

표본조사 실습을 위한 통계 소프트웨어의 개발

Development of Statistics Software for Sampling Learning

이범석

해천대학 컴퓨터통신계열

<요약>

표본조사 교육의 효과적인 학습을 위하여 컴퓨터를 이용한 다양한 표본조사 기법들과 이에 따른 추정이론들을 비교분석하는 소프트웨어를 개발하고자 한다. 또한 일반적인 표본조사 실습에는 시간과 비용이 많이 들게 되므로 충분한 실습을 하기 어렵다. 따라서 본 연구에서는 이러한 문제점을 극복하기 위하여 사용자에게 사용이 편리한 멀티미디어 툴과 멀티미디어 툴에서의 한계인 수학적 계산은 통계 패키지를 이용하여 보다 사용하기 쉽고 실시간에 모수추정이 가능한 표본조사 실습 프로그램을 개발하데 목적으로 하였다.

1. 서론

컴퓨터의 발전은 교육분야에 많이 활용되고 강좌의 성격에 따라 컴퓨터를 이용한 이론교육과 실습교육이 병행되어 진행될 때 더욱 효과적이다. 표본조사 교육을 실시하는 데에 컴퓨터의 도움 없이는 시간과 비용이 많이 들게되므로 충분한 실습이 어려운 실정이다. 또한 컴퓨터를 이용한다고 하여도 단순히 연산이나 통계패키지의 이용으로는 사용자의 지식 한계로 실습에 크게 도움이 되지 않는다. 표본조사 강의에서 실습을 통한 이해가 이루어지지 못하고 이론강의에만 치우치면 흥미를 유발시키지도 못하고 실제자료처리도 못할 우려가 있다.

최근 컴퓨터를 이용한 통계교육에 많은 연구가 고조되어 안기수(1996)는 회귀분석을 위한 학습시스템으로 HyperRClass를 개발하였고, 정남철(1996)은 인터넷에서 통계 학습시스템을 연구하였고, 이범석(1997)은 표본조사 전문가 시스템으로 SLES를 개발하는 등 여러분야에서 개발되고 있는 현실이다. 그러나, 기존에 사용되고 있는 대부분의 통계관련 소프트웨어들은 통계학도들을 위한 통계교육과 전문가들을 위한 것으로

산업공학등 통계학도이외의 학생과 일반인을 위한 소프트웨어와 통계학의 가장 기본이 되는 표본론에 관한 교육용 프로그램의 개발은 유용하다고 여겨진다.

본 논문에서는 표본조사에 관한 통계학 전공자, 산업공학 전공자, 사회과학 전공자를 위한 한글 소프트웨어를 개발하여 소개하고자 한다. 본 소프트웨어는 어프-라인상 태에서 한글로 개발된 것으로 누구나 쉽게 사용할 수 있다. 2절에서는 소프트웨어의 개요를 소개하고, 3절에서는 소프트웨어의 주요내용으로 표본추출법, 추정 등을 다루고 있다.

2. 소프트웨어의 개요

본 연구에서 개발된 소프트웨어는 표본조사에 관한 지식이 없는 사용자에게 능동적으로 표본조사 이론을 학습하고 실습하도록 한 것이다. 본 소프트웨어는 표본추출방법과 난수발생법으로 난수에의한 모집단에서 표본추출, 모추추정으로 이루어져 있으며 사용자는 제시된데이터를 가지고 모집단을 구성할 수도 있고, 표본추출법에 따라

다른 추출법을 사용하여 추정값을 비교할 수도 있다.

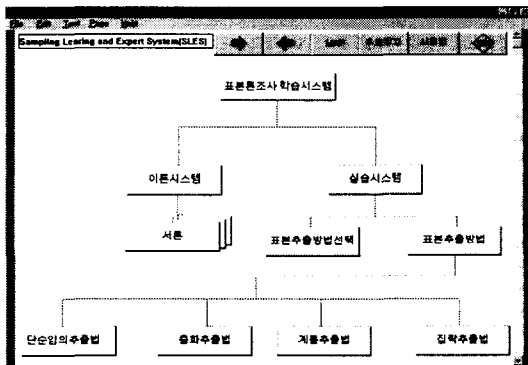
본 소프트웨어는 틀북을 이용하여 사용자 인터페이스를 쉽게하였고 통계계산을 위한 통계 패키지는 패키지의 용량이나 저작권의 문제등을 고려하여 Xlisp-Stat를 사용하였다. 멀티미디어 틀인 틀북과 통계패키지의 동적자료교환이 원활한 교육용 소프트웨어의 개발에

중점을 두었으며 본 시스템의 운영을 위하여는 시스템용량이 7MB이상이며 펜티엄 166MHZ이상인 PC기종 이면 된다.

