는 것으

로 프로그램에서는 명령단추(CommandButton), 확인란(CheckBox), 옵션단추(OptionButton), 그룹상자(GroupBox)를 사용하였다.

VBA Project의 모듈 창에는 다음과 같이 코드를 작성하였다.

- ① Sub 문을 사용하여 일반프로시저로 작성하였다.
- ② 셀을 지정하여 입력받은 값을 수식을 이용하여 기록하는 프로시저를 작성하였다.
- ③ 엑셀 자체에서 제공되는 분석기능과 VBA 등으로 작성된 프로그램을 연결하였다.
- ④ 설명, 수식, 그래프 등은 Rectangle, Object, Group, Chartobjects 등으로 작성하여 ActiveSheet를 사용하여 활성화하였다.
- ⑤ 명령단추를 사용하여 단추를 누르면 바로 매크로가 실행되게 하였다.

3. KESS, XLSTAT와 비교

엑셀 Add-in 프로그램으로 많이 사용하는 것이 KESS와 XLSTAT 프로그램이다.

KESS(Korean Educational Statistics Software)는 기초통계교육의 보조수단으로 사용하기 위한 교육용 통계소프트웨어로 서울대학교 통계학과에서 제작한 엑셀 Add-in 프로그램이다. 엑셀을 구동 하면 새로운 메뉴(Kess)가 생성되어 대화상자를 활성화하여 자료를 분석할 수 있도록 하였다.

XLSTAT는 Thierry Fahmy(<http://www.xlstat.com>)에 의해 개발된 프로그램으로 마이크로소프트 엑셀을 기반으로 해서 운영되며 데이터분석 및 통계솔루션의 일종으로 엑셀 Add-in 프로그램이다. 엑셀을 구동 하면 새로운 메뉴와 툴바(Toolbar)가 생겨 메뉴나 툴바를 클릭하여 데이터분석기법에 대응하는 대화상자를 활성화 시켜 자료를 분석할 수 있다.

<그림 2>는 KESS로 두 집단 자료에 대하여 분석한 독립표본에 대한 두 모집단의 추론 결과이다. 자료입력창과 결과창이 다르고 기초적인 통계량과 결과만 주어지기 때문에 학습자가 내용을 이해하고 결과 분석을 하는데 어려움이 있을 수도 있다.

도시와(과) 지방의 독립표본에 의한 t-(양측)검정 결과			
	평균	표준편차	개수
도시	83	4	12
지방	81	4.8819	10
t-검정결과			
H0 : $\mu_1 = \mu_2$			
	통계량	t-통계량	자유도
		1.0445	20
		1.023	17.1835
			p-값
			0.3087
			0.3205
두 변수의 독립성 (양측)검정			
H0 : 두 변수의 분산은 같다.			
	F통계량	자유도	p-값
	1.5255	(5,11)	0.4822
두 변수의 모평균의 차에 대한 95% 신뢰구간			
	하한	상한	
	통계량	-1.9943	5.9943
	통계량	-2.0408	5.0408

<그림 2> KESS 분석 결과

<그림 3>은 XLSTAT로 독립표본에 대한 두 모평균의 차에 대한 검정결과이다. 자료입력과 결과가 한 화면상에 나타나지만, 모든 과정이 한꺼번에 나타나며 결과도 값만 제공하므로 통계학을 처음 접하는 학생이나 일반인들이 이해하기 곤란할 수도 있다.

	A	B	C	D	E	F	G	H	I	J
1	도시	지방								
2	82	75	References for values of group 1: Workbook = Book1 / Sheet = Sheet1 / Range = \$A\$2:\$A\$13							
3	81	81	References for values of group 2: Workbook = Book1 / Sheet = Sheet1 / Range = \$B\$2:\$B\$11							
4	80	81								
5	89	82	General statistics							
6	86	76								
7	83	80								
8	80	77			Freq.	Mean	Std. dev.			
9	74	87	1	12	89	4				
10	86	77	2	10	81	4.966816516				
11	86	83								
12	86		Results for Student's t test / Variances assumed to be equal / two-tailed test :							
13	83									
14			Value of the t statistic :							
15			1.0448936							
16										
17			p-value corresponding to Student's t: 0.3087							
18			Critical value for t with a significance level of 0.0500 : 2.0660							
19										
20			Conclusions :							
21			Using Student's t test with a significance level of 0.0500, one CAN accept the hypothesis that Mean1 - Mean2 = 0							
22										
23			Nevertheless, the alternate hypothesis is not valid.							

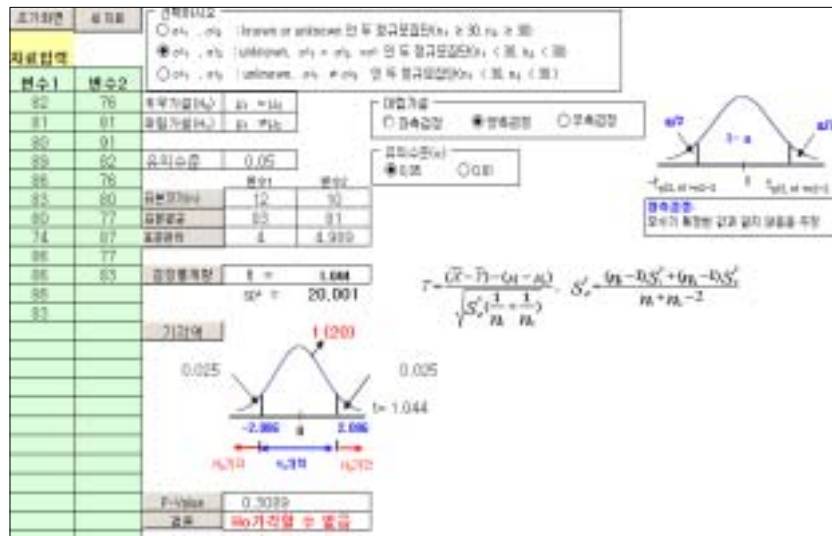
<그림 3> XLSTAT 분석결과

<그림 4>~<그림 6>은 KESS와 XLSTAT에서 분석한 자료를 본 프로그램으로 분석한 내용이다. <그림 4>는 모분산 σ_1^2 과 σ_2^2 이 알려져 있지 않은($\sigma_1^2 = \sigma_2^2 = \sigma^2$) 두 정규모집단에 대하여 분석한 결과이다. 옵션단추로 분포에 맞는 모집단, 대립가설, 유의수준을 선택하고, 단계적으로 명령단추를 클릭할 때마다 해당하는 통계량, 수식, 그림이 동시에 나타나 이론적인 내용을 효과적으로 학습하는데 도움을 줄 수 있다.

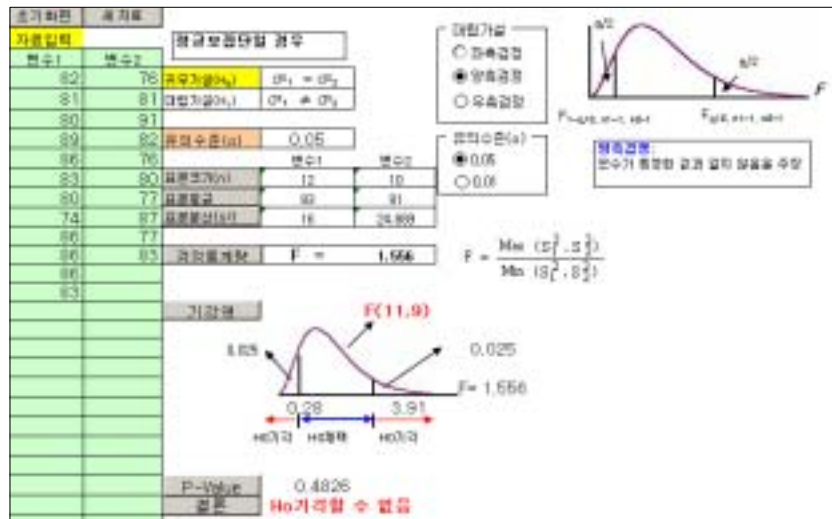
<그림 5>는 두 모분산의 비에 대한 검정결과이다. KESS로 분석한 결과는 F값과 p값의 결론만 주어졌지만 본 프로그램으로 분석한 결과는 자료입력과 결론이 동일화면상에 나타나고 단계적 절차에 의한 자료처리를 수식, 그림과 함께 제공하여 분석과정을 이해하도록 하였다. <그림 6>은 두 모평균의 차에 대한 신뢰구간을 본 프로그램으로 분석한 결과화면이다. KESS는 하나의 신뢰구간에 대한 결과만 제공하지만 본 프로그램은 해당 신뢰구간의 확인란을 클릭하면 90%, 95%, 99%에 대한 신뢰구간이 수식, 그림, 결과가 함께 나타나 통계학을 처음 접하는 학생이나 비전공자들도 신뢰구

간의 분석과정을 쉽고 정확하게 이해할 수 있다.