3. 소프트웨어의 주요구성내용

3.1 시스템의 구조

본 소프트웨어는 두분야로 나누어 구성되어 있으며 기본구조는 <그림1>과 같이 이론시스템과 실습시스템으로 나누었다. 이론시스템은 표본조사의 학습과정과 표본조사에서 필요로 하는 단계등을 다루었다. 실습시스템에서는 난수발생, 표본추출법 선정등을 다루며, 모수추정을 위하여는 멀티미디어틀과 통계패키지사이의 동적자료교환을 이용하여 실시간에 자료를 계산토록 하였다. 모든 구성은 버튼, 콤보박스(combo-box)와 하이퍼텍스트로 구성되어 엔터키와 마우스 클릭으로 모든 분야의 접근이 가능토록 하였다



<그림1>초기화면

3.2 표본의 크기결정

표본설계에서 조사결과의 정도를 미리 정하여야 하는데 정도는 표본의 크기에 의해서 결정된다. 표본크기 결정은 모평균, 모집단총합, 모비율 추정을 위한 표본크기 결정 방법으로 세 가지가 있는데 이중 첫째 모평균 추정을 위한 표본크기 결정을 대표로 설명하다.

표본의 크기를 무조건 크게 하였다고 정도가 높아지는 것이 아니므로 모집단의 크기 N 과 표본평균의 분산 σ^2 그리고 오차한계 B 를 사전에 확인하고 표본의 크기를 결정해야 한다.

즉, 모평균추정을 위한 표본크기

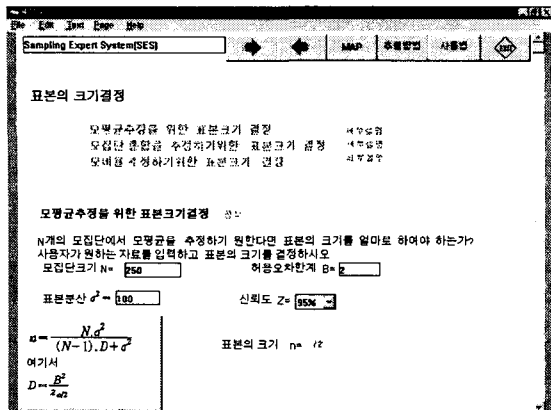
$$n = \frac{N \cdot \sigma^2}{(N-1) \cdot D + \sigma^2}$$

$$D = \frac{B^2}{(z_{\alpha/2})^2}$$

에 의해서 표본의 크기가 결정된다. 이는 각화면에 “공식”으로 하이퍼텍스트로 처리하여 보여주며, 이상의 식을 프로그래밍 하면, 사용자 인터페이스가 원활해지고 신뢰도의 선택된 값에 따라서 표본크기가 변화하는 것을 <그림2>에서 볼 수 있다. 화면의 상단 중앙에 있는 “세부 설명”은 하이퍼텍스트로 구성하여 표본크기 결정방법의 보충설명과 학습이 필요할 때 클릭하여 관련된 자료를 참조할 수 있다.

예를 들어

- 모집단의 크기: $N = 250$
- 표본평균의 분산: $\sigma^2 = 100$
- 허용 오차의 한계: $B = 2$
- 신뢰도: 95%



<그림2> 단순임의추출법의
표본크기결정

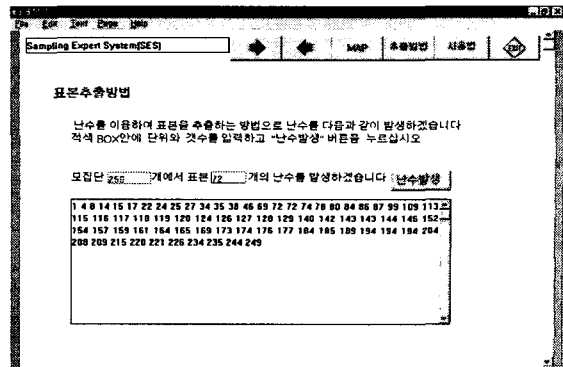
일 때 표본의 크기 $n=72$ 이다.

즉, 250개의 모집단에서 72개의 표본의 크기를 추출하면 95%의 신뢰도에서 오차한계 2이내의 모평균을 추정할 수 있다.