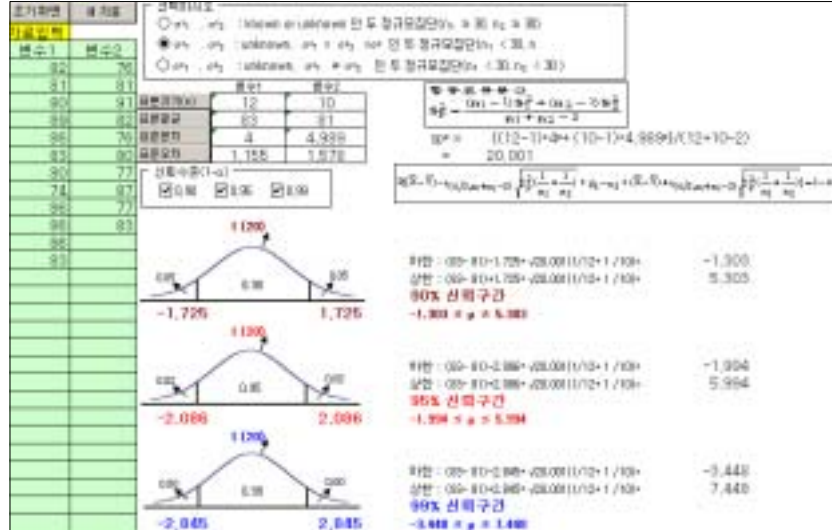
엑셀 Add-in 프로그램인 KESS와 XLSTAT 이외의 다른 통계패키지들은 통계학 전반에 걸쳐서 많은 분석을 제공하여 대규모의 자료분석에 대한 결과만 제공할 뿐이지만, 본 프로그램은 두 집단 비교를 할 때 다른 통계패키지에는 없는 단계적 절차, 수식, 그림 등을 함께 제공하여 통계적 추론에 대한 기본원리와 분석과정을 터득하게 하는 학습용 또는 교육용 프로그램이다.



<그림 4> 두 모평균들의 차에 대한 검정결과



<그림 5> 두 모분산들의 비에 대한 검정 결과



<그림 6> 두 모평균 차에 대한 신뢰구간 분석결과

4. 프로그램의 특징과 구성

본 프로그램은 대규모의 데이터를 입력하여 결과만 제공하는 다른 패키지와는 달리 통계적 추론에 관한 과정을 단계적이고 시각적인 기법을 잘 활용하여 학습자들의 이해를 돕는 프로그램이다.

본 프로그램은 추정과 가설검정에 대하여 <표 1>의 경우에 분석할 수 있다.

<표 1> 분석범위

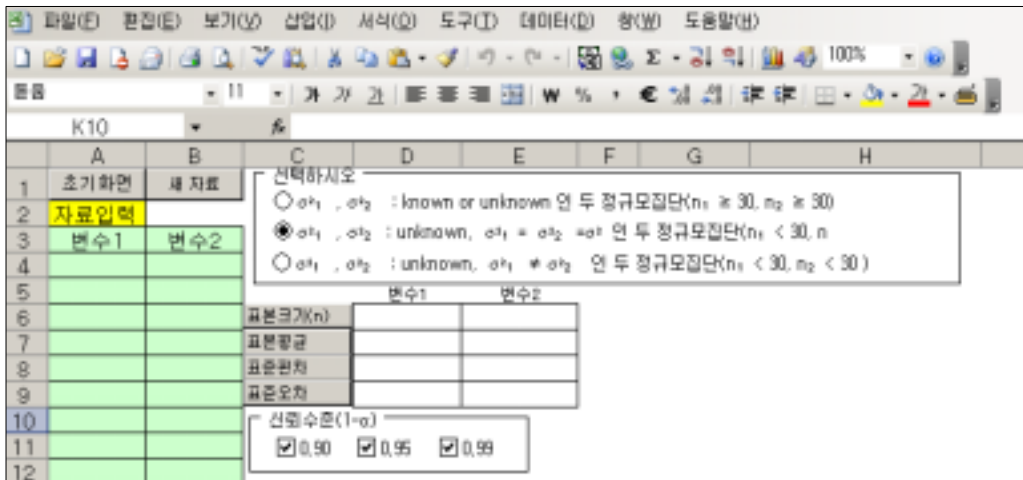
통계적 추론	조건
두 모평균들의 차 $\mu_1 - \mu_2$	① 모분산 σ_1^2 과 σ_2^2 이 알려져 있는 두 정규모집단 ② 모분산 σ_1^2 과 σ_2^2 이 알려져 있지 않은 ($\sigma_1^2 = \sigma_2^2 = \sigma^2$) 두 정규모집단 ③ 모분산 σ_1^2 과 σ_2^2 이 알려져 있지 않은 ($\sigma_1^2 \neq \sigma_2^2$) 두 정규모집단
짝지워진 표본을 이용한 모평균 $\mu_D = \mu_1 - \mu_2$	모분산 σ_D^2 이 알려져 있지 않은 정규모집단
두 모분산들의 비 σ_1^2 / σ_2^2	두 정규모집단
두 모비율들의 차 $p_1 - p_2$	표본크기 n_1 과 n_2 이 충분히 큰 경우