3.3 표본추출법

표본의 크기를 결정한 후에 표본을 추출하기 위하여 난수표를 이용하거나 난수를 발생시켜야 된다. <그림2>에서 표본의 크기가 결정되면 다음 화면으로 이동하는 아이콘을 클릭하여 <그림3>에서 난수발생 화면으로 가면 모집단의 크기와 표본의 크기는 자동적으로 이동되어 온다. 만약 사용자가 크기가 다른 모집단 크기와 표본 크기를 원한다면 모집단의 크기와 표본 크기를 입력하여도 된다.

이 화면에서 “난수발생”버튼을 클릭하면 난수는 0보다 크고 모집단의 크기 250보다 작은 72개의 난수를 발생한다. 이상의 난수를 기준으로 표본을 추출하여 DOS 또는 WINDOWS 응용 프로그램을 이용하여 파일을 만들거나 모수추정을 위한 화면의 입력필드에 자료를 입력하여 모수를 추정할 수 있다.

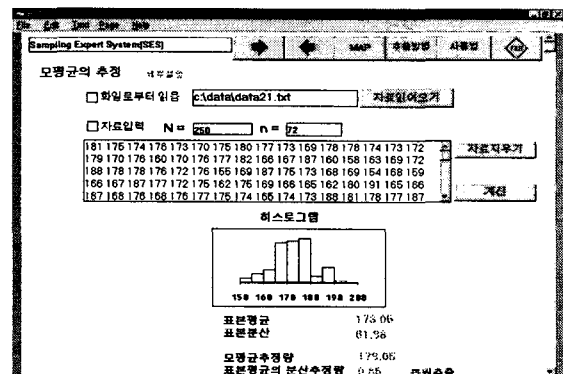


<그림3> 표본의 크기 결정후 난수발생 화면

3.4 모평균추정

<그림4>는 모평균 추정을 위한 화면으로 파일을 만들었을 경우 “파일로부터 읽음”이란 필드에 파일이 보관되어 있는 곳을 화면같이 “C:\data\data21.txt” 입력하고 “자료읽어오기” 버튼을 클릭하여 자료를 입력 필드에 채운다. 만약 파일을 만들지 않고 직접 입력하고자 할 때는 “자료지우기” 버튼을 클릭하여 입력 필드를 지운 후 자료를 입력한다.

모든 자료를 입력 후 “계산” 버튼을 클릭하면 표본평균, 표본분산, 모평균 추정량, 표본평균의 분산추정량이 화면에 표본조사 자료의 히스토그램과 함께 적색 글씨로 출력된다. 또한 신뢰도에 따른 신뢰구간을 계산하기 위하여 콤보박스에서 적절한 신뢰도 90%, 95%, 99%를 선택하면 모평균 추정에 대한 신뢰구간을 계산하여 출력할 수 있다.

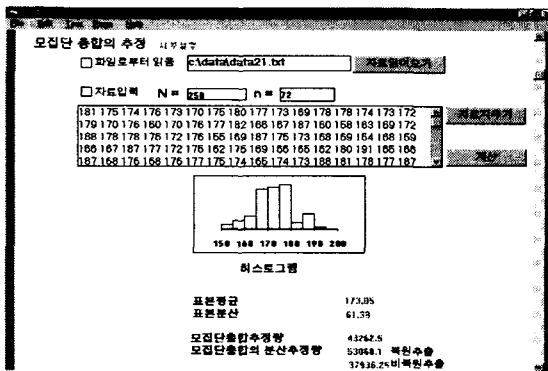


<그림4> 모평균 추정 화면

$$\hat{p} = \sum_{i=1}^n \frac{y_i}{n} = \bar{y}$$

3.5 모집단총합 추정

<그림5>는 모집단 총합 추정을 위한 화면으로, 사용자가 다른 응용 프로그램으로 작성된 파일을 사용할 경우 읽음 입력 필드에 파일의 path와 이름을 적으면 파일에서 자료를 읽고 통계 소프트웨어에서 계산되어 사용자에게 제공한다.



<그림5> 모집단총합추정 화면

3.6 모비율추정

<그림6>은 모비율의 추정에 관한 것이다. 모집단에서 특정한 속성을 갖는 경우의 추정 값으로 특정한 속성을 가질 경우는 1, 속성을 갖고있지 않을 경우는 0으로 하여 모수를 추정한다.