본 프로그램의 초기화면은 <그림 7>과 같다. 프로그램을 실행하면 초기화면이 나타난다. 분석하고자 하는 단원의 단추를 누르면 해당단원으로 넘어가고, 각 단원에서 초기화면 단추를 누르면 아래의 초기화면이 나타나 자료를 입력하거나 표본크기(n), 표본평균(\bar{X}), 표본분산(S^2) 등의 통계량이 주어진 경우에 대하여 분석할 수 있다.

두 모집단의 추정과 가설검정	
두 모평균의 차 $\mu_1 - \mu_2$ 의 추정 (독립인 표본)	두 모평균의 차 $\mu_1 - \mu_2$ 의 가설검정 (독립인 표본)
두 모평균의 차 $\mu_1 - \mu_2$ 의 추정 (대응비교)	두 모평균의 차 $\mu_1 - \mu_2$ 의 가설검정 (대응비교)
두 모비율의 차 $p_1 - p_2$ 의 추정	두 모비율의 차 $p_1 - p_2$ 의 가설검정
두 모분산의 비 σ_1^2 / σ_2^2 의 추정	두 모분산의 비 σ_1^2 / σ_2^2 의 가설검정

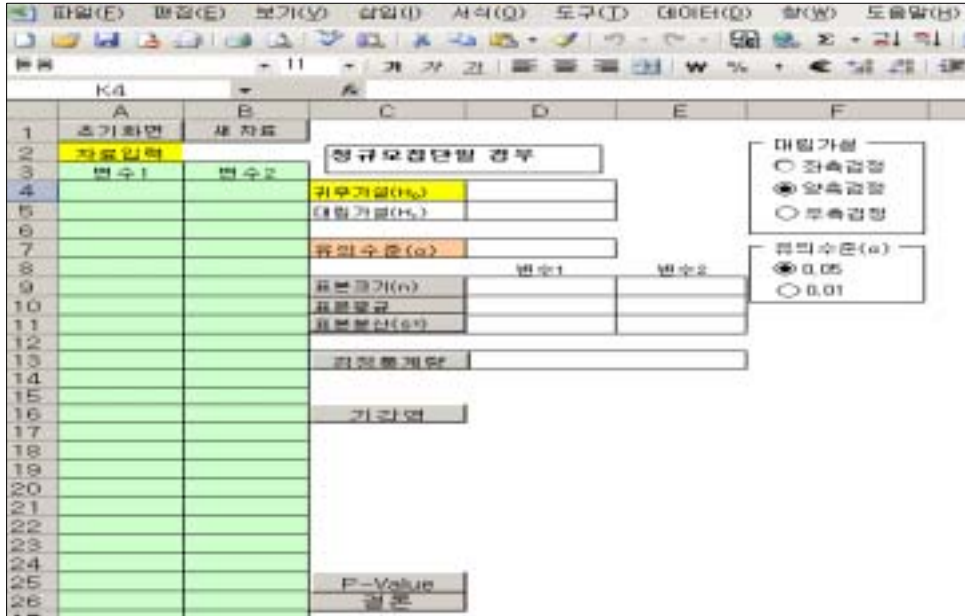
<그림 7> 초기화면

초기화면에서 두 모평균의 차의 추정을 클릭하면 <그림 8>이 나타나서 자료입력 후 옵션단추의 선택과 확인란의 클릭만으로 독립인 두 확률표본들의 차에 구간추정을 실행할 수 있다. 자료를 입력하여 실행한 결과가 <그림 6>에 나타나 있다. 단추를 누르면 다음단계로 넘어가는 학습자가 단지 바라보는 것이 아닌 직접 실행하는 학습이 되어 학습의욕을 고취시킬 수 있다.



<그림 8> 두 모평균들의 차에 대한 추정의 초기화면

<그림 9>는 두 정규모집단을 대상으로 두 모분산들의 비에 대한 검정의 초기화면 이고 자료를 입력하여 단계적인 과정을 거쳐서 나온 결과가 <그림 5>이다. 가설의 설정부분에서는 대립가설부분을 옵션단추로 처리하여 좌측, 양측, 우측 중 하나를 택 하면 분포형태와 함께 주어진다. 유의수준, 검정통계량, 기각역, p값, 결론 단추를 클릭하여 단계적으로 결론에 이르므로 쉽게 개념을 습득하고 분석과정을 이해할 수 있다.



<그림 9> 두 모분산들의 비에 대한 검정의 초기화면

본 프로그램의 특징은 다음과 같다.

학습 또는 교육용 프로그램이므로 효율적인 교육을 위하여 보다 손쉽게 접근할 수 있는 엑셀 프로그램으로, 분석화면과 출력화면이 동일화면상에 나타나고, 분석과정을 자료입력에서 결론에 이르기까지 단추를 누르면 한 단계씩 계산이 되도록 하는 방법을 택하였다.

- ① 사회 각 분야에서 가장 널리 활용되고 있는 엑셀을 이용한 프로그램으로 어떤 특수한 언어를 배울 필요가 없으며 사용하기가 편리하다.
- ② 추정과 가설검정을 위주로 하였기 때문에 다른 프로그램에 비하여 두 모집단의 특성을 분석하는데 체계적이고 효율적인 분석결과를 제공한다.
- ③ 동일 화면상에서 자료입력과 결과가 나타나고, 수식, 그림과 함께 단계적으로 결과를 확인할 수 있게 하여 통계분석의 기본원리와 분석과정을 스스로 이해하도록 하였다.
- ④ GUI(Graphic User Interface)환경으로 단추를 누름으로서 진행이 되므로 직접 실행하는 학습이 되도록 하였다.

4. 결론

본 연구에서는 엑셀매크로를 사용하여 두 집단에서의 통계적 추론을 학습할 때 스스로 체계적이고 효율적인 통계학습을 할 수 있는 엑셀 매크로 프로그램을 개발하여

분석내용 및 방법을 소개하였다. 다른 통계패키지로도 분석이 가능하나 대부분 모든 통계학 단원에 대하여 대규모의 자료를 분석하여 결론을 제공하는데 반해, 본 프로그램은 두 모집단의 추론에 대한 단계적 절차에 의한 자료처리를 수식, 그림과 함께 제공하여 엑셀 프로그램만 있으면 학습할 수 있는 장점이 있다. 즉 결과가 나오는 과정을 GUI(Graphic User Interface)환경으로 버튼을 누름으로써 진행이 되는 능동적으로 직접 실행하는 학습이 되도록 하였다. 본 프로그램은 <http://home.kmu.ac.kr/~statexe>의 자료실에서 다운받아서 사용할 수 있다. 엑셀 매크로로 프로그램을 작성하면 고급 통계기법들을 사용할 수 있으므로 통계적 추론을 비롯한 여러 단원에서도 분석프로그램을 개발하여 누구라도 쉽게 통계분석을 이해하고 활용할 수 있게 하여야 할 것이다.

참고문헌

1. 송문섭, 조신섭 (1999). *엑셀에 기초한 통계학 입문*, 자유아카데미, 서울.
2. 조신섭, 송문섭, 이윤모, 성병찬, 윤영주, 이현부 (1998). 기초통계교육을 위한 통계패키지의 비교 연구 및 엑셀을 이용한 한글 통계패키지의 구현. *한국통계학회 춘계학술발표회 논문집*, 75-79.
3. 조신섭, 송문섭, 이윤모, 성병찬, 윤영주, 이현부 (1999). 기초통계교육을 위한 통계소프트웨어의 개발 -Excel에 기초한-. *품질경영학회지*, 제27권 제2호, 277-290.
4. 최현석 (2004a). The Development of Program for Teaching on Statistical Inference at One Population, *Journal of the Korean Data & Information Science Society*, vol 15, 543-554.
5. 최현석 (2004b). The Program for Teaching on Type I error and Type II error, *한국데이터정보과학회 추계학술발표회 논문집*, 17-23.
6. Jacobson, R. (1997). *Microsoft Excel 97 Visual Basic Step by Step*. Redmond, W.A.: Microsoft Press.
7. <http://www.xlstat.com>
8. <http://stats.snu.ac.kr/time>
9. <http://www.unistat.com>
10. <http://home.kmu.ac.kr/~statexe>

[2004년 11월 접수, 2005년 3월 채택]