즉,

$y_i = 0$: i 번째 원소가 특정한 속성을 갖고 있지 않는 경우

$y_i = 1$: i 번째 원소가 특정한 속성을 갖고 있는 경우

본 화면에서 신뢰도에 대한 모비율의 신뢰구간은 여타 화면에서와 같이 콤보박스에서 신뢰도를 선택하면 된다.

모비율 추정량

표본비율의 분산 추정량

$$\hat{V}(\hat{p}) = \frac{\hat{p}\hat{q}}{n} \left(\frac{N-n}{N} \right)$$

여기서 $\hat{q} = 1 - \hat{p}$

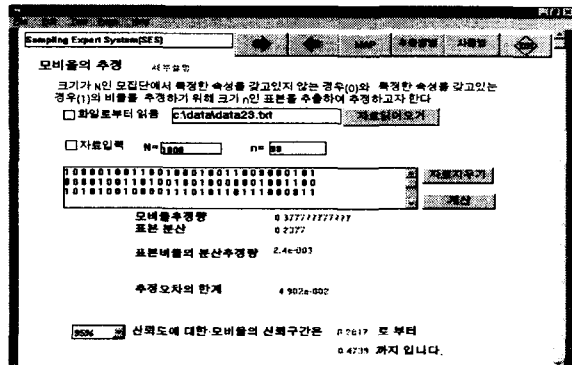
추정오차의 한계

$$z_{\alpha/2} \sqrt{\hat{V}(\hat{p})} = z_{\alpha/2} \sqrt{\frac{\hat{p}\hat{q}}{(n-1)} \left(\frac{N-n}{N} \right)}$$

신뢰구간

$$\hat{p} \pm z_{\alpha/2} \sqrt{\frac{\hat{p}\hat{q}}{n-1} \frac{N-n}{N}}$$

로 구성되어 졌다.



<그림6> 모비율추정화면

4. 결론

각종 학습을 위한 소프트웨어가 많이 개발되고 통계 교육을 위한 소프트웨어의 국내 개발도 활발하게 추진되고 있는 새로운 상황이 전개되고 있다. 이에 비해볼 때 통계학의 중요부분중 하나인 표본조사의 실습은 중요함에도 불구하고 이에 적절한 실습 소프트웨어 개발이 미비한 상태이다.

이 소프트웨어 개발의 의미는 표본조사의 개념을 효과적으로 이해할 수 있으며 실제 자료로 실습을 할 수 있으므로 강의실의 강의와 실습이 현장에 즉시 투입할 수 있다는

장점이 있다. 향후 연구방향으로는 인터넷으로 해당 데이터베이스와의 연결을 시도하여 보다 효율적으로 교육과 실습 및 실 현장에서 사용하도록 하는 것이다. 또한 통계학의 다른분야에서도 활용하며 음성, 동영상 등의 멀티미디어를 포함시키므로 교육의 효율을 높이는데 있다.

참고문헌

- [1] 남궁 평(1997), “현대표본이론“, 탐진.
- [2] 이범석(1997), “하이퍼미디어를 이용한 표본조사 전문가시스템의 구현“, 박사 학위논문, 성균관대학교, 1997.
- [3] 이범석, 남궁평, 변종석(1998), “표본조사교육 및 실습을 위한 하이퍼미지어 전문가시스템“, 한국통계학회 논문지, 제5권 제3호, pp.659-673
- [4] 허문열(1995), “XLISP-STAT : 객체지향통계언어“, 자유아카데미.
- [5] Asymetrix(1994), “Multimedia Toolbook user manual,” Aysmetrix Corp.
- [6] Berger, R. G., Ajmani, A. K.(1993), “Multimedia Case Authoring and Simulation System,” In Proceedings of the 17th Annual Symposium in Medical Care, Washington D. C.
- [7] Chua, Tat-Seng(1991), “Issues in Hypermedia Research Multimedia Technology and Application,” Ellis Horwood Limeted.
- [8] Emily B. and Joseph D.(1991), “Hypertext / Hypermedia Handbook”, McGraw Hill.
- [9] Horn, R. E.(1989), “Mapping Hypertext,” A Publication of The Lexington Institute.
- [10] Leslie, Kish(1995), “Survey

Sampling,” John Wiley & Sons, INC.

[11] Nigel, Woodhead(1990), “Hypertext and Hypermedia : Theory and Applications,” Sigma Press.

[12] MCP Project (University of North Carolina at Wilmington), <http://smec.uncwil.edu/mcp/>

[13] Technology College ; Multimedia Browser Test, <http://www.uncwil.edu/mmtest/>

[14] XperRule: Attar Software product summary, <http://www.attar.com/pages/products.htm>